INSTITUT POLYTECHNIQUE DE PARIS

Habilitation à diriger des recherches

Contributions to Mathematical Morphology

Habilitation à diriger des recherches de l'Institut Polytechnique de Paris
Spécialité de doctorat : Informatique, Données, et IA

HDR présentée et soutenue à Paris, le 28 Septembre 2022 , par

SANTIAGO VELASCO-FORERO

Composition du Jury :

Isabelle BLOCH Professeur, Sorbonne Université	Présidente
Pierre SOILLE Project Leader, HDR, Joint Research Centre of the European Commission, Italie	Rapporteur
Yann GOUSSEAU Professeur, Télécom ParisTech	Rapporteur
Nicolas PASSAT Professeur, Université de Reims Champagne-Ardenne	Rapporteur
Jean-Michel MOREL Professeur, École Normale Supérieure, Paris-Saclay	Examinateur
Stéphane MALLAT Professeur, Collège de France	Examinateur

Pour Lolout, Emiliout et Mamout.

Acknowledgements

I would like to thank all the people who throughout my life have supported me and helped me overcome difficult times. In this list, which is long, I must start with those who have always motivated me and believed in me: my family. Thanks to my super grandma, my parents and siblings for their support afar. My nieces who with their smiles motivate me to be a better uncle every day. And of course, my wife and my children, who with their love and affection make me get the best of me. I mention my wife again, who has influenced me positively by her example, always does my part, without thinking about what others do or say.

The development of this document could not have been completed without the support of my CMM colleagues, my current and former Ph.D. and Master students. In addition, many of the applications of the methods developed have only been possible thanks to the hard work of my academic and industrial collaborators.

In closing, I thank the members of the jury, for their review and comments, and for considering this manuscript to have merit for the award of an HDR.

Contents

List of Symbols

Symbol	Meaning
Ω	Spatial Domain of points.
$\mathbf{f}(x) \in \mathcal{F}(\Omega, \mathbb{V}))$	Signal of the domain of the form $\mathbf{f} : \Omega \to \mathbb{V})$.
$\psi \in \mathcal{F}(\Omega, \mathbb{V}) \to \mathcal{F}(\Omega, \mathbb{V})$	A mapping on signals.
n, p	Number of pixels and Dimension of the vector space.
$x \in \Omega$	A pixel in the original image.
$\mathbf{x} \in \mathbb{R}^p$	A vector representation of $\mathbf{x} = \mathbf{f}(x)$
$\mathtt{x}$	A coordinate in the vector space, $\mathbf{x} = [\mathtt{x}_1, \ldots, \mathtt{x}_p]$
$\mathbf{X} \in Mat_{n,p}(\mathbb{R})$	A matrix representation of a signal, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$
$\mathbb{V}$	Vector space for spectral information.
$\mathcal{L}$	A lattice
$\mathtt{SE}$	Structuring Element.
$\mathtt{B}$	Ball centred at x.
$\delta_{\mathtt{SE}}(\cdot), \varepsilon_{\mathtt{SE}}(\cdot)$	Dilation, Erosion operator.
$\gamma(\cdot), \varphi(\cdot)$	Opening, Closing operator
$\tau(\cdot)$	Toggle mapping operator
$\mathtt{HMAX}(\cdot)$	H-max transformation
$\mathtt{RMAX}(\cdot)$	Regional maxima transformation
$\mathtt{REC}(\cdot, \cdot))$	Morphological Reconstruction
$\mathcal{M}$	Manifold.
$\mathtt{Ret}_\theta$	Retraction from the tangent space of θ onto $\mathcal{M}$
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$	Graph with nodes in $\mathcal{V}$ and edges in $\mathcal{E}$
$\mathcal{MST}$	Minimum Spanning Tree
$\mathcal{FOREST}$	Forest
$\mathcal{H}$	Hierarchy
$\Lambda : \mathcal{H} \to \mathbb{R}^+$	Ultrametric distance
$\mathbf{W}$	Matrix of Weights
θ	Parameter
$\mathtt{loss}$	loss function
L	Number of layers
$\mathtt{Fix}$	Set of Fixed Points
$\mathtt{BA}$	Attraction Basin
$\mathbf{Jac}$	Jacobian Matrix

Introduction

This manuscript is intended to obtain the diploma of “Habilitation à Diriger des Recherches", issued by the Institut Polytechnique de Paris, in the “Informatique, Données, et IA" discipline. It describes research activities pursued after obtaining a Doctoral thesis in Mathematical Morphology, in June 2012. These activities essentially relate to mathematical morphology approaches for image analysis, multivariate analysis for anomaly detection and machine learning methods, from a theoretical and methodological point of view, but also in the context of applications, for remote sensing, astronomy, material science among others.

0.1 Context

My research career began at the end of my master's degree with the participation in various projects at the University of Puerto Rico, continuing with my Ph.D. thesis in the *Center for Mathematical Morphology* at Mines Paris - PSL, which I defended in June 2012. Followed by one year as postdoctoral researcher at the *Fraunhofer-Institut für Techno-und Wirtschaftsmathematik* - ITWM in Kaiserslautern, Germany, and 13 months as a postdoctoral fellow at the Department of Mathematics in the University of Singapore. In both cases, my responsibilities include participation in industrial projects that allowed theoretical contributions to be confronted with real problems. The second part of my research experience (which continues until now) begins upon my return to France in October 2014, as a three years tenure track followed by a permanent research position at the Center for Mathematical Morphology of Mines Paris - PSL Research University. This context, rich in both academic developments and industrial collaborations, has motivated me to apply my skills in a wide range of applications including astronomy, remote sensing, biology, radar technology and material science among others. Additionally, the daily interaction with other researchers, master's and doctoral thesis students, has enriched my knowledge and understanding of various aspects of image processing, multivariate statistics, and machine learning. This document attempts to collect the results obtained in the last ten years of research.

0.2 Structure

The topics presented here mainly study the interaction between mathematical morphology, discrete mathematics, multivariate statistics and machine learning. The first chapter studies the extension of mathematical morphology for multivariate images by means of lattice theory and

machine learning. The second chapter is devoted to adaptive structuring elements in general and includes as a practical example the study of a morphological salt and pepper noise reduction method via conditional morphology. Chapter three includes some contributions of hierarchical analysis of images. Chapter four presents some links between mathematical morphology and machine learning. Chapter five presents the problem of anomaly detection from a statistical point of view, and some extension using machine learning methods. Chapter six is a compendium of applications. The appendices include an exhaustive list of teaching experiences, list of collaborators, publications and industrial projects.

0.3 A personal point of view about Research

In this *Habilitation à diriger des recherches* (HDR) includes most of the results of my last eleven years of interaction with a magnificent group of scientists including colleagues, former Ph.D. and master's students. The readers should note that the my works can be considered as application-inspired, since most of the contributions are motivated by a real need that has been explored during industrial and academical collaborations. Additionally, I would like to include some statement to summarise what I think about the *Research*. I consider that

- "Science is not a business" and as education should be a human right, science should be common for all.
- a dependence on single numbers to quantify scientists contribution and make administrative decisions is a bad habit, and it may force people to somehow enhance their quality indexes (as for instance the h-index) instead of focusing on their more legitimate activity.
- *Research* is a collective adventure, and by sharing ideas, new solutions emerges.
- while results contained in this work have been mainly the effort of this HDR's author, it could not have been concluded without the collaboration of the corresponding co-authors.

1 Mathematical Morphology on Vector Images

> An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem.
>
> John Tukey

1.1 Introduction

From the creation of the first photograph in 1826[1], to the invention of the first devices able to produce a digital image in the 70th's, a long time passed. Nowadays, images are everywhere, and it has been estimated that more than 1.2 trillion digital photos have been taken worldwide in 2017[2]. Automatic analysis of large set of image is the main focus of many digital technology companies. The development of tools that allow the processing of digital images is the main focus of this manuscript. We start this chapter with the presentation of the notation and the interest for a professional image analyst[3] of having a description by a lattice of a digital image.

1.1.1 Notations

In this document, an *image* is defined as a function $\mathbf{f}$ which associates a vector information to each pixel, we use the notation $\mathbf{f} : \Omega \to \mathbb{V}$, where Ω is the support of pixel values, usually $\mathbb{Z}^2$ or $\mathbb{Z}^3$ for digital images, and $\mathbb{V}$ denotes the vector space of spectral information. For instance, for 2D colour images, $\Omega = \mathbb{Z}^2$ and $\mathbb{V} = \mathbb{R}^3$, *i.e.* each pixel $x = (i, j)$, $\mathbf{f}(x)$ is a triplet of values $[\mathsf{x}_1, \mathsf{x}_2, \mathsf{x}_3]$. An illustration of the used notation is given in Figure 1.1.

Additionally, we write $\mathcal{F}(\Omega, \mathbb{V})$ for the set of functions $\Omega \to \mathbb{V}$. For grey scale images, $\mathbf{f}, \mathbf{g} \in \mathcal{F}(\Omega, \mathbb{R})$ is naturally endowed with a partial order relation $\leq$ defined by setting $\mathbf{f} \leq \mathbf{g}$ for two functions $\mathbf{f}$ and $\mathbf{g}$, if and only if for every pixel $x \in \Omega$ we have $\mathbf{f}(x) \leq \mathbf{g}(x)$. For any two functions $\mathbf{f}, \mathbf{g} \in \mathcal{F}(\Omega, \mathbb{R} \cup \{-\infty, \infty\})$, we write $\mathbf{f} \vee \mathbf{g}$ and $\mathbf{f} \wedge \mathbf{g}$ for the *join* and the *meet*, *i.e.*, $\forall x \in \Omega \ (\mathbf{f} \vee \mathbf{g})(x) = \max(\mathbf{f}(x), \mathbf{g}(x))$, and, $\forall x \in \Omega \ (\mathbf{f} \wedge \mathbf{g})(x) = \min(\mathbf{f}(x), \mathbf{g}(x))$

[1] *View from the Window at Le Gras* is a heliographic image and the oldest surviving camera photograph according to (259)

[2] The estimation has been taking from https://www.infotrends.com/

[3] The term *professional image analyst* is used in this document to refer to the person who needs to use imaging methods in practice for a given application.

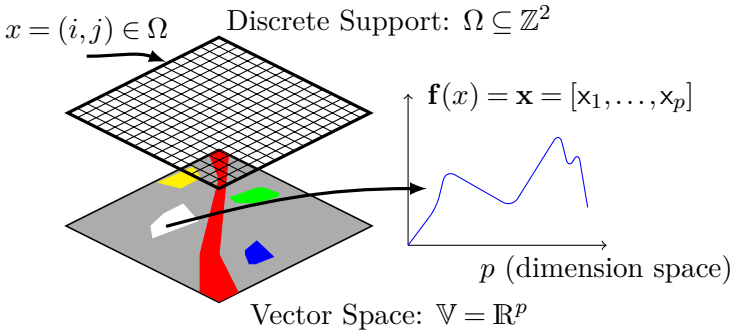
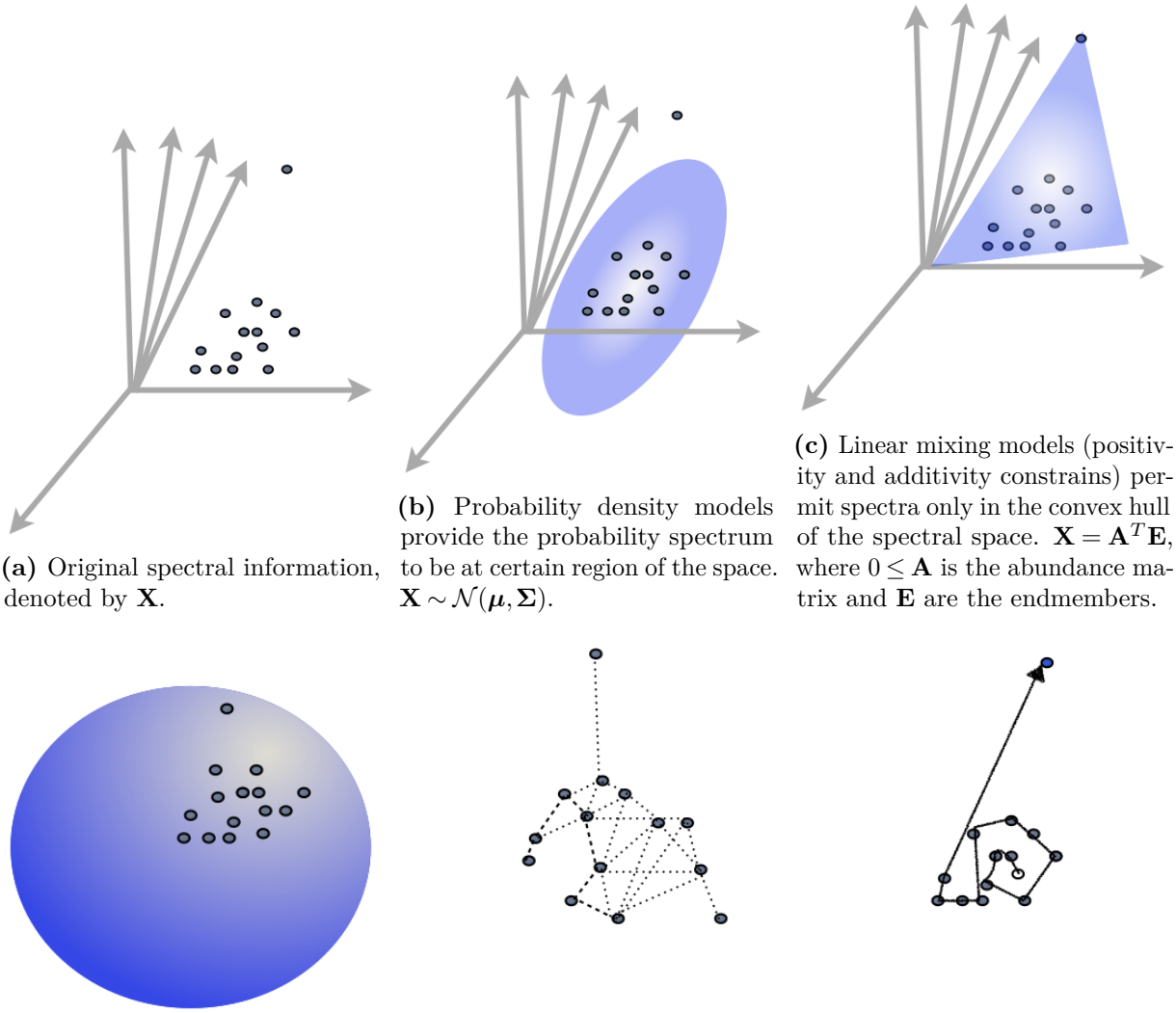


Figure 1.1 Mathematical notation for a 2D multivariate image, $\mathbf{f} : \Omega \to \mathbb{V}$. The matrix of dimension $n_1 \times n_2 \times p$ containing the vector information on $n := n_1 \times n_2$ pixels in dimension p of $\mathbf{f}$ is denoted by $\mathbf{X}_\mathbf{f}$.

Depending on the application, one may be interested in enriching the information in an image, going beyond colour information. This is what multispectral sensors do. They collect information at different values of wavelength ranges across the electromagnetic spectrum, including additional information the human eye fails to capture given that its receptors operate only in the visible light spectrum.

A variety of representations can be used to characterise the spectral information contained in an image, for instance:

- Probability density models. It is the classical assumption of the statistical modelling for a multivariate image. The classical assumption is a Multivariate Gaussian Distribution, *i.e.*, the probability of a spectrum to be at certain region of the space follows $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for some mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.
- Linear mixing models: In such a way that each pixel in a scene may be decomposed into a sum of finite number of constituent endmembers, which represent the purest pixels in the scene. The abundances (weights in the sum) are subject to non-negativity constraint and, in some cases, sum-to-one constraint.
- Manifold models: The idea of capturing the complex geometry in the spectral representation of an image is the core of a (non-Euclidean) manifold representation (212). Manifold learning methods are becoming a standard to embedding data onto their new transformed spaces. For instance, on the $p-1$ hypersphere by normalising by the norm per pixel, or in supervised projections on the sphere as in (153). Another manifold structure in a local-graph. It is denoted by the graph $\mathcal{G} = (\mathcal{E}, \mathcal{V})$, where $\mathcal{E} = \mathbf{X}$ is the set of vertices and $\mathcal{V} = [v_{ij}] \in \mathbb{R}^n \times \mathbb{R}^n$ the edge weight matrix. The graph is constructed in an unsupervised manner, with a goal of automatically determining the neighborhood structure as well as the corresponding connection weight for each datum. Examples of this type of representation are k-graph, ϵ-graph and ℓ^1-graph (299). Clustering, dimensionality reduction, image segmentation and analysis can be performed in this representation space.
- Complete lattice model: A total ordering definition for a cloud of points. The relation $x \leq y$ have to be know for all pair of pixels x and y of the multivariate image. That is the



(a) Original spectral information, denoted by $\mathbf{X}$.

(b) Probability density models provide the probability spectrum to be at certain region of the space. $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

(c) Linear mixing models (positivity and additivity constrains) permit spectra only in the convex hull of the spectral space. $\mathbf{X} = \mathbf{A}^T\mathbf{E}$, where $0 \leq \mathbf{A}$ is the abundance matrix and $\mathbf{E}$ are the endmembers.

(d) Manifold models, for instance, representations on the sphere, allow a representation invariant to intensity changes (norm). $\mathbf{x}_i \rightarrow \mathbf{x}_i/||\mathbf{x}_i||$, for all $i = 1 \ldots n$.

(e) Graph models provides a structure representation invariant to small perturbations in the spectral space. $\mathbf{X} \rightarrow (V, E)$.

(f) In the complete lattice representation, the set of pixels are analysed through a total order relation. $\mathbf{X} \rightarrow \mathcal{L}$.

Figure 1.2 Scheme of different representation for the spectral information of a multivariate image. This thesis deals with spectral representation based on complete lattice representation as in (f).

main goal in Chapter one. Mathematical morphology requires this kind of representation to ensure the appropriate application of lattice based transformations.

Fig. 1.2 shows a scheme of the some representations for the spectral information of a given multivariate image.

1.1.2 Mathematical morphology

Mathematical morphology is an approach created in the 1960s by George Matheron and Jean Serra, which offers a theory for nonlinear image analysis. Mathematical morphology has been used by many professional image analysts due to its sound theoretical basis and the fact that the

operators have a natural geometric interpretation, which facilitates the design and interpretation of automatic methods. Mathematical morphology (197; 246; 251), is part of the image analysis techniques that consider images as a *topographical surface*(48).

Mathematical morphology, developed originally for sets, can be applied to numerical functions either by means of umbras, or directly via the complete lattice approach.

1.1.3 Mathematical morphology on complete lattices

In this HDR, two family of mathematical morphological transformations are considered: 1) *connection based morphology* and 2) *adjunction based morphology*. The first strategy deals with simplification of a given image in the partition space induced by its connected components (227; 237; 246). The second perspective analyses an image by composition of two basic transformations, dilation and erosion, which form a *Galois connection* (108). Another approach studying the links between mathematical morphology and Partial Differential Equations (PDEs) (102; 162) is not considered in this manuscript. Other representation of morphological operators in max-plus algebras studied in (35) are neither considered in this document.

In this section we provide the theoretical background of mathematical morphology in its formulation based on complete lattices. For a more detailed exposition, we refer to Chapter two by J. Serra and C. Ronse in (197).

Definition 1.1. A set $\mathcal{L}$ with a binary relation $\leq$ that satisfies the following properties:

1. reflexivity: $x \leq x$
2. anti-symmetry: $x \leq y$ and $y \leq x \Rightarrow x = y$
3. transitivity: $x \leq y$ and $y \leq z \Rightarrow x \leq z$

for all $x, y, z \in \mathcal{L}$, is called a *partially ordered set* (poset) by the *partial order* $\leq$. The order becomes *total* when: $\forall x, y \in \mathcal{L}, x \leq y$ or $y \leq x$ $\mathcal{L}$ is totally ordered by the relation $\leq$.

Definition 1.2. A poset $(\mathcal{L}, \leq)$ is called a *lattice* is each two-element $\{r, s\}$ subset in $\mathcal{L}$ has a least upper bound *join* $(r \vee s)$ and a greatest lower bound *meet* $(r \wedge s)$.

Note that the previous definitions makes $\wedge$ and $\vee$ binary operations monotone with respect to the given order $\leq$. Additionally, by induction argument every non-empty finite subset in $\mathcal{L}$ has a join and a meet. The lattice $\mathcal{L}$ is *complete* if all non-empty subsets finite or not of $\mathcal{L}$ has a join and a meet.

A *minimum (or least)* $\perp \in \mathcal{L}'$ is an element which is less than or equal to any other element of $\mathcal{L}'$, that is, $r \in \mathcal{L}' \Rightarrow \perp \leq r$. We denote the minimum of $\mathcal{L}$ by $\perp$. Equivalently, a *maximum (largest)* $\top$ in $\mathcal{L}'$ is the greatest element of $\mathcal{L}'$, that is, $r \in \mathcal{L}' \Rightarrow r \leq \top$. We denote the maximum of $\mathcal{L}$ by $\top$.

Definition 1.3 (*Dilation/Erosion*)**.** A mapping $\psi : \mathcal{L}_1 \to \mathcal{L}_2$ of a complete lattice $\mathcal{L}_1$ into a complete lattice $\mathcal{L}_2$ is said to be a *dilation* if $\psi(\bigvee_{j \in J} r_j) = \bigvee_{j \in J} \psi(r_j)$ for all families $(r_j)_{j \in J}$ of elements in $\mathcal{L}_1$. A mapping is said to be an *erosion* if $\psi(\bigwedge_{j \in J} r_j) = \bigwedge_{j \in J} \psi(r_j)$ for all families $(r_j)_{j \in J}$ of elements in $\mathcal{L}_1$.

The important relationship between dilation and erosion is that they are dual concepts from the lattice point of view. Heijmans and Ronse (108) showed that for any complete lattice $\mathcal{L}$, we always have a dual isomorphism between the complete lattice of dilation on $\mathcal{L}$ and the complete lattice of erosions on $\mathcal{L}$. This dual isomorphism is called by (247) (Chapter 1) the *morphological duality*. In fact it is linked to the concept of *Galois connection*s in lattice theory, as we will see at the end of this section.

Definition 1.4 (*Adjunction*)**.** Let $\delta \in \mathcal{F}(\mathcal{L}_1, \mathcal{L}_2)$ and $\varepsilon \in \mathcal{F}(\mathcal{L}_2, \mathcal{L}_1)$ be two mappings and $(\mathcal{L}_1, \leq_1), (\mathcal{L}_2, \leq_2)$ two complete lattices. Then we say that (ε, δ) is an *adjunction* if for every $r \in \mathcal{L}_1, s \in \mathcal{L}_2$, we have

$$\delta(r) \leq_2 s \iff r \leq_1 \varepsilon(s) \tag{1.1}$$

In an adjunction (ε, δ), ε is called the *upper adjoint* and δ the *lower adjoint*.

Proposition 1.5. *If (ε, δ) is an adjunction, then δ is a dilation and ε is an erosion.*

Definition 1.6 (*Galois connection*)**.** Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be lattices and let $\alpha : \mathcal{L}_1 \to \mathcal{L}_2$ and $\beta : \mathcal{L}_2 \to \mathcal{L}_1$ satisfy the following conditions.

1. For $r, s \in \mathcal{L}_1$, if $r \leq s$, then $\alpha(s) \leq \alpha(r)$;
2. For $r, s \in \mathcal{L}_2$, if $r \leq s$, then $\beta(s) \leq \beta(r)$;
3. For $r \in \mathcal{L}_1$, $\beta\alpha(r) \leq r$;
4. For $r \in \mathcal{L}_2$, $\alpha\beta(r) \leq r$.

Then (α, β) is a *Galois connection* between $\mathcal{L}_1$ and $\mathcal{L}_2$.

Proposition 1.7. *Let the lattices $\mathcal{L}_1$ and $\mathcal{L}_2$, maps $\alpha : \mathcal{L}_1 \to \mathcal{L}_2$ and $\beta : \mathcal{L}_2 \to \mathcal{L}_1$ a Galois connection. Then the following condition holds for all $r \in \mathcal{L}_1$ and $s \in \mathcal{L}_2$:*

$$s \leq_2 \alpha(r) \iff r \leq_1 \beta(s). \tag{1.2}$$

Clearly an adjunction in $\mathcal{L}$ is a Galois connection between the dual $(\mathcal{L}_1, \geq_1)$ and $(\mathcal{L}_2, \geq_2)$ (indeed, compare definition 1.4 and proposition 1.7).

At this point, we can see that definition of erosion/dilation on a image requires a complete lattice structure, *i.e.*, a total ordering[4] among the pixels to be analysed.

Accordingly, the extension of mathematical morphology to vector spaces, for instance, colour/multi/hyper/ultraspectral images, is neither direct nor trivial because the pixels in the images are vectors. We refer keen readers to (8; 14) for a comprehensive review of vector valued mathematical morphology.

1.1.4 Preorder by h-function

Let $\mathcal{S}$ be a nonempty set and assume that $\mathcal{L}$ is a complete lattice. Let $h : \mathcal{S} \to \mathcal{L}$ be a surjective mapping in $\mathcal{F}(\mathcal{S}, \mathcal{L})$. Define an equivalence relation $=_h$ on $\mathcal{S}$ as follows: $r =_h s \Leftrightarrow h(r) =$

[4]Theoretically, a partial ordering is enough but to make easier the presentation we analyse the case of total ordering.

$h(s) \quad \forall r, s \in \mathcal{S}$. As it was defined in (97), we refer by $\leq_h$ the *h-ordering* given by the following relation in $\mathcal{S}$

$$\forall r, s \in \mathcal{S}, \quad r \leq_h s \Leftrightarrow h(r) \leq h(s)$$

Note that $\leq_h$ preserves reflexivity ($r \leq_h r$) and transitivity ($r_1 \leq_h r_2$ and $r_2 \leq_h r_3 \Rightarrow r_1 \leq_h r_3$). Even so, $\leq_h$ is not a partial ordering because $r \leq_h s$ and $s \leq_h r$ implies only that $r =_h s$ but not $r = s$. Note that h-ordering is a preorder in $\mathcal{S}$.

An operator $\psi : \mathcal{S} \to \mathcal{S}$ is *h-increasing* if $r \leq_h s$ implies that $\psi(r) \leq_h \psi(s)$. Additionally, since h is surjective, an equivalence class is defined by $\mathcal{S}[r] = \{s \in \mathcal{S} | h(s) = r\}$, where $r \in \mathcal{L}$. The Axiom of Choice (97) implies that there exist mappings $h^{\leftarrow} : \mathcal{L} \to \mathcal{S}$ such that $hh^{\leftarrow}(r) = r$, for $r \in \mathcal{L}$. Unless h is injective, there exist more than one such $h^{\leftarrow}$ mappings: $h^{\leftarrow}$ is called the semi-inverse of h. Note that $h^{\leftarrow}h$ is not the identity mapping in general (but $h^{\leftarrow}h =_h \mathrm{id}$). However, we have that for any *h-increasing* $\psi : \mathcal{S} \to \mathcal{S}$ the result $\psi h^{\leftarrow} h =_h \psi$ and hence $h\psi h^{\leftarrow} h = h\psi$. Let us introduce $\widetilde{\psi}$ the operator associated to ψ in the lattice $\mathcal{L}$. A mapping $\psi : \mathcal{S} \to \mathcal{S}$ is *h-increasing* if and only if there exists an increasing mapping $\widetilde{\psi} : \mathcal{L} \to \mathcal{L}$ such that $\widetilde{\psi} h = h\psi$. The mapping $\widetilde{\psi}$ is uniquely determined by ψ and can be computed from

$$\widetilde{\psi} = h\psi h^{\leftarrow}$$

We can now define the *h-erosion* and *h-dilation*. Let $\varepsilon, \delta : \mathcal{S} \to \mathcal{S}$ be two mappings with the property

$$\delta(x) \leq_h y \Leftrightarrow x \leq_h \varepsilon(y), \quad \forall x, y \in \mathcal{S}$$

then the pair (ε, δ) is called an *h-adjunction*. Moreover, let (ε, δ) be h-increasing mappings on $\mathcal{S}$, and let $\varepsilon \to^h \widetilde{\varepsilon}$, $\delta \to^h \widetilde{\delta}$. Then (ε, δ) is an h-adjunction on $\mathcal{S}$ if and only if $(\widetilde{\varepsilon}, \widetilde{\delta})$ is an adjunction on the lattice $\mathcal{L}$. Therefore a mapping δ (resp. ε) on $\mathcal{S}$ is called *h-dilation* (resp. *h-erosion*) if $\widetilde{\delta}$ (resp. $\widetilde{\varepsilon}$) is a dilation (resp. erosion) on $\mathcal{L}$. h-adjunctions inherit a large number of properties from ordinary adjunctions between complete lattices. Assume that (ε, δ) is an h-adjunction then

$$\gamma = \delta\varepsilon \leq_h \mathrm{id} \leq_h \varphi = \varepsilon\delta.$$

Hence, γ is h-anti-extensive and ϕ is h-extensive. The operator γ on $\mathcal{S}$ is called *h-opening* if the operator $\widetilde{\gamma}$ on $\mathcal{L}$ determined by $\gamma \to^h \widetilde{\gamma}$ is an opening. The operator γ is also h-increasing and satisfies $\gamma\gamma =_h \gamma$ (h-idempotency). The *h-closing* is similarly defined.

1.1.5 Morphological analysis on the h-function

For multiband imagery, as colour, multispectal, hyperspectral or ultraspectral images, pixel values are vectors defined in $\mathbb{V} := \mathbb{R}^p$, for an integer $p > 1$ the dimension of the vector space of the image. Consequently the main challenge to build complete lattice structures is to define a mapping $h : \mathbb{R}^p \to \mathcal{L}$, where $\mathcal{L}$ can be the lattice of the extended real line $(\overline{\mathbb{R}}, \leq)$ using $\overline{\mathbb{R}} = \mathbb{R} \bigcup \{-\infty, +\infty\}$ and $\leq$ as the "less than or equal to" operation.

From the previous section we have the ingredients to define morphological colour ($\mathbb{V} = \mathbb{R}^3$) and multispectral ($\mathbb{V} = \mathbb{R}^p$) erosion and dilation. We limit here our developments to the flat

operators, *i.e.*, the structuring elements are planar shapes. The non-planar structuring functions are defined by weighting values on their support (246). Let us assume that we have an adaptive[5] mapping $h : \mathbb{R}^p \to \mathbb{R}$. The *h-erosion* $\varepsilon_{\mathtt{SE},h}(\mathbf{f})$ and *h-dilation* $\delta_{\mathtt{SE},h}(\mathbf{f})$ of an image $\mathbf{f}$ at pixel $x \in \Omega$ by the structuring element $\mathtt{SE} \subset \Omega$ are the two mappings $\mathcal{F}(\Omega, \mathbb{V}) \to \mathcal{F}(\Omega, \mathbb{V})$ defined respectively by

$$h\left(\varepsilon_{\mathtt{SE},h}(\mathbf{f})(x)\right) = \varepsilon_{\mathtt{SE}}\left(h(\mathbf{f})\right)(x), \tag{1.3}$$

and

$$h\left(\delta_{\mathtt{SE},h}(\mathbf{f})(x)\right) = \delta_{\mathtt{SE}}\left(h(\mathbf{f})\right)(x), \tag{1.4}$$

where $\varepsilon_{\mathtt{SE}}(\mathbf{f})$ and $\delta_{\mathtt{SE}}(\mathbf{f})$ are the standard numerical flat erosion and dilation of image $\mathbf{f} \in \mathcal{F}(\Omega, \mathcal{L})$:

$$\varepsilon_{\mathtt{SE}}(\mathbf{f})(x) = \left\{\mathbf{f}(y) : \mathbf{f}(y) = \bigwedge \left[\mathbf{f}(z)\right], z \in \mathtt{SE}_x\right\} \tag{1.5}$$

$$\delta_{\mathtt{SE}}(\mathbf{f})(x) = \left\{\mathbf{f}(y) : \mathbf{f}(y) = \bigvee \left[I(z)\right], z \in \hat{\mathtt{SE}}_x\right\} \tag{1.6}$$

with $\mathtt{SE}_x$ being the structuring element centred at point x and $\hat{\mathtt{SE}}$ is the reflected structuring element. If the inverse mapping h^{-1} is defined, the h-erosion and dilation can be explicitly written as:

$$\varepsilon_{\mathtt{SE},h}(\mathbf{f})(x) = h^{-1}\left(\varepsilon_{\mathtt{SE}}\left(h(\mathbf{f})\right)\right)(x),$$

and

$$\delta_{\mathtt{SE},h}(\mathbf{f})(x) = h^{-1}\left(\delta_{\mathtt{SE}}\left(h(\mathbf{f})\right)\right)(x).$$

Of course, the inverse h^{-1} only exists if h is injective. In practice, we can impose the invertibility of h by considering a lexicographic ordering for equivalence class $\mathcal{L}[\mathbf{x}]$. In fact, this solution involves a structure of total ordering which allows to compute directly the h-erosion and dilation without using the inverse mapping, *i.e.*,

$$\varepsilon_{\mathtt{SE},h}(\mathbf{f})(x) = \left\{\mathbf{f}(y) : \mathbf{f}(y) = \bigwedge_h \left[\mathbf{f}(z)\right], z \in \mathtt{SE}_x\right\}, \tag{1.7}$$

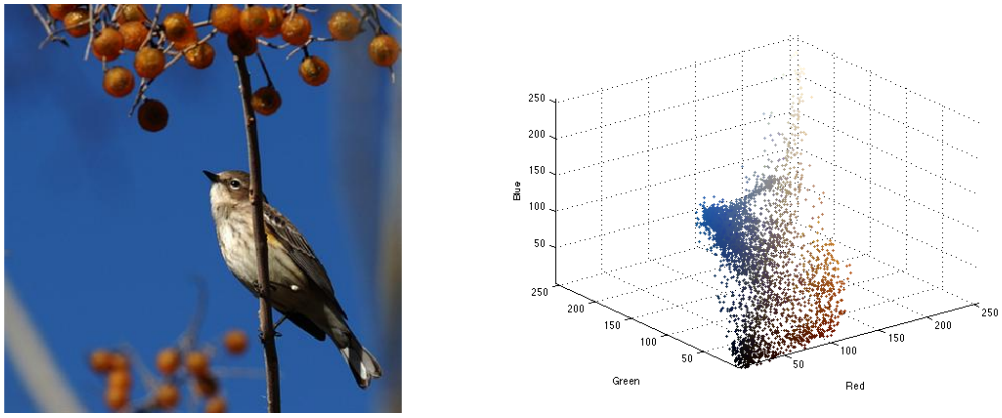
and

$$\delta_{\mathtt{SE},h}(\mathbf{f})(x) = \left\{\mathbf{f}(y) : \mathbf{f}(y) = \bigvee_h \left[\mathbf{f}(z)\right], z \in \hat{\mathtt{SE}}_x\right\}, \tag{1.8}$$

where $\bigwedge_h$ and $\bigvee_h$ are respectively the infimum and supremum according to the ordering $\leq_h$. Starting from the h-adjunction $(\varepsilon_{\mathtt{SE},h}(\mathbf{f}), \delta_{\mathtt{SE},h}(\mathbf{f}))$, all the morphological filters such as the opening and closing have their h-counterpart, *e.g.*, the h opening and closing are defined as

$$\gamma_{\mathtt{SE},h}(\mathbf{f}) = \delta_{\mathtt{SE},h}(\varepsilon_{\mathtt{SE},h}(\mathbf{f})), \quad \varphi_{\mathtt{SE},h}(\mathbf{f}) = \varepsilon_{\mathtt{SE},h}(\delta_{\mathtt{SE},h}(\mathbf{f})) \tag{1.9}$$

[5] Adaptive in the sense that the mapping depend on the information contained in a multivariate image $\mathbf{f}$. The correct notation should be $h(\cdot;\mathbf{f})$. However, in order to make easier the understanding of the section we use h for adaptive mapping.



(a) Original colour image denoted by $\mathbf{f}$ **(b)** Scatterplot of the three-channel image $\mathbf{X_f}$

Figure 1.3 Spectral representation of a colour image in the RGB space. A spatial position x in the image $\mathbf{f}$ contains three coordinates in the RGB-space represented by $\mathbf{x}$.

Similarly, any other mathematical morphology operator based on adjunction operators can be also extended to multivariate images. For instance, geodesic operators as opening by reconstruction(251), levelings (176), additive morphological decompositions (278) and so on.

1.2 Pre-ordering a vector space

Let $\mathbf{X_f}$ be the set of vector values of a given image $\mathbf{f}$, which can be viewed as a cloud of points in $\mathbb{V} = \mathbb{R}^p$. Fig. 1.3 shows an example of colour image $\mathbf{f}$, and its spectral representation as points $\mathbf{X_f}$. In general, pixel values in multispectral images are vectors defined in $\mathbb{R}^p$. From previous section, for a given multivariate image $\mathbf{f} : \Omega \to \mathbb{R}^p$, the challenge to build complete lattice structures is to define a mapping $h : \mathbb{R}^p \to \mathcal{L}$, to obtain a mapping $\Omega \to \mathcal{L}$, where $\mathcal{L}$ is a lattice.

Many authors have already worked on this idea (8; 14; 19; 277). We present three families of mappings h for a given $\mathbf{x} \in \mathbb{R}^p$ in the following subsections.

1.2.1 Unsupervised ordering

That can be obtained by using the more representative projection in a statistical dimensional reduction technique, for example a linear approach as PCA (132) or some non-linear projections approach (149). To illustrate, we consider the first projection to induce the ordering, *i.e.*, $\mathbf{x}_1 \leq \mathbf{x}_2 \iff h_{\texttt{PCA}}(\mathbf{x}) \leq h_{\texttt{PCA}}(\mathbf{x}_2)$, where $h_{\texttt{PCA}}$ is the first eigenvector of the centred covariance matrix $\mathbf{X_f}$. The intuition behind this approach is simple and clear: pixels are ordered according to their representation in the projection with greatest variance. An example is illustrated in Fig. 1.4(b). In this example, we can see that the induced minimum and maximum have no practical interpretation.

1.2.2 Distance based ordering

Let us focus on the case of h-ordering based on distances. This approach is motivated by the intuition that order computation should be adaptive to prior information given by application interests.

Referenced ordering

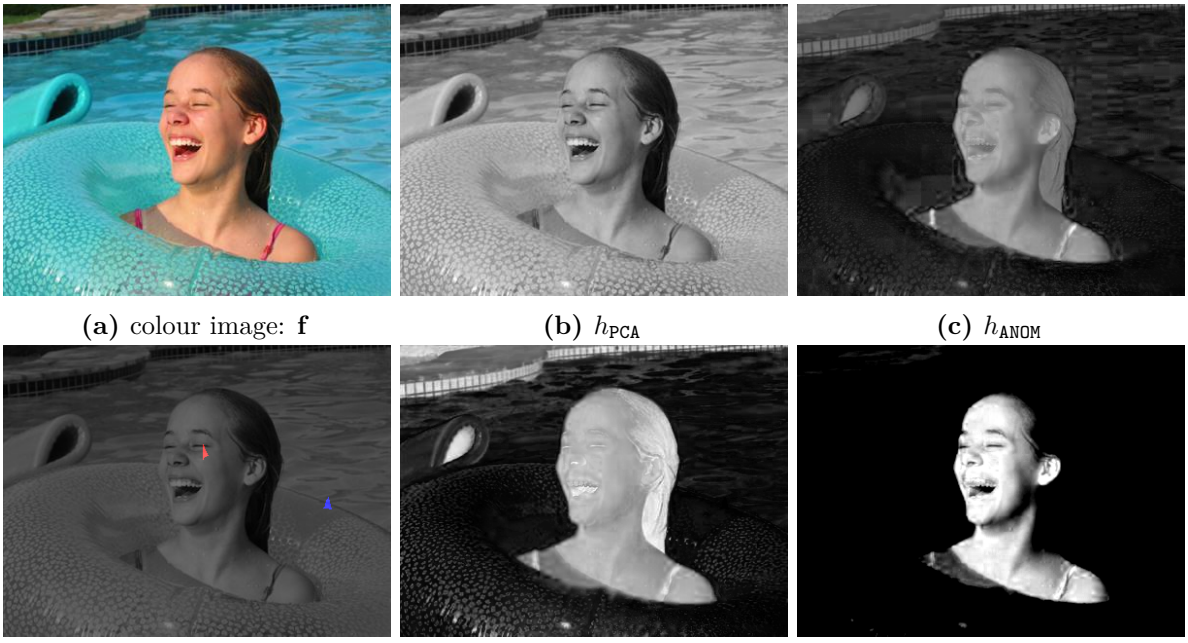
As a starting point for distance based ordering, we consider (8) defining a function $h_{\texttt{REF}}(\cdot, \mathbf{t})$ that computes the similarity for a given pixel $\mathbf{x}$ to a colour reference $\mathbf{t}$ by measuring its spectral distance, *i.e.*, $\mathbf{x}_1 \leq_{h_{\texttt{REF}}} \mathbf{x}_2 \iff \mathbf{K}(\mathbf{x}_1, \mathbf{t}) \leq \mathbf{K}(\mathbf{x}_2, \mathbf{t})$, where $\mathbf{K} : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^+$ is a kernel-induced distance(190). The original formulation in (8) uses the case of Euclidean distance in the colour space as kernel-induced distance[6]. Thus, the ordering based on a reference spectrum exhibits a lattice where the minimum has been fixed. However, that maximum is associated with the "farthest" vector but that does not have a simple interpretation. To illustrate the result of this approach, we generalise the definition of a referenced order for a training set T as follows, $\mathbf{x}_1 \leq_{h_{\texttt{REF}}} \mathbf{x}_2 \iff \min_i ||\mathbf{x}_1 - \mathbf{t}_i|| \geq \min_i ||\mathbf{x}_2 - \mathbf{t}_i||$ for all $\mathbf{t}_i \in T$. The geometric interpretation is that $h_{\texttt{REF}}(\mathbf{x}; T)$ is basically the distance in L_∞ of $\mathbf{x}$ to the convex hull of vectors in T (if $\mathbf{x}$ is not in the convex hull). Thus, is not so difficult to see that $h_{\texttt{REF}}$ can be expressed as $h_{\texttt{REF}}(\mathbf{x}; T) = \sum_{i=1}^{|T|} \theta_{\mathbf{x}}^i \mathbf{K}(\mathbf{t}_i, \mathbf{x})$ where $\theta_{\mathbf{x}}^i \neq 0$ only for $\arg\min_i ||\mathbf{x} - \mathbf{t}_i||$. Fig. 1.4(e) shows the referenced mapping for the colour image in Fig. 1.4(a). The training set are the pixel in the red region of Fig 1.4(d). Note that $h_{\texttt{REF}}$ "detects" the girl but at the same time the border of the swimming-pool. Associated morphological adjunction and gradient are illustrated in Fig. 1.5(g-i).

Supervised ordering

A more general formulation for distance based ordering has been introduced in (277). It defines a *h-supervised ordering* for every vector $\mathbf{x} \in \mathbb{R}^p$ based on the subsets $\mathrm{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_{|\mathrm{B}|}\}$ and $\mathrm{F} = \{\mathbf{f}_1, \ldots, \mathbf{f}_{|\mathrm{F}|}\}$, as a h-ordering that satisfies the following conditions: $h(\mathbf{b}) = \bot$ then $\mathbf{b} \in \mathrm{B}$, and $h(\mathbf{f}) = \top$ then $\mathbf{f} \in \mathrm{F}$. Note that $\bot, \top$ are the smallest and largest element in the lattice $\mathcal{L}$. Such an *h-supervised ordering* is denoted by $h_{\texttt{SUPER}}(\cdot; \mathrm{B}, \mathrm{F})$.

The main motivation of defining this supervised ordering schema is to obtain maximum and minimum in the lattice $\mathcal{L}$ interpretable with respect to sets B and F. It is important to remind that max and min are the basic words in the construction of all mathematical morphology operators. At this point, the problem is how to define an adequate supervised ordering for a given vector space $\mathbb{R}^p$ and two pixel sets B, F. The approach introduced by (277) involves the computation of standard support vector machine (SVM) to solve a supervised classification problem to define the function $h_{\texttt{SUPER}}(\mathbf{x}; \mathrm{B}, \mathrm{F})$. An amusing geometrical interpretation is based on results from (28), in where the ordering induced by $h_{\texttt{SUPER}}$, corresponds to the signed distance to the separating plane between the convex hull associated to F and the one containing the B.

[6] In this case the sense of the inequality change, *i.e.*, $\mathbf{x}_1 \leq_{h_{\texttt{REF}}} \mathbf{x}_2 \iff ||\mathbf{x}_1 - \mathbf{t}||^2 \geq ||\mathbf{x}_2 - \mathbf{t}||^2$.



(a) colour image: $\mathbf{f}$ **(b)** $h_{\texttt{PCA}}$ **(c)** $h_{\texttt{ANOM}}$

(d) B and F sets **(e)** $h_{\texttt{REF}}$ **(f)** $h_{\texttt{SUPER}}$

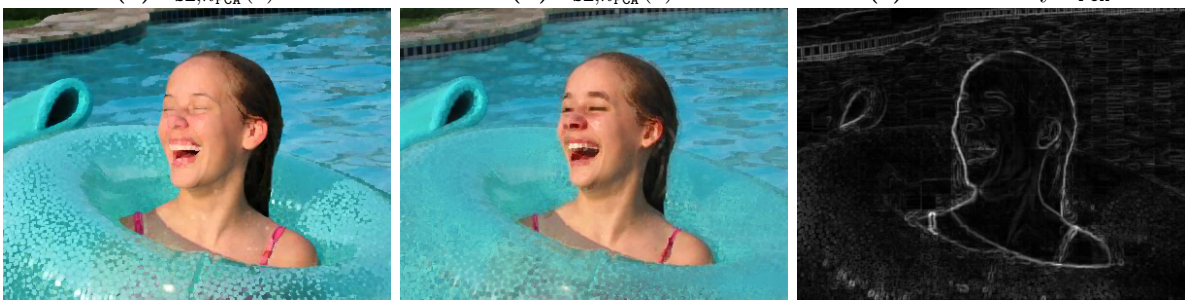
Figure 1.4 Comparison of different h-mappings considered in this section for a given colour image. Referenced h-mapping requires prior information given by an one class training set T, in this example $T = \mathrm{F}$. Supervised h-mappings requires prior information given by the sets B and F. Anomaly based ordering is intrinsically adapted to the image.

From (64), the solution of the classification case of SVM can be expressed as follows:

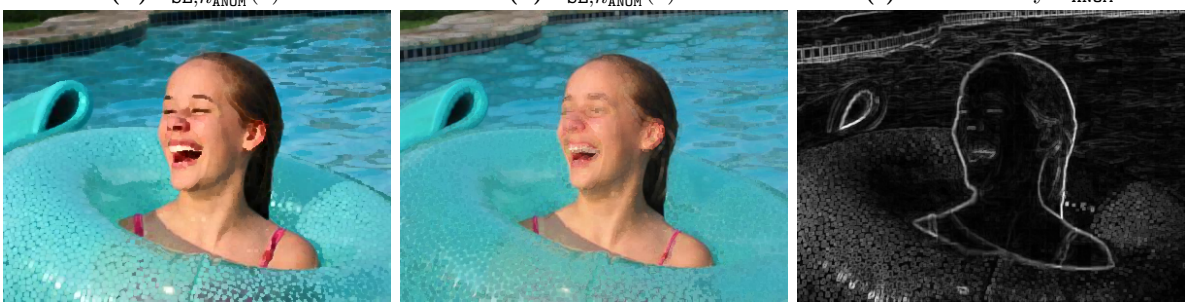
$$h_{\texttt{SUPER}}(\mathbf{x};\mathrm{B},\mathrm{F}) = \sum_{k=1}^{|\mathrm{B}|} \theta^k \mathbf{K}(\mathbf{b}_k,\mathbf{x}) + \sum_{j=1}^{|\mathrm{F}|} \theta^j \mathbf{K}(\mathbf{f}_j,\mathbf{x}) \tag{1.10}$$

where θ^k are computed simultaneous as a quadratic programming optimisation problem (64). For all the examples, given in this section we have used a Gaussian Kernel, with the Euclidean distance between colour or spectra, *i.e.*, $\mathbf{K}(\mathbf{x}_i,\mathbf{x}_j) = \exp(-c||\mathbf{x}_i - \mathbf{x}_j||^2)$, where the constant c is obtained by cross-validation on the training set (64). Results of this supervised ordering are illustrated in Fig. 1.4(f). The $h_{\texttt{SUPER}}$ matches our intuition of what should be maximum and what should be minimum in the image according to the couple $\{\mathrm{B},\mathrm{F}\}$ in 1.4(d). The supervised adjunction is shown in Fig. 1.5(j-k). Note that the supervised gradient in Fig.1.5(l) is better defined on the contour of the girl in comparison to unsupervised and referenced orders. A second example is presented for the RGB image in Fig. 1.3 considering the training sets in Fig. 1.6(a). Note that the supervised lattice in Fig. 1.6(c), is a mapping from the spectral information to a linear ordering (from top-left corner to bottom right corner). One advantage of the definition of h-ordering on vector space is that it can be applied directly to multispectral or even hyperspectral images.

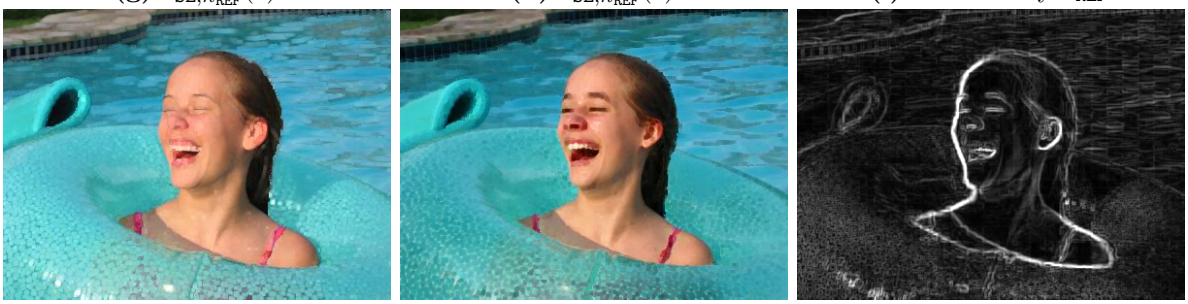
(a) $\delta_{\mathtt{SE},h_{\mathtt{PCA}}}(\mathbf{f})$ **(b)** $\varepsilon_{\mathtt{SE},h_{\mathtt{PCA}}}(\mathbf{f})$ **(c)** Gradient by $h_{\mathtt{PCA}}$



(d) $\delta_{\mathtt{SE},h_{\mathtt{ANOM}}}(\mathbf{f})$ **(e)** $\varepsilon_{\mathtt{SE},h_{\mathtt{ANOM}}}(\mathbf{f})$ **(f)** Gradient by $h_{\mathtt{ANOM}}$



(g) $\delta_{\mathtt{SE},h_{\mathtt{REF}}}(\mathbf{f})$ **(h)** $\varepsilon_{\mathtt{SE},h_{\mathtt{REF}}}(\mathbf{f})$ **(i)** Gradient by $h_{\mathtt{REF}}$

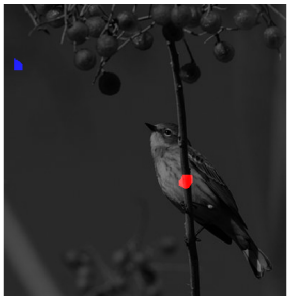


(j) $\delta_{\mathtt{SE},h_{\mathtt{SUPER}}}(\mathbf{f})$ **(k)** $\varepsilon_{\mathtt{SE},h_{\mathtt{SUPER}}}(\mathbf{f})$ **(l)** Gradient by $h_{\mathtt{SUPER}}$

Figure 1.5 Comparison of colour dilation, erosion and associated gradient using different h-orderings. Gradients have been normalised from zero to one to make easier the visual comparison.

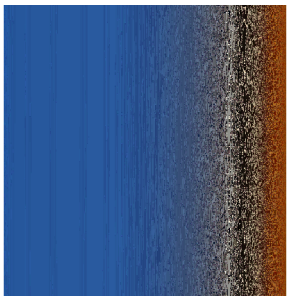
1.2.3 Ordering based on anomalies

Distance based ordering approaches discussed above are valid if the pair set (B,F) is available. Obviously, one cannot realistically believe that for every application the exact spectral information about the background of the image is available. Thus, if one gives up this paradigm, no other option different to unsupervised ordering remains. Therefore, in order to take advantage of the physical structure of an image, it was introduced in (279) an ordering based on "anomalies" with respect to a background associated with a majority of points. It is called *depth ordering*



(a) Background/foreground training set

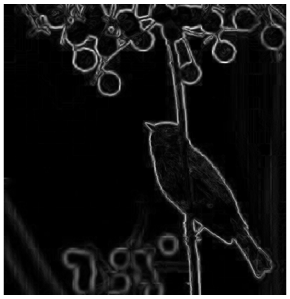
(b) $h_{\texttt{SUPER}}(\cdot;\mathrm{B},\mathrm{F})$



(c) Learned Order from (b).

(d) $\delta_{\texttt{SE},h_{\texttt{SUPER}}}(\mathbf{f})$

(e) $\varepsilon_{\texttt{SE},h_{\texttt{SUPER}}}(\mathbf{f})$



(f) Gradient by $h_{\texttt{SUPER}}$ in (b)

Figure 1.6 Background pixels are in blue, and foreground ones in red. c) The minimum in the supervised ordering is placed at the top left corner and the maximum at the bottom right corner. Morphological operators are computed by using a square of side three pixels as `SE`.

and is maximal in the "center" of the spectral representation of a image $\mathbf{f}$ and it produces a vector ordering "centre-outward" to the outliers in the vector space $\mathbb{R}^p$. In this paradigm, the assumption of existence of an intrinsic background/foreground representation is required, *i.e.*, given a vector image $\mathbf{f}:\Omega\to\mathbb{R}^p$, $\mathbf{X_f}$ has can be decomposed as $\mathbf{X_f}=\{\mathbf{X}_{\mathrm{B}(\mathbf{f})},\mathbf{X}_{\mathrm{F}(\mathbf{f})}\}$ such that $\mathbf{X}_{\mathrm{B}(\mathbf{f})}\cap\mathbf{X}_{\mathrm{F}(\mathbf{f})}=\emptyset$ and $\mathbf{card}\{\mathbf{X}_{\mathrm{B}(\mathbf{f})}\}>\mathbf{card}\{\mathbf{X}_{\mathrm{F}(\mathbf{f})}\}$. Roughly speaking, the assumption means: (1) the image has two main components: the background and the foreground; (2) There are more pixels in the background than in the foreground. Several examples of these kinds of functionals have been analysed in (276). However, we limited ourselves to the *statistical projection depth* case presented in (279) and defined by

$$h_{\texttt{ANOM}}(\mathbf{x};\mathbf{f})=\sup_{||\mathbf{u}||=1}\frac{|\mathbf{u}^T\mathbf{x}-\texttt{MED}(\mathbf{u}^T\mathbf{X_f})|}{\texttt{MAD}(\mathbf{u}^T\mathbf{X_f})} \tag{1.11}$$

where `MED` denoted the *univariate median* and `MAD` the *median absolute deviation*, *i.e.*, the median of the differences with respect to the median. Note that the superscript ^{T} denotes matrix transposition. Let us now point out some aspects of 1.11 in order to better characterise it. First, it is a anomaly based ordering, due to the fact that if $\mathbf{X_f}\sim\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ a Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ then $h_{\texttt{ANOM}}(\mathbf{x};\mathbf{f})^2\propto(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$, the Mahalanobis distance (see 5 to details). Secondly, 1.11 is invariant to affine transformations in the vector space $\mathbb{R}^p$. Third, unfortunately, the *exact computation* of 1.11 is computationally intensive except when the number of pixels n is very small. However, we can compute a stochastic approximation by using a large number of random projections $\mathbf{u}$ and computing the maximum for a given $\mathbf{x}$ (279).

To summarise the above, the statistical projection depth function in 1.11 induces an anomaly based ordering for images with background/foreground representation. That is an ordering based on a data-adapted function and in such a way that the interpretation of supremum and infimum operations is known a priori, because max values can be associated with “outlier” pixels in the high-dimensional space and min are “central” pixels in $\mathbb{R}^p$ space. A simple example is illustrated in Fig. 1.4(c) where 1.11 “detects" the girl thanks to the fact that her spectral information is unusual in comparison to the one from the swimming pool.

1.2.4 Implementation

Once a h-ordering has been defined, it becomes easy in practice to implement morphological transformations on multidimensional images such as colour or multispectral ones. Actually, we can use a scalar to code each pixel on the image, and the standard morphological transformations for greyscale images can be used directly. The result is deciphered by mapping back the total ordering in to the vector space. An effective implementation using a look-up table has been presented in (263). A pseudo-code for a multivariate erosion is shown in (281). The index image and the sorted vector look-up table constructed above are used to generate an ordered table. At this point, any morphological transformation can be performed on the lattice image, which can be considered as a greyscale image. The output of the morphological transformation is converted back to the original vector space by replacing each pixel by its corresponding vector using a look-up table.

1.3 The False colours Problem Versus the Irregularity Issue

One problem on vector-valued mathematical morphology is the creation of “false colours” or, more generally, false values (249). An operator $\psi : \mathbb{R}^p \to \mathbb{R}^p$ introduces false values whenever there are values in $\psi(\mathbf{f})$ which do not belong to the original image $\mathbf{f}$. The abnormal false values can be a problem in some applications such as when dealing with remote sensing images (249).

A total order, such as the lexicographical order and a reduced ordering combined with a look-up table, circumvents the problem of false values (249). Using a total order, the supremum and the infimum of a finite set are elements of the set, *i.e.*,they coincide with the maximum and minimum operations, respectively. On the downside, a total order can be irregular in a metric space. According to Chevallier and Angulo, the irregularities follow because the topology induced by a total order may not reproduce the topology of a metric space (57). Specifically, let the value set $\mathbb{V}$ be a totally ordered set as well as a metric space, with metric $d : \mathbb{V} \times \mathbb{V} \to [0, +\infty)$. (57) showed that there exist $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{V}$ such that $\boldsymbol{u} \leq \boldsymbol{v} \leq \boldsymbol{w}$ but $d(\boldsymbol{u}, \boldsymbol{w}) < d(\boldsymbol{u}, \boldsymbol{v})$ under mild conditions with respect to the connectivity of $\mathbb{V}$. In words, although $\boldsymbol{u}$ is closer to $\boldsymbol{w}$ than to $\boldsymbol{v}$, the inequalities $\boldsymbol{u} \leq \boldsymbol{v} \leq \boldsymbol{w}$ suggest $\boldsymbol{w}$ is farther from $\boldsymbol{u}$ than $\boldsymbol{v}$. Since the morphological operators are defined using the extrema operators, they do not take the metric of $\mathbb{V}$ into account.

A visual interpretation of the irregularity is shown in Figure 1.7, which is very similar to an example provided in (57). Figure 1.7a) shows an image with three RGB colours, namely $\boldsymbol{u} = (0,0,0)$, $\boldsymbol{v} = (0,0,1)$, and $\boldsymbol{w} = (1/255,0,0)$. The toy image $\mathbf{f}$ is obtained by replacing pure black values $\boldsymbol{u}$ by $\boldsymbol{w}$ with probability 0.3 from an image of size 32×64 with two stripes of colours

(a) Toy Image $\mathbf{f}$. $\mathbf{f}$ has two types of dark pixels.

(b) Dilated Image $\delta^L_{\mathtt{SE}}(\mathbf{f})$ (Lexicographical RGB)

(c) Dilated Image $\delta^M_{\mathtt{SE}}(\mathbf{f})$ (Marginal)

Figure 1.7 Illustrative example of the irregularity issue. Image with three colours and its corresponding dilation by a cross structuring element using the RGB lexicographical and marginal orderings.

blue and black. The dilations $\delta^L_{\mathtt{SE}}(\mathbf{f})$ and $\delta^M_{\mathtt{SE}}(\mathbf{f})$ by a cross structuring element $\mathtt{SE}$ obtained using the lexicographical RGB and the marginal ordering schemes are also depicted in Figure 1.7. Visually, $\boldsymbol{u}$ and $\boldsymbol{w}$ are black colours while $\boldsymbol{v}$ is a pure blue. Using the Euclidean distance, we obtain $d(\boldsymbol{u}, \boldsymbol{v}) = 1$ and $d(\boldsymbol{u}, \boldsymbol{w}) = 0.005$. These distances agree with our colour perception. However, using the lexicographical ordering, we obtain $\boldsymbol{u} \leq_L \boldsymbol{v} \leq_L \boldsymbol{w}$. As a consequence, the following happens when we compute the dilation $\delta^L_{\mathtt{SE}}(\mathbf{f}) = \mathbf{g}_L$ using the lexicographical ordering: the blue $\boldsymbol{v}$ advances over the black $\boldsymbol{u}$ but it is overlaid by the black $\boldsymbol{w}$, resulting in the irregularities shown in Figure 1.7b). In contrast, the dilated image depicted in Figure 1.7c) obtained using the marginal ordering does not present any visual irregularity.

Although we know that the irregularity results from a divergence between the topologies induced by the metric and the total order, no consensual measure agrees with our visual perception. A measure for the irregularity can help to choose an appropriate ordering scheme for vector-valued mathematical morphology. The following section proposes a measure of irregularity using the Wasserstein metric.

1.3.1 The Wasserstein Metric and the generalised Sum of Pixel-wise Distances

In this subsection, we present a measure for the irregularity, referred to as the *global irregularity index*. Although we are interested in measuring the irregularity implied by a total ordering, we will not assume $\mathbb{V}$ is totally ordered. Indeed, the proposed irregularity measure is well defined whenever the Ω is finite and $\mathbb{V}$ is a metric space. For simplicity, however, the value set $\mathbb{V}$ corresponds to the RGB colour space equipped with the Euclidean distance in the following examples and computational experiments.

The global irregularity index is given by the relative gap between the Wasserstein metric and a generalised sum of pixel-wise distances.

Given an input image $\mathbf{f} \in \mathcal{F}(\Omega, \mathbb{V})$, let $\mathbf{g} = \psi(\mathbf{f})$ denote the output of the image operator. The generalised sum of pixel-wise distances of $\mathbf{f}$ and $\mathbf{g}$ is an operator $\mathcal{D}_{p_0} : \mathcal{F}(\Omega, \mathbb{V}) \times \mathcal{F}(\Omega, \mathbb{V}) \to [0, +\infty)$ given by

$$\mathcal{D}_{p_0}(\mathbf{f}, \mathbf{g}) = \left(\sum_{x \in \Omega} d^{p_0}((\mathbf{f})(x), \mathbf{g}(x)) \right)^{\frac{1}{p_0}}, \tag{1.12}$$

with a parameter $p_0 \geq 1$. The generalised sum of pixel-wise distances is one of the simplest measures that considers the metric d and the pixel locations. However, $\mathcal{D}_{p_0}$ is usually not properly scaled, possibly because its dimension is the same as the metric d. For example, the images shown in Figure 1.7 yield the values $\mathcal{D}_1(\mathbf{f}, \mathbf{g}_L) = 34.12$ and $\mathcal{D}_1(\mathbf{f}, \mathbf{g}_M) = 66.05$. Note that the inequality $\mathcal{D}_1(\mathbf{f}, \mathbf{g}_L) < \mathcal{D}_1(\mathbf{f}, \mathbf{g}_M)$ holds true although $\mathbf{g}_L$ is more irregular than $\mathbf{g}_M$. Hence, the generalised sum of pixel-wise distances is not an appropriate measure for the irregularity.

Let us now review the Wasserstein metric, also known as the Earth mover's distance or the Kantorovich-Rubinstein distance in some contexts (230; 289). The Wasserstein metric is formulated as a transport problem and can measure distances between probability distributions (289).

The objective of a transport problem is to minimise the cost to deliver items from n_1 factories to n_2 shops (213). In our context, the transport problem minimises the cost to transform the input image $\mathbf{f}$ into the output image $\mathbf{g}$. The cost is defined using the metric on the value set $\mathbb{V}$. Precisely, let $V(\mathbf{f}) = \{v_1, \ldots, v_{n_1}\}$ and $V(\mathbf{g}) = \{u_1, \ldots, u_{n_2}\}$ be the sets of colour values of $\mathbf{f}$ and $\mathbf{g}$, respectively. Also, let

$$f_i = \texttt{Card}(\{x : (\mathbf{f})(x) = v_i\}) \quad \text{and} \quad g_j = \texttt{Card}(\{x : \mathbf{g}(x) = u_j\}), \tag{1.13}$$

denote respectively the number of pixels of value v_i in the image $\mathbf{f}$ and the number of occurrences of the value u_j in $\mathbf{g}$, for $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$. Given $p_0 \geq 1$, the cost to transform a value v_i of $\mathbf{f}$ into a value u_j of $\mathbf{g}$ is defined by

$$c_{ij} = d^{p_0}(v_i, u_j), \quad \forall i = 1, \ldots, n_1, \, \forall j = 1, \ldots, n_2. \tag{1.14}$$

The Wasserstein metric, denoted by $\mathcal{W}_{p_0} : \mathcal{F}(\Omega, \mathbb{V}) \times \mathcal{F}(\Omega, \mathbb{V}) \to [0, \infty)$ for $p_0 \geq 1$, is given by

$$\mathcal{W}_{p_0}(\mathbf{f}, \mathbf{g}) = \left(\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} c_{ij} x_{ij} \right)^{\frac{1}{p_0}}, \quad p_0 \geq 1, \tag{1.15}$$

where x_{ij} solves the linear programming problem

$$\begin{cases} \text{minimise} & \displaystyle\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} c_{ij} x_{ij} \\ \text{subject to} & \displaystyle\sum_{j=1}^{n_2} x_{ij} = f_i, \quad \forall i = 1, \ldots, n_1, \\ & \displaystyle\sum_{i=1}^{n_1} x_{ij} = g_j, \quad \forall j = 1, \ldots, n_2, \\ & x_{ij} \geq 0, \quad \forall i = 1, \ldots, n_1, \, \forall j = 1, \ldots, n_2. \end{cases} \tag{1.16}$$

The Wasserstein metric is the p_0-th root of the minimal cost to transform $\mathbf{f}$ into $\mathbf{g}$. In the transport problem (1.16), the variable x_{ij} represents the (optimal) number of pixels with value v_i of $\mathbf{f}$ transformed to pixels with value u_j of $\mathbf{g}$. Moreover, the solution of (1.16), which can be arranged in a matrix $X \in \mathbb{R}^{n_1 \times n_2}$, is an optimal transport plan. An optimal transport plan is a cyclically monotone plan in the sense that the cost $\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} c_{ij} x_{ij}$ cannot be improved by

changing the number of pixels with value v_i transformed to pixels with value u_j (289). For the images shown in Figure 1.7, we obtain $\mathcal{W}_1(\mathbf{f},\mathbf{g}_L) = 6.18$ and $\mathcal{W}_1(\mathbf{f},\mathbf{g}_M) = 65.94$. Note that the inequality $\mathcal{W}_1(\mathbf{f},\mathbf{g}_L) < \mathcal{W}_1(\mathbf{f},\mathbf{g}_M)$ holds despite $\mathbf{g}_L$ being more irregular than $\mathbf{g}_M$. Like the generalised sum of pixel-wise distances, the Wasserstein metric is not appropriate for measuring the irregularity.

1.3.2 The Global Irregularity Index

Although both the generalised sum of pixel-wise distances and the Wasserstein metric are, *per se*, not appropriate to evaluate the irregularity, we advocate that they can be combined to yield a useful measure.

First of all, note that the generalised sum of pixel-wise distances satisfies

$$\mathcal{D}_{p_0}(\mathbf{f},\mathbf{g}) = \left(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} c_{ij}y_{ij}\right)^{\frac{1}{p_0}}, \quad p_0 \geq 1, \tag{1.17}$$

where

$$y_{ij} = \mathrm{Card}\left(\{x : (\mathbf{f})(x) = v_i \text{ and } \mathbf{g}(x) = u_j, x \in \Omega\}\right), \tag{1.18}$$

for all $i = 1,\ldots,n_1$ and $j = 1,\ldots,n_2$. Moreover, it is not hard to see that $y_{ij} \geq 0$ and the identities

$$\sum_{j=1}^{n_2} y_{ij} = f_i \quad \text{and} \quad \sum_{i=1}^{n_1} y_{ij} = g_j, \tag{1.19}$$

where f_i and g_j are given by (1.13), hold for all $i = 1,\ldots,n_1$ and $j = 1,\ldots,n_2$. Therefore, the generalised sum of pixel-wise distances also measures the cost of transforming $\mathbf{f}$ into $\mathbf{g}$. Because $\mathcal{W}_{p_0}$ is the minimal cost, the inequality $\mathcal{W}_{p_0}(\mathbf{f},\mathbf{g}) \leq \mathcal{D}_{p_0}(\mathbf{f},\mathbf{g})$ holds for any $\mathbf{f}$ and $\mathbf{g} = \psi(\mathbf{f})$. The y_{ij}'s given by (1.18), which can be arranged in a matrix $Y \in \mathbb{R}^{n_1\times n_2}$, is called the operator's plan. The operator's plan is probably not an optimal transport plan. Indeed, one usually can reduce the cost $\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} c_{ij}y_{ij}$ by rerouting the number of pixels with value v_i in $\mathbf{f}$ transformed to pixels with value u_j in $\mathbf{g}$. In some sense, the difference $\mathcal{D}_{p_0}(\mathbf{f},\mathbf{g}) - \mathcal{W}_{p_0}(\mathbf{f},\mathbf{g})$ measures the cycles in the operator's plan that can be reduced.

In order to reduce the impact of the metric d on the value set $\mathbb{V}$ and the impact on the choice of the parameter $p_0 \geq 1$, we propose to measure the irregularity using the mapping $\Phi^g_{p_0} : \mathcal{F}(\Omega,\mathbb{V}) \times \mathcal{F}(\Omega,\mathbb{V}) \to [0,1]$ given by the relative gap between $\mathcal{D}_{p_0}$ and $\mathcal{W}_{p_0}$. Precisely, given images $\mathbf{f},\mathbf{g} \in \mathcal{F}(\Omega,\mathbb{V})$, we define the global irregularity index by means of the equation

$$\Phi^g_{p_0}(\mathbf{f},\mathbf{g}) = \frac{\mathcal{D}_{p_0}(\mathbf{f},\mathbf{g}) - \mathcal{W}_{p_0}(\mathbf{f},\mathbf{g})}{\mathcal{D}_{p_0}(\mathbf{f},\mathbf{g})}, \quad \text{if} \quad \mathcal{D}_{p_0}(\mathbf{f},\mathbf{g}) \neq 0, \tag{1.20}$$

and $\Phi^g_{p_0}(\mathbf{f},\mathbf{g}) = 0$ if $\mathcal{D}_{p_0}(\mathbf{f},\mathbf{g}) = 0$. Note that the larger the gap between $\mathcal{W}_{p_0}(\mathbf{f},\mathbf{g})$ and $\mathcal{D}_{p_0}(\mathbf{f},\mathbf{g})$, the larger the global irregularity index. Equivalently, we have

$$\Phi^g_{p_0}(\mathbf{f},\mathbf{g}) = \begin{cases} 0, & \text{if } \mathcal{D}_{p_0}(\mathbf{f},\mathbf{g}) = 0, \\ 1 - \dfrac{\mathcal{W}_{p_0}(\mathbf{f},\mathbf{g})}{\mathcal{D}_{p_0}(\mathbf{f},\mathbf{g})}, & \text{otherwise.} \end{cases} \tag{1.21}$$

The irregularity index is symmetric and bounded, that is, $\Phi_{p_0}(\mathbf{f},\mathbf{g}) = \Phi_{p_0}(\mathbf{g},\mathbf{f})$ and $0 \leq \Phi_{p_0}(\mathbf{f},\mathbf{g}) \leq 1$. Moreover, because $\mathcal{D}_{p_0}$ and $\mathcal{W}_{p_0}$ have the same units and similar magnitudes, $\Phi^g_{p_0}(\mathbf{f},\mathbf{g})$ is a dimensionless quantity. The more irregular is $\mathbf{g} = \psi(\mathbf{f})$, the larger the value of $\Phi^g_{p_0}(\mathbf{f},\mathbf{g})$ is expected to be. For example, the irregularity index of the dilated images $\mathbf{g}_L$ and $\mathbf{g}_M$ shown in Figure 1.7b) and 1.7c) are $\Phi^g_1(\mathbf{f},\mathbf{g}_L) = 81.9\%$ and $\Phi^g_1(\mathbf{f},\mathbf{g}_M) = 0.17\%$, respectively. The following examples explores further the global irregularity index using toy images.

As example, some computational experiments with tiny colour images are provided. Precisely, the global irregularity index Φ^g_1 of morphological operators applied on one hundred colour images from the CIFAR10 dataset is computed. The measures have been computed using erosion, dilation, opening, and closing by a 3×3 square structuring element. Specifically, we have computed twenty-five dilations, erosions, openings, and closings on different images from the CIFAR10 dataset. Three approaches based on total orderings are considered. The first is the RGB lexicographical order in which the colours are ranked sequentially according to the red, green, and blue channels. The other two approaches are based on reduced orderings (281). Specifically, we considered the supervised reduced ordering based on an SVM with radial basis function kernel (277). The SVM is trained to distinguish the central object on an image from the rest. The last approach uses an unsupervised reduced ordering based on the random projection depth, which aims to discriminate between background and foreground pixels (279). For comparison, two approaches based on partial orderings are also included. Namely, the marginal approach and the approach based on Loewner order (44). Because these two approaches are not based on total orders, their output images are not expected to be very irregular.

Figure 1.8 depicts some original images, the outcome of a morphological operator, and the corresponding global irregularity index. The images in the first column correspond to the original colour images. The following columns present the output of morphological operators defined using the marginal, Loewner, lexicographical RGB, supervised SVM-based, and projection depth approaches, respectively. We provide the global irregularity index below the images produced by the morphological operators. As expected, the marginal and the Loewner approaches yielded global irregularity indexes smaller than the lexicographical, SVM-based, and projection depth approaches. The supervised SVM-based approach yielded the most irregular dilated image. The irregularity index of 5.38% produced by projection depth's dilation of the cat image is an outlier of the global irregularity index produced by this unsupervised morphological approach. Indeed, the median of the global irregularity index produced by the morphological operators based on the projection depth is 45.22%. The eroded image depicted in the last column of Figure 1.8 is a typical outcome of the projection depth approach. The median of the irregularity indexes produced by the marginal, Loewner, lexicographical, and the SVM-based approaches are 2.56%, 2.81%, 6%, and 22.98%, respectively. The truck image's openings and the car image's closings provide typical examples of the global irregularity index produced by the marginal, Loewner, and lexicographical approaches. The opening of the car image produced by the projection depth approach yielded the largest global irregularity index of this experiment.

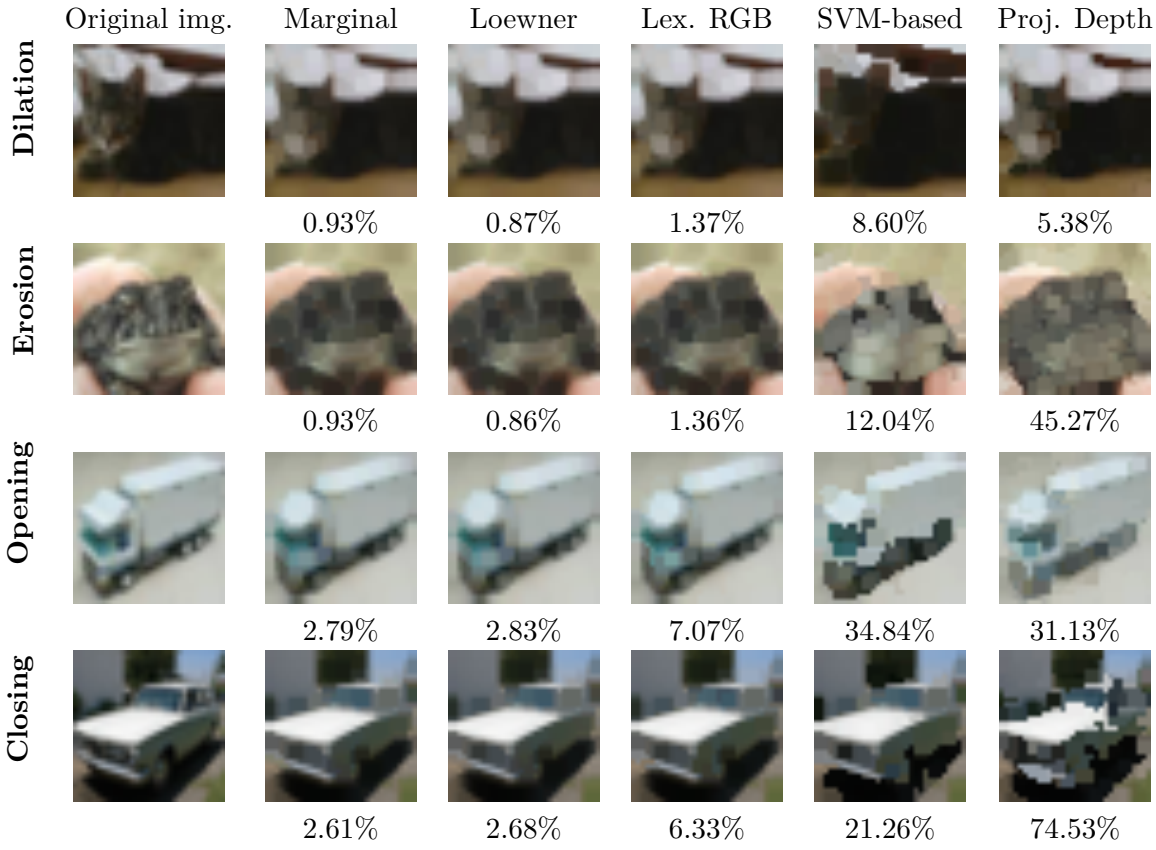


Figure 1.8 Illustrative examples of the global irregularity index Φ_1^g computed for several tiny colour images using different morphological operators.

1.4 Perspectives

Considering the works presented in this chapter, one can imagine the continuation of the idea of order learning in different aspects:

1. How to learn an order function from an database of images?
2. Is it possible to learn order function in the deep learning paradigm? What would be the correct loss functions in this context? How to create a ground-truth of order functions?
3. Can classical convolution neural networks be interpreted as models that learn h-order in a multivariate sense?

Related Publications

1. **Supervised ordering in $\mathbb{R}^p$: Application to morphological processing of hyperspectral images**, S. Velasco-Forero and J. Angulo, vol. 20, num. 11, Oct. 2011, IEEE Transactions on Image Processing.
2. **Random projection depth for multivariate mathematical morphology**, S. Velasco-Forero and J. Angulo, vol. 6, num. 7, Oct. 2012, IEEE Journal of Selected Topics in Signal Processing.

3. **Hit-or-Miss Transform in Multivariate Images**, S.Velasco-Forero and J.Angulo in Advanced Concepts for Intelligent Vision Systems, vol. 6474, 2010, pp. 452-463, LNCS, 2010.
4. **Supervised morphology for tensor structure-valued images based on symmetric divergence kernels**, S. Velasco-Forero and J. Angulo, in Geometric Science of Information, vol. 8085, pp. 543-550, 2013.
5. **Vector Ordering and Multispectral Morphological Image Processing**, J. Angulo and S. Velasco-Forero, Advances in Low-Level colour Image Processing, Signals and Communication Technology, pp. 223-239, 2014.
6. **Irregularity Index for Vector-Valued Morphological Operators**, M. Valle, S. Francisco, M.A. Granero and S. Velasco-Forero, Journal of Mathematical Imaging and Vision (2022)

2 Extensions of Mathematical Morphology

> The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.
>
> William Lawrence Bragg

2.1 Introduction

In Chapter one the modelling of a multivariate image based as a lattice is presented. That leads to the problem of morphological treatment of images to determine an order function for vector data. In this chapter, most well-known morphological operators are reviewed. Additionally, adaptive structuring elements are introduced and an application of salt and pepper denoising is illustrated used conditional morphology. One can define a continuous image $\tilde{\mathbf{f}}$ on $[0, n_1] \times [0, n_2]$ by

$$\forall x \in \Omega, \quad \tilde{\mathbf{f}}(x) = \begin{cases} \mathbf{f}(x - (\frac{1}{2}, \frac{1}{2})) & \text{if} \quad x - (\frac{1}{2}, \frac{1}{2}) \in \Omega \\ \bot \quad \text{otherwise} \end{cases} \tag{2.1}$$

where $\bot$ represent the minimum possible value of any image. Then by considering the dilation of $\tilde{\mathbf{f}}$ by a closed unit square $\mathtt{S} = [-1/2, 1/2]^2$, $\mathbf{f}_c = \sup_{y \in \mathtt{S}} \tilde{\mathbf{f}}(x + y)$, one can obtain a continuous representation of the digital image that is upper semicontinuous and contrast invariant (48). In the sequel, by a convenient abuse of notation, we will use $\mathbf{f}$ to refer $\mathbf{f}_c$.

2.2 Inf/Sup convolutions

We study here functions $\mathbf{f} : \Omega \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}}$ it allowed to be *extended-real-valued*, *i.e.*, to take values in $\overline{\mathbb{R}} = [-\infty, \infty]$. Accordingly, the set of all such functions is denoted by $\mathcal{F}(\Omega, \overline{\mathbb{R}})$.

Definition 2.1. In mathematical morphology (246), the dilation (*sup-convolution*) $\delta_{\mathtt{SE}}(\mathbf{f})$ of $\mathbf{f}$ is given by:

$$\forall x \in \Omega, \quad \delta_{\mathtt{SE}}(\mathbf{f})(x) := \sup_{y \in \Omega} \{\mathbf{f}(y) + \mathtt{SE}(x - y)\} = \sup_{w \in \Omega} \{\mathbf{f}(x - w) + \mathtt{SE}(w)\}, \tag{2.2}$$

where $\mathtt{SE} \in \mathcal{F}(\Omega, \overline{\mathbb{R}})$ is the *structuring function* which determines the effect of the operator. Here the inf-addition rule $\infty - \infty = \infty$ is to be used in case of conflicting infinities. $\sup(\mathbf{f})$ and $\inf(\mathbf{f})$ refer to the *supremum* (least upper bound) and *infimum* (greatest lower bound) of $\mathbf{f}$.

The erosion (246) $\varepsilon_{\mathtt{SE}}(\mathbf{f})$, known as *inf-convolution* in convex analysis (189), is the adjoint operator to 2.2, and it is defined as

$$\forall x \in \Omega, \quad \varepsilon_{\mathtt{SE}}(\mathbf{f})(x) := -\delta_{\hat{\mathtt{SE}}}(-(\mathbf{f})(x) = \inf_{y \in \Omega} \{\mathbf{f}(y) - \mathtt{SE}(y-x)\} = \inf_{w \in \Omega} \{\mathbf{f}(x+w) - \mathtt{SE}(w)\}, \quad (2.3)$$

where the transposed structuring function is $\hat{\mathtt{SE}}(x) = \mathtt{SE}(-x)$.

Remark 2.2. $\forall \mathbf{f}, \mathbf{g} \in \mathcal{F}(\Omega, \overline{\mathbb{R}})$

1. The operators (2.2) and (2.3) are translation invariant.
2. (2.2) and (2.3) correspond to one another through the duality relation $\delta_{\mathtt{SE}}(\mathbf{f})(x) \leq \mathbf{g}(x) \iff (\mathbf{f})(x) \leq \varepsilon_{\mathtt{SE}}(\mathbf{g})(x)$, called *adjunction* (96).
3. An operator ψ is called *increasing* if $(\mathbf{f})(x) \geq \mathbf{g}(x) \Rightarrow \psi(\mathbf{f})(x) \geq \psi(\mathbf{g})(x)$ $\forall x$. The dilation (2.2) and erosion (2.3) are increasing for all $\mathtt{SE}$.
4. An operator ψ is called *extensive* (resp. *antiextensive*) if $\psi(\mathbf{f})(x) \geq (\mathbf{f})(x)$ (resp. $\psi(\mathbf{f})(x) \leq (\mathbf{f})(x)$), $\forall x$. The dilation (2.2) (resp. erosion (2.3)) is extensive (resp. antiextensive) if and only if $\mathtt{SE}(0) \geq 0$, *i.e.*, the structuring function evaluated at the origin is non-negative.
5. $\varepsilon_{\mathtt{SE}}(\mathbf{f})(x) \leq (\mathbf{f})(x) \leq \delta_{\mathtt{SE}}(\mathbf{f})(x)$ if and only if $\mathtt{SE}(0) \geq 0$.
6. $\delta_{\mathtt{SE}}$ (resp. $\varepsilon_{\mathtt{SE}}$) does not introduce any local maxima (resp. local minima) if $\mathtt{SE} \leq 0$ and $\mathtt{SE}(0) = 0$. In this case, we say that $\mathtt{SE}$ is *centered*.

Proof. (1) and (2) are classical results from (246). As explained in (108) and (163), the *adjunction* is related to a well-known concept in group and lattice theory, the *Galois connection*. (3) and (6) are easy to prove directly from the definition of the operators. It has been also proved in the original paper of inf-convolution (Proposition 6.d) in (189). (4) $\forall \mathbf{f}, \delta_{\mathtt{SE}}(x) \geq (\mathbf{f})(x) \Rightarrow \forall f, \sup(\mathbf{f}(x-w) + \mathtt{SE}(w) - \mathbf{f}(x))(x) \geq 0 \Rightarrow \mathtt{SE}(0) \geq 0$. Now, $\sup(\mathbf{f}(x-w) + \mathtt{SE}(w)) \geq (\mathbf{f})(x) + \mathtt{SE}(0)$, if $\mathtt{SE}(0) \geq 0 \Rightarrow \sup(\mathbf{f}(x-w) + \mathtt{SE}(w)) \geq \mathbf{f}(x)$. From (3) and (4) is easy to prove (5). □

Usually in practice the function $\mathbf{f}$ is a countable set of points, then max and min are used instead of supremum and infimum.

Definition 2.3. The opening (resp. closing) of $\mathbf{f}$ by the structuring function $\mathtt{SE}$ is denoted by $\gamma_{\mathtt{SE}}(\mathbf{f})$ (resp. $\varphi_{\mathtt{SE}}(\mathbf{f})$), and it is defined by $\gamma_{\mathtt{SE}}(\mathbf{f}) := \varepsilon_{\mathtt{SE}}(\delta_{\mathtt{SE}}(\mathbf{f}))$ (resp. $\varphi_{\mathtt{SE}}(\mathbf{f}) := \delta_{\mathtt{SE}}(\varepsilon_{\mathtt{SE}}(\mathbf{f}))$)

Remark 2.4. $\forall \mathbf{f}, \mathbf{g} \in \mathcal{F}(\Omega, \overline{\mathbb{R}})$

1. Opening and closing are *duals*, $\varphi(\mathbf{f}) = -\gamma(-\mathbf{f})$
2. Opening and closing are idempotents, $\varphi(\varphi(\mathbf{f})) = \varphi(\mathbf{f})$ and $\gamma(\gamma(\mathbf{f})) = \gamma(\mathbf{f})$
3. Opening is *anti-extensive* and closing is *extensive*, $\mathbf{f} \leq \varphi(\mathbf{f})$ and $\gamma(\mathbf{f}) \leq \mathbf{f}$

2.3 Geodesic Morphological Operations

In this section, we firstly provide an overview of fixed-point geodesic morphological operations which achieve reconstructions from a *marker* function under a constraining second image called *mask*. Then, we explain how these operations can be included as layers in a neural network, so that backpropagation can be performed. Finally, we present the main property of the Jacobian, that allows both a geometric interpretation and a strong robustness through noise. The main ingredient of geodesic transformations is the *geodesic dilation*. Let us consider numerical functions $\mathbf{f}, \mathbf{g} \in \mathcal{F}(\Omega, \mathbb{R})$, the set of functions mapping from space of points Ω to $\mathbb{R}$, the set of different possible values of the image.

Definition 2.5. Let $\mathbf{f}, \mathbf{g}$ be such that $\mathbf{f} \leq \mathbf{g}$, $\mathbf{f}$ is called the *marker* and $\mathbf{g}$ the *mask* (251). The *geodesic dilation* of size one of $\mathbf{f}$ with respect to $\mathbf{g}$ is denoted by $\delta^{(1)}_{\mathbf{g}}(\mathbf{f})$ and is defined as the point-wise minimum between $\mathbf{g}$ and the elementary dilation in a given spatial space (a.k.a. structuring element $\mathtt{SE}$), $\delta^1 := \delta_{\mathtt{SE}}$ of the marker image, i.e.,

$$\forall x \in \Omega, \quad \delta^{(1)}(\mathbf{f}, \mathbf{g})(x) := \delta^{(1)}_{\mathbf{g}}(\mathbf{f})(x) := \delta^{(1)}(\mathbf{f})(x) \wedge \mathbf{g}(x) \tag{2.4}$$

where $\wedge$ denotes the minimum coordinate-wise operation.

Definition 2.6. The *reconstruction by dilation*(251) of a mask $\mathbf{g}$ from a marker $\mathbf{f}$ is defined as the geodesic dilation of $\mathbf{f}$ with respect to $\mathbf{g}$ iterated until stability and is denoted by $\mathtt{REC}^{\delta}_{\mathbf{g}}(\mathbf{f})$:

$$\forall x \in \Omega, \quad \mathtt{REC}^{\delta}(\mathbf{f}, \mathbf{g})(x) := \mathtt{REC}^{\delta}_{\mathbf{g}}(\mathbf{f})(x) := \lim_{k \to +\infty} \delta^{(k)}_{\mathbf{g}}(\mathbf{f}(x)) = \lim_{k \to +\infty} \underbrace{\delta^{(1)}_{\mathbf{g}} \circ \dots \delta^{(1)}_{\mathbf{g}}}_{k\text{times}}(\mathbf{f}(x)) \tag{2.5}$$

where k is such that $\delta^{(k)}_{\mathbf{g}}(\mathbf{f}(x)) = \delta^{(k+1)}_{\mathbf{g}}(\mathbf{f}(x))$.

The reconstruction by dilation extracts the *domes* or *peaks* of the mask which are marked by the marker. This is illustrated in Fig. 2.1(a).

Definition 2.7. The *h-maxima* transform provides a filter to select signal maxima using a contrast criterion (251),

$$\forall x \in \Omega, \quad \mathtt{HMAX}_h(\mathbf{f})(x) = \mathtt{REC}^{\delta}_{\mathbf{f}}(\mathbf{f}(x) - h) \tag{2.6}$$

where $h \in \mathbb{R}$ is a parameter.

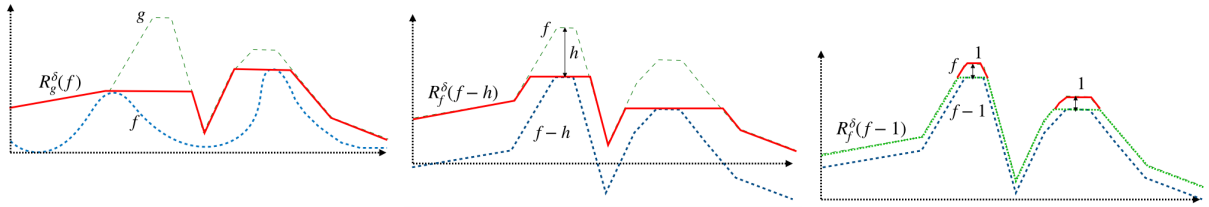
$\mathtt{HMAX}_h$ transform cuts local maxima overall the image given a parameter h. Finally, a common way to detect peaks (251), is the *regional maxima transformation*.

Definition 2.8. The *regional maxima transformation* for $\mathbf{f}$ is defined by

$$\forall x \in \Omega, \quad \mathtt{RMAX}(\mathbf{f})(x) = \mathbf{f}(x) - \mathtt{REC}^{\delta}_{\mathbf{f}}(\mathbf{f}(x) - \epsilon), \tag{2.7}$$

$\epsilon = 1$ in the case of discrete values for $\mathbf{f}$.

Some examples of (2.6) and (2.7) are shown in Fig. 2.1(b) and (c).



(a) $\mathtt{REC}^\delta(\mathbf{f},\mathbf{g})$ **(b)** $\mathtt{HMAX}_h(\mathbf{f})$ **(c)** $\mathtt{RMAX}(\mathbf{f})$.

Figure 2.1 Examples of different geodesic morphological operators in a mask $\mathbf{g}$ from a marker $\mathbf{f}$. The illustrated example in (c) uses $\epsilon = 1$, but in practical implementation it should be a small number.

2.4 Toggle mapping

The *toggle contrast mapping* is based on the idea of using a dilation process near a local maximum and an erosion process around a local minimum. The decision whether a pixel belongs to the influence zone of a maximum or a minimum is made on the basis of the morphological Laplacian Δ defined in (274) as the difference between the gradients by dilation Δ^δ and erosion Δ^ε (220):

$$\forall x \in \Omega, \quad \Delta_{\mathtt{SE}}(\mathbf{f})(x) = \Delta^\delta_{\mathtt{SE}}(\mathbf{f}) - \Delta^\varepsilon_{\mathtt{SE}}(\mathbf{f})(x),$$

where $\Delta^\delta_{\mathtt{SE}}(\mathbf{f})(x) = (\delta_{\mathtt{SE}}(\mathbf{f})(x) - \mathbf{f}(x))$ and $\Delta^\varepsilon_{\mathtt{SE}}(\mathbf{f})(x) = (\mathbf{f}(x) - \varepsilon_{\mathtt{SE}}(\mathbf{f})(x))$.

If the Laplacian is negative, then the pixel is considered to be in the influence zone of a maximum, while it is regarded to belong to the influence zone of a minimum if the Laplacian is positive. With this framework, we adopt the three-state toggle contrast mapping proposed in (243) and based on the erosion, dilation, and identity transformations:

Definition 2.9. (243) Given an image $\mathbf{f}$ the toggle contrast mapping is defined as follows:

$$\forall x \in \Omega, \quad \tau_{\mathtt{SE}}(\mathbf{f})(x) = \begin{cases} \varepsilon_{\mathtt{SE}}(\mathbf{f})(x) & \text{if} \quad \Delta_{\mathtt{SE}}(\mathbf{f})(x) > 0, \\ \delta_{\mathtt{SE}}(\mathbf{f})(x) & \text{if} \quad \Delta_{\mathtt{SE}}(\mathbf{f})(x) < 0, \\ \mathbf{f}(x) & \text{otherwise.} \end{cases} \tag{2.8}$$

Note that the original shock filter proposed of Kramer & Bruckner (139) corresponds to a two-state toggle contrast mapping where the identity transformation is not considered and the strict inequality $<$ in (2.8) is replaced by $\leq$, see also (181). The three-state toggle contrast mapping is more robust than the two-state one because it is self-dual and preserve the original signal in a single-slope signal, *i.e.*, signals such that $\forall x : \Delta^\varepsilon_{\mathtt{SE}}(\mathbf{f}(x)) = \Delta^\delta_{\mathtt{SE}}(\mathbf{f}(x))$. The original method of Kramer and Bruckner was formulated in a fully discrete way. The term *shock filtering* was introduced later by Osher and Rudin (201). Shock filtering constitutes an example of a PDE that is difficult to analyse in the continuous setting, while for a 1-D space discretisation it has been shown in (293) that this process is well-posed and satisfies a maximum-minimum principle. An analytic solution of the corresponding dynamical system was even found (293). Different modifications have been proposed in order to improve the performance of shock filters. Alvarez

and Mazorra (4) replaced the Laplacian as edge detector by $K_\sigma * \Delta(\mathbf{f})$, where K_σ is a Gaussian with standard deviation σ, $\Delta\mathbf{f}$ is the gradient of $\mathbf{f}$, and $*$ denotes convolution.

By iterating toggle contrast mappings, a sharp discontinuity (shock) at the borderline between two influence zones is produced. Within each zone, a constant segment is created. Iterations are usually performed until idempotence[1], i.e.,

$$\forall x \in \Omega, \quad \tau^{\infty}_{\mathtt{SE}}(\mathbf{f})(x) = \lim_{i \to \infty} \tau^{i}_{\mathtt{SE}}(\mathbf{f})(x) \tag{2.9}$$

where $\tau^{i}_{\mathtt{SE}}(\mathbf{f}) = \tau_{\mathtt{SE}}(\tau^{i-1}_{\mathtt{SE}}(\mathbf{f}))$, $\tau^{0}_{\mathtt{SE}}(\mathbf{f})(x) = \mathbf{f}(x)$ and $\tau_{\mathtt{SE}}(\mathbf{f})(x)$ is given by (2.8). The iterations of toggle contrast converge to a *fixed point* (139) reached after a finite number of iterations. In (111) are defined self-dual operators based on the morphological centre and, more generally, self-dual filters. Basically, it states that every increasing, self-dual operator can be modified in such a way that when iterated on any given image, it leads to a pixel-wise monotone sequence (*i.e.*, strictly increasing or decreasing in each pixel). This implies a convergence to a limit operator, thus avoiding oscillation problems which typically hold for non-convergent filters such as median filtering. The convergence is obtained when all the pixels are fixed points, as shown in Fig. 2.2. To illustrate the evolution until convergence, Figs. 2.2(a-b) show the first and second iteration of (2.9) in a bi-dimensional representation of image value gradient by erosion and dilation. When convergence is reached, Fig. 2.2(c), all the pixels are fixed points according to the criterion (2.8). Note the presence of pixels in the line $\Delta^{\varepsilon}_{\mathtt{SE}}(\mathbf{f})(x) = \Delta^{\delta}_{\mathtt{SE}}(\mathbf{f})(x)$ revealing that some pixels have identical values for their gradients by erosion and dilation. Shock filters following (2.9) have two main drawbacks. First, they suffer from the halo effect (90; 181) due to the tie case in the definition. In addition, they require a large number of iterations to converge. For example, in the case of the Cameraman image, 96 iterations were needed to reach the stable result shown in Fig. 2.2(f). These two drawback motived us to define the conditional toggle mapping introduced in section 2.6.

2.5 Type of Structuring Elements

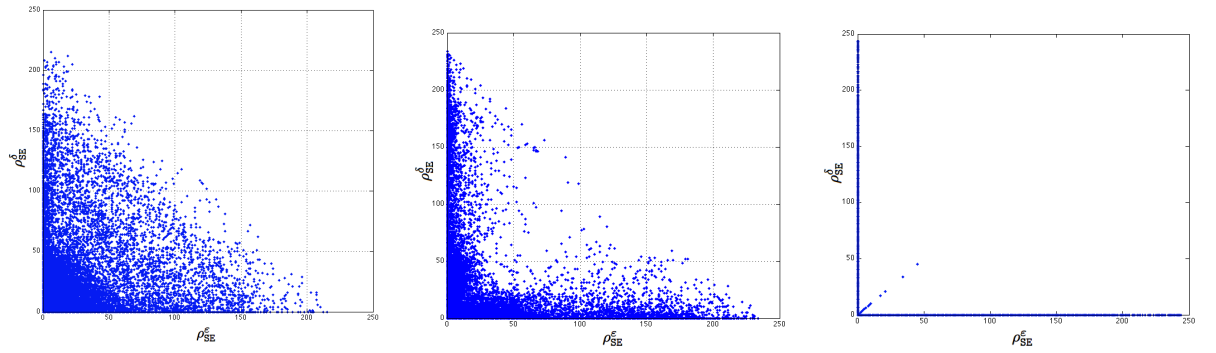
2.5.1 Flat Structuring Element

The most commonly studied framework for dilation/erosion of functions is based on *flat structuring functions*, where structuring elements are viewed as *shapes*. More precisely, given the structuring element $\mathtt{B} \subseteq \Omega$, its associated structuring function is

$$\forall x \in \Omega, \quad B(x) = \begin{cases} 0 & \text{if } x \in \mathtt{B}, \\ -\infty & \text{if } x \in \mathtt{B}^c. \end{cases} \tag{2.10}$$

Hence, the flat dilation $\delta_B(\mathbf{f})$ and flat erosion $\varepsilon_B(\mathbf{f})$ can be computed respectively by the moving local maxima and minima filters with a neighbourhood induced by B. The shape of $\mathtt{B}$ is often a

[1] An operator $\psi : \Omega \to \mathbb{R}^p$ is idempotent if $\forall x \in \Omega, \psi^2(x) = \psi(\psi(x)) = \psi(x)$.



(a) $\Delta^{\varepsilon}_{\mathtt{SE}}(\mathbf{f})$ vs $\Delta^{\delta}_{\mathtt{SE}}(\mathbf{f})$ (b) $\Delta^{\varepsilon}_{\mathtt{SE}}(\tau_{\mathtt{SE}}(\mathbf{f}))$ vs $\Delta^{\delta}_{\mathtt{SE}}(\tau_{\mathtt{SE}}(\mathbf{f}))$ (c) $\Delta^{\varepsilon}_{\mathtt{SE}}(\tau^{\infty}_{\mathtt{SE}}(\mathbf{f}))$ vs $\Delta^{\delta}_{\mathtt{SE}}(\tau^{\infty}_{\mathtt{SE}}(\mathbf{f}))$

(d) $\mathbf{f}$ (e) $\tau_{\mathtt{SE}}(\mathbf{f})$ (f) $\tau^{\infty}_{\mathtt{SE}}(\mathbf{f})$

Figure 2.2 (a)–(c) Bi-dimensional representation of the evolution in the classical shock filter (2.8) for the Cameraman grey-scale image shown in (d). (e) One iteration of the shock filter in (2.8) (f) Image obtained after iterating (2.8) until stability.

disk of radius θ, denoted by B_θ.

$$\forall x \in \Omega, \quad B_\theta(x) = \begin{cases} 0 & \text{if } \|x\| \leq \theta, \\ -\infty & \text{if } \|x\| > \theta. \end{cases} \tag{2.11}$$

2.5.2 Quadratic Structuring Element

From the theory of morphological scale-spaces, the most useful non-flat structuring functions are those which depend on a scale parameter (109; 244). The only separable and rotationally invariant structuring functions is the called *quadratic structuring function*(272):

$$\forall x \in \Omega, q_\theta(x) = -\frac{\|x\|^2}{2\theta}, \tag{2.12}$$

such that the corresponding dilation and erosion are equal to the Lax–Oleinik operators or viscosity solutions of the standard Hamilton–Jacobi PDE, also known as morphological PDE: $\mathbf{u}_t(t,x) \mp \|\mathbf{u}_x(t,x)\|^2 = 0$, $(t,x) \in (0,+\infty) \times \Omega$; $\mathbf{u}(0,x) = (\mathbf{f})(x)$, $x \in \Omega$. The morphological PDE was proposed and analysed using 2D boundary propagation in (273) and further analysed using the morphological slope transform in (107).

Remark 2.10. The erosion by a quadratic structuring function with parameter $\theta \in \mathbb{R}^+$ is defined by

$$\varepsilon_{q_\theta}(\mathbf{f})(x) := \inf_{y\in\Omega}\{\mathbf{f}(y) - q_\theta(y-x)\} = \inf_{z\in\Omega}\{\mathbf{f}(z-x) - q_\theta(z)\} = \inf_{z\in\Omega}\left\{\mathbf{f}(z-x) + \frac{\|z\|^2}{2\theta}\right\}. \quad (2.13)$$

The erosion of a function $\mathbf{f}$ by a quadratic structuring function with parameter θ is known as the *Moreau envelope* or *Moreau-Yosida approximation* (189; 205; 222), which offers many benefits specially for optimisation purposes (188). Additionally, 2.13 induces an additive scale-space (107; 124)„ $\varepsilon_{q_{\theta_1}}(\varepsilon_{q_{\theta_2}}(\mathbf{f})) = \varepsilon_{q_{\theta_1+\theta_2}}(\mathbf{f})$ and they form a $(S^{1/2}, +)$ scale-spaces that can be regarded as morphological counterparts to the Gaussian scale-space (273). It also plays a canonical role in the definition of dilation and erosion on Riemannian manifolds (13) and their behaviour with respect to the maxima/minima is well understood (125). Besides their feature extraction properties, morphological dilation and erosion using quadratic structuring functions are a powerful tool for Lipschitz regularisation. For the nonconvex case, the Lasry–Lions double envelope is defined as the composition of two different Moreau envelopes, or using the morphological vocabulary, the composition of an erosion followed by a dilation with quadratic structuring functions. For all $0 < c < 1$ and $0 < \lambda$, the so-called Lasry–Lions regularisers (130) are defined as

$$\begin{aligned}\gamma_\lambda^c(\mathbf{f})(x) &:= \delta_{q_{c\lambda}}(\varepsilon_{q_\lambda}(\mathbf{f}))(x),\\ \varphi_\lambda^c(\mathbf{f})(x) &:= \varepsilon_{q_{c\lambda}}(\delta_{q_\lambda}(\mathbf{f}))(x),\end{aligned}$$

such that if $\mathbf{f}$ is bounded, the functions γ_λ^c and φ_c^λ are bounded and one has the ordering properties for the following envelopes:

- if $\lambda_1 \geq \lambda_2 > 0$, for any $0 < c < 1$ then

$$\gamma_{\lambda_1}^c(\mathbf{f})(x) \leq \gamma_{\lambda_2}^c(\mathbf{f})(x) \leq f \leq \varphi_{\lambda_2}^c(\mathbf{f})(x) \leq \varphi_{\lambda_1}^c(\mathbf{f})(x);$$

- if $0 < c_2 < c_1 < 1$, for any $\lambda > 0$ then

$$\gamma_\lambda^{c_2}(\mathbf{f})(x) \leq \gamma_\lambda^{c_1}(\mathbf{f})(x) \leq f \leq \varphi_\lambda^{c_1}(\mathbf{f})(x) \leq \varphi_\lambda^{c_2}(\mathbf{f})(x).$$

For any bounded function $\mathbf{f}$, Lasry–Lions regularisers provide a function with a Lipschitz continuous gradient, *i.e.*,

$$|\nabla\gamma_\lambda^c(\mathbf{f})(x) - \nabla\gamma_\lambda^c(\mathbf{f})(y)| \leq M_{\lambda,c}\|x-y\|, \qquad |\nabla\varphi_\lambda^c(\mathbf{f})(x) - \nabla\varphi_\lambda^c(\mathbf{f})(y)| \leq M_{\lambda,c}\|x-y\|.$$

where the Lipschitz constant is $M_{\lambda,c} = \max\left((c\lambda)^{-1}, ((1-c)\lambda)^{-1}\right)$. If $\mathbf{f}$ is bounded and Lipschitz continuous, one has

$$\mathrm{Lip}(\gamma_\lambda^c(\mathbf{f})) \leq \mathrm{Lip}(\mathbf{f}) \text{ and } \mathrm{Lip}(\varphi_\lambda^c(\mathbf{f})) \leq \mathrm{Lip}(\mathbf{f}),$$

with

$$\mathrm{Lip}(g) = \sup\left\{\frac{|g(x)-g(y)|}{\|x-y\|};\ x, y \in \mathbb{R}^p,\ x \neq y\right\}.$$

For more details on the properties of Lasry–Lions regularisers in the context of mathematical morphology, see (10).

Remark 2.11. The following statements are interesting about the composition of quadratic morphological operators (52; 222). Let $0 < \mu < \lambda$,

1. $\varepsilon_{q_\lambda}(\gamma_{q_\lambda}(\mathbf{f})) = \varepsilon_{q_\lambda}(\mathbf{f})$;
2. $\gamma_{q_\mu}(\varepsilon_{q_{\lambda-\mu}}(\mathbf{f})) = \varepsilon_{q_{\lambda-\mu}}(\gamma_{q_\lambda}(\mathbf{f}))$;
3. $\gamma_{q_{\lambda-c\lambda}}\varphi^c_\lambda(\mathbf{f}) = \varphi^c_\lambda(\mathbf{f})$.

2.5.3 Adaptive mathematical morphology

The formulation contained in the previous subsection is translation invariant in the space and in the intensity, *i.e.*, the same processing is considered for each pixel x in the image $\mathbf{f}$. Several ways have been analysed to define local characteristics of the image in order to locally design the SE at each point of the product space $(x \times t) \in \Omega \times \mathbb{R}^p$. There are different ways to define a hierarchy of approaches proposed on adaptive morphology. We use the scheme introduced by Roerdink (223). According to the adaptivity considered by the construction of the structuring element, we have two main types:

1. Location-adaptive structuring elements (variability on Ω (12)): The structuring element $\mathtt{SE}(x)$, depends on the location x in the image. It does not depend on the input image $\mathbf{f}(x)$. One of the earliest applications that required the use of variable size SEs is the traffic control camera system (32). This application inspired (32) to consider the perspective effect in the morphological analysis. Vehicles at the bottom of the image are closer and they appear larger than those higher in the camera. Thus, the SE should follow a law of perspective, for instance, vary linearly with its vertical position in the image. Another example is the term "locally adaptable" used in (65), for SEs as a disk where the radius depends on the position of the image.
2. Input-adaptive structuring elements (variability on $\mathbb{R}^p$ (12)): The shape of the $\mathtt{SE}(x)$ at x depends on the local features of an image $\mathbf{f}$. We denote this kind of structuring element by $\mathtt{SE}_\mathbf{f}(x)$. Examples of this type of adaptive are morphological amoebas (148), intrinsic structuring elements (68), region growing structuring element (187), nonlocal structuring elements (280), geodesic neighbourhood (99), bilateral inspired structuring elements (9) and adaptive anisotropic structuring element (34).

The question one can ask is what kind of structuring elements are valid in the sense that adjunction properties are preserved. The following theorem has been presented in (280):

Definition 2.12. A *morphological weight system* $\mathcal{W}_\mathbf{f} : \Omega \times \Omega \to \mathbb{R}^+$ on $\mathbf{f}$ is a weight function such for all $x, y \in \Omega$,

1. $\mathcal{W}_\mathbf{f}(x,x) = 0 \quad \forall x \in \Omega$,
2. $\mathcal{W}_\mathbf{f}(x,y) = \mathcal{W}_\mathbf{f}(y,x) \quad \forall x, y \in \Omega$,
3. $-\infty \leq \mathcal{W}_\mathbf{f}(x,y) \leq 0 \quad \forall x, y \in \Omega$.

Figure 2.3 Example a) Original evaluated graph b) Example of Structuring element considering the geodesic distance approximated in the graph (a) with a center in the point of the noise. c) Riemannian Dilation (2.14) d) Riemannian Erosion (2.15) e) Riemannian Closing and f) Riemannian Opening. The operators are applied marginally.

Theorem 2.13. *If* $\mathcal{W}_{\mathbf{f}}$ *a morphological weight system on* $\mathbf{f}$ *then* $\delta_{\mathcal{W}_{\mathbf{f}}}(\mathbf{g}_1) \leq \mathbf{g}_2 \iff \mathbf{g}_1 \leq \varepsilon_{\mathcal{W}_{\mathbf{f}}}(\mathbf{g}_2)$, *for all* $\mathbf{f}, \mathbf{g}_1, \mathbf{g}_2 \in \mathcal{F}(\mathcal{L}, \mathcal{L})$

Finally, an extension when the Ω is a Riemannian manifold (13) is the following:

Definition 2.14. Let $\mathcal{M}$ a complete Riemannian manifold [1] and $d_{\mathcal{M}} : \mathcal{M} \times \mathcal{M} \to \mathbb{R}^+$, $(x, y) \mapsto d_{\mathcal{M}}(x, y)$, is the geodesic distance on $\mathcal{M}$, for any image $\mathbf{f} : \mathcal{M} \to \overline{\mathbb{R}}$, $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, so $\mathbf{f} \in \mathcal{F}(\mathcal{M}, \overline{\mathbb{R}})$ and for $\theta > 0$ we define for every $x \in \mathcal{M}$ the canonic *Riemannian dilation* of $\mathbf{f}$ of scale parameter θ as

$$\delta_\theta(\mathbf{f})(x) = \sup_{y \in \mathcal{M}} \left\{ \mathbf{f}(y) - \frac{1}{2\theta} d_{\mathcal{M}}(x, y)^2 \right\} \tag{2.14}$$

and the canonic *Riemannian erosion* of $\mathbf{f}$ of parameter θ as

$$\varepsilon_\theta(\mathbf{f})(x) = \inf_{y \in \mathcal{M}} \left\{ \mathbf{f}(y) + \frac{1}{2\theta} d_{\mathcal{M}}(x, y)^2 \right\} \tag{2.15}$$

An example for colored surfaces is illustrated in 2.3.

[1] A Riemannian manifold $\mathcal{M}$ is *complete* if starting at any point $p \in \mathcal{M}$, all geodesics are defined for all $t \in \mathbb{R}$.

A final example of adaptive morphology is included in the next section. In this case, the adaptability is induced by a mask image, called conditional morphology. The case of salt-and-pepper denoising is illustrated by means of a conditional shock filter.

2.6 Conditional Mathematical Morphology

In this section, the *conditional toggle mapping* is defined. That is the toggle mapping operator defined in (2.8) using an adaptive structuring element that varies in Ω according to a mask. This term is used as reference of the role played by the mask and should not be confused with the one in conditional geodesic dilation, where a conditional dilation is defined (in the binary case) as the intersection of the mask image with a dilation of the marker image, see for instance (247, Sec. 4.4). Let $\mathbf{m}$ be the characteristic function of a mask, *i.e.*, $\mathbf{m} \in \mathcal{F}(\Omega, \{0,1\})$ maps each pixel $x \in \Omega$ into $\{0,1\}$. Our approach is based on a neighbourhood associated with a structuring element $\mathtt{SE}$ and to the mask $\mathbf{m}$, denoted by N and defined as follows:

$$\forall x \in \Omega, \quad N_{(\mathtt{SE},\mathbf{m})}(x) = \{y \in \mathtt{SE}(x) \text{ and } \mathbf{m}(y) = 1\}.$$

That allows us to define the conditional version of the dilation as the following morphological mapping $\mathcal{F}(\Omega, \mathcal{L}) \to \mathcal{F}(\Omega, \mathcal{L})$.

Definition 2.15. The *conditional dilation* of an image $\mathbf{f}$ with respect to $\mathbf{m}$ is defined by

$$\forall x \in \Omega, \quad \delta_{\mathtt{SE}}(\mathbf{f}, \mathbf{m})(x) = \begin{cases} \bigvee\limits_{y \in N_{(\mathtt{SE},\mathbf{m})}(x)} \mathbf{f}(y) & \text{if} \quad x \notin \mathbf{m} \text{ and } N_{(\mathtt{SE},\mathbf{m})}(x) \neq \emptyset, \\ \mathbf{f}(x) & \text{otherwise.} \end{cases} \tag{2.16}$$

and similarly for the erosion,

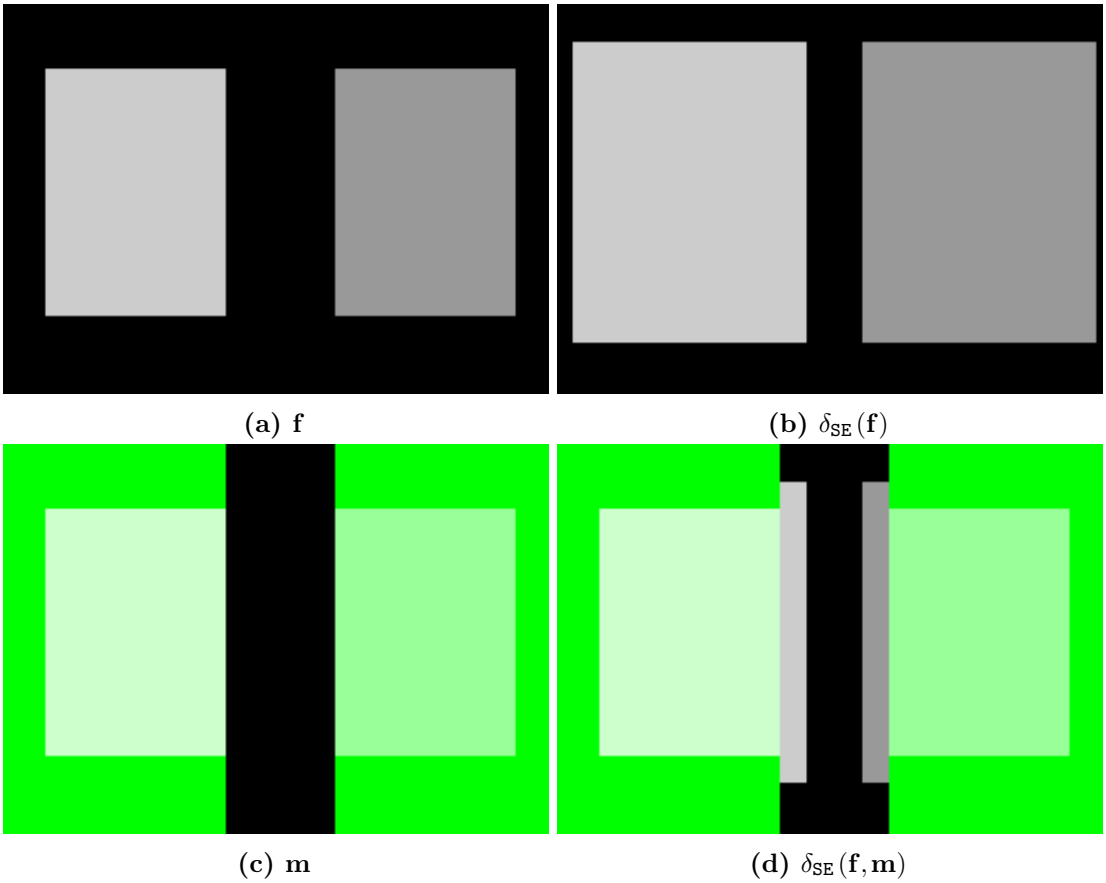
Definition 2.16. The *conditional erosion* of an image $\mathbf{f}$ with respect to a binary mask $\mathbf{m}$ is defined by the following expression:

$$\forall x \in \Omega, \quad \varepsilon_{\mathtt{SE}}(\mathbf{f}, \mathbf{m})(x) = \begin{cases} \bigwedge\limits_{y \in N_{(\hat{\mathtt{SE}},\mathbf{m})}(x)} \mathbf{f}(y) & \text{if} \quad x \notin \mathbf{m} \text{ and } N_{(\hat{\mathtt{SE}},\mathbf{m})}(x) \neq \emptyset, \\ \mathbf{f}(x) & \text{otherwise.} \end{cases} \tag{2.17}$$

A similar idea of conditional morphology was presented in (131) in the case of binary images. Expression 2.16 and 2.17 are equivalent to the ones introduced in (131) for binary images but differs in greyscale images due to the "otherwise" case. The motivation of this idea is that pixels in the mask are considered as sources in the morphological operation, and they are invariants (see Fig. 2.4 for an illustrative example).

We can now note that the pair $(\varepsilon_{\mathtt{SE}}(\cdot, \mathbf{m}), \delta_{\mathtt{SE}}(\cdot, \mathbf{m}))$ is not an adjunction 1.4as it is illustrated in Fig. 2.5.

However, we can calculate the algebraic adjunction of the conditional dilation. It is important because it produces a link in a unique way between morphological operators and idempotent filtering, achieved by composition of dilation/erosion (opening/closing). Additionally, it guarantees



(a) $\mathbf{f}$ **(b)** $\delta_{\mathtt{SE}}(\mathbf{f})$

(c) $\mathbf{m}$ **(d)** $\delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m})$

Figure 2.4 Conditional vs standard operators. $\mathbf{f}$ is a greyscale image of 175×245 pixels (a), $\mathtt{SE}$ is a square of side 25. Pixels of the mask are displayed in green (c). Note that objects in the mask are neither dilated nor eroded for the conditional operators.

that the composition reduces the information content. This kind of analysis has been carried out in morphological operators applied to images (108), graphs (112), pyramids (136), and curve evolution (136).

Proposition 2.17. *Let* $\tilde{\varepsilon}_{\mathtt{SE}}(\mathbf{f},\mathbf{m})$ *be defined by:*

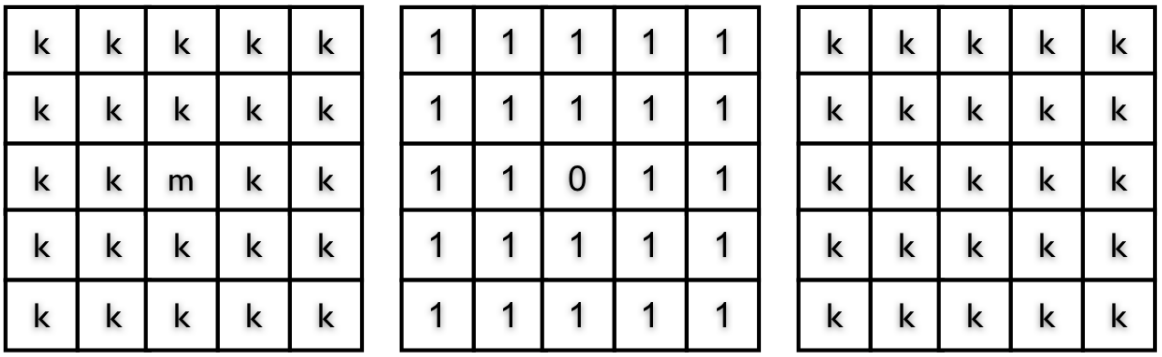
$$\forall x \in \Omega, \quad \tilde{\varepsilon}_{\mathtt{SE}}(\mathbf{f},\mathbf{m})(x) = \begin{cases} \bigwedge_{y \in N_{(\hat{\mathtt{SE}},\mathbf{m}^c)}(x)} \mathbf{f}(y) & \text{if } x \in \mathbf{m}, \text{ and } N_{\hat{\mathtt{SE}} \cap \mathbf{m}^c}(x) \neq \emptyset, \\ \mathbf{f}(x) & \text{otherwise,} \end{cases} \tag{2.18}$$

then the pair $(\tilde{\varepsilon}_{\mathtt{SE}}(\mathbf{f},\mathbf{m}), \delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m}))$ *is an adjunction.*

Proof. For every $\mathtt{SE} \in \Omega$ and a given $\mathbf{m} \in \mathcal{F}(\Omega,\{0,1\})$, the pair $(\tilde{\varepsilon}_{\mathtt{SE}}(\cdot,\mathbf{m}), \delta_{\mathtt{SE}}(\cdot,\mathbf{m}))$ defines an adjunction on $\mathcal{F}(\Omega,\mathcal{L})$ [Proposition 4.33, (110)]. In other words

$$\delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) \leq \mathbf{g} \iff \mathbf{f} \leq \tilde{\varepsilon}_{\mathtt{SE}}(\mathbf{g},\mathbf{m})$$

for all $\mathbf{f},\mathbf{g} \in \mathcal{F}(\Omega,\mathcal{L})$. Following the suggestion by Jos Roerdink (224) a simple proof is obtained based on the observation that the conditional dilation in equation 2.16 can be rewritten as a



(a) $\mathbf{f}$ **(b)** $\mathbf{m}$ **(c)** $\delta_{\mathtt{SE}}(\varepsilon_{\mathtt{SE}}(\mathbf{f},\mathbf{m}),\mathbf{m})$

Figure 2.5 Example showing that the pair conditional operators are not an adjunction in algebraic sense. If $k \neq m$, $\delta_{\mathtt{SE}}(\varepsilon_{\mathtt{SE}}(\mathbf{f},\mathbf{m}),\mathbf{m}) = \varepsilon_{\mathtt{SE}}(\delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m}),\mathbf{m})$ but not equal to $\mathbf{f}$, so $(\varepsilon_{\mathtt{SE}}(\cdot,\mathbf{m}),\delta_{\mathtt{SE}}(\cdot,\mathbf{m}))$ is not an adjunction 1.4. In the example, $\mathtt{SE}$ is a square of three pixels (8-connectivity).

space-adaptive dilation, *i.e.*,

$$\forall x \in \Omega, \delta_{\mathtt{SE}(\cdot,\mathbf{m})}(\mathbf{f})(x) = \bigwedge_{y \in N(x)} \mathbf{f}(y),$$

where the spatial-varying structuring element $N(x)$ is defined as (using the set notation for the mask $\mathbf{m}$):

$$\forall x \in \Omega, N(x) = \begin{cases} \mathtt{SE}(x) \cap \mathbf{m} & \text{if } x \notin \mathbf{m} \text{ and } \mathtt{SE}(x) \cap \mathbf{m} \neq \emptyset, \\ \{x\} & \text{otherwise.} \end{cases}$$

Then the space-varying adjoint erosion (223), written $\tilde{\varepsilon}_{\mathtt{SE}(;\mathbf{m})}$, is defined as

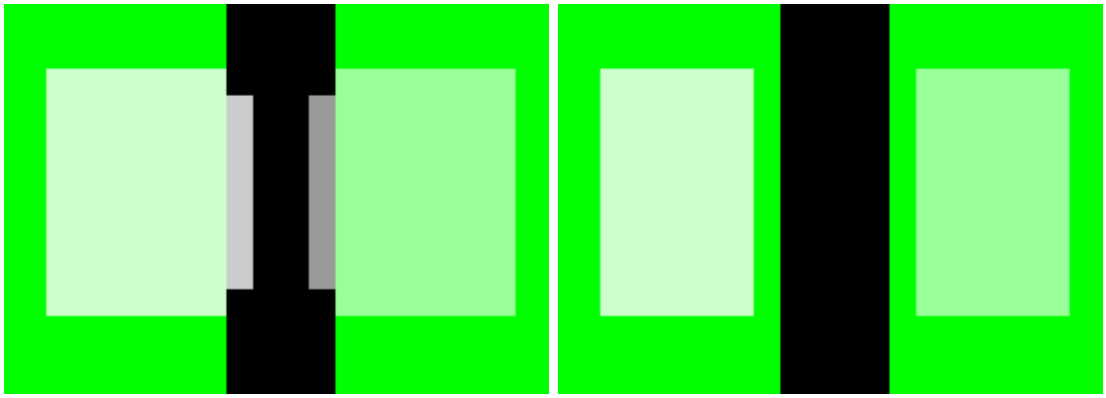
$$\forall x \in \Omega, \tilde{\varepsilon}_{\mathtt{SE}(\cdot,\mathbf{m})}(x) = \bigwedge_{y \in \check{N}(x)} \mathbf{f}(y)$$

where $\check{N}(x)$ is the reflected structuring element defined by $y \in \check{N}(x) \iff x \in N(y)$. Accordingly, it can be seen that

$$\forall x \in \Omega, \check{N}(x) = \begin{cases} \hat{\mathtt{SE}}(x) \cap \mathbf{m}^c & \text{if } x \notin \mathbf{m}^c \text{ and } \hat{\mathtt{SE}}(x) \cap \mathbf{m}^c \neq \emptyset, \\ \{x\} & \text{otherwise.} \end{cases}$$

which corresponds to the neighbourhood in 2.18. □

The adjoint operator only changes pixels on the mask $\mathbf{m}$ as it is illustrated in Fig. 2.6 for the same example of Fig. 2.4. However, in the practical applications considered in this paper the adjoint operator $\tilde{\varepsilon}_{\mathtt{SE}}(\cdot,\cdot)$ does not have any interest. Finally, we present a list of properties for the conditional morphological operators defined in 2.16 and 2.17. Let $\varepsilon_{\mathtt{SE}}(\cdot,\cdot)$, $\delta_{\mathtt{SE}}(\cdot,\cdot)$ be the pair of conditional operators. Let define the composition of conditional operation as follows, $\phi_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) = \varepsilon_{\mathtt{SE}}(\delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m}),\mathbf{m})$ and $\gamma_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) = \delta_{\mathtt{SE}}(\varepsilon_{\mathtt{SE}}(\mathbf{f},\mathbf{m}),\mathbf{m})$, Accordingly, the



(a) $\varepsilon_{\mathtt{SE}}(\mathbf{f},\mathbf{m})$ **(b)** $\tilde{\varepsilon}_{\mathtt{SE}}(\mathbf{f},\mathbf{m})$

Figure 2.6 $(\tilde{\varepsilon}_{\mathtt{SE}}(\mathbf{f},\mathbf{m}),\delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m}))$ is an adjunction whereas $(\varepsilon_{\mathtt{SE}}(\mathbf{f},\mathbf{m}),\delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m}))$ is not. $\tilde{\varepsilon}_{\mathtt{SE}}(\mathbf{f},\mathbf{m})$ only operates on the mask $\mathbf{m}$ in contrast with $\varepsilon_{\mathtt{SE}}(\mathbf{f},\mathbf{m})$ that performs outside of $\mathbf{m}$.

following properties hold:

$$
\begin{array}{ll}
(a) & \varepsilon_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) = \varepsilon_{\mathtt{SE}}(\varepsilon_{\mathtt{SE}}(\mathbf{f},\mathbf{m}),\mathbf{m}), \text{(idempotence)};\\
(b) & \delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) = \delta_{\mathtt{SE}}(\delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m}),\mathbf{m}), \text{(idempotence)};\\
(c) & \bigwedge_{i=1}^{i}\varepsilon_{\mathtt{SE}}(\mathbf{f}_i,\mathbf{m}) = \varepsilon_{\mathtt{SE}}\left(\bigwedge_{i=1}^{i}\mathbf{f}_i,\mathbf{m}\right), \text{(distributivity)};\\
(d) & \bigvee_{i=1}^{i}\delta_{\mathtt{SE}}(\mathbf{f}_i,\mathbf{m}) = \delta_{\mathtt{SE}}\left(\bigvee_{i=1}^{i}\mathbf{f}_i,\mathbf{m}\right), \text{(distributivity)};\\
(e) & \mathbf{f}\leq\mathbf{g}\Rightarrow\varepsilon_{\mathtt{SE}}(\mathbf{f},\mathbf{m})\leq\varepsilon_{\mathtt{SE}}(\mathbf{g},\mathbf{m}), \text{(increasingness)};\\
(f) & \mathbf{f}\leq\mathbf{g}\Rightarrow\delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m})\leq\delta_{\mathtt{SE}}(\mathbf{g},\mathbf{m}), \text{(increasingness)};\\
(g) & \delta_{\mathtt{SE}}(\cdot,\cdot)\text{and }\varepsilon_{\mathtt{SE}}(\cdot,\cdot)\text{ are morphological filters};\\
(h) & \gamma_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) = \delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m});\\
(i) & \phi_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) = \varepsilon_{\mathtt{SE}}(\mathbf{f},\mathbf{m});\\
(j) & \varepsilon_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) \leq \delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m});\\
(k) & \varepsilon_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) = \mathbf{t}_{max} - (\delta_{\mathtt{SE}}(\mathbf{t}_{max}-\mathbf{f},\mathbf{m})), \text{(duality)}.
\end{array}
$$

where, $\mathbf{t}_{max} = \max(\mathbf{f})$. Note that properties (a) and (b) are unusual, but they illustrate the mask effect included in $\mathbf{m}$. Additionally, note the fact that $\delta_{\mathtt{SE}}(\cdot,\cdot)$ (resp. $\varepsilon_{\mathtt{SE}}(\cdot,\cdot)$) is not extensive (resp. anti-extensive), even if the $\mathtt{SE}$ contains the origin. The demonstration of these properties is straightforward from the definition of conditional operators and therefore omitted. As discussed before, we keep the definition $\varepsilon_{\mathtt{SE}}(\cdot,\cdot)$ for the definition of the conditional toggle criterion and define the *conditional Laplacian* as follows:

$$\delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) = (\delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m})-\mathbf{f})-(\mathbf{f}-\varepsilon_{\mathtt{SE}}(\mathbf{f},\mathbf{m})).$$

Definition 2.18. The conditional toggle criterion based on $\mathbf{m}$ is defined as follows:

$$\forall x \in \Omega, \quad \tau_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) = \begin{cases} \varepsilon_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) & \text{if} \quad \delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) > 0, \\ \delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) & \text{if} \quad \delta_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) < 0, \\ \mathbf{f} & \text{otherwise.} \end{cases} \tag{2.19}$$

The motivation behind definition 2.19 is that the mask $\mathbf{m}$ plays the role of a seed indicator, where the pixel values spread through the image $\mathbf{f}$ according to the toggling criterion. Similarly to non-conditional toggle mapping, conditional toggle mapping should be applied iteratively. In this point, $\mathbf{m}$ has to spread their values through the image. Thus, we define a mapping $\tilde{\tau}_{\mathtt{SE}}(\cdot,\cdot)$ from and onto the pair image $\mathbf{f}$ and the mask $\mathbf{m}$ by taking 2.19 on $\mathbf{f}$ and a dilation on $\mathbf{m}$, *i.e.*, $\tilde{\tau}_{\mathtt{SE}}$ is a mapping $\mathcal{F}(\Omega,\mathcal{L}) \times \mathcal{F}(\Omega,\{0,1\}) \to (\mathcal{F}(\Omega,\mathcal{L}),\mathcal{F}(\Omega,\{0,1\}))$ such that

$$\tilde{\tau}_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) = (\tau_{\mathtt{SE}}(\mathbf{f},\mathbf{m}),\delta_{\mathtt{SE}}(\mathbf{m})). \tag{2.20}$$

Accordingly, the next iteration can be calculated as follows:

$$\begin{aligned} \tilde{\tau}^2_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) &= \tilde{\tau}_{\mathtt{SE}}(\tilde{\tau}_{\mathtt{SE}}(\mathbf{f},\mathbf{m})) \\ &= \tilde{\tau}_{\mathtt{SE}}(\tau_{\mathtt{SE}}(\mathbf{f},\mathbf{m}),\delta_{\mathtt{SE}}(\mathbf{m})) \\ &= (\tau_{\mathtt{SE}}(\tau_{\mathtt{SE}}(\mathbf{f},\mathbf{m}),\delta_{\mathtt{SE}}(\mathbf{m})),\delta_{\mathtt{SE}}(\delta_{\mathtt{SE}}(\mathbf{m}))) \\ &= (\tau_{\mathtt{SE}}(\tilde{\tau}_{\mathtt{SE}}(\mathbf{f},\mathbf{m})),\delta^2_{\mathtt{SE}}(\mathbf{m})) \end{aligned}$$

Finally, the *conditional toggle mapping* is defined by iteration until convergence $\tilde{\tau}^{\infty}_{\mathtt{SE}}$ as follows.

Definition 2.19. The conditional toggle mapping based on $\mathbf{m}$ is defined by

$$\forall x \in \Omega, \quad \tilde{\tau}^{\infty}_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) = \lim_{i\to\infty} \tilde{\tau}^{m}_{\mathtt{SE}}(\mathbf{f},\mathbf{m}), \tag{2.21}$$

where $\tilde{\tau}^{i}_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) = \tau_{\mathtt{SE}}(\tilde{\tau}^{i-1}_{\mathtt{SE}}(\mathbf{f},\mathbf{m}),\delta^{i}_{\mathtt{SE}}(\mathbf{m}))$ and $\tilde{\tau}^{0}_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) = (\mathbf{f},\mathbf{m})$.

We can also prove that the convergence of the conditional toggle mapping depends on the mask $\mathbf{m}$.

Proposition 2.20. *If $i^{\mathbf{D}} \geq \max(\mathbf{D}_{\mathtt{SE}}(\mathbf{m}))$ then $\tilde{\tau}^{i^{\mathbf{D}}}_{\mathtt{SE}}(\mathbf{f},\mathbf{m})$ converges to $\tilde{\tau}^{\infty}_{\mathtt{SE}}$, where $\mathbf{D}_{\mathtt{SE}}(\mathbf{m})$ is the distance transform of the binary image $\mathbf{m}$ associated with connectivity induced by the structuring element $\mathtt{SE}$ containing the origin.*

Proof. Let $\mathbf{m} : \mathbb{Z}^2 \to [0,1]$ be a binary image. The distance transform (DT) is the transformation that generates a map $\mathbf{D}$ whose value in each pixel x is the smallest distance from this pixel to $\mathbf{m}$, *i.e.*,

$$\forall x \in \Omega, \quad \mathbf{D}_{\mathtt{SE}}(\mathbf{m})(x) = \min\{i \in \mathbb{N} \mid x \in \delta^{i}_{\mathtt{SE}}(\mathbf{m})\}. \tag{2.22}$$

It is important to note that $\mathtt{SE}$ in 2.22 must contain the origin, otherwise the standard dilation $\delta_{\mathtt{SE}}(\mathbf{m})$ is not extensive and cannot be used to define a distance transformation $\mathbf{D}_{\mathtt{SE}}$. Let $i \geq i^{\mathbf{D}} = \max(\mathbf{D}_{\mathtt{SE}}(\mathbf{m}))$ so we have $\delta^{i+1}_{\mathtt{SE}}(\mathbf{m}) = \delta^{i}_{\mathtt{SE}}(\mathbf{m}) = \mathbf{1}$, where $\mathbf{1}$ is the indicator image of

$\mathbb{Z}^2$. Accordingly, by definition of conditional morphology, $\varepsilon_{\mathtt{SE}}(\mathbf{f},\mathbf{1}) = \delta_{\mathtt{SE}}(\mathbf{f},\mathbf{1}) = \tau_{\mathtt{SE}}(\mathbf{f},\mathbf{1}) = \mathbf{f}$. Therefore, $\tilde{\tau}^i_{\mathtt{SE}}(\mathbf{f},\mathbf{m}) = \tilde{\tau}^i_{\mathtt{SE}}(\mathbf{f},\mathbf{1}) = \tilde{\tau}^{i+1}_{\mathtt{SE}}(\mathbf{f},\mathbf{1}) = \tilde{\tau}^{i+1}_{\mathtt{SE}}(\mathbf{f},\mathbf{m})$. □

Proposition 2.20 means that expression 2.21 converges in no more iterations than $i^{\mathbf{D}} = \max(\mathbf{D}_{\mathtt{SE}}(\mathbf{m}))$, where $\mathbf{D}_{\mathtt{SE}}(\mathbf{m})$ is the distance transform of $\mathbf{m}$ with connectivity induced by the structuring element SE. That point is important because the non-conditional toggle mapping requires a large number of iterations until convergence.

Salt & pepper noise reduction An advantage of the conditional toggle mapping in 2.21, is that its convergence and mathematical properties are valid for any definition of the binary image $\mathbf{m}$. To illustrate this benefit, we introduce a mask definition for impulse noise removal applications. Ideally $\mathbf{m}$ should have zero values on pixels that are not corrupted by the impulse noise. We propose a simple noise detector as follows,

$$\forall x \in \Omega, \quad \mathbf{m}_N(x) = \begin{cases} 1 & \text{if} \quad \min(\Delta^{\varepsilon}_{\mathtt{SE}}(x), \Delta^{\delta}_{\mathtt{SE}}(x)) > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.23}$$

However, we note that other impulse detectors have been proposed in (122; 262; 300; 304) (see also the review (152)) and can be used for $\mathbf{m}_N$. Images contaminated with salt-and-pepper noise in range from 20% to 95%, with increment steps of 5% are considered. In the case of colour images, a channel-independent salt-and-pepper noise was simulated according to the following rule (37): the value of pixels was replaced by 0 with probability $\xi/2$ and replaced by 1 with probability $\xi/2$ with $\xi \in [0,1]$. A quantitative comparison of salt-and-pepper denoising methods has been presented in (284)

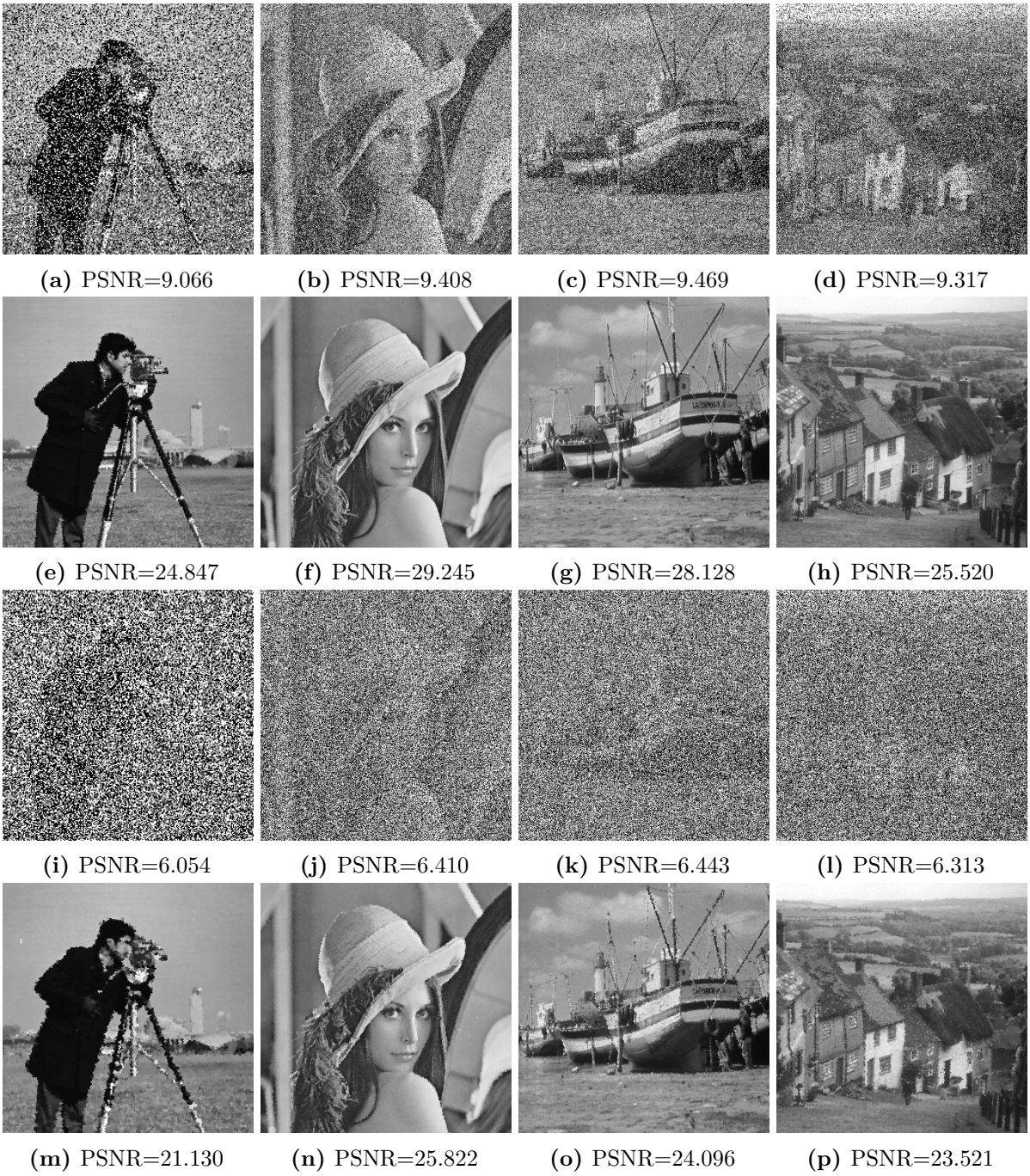
2.7 Perspectives

Methods based on adaptivity morphology allow the use of morphological operators in different types of structures (not only in images). A research topic that should be explored is the simultaneous learning of the adaptive neighbourhood for morphological operator in the sense of Chapter 4. For the particular case of conditional morphology, one can learn a mask of uncontaminated pixels, on which the conditional dilation can be applied [2]. An aspect that has not been presented in this document is the multiscale morphological decompositions that can serve to give a multilevel structure to characterise objet of the image. This is the subject of the recent work (238), where morphological based multiscale decompositions have been used for scale equivariant neural network.

Related Publications

1. **Tropical and morphological operators for signals on graphs**, S. Blusseau, S. Velasco-Forero, J. Angulo and I. Bloch, 25th IEEE International Conference on Image Processing (ICIP), pp. 1198-1202, 2018.
2. **Riemannian mathematical morphology**, J. Angulo and S. Velasco-Forero, Volume 47, Nro. 1, October 2014, 93–101, Pattern Recognition Letters.

[2] An example of this type of interaction is the main reason for the development of the library morpholayers



(a) PSNR=9.066 **(b)** PSNR=9.408 **(c)** PSNR=9.469 **(d)** PSNR=9.317

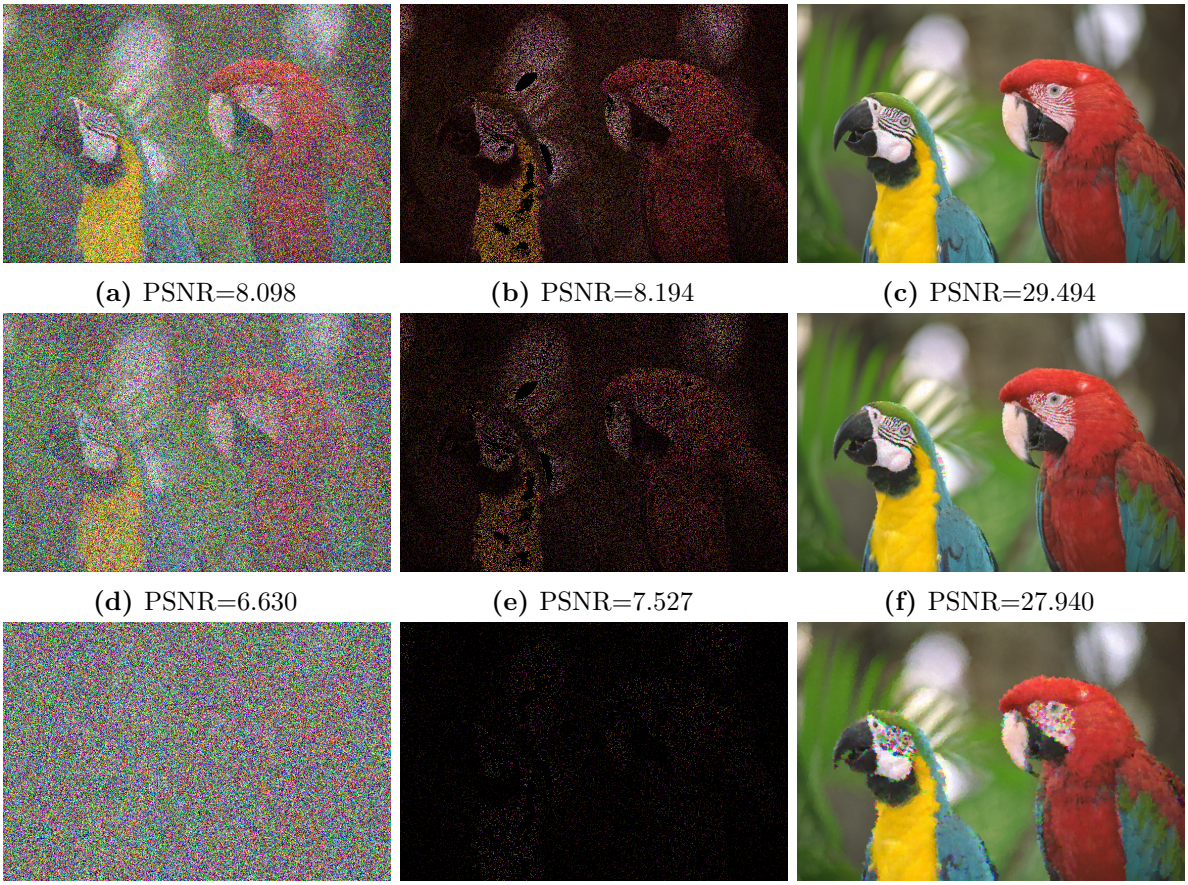
(e) PSNR=24.847 **(f)** PSNR=29.245 **(g)** PSNR=28.128 **(h)** PSNR=25.520

(i) PSNR=6.054 **(j)** PSNR=6.410 **(k)** PSNR=6.443 **(l)** PSNR=6.313

(m) PSNR=21.130 **(n)** PSNR=25.822 **(o)** PSNR=24.096 **(p)** PSNR=23.521

Figure 2.7 Example of noise removal by proposed method. Images contaminated by salt-and-pepper noise. (a)-(d) Noisy images (40%.) (e)-(h) Results restored by using the conditional toggle mapping. (i)-(l) Noisy images (80%.). (m)-(p) Image restored by the conditional toggle mapping.

3. **Morphological processing of univariate Gaussian distribution valued images based on Poincarè upper-half plane representation**, J. Angulo and S. Velasco-Forero, Geometric Theory of Information Signals and Communication Technology, pp. 331-366, 2014.



(a) PSNR=8.098 **(b)** PSNR=8.194 **(c)** PSNR=29.494

(d) PSNR=6.630 **(e)** PSNR=7.527 **(f)** PSNR=27.940

(g) PSNR=5.301 **(h)** PSNR=6.855 **(i)** PSNR=23.958

Figure 2.8 Example of noise removal by proposed method.(a)-(c) Experiment (50%). (d)-(f) Experiment (70%). (g)-(i) Experiments (95%.). First column shows images contaminated by marginal salt-and-pepper noise, second column includes mask image by using Eq. (2.23) and third column displays the result of the conditional toggle mapping.

4. **Morphological Semigroups and Scale-Spaces on Ultrametric Spaces**, J. Angulo and S. Velasco-Forero, ISMM 2017
5. **On nonlocal mathematical morphology**, S. Velasco-Forero and J. Angulo, Mathematical Morphology and Its Applications to Signal and Image Processing, LNCS, vol. 7883, pp. 219-230, 2013.
6. **Stochastic morphological filtering and Bellman-Maslov chains**, J. Angulo and S. Velasco-Forero, Mathematical Morphology and Its Applications to Signal and Image Processing, LNCS, vol. 7883, pp. 171-182, 2013.
7. **Structurally adaptive mathematical morphology based on nonlinear scale-space decompositions**, J. Angulo and S. Velasco-Forero, num. 30, 101-112, 2011, Image Analysis and Stereology.
8. **Conditional toggle mappings: principles and applications**, S. Velasco-Forero, J. Angulo and P. Soille, March 2014, Volume 48, Issue 3, pp. 544-565, Journal of Mathematical Imaging and Vision.

9. **Inner-Cheeger Opening and Applications**, S. Velasco-Forero, Mathematical Morphology and Its Applications to Signal and Image Processing, LNCS, vol. 9082, pp. 75-85, 2015
10. **Nonlinear Operators on Graphs via Stacks**, Geometric Science of Information, S. Velasco-Forero and J. Angulo, pp. 654–663, 2016.
11. **Scale equivariant neural networks with morphological scale-spaces**, 2021, International Conference on Discrete Geometry and Mathematical Morphology, pp. 483–495, 2021.

3 Connection based Mathematical Morphology

> We are like islands in the sea, separate on the surface but connected in the deep.
>
> William James

3.1 Introduction

A professional image analyst can be interested in finding objects in an image with a condition that depends on the local gradient of the image, for example an upper bound on the maximum value of the difference between the pixels of the found object, and/or an upper limit to the size of the object, and/or objects that have a circular shape. In mathematical morphology there are multiple structures that fully describe the image, such as: Tree-of-Shapes (184), Max-Tree (235), Minimum Spanning Forest (175), among others, where the object extraction procedure becomes a filter step from the component of the structure. This section starts with the definition of necessary elements for the definition of minimum spanning tree based hierarchies, and then presents some contributions in a modern approach including: a) A simplification for combining MST hierarchies with different geometrical conditions, b) A greedy algorithm for finding an optimal MST hierarchy c) An approach for Streaming a MST hierarchies d) An end-to-end learning process for MST-hierarchies by deep metric learning.

3.1.1 Notations

The notion of dissimilarity based connective segmentation, widely described in the literature (180; 193; 195; 247; 248; 252; 253), is recalled in the following subsection.

Let $\mathbf{f} : \Omega \to \mathbb{V}$ be a image and $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ an undirected simple finite graph with vertex set $\mathcal{V}$ matching the image pixels and edge set $\mathcal{E}$ consisting of unordered pairs of vertices indicating the adjacency relations between the image pixels. A path between two pixels x and y in $\mathcal{G}$ is a sequence of $m > 1$ pixels $\langle x = x_1, \ldots, x_m = y \rangle$ such that any two successive pixels of the sequence are adjacent: $\{x_i, x_{i+1}\} \in \mathcal{E}$ for $1 \leq i < m$. The image $\mathbf{f}$ assigns to each element of $\mathcal{V}$ (that is, the pixels), a vector of values.

The partition of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ into connected components relies on a function associating a *weight* to its edges. This function can be viewed as a measure of the degree of dissimilarity between adjacent vertices. For gray level images, the most common dissimilarity is the absolute difference. A dissimilarity measure involving a larger neighbourhood to prevent chaining through transitions while favouring it within homogeneous regions is presented in (253).

Definition 3.1. A *partition* $\mathtt{P}$ of the set $\mathcal{V}$ is a set of nonempty disjoint subsets of $\mathcal{V}$ whose union is equal to $\mathcal{V}$.

Definition 3.2. A *hierarchy*, $\mathcal{H}$ is a chain of nested partitions $\mathcal{H} = \{\mathtt{P}_0, \mathtt{P}_1, \ldots, \mathtt{P}_n | \forall j, k, \quad 0 \leq j \leq k \leq n \Rightarrow \mathtt{P}_j \sqsubseteq \mathtt{P}_k\}$, with $\mathtt{P}_n = \mathcal{V}$, the single-region partition, and $\mathtt{P}_0$ the finest partition on $\mathcal{V}$

Definition 3.3 (Tree and Forest)**.** A *tree* is a connected graph $\mathcal{G}$, that does not contain any cycle. A graph $\mathcal{FOREST}$ that contains no cycles is called *forest*. Each connected sub-graph of the forest is a tree.

Definition 3.4 (Spanning Tree)**.** Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a connected graph. A subgraph $\mathcal{TREE} = (\mathcal{V}', \mathcal{E}') \subseteq \mathcal{G}$ is a spanning tree for $\mathcal{G}$ if $\mathcal{TREE}$ is a tree and $\mathcal{V}' = \mathcal{V}$. More generally, a *spanning forest* $\mathcal{TREE}$ for a graph $\mathcal{G}$ is a forest that spans all the nodes of the graph $\mathcal{G}$, and such that each tree $\mathcal{TREE} \subseteq \mathcal{FOREST}$ in the forest is a spanning tree for a connected component in $\mathcal{G}$.

Definition 3.5 (Minimum Spanning Tree)**.** Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ be a connected weighted graph. A Minimum Spanning Tree (MST) of a $\mathcal{G}$, hereafter called $\mathcal{MST}(\mathcal{G})$, is a subgraph $\mathcal{TREE} = (\mathcal{V}', \mathcal{E}', \mathbf{W})$ such that:

i) $\mathcal{TREE}$ is a tree

ii) $\mathcal{TREE}$ spans all the vertices of $\mathcal{G}$, i.e. $\mathcal{V} = \mathcal{V}'$

iii) the sum of its weights $\sum_{e \in \mathcal{E}'} \mathbf{W}(e)$ is minimum among all the possible spanning trees.

In this document, we consider dissimilarity measures obtained through a dissimilarity function defined for any pair of value vectors.

Definition 3.6 ((101))**.** Let $\mathbb{V} = \mathbb{R}^p$ be the space of image value vectors. A dissimilarity function indexed by the intensity values of the image $\mathbf{f}$ is defined as any function $\mathtt{diss}$ from $\mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}_0^+$ such that $\mathtt{diss}(\mathbf{f}(x), \mathbf{f}(y)) = \mathtt{diss}(\mathbf{f}(y), \mathbf{f}(x))$, where x, y are two pixels of the image. This latter property is imposed by the symmetry property of dissimilarity measures.

The most natural dissimilarity functions are obtained by considering the norm of the difference of the input value vectors. Another common choice for multispectral images is the spectral angular distance (140).

Given a dissimilarity function $\mathtt{diss}$, the weight of an edge $\{x, y\}$ of the adjacency graph underlying an image $\mathbf{f}$ is denoted by $\mathbf{W}_{\mathtt{diss}}$: $\mathbf{W}_{\mathtt{diss}}(\{x, y\}) = \mathtt{diss}(\mathbf{f}(x), \mathbf{f}(y))$. Given a dissimilarity threshold λ, two distinct pixels x and y of an image $\mathbf{f}$ are $\mathtt{diss}_\lambda$-connected if there exists a path going from x to y such that the weight $\mathbf{W}_{\mathtt{diss}}$ between any two successive pixels of this path does not exceed the value of the dissimilarity threshold level λ. In addition, to ensure the reflexivity property of an equivalence relation, a pixel is always said to be $\mathtt{diss}_\lambda$-connected to itself.

Definition 3.7 ((101))**.** Let an image $\mathbf{f}$ be represented by a graph $\mathcal{G}$, where the edge weights are given by a dissimilarity function $\texttt{diss}(\mathbf{f}(x),\mathbf{f}(y)) = \mathbf{W}_{\texttt{diss}}(x,y)$, $(x,y) \in \mathcal{E}$. The *connected component* $\texttt{diss}_\lambda$-CC(x) of a pixel x is the set of pixels which are $\texttt{diss}_\lambda$-connected to this pixel:

$$\begin{aligned}\texttt{diss}_\lambda\text{-CC}(x) = \{x\} \cup \{y \mid \exists \langle x = x_1, \ldots, x_m = y\rangle,\\ \texttt{diss}(\mathbf{f}(x_i),\mathbf{f}(x_{i+1})) \leq \lambda, \forall\, 1 \leq i < m\}.\end{aligned} \tag{3.1}$$

The $\texttt{diss}_\lambda$-connectivity being an equivalence relation, it induces a unique partition of the image $\mathbf{f}$ support into maximal connected regions being the $\texttt{diss}_\lambda$-connected components (252). This approach for segmentation was put forward implicitly in (193) for the segmentation of multispectral images using the ℓ^1 norm (see also (18)). For grey level images, the ℓ^p norm boils down to the absolute difference. The corresponding connected components are called quasi-flat zones (179) or *lambda-flat zones* (303) in mathematical morphology. Because the dissimilarity value between adjacent pixels of a connected component can be arbitrarily large, the terminology α-connected components was suggested in (252; 253). Rather than using the notion of equivalence relation and path-based connectivity, Serra (247) proposes a more general approach to image segmentation based on the lattice theory and that does not require the definition of paths. With this theory, the existence of a maximum partition is secured if and only if the homogeneous classes form a connection based on the so-called connective criterion. The resulting segmentation was called a connection in (247) and lately referred to as a connective segmentation (226; 228). The Boolean connective criterion underlying the $\texttt{diss}_\lambda$-connectivity is detailed in (101; 202).

A fundamental property of the $\texttt{diss}_\lambda$-connected components of a pixel is that they form an ordered sequence (hierarchy) when increasing the dissimilarity threshold value λ (252):

$$\texttt{diss}_{\lambda_1}\text{-CC}(p) \subseteq \texttt{diss}_{\lambda_2}\text{-CC}(p), \forall \lambda_1 \leq \lambda_2. \tag{3.2}$$

This hierarchy is at the root of the greedy algorithm by Kruskal (141) for solving the minimum spanning tree problem and at the very basis of the dendrogram representation of the single linkage clustering (98). Cutting all edges of the $\mathcal{MST}(\mathcal{G})$ having a valuation superior to a threshold λ leads to a *minimum spanning forest* (MSF) $\mathcal{FOREST}(\mathcal{G},\lambda)$, *i.e.*, to a partition of the graph. Note that the obtained partition is the same that one would have obtained by cutting edges superior to λ directly on $\mathcal{G}$ [29]. Since working on the $\mathcal{MST}(\mathcal{G})$ is less costly and provides identical results regarding graph-based segmentation, we work only with the $\mathcal{MST}(\mathcal{G})$ in the remainder of this document. So cutting edges by decreasing valuations gives an *indexed hierarchy of partitions* $(\mathcal{H},\Lambda)$, with $\mathcal{H}$ a *hierarchy of partitions*, and $\Lambda : \mathcal{H} \to \mathbb{R}^+$ being a stratification index that corresponds to the *ultrametric distance* defining the hierarchy and verifying $\Lambda(\mathtt{P}) < \Lambda(\mathtt{P}')$ for two nested partitions $\mathtt{P} \subset \mathtt{P}'$. This process is otherwise called *single-linkage hierarchical clustering* (133). In that case, the internal nodes of $\mathcal{H}$ correspond to clusters of pixels at various levels of granularity.

This increasing map allows us to value each contour according to the level of the hierarchy for which it disappears: this is the *saliency* of the contour (corresponding to the ultrametric distance between the two regions it separates), and we consider that the higher the saliency, the

stronger the contour. For a given hierarchy, the image in which each contour takes as value its saliency is called *Ultrametric Contour Map* (UCM)(15) or *saliency map* (63).

We refer to a hierarchy built on a graph with edge weights expressing local contrast as to a *trivial* hierarchy. Whatever the intended use of hierarchical representations, for example the extraction of a segmentation out of a hierarchy (103; 137), the trivial hierarchy is usually not the more adapted one to work with in order to obtain the best results. It is thus interesting to look for more informative dissimilarities adapted to the content of images, so that the simplest methods are sufficient to obtain the desired results, for example computing interesting partitions. As these hierarchies are defined as ultrametric distances on a set of nodes, we can either aim at learning these ultrametrics (295) or at designing them in order for them to capture certain types of information. Several morphological hierarchical techniques exist to do the latter.

3.2 A variety of morphological hierarchies

Morphological hierarchies are representations capturing information across scales with an emphasis put on shape and size features. We hereby remind the reader of some of them known as watershed hierarchies, while insisting on the fact that approaches and methods proposed throughout the rest of this section can be used with any type of hierarchy.

Seminal works on morphological hierarchies include the dynamics hierarchy exhibiting contrasted regions (100) (and corresponding to the trivial hierarchy), or the area-based and volume-based watershed hierarchies (270) extending the dynamics hierarchy by taking into account sizes of regions as well. The waterfall hierarchy, first described in the context of a topographic surface flooding (31), has then been extended on graphs (178). The waterfall hierarchy highlights the nested structure of the catchment basins of a topographic surface. By flooding each catchment basin of a topographic surface up to the level of its lowest pass point, a new simpler topographic surface is produced, whose catchment basins result from the merging of catchment basins of the initial surface. The *stochastic watershed* (SWS), introduced in (11) on a simulation basis and extended with a graph-based approach in (177), is a versatile tool to construct hierarchies. The seminal idea is to operate multiple times marker-based segmentation with random markers and valuate each edge of the image by its frequency of appearance in the resulting segmentations. The output of the SWS algorithm is a hierarchy highlighting specific types of regions at different scales. It is very versatile as the type of markers spread, as long as the probabilistic law governing markers distribution, can be adapted for various tasks (80).

3.2.1 Sequential combinations of hierarchies through chaining

We have studied in (81) combination by chaining of morphological hierarchies. More specifically, from a given input image $\mathbf{f}$, a minimum spanning tree is computed $\mathcal{MST}(\mathcal{G}(\mathbf{f}))$ and operation of this tree can provide a large variability of hierarchies. We should note that we have limited ourself to a particular subset of hierarchies sharing the same edges than $\mathcal{MST}(\mathcal{G}(\mathbf{f}))$. However, new evaluation on edges can be performed by following rule inspired in classical morphological hierarchies. The new evaluation can be computed during the fusion of components, for instance via the *union-find data structure* in the Kruskal algorithm on the $\mathcal{MST}$(196). A pretty efficient

Hierarchy Name	Formula
Area Absorption	$\texttt{Area}_x \wedge \texttt{Area}_y$
Stochastic Area Absorption	$\theta_{x,y} = 1-(1-\frac{\texttt{Area}_x \wedge \texttt{Area}_y}{S})^K$
Stochastic Area Watershed	$\theta_{x,y} = 1-(1-\frac{\texttt{Area}_x}{S})^K-(1-\frac{\texttt{Area}_y}{S})^K+(1-\frac{\texttt{Area}_x - \texttt{Area}_y}{S})^K$

Table 3.1 Three examples of computation formula from area based morphological hierarchies.$K \in \mathbb{R}^+$ is a parameter called the *number of random markers*(81). $\texttt{Area}_x$ denoted the size of ultrametric ball and S a reference of the total area, usually equal to the number of pixels in the image.

implementation of different morphological hierarchies is available in (208). As example, the formula for some area-based hierarchies are given in Table 3.1(81). For an edge (x, y) the new evaluation depends on the area of their correspondent ultrametric ball. In the Stochastic Area Watershed, this evaluation is weighted according to the probability of having at least one over K random markers in ultrametric ball (81). In this way, one can include a priori information, about the shape of the objects that are desired to be highlighted in a hierarchy, since instead of punctual processes, one can have horizontal, vertical lines, or another information about the area or the volume of the region determined by the ultrametric ball.

3.2.2 Gromov-Hausdorff distance as feature

In the following experiment, we consider a classification problem on a set of simulated images from different dead leaves process (126; 170), namely five classes with 100 images each with different primary grains: circles, crosses, flowers, horizontal and vertical lines. In a dead leaves model, two dimensional textured surfaces (which are called "leaves" or "primary grains") are sampled from a shape and size distribution and then placed on the image plane at random positions, occluding one another to produce an image. It is well-known that such a model creates images which share many properties with natural images such as scale invariance and other statistical properties (214). For each of these images, we compute the following hierarchies: trivial, area-based SWS hierarchies with structuring elements of various sizes and forms (cross, circle, diagonals, horizontal and vertical lines), as long as logical combinations "AND" and "OR" of these SWS hierarchies. Then we generate as features the called *inter-hierarchy distance matrices* in (81), which is the Gromov-Hausdorff distance between hierarchies sharing the MST (*i.e.*

$$\mathrm{d_{GH}}(\mathcal{H}_1, \mathcal{H}_2) := \max_{(x,y)\in\mathcal{V}} |\max(\texttt{Path}_{\mathcal{H}_1}(x,y)) - \max(\texttt{Path}_{\mathcal{H}_2}(x,y))| \tag{3.3}$$

where $\texttt{Path}_{\mathcal{H}}(x, y)$ denote the values of the weights in the path from x to y in $\mathcal{H}$.

We can then use (3.3) as features in a linear support vector machines (SVM) to classify images of each class. We notice that the system can learn with very few examples how to discriminate properly these five classes. For comparison, we conduct the same experiment using a Convolutional Neural Network (CNN) with a two-layers architecture[1] without image augmentation for a fair comparison. In Fig. 3.1(f)(g) are represented for both experiments the

[1] (12 Conv2D + 12 Conv2D + MaxPooling2D(3×3) + Dropout(0.3)) + (24 Conv2D + 24 Conv2D + MaxPooling(3×3)+ Dropout(.5)) + Dense(NumClasses) + SoftMax. Categorical cross-entropy as loss function and adaptive gradient (Adagrad) as optimiser.

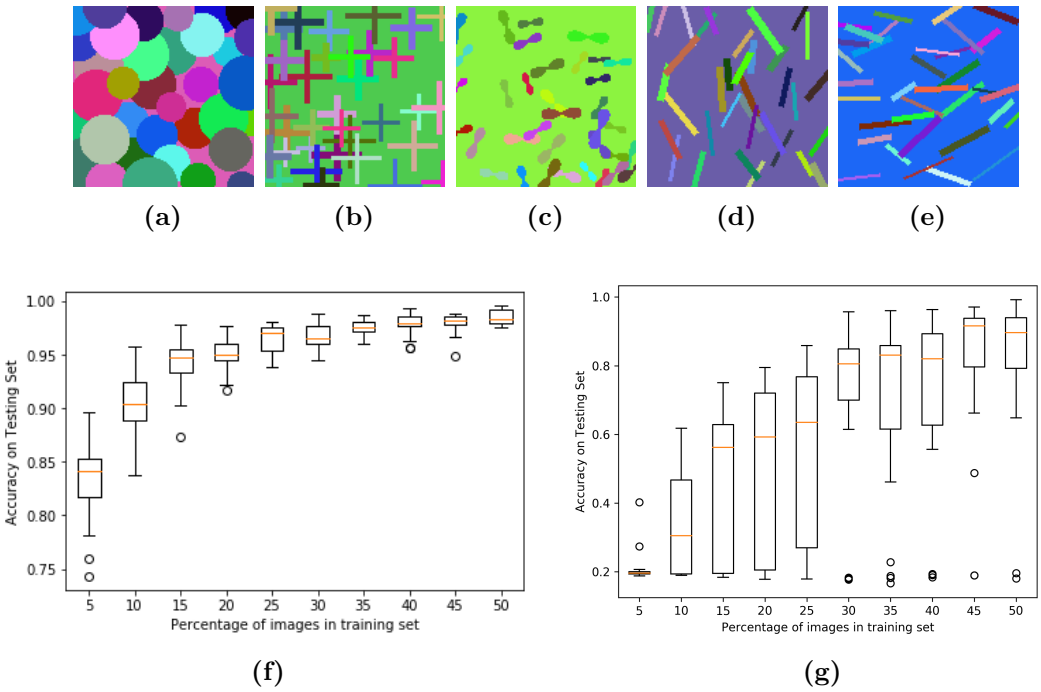
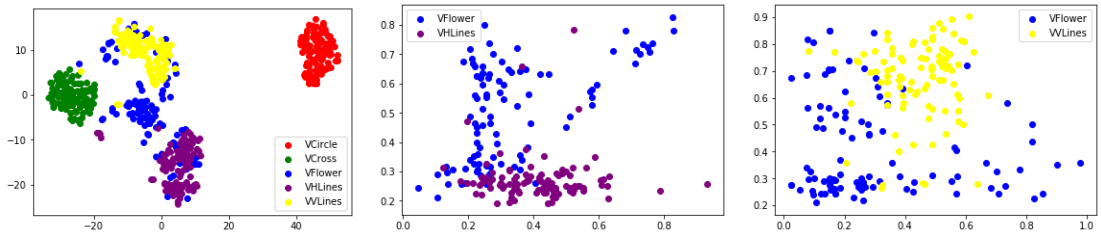


Figure 3.1 (a)-(e) False-colour representation of simulated images by dead leaves model with different primary grains. Linear SVM on proposed features. CNN[1] on proposed features (f)-(g) Accuracy vs the number of images in the training set for 25 repetitions.

evolutions of the average F-score with respect to the percentage of images used in the training set. In the first experiment (using the distance matrices as features), we notice that using only 5% of them (so 25 images out of 500) already leads to a 85% F-score over the remaining images, and that this figure quickly goes up. In the CNN experiment, the number of required training images to get to the same results is significantly larger (≈ 225). It is thus as if, on the contrary to CNN that have a black-box behaviour, our approach shows what is often referred to as an "aha moment"[2], i.e. a moment of sudden realisation and comprehension (301). This translates a form of understanding of the content of the image, which is corroborated by the study of the importance of which specific interhierarchy distances were the more useful to discriminate between two types of classes. For example, discriminating between horizontal and vertical lines will mainly be due to $\mathrm{d_{GH}}(\mathcal{H}_{surf-VertSE}, \mathcal{H}_{surf-HorizSE})$, while discriminating between crosses and circles will mainly be due to $\mathrm{d_{GH}}(\mathcal{H}_{surf-CrossSE}, \mathcal{H}_{surf-HexSE})$. A visualisation of the quality of the features space thus generated can be found in Fig. 3.2(a), where we project the features in a space of two dimensions using the t-SNE algorithm (155). Furthermore, using the variable selection method ℓ^1-SVM (308), we can isolate the more discriminative distances for two specific classes to separate. For example, the t-SNE visualisation in Fig. 3.2(a) shows us that discriminating between the classes "Flowers" and "Horizontal Lines" is not straightforward. The more discriminative variable between these two classes is the distance between $\mathcal{H}_{surf-VertSE}$ and

[2]Aha! moment also known as *eureka moment* or *eureka effect* refers to the common human experience of suddenly understanding a previously incomprehensible problem or concept. *Insight* or *Epiphany* is a psychological term that attempts to describe the process in problem solving when a previously unsolvable puzzle becomes suddenly clear and obvious



(a) (b) (c)

Figure 3.2 (a) t-SNE .We notice that the classes "Flowers" and "Horizontal Lines" are not well separated (b) ℓ^1-SVM. These two distances between hierarchies provide a geometrical understanding of the images content. Projecting along these features does indeed separate these classes efficiently. (c) The same can be done for example for the classes "Flowers" and "Vertical Lines"

$\mathcal{H}_{AND(surf-VertSE,surf-HexSE)}$: this is a geometrical interpretation of the image content, as they respectively capture straight lines and lines with a protuberance (i.e. flowers). Projecting the distances features onto the subspace of the two more discriminative variables properly separates these two classes, as can be seen in Fig. 3.2(b).

3.3 Looking for a good horizontal cut

Let us suppose we have at our disposal a $\texttt{score}(\mathbf{f},(\mathcal{H},\lambda))$[3] to judge the quality of a segmentation $(\mathcal{H},\lambda)$ obtained for an image $\mathbf{f}$. Note that $(\mathcal{H},\lambda)$ is the partition obtained after setting the value of the indexed hierarchy $(\mathcal{H},\Lambda)$ to λ (corresponding to a horizontal cut of the hierarchy). We would like to find the best hierarchy and the best cut level λ according to the score evaluated on a training set of images. Let us consider a training set $T=\{\mathbf{f}_1,\ldots,\mathbf{f}_{|T|}\}$ and a set of indexed hierarchies $\mathrm{H}=\{(\mathcal{H}_1,\Lambda_1),(\mathcal{H}_2,\Lambda_2),\ldots,(\mathcal{H}_{|\mathrm{H}|},\Lambda_{|\mathrm{H}|})\}$.

For any image, there is a best hierarchy and cut level that minimises the score, that we call *oracle*:

$$(\mathcal{H}^{\texttt{oracle}}(\mathbf{f}),\lambda^{\texttt{oracle}}(\mathbf{f})) := \underset{(\mathcal{H}\in\mathrm{H},\lambda\in\Lambda)}{\arg\min}\ \texttt{score}(\mathbf{f},(\mathcal{H},\lambda)). \tag{3.4}$$

Let us consider a set of homogeneous images, that we subdivide into training and testing subsets, and a set of indexed hierarchies H (possibly composition of hierarchies as in 3.2.1). During the training phase, we are interested in finding the hierarchy $\mathcal{H}$ and cut level λ that minimise the score on average over the whole set, *i.e.*,

$$(\mathcal{H}^*,\lambda^*) := \underset{(\mathcal{H},\lambda\in\Lambda)\in\mathrm{H}}{\arg\min} \sum_{i=1}^{|T|} \texttt{score}(\mathbf{f}_i,(\mathcal{H},\lambda)). \tag{3.5}$$

We call this learned hierarchy the *model* hierarchy.

To sum up, we follow a two-steps procedure, given a segmentation score and a set of homogeneous images :

[3] A description of the `score` function used in the experiments is given in (79)

1. For each image, we extract a wide variety of structured contours information using morphological hierarchies.
2. We select the best hierarchical segmentation and cut level among all possible ones using a greedy feedforward search.

To test the pertinence of this learned model, we compare its result, on each image of the test set, with the oracle model computed for this image.

One can say we have effectively found a good model hierarchy for the set of images if the difference between the scores obtained for the model (3.5) and the oracle (3.4) is on average low on the test subset.

(a) Some examples in the training set

(b) f **(c)** `model` **(d)** `oracle`

(e) f **(f)** `model` **(g)** `oracle`

(h) f **(i)** `model` **(j)** `oracle`

Figure 3.3 Results on some examples of Intrinsic Images in the Wild, for a WHDR score.(b),(e),(h) are images from the testing set, (c),(f),(i) the model segmentations and (d),(g),(j) the oracle segmentations.

3.4 Streaming of Hierarchies based on MST

In many applications, one can access partial information of the scene of interest. Which can update as the sensor, usually a satellite, travels through the area of interest. Usually the segmentation algorithms require to see the whole scene in order to determine the results. In the case of the morphological hierarchy, one can ask, if it is possible to have the hierarchy of interest with partial information, and furthermore, what are the elements that can be determined as stable, in the sense that they will not change with the appearance of new objects.

Let us introduce the streaming image problem. Consider the simple case of an image $\mathbf{f}$ decomposed in two blocks $\mathbf{b}_1, \mathbf{b}_2$ and sent one after another. Let $\mathbf{b}_1$ be the first block arriving. Suppose that we compute its MST. The question is, how to compute the MST for the whole image, $\mathbf{f} = \mathbf{b}_1 \cup \mathbf{b}_2$, when $\mathbf{b}_2$ arrives and exploit the information extracted from $\mathbf{b}_1$?

In (89), we introduce the formulation for stream a hierarchy on images. Let $\mathbf{f}_t$ be an image streaming over time. Without loss of generality, assume that new pixels come from one side of the image, for example the right side of the image. If $\mathbf{b}_t$ is the new block at time t, for $t = 0, \ldots, T$, we have $\mathbf{f}_t = \mathbf{f}_{t-1} \cup \mathbf{b}_t$. The last column of $\mathbf{f}_{t-1}$ is also the first column of $\mathbf{b}_t$.

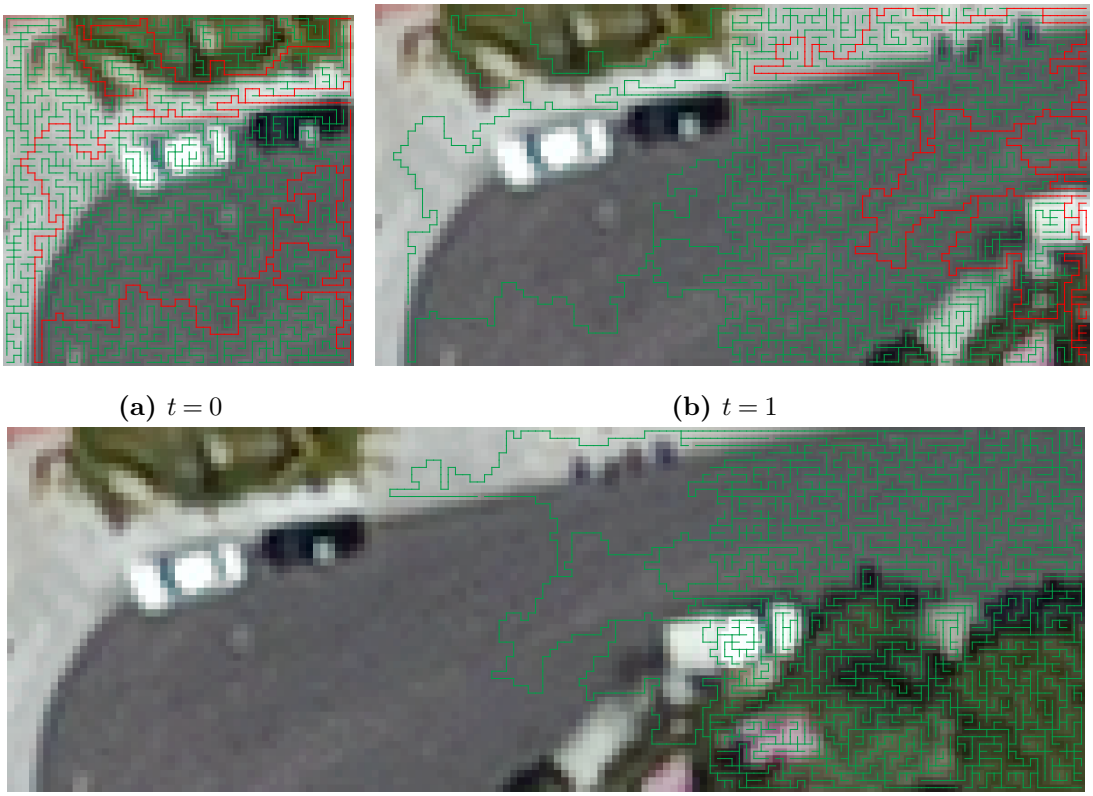
The first graph, is indeed made by edges that we call *unstable*. Mostly because we could eliminate some of them in the next step $t+1$. The second graph is made by edges that we call *stable*, since they will belong to all MSTs from now on. This is important for two reasons. 1) At each step, the memory footprint is reduced by discarding edges that are no longer necessary to compute further MSTs. 2) The *stable* edges can be used for further tasks, as we will see below. In Figure 3.4, we report an example that shows the evolution of the *stable + unstable* decomposition of minimum spanning trees through the time. In green, we represent the forest $\mathcal{FOREST}_t$ over time, while in red the graph $\mathcal{E}_{\mathcal{G}_t}$.

In (89), we exploited the *stable + unstable* decomposition of our methods to implement a streaming version of λ-quasi-flat zones (196)(303), watershed-cuts (62) and constrained connectivity (252) for large images. An example of streaming λ-quasi-flat zones is given in Fig. 3.5

3.5 End-to-End Similarity Learning and Hierarchical Clustering

Let $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ a set of points obtained as realisation of the k random variables. Moreover, (88) assume to be in a semi-supervised setting. Without loss of generality, we expect to know the associated labels of the first l points in $\mathbf{X}$. Each label takes value in $[k] = \{1, \ldots, k\}$, and indicates from which distribution the point has been sampled. In our work, we aim to obtain at the same time a good similarity function $\texttt{diss} : \mathbb{V} \times \mathbb{V} \to \mathbb{R}_+$ that permits us to discriminate the points according to the distribution they have been drawn and an optimal hierarchical clustering for each set $\mathbf{X}$. Our idea to achieve this goal is to combine the continuous optimisation framework proposed by Chami (49) along with deep metric learning to learn the similarities between points. Hence, we look for a function $\texttt{diss}_\theta : \mathbb{V} \times \mathbb{V} \to \mathbb{R}_+$ such that

$$\min_{\theta, Z \in \mathcal{Z}} C_{\text{HYPHC}}(Z, \texttt{diss}_\theta, \tau) + \texttt{loss}_{\text{triplet}}(\texttt{diss}_\theta; \alpha). \tag{3.6}$$



(a) $t = 0$ **(b)** $t = 1$

(c) $t = 2$

Figure 3.4 An example of *stable + unstable* decomposition of minimum spanning tree. The green graph is the forest $\mathcal{FOREST}_t$ that contains only *stable edges*, while the red graph is $\mathcal{E}_{\mathcal{G}_t}$ that contains only *unstable edges*. (b-c) Pixels without edges are stable, so is possible to store that part of the graph and do not need to consider in following intervals.

where the first term is a continuous version of Dasgupta's cost function (49). Let $Z = \{z_1, \ldots, z_n\} \subset \mathtt{B}^p$ be an embedding of a tree $\mathcal{TREE}$ with n leaves, they define their cost function as:

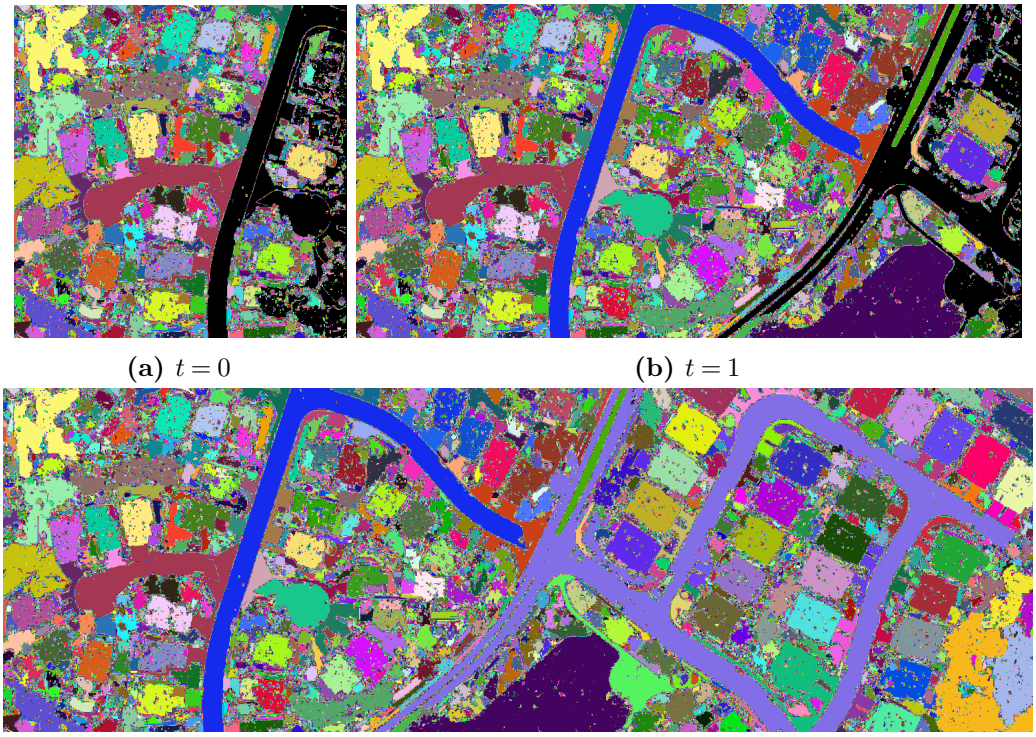
$$C_{\text{HYPHC}}(Z;\theta,\tau) = \sum_{ijk}(\theta_{ij} + \theta_{ik} + \theta_{jk} - \theta_{\text{HYPHC},ijk}(Z;\theta,\tau)) + \sum_{ij}\theta_{ij}, \tag{3.7}$$

where $\theta_{\text{HYPHC},ijk}(Z;\theta,\tau) = (\theta_{ij},\theta_{ik},\theta_{jk}) \cdot \sigma_\tau(d_o(z_i \vee z_j), d_o(z_i \vee z_k), d_o(z_j \vee z_k))^\top$, and $\sigma_\tau(\cdot)$ is the scaled softmax function $\sigma_\tau(\theta)_i = e^{\theta_i/\tau} / \sum_j e^{\theta_j/\tau}$. We recall that θ_{ij} are the pair-wise similarities, which in (49) are assumed to be known, but in this work are learned. The distance between two points in the Poincaré ball[4] $x, y \in \mathtt{B}^p$ is given by

$$d_{\mathtt{B}}(x,y) = \cosh^{-1}\left(1 + 2\frac{\|x-y\|_2^2}{(1-\|x\|_2^2)(1-\|y\|_2^2)}\right). \tag{3.8}$$

It is thus straightforward to prove that the distance of a point to the origin is $d_o(x) \coloneqq d(o,x) = 2\tanh^{-1}(\|x\|_2)$.

[4]The Poincaré ball is the Riemannian manifold $\mathcal{M} = (\mathcal{B}, d_p)$, where $\mathtt{B}^p = \{y \in \mathbb{R}^p : ||y|| < 1\}$ is the open p-dimensional unit ball. The distance function on $\mathcal{M}$ is defined as (3.8)



(a) $t = 0$ **(b)** $t = 1$

(c) $t = 2$

Figure 3.5 An example of one level of λ-quasi-flat zones in streaming, with $\lambda = 10$ for image in Fig. 3.4. Black pixels in Figures (a) and (b) are those that do not have a stable label in that iteration.

The second term of (3.6) is the sum over the set $\mathcal{T}$ of triplets:

$$\texttt{loss}_{\text{triplet}}(\texttt{diss}_\theta; \alpha) = \sum_{(a_i, p_i, n_i) \in \mathcal{T}} \max(\texttt{diss}_\theta(a_i, p_i) - \texttt{diss}_\theta(a_i, n_i) + \alpha, 0), \tag{3.9}$$

where a_i is the anchor input, p_i is the positive input of the same class as a_i, n_i is the negative input of a different class from a_i and $\alpha > 0$ is the margin between positive and negative values. One advantage of our formalism is that it allows us to use deep learning approach, *i.e.*, backpropagation and gradient descend optimisation to optimise the model's parameters. As explained before, we aim to learn a similarity function and at the same time find an optimal embedding for a family of point sets into the hyperbolic space which implicitly encodes a hierarchical structure. To achieve this, our idea is to model the function δ_θ using a neural network whose parameters we fit to optimise the loss function defined in (3.6). Our implementation consists of a neural network NN_θ that carries out a mapping $\text{NN}_\theta : \mathbb{V} \to \mathbb{R}^2$. The function $\texttt{diss}_\theta$ is thus written as:

$$\texttt{diss}_\theta(x, y) = \cos(\angle(\text{NN}_\theta(x), \text{NN}_\theta(y))), \tag{3.10}$$

We use the cosine similarity for two reasons. The first comes from the intuition that points belonging to the same cluster will be forced to have small angles between them. As a consequence, they will be merged earlier in the hierarchy. The second reason regards the optimisation process. Since the hyperbolic metric is *conformal* to the Euclidean metric, the cosine similarity allows

us to use the Riemannian[5] Adam optimiser (23) in (3.6). Once computed the similarities, the points are all normalised at the same length to embed them into the Hyperbolic space. The normalisation length is also a trainable parameter of the model. Accordingly, we have selected two architectures. The first is a Multi-Layer-Perceptron (MLP) composed of four hidden layers, and the second is a model composed of three layers of Dynamic Graph Edge Convolution (DGCNN) (292).

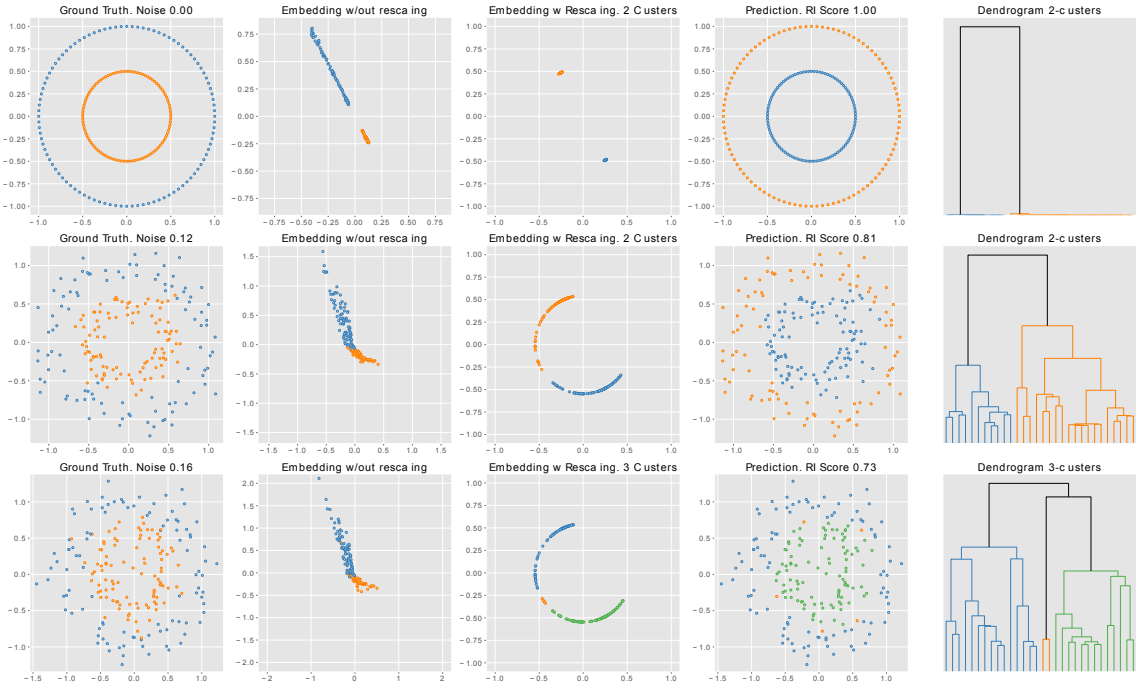


Figure 3.6 Effect of noise on predictions in the *circles* database. The model used for prediction is an MLP trained without noise. From top to bottom, each row is a case with an increasing level of noise. In the first column input points, while in the second column we illustrate hidden features. Points are coloured according to ground truth. The third column illustrates hidden features after projection to Poincaré Disk. The fourth column shows predicted labels, while the fifth column shows associated dendrograms.

3.6 Perspectives

This chapter has presented some of the work related to morphological hierarchical segmentation. There are still many open questions on this subject:

- It is possible to train networks to produce morphological hierarchical segmentation, this has been the subject of recent articles such as (49; 59), but the application to real problems is far from being validated.
- What kind of invariances should be sought in hierarchical methods? For example, is it possible to have hierarchical methods invariant to rotations or changes in scale?

[5]The update in epoch k is $\theta^{k+1} = \texttt{Ret}_{\theta^t}(-\eta \texttt{grad}(\mathcal{M}, \theta^t))$, where $\texttt{grad}(\mathcal{M}, \theta)$ denotes the Riemannian gradient with respect to θ, $\texttt{Ret}$ denotes a *retraction* from the tangent space of θ onto $\mathcal{M}$, and $\eta > 0$ is the learning rate.

- Is it possible to perform hierarchy streaming with low resolution images and then fine tune the accuracy if necessary by increasing the resolution?
- Additionally, how can one determine the best hierarchy in a video sequence with a predictive logic, that is, being able to predict the hierarchy in a frame that has not been seen yet?

Related Publications

1. **Local mutual information for dissimilarity based image segmentation**, L. Gueguen, S. Velasco-Forero and P. Soille, March 2014, Volume 48, Issue 3, pp 625-644, Journal of Mathematical Imaging and Vision.
2. **Automatic selection of Stochastic Watershed Hierarchies**, A. Fehri, S. Velasco-Forero and F. Meyer, EUSIPCO 2016
3. **Prior-based Hierarchical Segmentation Highlighting Structures of Interest**, A. Fehri, S. Velasco-Forero and F. Meyer, Mathematical Morphology-Theory and Applications, 3 (1), 29–44, 2019
4. **On minimum spanning tree streaming for hierarchical segmentation**, L. Gigli, S. Velasco-Forero and B. Marcotegui, Pattern Recognition Letters 138, 155-162, 2021
5. **End-to-End Similarity Learning and Hierarchical Clustering for unfixed size datasets**, Gigli, L., Marcotegui, B., and Velasco-Forero, S. (2021, July), In International Conference on Geometric Science of Information (pp. 596-604). Springer, Cham.

4 Links between Deep/Machine Learning and Mathematical Morphology

Opportunity and risk come in pairs

Bangambiki Habyarimana

4.1 Introduction

Deep learning has shown strong predictive accuracy in a wide range of applications. In particular, it has achieved and, in some cases, surpassed human-level performance on many cognitive tasks, for example, supervised classification, object detection and recognition, semantic and instance segmentation. This success can be attributed in part to the ability of a neural network (NN) to construct an arbitrary function by means of the composition of simple functions. Remarkably, the essence of deep learning is built from two simple algorithmic intuitions: first, the notion of feature learning, and second, learning by local gradient-descent, typically implemented as backpropagation. Another part of its success is due to its easy adaptation to different contexts, *i.e.*, adjusting the base operators (convolution, pooling, *etc.*) to different domains, such as graphs, texts, surfaces and so on.

Nowadays, state-of-the-art deep learning algorithms use convolution as their fundamental operation, in the so-called *convolutional neural network* (CNNs). Convolution has a long and proud history in signal/image processing, e.g. extracting low-level features like edges, noise filtering, frequency orientation filtering via Gabor among others. The CNN learns more and more features progressively in depth from the features learned in previous layers. However, (86) shows that CNNs are strongly biased toward recognition of texture over form, which shows fundamentally different classification strategies in comparison to human beings. Recently, (114) indicates that this bias can be reduced by an adequate image augmentation technique.

In the following subsection, an extremely short introduction of Convolutional NNs (CNNs) is presented.

4.1.1 Convolutional Neural Networks

The simplest form of a deep[1] neural network is the called *multilayer architecture*, which is a stack by composition of modules, each module implements a function $Z_L = F_L(\theta_L, Z_{L-1})$, where Z_L is a vector representing the output of module, θ_L is the vector of learnable parameters in the module, and Z_{L-1} is the module input vector (as well as the output of the previous module) and L is the number of *layers* in the neural networks. The input of the first module Z_0 is an input pattern, the output of the whole system is the one of the last module which denoted Z_l, where l is the *number of layers*. In *gradient-based learning methods*, given a objective function or *loss function*, $\texttt{loss}(\cdot,\cdot) \to \mathbb{R}^+$ measuring the discrepancy between the output of the system Z_l^k and D^k the "correct" or desired output for the k-th input pattern. One is interested on minimising the average discrepancy over a set of input/output pairs called the *training set*, $\{(Z_0^0, D^0), (Z_0^1, D^1), \ldots, (Z_0^n, D^n)\}$, where n is the number of samples in the training set. The network is initialised with randomly chosen weights θ^0. The gradient of the error function with respect to each parameter is computed and gradient descent is used to update the weights in each layer, *i.e.*, for the i-th iteration, $\theta^{i+1} = \theta^i - \eta \frac{\partial \texttt{loss}(\theta)}{\partial \theta^i}$ where η is a learning rate, and the computation of $\frac{\partial \texttt{loss}(\theta)}{\partial \theta^i}$, is performed by *backpropagation algorithm* through the layers (225). Additionally, for structured data like images, *convolutional neural networks* are nowadays the recommended solution. In CNNs, the same operator is computed in each pixel of the image. This mechanism is called *weight sharing*, and it has several advantages such as it can reduce the model complexity and make the network easier to train (200). Additionally, the weight sharing makes the network translation invariant and allows overfitting to be reduced, and it can be used in prediction on images of different sizes. For keen readers, some reference can be recommended (94; 144; 221)

4.2 Morphological Inspired Activation functions and Poolings

Introduction

The basic component in the NN introduced in previous subsection is the *linear perceptron* which is a linear combination of weights with biases followed by a nonlinear function called *activation function*, *i.e.*, $F(X) = \sigma(\theta^T X + \theta_b)$, where $\sigma \in \mathcal{F}(\mathbb{R}, \mathbb{R})$ is the *activation function*, and θ, θ_b are parameter to learn. The most famous activation function is the Rectified Linear Unit (ReLU) proposed by (194), which is simply defined as $\texttt{ReLU}(x) = \max(x, 0)$. A clear benefit of ReLU is that both the function itself and its derivatives are easy to implement and computationally inexpensive. However, ReLU has a potential loss during optimisation because the gradient is zero when the unit is not active. This could lead to cases where there is a gradient-based optimisation algorithm that will not adjust the weights of a unit that was never initially activated. An approach purely computational motivated to alleviate potential problems caused by the hard zero activation of ReLU, proposed a leaky ReLU activation (154): $\texttt{LeakyReLU}(x) = \max(x, .01x)$. A simple generalisation is the Parametric ReLU proposed by (105), defined as $\texttt{PReLU}_\beta(x) = \max(x, \beta x)$, where $\beta \in \mathbb{R}$ is a learnable parameter. In general, the use of *piecewise-linear functions* as

[1] The term *deep* is use when the number of layers, L, is larger than three.

activation function has been initially motivated by neurobiological observations; for instance, the inhibiting effect of the activity of a visual-receptor unit on the activity of the neighbouring units can be modelled by a line with two segments (104). On the other hand, for the particular case of structured data as images, a translation invariant DNN called Deep Convolutional Neural Networks (DCNN) is the most used architecture. In the conventional DCNN framework interspersed convolutional layers and pooling layers to summarise information in a hierarchical structure. The common choice is the pooling by a maximum operator called *max-pooling*, which is particularly well suited to the separation of features that are very sparse (40).

Morphological operators have been used in the context of DCNNs following the paradigm of replacing lineal convolutions by non-linear morphological filters (115; 123; 185; 219; 287), or hybrid variants between linear and morphological layers (116; 211; 261; 271). The contribution of this subsection is more in the sense of (83) where the authors show favourable results in quantitative performance for some applications when seeing the max-pooling operator as a dilation layer. However, we go further to study both nonlinear activation and max-pooling operators in the context of morphological representation theory of nonlinear operators.

Remark 4.1. ReLU activation function and max-pooling are dilation operators on the lattice of functions.

Theorem 4.2 ((161))**.** *Consider an upper semi-continuous operator ψ acting on an upper semi-continuous*[2] *function $\mathbf{f}$. Let $Bas(\psi) = \{g_i\}_{i\in I}$ be its basis and $Bas(\bar{\psi}) = \{h_j\}_{j\in J}$ the basis of the dual operator. If ψ is a translation invariant and increasing operator then it can be represented as*

$$\psi(\mathbf{f})(x) = \sup_{i\in I}(\mathbf{f}\ominus\mathbf{g}_i)(x) = \sup_{i\in I}\inf_{y\in\mathbb{R}^n}\{\mathbf{f}(x+y)-\mathbf{g}_i(y)\} \tag{4.1}$$

$$= \inf_{j\in J}(\mathbf{f}\oplus\check{\mathbf{h}}_j)(x) = \inf_{j\in J}\sup_{y\in\mathbb{R}^n}\left\{\mathbf{f}(x-y)+\check{\mathbf{h}}_j(y)\right\} \tag{4.2}$$

The converse is true. Given a collection of functions $\mathcal{B} = \{\mathbf{g}_i\}_{i\in I}$ such that all elements of it are minimal in $(\mathcal{B},\leq)$, the operator $\Psi(\mathbf{f}) = \sup_{i\in I}\{\mathbf{f}\ominus\mathbf{g}_i\}$ is a translation invariant increasing operator whose basis is equal to $\mathcal{B}$.

The extension to translation invariant *non necessarily increasing mappings* was presented by Bannon and Barrera in (17), which involves a supremum of an operator involving an erosion and an anti-dilation.

The previous theorem characterises increasing operator. However, to consider activation functions, we take advantage of the following results from (234)

Definition 4.3 ((234))**.** Let $\mathbf{f}\in\mathcal{F}(\mathbb{R}^n\to\mathbb{R})$ and $\mathbf{f}_1,\ldots,\mathbf{f}_m\in\mathcal{F}(\mathbb{R}^n\to\mathbb{R})$ be continuous functions. If $I(x) = \{i|\mathbf{f}_i(x) = (\mathbf{f})(x)\}$ is nonempty at every point $x\in\mathbb{R}^n$, then $\mathbf{f}$ is called a *continuous selection* of the functions $\mathbf{f}_1,\ldots,\mathbf{f}_m$. We denote by $CS(\mathbf{f}_1,\ldots,\mathbf{f}_m)$ the set of all continuous selections of $\mathbf{f}_1,\ldots,\mathbf{f}_m$. The set $I(x)$ is called the *active index set* of f at the point x.

[2] A function $\mathbf{f}:\mathbb{R}^n\to\bar{\mathbb{R}}$ is *upper semi-continuous* (u.s.c) (resp. lower semi-continuous (l.s.c.)) if and only if, for each $x\in\mathbb{R}^n$ and $t\in\bar{\mathbb{R}},\mathbf{f}(x)<t$ (resp. $(\mathbf{f})(x)>t$) implies that $\mathbf{f}(y)<t$ (resp. $\mathbf{f}(y)<t$) for all in some neighborhood of x. Similarly, $\mathbf{f}$ is u.s.c. (resp. l.s.c.) if and only if all its level sets are closed (resp. open) subsets of $\mathbb{R}^n$. A function is continuous iff is both u.s.c and l.s.c.

Definition 4.4 (Max-Min normal forms)**.** Max-Min type functions may be transformed to the following two normal forms:

$$\bigvee_{i\in 1}^{r}\bigwedge_{j\in M_i}\mathbf{f}_i(\text{max-min normal form}), \tag{4.3}$$

$$\bigwedge_{i\in 1}^{r}\bigvee_{j\in M_i}\mathbf{f}_i(\text{min-max normal form}), \tag{4.4}$$

where the index sets $M_i \subseteq \{1,\dots,M\}$ are such that $M_i \subseteq M_j$ if and only if $i=j$.

At this point, one can be interested in knowing if any continuous selection of affine functions admits a max-min representation, and under which conditions the corresponding normal forms are unique.

Theorem 4.5. *(234) Let $\mathbf{f}\in\mathcal{F}(\mathbb{R}^n,\mathbb{R})$ be a max-min-type continuous selection of affine functions $a_1^Tx+\alpha_1,\dots,a_m^Tx+\alpha_m$. If the vectors $a_1,\dots,a_m$ are affine independent, then both normal forms of $\mathbf{f}$ are unique.*

Remark 4.6. Let $\mathbf{f}\in \mathrm{CS}(a_1^T+\alpha_1,\dots,a_m^T+\alpha_m)$. Then $\mathbf{f}$ can be expressed as a max-min type selection.

Remark 4.7. Any continuous selection of linear functions is representable as the difference of two convex continuous selections of linear functions.

Experimental Section

Definition 4.8. (283) defines the *MorphoActivation* layers, by considering the activation functions and Pooling Morphological Operator, as one transformation as follows. $\mathbf{f}\mapsto \Psi^{\text{Morpho}}:\mathbb{R}^n\to\mathbb{R}^{n'}$ either by composition $[\pi\circ\sigma(\mathbf{f})](x)$ or $[\sigma\circ\pi(\mathbf{f})](x)$ as follows:

$$\Psi_1^{\text{Morpho}}(\mathbf{f}) = \bigwedge_{1\le j\le M}\left\{\delta_{R,b_j}^{\texttt{MaxPool}}\left(\bigvee_{1\le i\le N}(\beta_i^j f+\alpha_i^j)\right)\right\}, \tag{4.5}$$

$$\Psi_2^{\text{Morpho}}(\mathbf{f}) = \bigwedge_{1\le i\le N}\left\{\bigvee_{1\le j\le M}\left(\beta_i^j\delta_{R,b_i}^{\texttt{MaxPool}}(\mathbf{f})+\alpha_i^j\right)\right\}, \tag{4.6}$$

where

$$\begin{cases}\delta_{R,b_j}^{\texttt{MaxPool}}(\mathbf{f})(x)=\delta_{b_j}(\mathbf{f})(R\cdot x), \text{ with}\\ \\ \delta_{b_j}(\mathbf{f})(x)=(\mathbf{f}\oplus b_j)(x)=\bigvee_{y\in W}\{\mathbf{f}(x-y)+b_j(y)\}\end{cases}$$

In the context of an end-to-end learning DCNN, the parameters β_j, α_j and structuring functions b_j are learnt by backpropagation (287). The learnable structuring functions $\mathbf{b}_j$ play the same role as the kernel in the convolutions. Note that one can have $R=1$, the pooling does not involve downsampling.

Firstly, to illustrate the kind of activation functions that our proposition can learn, we use the MNIST dataset as a ten class supervised classification problem and an architecture

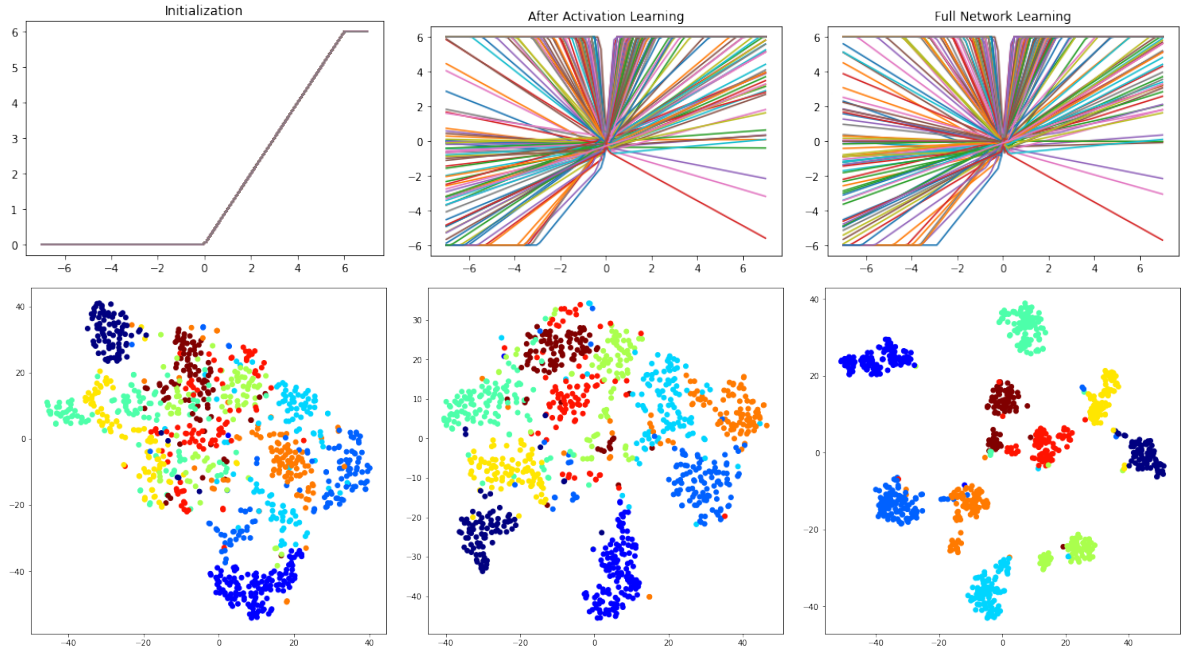


Figure 4.1 First Row: Left: Random initialisation with (14%) of accuracy on the test set, We use a simplified version of proposed activation $\min(\max(\beta_0 x+\alpha_0, \beta_1 x+\alpha_1, \alpha_2), \alpha_3)$, with initialisation $\max(\min(\texttt{ReLU}(x), 6), -6)$ Center: Training only activations (92.38%), Right: Training Full Network (98,58%). Second Row: t-SNE visualisation of last layer is the 10-classes MNIST prediction for a CNN.

composed of two convolutional layers and dense layer for reducing to the number of classes. The activation functions that we optimise by stochastic gradient descent have as general form $\min(\max(\beta_0 x+\alpha_0, \beta_1 x+\alpha_1, \alpha_2), \alpha_3)$, which corresponds to (4.5) and (4.6) where $R=1$, i.e. without pooling. We have initialised all the activations to be equal to $\max(\min(\texttt{ReLU}(x), 6), -6)$ as it is illustrated in Fig.4.1(left). The accuracy of this network without any training is 14%. Surprisingly when one optimises [3] *only* the parameters of activation functions the network accuracy increases to the acceptable performance of 92.38% and a large variability of activations are found Fig.4.1(center). This is a way to assess the expressive power[4] of the parameter of the activation as it has been proposed in (84). Additionally, an adequate separation among classes is noted by visualising the projection to two-dimensional space of the last layer via the t-SNE (155) algorithm. A much better accuracy(98,58%) and inter-class separation is obtained by optimising all the parameters of the network Fig.4.1(right).

Secondly, we compare the performance of (4.5) and (4.6) following the common practice and train all the models using a training set and report the standard top-one error rate on a testing set. We use as architecture a classical two-layer CNN (without bias for (4.5) and (4.6)) with 128 filters of size (3×3) per layer, and a final dense layer with dropout. After each convolution the different propositions are used to both produce a nonlinear mapping and reduce spatial dimension via pooling stride of two. For comparison, we include the case of a simple `ReLU` activation followed by a `MaxPool` with stride two. The difference in top-one error rate on

[3] We use ADAM optimiser with a categorical entropy as loss function, a batch size of 256 images and a learning rate of 0.001.

[4] The expressive power describes neural networks ability to approximate functions.

	Fashion MNIST			CIFAR10			CIFAR100		
`MaxPool(ReLU)`	93.11			78.04			47.57		
MorphoActivation in (4.5)	N=2	N=3	N=4	N=2	N=3	N=4	N=2	N=3	N=4
M=2	-0.06	-0.05	-0.1	-0.42	0.02	-0.02	0.44	0.7	0.4
M=3	-0.14	-0.14	-0.06	-0.57	-0.4	-0.35	0.56	0.49	0.61
M=4	-0.02	-0.08	-0.01	0.05	-0.62	-0.5	0.41	0.35	0.73
MorphoActivation in (4.6)	N=2	N=3	N=4	N=2	N=3	N=4	N=2	N=3	N=4
M=2	0.04	-0.16	-0.12	1.84	2.02	1.49	3.31	3.5	3.45
M=3	0.08	-0.09	**0.12**	2.39	1.96	1.82	3.48	3.55	**3.86**
M=4	-0.02	0.09	-0.03	**2.49**	2.25	2.13	3.47	3.73	3.58

Table 4.1 Relative difference with respect to our baseline (`ReLU` followed by a `MaxPool`). Architecture used is a CNN with two layers. ADAM optimiser with an early stopping with patience of ten iterations. Only Random Horizontal Flip has been used as image augmentation technique for CIFARs. The results are the average over three repetitions of the experiments.

a testing set is reported in Table 4.1 for CIFAR10, CIFAR100 and Fashion-MNIST databases. Additionally, (4.6) performs better than (4.5), and it improves the accuracy in comparison with our baseline in all the considered databases.

4.3 Max-plus Operator as a Morphological Unit

4.3.1 Introduction

In traditional literature on machine learning and neural networks, a perceptron (229) is defined as a linear computational unit, possibly followed by a non-linear activation function. Among all popular choices of activation functions, such as logistic function, hyperbolic tangent function and rectified linear unit (ReLU) function (93), this last generally achieves better performance due to its simple formulation and non-saturating property. Instead of multiplication and addition, the morphological perceptron employs addition and maximum, which results in a non-linear computational unit. A simplified version (51) of the initial formulation (67; 219) is defined as follows.

Definition 4.9. (Morphological Perceptron). Given an input vector $\mathbf{x} \in \mathbb{R}_{max}^n$ (with $\mathbb{R}_{max} = \mathbb{R} \cup \{-\infty\}$), a weight vector $\theta \in \mathbb{R}_{max}^n$, and a bias $\mathbf{b} \in \mathbb{R}_{max}$, the morphological perceptron computes its activation as:

$$\mathbf{f}(\mathbf{x}) = \max \left\{ \mathbf{b}, \bigvee_{i \in \{1,\dots,n\}} \{\mathbf{x}_i + \theta_i\} \right\} \tag{4.7}$$

where $\mathbf{x}_i$ (resp. θ_i) denotes the i-th component of $\mathbf{x}$ (resp. θ).

This model may also be referred to as $(\max, +)$ perceptron since it relies on the $(\max, +)$ semi-ring with underlying set $\mathbb{R}_{max}$. It is a dilation (1.3) on the complete lattice $((\mathbb{R} \cup \{\pm\infty\})^n, \leq_n)$ with $\leq_n$ the Pareto ordering.

4.3.2 Max-plus Block

Based on the formulation of the morphological perceptron, we define the Max-plus block as a standalone module that combines a fully-connected layer (or convolutional layer) with a Max-plus layer (51). Let us denote the input vector of the fully-connected layer[5], the input and output vectors of the Max-plus layer respectively by $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, whose components are indexed by $i \in \{1,...,I\}$, $j \in \{1,...,J\}$ and $k \in \{1,...,K\}$, respectively. The corresponding weight matrices are denoted by $\mathbf{w}^f \in \mathbb{R}_{max}^{I \times J}$ and $\theta^m \in \mathbb{R}_{max}^{J \times K}$. Then the operation performed in this Max-plus block is:

$$\begin{array}{rcl} \mathbf{y}_j & = & \sum_{i \in \{1,...,I\}} \mathbf{x}_i \cdot \theta_{ij}^f \\ \mathbf{z}_k & = & \bigvee_{j \in \{1,...,J\}} \left\{ \mathbf{y}_j + \theta_{jk}^m \right\} \end{array} \tag{4.8}$$

Note that the bias vector of the fully-connected layer (convolutional layer) is removed in our formulation, since its effect overlaps with that of the weight matrix θ^m. In addition, the bias vector of the Max-plus layer is shown to be ineffective in practice and is therefore not used here.

4.3.3 Universal Function Approximator Property

The result presented here is very similar to the approximation theorem on Maxout networks[6] (95), based on Wang's work (290). As shown in (95), Maxout networks with enough affine components in each Maxout unit are universal function approximators. Recall that a model is called a universal function approximator if it can approximate arbitrarily well any continuous function provided enough capacity. Similarly, provided that the input vector (or input feature maps) $\mathbf{y} \in \mathbb{R}_{max}^J$ of the Max-plus layer may have arbitrarily many affine components (or affine feature maps), we show that a Max-plus model with just two output units in its Max-plus block can approximate arbitrarily well any continuous function of the input vector (or input feature maps) $\mathbf{x} \in \mathbb{R}^I$ of the block on a compact domain.

Theorem 4.10. *(305) (*Universal function approximator*) A Max-plus model with two output units in its Max-plus block can approximate arbitrarily well any continuous function of the input of the block on a compact domain.*

Filter selection property for model pruning have been explored in (305).

4.4 Learning a Morphological Empirical Mode

4.4.1 Empirical Mode Decomposition (EMD)

EMD is an algorithm introduced by Huang et al. (121) for analysing linear and non-stationary time series. It is a way to decompose a signal in order to obtain instantaneous frequency data. In this original version of the EMD is an iterative process which decomposes real signals $\mathbf{f}$ into simpler signals (modes), $(\mathbf{f})(x) = \sum_{i=1}^{M} \Phi_j(x)$, where each *mono-component* signal Φ should be written in the form $\Phi(x) = r(x)\cos(\theta x)$, where the amplitude and phase are both physically

[5]This formulation can be easily generalised to the case of convolutional layers.

[6]Note that the classical universal approximation theorems for neural networks (see for example (118)) do not hold for networks containing max-plus units.

and mathematically meaningful (250). Unlike some other common transforms like the Fourier transform for example, the EMD was built as an algorithm and lacks theoretical background. The problem of EMD to represent a signal as a sum of amplitude modulation (AM) and frequency modulation (FM) components at multiple scales was first proposed in (164) where the problem of finding the AM-FM components and their envelopes was solved using multiscale Gabor filters and nonlinear Teager-Kaiser Energy Operators via an Energy Separation Algorithm (ESA). In the original EMD, there is no parametric family of filters used to estimate the envelopes.

An alternative characterisation of the EMD computation was introduced by Diop *et al.* in (70; 71) according to the definition of *local mean*, *i.e.*, the sifting process is fully determined by the sequence $(h_k)_{k\in\mathbb{N}}$ defined by :

$$\begin{cases} h_{k+1} = h_n - \Phi(h_k) = (\mathrm{Id} - \Phi)\, h_k \\ h_0 = \mathbf{f} \end{cases} \tag{4.9}$$

where $\Phi(h_k) = \frac{\hat{h}_k + \check{h}_k}{2}$, and $\hat{h}_k$ (resp. $\check{h}_k$) denotes a continuous interpolation of the maxima (resp. minima) of h_k.

The main motivation of this section is to define EMD learnable in the sense of neural networks approaches. Note that last property in remark 2.2 together with the extensivity/antiextensivity (*i.e.*, upper/lower envelopes) imply that the pair of operators $(\varepsilon_{\mathtt{SE}}, \delta_{\mathtt{SE}})$ are candidate functions for $(\hat{h}, \check{h})$ in 4.9. Accordingly, we proposed a simple generalisation by considering non-flat structuring functions.

Definition 4.11. The Morphological Empirical Mode (MEM) is defined as

$$\forall x \in \Omega, \quad \Phi_{\varepsilon,\delta,\mathtt{SE}}(\mathbf{f}) = \frac{\delta_{\mathtt{SE}}(\mathbf{f})(x) + \varepsilon_{\mathtt{SE}}(\mathbf{f})(x)}{2} \tag{4.10}$$

This operator can be formulated in any dimension (from 1D to nD signals) and avoid using an interpolation method which is the bottleneck of the original definition of EMD. The (4.10) with a flat structuring element, *i.e.*, $(\varepsilon_{B_\lambda}, \delta_{B_\lambda})$ has been proposed in (71).

Definition 4.12. The Flat Morphological Empirical Mode (71) is defined as

$$\forall x \in \Omega, \quad \Phi_{\varepsilon,\delta,\mathtt{B}_\lambda}(\mathbf{f})(x) := \frac{\delta_{\mathtt{B}_\lambda}(\mathbf{f})(x) + \varepsilon_{\mathtt{B}_\lambda}(\mathbf{f})(x)}{2} \tag{4.11}$$

The operator 4.11 was proposed to generate an EMD based on solving a morphological PDE (71).

Remark 4.13. Note that using 4.11 twice, the first residual 4.9 is $2(\mathbf{f} - \Phi_\lambda(\mathbf{f})) = (\mathbf{f} - \delta_{\mathtt{B}_\lambda}(\mathbf{f})) + (\mathbf{f} - \varepsilon_{\mathtt{B}_\lambda}(\mathbf{f})) = 2\mathbf{f} - \delta_{\mathtt{B}_\lambda}(\mathbf{f}) - \varepsilon_{\mathtt{B}_\lambda}(\mathbf{f})$. This expression, up to a minus sign, corresponds just to the so-called *morphological Laplace operator (275)*, and therefore provides an interpretation of the EMD as an iterated second-order derivative decomposition of the function $\mathbf{f}$.

One of the main advantages of EMD is that it can be considered as a parameter-free decomposition (255) and, for this reason, the inclusion of the structuring function and the parameter α can be seen as inconvenient. However, in the following, we consider EMD in the context of learning from data (151), where one would be interested in using EMD decomposition as

a preprocessing of an input signal before using machine learning or deep learning methods (24; 129; 215). Including any new layer, like EMD, requires therefore the computation of the corresponding gradient of the layer with respect to the parameters to be learnt.

4.4.2 Varying the Envelope

In this subsection, we explore several possibilities for $\check{h}, \hat{h}$. Additionally, one should note that the structuring elements SEs can be learned during the training stage.

Opening/Closing MEM

The theory of morphological filtering is based on the opening $\gamma_{\mathtt{SE}}(\mathbf{f})(x)$ and closing $\varphi_{\mathtt{SE}}(\mathbf{f})(x)$ operators, obtained respectively by the composition product of erosion-dilation and dilation-erosion, *i.e.*, $\gamma_{\mathtt{SE}}(\mathbf{f})(x) = \delta_{\mathtt{SE}}\left(\varepsilon_{\mathtt{SE}}(\mathbf{f})\right)(x)$ and $\varphi_{\mathtt{SE}}(\mathbf{f})(x) = \varepsilon_{\mathtt{SE}}\left(\delta_{\mathtt{SE}}(\mathbf{f})\right)(x)$. Opening (resp. closing) is increasing, idempotent and anti-extensive (resp. extensive), independently of the properties of the structuring function. The opening can be seen as the supremum of the invariant parts of $\mathbf{f}$ under-swept by SE and it can be again rewritten as a maximal lower envelope of structuring functions (resp. minimal upper envelope of negative symmetric structuring functions). We highlight that the *quadratic envelope* also called as *proximal hull* (47) is an opening with a quadratic structuring function, *i.e.*, a quadratic erosion followed by a quadratic dilation.

Definition 4.14. The opening/closing morphological empirical mode (OCMEM) is defined as a MEM where the pair $(\hat{h}, \check{h})$ corresponds to $(\gamma_{\mathtt{SE}}, \varphi_{\mathtt{SE}})$, *i.e.*,

$$\forall x \in \Omega, \quad \Phi_{\gamma,\varphi,\mathtt{SE}}(\mathbf{f})(x) = \frac{\gamma_{\mathtt{SE}}(\mathbf{f})(x) + \varphi_{\mathtt{SE}}(\mathbf{f})(x)}{2}. \tag{4.12}$$

For the case of flat disks B_λ, the operator 4.12 was called a morphological locally monotonic (LOMO) filter in (39). A signal is monotonic over an interval if it is either non-increasing or non-decreasing over that interval. A 1D signal is *locally monotonic* of degree n (LOMO-n) if and only if the signal is monotonic within every interval of length n. In the general case, a LOMO filter of $\mathbf{f}$ is defined as the fixed point of iterating $\Phi_{\gamma,\varphi,B_\lambda}(\mathbf{f})$, which is simultaneously idempotent to both the opening and closing by a flat disk as structuring function. Two examples of 4.12 for both flat and quadratic structuring function for the 1D signal with noise are shown in 4.3.

Lasry–Lions MEM

Definition 4.15. The Lasry-Lions morphological empirical mode (LLMEM) is defined as a MEM where the pair $(\hat{h}, \check{h})$ corresponds to $(\gamma^c_\lambda, \varphi^c_\lambda)$, *i.e.*,

$$\Phi_{\gamma,\varphi,c,\lambda}(\mathbf{f}) := \frac{\gamma^c_\lambda(\mathbf{f}) + \varphi^c_\lambda(\mathbf{f})}{2}. \tag{4.13}$$

An example of 4.13 for a 1D signal is shown in 4.3(c).

Definition 4.16. Let α be a real value with $0 \leq \alpha \leq 1$, the α-MEM based on the pair $(\check{h}, \hat{h})$ is defined as:

$$\Phi^\alpha_{\check{h},\hat{h}}(\mathbf{f}) = \alpha\hat{h}(\mathbf{f}) + (1-\alpha)\check{h}(\mathbf{f}). \tag{4.14}$$

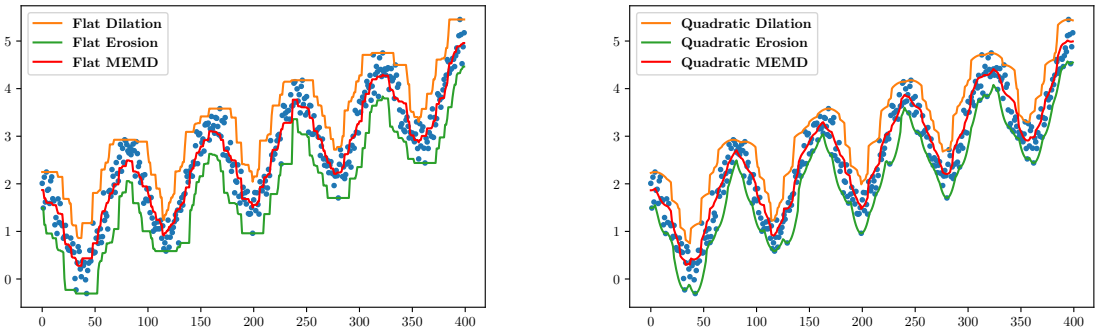


Figure 4.2 The blue points corresponds to the observed signal, a) Flat dilation/erosion based MEM 4.11 with a disk of $\lambda = 5$, b) Quadratic dilation/erosion based MEM with $\lambda = 3$.

In this context, an important fact to consider are the invariances of the operator 4.14.

Remark 4.17. For any $\mathtt{SE} \in \mathcal{F}(\Omega, \overline{\mathbb{R}})$, $\forall 0 \leq \alpha \leq 1$, and all the pairs $(\check{h}, \hat{h})$ previously considered, the operator 4.14 is increasing, invariant to translation, and the sifting process $\mathbf{f} - \Phi^{\alpha}_{\check{h},\hat{h}}(\mathbf{f})$ is invariant to additive intensity shifts, *i.e.*, $\forall c \in \mathbb{R}$ and $\forall \mathbf{f} \in \mathcal{F}(\Omega, \overline{\mathbb{R}})$,

$$((\mathbf{f})(x) + c) - \Phi^{\alpha}_{\check{h},\hat{h}}((\mathbf{f})(x) + c) = (\mathbf{f})(x) - \Phi^{\alpha}_{\check{h},\hat{h}}((\mathbf{f})(x)).$$

4.4.3 Derivatives of Morphological EMD in discrete domains

Derivative of dilation and erosion

Our approach involves dilation and erosion operators as defined in 2.2 and 2.3. However, in the discrete domain as it is the case of discrete images, the sup operator is computed via max. Consequently, for dilation operator 2.2, is computed by $\delta_\theta(x) = \max_\theta \{\mathbf{f}(x-y) + \mathtt{SE}_\theta(y)\}$. To understand how to compute the derivative of $\delta_\theta(x)$ with respect to θ, we rewrite $\delta_\theta(x) = \max_{y \in \mathtt{SE}_\theta} u(y)$. The max operator has no gradient with respect to non-maximum values, since changing them slightly does not affect the output. In general for rank operators, their derivative is zero in every coordinate, except for that of the value attending the desired rank (198; 210). Accordingly, the derivative with respect to a parameter in the additive structuring function is given by

$$\frac{\partial \delta_\theta(x)}{\partial \theta} = \frac{\partial \delta_\theta(x)}{\partial u(y)} \frac{\partial u(y)}{\partial \theta} = \begin{cases} \frac{\partial \mathtt{SE}_\theta(y)}{\partial \theta} & \text{if } y \in \arg\max_x \delta_\theta(x) \\ 0 & \text{otherwise} \end{cases} \tag{4.15}$$

where the operator $\arg\max_x(\mathbf{f})(x) := \{x \,|\, \forall y : \mathbf{f}(y) \leq (\mathbf{f})(x)\}$. In other words, argmax is the set of points x, for which $(\mathbf{f})(x)$ attains the largest value of the function. Note that we do not regard maximum as being attained at any x when $(\mathbf{f})(x) = \infty$, nor do we regard the minimum as being attained at any x when $(\mathbf{f})(x) = -\infty$. Similarly for the erosion, $\varepsilon_\theta(x) = \min_y[\mathbf{f}(x+y) - \mathtt{SE}_\theta(y)] =$

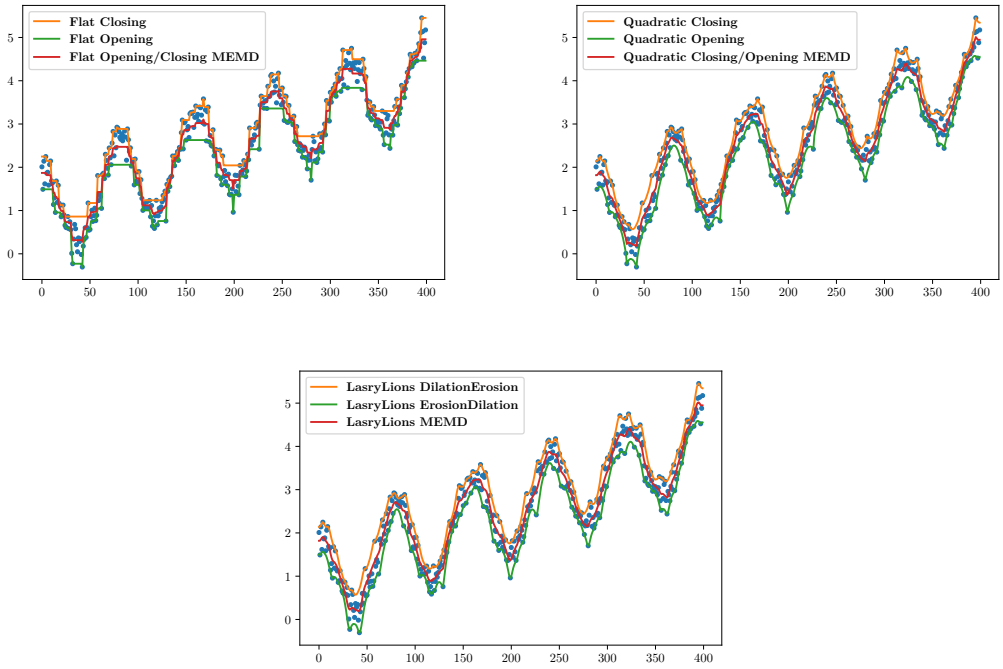


Figure 4.3 a) Flat OCMEM with a disk of $\lambda = 5$, b) Quadratic OCMEM with $\lambda = 3$ and c) Lasry-Lions MEM with $\lambda = 3$ and $c = .9$

$\min_{y \in \mathtt{SE}_\theta} u(y)$

$$\frac{\partial \varepsilon_\theta(x)}{\partial \theta} = \frac{\partial \varepsilon_\theta(x)}{\partial u(y)} \frac{\partial u(y)}{\partial \theta} = \begin{cases} -\frac{\partial \mathtt{SE}_\theta(y)}{\partial \theta} & \text{if } y \in \arg\min_x \varepsilon_\theta(x) \\ 0 & \text{otherwise} \end{cases} \tag{4.16}$$

there is only gradient with respect to minimum values. As example, for the dilation by quadratic structuring element 2.12, one has

$$\frac{\partial q_\theta(z)}{\partial \theta} = (2\theta^2)^{-1}\|z\|^2 \implies \frac{\partial \delta_\theta(x)}{\partial \theta} = \begin{cases} \frac{\|y\|^2}{2\theta^2} & \text{if } y \in \arg\max_x \delta_\theta(x) \\ 0 & \text{otherwise} \end{cases}$$

Experimental results

We will focus in the case of supervised classification of high-dimensional 1D signals in hyperspectral images. The architecture chosen as baseline is the one recommended in (204) and illustrated in 4.5. More specifically, the network is composed of convolution layers, RELU, max-pooling. Each stage consists of twenty convolution layers with a kernel size of 24 channels followed by ReLU activation, and a dense layer with batch normalisation. We explore the use of proposed EMDs as preprocessing layers, that means instead of learning the classification task from the original spectral signals, we will use the residual of a single step of the decomposition by MEMD. The parameters of the MEMD are learned in a gradient-based learning method. As it is common in supervised classification problems, we have used categorical cross-entropy as loss function.

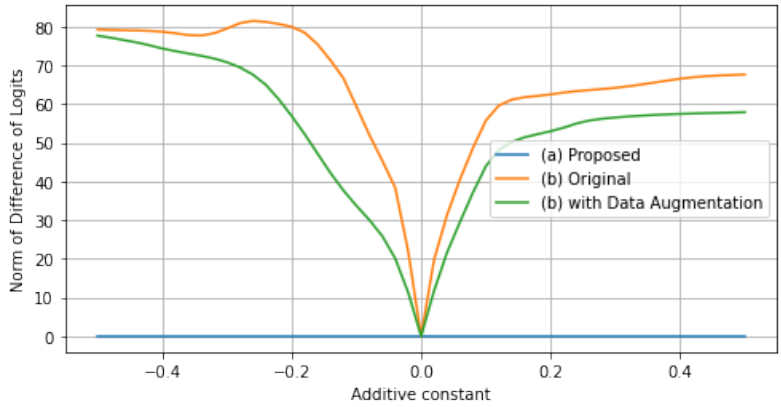


Figure 4.4 Analysis of invariance against additive shift for the training sample of Indian Pines. Norm of the Difference in the predictions with and without additive shift, *i.e.*, $||pred(x) - pred(x+c)||_2^2$ for different values of c is given for three models: a) MEMD by (ε, δ), b) baseline model, c) baseline model with a data augmentation by random additive constant. We highlight that by 4.17 all the MEMD based models are invariant to additive shifts.

Additionally, for quantitative comparisons, we have reported best, mean and standard deviation after ten repetitions on both Indian Pines HSI (4.2) and Pavia University HSI (4.4). In general, the following results can be highlighted:

- Learning the parameter in the α-MEM 4.14 improves the performance. This can be observed in 4.2) and Pavia University HSI (4.4) by comparing the performance of models trained with $\alpha = 0.5$ and models where this parameter is learned.
- Quadratic MEMDs perform significantly worse than non-flat ones. However, we would like to highlight that the number of parameters is less in the first case.
- In the considered HSI supervised classification problems, the best of the proposed approaches have a performance equivalent to our baseline, which is the state-of-the-art for the considered problems (Table. 4.5). However, we remark that the inclusion of morphological EMDs induces an invariant to additive intensity shifts in the classification model. To illustrate this fact, we have trained a classical model Fig. 4.5 with and without a random data augmentation by using an additive shift as transformation. That is the usual approach to include some invariance in deep learning models. This gives an improvement in the invariance measure in Fig. 4.4. We highlight that by 4.17 all the MEMD based models are invariant to additive shifts, which is illustrated in Fig. 4.4.

4.5 Geodesic Operations for DCNNs

In this section, we study morphological reconstruction in (2.5) as layer in Deep CNNs (288). We show they can be implemented as part of DCNN architectures, and how the backpropagation can be achieved during the neural network optimisation. To make easier the presentation of our results, we consider `SE` a flat structuring element and we considered functions having length n. However, for grey-scale images in 2D or 3D the implementation is equivalent by considering the connectivity induced by the `SE`s, and n the total number of pixels.

Layer (type)	Output Shape	Param #
InputLayer	(None, 103, 1, 1)	0
conv2d (Conv2D)	(None, 80, 1, 20)	500
max_pooling2d	(None, 16, 1, 20)	0
flatten	(None, 320)	0
dense	(None, 100)	32100
batch_normalization	(None, 100)	400
activation	(None, 100)	0
dense_1 (Dense)	(None, 9)	909

Total params: 33,909

Layer (type)	Output Shape	Param #
input (InputLayer)	[(None, 103, 1, 10)]	0
conv2d (Conv2D)	(None, 80, 1, 20)	4820
max_pooling2d	(None, 16, 1, 20)	0
flatten (Flatten)	(None, 320)	0
dense (Dense)	(None, 100)	32100
batch_normalization	(None, 100)	400
activation (Activation)	(None, 100)	0
dense (Dense)	(None, 9)	909

Total params: 38,229

Figure 4.5 Baseline architecture (204) vs Baseline architecture applied to EMD. The baseline uses a 20 convolutions 2D with a kernel size of $(24,1)$ followed by a max-pooling reduction of size $(5,1)$ and a RELU activation. For the case presented in the experimental section the same baseline architecture is used. In (b) is the same baseline architecture adapted for ten empirical modes.

4.5.1 Interpretation of Jacobian matrix

One should note, that the morphological reconstruction in (2.5) has not parameter to learn. Accordingly, we study what is its effect in the backpropagation by means of it Jacobian.

For a multivariate vector-valued function $\tau : \mathbb{R}^n \mapsto \mathbb{R}^n$, the Jacobian is a $n \times n$ matrix denoted by $\mathbf{Jac}_\tau$, containing all first order partial derivatives of the transformation τ. The row i of the Jacobian corresponds to the gradient of the i-th component of the output vector. It tells how the variation of each input variable affects the variation of the i-th component of the output of τ. We are interested in giving an interpretation for the Jacobian of reconstruction operation (2.5), because this is the fundamental element to understand the evolution process in gradient-based learning methods (92; 145). The geodesic reconstruction (2.5) is the composition of two operations, so below we describe the Jacobian of each of them. Firstly, for the minimum-wise operation,

$$\mathbf{Jac}_{\mathbf{f}\wedge\mathbf{g}}((\mathbf{f})(x)) = \begin{cases} 1 \text{ in } (i,i) & \text{if } \mathbf{f}(x_i) \leq \mathbf{g}(x_i) \\ 0 & \text{otherwise.} \end{cases} \tag{4.17}$$

Secondly, for the elementary dilation,

$$\mathbf{Jac}_{\delta_{\mathtt{SE}}}((\mathbf{f})(x)) = \begin{cases} 1 \text{ in } (i,j) & \text{for } x_j = \arg\max \delta_{\mathtt{SE}}(\mathbf{f}(x_i)) \\ 0 & \text{otherwise.} \end{cases} \tag{4.18}$$

In (4.17), the Jacobian in not null is pixels satisfying the condition *less or equal than*. In (4.18), the Jacobian indicates from which pixel comes the maximum values that the dilation has locally. We should note that in implementation by auto-differentiation in DL modules as Tensorflow or Pytorch, the Jacobian in (4.17) and (4.18) will have values different from zero *only* for the first element equal to the $\wedge$ or $\arg\max$ instead of the complete equivalence class. That is the same as local pooling by maximum are implemented nowadays. By using the chaining rule for the

Table 4.2 Experiment on hyperspectral *Indian Pines Disjoint* classification problem. Each experiment has been repeated ten times varying the initialisation of base architecture. Twenty filters of MEMD in a single level of simplification. The training was performed without any data augmentation technique. The constraint $\mathtt{SE}(0) \geq 0$ is used to assure the order relation among envelopes (See Theorem 2.2)

			Overall Val. Acc.		Overall Training Acc.	
Type	Operator	α	Best	$\mu \pm \sigma$	Best	$\mu \pm \sigma$
Baseline	—	—	85.035	83.929±0.654	93.443	91.413±1.696
NonFlat	(γ, φ)	.5	84.080	83.239 ± 0.512	97.012	95.495±1.184
		True	84.420	83.490 ± 0.656	97.223	96.012 ± 0.847
	(ε, δ)	.5	83.252	82.764 ± 0.576	97.451	95.226 ± 2.065
		True	**85.311**	84.052 ± 1.227	95.922	94.015 ± 2.717
	$(\varepsilon, \delta\ \mathtt{SE}(0) \geq 0$	.5	83.379	82.870 ± 0.261	96.889	95.621 ± 1.043
		True	85.247	83.821 ± 0.787	96.168	94.874 ± 1.120
Quadratic	(γ, φ)	.5	79.495	78.024 ± 0.754	96.080	93.580 ± 2.625
		True	80.959	77.971 ± 1.563	97.645	95.043 ± 1.565
	(ε, δ)	.5	81.363	79.798 ± 1.006	96.484	94.964± 1.111
		True	81.596	80.847 ± 0.537	97.223	95.066± 1.191
	Lasry-Lions	.5	81.384	79.909 ± 0.876	96.924	95.273 ± 1.336
		True	*82.424*	81.299 ± 0.983	96.941	95.674 ± 0.927

composition of functions, the Jacobian the geodesic dilation in (2.4) is

$$\mathbf{Jac}_{\delta^{(1)}(\mathbf{f},\mathbf{g})}(\mathbf{f}(x)) = \mathbf{Jac}_{\mathbf{f}\wedge\mathbf{g}}(\delta_{\mathtt{SE}}((\mathbf{f})(x)))\mathbf{Jac}_{\delta_{\mathtt{SE}}}((\mathbf{f})(x)) = \begin{cases} 1 \text{ in } (i,j) & \text{if } x_j = \arg\max \delta_{\mathtt{SE}}(\mathbf{f}(x_i)) \\ & \text{and } \delta_{\mathtt{SE}}(\mathbf{f}(x_i)) \leq \mathbf{g}(x_i) \\ 0 & \text{otherwise.} \end{cases} \tag{4.19}$$

To compute the Jacobian of (2.5), one should consider operation in convergence, that is, when it is idempotent. Firstly, the Jacobian with respect to the marker $\mathbf{f}$ has zero value, in values that come from dilation, creating flat areas that can be associated with a local maximum in $\mathbf{f}$. Specifically using the concept of *basin of attraction* [7], the Jacobian of (2.5) with respect to $\mathbf{f}$ is determined by

$$\mathbf{Jac}_{\mathtt{REC}^{\delta}(\mathbf{f},\mathbf{g})}((\mathbf{f})(x)) = \begin{cases} 1 \text{ in } (i,i) & \text{if } \mathbf{f}(x_i) = \mathtt{REC}^{\delta}(\mathbf{f},\mathbf{g})(x_i) \\ 1 \text{ in } (i,j) & \text{if } x_j \in \mathtt{BA}_{x_i}(\delta^{(1)}_{\mathbf{g}}(\mathbf{f})) \\ 0 & \text{otherwise,} \end{cases} \tag{4.20}$$

[7] The *basin of attraction* of a fixed point x_α for $\mathbf{f}$, denoted by $\mathtt{BA}_{x_\alpha}(\mathbf{f})$, is the interval $[a,b]$ if for all $x_0 \in [a,b] \Rightarrow \lim_{k\to+\infty} \underbrace{(f \circ f \circ \cdots \circ f)}_{k \text{ times}}(x_0) = \mathbf{f}^{(k)}(x_0) = f(x_\alpha)$, where k is such that $\mathbf{f}^{(k)}(x) = \mathbf{f}^{(k+1)}(x)$ for all $x \in \Omega$.

Table 4.3 Experiment on hyperspectral *Pavia University* for a *disjoint training sample*. Nine different classes. Each experiment has been repeated ten times varying the initialisation of base architecture. Twenty filters of MEMD in a single level of simplification. The training was performed without any data augmentation technique. The constraint $\texttt{SE}(0) \geq 0$ is used to assure the order relation among envelopes (See Theorem 2.2)

			Overall Val. Acc.		Overall Training Acc.	
Type	Operator	α	Best	$\mu \pm \sigma$	Best	$\mu \pm \sigma$
Baseline	—	—	85.468	83.396± 2.420	92.527	86.447 ± 8.960
NonFlat	(γ, φ)	.5	79.543	78.189 ± 0.726	95.715	92.219 ± 2.408
		True	82.353	79.293 ± 1.767	96.353	91.525 ± 4.335
	$(\varepsilon\delta)$	.5	84.261	82.681 ± 0.798	93.726	88.794 ± 4.998
		True	84.133	82.529 ± 1.131	93.879	89.735 ± 2.118
	(ε, δ), $\texttt{SE}(0) \geq 0$	.5	83.908	81.740 ± 1.295	93.216	84.575 ± 7.334
		True	**85.483**	83.994 ± 1.238	94.389	89.617 ± 3.289
Quadratic	(γ, φ)	.5	74.516	70.821 ± 2.023	91.201	80.951 ± 6.432
		True	73.539	69.399 ± 2.339	93.828	87.360 ± 6.443
	(ε, δ)	.5	77.411	75.432 ± 1.193	95.052	86.470 ± 5.939
		True	*81.196*	77.923 ± 1.700	92.476	86.593 ± 6.585
	Lasry-Lions	.5	77.461	76.396 ± 0.614	97.067	90.826 ± 6.223
		True	80.971	78.501 ± 1.332	96.123	87.082 ± 8.221

and equivalent with respect to the mask $\mathbf{g}$ is

$$\mathbf{Jac}_{\texttt{REC}^\delta(\mathbf{f},\mathbf{g})}(\mathbf{g}(x)) = \begin{cases} 1 \text{ in } (i,i) & \text{if } \mathbf{g}(x_i) = \texttt{REC}^\delta(\mathbf{f},\mathbf{g})(x_i) \\ 1 \text{ in } (i,j) & \text{if } x_j \in \texttt{BA}_{x_i}(\delta^{(1)}_{\mathbf{g}}(\mathbf{f})) \\ 0 & \text{otherwise.} \end{cases} \tag{4.21}$$

We highlight that the basin of attraction in both (4.20) and (4.21) are flat zones, *i.e.*, $x_j \in \texttt{BA}_{x_i}(\delta^{(1)}_{\mathbf{g}}(\mathbf{f})) \Rightarrow \texttt{REC}^\delta(\mathbf{f},\mathbf{g})(x_j) = \texttt{REC}^\delta(\mathbf{f},\mathbf{g})(x_i)$. For the marker, (4.20), these flat zones are related to some local maxima in $\mathbf{f}$, and in (4.21) are related to some local minima of $\mathbf{g}$. In other words, the flow of the gradient in the reconstruction layer depends on the membership of the basin of attraction, that is, for x_j from which x_i comes the value of the reconstruction, and if this value come from $\mathbf{f}$ or $\mathbf{g}$. Additionally, the number of ones in the i-th row, correspond to the cardinality of the basin of attraction of x_i. An illustrative example is given in Fig. 4.6. Thus, if one uses the geodesic reconstruction in a DL architecture, the backpropagation of the gradient pass through some maximum of $\mathbf{f}$ and some minimum of $\mathbf{g}$, and the gradient value will be proportional to the cardinality of the basin of attraction. As a final observation, the number of zero values in $\mathbf{Jac}_{\texttt{REC}^\delta_{\mathbf{g}}(\mathbf{f})}((\mathbf{f})(x)) + \mathbf{Jac}_{\texttt{REC}^\delta_{\mathbf{g}}(\mathbf{f})}(\mathbf{g}(x))$ is equal to n. Similarly, for the h-maxima transform (2.6), where both marker and mask depend on $\mathbf{f}$, the Jacobian indicates the position x_i from where each x_j has taken the value in convergence, *i.e.*, $\mathbf{Jac}_{\texttt{HMAX}_h(\mathbf{f})}((\mathbf{f})(x)) = 1$ in (i,j) if $x_j \in \texttt{BA}_{x_i}(\delta^{(1)}_{\mathbf{f}-h}(\mathbf{f}))$, and 0 otherwise.

Table 4.4 Experiment on hyperspectral *Pavia University* for a *disjoint training sample*. Nine different classes. Each experiment has been repeated ten times varying the initialisation of base architecture. Twenty filters of MEMD in a single level of simplification. The training was performed without any data augmentation technique. The constraint $\mathtt{SE}(0) \geq 0$ is used to assure the order relation among envelopes (See 2.2)

			Overall Val. Acc.		Overall Training Acc.	
Type	Operator	α	Best	$\mu \pm \sigma$	Best	$\mu \pm \sigma$
Baseline	—	—	85.468	83.396± 2.420	92.527	86.447 ± 8.960
NonFlat	(γ, φ)	.5	79.543	78.189 ± 0.726	95.715	92.219 ± 2.408
		True	82.353	79.293 ± 1.767	96.353	91.525 ± 4.335
	(ε, δ)	.5	84.261	82.681 ± 0.798	93.726	88.794 ± 4.998
		True	84.133	82.529 ± 1.131	93.879	89.735 ± 2.118
	(ε, δ), $\mathtt{SE}(0) \geq 0$	.5	83.908	81.740 ± 1.295	93.216	84.575 ± 7.334
		True	**85.483**	83.994 ± 1.238	94.389	89.617 ± 3.289
Quadratic	(γ, φ)	.5	74.516	70.821 ± 2.023	91.201	80.951 ± 6.432
		True	73.539	69.399 ± 2.339	93.828	87.360 ± 6.443
	(ε, δ)	.5	77.411	75.432 ± 1.193	95.052	86.470 ± 5.939
		True	*81.196*	77.923 ± 1.700	92.476	86.593 ± 6.585
	Lasry-Lions	.5	77.461	76.396 ± 0.614	97.067	90.826 ± 6.223
		True	80.971	78.501 ± 1.332	96.123	87.082 ± 8.221

Table 4.5 Comparison (in terms of OA) between different HSI classification models trained on spatial-disjoint samples. The performance for first four models are included for comparison from (204).

Method	Indian Pines	Pavia University
Random Forest	65.79	69.64
Multinomial Logistic regression	83.81	72.23
Support Vector Machines	85.08	77.80
MLP	83.81	81.96
CNN1D	85.03	85.47
$\Phi^{\alpha}_{\varepsilon,\delta}$+ CNN1D	85.31	85.48

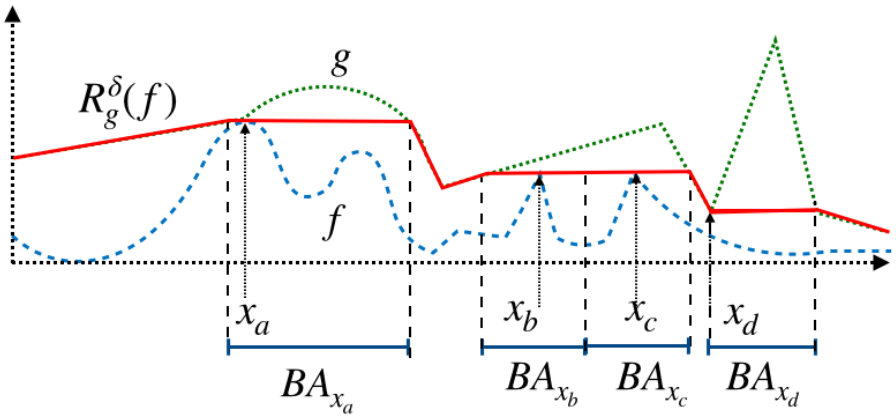


Figure 4.6 Basins of attraction with cardinality greater that one. $\mathtt{BA}_{x_a}, \mathtt{BA}_{x_b}$ and $\mathtt{BA}_{x_c}$ contribute to the Jacobian with respect to the mask $\mathbf{f}$ in (4.20) and are associated to local maxima of $\mathbf{f}$. The $\mathtt{BA}_{x_d}$ contributes to the Jacobian with respect to the marker $\mathbf{g}$ (4.21), and is associated to a local minimum of $\mathbf{g}$.

4.5.2 Experimental section

This subsection seeks to illustrate the benefits of using reconstruction operators in the context of CNNs. One can think that the geometrical understanding of the effect that the reconstruction layer has allows the design of architectures that are better adapted to specific problems. In this way two examples are presented: 1. The use of the reconstruction operator `REC` (2.5) in the case of networks that learn attributes about objects 2. The use of `HMAX` (2.6) and `RMAX` (2.7) to train a denoising layer which is robust to different levels and types of noise.

Learning geometrical attributes on simple objects

In this experiment, CNNs are trained in order to learn geometric attributes on images composed of simple geometric objects as it is shown in Fig 4.7. Each example is a random image with no overlap objects with size distributed as a uniform distribution between $[3, 20]$. As example, we consider the following geometrical attributes: Area, Perimeter, Area of Bounding-Box and Eccentricity[8]. Two models are trained[9] with the same number of parameters, but with the difference that a reconstruction of the last layer with the input image is used as prediction for the model denoted as $\texttt{CNN}_{\texttt{REC}}$. This has some benefits: 1) The result of the prediction is homogeneous within each of the objects due to the reconstruction process (Fig. 4.7) 2) According to the analysis of the Jacobian matrix, only the local maxima of the prediction will have an effect during learning, which should simplify the task. Note that each simple object forms a basin of attraction with size equal to the area of the object. Consequently, $\texttt{CNN}_{\texttt{REC}}$ homogenises the results of the `CNN` inside each object, reducing drastically the validation loss in Fig. 4.8

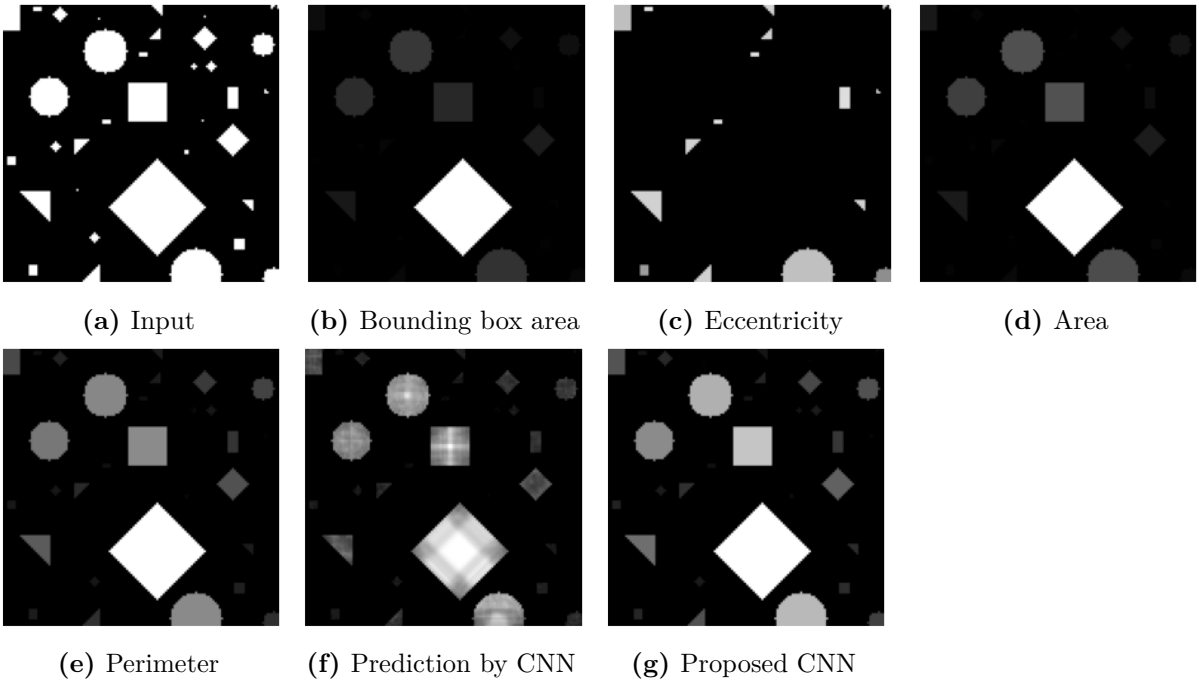
Denoising (Only one noise level and only one database)

Let us consider an image perturbed by additive positive random noise. This noise implies the creation of local maximum over the image. These peaks contain strong information about the noise distribution. The `HMAX` transform in (2.6) gives more importance in the gradient to fixed points, thus using information about noise itself. This allows you to learn the noise directly from the peaks. Accordingly, our experiments aim to show that it is possible to train a CNN to learn the parameter h following the noise strength. At the test stage, we analyse the capacity of generalisation for a wide range of dataset and noise levels. Our proposed model has two components (Fig. 4.9): 1) a reconstruction block, which for an input image f estimates a value of h with a CNN and computes the $\texttt{HMAX}_h(\mathbf{f})$ transformation on an input image; 2) A classification block, using training a CNN using as input the $\texttt{HMAX}_h(\mathbf{f})$.

At first impression one might think that it is enough to train the network in the supervised case, that is, using categorical cross entropy. However, you should train a large range of noise levels. To avoid this issue, we train on *only one noise level and only one database* (as an example, we have used the MNIST database). In all the following experiments, the ADAM optimiser has

[8] Eccentricity of the ellipse that has the same second-moments as the object. The eccentricity is the ratio of the focal distance (distance between focal points) over the major axis length. The value is in the interval [0, 1). When it is 0, the ellipse becomes a circle.

[9] The mean squared error is used as loss function, Adam optimiser, learning rate of 0.001, learning rate schedule by a factor of 0.1 with a patience of five epochs, and an early stopping with patience of ten epochs.



(a) Input **(b)** Bounding box area **(c)** Eccentricity **(d)** Area

(e) Perimeter **(f)** Prediction by CNN **(g)** Proposed CNN

Figure 4.7 a) Example of a random image containing geometrical shapes as Circles, Squares, Triangles and Rectangles. b) Bounding box Area c) Eccentricity d) Area e) Perimeter. f) Example of prediction for the attribute perimeter e) for a `CNN` in f) and the proposed $\texttt{CNN}_{\texttt{REC}}$ in g). Both trained models in f) and g) have the same number of parameters. $\texttt{CNN}_{\texttt{REC}}$ homogenises the results of the `CNN` inside each object, reducing drastically the validation loss in Fig. 4.8

been used during 50 epochs, with an initial learning rate of 0.001 and reducing it by a factor of 0.1 after ten epochs without improving on the validation loss. Models and code are available in the authors repository[10]. In the training stage of the reconstruction block (Fig. 4.9), we propose to minimise the following loss function, $\texttt{loss}(\mathbf{f},\hat{\mathbf{f}}) = ||\mathbf{f}-\hat{\mathbf{f}}||_2^2 + \alpha||\texttt{RMAX}(\mathbf{f})\mathbf{f} - \texttt{RMAX}(\hat{\mathbf{f}})\hat{\mathbf{f}}||_2^2$ denoting $\hat{\mathbf{f}} = \texttt{HMAX}_h(\mathbf{f})$. The first term is an *attached to data*, the second one looks for local large maxima to be preserved after `HMAX` transformation and α is a hyperparameter to trade-off the effect of each term ($\alpha = 0.75$ in our experiments). The pertinence of our proposition can be observed in Figure 4.12 when we have used the reconstruction block trained only in MNIST to estimate the value of h on different image databases at different noise levels (which have not been shown in the training phase). The estimated value of h follows the noise level, in pretty different datasets including CIFAR10, 91IMAGES(302) and BSD300 (169). We highlight that for colour images the reported h is the average of the estimation channelwise. For a quantitative comparison, we illustrated the impact of the `HMAX` layer in a classification network, and show that it can provide a better robustness through noise than a classic CNN classification network. For results shown in Fig. 4.10, reconstruction block has been pretrained on MNIST, and only the classification block is updated by a categorical cross-entropy loss function. In Fig. 4.10, we have also included a training with data augmentation by the type of noise with σ between 0 and 0.2 for fair comparison. In most of the explored scenarios, the proposed model is more robust than classical and data augmentation approaches.

[10] https://github.com/Jacobiano/morpholayers

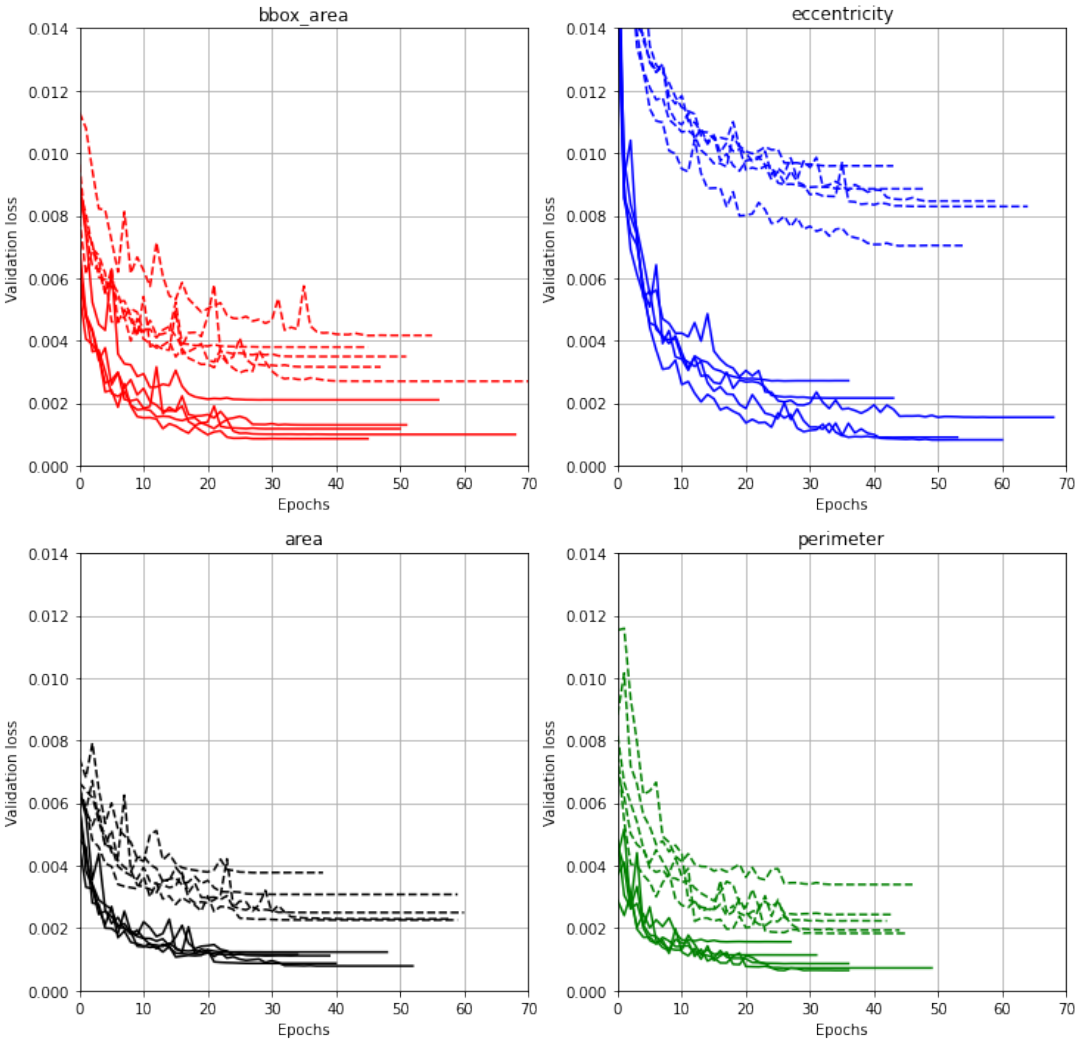


Figure 4.8 Several repetitions of the training protocol by varying the random initialisation. Dotted lines correspond to traditional CNN, and plain lines to $\texttt{CNN}_{\texttt{REC}}$, i.e, with a reconstruction as the last layer. Note that CNN and $\texttt{CNN}_{\texttt{REC}}$ models have the same number of parameters. In the explored scenarios, the inclusion of the geodesic reconstruction helps to converge to a model with lower loss on the validation set.

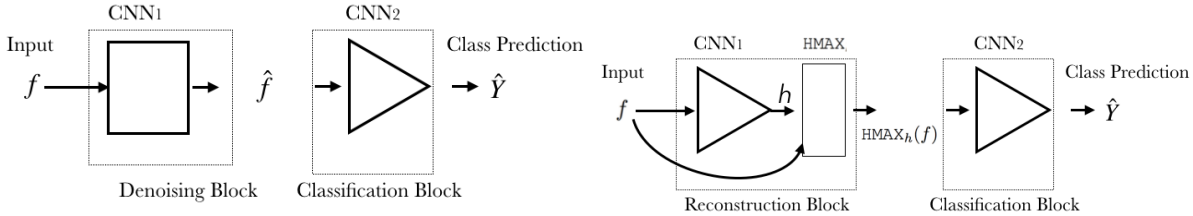
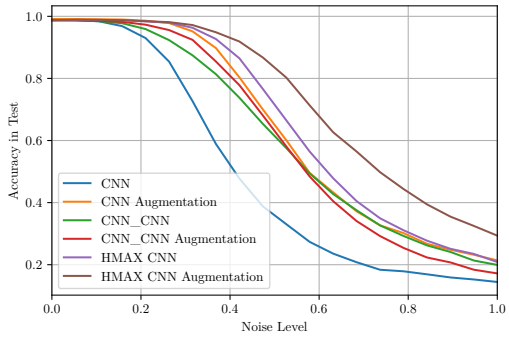
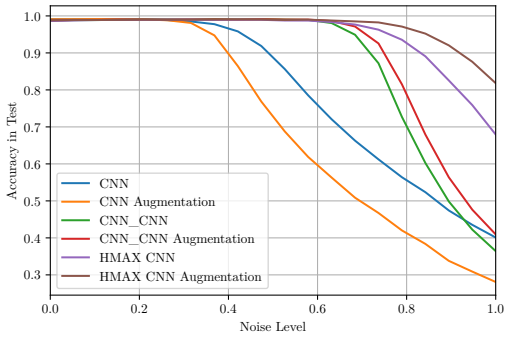


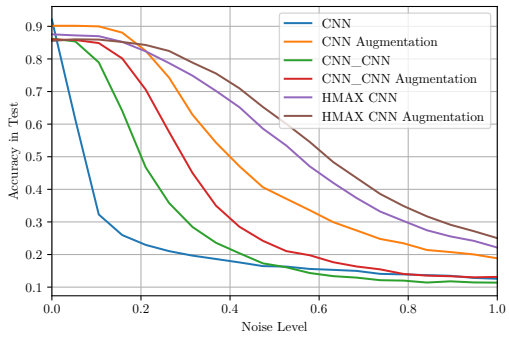
Figure 4.9 Left: $\texttt{CNN} - \texttt{CNN}$ model for denoising. The first block is trained to learn how to denoise an input image. The second is trained for classification. Right: Proposed model is composed of two blocks. Reconstruction block: it is composed of a CNN which computes from an input $\mathbf{f}$ a single real value h. The output is $\texttt{HMAX}_h(\mathbf{f})$. Classification block: A block of supervised classification. In both models, the second block is trained for classification without updating the denoising block.



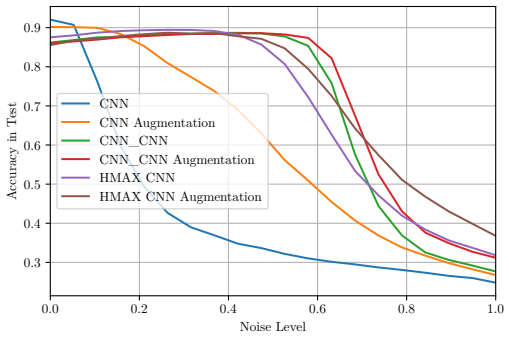
(a) MNIST Gaussian noise



(b) MNIST Uniform noise



(c) FASHION MNIST Gaussian noise



(d) FASHION MNIST Uniform noise

Figure 4.10 Classification accuracy of classical and proposed model for MNIST and Fashion MNIST with additive Gaussian and Uniform noise with $\mu = 0$ and $\sigma \in \{0., 0.05, \cdots, 1\}$. The original images have been normalised from zero to one. Examples of images with different levels of noise are shown in Fig. 4.11. The reconstruction block (Fig.4.9) has been trained *only* on MNIST with additive noise distributed as an absolute value zero-mean Gaussian with $\sigma = 0.1$. We have included a training with augmentation by additive random Gaussian noise at $\mu = 0$ and σ between 0 and 0.2 for fair comparison.

4.6 Sparse NMF representation and Mathematical Morphology

Mathematical morphology is strongly related to the problem of data representation. Applying a morphological filter can be seen as a test on how well the analysed element is represented by the set of invariants of the filter. For example, applying an opening by a structuring element `SE` tells how well a shape can be represented by the supremum of translations of `SE`. The morphological skeleton (165; 251) is a typical example of description of shapes by a family of building blocks, classically homothetic spheres. It provides a disjunctive decomposition where components - for example, the spheres - can only contribute positively as they are combined by supremum. A natural question is the optimality of this additive decomposition according to a given criterion, for example its sparsity - the number of components needed to represent an object. Finding a sparse disjunctive (or part-based) representation has at least two important features: first, it allows *saving resources* such as memory and computation time in the processing of the represented object; secondly, it provides a *better understanding* of this object, as it reveals its most elementary components, hence operating a dimensionality reduction that can alleviate the issue of model

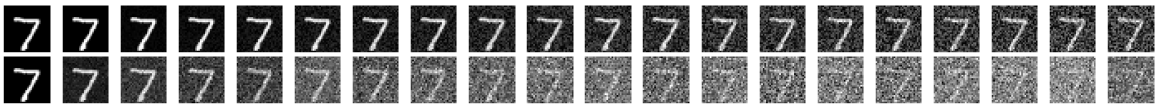


Figure 4.11 Example of images with level of noise between zero to ones where it has been used in Fig. 4.10. First row is a random uniform noise. Second row is a random Gaussian noise. Note that even with high noise levels the digit is perceived.

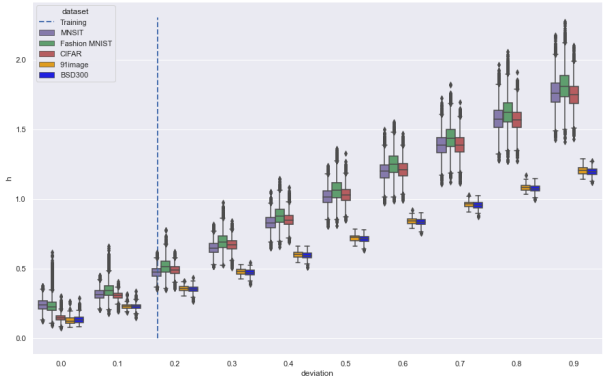


Figure 4.12 The figure shows the predicted h values for same type of noise on dataset MNIST, Fashion MNIST, CIFAR10, 91IMAGES(302) and BSD300 (169) for $\sigma \in \{0, 0.1, \ldots, 0.9\}$. The proposed reconstruction block network succeed in predict the noise strength for several datasets despite having been trained *only* at the level $\sigma = .2$ and *only* on MNIST database.

over-fitting. Such representations are also believed to be the ones at stake in human object recognition (264).

Similarly, the question of finding a sparse disjunctive representation of a whole database is also of great interest and will be the main focus of the present section. More precisely, we will approximate such a representation by a non-negative, sparse linear combination of non-negative components, and we will call *additive* this representation. Given a large set of images, our concern is then to find a smaller set of non-negative image components, called dictionary, such that any image of the database can be expressed as an additive combination of the dictionary components. As we will review in the next section, this question lies at the crossroad of two broader topics known as sparse coding and dictionary learning (157).

Besides a better understanding of the data structure, this section approach is also more specifically linked to mathematical morphology applications (33). The main goal is to be able to apply morphological operators to massive sets of images by applying them only to the reduced set of dictionary images. This is especially relevant in the analysis of remote sensing hyperspectral images where different kinds of morphological decomposition are widely used. For reasons that will be explained later, sparsity and non-negativity are sound requirements to achieve this goal. What is more, whereas the representation process can be learned offline on a training dataset, we need to compute the decomposition of any new sample *online*. Hence, we take advantage of the recent advances in deep, sparse and non-negative auto-encoders to design a new framework able to learn part-based representations of an image database, compatible with morphological processing.

4.6.1 Introduction

Consider a family of M images (binary or greyscale) $\mathbf{f}^{(1)}$, $\mathbf{f}^{(2)}, \ldots, \mathbf{f}^{(M)}$ of n pixels each, aggregated into a $M \times n$ data matrix $\mathbf{X} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(M)})^T$ (the i^{th} row of $\mathbf{X}$ is the transpose of $\mathbf{x}^{(i)}$, which is the image $\mathbf{f}^{(i)}$ seen as a vector). Given a feature dimension $k \in \mathbb{N}^*$ and two numbers $s_{\mathbf{L}}$ and $s_{\mathbf{A}} \in [0,1]$, a sparse NMF[11] of $\mathbf{X}$ with dimension k, as defined in (119), is any solution $(\mathbf{L}, \mathbf{A})$ of the problem

$$\min ||\mathbf{X} - \mathbf{L}\mathbf{A}||_2^2 \quad \text{s.t.} \quad \begin{cases} \mathbf{L} \in \mathbb{R}^{M \times k}, \mathbf{A} \in \mathbb{R}^{k \times M} \\ \mathbf{L} \geq 0,\ \mathbf{A} \geq 0 \\ \sigma(\mathbf{L}_{:,j}) = s_{\mathbf{L}}, \sigma(\mathbf{A}_{j,:}) = s_{\mathbf{A}},\ 1 \leq j \leq k \end{cases} \tag{4.22}$$

where the second constraint means that both $\mathbf{L}$ and $\mathbf{A}$ have non-negative coefficients, and the third constraint imposes the degree of sparsity of the columns of $\mathbf{L}$ and lines of $\mathbf{A}$ respectively, with σ the function defined by

$$\forall \mathbf{v} \in \mathbb{R}^p, \quad \sigma(\mathbf{v}) = \frac{\sqrt{p} - ||\mathbf{v}||_1 / ||\mathbf{v}||_2}{\sqrt{p} - 1}. \tag{4.23}$$

Note that σ takes values in $[0,1]$. The value $\sigma(\mathbf{v}) = 1$ characterises vectors $\mathbf{v}$ having a unique non-zero coefficient, therefore the sparsest ones, and $\sigma(\mathbf{v}) = 0$ the vectors whose coefficients all have the same absolute value. Hoyer (119) designed an algorithm to find at least a local minimiser for the problem (4.22), and it was shown that under fairly general conditions (and provided the ℓ^2 norms of $\mathbf{L}$ and $\mathbf{A}$ are fixed) the solution is unique (268).

In representation learning, each row $\mathbf{L}^{(i)}$ of $\mathbf{L}$ is called the *encoding* or *latent features* of the input image $\mathbf{x}^{(i)}$, and $\mathbf{A}$ holds in its rows a set of k images called the *dictionary*. In the following, we will refer to the images $\mathbf{a}_j = \mathbf{A}_{j,:}$ of the dictionary as *atom images* or *atoms*. As stated by 4.22, the atoms are combined to approximate each image $\mathbf{x^{(i)}} := \mathbf{x}_{i,:}$ of the dataset by an estimate $\hat{\mathbf{f}}^{(i)}$, which writes as follows:

$$\forall i \in \{1, ..., M\},\ \hat{\mathbf{f}}^{(i)} = \mathbf{L}_{i,:}\mathbf{A} = \mathbf{L}^{(i)}\mathbf{A} = \sum_{j=1}^{k} l_{i,j}\mathbf{a}_j, \tag{4.24}$$

where $l_{i,j}$ is the coefficient at row i and column j in matrix $\mathbf{L}$

By choosing the sparse NMF representation, the (282) aim at approximating a morphological operator ϕ on the data $\mathbf{x}$ by applying it to the atom images $\mathbf{A}$ only, before projecting back into the input image space. That is, they want $\phi(\mathbf{x}^{(i)}) \approx \Phi(\mathbf{x}^{(i)})$, with $\Phi(\mathbf{x}^{(i)})$ defined by

$$\Phi(\mathbf{x}^{(i)}) := \sum_{j=1}^{k} l_{i,j}\phi(\mathbf{a}_j), \tag{4.25}$$

where the $l_{i,j}$ and $\mathbf{a}_j$ are the same as in 4.24. The operator Φ in 4.25 is called a **part-based approximation** to ϕ. To understand why non-negativity and sparsity help this approximation to be a good one, we can point out a few key arguments. First, sparsity favours the support of

[11] A complete description and analysis of the nonnegative matrix factorization (NMF) is given in (91)

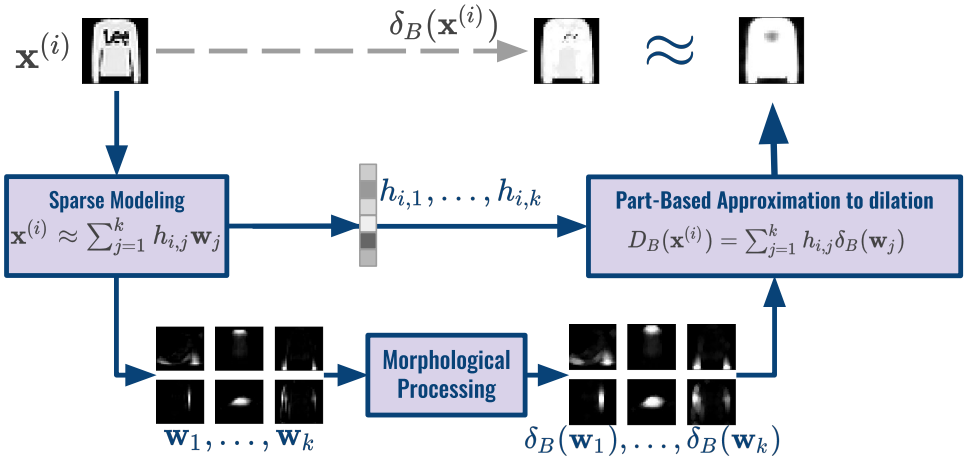


Figure 4.13 Process for computing the part-based approximation to dilation, based on Equations (4.24) and (4.26).

the weighted atom images to have little pairwise overlap. Secondly, a sum of images with disjoint supports is equal to their (pixel-wise) supremum. Finally, dilations commute with the supremum and, under certain conditions that are favoured by sparsity, this also holds for the erosions. This will be developed in more details in Section 4.6.2. For now, Figure 4.13 illustrates the part-based approximation $\delta^*_{\mathtt{SE}}$ of the dilation $\delta_{\mathtt{SE}}$ by a structuring element $\mathtt{SE}$, expressed as:

$$\delta^*_{\mathtt{SE}}(\mathbf{x}^{(i)}) := \sum_{j=1}^{k} l_{i,j}\delta_{\mathtt{SE}}(\mathbf{a}_j). \tag{4.26}$$

4.6.2 Morphological operators on non-negative linear combinations

In this section we precise the intuitions about the part-based approximation of morphological operators. Let $\mathcal{L}$ be the complete lattice of images with n pixels and with values in $[0, +\infty]$ ordered by the Pareto ordering ($\mathbf{f} \leq \mathbf{g}$ iff for any q, $1 \leq q \leq M$, $\mathbf{f}_q \leq \mathbf{g}_q$). Consider a flat, extensive dilation $\delta_{\mathtt{SE}}$ on $\mathcal{L}$ and its adjoint anti-extensive erosion $\varepsilon_{\mathtt{SE}}$, $\mathtt{SE}$ being a flat structuring element. Let $\mathbf{f}$ be an image approximated by the non-negative combination $\hat{\mathbf{f}} = \sum_{j=1}^{k} l_j\mathbf{a}_j$ of k atom images $\mathbf{a}_1, \ldots, \mathbf{a}_k$. Following 4.25, we define the part based approximations of the four operators $\delta_{\mathtt{SE}}$, $\varepsilon_{\mathtt{SE}}$, $\gamma_{\mathtt{SE}} = \delta_{\mathtt{SE}}\varepsilon_{\mathtt{SE}}$ and $\varphi_{\mathtt{SE}} = \varepsilon_{\mathtt{SE}}\delta_{\mathtt{SE}}$ as:

$$\begin{aligned} \delta^*_{\mathtt{SE}}(\mathbf{f}) &:= \textstyle\sum_{j=1}^{k} l_j\delta_{\mathtt{SE}}(\mathbf{a}_j), \quad & \varepsilon^*_{\mathtt{SE}}(\mathbf{f}) &:= \textstyle\sum_{j=1}^{k} l_j\varepsilon_{\mathtt{SE}}(\mathbf{a}_j) \\ \gamma^*_{\mathtt{SE}}(\mathbf{f}) &:= \textstyle\sum_{j=1}^{k} l_j\gamma_{\mathtt{SE}}(\mathbf{a}_j), \quad & \varphi^*_{\mathtt{SE}}(\mathbf{f}) &:= \textstyle\sum_{j=1}^{k} l_j\varphi_{\mathtt{SE}}(\mathbf{a}_j). \end{aligned} \tag{4.27}$$

We focus on establishing whether these expressions approximate well their exact counterparts $\delta_{\mathtt{SE}}(\mathbf{f})$, $\varepsilon_{\mathtt{SE}}(\mathbf{f})$, $\gamma_{\mathtt{SE}}(\mathbf{f})$ and $\varphi_{\mathtt{SE}}(\mathbf{f})$, assuming $\mathbf{f}$ is well approximated by $\hat{\mathbf{f}} = \sum_{j=1}^{k} l_j\mathbf{a}_j$. It is likely to be so as soon as $\delta^*_{\mathtt{SE}}(\mathbf{x}) = \delta_{\mathtt{SE}}(\hat{\mathbf{f}})$, $\varepsilon^*_{\mathtt{SE}}(\mathbf{x}) = \varepsilon_{\mathtt{SE}}(\hat{\mathbf{f}})$, $\gamma^*_{\mathtt{SE}}(\mathbf{x}) = \gamma_{\mathtt{SE}}(\hat{\mathbf{f}})$ and $\varphi^*_{\mathtt{SE}}(\mathbf{x}) = \varphi_{\mathtt{SE}}(\hat{\mathbf{f}})$, which is to say as soon as the four operators commute with the non-negative linear application $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_k] \mapsto \mathbf{Al} := \sum_{j=1}^{k} l_j\mathbf{a}_j$. As sketched earlier, sums can be identified to suprema if the involved images have disjoint supports, and this also favours the commutation of the erosion

with the supremum. This is why we introduce the following hypothesis that characterises the disjunction of supports (*i.e.*, the regions where the image is non-zero) of the $l_j\mathbf{a}_j$.

Let $\mathcal{H}_1$ denote the hypothesis:

$$\mathcal{H}_1\text{: ``For any } 1 \leq i \leq k, 1 \leq j \leq k, i \neq j,\ \delta_{\mathtt{SE}}(l_i\mathbf{a}_i) \bigwedge \delta_{\mathtt{SE}}(l_j\mathbf{a}_j) = 0\text{''}$$

where 0 denotes an image equal to zero everywhere (*i.e.* with empty support), and more generally, for an integer n,

$$\mathcal{H}_n\text{: ``For any } 1 \leq i \leq k, 1 \leq j \leq k, i \neq j,\ \delta^n_{\mathtt{SE}}(l_i\mathbf{a}_i) \bigwedge \delta^n_{\mathtt{SE}}(l_j\mathbf{a}_j) = 0\text{''},$$

where $\delta^n_{\mathtt{SE}} = \delta_{\mathtt{SE}} \circ \cdots \circ \delta_{\mathtt{SE}} = \delta_{n\mathtt{SE}}$, denoting by $n\mathtt{SE}$ the n-terms Minkowski sum $\mathtt{SE} \oplus \mathtt{SE} \oplus \cdots \oplus \mathtt{SE}$ for $n > 0$, and $\delta^n_{\mathtt{SE}}$ is the identity for $n = 0$. Note that, since $\delta_{\mathtt{SE}}$ is extensive, $\mathcal{H}_n$ implies any $\mathcal{H}_p$ with $p \leq n$. In particular, any $\mathcal{H}_n$ implies $\mathcal{H}_0$, which simply states the disjunction of the supports of any two images $l_i\mathbf{a}_i$ and $l_j\mathbf{a}_j$, $i \neq j$. We can now state the following result:

Proposition 4.18. *If $\mathcal{H}_1$ holds for the representation $\hat{\mathbf{f}} = \sum_{j=1}^k l_j\mathbf{a}_j$, then:*

$$\delta^*_{\mathtt{SE}}(\mathbf{f}) = \delta_{\mathtt{SE}}(\hat{\mathbf{f}}), \quad \varepsilon^*_{\mathtt{SE}}(\mathbf{f}) = \varepsilon_{\mathtt{SE}}(\hat{\mathbf{f}}), \quad \gamma^*_{\mathtt{SE}}(\mathbf{f}) = \delta_{\mathtt{SE}}(\varepsilon_{\mathtt{SE}}(\hat{\mathbf{f}})) = \gamma_{\mathtt{SE}}(\hat{\mathbf{f}}). \tag{4.28}$$

If additionally $\mathcal{H}_2$ holds, then we also have:

$$\varphi^*_{\mathtt{SE}}(\mathbf{f}) = \varepsilon_{\mathtt{SE}}(\delta_{\mathtt{SE}}(\hat{\mathbf{f}})) = \varphi_{\mathtt{SE}}(\hat{\mathbf{f}}). \tag{4.29}$$

A proof of this result is detailed in (33).

Proposition 4.18 implies that under the H_n hypothesis the error $||\phi_{\mathtt{SE}}(\mathbf{f}) - \Phi_{\mathtt{SE}}(\mathbf{f})||^2$ between the actual transformed image and its part-based approximation only depends on the quality of the reconstruction, that is to say on the error $||\mathbf{f} - \hat{\mathbf{f}}||^2$. Indeed, if $\mathbf{f} = \hat{\mathbf{f}}$ then $\delta^*_{\mathtt{SE}}(\mathbf{f}) = \delta_{\mathtt{SE}}(\mathbf{f})$, $\varepsilon^*_{\mathtt{SE}}(\mathbf{f}) = \varepsilon_{\mathtt{SE}}(\mathbf{f})$ and so on. Obviously, the more constrained the representation, the smaller the class of images that can be accurately represented. The non-negativity and sparsity constraints are therefore likely to increase the representation error $||\mathbf{f} - \hat{\mathbf{f}}||^2$. Hence, unless the data can be perfectly represented by non-negative combinations of atoms complying with a hypothesis H_n, a trade-off needs to be found to achieve a good approximation of morphological operators. This is the target of the asymmetric auto-encoder presented in (33).

We shall now generalise Proposition 4.18 by applying it to the representation that we note $\hat{\mathbf{f}}^{(n-1)} = \sum_{j=1}^k l_j \delta_{(n-1)\mathtt{SE}}(\mathbf{a}_j)$. Notice that H_1 holds for $\hat{\mathbf{f}}^{(n-1)}$ if and only if H_n holds for $\hat{\mathbf{f}}$. This yields the following corollary.

Remark 4.19. If H_n holds for the representation $\hat{\mathbf{f}} = \sum_{j=1}^k l_j\mathbf{a}_j$, then for any integer $p \leq n$:

$$\delta^*_{p\mathtt{SE}}(\mathbf{f}) = \delta_{p\mathtt{SE}}(\hat{\mathbf{f}}), \quad \varepsilon^*_{p\mathtt{SE}}(\mathbf{f}) = \varepsilon_{p\mathtt{SE}}(\hat{\mathbf{f}}), \quad \gamma^*_{p\mathtt{SE}}(\mathbf{f}) = \delta_{p\mathtt{SE}}(\varepsilon_{p\mathtt{SE}}(\hat{\mathbf{f}})) = \gamma_{p\mathtt{SE}}(\hat{\mathbf{f}}),$$

and for any integer $p \leq n - 1$

$$\varphi^*_{p\mathtt{SE}}(\mathbf{f}) = \varepsilon_{p\mathtt{SE}}(\delta_{p\mathtt{SE}}(\hat{\mathbf{f}})) = \varphi_{p\mathtt{SE}}(\hat{\mathbf{f}}).$$

Quantitative results from (33) are given in Table 4.6 and in Fig. 4.14 for the Fashion MNIST database.

Table 4.6 Comparison of the reconstruction error, sparsity of encoding and part-based approximation error to dilation produced by the sparse-NMF, the NNSAE, the NCAE and the AsymAE, for both MNIST and Fashion-MNIST datasets.

Model	Reconstruction error	Sparsity of code	Part-based approximation error to dilation
MNIST			
Sparse-NMF	0.011	**0.66**	**0.012**
NNSAE	0.015	0.31	0.028
NCAE	0.010	0.35	0.18
AsymAE	**0.007**	0.54	0.069
Fashion MNIST			
Sparse-NMF	0.011	**0.65**	**0.022**
NNSAE	0.029	0.22	0.058
NCAE	0.017	0.60	0.030
AsymAE	**0.010**	0.52	0.066

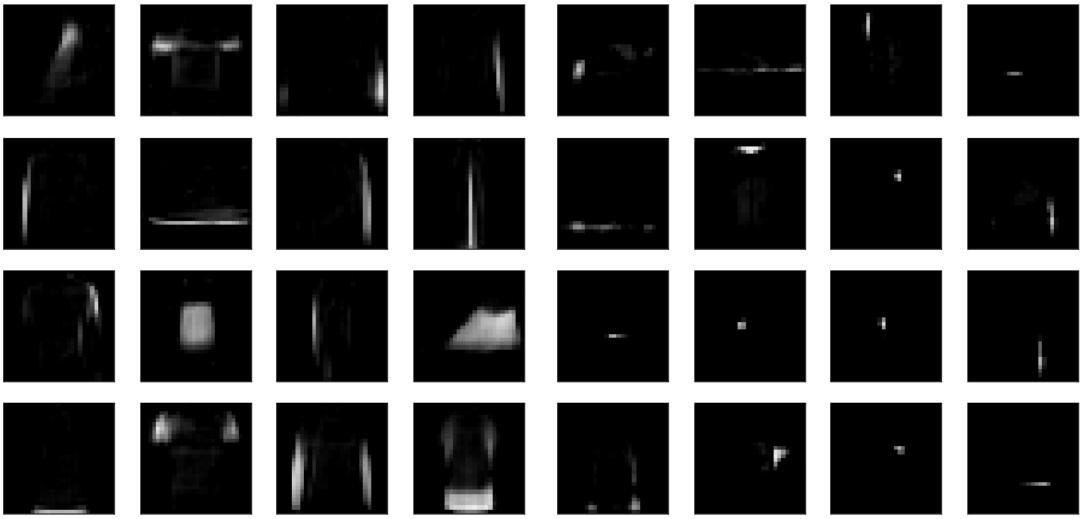
Related Publications

- **Max-plus operators applied to filter selection and model pruning in neural networks**, Y. Zhang, S. Blusseau, S. Velasco-Forero, I. Bloch, J. Angulo, ISMM 2019.

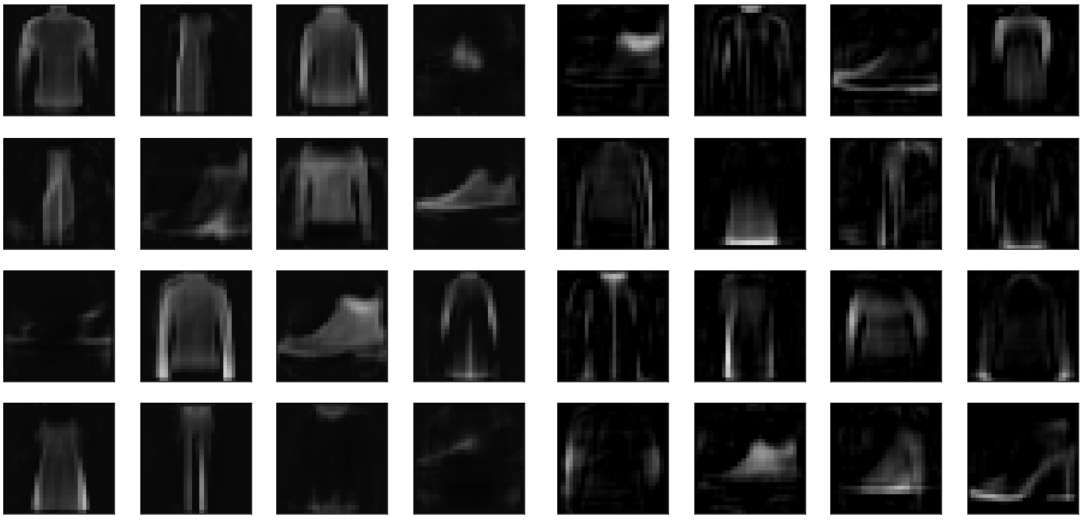
- **Non-Negative Sparse Mathematical Morphology**, J. Angulo and S. Velasco-Forero, Advances in Imaging and Electron Physics 202, 1-37, 2017.
- **Part-based approximations for morphological operators using asymmetric auto-encoders**, Bastien Ponchon, Santiago Velasco-Forero, Samy Blusseau, Jesus Angulo, and Isabelle Bloch, ISMM 2019.
- **Approximating morphological operators with part-based representations learned by asymmetric auto-encoders**, S. Blusseau, B. Pochon, S. Velasco-Forero, J. Angulo and I. Bloch, Mathematical Morphology-Theory and Applications, 2020
- **Learnable Empirical Mode Decomposition based on Mathematical Morphology**, Velasco-Forero, S., Pagès, R., and Angulo, J., SIAM Journal on Imaging Sciences, 15(1), 23-44, (2022).
- **MorphoActivation: Generalizing ReLU activation function by mathematical morphology**, Velasco-Forero, S., and Angulo, J. DGMM, (2022).
- **Fixed Point Layers for Geodesic Morphological Operations**, Velasco-Forero, S., Rhim, A., and Angulo, J., British Machine Vision Conference, (2022).

4.6.3 Other contributions

- **On power Jaccard losses for semantic segmentation**, D. Duque-Arias et al, VISAPP 2021.



(a) Sparse-NMF **(b)** NNSAE



(c) NCAE **(d)** AsymAE

Figure 4.14 16 of the 100 atom images of the four compared representations of Fashion-MNIST dataset.

In this work, a new generalised loss function is proposed called power Jaccard to perform semantic segmentation tasks. It is compared with classical loss functions in different scenarios, including grey level and colour image segmentation, as well as 3D point cloud segmentation. The results show improved performance, stability and convergence. We made available the code with our proposal with a demonstrative example.

- **A New colour Augmentation Method for Deep Learning Segmentation of Histological Images**, Y. Xiao et al, ISBI 2019.
  This paper addresses the problem of labeled data insufficiency in neural network training for semantic segmentation of colour-stained histological images acquired via Whole Slide Imaging. It proposes an efficient image augmentation method to alleviate the demand for a large amount of labeled data and improve the network's generalisation capacity. Typical image augmentation in bioimaging involves geometric transformation. Here, we propose a

new image augmentation technique by combining the structure of one image with the colour appearance of another image to construct augmented images on-the-fly for each training iteration. We show that it improves performance in the segmentation of histological images of human skin, and also offers better results when combined with geometric transformation

- **Dealing with Topological Information within a Fully Convolutional Neural Network**, E. Decencière et al., ACIVS 2018.
  A fully convolutional neural network has a receptive field of limited size and therefore cannot exploit global information, such as topological information. A solution is proposed in this paper to solve this problem, based on pre-processing with a geodesic operator. It is applied to the segmentation of histological images of pigmented reconstructed epidermis acquired via Whole Slide Imaging.
- **Classification of hyperspectral images by tensor modelling and additive morphological decomposition**, S. Velasco-Forero and J. Angulo, vol. 46, num. 2, Feb. 2013, Pattern Recognition.
  Pixel-wise classification in high-dimensional multivariate images is investigated. The proposed method deals with the joint use of spectral and spatial information provided in hyperspectral images. Additive morphological decomposition (AMD) based on morphological operators is proposed. AMD defines a scale-space decomposition for multivariate images without any loss of information. AMD is modelled as a tensor structure and tensor principal components analysis is compared as a dimensional reduction algorithm versus classic approach. Experimental comparison shows that the proposed algorithm can provide better performance for pixel classification of hyperspectral image than many other well-known techniques.

4.7 Perspectives

Despite having excellent results in quantitative performance measurements, the main drawback of classical deep learning methods is that they are trained as *black boxes*, where neither imposing a priori problem knowledge nor understanding the decision rule are straightforward. In the following some aspects where mathematical morphology based models could be interesting for Deep Learning models.

Interpretability

One argument against convolution is that its filter does not lend itself to an *interpretable*[12] form. The filter weights do not indicate the absolute intensities/levels in the shape. Instead, it means a relative importance. Recently, research in deep network interpretation as for instance *guided backpropagation* (3; 245) made it possible to visualise what CNNs might be looking at. However, these algorithms are not guarantees, they do not report which features are more important, not what exact shape, texture, colour, or other characteristics led a machine to make the decision they

[12]Interpretability in the sense of (183): "Interpretability is the degree to which a human can understand the cause of a decision".

did. Another approach are the *attention-like* mechanisms introduced in the 90's as multiplicatives modules (74), it motivated by the idea that networks should devote more focus to the small, but important, parts of the data (199; 297).

Unlike convolution, morphology-based operations are more interpretable, in the sense that for a given layer, the flat morphological filters are easier to interpret than the convolutional ones. Morphological operators allow the characterisation of objects in the image with respect to the shape/size relationship, providing an absolute measure of fitness against the measure relative by convolution, which facilitates the interpretability of the structuring elements.

However, in general, as soon as the complexity of the model increases, the interpretation becomes more difficult. In deep models, usually one should use pooling to downsampling spatial domain, which is one of the key points in classical multiresolution analysis (106; 158).

Especially in shallow models, morphological models should give geometric rules that allow to characterise the important elements in the images. From this point of view, the morphological methods should be used at the moment of decision making, combining with classical linear convolutions in the feature extraction layers.

Invariance

A professional image analyst would like to include some geometrical knowledge during the training of CNNs. Especially when analysing images, one would like to have results that are *invariant* to shifts, symmetries, rotations, or scale changes (159). This makes sense since the characteristics of the object do not change (298). However, it is easy to see that due to the definition of the elementary operations that are being used in the classic deep learning models, this property is not satisfied. A mathematical approach is to define a group of transformation and defining layers whose results are invariant to the action of the group. Recently, a wide variety of articles have been presented in this regard including invariances to rotation and changes in the scale of objects (29; 43). The problem of defining morphological operators invariant to a group of transformation have been studied in (223; 241; 254). Finally, we should note that the morphological hierarchies based on local gradient are by definition invariant to shift in the contrast (184). Accordingly, invariant descriptors to groups such as additive and multiplicative can be easily constructed by means of range operators, and in their convolutions version by morphological operators. Some recent works on CNNs using Gaussian derivatives to produce equivariant layers and/or train with less examples look promising specially for industrial applications (206)(239)(240).

Shape/Size priors for segmentation models

In segmentation models, especially in images, one may be interested in including a priori knowledge about the objects of interest, for example constraints on size order relationships or other geometric attributes. Some of these possibilities are listed below, including some important references.

1. **Structure-driven priors**: This is a generalisation of classical marker based segmentation models (42; 236). The approach is to use semi-supervised formulation to include *marker information* in the loss function, *i.e.*, from a pair of image, seeds *(*$\mathbf{f}$*,seeds)*(167) to goal is

to segment an instance per marker. Some examples for medical applications can be found in (127) (216)

2. **Knowledge-driven priors** (e.g., anatomy): Actually in many applications one is interested in including information about the allowed size of objects to segmentation, order relationships (for instance, one segmented object must be larger than another, less eccentric than another), connectivity (for instance, two segments must be connected) and/or topology (for instance, segmented objects must not contain hulls.). One can characterise the prior-information for deep learning models in four types:
   (a) *Constraints*: Defining the limits of plausible segmentation according to a variable (for instance size), and or partially labeled data. (128; 134; 135; 217; 307). Note that regularising the loss function including shape constraint is pretty direct, but the model does not take into account this term in the same way during the optimisation process.
   (b) *Shape Models*: Most of the methods are hybrid solutions involving active contour or statistical shape models. A state-of-the art for medical applications can be found in (36)
   (c) *Morphological constraints*: Many of the geometric constraints can be described by means of mathematical morphological operators, which can be included in deep learning models in different ways:
      i. Preprocessing: For instance, including morphological reconstruction to consider nonlocal information in DL models. (69)
      ii. Morphological filters: The use of morphological transformation as layers in Deep Learning models to give an interpretation from the size/shape relationship of learned kernels. (287). Most of the morphological filtering can be used in a DL network to certify the result of the procedure morpholayers.
      iii. As a regulariser: One can consider to study to regularise the results of a DL model by considering morphological based loss functions (16).
   (d) *Topological regularisation*: According to the application, one may be interested in obtaining segmentation that preserves connectivity criteria. In (59; 88) the use ultrametric is used in gradient based method to penalise the results of hierarchical clustering (one can see hierarchical segmentation as a particular case of hierarchical clustering). Additionally, (209) presents a loss function that “motivates" the connection of the objects in the image. Another way to include topological information in the loss function is to use *Persistent Homology* (61)(120). Note: That the inclusion of terms in the cost function does not guarantee that the constraint is satisfied. An important connection between morphological and persistence-based methods has recently been presented in (41; 60)

Related Publications

- **Fully Trainable Gaussian Derivative Convolutional Layer**, V. Penaud-Polge, S. Velasco-Forero and J. Angulo, ICIP, 2022.

- **Differential invariants for SE (2)-equivariant networks**, M. Sangalli, S. Blusseau, S. Velasco-Forero and J. Angulo, ICIP, 2022.
- **Moving Frame Net: SE(3)-Equivariant Network for Volumes**, M. Sangalli, S. Blusseau, S. Velasco-Forero and J. Angulo, NEURIPS, 2022.

5 Contributions in Anomaly Detection

C'est avec la logique que nous prouvons et avec l'intuition que nous trouvons.

Henri Poincaré

5.1 Introduction

The job of the professional image analyst is to find things in images. Often the analyst knows in advance what kinds of things to look for: peoples, airplanes, industrial facilities, cars, and so on. But sometimes the analyst is faced with the more open task of finding "unusual" things, without knowing in advance what these unusual things will be.

When the object of interest are known, one can employ *supervised learning*. But a pretty different problem arise when the examples of the target class are unavailable, or simply when the target class are *unknown* objects. The analyst should choice "normal" images and use them from training. This is the *anomaly detection problem*: a *unsupervised learning* problem in which the learning from data proceeds without any example [1] of the target of interest.

Many AD methods have been proposed, and a few literature reviews or tutorials have been thoroughly done (38; 50; 172; 173; 242; 257). Recently, the tutorial by (173) gives a good overview of different AD methods in the literature. In this manuscript, we analyse the AD in the context of optimal statistical detection, where the covariance matrix of the background is required to be estimated. Additionally, the extension in the context of deep learning is motivated and introduced.

In the dictionary, the definition of anomaly is given by:

1. something different, abnormal, peculiar, or not easily classified : something anomalous.
2. deviation from the common rule.

The first definition motivates many approaches using compression techniques, autoencoders, among others as anomalies detector (5; 306). Where one has a normal information encoding/decoding system, and one hopes that the anomalies are the objects that cannot be decoded

[1] In case one has some examples of anomalies, we will use the term semisupervised

well. The second definition is more in the statistical sense, where one fits a model on normal data, and expects anomalies to be objects with low probability according to the model.

The following section briefly describes the RX-detector before reviewing some covariance matrix estimation methods in the literature.

5.2 The RX-detector

AD may be considered as a binary hypothesis testing problem at every pixel as follows:

$$\begin{aligned} \mathcal{H}_0 : &\quad \mathbf{x} \sim f_{\mathbf{x}|\mathcal{H}_0}(\mathbf{x}), \\ \mathcal{H}_1 : &\quad \mathbf{x} \sim f_{\mathbf{x}|\mathcal{H}_1}(\mathbf{x}), \end{aligned} \tag{5.1}$$

where $f_{\mathbf{x}|\mathcal{H}_i}(\cdot)$ denotes the probability density function (PDF) conditioned on the hypothesis i, *i.e.*, $\mathcal{H}_0$ when the target is absent (*background*), and $\mathcal{H}_1$ when the target is present. Usually, For $\mathcal{H}_0$, the background distribution $f_{\mathbf{x}|\mathcal{H}_0}(\mathbf{x})$ is assumed to be a multivariate Gaussian model (MGM) due to theoretical simplicity. The distribution in the presence of the target can be assumed to have a multivariate uniform PDF (266). The well-known *RX anomaly detector* (Reed and Xiaoli Yu (218)) was based on these two assumptions, its test statistics is as follows:

$$\begin{aligned} \frac{p}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) &\underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \tau_0, \\ \Rightarrow \mathtt{AD}_{\mathtt{RX}}(\mathbf{x},\tau_1) := (\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) &\underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau_1, \end{aligned} \tag{5.2}$$

where $|\boldsymbol{\Sigma}|$ is the determinant of matrix $\boldsymbol{\Sigma}$, and τ_0 and τ_1 are thresholds, above which $\mathcal{H}_0$ is rejected in favour of $\mathcal{H}_1$. In other words, the RX-detector is a threshold test on the *Mahalanobis distance* (156). Thresholding the likelihood ratio provides the hypothesis test that satisfies various optimality criteria including: maximum probability of detection for the given probability of false alarm, minimum expected cost, and minimisation of maximal expected cost (160). However, in most of the cases, $\boldsymbol{\Sigma}$ is unknown and needs to be estimated. It is well-known (7) that given n independent samples, $\mathbf{X}_{n\times p} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathbb{R}^p\}$ from a p-variate Gaussian distribution with known mean $\boldsymbol{\mu} \in \mathbb{R}^p$, the *sample covariance matrix* (SCM) defined by

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T, \tag{5.3}$$

is the maximum likelihood estimator (MLE) of $\boldsymbol{\Sigma}$. The *sample covariance matrix* (SCM) is the maximum likelihood estimator, but it tends to overfit the data when n does not greatly exceed p. However, in the presence of multiple clusters, this estimation fails to characterise the background correctly. For these reasons, a variety of regularisation schemes have been proposed (55; 56), as well as several robust estimation approaches (1; 30; 85; 117; 171; 260; 267; 269; 296).

5.2.1 The RX-detector in High Dimensional Space

To help better understand the implication of high dimensionality in the RX-detector, we develop an alternative expression for (5.2) based on the *Singular Value Decomposition* (SVD) of the covariance matrix $\mathbf{\Sigma}$, as follows:

$$\mathtt{AD_{RX}}(\mathbf{x}, \tau_1) = (\mathbf{x}-\boldsymbol{\mu})^T \mathbf{U}^{-1}\mathbf{\Lambda}^{-1}\mathbf{U}(\mathbf{x}-\boldsymbol{\mu}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau_1,$$

where $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$ with $\mathbf{\Lambda}$ a diagonal matrix and $\mathbf{U}$ an orthogonal matrix. The eigenvalues $\{\lambda_i\}_{i=1}^p$ in $\mathbf{\Lambda}$ correspond to the variances along the individual eigenvectors and sum up to the total variance of the original data. Let the diagonal matrix $\mathbf{\Omega} = \{\Omega_{ii}\}_{i=1}^p = \{1/\sqrt{\lambda_i}\}_{i=1}^p$, then $\mathbf{\Omega}^2 = \mathbf{\Lambda}^{-1}$. Additionally, since $\mathbf{U}$ is a rotation matrix, *i.e.*, $\mathbf{U}^{-1} = \mathbf{U}^T$, we can rewrite the RX-detector as follows:

$$\begin{aligned}
&= (\mathbf{x}-\boldsymbol{\mu})^T \mathbf{U}\mathbf{\Omega}\mathbf{\Omega}\mathbf{U}^T(\mathbf{x}-\boldsymbol{\mu}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau_1 \\
&= ||\mathbf{\Omega}\mathbf{U}^T(\mathbf{x}-\boldsymbol{\mu})||_2^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau_1.
\end{aligned} \tag{5.4}$$

As we can see from this decomposition, the RX-detector in 5.2 is equivalent to the weighted Euclidean norm by the eigenvalues along the principal components. Note that as $\lambda_i \to 0$, the detector $\mathtt{AD_{RX}}(\mathbf{x}, \tau_1) \to \infty$, $\forall \mathbf{x}$, resulting in an unreasonable bias towards preferring $\mathcal{H}_1$ to $\mathcal{H}_0$. This fact is well-known in the literature as bad conditioning, *i.e.*, the condition number [2] of $cond(\mathbf{\Sigma}) \to \infty$. Before looking at the possible solutions to the ill-conditioning issue, we would like to have a more detailed analysis of the eigenvalue distribution of covariance matrices in the theory of random matrices (6; 75; 166).

5.3 Robust Estimation in Non-Gaussian Assumptions

Presence of outliers can distort both mean and covariance estimates in computing Mahalanobis distance. In the following, we describe two types of robust estimators for covariance matrix.

5.3.1 M-estimators

In a Gaussian distribution, the SCM $\widehat{\mathbf{\Sigma}}$ in (5.3) is the MLE of $\mathbf{\Sigma}$. This can be extended to a larger family of distributions. *Elliptical distributions* is a broad family of probability distributions that generalise the multivariate Gaussian distribution and inherit some of its properties (7; 82). The p-dimension random vector $\mathbf{X}$ has a multivariate elliptical distribution, written as $\mathbf{X} \sim E_p(\boldsymbol{\mu}, \mathbf{\Sigma}, \psi)$, if its characteristic function can be expressed as, $\psi_{\mathbf{X}} = \exp(i\mathbf{t}^T\boldsymbol{\mu})\psi\left(\frac{1}{2}\mathbf{t}^T\mathbf{\Sigma}\mathbf{t}\right)$ for some vector $\boldsymbol{\mu}$, positive-definite matrix $\mathbf{\Sigma}$, and for some function ψ, which is called the characteristic generator. From $\mathbf{X} \sim E_p(\boldsymbol{\mu}, \mathbf{\Sigma}, \psi)$, it does not generally follow that $\mathbf{X}$ has a density $f_{\mathbf{X}}(\mathbf{x})$, but, if it exists,

[2] The condition number of a real matrix $\mathbf{\Sigma}$ is the ratio of the largest singular value to the smallest singular value. A well-conditioned matrix means its inverse can be computed with good accuracy.

it has the following form:

$$f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},g_d)=\frac{c_p}{\sqrt{|\boldsymbol{\Sigma}|}}g_p\left[\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] \tag{5.5}$$

where c_p is the normalization constant and g_p is some non-negative function with $(\frac{p}{2}-1)$-moment finite. In many applications, including AD, one needs to find a robust estimator for data sets sampled from distributions with heavy tails or outliers. A commonly used robust estimator of covariance is the Maronna's M estimator (168), which is defined as the solution of the equation

$$\widehat{\boldsymbol{\Sigma}}_M=\frac{1}{n}\sum_{i=1}^{n}u((\mathbf{x}_i-\boldsymbol{\mu})^T\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_i-\boldsymbol{\mu}))((\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^T, \tag{5.6}$$

where the function $u:(0,\infty)\to[0,\infty)$ determines a whole family of different estimators. In particular, a special case $u(x)=\frac{p}{x}$ is shown to be the most robust estimator of the covariance matrix of an elliptical distribution with form (5.5), in the sense of minimising the maximum asymptotic variance. This is the called *Tyler's method* (269) which is given by

$$\widehat{\boldsymbol{\Sigma}}_{\text{Tyler}}=\frac{p}{n}\sum_{i=1}^{n}\frac{(\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^T}{(\mathbf{x}_i-\boldsymbol{\mu})^T\widehat{\boldsymbol{\Sigma}}_{\text{Tyler}}^{-1}(\mathbf{x}_i-\boldsymbol{\mu})}. \tag{5.7}$$

(269) established the conditions for the existence of a solution of the fixed point equation (5.7). Additionally, (269) shows that the estimator is unique up to a positive scaling factor, *i.e.*, that $\boldsymbol{\Sigma}$ solves (5.7) if and only if $c\boldsymbol{\Sigma}$ solves (5.7) for some positive scalar $c>0$. Another interpretation to (5.7) can be found by considering normalised samples defined as $\{\mathbf{s}_i=\frac{\mathbf{x}_i-\boldsymbol{\mu}}{||\mathbf{x}_i-\boldsymbol{\mu}||}\}_{i=1}^n$. Then, the PDF of $\mathbf{s}$ takes the form (82):

$$f_{\mathbf{S}}(\mathbf{s})=\frac{\Gamma(\frac{p}{2})}{2\pi^{\frac{p}{2}}}\det(\boldsymbol{\Sigma})^{-\frac{1}{2}}(\mathbf{s}^T\boldsymbol{\Sigma}^{-1}\mathbf{s})^{\frac{-p}{2}},$$

and the MLE of $\boldsymbol{\Sigma}$ can by obtained by minimizing the negative log-likelihood function:

$$\mathcal{L}(\boldsymbol{\Sigma})=\frac{p}{2}\sum_{i=1}^{n}\log(\mathbf{s}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{s}_i)+\frac{n}{2}\log\det(\boldsymbol{\Sigma}). \tag{5.8}$$

If the optimal estimator $\widehat{\boldsymbol{\Sigma}}>0$ of (5.8) exist, it needs to satisfy the equation (5.7) (82). When $n>p$, Tyler proposed the following iterative algorithm based on $\{\mathbf{s}_i\}$:

$$\widetilde{\boldsymbol{\Sigma}}_{k+1}=\frac{p}{n}\sum_{i=1}^{n}\frac{\mathbf{s}_i\mathbf{s}_i^T}{\mathbf{s}_i^T\widehat{\boldsymbol{\Sigma}}_k^{-1}\mathbf{s}_i},\widehat{\boldsymbol{\Sigma}}_{k+1}=\frac{\widetilde{\boldsymbol{\Sigma}}_k}{\text{tr}(\widetilde{\boldsymbol{\Sigma}}_k)}. \tag{5.9}$$

It can be shown (269) that the iteration process in (5.9) converges and does not depend on the initial setting of $\widehat{\boldsymbol{\Sigma}}_0$. Accordingly, the initial $\widehat{\boldsymbol{\Sigma}}_0$ is usually set to be the identity matrix of size p. We have denoted the iteration limit $\widehat{\boldsymbol{\Sigma}}_\infty=\widehat{\boldsymbol{\Sigma}}_{\text{Tyler}}$. Note that the normalisation by the trace in the right side of (5.9) is not mandatory but it is often used in Tyler based estimation to make easier the comparison and analysis of its spectral properties without any decrement in the detection performance.

5.3.2 Multivariate t-distribution Model

Firstly, we evoke a practical advice to perform AD in real-life HS images from (50). They have indicated that the quality of the AD can be improved by means of considering the correlation matrix $\mathbf{R}$ instead of the covariance matrix $\mathbf{\Sigma}$, also known as the *R-RX-detector* (66). However, notice that writing the j-th coordinate of the vector $\mathbf{z}$ as $z_{(j)} = \frac{\mathbf{x}_{(j)} - \mu_{(j)}}{\sqrt{\sigma_{(jj)}}}$, we have $\mathbf{z} = (z_1, \ldots, z_p) = \boldsymbol{\sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$, where $\boldsymbol{\sigma} = \text{diag}(\sqrt{\sigma_1, \ldots, \sigma_p})$. Now, $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]$ is zero-mean, and $cov(\mathbf{Z}) = \boldsymbol{\sigma}^{-1/2}\mathbf{\Sigma}\boldsymbol{\sigma}^{-1/2} = \mathbf{R}$, the correlation matrix of $\mathbf{X}$. Thus, the correlation matrix of $\mathbf{X}$ is the covariance matrix of $\mathbf{Z}$, *i.e.* the standardisation ensuring that all the variable in $\mathbf{Z}$ are on the same scale. Additionally, note that (66) gives a characterization of the performance of the R-RX-detection. They conclude that the performance of R-RX depends not only on the dimensionality p and the deviation from the anomaly to the background mean but also on the squared magnitude of the background mean. That is an unfavourable point in the case that $\boldsymbol{\mu}$ needs to be estimated. At this point, we are interested in characterising the MLE solution of the correlation matrix $\mathbf{R}$ by means of *t-distribution*. A p-dimensional random vector $\mathbf{x}$ is said to have the p-variate $t-$distribution with degrees of freedom v, mean vector $\boldsymbol{\mu}$, and correlation matrix $\mathbf{R}$ (and with $\mathbf{\Sigma}$ denoting the corresponding covariance matrix) if its joint PDF is given by:

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{\Sigma}, v) = \frac{\Gamma(\frac{v+p}{2})|\mathbf{R}|^{-1/2}}{(\pi v)^{\frac{p}{2}}\Gamma(\frac{v}{2})\left[1 + \frac{1}{v}(\mathbf{x} - \boldsymbol{\mu})^T\mathbf{R}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right]^{\frac{v+p}{2}}},$$

where the degree of freedom parameter v is also referred to as the shape parameter, because the peakedness of (5.10) may be diminished or increased by varying v. Note that if $p = 1, \boldsymbol{\mu} = 0$, and $\mathbf{R} = 1$, then (5.10) is the PDF of the *univariate Student's t distribution* with degrees of freedom v. The limiting form of (5.10) as $v \to \infty$ is the joint PDF on the p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$. Hence, (5.10) can be viewed as a generalization of the multivariate normal distribution. The particular case of (5.10) for $\boldsymbol{\mu} = 0$ and $\mathbf{R} = \mathbf{I}_p$ is a normal density with zero means and covariance matrix $v\mathbf{I}_p$ in the scale parameter v. However, the MLE does not have closed form and it should be found through *expectation-maximization algorithm* (EM) (186)(150). The EM algorithm takes the form of iterative updates, using the current estimates of $\boldsymbol{\mu}$ and $\mathbf{R}$ to generate the weights. The iterations take the form:

$$\widehat{\boldsymbol{\mu}}_{k+1} = \frac{\sum_{i=1}^n w_k^i \mathbf{x}_i}{\sum_{i=1}^n w_k^i}, \text{ and} \tag{5.10}$$

$$\widehat{\mathbf{R}}_{k+1} = \frac{1}{n}\sum_{i=1}^n (w_k^i(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_{k+1})(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_{(k+1)})^T), \tag{5.11}$$

where $w_{k+1}^i = \frac{v+p}{v+(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)^T\mathbf{R}_k^{-1}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_k)}$. For more details of this algorithm, interested readers may refer to (150), and (192) for faster implementations. In our case, of known zero mean, this approach becomes:

$$\widehat{\mathbf{R}}_{k+1} = \frac{v+p}{n}\sum_{i=1}^n \frac{\mathbf{x}_i\mathbf{x}_i^T}{v + \mathbf{x}_i^T\widehat{\mathbf{R}}_k^{-1}\mathbf{x}_i} \tag{5.12}$$

For the case of unknown v, (143) showed how to use EM to find the joint MLEs of all parameters $(\boldsymbol{\mu}, \mathbf{R}, v)$. However, our preliminary work (285) shows that the estimation of v does not give any improvement in AD task. Therefore, we limited ourselves to the case of t-distribution with known value of degrees of freedom v.

5.4 Estimators in High Dimensional Space

The SCM $\widehat{\boldsymbol{\Sigma}}$ in (5.3), offers the advantages of easy computation and being an unbiased estimator, *i.e.*, its expected value is equal to the covariance matrix. However, as illustrated in Section 5.2.1, in high dimensions the eigenvalues of the SCM are poor estimates for the true eigenvalues. The sample eigenvalues spread over the positive real numbers. That is, the smallest eigenvalues will tend to zero, while the largest tend toward infinity (76; 146). Accordingly, SCM is unsatisfactory for large covariance matrix estimation problems.

Shrinkage Estimator

To overcome this drawback, it is common to regularize the estimator $\widehat{\boldsymbol{\Sigma}}$ with a highly structured estimator $\mathbf{T}$ via a linear combination $\alpha\widehat{\boldsymbol{\Sigma}} + (1-\alpha)\mathbf{T}$, where $\alpha \in [0,1]$. This technique is called regularization or *shrinkage*, since $\widehat{\boldsymbol{\Sigma}}$ is "shrunk" towards the structured estimator. The shrinkage helps to condition the estimator and avoid the problems of ill-conditioning in (5.4). The notion of shrinkage is based on the intuition that a linear combination of an *over-fit* sample covariance with some *under-fit* approximation will lead to an intermediate approximation that is "just-right" (267). A desired property of shrinkage is to maintain eigenvectors of the original estimator while conditioning on the eigenvalues. This is called *rotationally-invariant estimators* (256). Typically, $\mathbf{T}$ is set to $\rho\mathbf{I}$, where $\mathbf{I}$ is the identity matrix for some $\rho > 0$ and ρ is set by $\rho = \sum_{i=1}^{p} \sigma_{ii}/p$. In this case, the same shrinkage intensity is applied to all sample eigenvalue, regardless of their position. To illustrate the eigenvalues behaviour after shrinkage, let us consider the case of linear shrinkage intensity equal to 1/4, 1/2 and 3/4. Fig. 5.1 illustrates this case. As it was shown in (147), in the case of $\alpha = 1/2$, every sample eigenvalue is moved half-way towards the grand mean of all sample eigenvalues. Similarly, for $\alpha = 1/4$ eigenvalues are moved a quarter towards the mean of all sample eigenvalues. An alternative is the non-rotationally invariant shrinkage method, proposed by Hoffbeck and Landgrebe (117), uses the diagonal matrix $\mathbf{D} = \mathrm{diag}(\widehat{\boldsymbol{\Sigma}})$ which agrees with the SCM the diagonal entries, but shrinks the off-diagonal entries toward zero:

$$\widehat{\boldsymbol{\Sigma}}^{\alpha}_{\mathrm{diag}} = (1-\alpha)\widehat{\boldsymbol{\Sigma}} + \alpha\,\mathrm{diag}(\widehat{\boldsymbol{\Sigma}}) \tag{5.13}$$

However, in the experiments, we use a normalised version of (5.13), considering the dimension of the data, *i.e.*

$$\widehat{\boldsymbol{\Sigma}}^{\alpha}_{\mathrm{Stein}} = (1-\alpha)\widehat{\boldsymbol{\Sigma}} + \alpha\mathrm{Id}(\widehat{\boldsymbol{\Sigma}}) \tag{5.14}$$

where $\mathrm{Id}(\boldsymbol{\Sigma}) = \frac{\mathrm{tr}(\widehat{\boldsymbol{\Sigma}})\mathbf{I}}{p}$. This is sometimes called *ridge* regularization.

Regularized Tyler-estimator

Similarly, shrinkage can be applied to other estimators such as the robust estimator in (5.9). The idea was proposed in (2; 55; 294). Wiesel (294) gives the fixed point condition to compute a robust and well-conditioned estimator of $\boldsymbol{\Sigma}$ by

$$\begin{aligned}\tilde{\boldsymbol{\Sigma}}_{k+1} &= \frac{p}{n(1+\alpha)}\sum_{i=1}^{n}\frac{(\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^T}{(\mathbf{x}_i-\boldsymbol{\mu})^T\tilde{\boldsymbol{\Sigma}}_k^{-1}(\mathbf{x}_i-\boldsymbol{\mu})} + \frac{\alpha}{1+\alpha}\frac{p\mathbf{T}}{\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_k^{-1}\mathbf{T})} \\ \widehat{\boldsymbol{\Sigma}}_{k+1} &:= \frac{\tilde{\boldsymbol{\Sigma}}_{k+1}}{\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_{k+1})}.\end{aligned} \tag{5.15}$$

This estimator is a trade-off between the intrinsic robustness from M-estimators in (5.9) and the well-conditioning of shrinkage based estimators in section 5.4. The existence and uniqueness of this approach has been shown in (260). Nevertheless, the optimal value of shrinkage parameter α in (5.15) is still an open question.

Geodesic Interpolation in Riemannian Manifold

The shrinkage methods discussed so far involve the linear interpolation between two matrices, namely, a covariance matrix estimator and a target matrix. It can be extended to other types of interpolations, *i.e.* other space of representation for $\widehat{\boldsymbol{\Sigma}}$ and $\mathbf{T}$ different to the Euclidean space. A well-known approach is the Riemannian manifold of covariance matrices, *i.e.* the space of symmetric matrices with positive eigenvalues (207). In general, Riemannian manifold are analytical manifolds endowed with a distance measure which allows the measurement of similarity or dissimilarity (closeness or distance) of points. In this representation, the distance, called *geodesic distance*, is the minimum length of the curvature path that connects two points (53), and it can be computed by

$$\mathrm{Dist}_{\mathrm{Geo}}(\mathbf{A},\mathbf{B}) := \sqrt{\mathrm{tr}(\log^2(\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}))}. \tag{5.16}$$

This nonlinear interpolation, here called a *geodesic path* from $\mathbf{A}$ to $\mathbf{B}$ at time t, is defined by $\mathrm{Geo}_t(\mathbf{A},\mathbf{B}) := \mathbf{A}^{1/2}\exp(t\mathbf{M})\mathbf{A}^{1/2}$, where $\mathbf{M} = \log(\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2})$ and exp and log are matrix exponential and logarithmic functions respectively. A complete analysis of (5.16) and the geodesic path via its representation as ellipsoids have been presented in (26). Additionally, (26) shows that the volume of the geodesic interpolation is smaller than linear interpolation and thus it can increase detection performance in HSI detection problems. Thus, we have included a *Geodesic Stein estimation* with the same intuition behind equation (5.14) as follows,

$$\widehat{\boldsymbol{\Sigma}}^{\alpha}_{\text{Geo-Stein}} = \mathrm{Geo}_{\alpha}(\widehat{\boldsymbol{\Sigma}}, \mathrm{Id}(\widehat{\boldsymbol{\Sigma}})), \tag{5.17}$$

where $\alpha \in [0,1]$ determines the trade-off between the original estimation $\widehat{\boldsymbol{\Sigma}}$ and the well-conditioning $\mathrm{Id}(\widehat{\boldsymbol{\Sigma}})$.

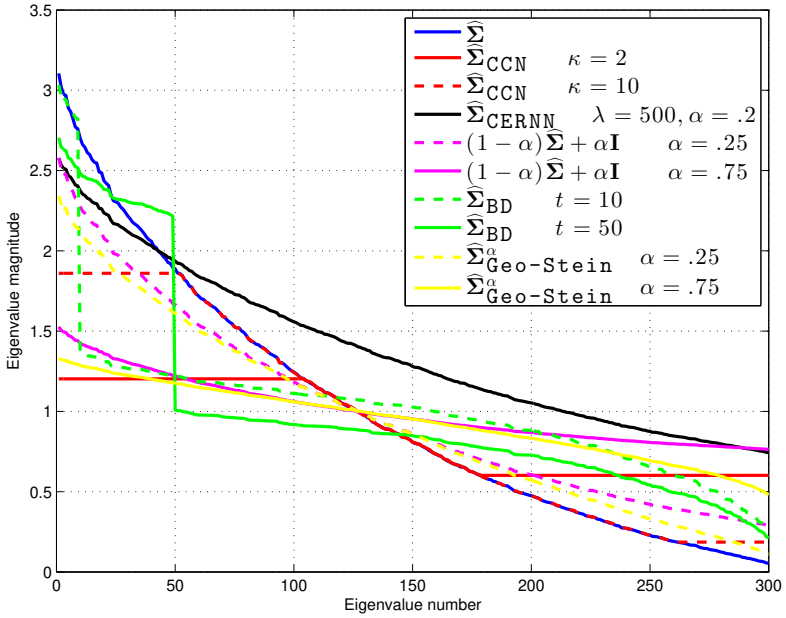


Figure 5.1 CCN truncates extreme sample eigenvalues and leaves the moderate ones unchanged and CERNN gives the contrary effect. Linear and geodesic shrinkages moves eigenvalues towards the grand mean of all sample eigenvalues. However, the effect of geodesic shrinkage is more attenuated for extreme eigenvectors than in linear case. The effect of BD-correction depends on the eigenvalues sets defined by t.

Constrained MLE

As we have shown in Section 5.2.1, even when $n > p$, the eigenstructure tends to be systematically distorted unless p/n is extremely small, resulting in ill-conditioned estimators for $\mathbf{\Sigma}$. Recently, several works have proposed regularising the SCM by explicitly imposing a constraint on the condition number. (296) proposes to solve the following constrained MLE problem:

$$\text{maximize } \mathcal{L}(\mathbf{\Sigma}) \text{ subject to } cond(\mathbf{\Sigma}) \leq \kappa \tag{5.18}$$

where $\mathcal{L}(\mathbf{\Sigma})$ stands for the log-likelihood function in the Gaussian distributions. This problem is hard to solve in general. However, (296) proves that in the case of rotationally-invariant estimators, (5.18) reduces to an unconstrained univariate optimisation problem. Furthermore, the solution of (5.18) is a nonlinear function of the sample eigenvalues given by:

$$\widehat{\lambda}_i = \begin{cases} \eta, & \lambda_i(\widehat{\mathbf{\Sigma}}) \leq \eta \\ \lambda_i(\widehat{\mathbf{\Sigma}}), & \eta < \lambda_i(\widehat{\mathbf{\Sigma}}) < \eta\kappa \\ \kappa\eta, & \lambda_i(\widehat{\mathbf{\Sigma}}) \geq \eta\kappa \end{cases} \tag{5.19}$$

for some η depending on κ and $\lambda(\widehat{\mathbf{\Sigma}})$. We refer this methodology as *Condition Number-Constrained* (CCN) estimation.

Covariance Estimate Regularized by Nuclear Norms

Instead of constrain the MLE problem in (5.18), (58) propose to penalise the MLE as follows,

$$\text{maximize } \mathcal{L}(\boldsymbol{\Sigma}) + \frac{\lambda}{2}[\alpha||\boldsymbol{\Sigma}||_* + (1-\alpha)||\boldsymbol{\Sigma}^{-1}||_*] \tag{5.20}$$

where the nuclear norm of a matrix $\boldsymbol{\Sigma}$, is denoted by $||\boldsymbol{\Sigma}||_*$, is the sum of the eigenvalues of $\boldsymbol{\Sigma}$, λ is a positive strength constant, and $\alpha \in (0,1)$ is a mixture constant. We refer this approach by the acronym CERNN (Covariance Estimate Regularised by Nuclear Norms).

Ben-David and Davidson correction

Given zero-mean[3] data with normal probability density $\mathbf{x} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, its sampled covariance matrix $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n-1}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T$ follows a central Wishart distribution with n degrees of freedom. The study of covariance estimators in Wishart distribution where the sample size (n) is small in comparison to the dimension (p) is also an active research topic (25; 174; 191). Firstly, Efron and Morris proposed a rotationally-invariant estimator of $\boldsymbol{\Sigma}$ by replacing the sampled eigenvalues with an improved estimation (77). Their approach is supported by the observation that for any Wishart matrix, the sampled eigenvalues tend to be more spread out than population eigenvalues, in consequence, smaller sampled eigenvalues are underestimated and large sampled eigenvalues are overestimated (25). Accordingly, they find the best estimator of inverse of the covariance matrix of the form $a\widehat{\boldsymbol{\Sigma}}^{-1} + b\mathbf{I}/\text{tr}(\widehat{\boldsymbol{\Sigma}})$ which is achieved by:

$$\widehat{\boldsymbol{\Sigma}}_{\text{Efron-Morris}} = \left((n-p-1)\widehat{\boldsymbol{\Sigma}}^{-1} + \frac{p(p+1)-2}{\text{tr}(\widehat{\boldsymbol{\Sigma}})}\mathbf{I} \right)^{-1}. \tag{5.21}$$

It is worth mentioning that other estimations have been developed following the idea behind Wishart modelling and assuming a simple model for the eigenvalue structure in the covariance matrix (usually two phases model). Recently, Ben-David and Davidson (25) have introduced a new approach for covariance estimation in HSI, called here *BD-correction*. From the SVD of $\widehat{\boldsymbol{\Sigma}} = \mathbf{U}\Lambda_{\widehat{\boldsymbol{\Sigma}}}\mathbf{U}^T$, they proposed a rotationally-invariant estimator by correcting the eigenvalues by means of two diagonal matrices,

$$\widehat{\boldsymbol{\Sigma}}_{\text{BD}} = \mathbf{U}\Lambda_{\text{BD}}\mathbf{U}^T, \quad \text{with} \quad \Lambda_{\text{BD}} = \Lambda_{\widehat{\boldsymbol{\Sigma}}}\Lambda_{\text{Mode}}\Lambda_{\text{Energy}}. \tag{5.22}$$

They firstly estimate the apparent multiplicity p_i of the i-th sample eigenvalue as $p_i = \sum_{j=1}^{p} card[a(j) \leq b(i) \leq b(j)]$, where $a(i) = \Lambda_{\widehat{\boldsymbol{\Sigma}}}(i)(1-\sqrt{c})^2$ and $b(i) = \Lambda_{\widehat{\boldsymbol{\Sigma}}}(i)(1+\sqrt{c})^2$. One can interpret the concept of "apparent multiplicity" as the number of distinct eigenvalues that are "close" together and thus represent nearly the same eigenvalue (25). Secondly, BD-correction affects the i-th sample eigenvalue via its apparent multiplicity p_i as $\Lambda_{Mode}(i) = \frac{(1+p_i/n)}{(1-p_i/n)^2}$ and as

$$\Lambda_{Energy}(i) = \begin{cases} \sum_{i=1}^{t} \Lambda_{\widehat{\boldsymbol{\Sigma}}}(i) / \sum_{i=1}^{t} (\Lambda_{\widehat{\boldsymbol{\Sigma}}}(i)\Lambda_{\text{Mode}}(i)) \\ \sum_{i=t+1}^{p} \Lambda_{\widehat{\boldsymbol{\Sigma}}}(i) / \sum_{i=t+1}^{p} (\Lambda_{\widehat{\boldsymbol{\Sigma}}}(i)\Lambda_{\text{Mode}}(i)) \end{cases} \tag{5.23}$$

[3] Or $\boldsymbol{\mu}$ known, in which case, one might subtract $\boldsymbol{\mu}$ from the data.

for a value $t \in [1, \min(n,p)]$ indicating the transition between large and small eigenvalues. Finally, reader can see (25) for an optimal selection of t. A comparison of correction in the eigenvalues by CCN, CERNN, the linear shrinkage in (5.14), the geodesic Stein in (5.17) and the BD-correction is illustrated in Fig. 5.1 for three values of regulation parameter. We can see that CCN truncates extreme sample eigenvalues and leaves the moderate ones unchanged. Compared to the linear estimator, both (5.18) and (5.20) pull the larger eigenvalues down more aggressively and pull the smaller eigenvalues up less aggressively.

5.4.1 Sparse Matrix Transform

Recently, (45; 267) introduced the *sparse matrix transform* (SMT). The idea behind is the estimation of the SVD from a series of *Givens rotations*, *i.e.*, $\widehat{\mathbf{\Sigma}}_{\text{SMT}} = \mathbf{V}_k \mathbf{\Lambda} \mathbf{V}_k^T$, where $\mathbf{V}_k = \mathbf{G}_1 \mathbf{G}_2 \cdots \mathbf{G}_k$ is a product of k *Givens rotation* defined by $\mathbf{G} = \mathbf{I} + \boldsymbol{\vartheta}(i,j,\vartheta)$ where

$$\boldsymbol{\vartheta}(a,b,\vartheta) = \begin{cases} \cos(\vartheta) - 1, & \text{if } r = s = a \text{ or } r = s = b \\ \sin(\vartheta), & \text{if } r = a \text{ and } s = b \\ -\sin(\vartheta), & \text{if } r = b \text{ and } s = a \\ 0, & \text{otherwise} \end{cases}$$

where each step $i \in \{1, \ldots, k\}$ of the SMT is designed to find the single Givens rotation that minimise $\text{diag}(\mathbf{V}_i^T \widehat{\mathbf{\Sigma}} \mathbf{V}_i)$ the most. The details of this transformation are given in (45; 46). The number of rotations k is a parameter and it can be estimated from heuristic Wishart estimator as in (267). However, in the numerical experiments, this method of estimating k tended to over-estimate. As such, SMT is compared with k as function of p in our experiments.

Table 5.1 Covariance matrix estimators considered in this section

Name	Notation	Formula
SCM	$\widehat{\mathbf{\Sigma}}$	$\frac{1}{n}\sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T$
Stein Shrinkage (146)	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Stein}}$	$(1-\alpha)\widehat{\mathbf{\Sigma}} + \alpha \text{Id}(\widehat{\mathbf{\Sigma}})$
Tyler (269)	$\widehat{\mathbf{\Sigma}}_{\text{Tyler}}$	$\widehat{\mathbf{\Sigma}}_{j+1} = \frac{p}{n}\sum_{i=1}^n \frac{(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T}{(\mathbf{x}_i - \boldsymbol{\mu})^T \widehat{\mathbf{\Sigma}}_j^{-1} (\mathbf{x}_i - \boldsymbol{\mu})}$
Tyler Shrinkage (56)	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Tyler}}$	$\tilde{\mathbf{\Sigma}}_{k+1} = \frac{1}{1+\alpha}\frac{p}{n}\sum_{i=1}^n \frac{\mathbf{x}\mathbf{x}^T}{\mathbf{x}_i^T \widehat{\mathbf{\Sigma}}_k^{-1} \mathbf{x}_i} + \frac{\alpha}{1+\alpha}\frac{p\mathbf{T}}{\text{tr}(\widehat{\mathbf{\Sigma}}_k^{-1}\mathbf{T})}$
Sparse Matrix Transform (SMT) (267)	$\widehat{\mathbf{\Sigma}}_{\text{SMT}}$	$\mathbf{G}_1 \mathbf{G}_2 \cdots \mathbf{G}_k \mathbf{\Lambda} (\mathbf{G}_1 \mathbf{G}_2 \cdots \mathbf{G}_k)^T$
t distribution(143)	$\widehat{\mathbf{\Sigma}}_{\text{t}}$	$\widehat{\mathbf{\Sigma}}_{j+1} = \frac{1}{n}\sum_{i=1}^n \frac{(v+p)(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T}{v + (\mathbf{x}_i - \boldsymbol{\mu})^T \widehat{\mathbf{\Sigma}}_j^{-1} (\mathbf{x}_i - \boldsymbol{\mu})}$
Geodesic Stein	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Geo-Stein}}$	$\text{Geo}_{\alpha}(\widehat{\mathbf{\Sigma}}, \text{Id}(\widehat{\mathbf{\Sigma}}))$
Constrained condition number(296)	$\widehat{\mathbf{\Sigma}}_{\text{CCN}}$	(5.19)
Covariance Estimate Regularized by Nuclear Norms(58)	$\widehat{\mathbf{\Sigma}}_{\text{CERNN}}$	(5.20)
Efron-Morris Correction (77)	$\widehat{\mathbf{\Sigma}}_{\text{Efron-Morris}}$	(5.21)
Ben-Davidson Correction (25)	$\widehat{\mathbf{\Sigma}}_{\text{BD}}$	(5.22)

Finally, as comparison in (286), RX-detector in (5.2) by using a large number of covariance matrix estimation is reported in 5.2

Table 5.2 Top-3 performances in different analysed scenarios in (286)

Distribution	Contamination	$c = \frac{p}{n}$	Top three performances		
Gaussian	No	0.2	$\widehat{\mathbf{\Sigma}}_{\text{BD}}$	$\widehat{\mathbf{\Sigma}}_{\text{Efron-Morris}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Geo-Stein}}$
Gaussian	1%	0.2	$\widehat{\mathbf{\Sigma}}_{\text{t}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Tyler}}$	—
Gaussian	10%	0.2	$\widehat{\mathbf{\Sigma}}_{\text{t}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Tyler}}$	—
Gaussian	No	0.9	$\widehat{\mathbf{\Sigma}}_{\text{SMT}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Stein}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Tyler}}$
Gaussian	1%	0.9	$\widehat{\mathbf{\Sigma}}_{\text{SMT}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Stein}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Tyler}}$
Gaussian	10%	0.9	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Tyler}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Stein}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Geo-Stein}}$
Cauchy	No	0.2	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Tyler}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Geo-Stein}}$	$\widehat{\mathbf{\Sigma}}_{\text{t}}$
Cauchy	1%	0.2	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Tyler}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Geo-Stein}}$	$\widehat{\mathbf{\Sigma}}_{\text{CERNN}}$
Cauchy	10%	0.2	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Tyler}}$	$\widehat{\mathbf{\Sigma}}_{\text{t}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Geo-Stein}}$
Cauchy	No	0.9	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Geo-Stein}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Tyler}}$	$\widehat{\mathbf{\Sigma}}_{\text{CERNN}}$
Cauchy	1%	0.9	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Tyler}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Geo-Stein}}$	$\widehat{\mathbf{\Sigma}}_{\text{CERNN}}$
Cauchy	10%	0.9	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Tyler}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Geo-Stein}}$	$\widehat{\mathbf{\Sigma}}_{\text{CCN}}$
Dirichlet	No	0.2	$\widehat{\mathbf{\Sigma}}_{\text{SMT}}$	$\widehat{\mathbf{\Sigma}}_{\text{BD}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Tyler}}$
Dirichlet	10%	0.2	$\widehat{\mathbf{\Sigma}}_{\text{CCN}}$	—	—
Dirichlet	No	0.9	$\widehat{\mathbf{\Sigma}}_{\text{SMT}}$	$\widehat{\mathbf{\Sigma}}_{\text{BD}}$	$\widehat{\mathbf{\Sigma}}^{\alpha}_{\text{Tyler}}$
Dirichlet	10%	0.9	$\widehat{\mathbf{\Sigma}}_{\text{CCN}}$	—	—

5.5 Approaches based on Machine and Deep Learning

5.5.1 Introduction

Instead of making assumptions about the nature of the data distribution, in machine learning, optimisation problems are built based on geometric intuitions. For the case of anomaly detection, the classic machine learning methods learn feature representation of normality. In this chapter only a second family of detectors based on the method of one-class SVM. For readers interested in a state of the art of anomaly detection methods should consult (203; 231).

5.5.2 One-class SVM

In (27) a mathematical model for learning the high-density areas of an unknown distribution from random points drawn according to this distribution is proposed. Theoretical bounds on the error were obtained for finite n and are independent of the underlying distribution. This is the based formalism for one-class SVM (265) (OCSVM), who aims at finding a minimum radius hypersphere to surround the majority of the data, allowing a small fraction of "outliers" to fall outside. In order to control this number a penalty for outliers must be incorporated into the objective function of the OCSVM learning model.

$$\min_{R,\mathbf{c},\epsilon} R^2 + \lambda \sum_{i=i}^{n} \epsilon_i \tag{5.24}$$

subject to :

$$||\mathbf{x}_i - \mathbf{c}||^2 \leq R^2 + \epsilon \quad \forall i = 1, \ldots, n$$
$$\epsilon_i \geq 0 \quad \forall i = 1, \ldots, n$$

where λ is penalty parameter, the slack variables ϵ create a soft-margin and the parameter $\mathbf{c}$ is the center of the hypersphere. The minimisation problem is solved by the introduction of Lagrange multipliers and the norm in the first constraint can consider kernel functions (265).

5.5.3 Deep support vector data description

The objective of Deep Support vector data description (SVDD) is to learn a neural network that minimises the volume of a data-enclosing hypersphere (232). It uses a neural network to project the supposedly normal training samples in a latent space so that all samples are within a normality hypersphere. The hypersphere is made as small as possible thanks to a suitable training loss, corresponding to

$$\min_{\boldsymbol{\theta}} \left[\frac{1}{n} \sum_{1}^{n} ||\Phi(\mathbf{x}_i; \boldsymbol{\theta}) - \mathbf{c}||^2 + \frac{\lambda}{2} \sum_{i=1}^{L} ||\boldsymbol{\theta}^L||_F^2 \right] \tag{5.25}$$

where $\mathbf{c}$ is the normality hypersphere center, determined by the mean latent coordinates of an initial forward pass of the training data. Φ represents the encoding neural network, $\boldsymbol{\theta}$ its weights, $\mathbf{x}_i$ the sample being projected, l a layer index. A common regularisation is performed on the network weights using the Frobenius norm, and is balanced through λ with the main training objective.

For a given test point $\mathbf{x}$, we can naturally define an anomaly score by the distance of the point to the center of the hypersphere, *i.e.*

$$\mathtt{AD}_{\mathtt{DSVDD}}(\mathbf{x}, \tau_2) = ||\Phi(\mathbf{x}; \boldsymbol{\theta}^*) - \mathbf{c}||^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau_2 \tag{5.26}$$

where $\boldsymbol{\theta}^*$are the network parameters of a trained model

5.5.4 Approximation by random projection depth

In this subsection, we analyse the case of anomaly detection when the projected subspace is unidimensional, *i.e.*, $K = 1$ and the projection matrix is in this case a ℓ^1 norm vector $\mathbf{u}$. We show that if the unidimensional detection is performed a number of times r, we can obtain a performance equivalent to RX-detection when r tends to infinity. Additionally, it allows to include a robust estimation and local anomaly detection on a RX-fashion without any covariance matrix estimation, *i.e.*, we introduce the detector:

$$\mathcal{D}_{\mathrm{PD}}(\mathbf{x}, \tau_3) := \sup_{||\mathbf{u}||=1} \frac{(\mathbf{u}^T\mathbf{x} - \mu(\mathbf{u}^T\mathbf{X}))^2}{\sigma^2(\mathbf{u}^T\mathbf{X})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau_3 \tag{5.27}$$

where $\mu(\mathbf{u}^T\mathbf{X})$ and $\sigma^2(\mathbf{u}^T\mathbf{X})$ denotes the mean and variance in the projection of the data $\mathbf{X}$ in $\mathbf{u}$. Note that $\mathbf{u}^T\mathbf{X}$ is a scalars vector, and (μ, σ^2) can be substituted by robust estimators of

first and second order statistic, for instance,

$$\mathcal{D}_{\mathrm{D}}(\mathbf{x},\tau_4) := \sup_{||\mathbf{u}||=1} \frac{(\mathbf{u}^T\mathbf{x} - \mathtt{MED}(\mathbf{u}^T\mathbf{X}))^2}{\mathtt{MAD}^2(\mathbf{u}^T\mathbf{X})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau_4 \tag{5.28}$$

where (MED,MAD) are the median and the median absolute deviation. These robust estimators are included to avoid "masking and swamping effect" in the detector. The absolute value of the square root of expression (5.28) was introduced by D. Donoho (72) as a multivariate measure of outlyingness.

Equivalence of Random Projections and RX Anomaly Detector

In this section, the equivalence of the proposed detectors (5.27) and (5.28) is proved for the family of Elliptically Symmetric Distributions (ESD) (78).

Definition 5.1. A random p-dimensional vector $\mathbf{x}$ is spherically distributed if $\mathbf{x}$ and $\mathbf{Q}^T\mathbf{x}$ have the same distribution for $\mathbf{O}_{p\times p} \in \mathcal{O}(p)$, the set of all orthogonal matrices of dimension p defined as $\mathcal{O}(p) = \{\mathbf{Q} \in \mathcal{M}(p) : \mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}_p\}$

The following necessary and sufficient condition hold(78). A random p-dimensional vector $\mathbf{x}$ is spherically distributed if and only if $\varphi_{\mathbf{x}}(\mathbf{Q}^T\mathbf{t}) = \varphi_{\mathbf{x}}(\mathbf{t}) = \phi(\mathbf{t}^T\mathbf{t})$ where $\varphi_{\mathbf{x}}(\mathbf{t})$ is its characteristic function and $\phi(\cdot)$ is a scalar function. Examples on this family of distributions include the Standard Normal, t, Cauchy, Bessel, Kotz type distribution among others. A complete list including the characteristic functions is due to D.R. Jensen in Table 2.1 of (138).

Definition 5.2. The p-dimensional random vector $\mathbf{x}$ is elliptically distributed with parameter $\boldsymbol{\mu}_{p\times 1}$ and a scale parameter $\mathbf{V}_{p\times p}$, written as $\mathbf{x} \sim E_p(\boldsymbol{\mu}, \mathbf{V}, \phi)$, if its characteristic function can be expressed as:

$$\varphi_{\mathbf{x}}(\mathbf{t}) = \exp(i\mathbf{t}^T\boldsymbol{\mu})\phi\left(\mathbf{t}^T\mathbf{V}\mathbf{t}\right) \tag{5.29}$$

for some vector $\boldsymbol{\mu}$, positive-definite matrix $\mathbf{V}$, and for some function φ, which is called the characteristic generator.

For $\mathbf{x} \sim E_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$, it does not generally follow that $\mathbf{x}$ has a density $f_{\mathbf{X}}(\mathbf{x})$, but, if it exists, it has the following form:

$$\mathbf{f}(\mathbf{x}) = \frac{c_p}{\sqrt{|\boldsymbol{\Sigma}|}} g_p\left[(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] \tag{5.30}$$

where c_p is the normalisation constant and g_p is some nonnegative function with $(\frac{p}{2}-1)$-moments finite. g_p is called *density generator* (78). In this case we shall use the notation $E_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g_p)$ instead of $E_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$. We now state the proposition for the case of standardised random projections.

Proposition 5.3. *Let $\mathbf{X}_{(p\times n)}$ be a i.i.d. random sample of size n, where $\mathbf{x}_i \sim E_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g_p)$, then:*

$$\mathcal{D}_{PD}(\mathbf{x}) = (\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \tag{5.31}$$

Proposition 5.4. *If* $\mathbf{x}$ *has a univariate symmetric probability density function* $f_{\mathbf{x}}$*, we denote* $\Phi_{\mathbf{x}}$ *the cumulative distribution that is continuous and positive on its support then*

$$\frac{\mathit{MAD}(\mathbf{x})}{\sigma(\mathbf{x})} = \Phi_{\mathbf{x}}^{-1}(3/4) \tag{5.32}$$

Finally, we provide the corresponding proposition for the case of projection depth function in elliptically symmetric random variables.

Proposition 5.5. *Let* $\mathbf{X}$ *be a i.i.d. random sample of size* n*, where* $\mathbf{x}_i \sim E_p(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Sigma}_{\mathbf{X}}, g_p)$*, then:*

$$c_{g_p}\mathcal{D}_D(\mathbf{x}) = (\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \tag{5.33}$$

with $c_{g_p} = (\Phi^{-1}(3/4))^2$.

5.5.5 From RPO to deep RPO

Whereas deep SVDD uses an Euclidean distance to the normality hypersphere center in the latent space, the present work evaluates the distance to various location estimators provided by a diversity of untrainable random projections. The outlyingness of (5.27) replaces the distance to a single hypersphere center to quantify abnormality in the latent space. The deep SVDD training objective of (5.25) therefore changes into:

$$\min_{\boldsymbol{\theta}} \left[\frac{1}{n}\sum_{i=1}^{n} \max_{\mathbf{u}\in\mathcal{S}^{p-1}} \frac{|\mathbf{u}^T\Phi(\mathbf{x}_i;\boldsymbol{\theta}) - \mu(\mathbf{u}^T\Phi(\mathbf{x}_i;\boldsymbol{\theta}))|}{\sigma(\mathbf{u}^T(\Phi(\mathbf{x}_i;\boldsymbol{\theta})))} + \frac{\lambda}{2}\sum_{l=1}^{L}||\boldsymbol{\theta}^l||_F^2 \right] \tag{5.34}$$

which (μ,σ) a pair of estimators of the first and second moment of the data on the projection space.

A second Deep SVDD derivative considered here is Deep RPO (21), which replaces the latent Euclidean distance to the normality centroid with a RPs-based outlyingness measure adapted in the latent space, leading to the following loss:

$$\min_{\boldsymbol{\theta}} \left[\frac{1}{n}\sum_{i=1}^{n} \left(\underset{\mathbf{u}\in\mathcal{S}^{p-1}}{mean} \frac{|\mathbf{u}^T\Phi(\mathbf{x}_i;\boldsymbol{\theta}) - \mathtt{MED}(\mathbf{u}^T\Phi(\mathbf{X};\boldsymbol{\theta}))|}{\mathtt{MAD}(\mathbf{u}^T\Phi(\mathbf{X};\boldsymbol{\theta}))} \right) + \frac{\lambda}{2}\sum_{l=1}^{L}||\boldsymbol{\theta}^l||_F^2 \right] \tag{5.35}$$

This training loss uses the outlyingness defined by random projections after the neural network encoding, with a max estimator transformed into a *mean* as suggested in (21) for better integration with the deep learning setup. The *mean* estimator computes a mean over the set of RPs available to compute the latent outlyingness, while the $\frac{1}{n}$ computes a mean over the batch samples. The use of a *mean* instead of a max removes the convergence to the square root of the Mahalanobis distance-inferred ellipsoid already mentioned in the RPO definition for a large set of RPs. The loss nonetheless still combines 1D outlyingness measures individually centred by their median and normalised by their median absolute deviation, but with no ellipsoid-like score distribution guarantee in the input space once integrated. No square was applied to the first loss term, in accordance with (72).

5.5.6 Multisphere case

An extension is Deep multi-sphere SVDD (MSVDD) (87) initialises numerous latent normality hyperspheres using k-means and progressively discards the irrelevant centroids during training. The relevance of latent hyperspheres is determined thanks to the cardinality of the latent cluster they encompass. The deep MSVDD training loss is:

$$\underset{\boldsymbol{\theta},R}{min}\left[\frac{1}{K}\sum_{k=1}^{K}R_k^2+\frac{1}{\nu n}\sum_{i=1}^{n}\max(0,||\Phi(\mathbf{x}_i;\boldsymbol{\theta})-\mathbf{c}_{j(i)}||^2-R_{j(i)}^2)+\frac{\lambda}{2}\sum_{l=1}^{L}||\boldsymbol{\theta}^l||_F^2\right] \tag{5.36}$$

The first term minimises the volume of hyperspheres of radius R_k, while the second is controlled by $\nu\in[0,1]$ and penalises points lying outside of their assigned hypersphere, training samples being assigned to the nearest hypersphere denoted by $\mathbf{c}_{j(i)}$.

5.5.7 Semisupervised case

SAD is achieved through outlier exposure (113; 233), which adds supervision to the training of the model thanks to the availability of few and non representative labeled anomalies. To take into account anomalies during training, Deep SAD (233) repels the outliers from the normality centroid by replacing the minimisation of the distance to the centroid with the minimisation of its inverse in the training loss. With m labeled anomalies $\tilde{\mathbf{x}}$ in a batch the Deep SVDD loss thus becomes, with a training objective balancing parameter η:

$$\min_{\theta}\left[\frac{1}{n+m}\sum_{i=1}^{n}||\Phi(\mathbf{x}_i;\boldsymbol{\theta})-\mathbf{c}||^2+\frac{\eta}{n+m}\sum_{j=1}^{m}(||\Phi(\widetilde{\mathbf{x}}_j;\theta)-\mathbf{c}||^2)^{-1}+\frac{\lambda}{2}\sum_{l=1}^{L}||\boldsymbol{\theta}^l||_F^2\right] \tag{5.37}$$

Labeled anomalies in the training set need to be distinguished from potential unlabelled anomalies that are considered to be normal samples, which confuse the AD by contaminating the training set instead of providing supervision. This adaptation can be repeated for both Deep RPO and Deep MSVDD, although for Deep MSVDD the multiplicity of normality centres calls for an additional consideration on how to choose from which centroid the labeled anomalies should be repelled as long as several centroids are kept active. The experiments implementing Deep MSVDD adapted to SAD with an additional loss term for labeled anomalies were inconclusive, such an adaptation will therefore not be part of the presented results. The additional loss term either minimised the latent distance between anomalies and dedicated centroids, or maximised the latent distance between anomalies and normality centroids. Once adapted to SAD, the Deep RPO loss becomes, similarly to the transformation that led to Eq. 5.37:

$$\begin{aligned}\min_{\theta}\Bigg[&\frac{1}{n+m}\sum_{i=1}^{n}\left(\underset{\mathbf{u}\in\mathcal{S}^{p-1}}{mean}\frac{|\mathbf{u}^T\Phi(\mathbf{x}_i;\boldsymbol{\theta})-\texttt{MED}(\mathbf{u}^T\Phi(\mathbf{X};\boldsymbol{\theta}))|}{\texttt{MAD}(\mathbf{u}^T\Phi(\mathbf{X};\boldsymbol{\theta}))}\right)\\&+\frac{\eta}{n+m}\sum_{j=1}^{m}\left(\underset{\mathbf{u}\in\mathcal{S}^{p-1}}{mean}\frac{|\mathbf{u}^T\Phi(\widetilde{\mathbf{x}_j};\boldsymbol{\theta})-\texttt{MED}(\mathbf{u}^T\Phi(\mathbf{X};\boldsymbol{\theta}))|}{\texttt{MAD}(\mathbf{u}^T\Phi(\mathbf{X};\boldsymbol{\theta}))}\right)^{-1}+\frac{\lambda}{2}\sum_{l=1}^{L}||\boldsymbol{\theta}^l||_F^2\Bigg]\end{aligned} \tag{5.38}$$

Arbitrary sets of outliers could not be completely gathered around a reference point since they are not necessarily concentrated in a common mode (233; 258). However, this does not forbid the concentration of identified modes among labeled anomalies close to dedicated centroids to provide additional supervision during training, a case which is part of the experiments presented. The possibly arbitrary distribution of normal and anomalous centroids and the relative distance between the centroids adds a way to use prior information regarding the proximity between the training samples. Such a setup can seem close to classification with rejection (20; 113), since the concentration of data points around dedicated normal and anomalous centroids can be interpreted as classification while the data points attached to no centroid and thus supposedly repelled from all centroids by the trained network constitutes a rejection. This parallel with classification with rejection is not necessarily relevant since the availability of labeled anomalies to train AD methods is usually very limited if not nonexistent. In contrast, supervised classification of identified data modes would imply rich, representative and relatively balanced datasets for each latent mode. The limited availability of labeled anomalies applies to actual anomalies and not to artificial anomalies provided by the transformation of existing training samples i.e. through self-supervision. With proper transformations self-supervision can produce as many labeled anomalies for training as there are normal samples, or even more if each normal sample is transformed multiple times. However this does not overcome the lack of representativeness of labeled anomalies. This is also made difficult since the choice of transformations requires expert knowledge.

The reunion of normal latent representations achieved through the deep one-class classification methods mentioned is analogous to the alignment principle put forward in (291), which also argued for a latent uniformity. Whereas the alignment principle compels similar samples to be assigned similar representations, the uniformity principle demands the preservation of maximal information. One way to achieve that according to (291) is to push all features away from each other on the unit hypersphere to intuitively facilitate a uniform distribution. The extension of the Deep SVDD loss to encourage a form of latent uniformity using the pairwise distance between normal samples during training was investigated without ever improving the baselines. The experiments conducted to evaluate the contribution of a pairwise distance of normal samples latent representations loss term revolved around the following training loss format, where the term tasked with enforcing latent uniformity is weighted using β and was expected to be judiciously balanced with the overall latent concentration:

$$\min_{\theta} \left[\frac{1}{n} \sum_{i}^{n} |\Phi(\mathbf{x}_i; \boldsymbol{\theta}) - \mathbf{c}|^2 + \frac{\beta}{n} \sum_{i \neq j}^{n} \left(|\mathbf{x}_i - \mathbf{x}_j|^2 \right)^{-1} + \frac{\lambda}{2} \sum_{l=1}^{L} ||\boldsymbol{\theta}^l||_F^2 \right] \tag{5.39}$$

The failure to make a loss term enforce a form of latent uniformity could signal the necessity of associating such a constraint with latent representations confined to a relevant manifold.

As example, unsupervised AD results, for which the training is only supervised by normal training samples, are presented in Table 5.3 for low-resolution radar micro-doppler signature in (22). These results indicate the superiority of deep learning for the OOD task considered, while demonstrating the substantial contribution of geometry-aware dimensionality reduction through the use of tPCA for non-deep AD. RPO is kept in Table 5.3 even though it does not achieve

Table 5.3 Unsupervised AD experiments results (average test AUCs in % ± StdDevs over ten seeds). These machine learning methods are trained on fully normal training sets, without labeled anomalies for SAD or self-supervision transformations. The four last methods are our deep AD baselines, trained on normalised spectral representations only. Deep MSVDD "mean best" indicates the neural network was trained using a simpler loss, analogous to the Deep SVDD loss, where only the distance to the best latent normality centroid is minimised. PCA and tPCA indicate that the AD model is trained after an initial dimensionality reduction, which is either PCA or tangent PCA.

AD method (input format)	Mean test AUC (1 mode)	Mean test AUC (2 modes)
OC-SVM (SP-PCA)	49.16 ± 26.69	45.48 ± 27.53
OC-SVM (SPD-PCA)	64.68 ± 9.10	58.23 ± 15.12
OC-SVM (SPD-tPCA)	57.59 ± 3.91	55.33 ± 9.48
IF (SP-PCA)	50.96 ± 17.37	48.50 ± 18.76
IF (SPD-PCA)	52.36 ± 22.47	47.50 ± 20.32
IF (SPD-tPCA)	66.91 ± 9.65	61.23 ± 12.65
LOF (SP-PCA)	56.80 ± 2.38	61.55 ± 10.29
LOF (SPD-PCA)	66.44 ± 21.37	65.83 ± 19.52
LOF (SPD-tPCA)	78.38 ± 8.86	73.56 ± 10.09
RPO (SP-PCA)	49.61 ± 6.89	50.43 ± 7.13
RPO (SPD-PCA)	51.08 ± 19.66	54.95 ± 17.58
RPO (SPD-tPCA)	33.97 ± 7.36	38.08 ± 14.58
Deep SVDD (SP)	83.03 ± 6.83	**78.29 ± 6.68**
Deep MSVDD (SP)	82.27 ± 9.67	**78.30 ± 8.28**
Deep MSVDD "mean best" (SP)	82.29 ± 7.20	78.02 ± 6.80
Deep RPO (SP)	**83.60 ± 5.35**	78.13 ± 6.02

useful discrimination because it is the shallow equivalent of Deep RPO, one of the highlighted deep AD methods, deprived of the neural network encoder and with a *max* estimator instead of a *mean*, as was previously justified. Deep MSVDD does not lead to the best performances, and is as effective as Deep SVDD and Deep RPO, which could have seemed surprising at least when normality is made of two target classes. Indeed, since Deep MSVDD has the possibility to use several disjointed hyperspheres to capture the latent normality distribution, one could expect it to better model more complex, e.g. multimodal, normality.

5.6 Perspectives

- Anomaly Detection and Continual Learning: Most of the industrial anomaly data are presented in an assembly line. Accordingly the anomaly detection model should overcome the catastrophic forgetting phenomenon, considering for instance decision from multiple detectors.
- Anomaly detection can be seen as an AutoML problem, since one want to produce methods that are able to work well for different types of data distribution. In this context, is it important to define statistics (determine the optimal ones) that allow to establish when an anomaly detection algorithm is well adapted to a problematic data?

- Invariance and interpretation of the decision is fundamental in AD for industrial problems. Methods indicating why an object is determined to be abnormal and the confidence that can be placed in this decision will be explored in the near future.
- The methods for the detection of multimodal anomalies in the case of contaminated bases must be studied to approach the industrial needs of today.
- Covariance matrix estimations for AD will be explored in the context of robust deep learning methods.

Related Publications

1. **Robust RX Anomaly Detector without covariance matrix estimation**, S.Velasco-Forero and J. Angulo, 2012 4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), June, pp. 1-4, 2012.
2. **Robust Anomaly Detection in Hyperspectral imaging**, J. Frontera-Pons et al., IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4604-4607, 2014
3. **Comparative Analysis of Covariance Matrix Estimation for Anomaly Detection in Hyperspectral Images**, S. Velasco-Forero, M.Chen, A. Goh and S.K. Pang, Volume 9, nro. 6, Sept 2015, 1061–1073, IEEE Journal of Selected Topics in Signal Processing.
4. **From unsupervised to semi-supervised anomaly detection methods for High Resolution Range Profiles**, M. Bauw et al., IEEE Radar Conference 2020.
5. **Near out-of-distribution detection for low-resolution radar micro-Doppler signatures**, M.Bauw et al., European Conference on Machine Learning PKDD, 2022.

6 Applications

> A picture may be worth a thousand words, a formula is worth a thousand pictures
>
> *Edsger Dijkstra*

Much of the motivation to seek solutions from a theoretical point of view comes from the need to solve real problems. This section contains a non-exhaustive list of some of the applications considered during my research. The abstracts of the contributions grouped by theme have been included in this chapter.

6.1 Image processing problems:

1. **Objects co-segmentation: Propagated from simpler images**, M. Chen, S. Velasco-Forero, I. Tsang and T.J. Cham, IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 1682-1686, 2015.
   Recent works on image co-segmentation aim to segment common objects among image sets. These methods can co-segment simple images well, but their performance may degrade significantly on more cluttered images. In order to co-segment both simple and complex images well, this paper proposes a novel paradigm to rank images and to propagate the segmentation results from the simple images to more and more complex ones. In the experiments, the proposed paradigm demonstrates its effectiveness in segmenting large image sets with a wide variety in object appearance, sizes, orientations, poses, and multiple objects in one image. It outperforms the current state-of-the-art algorithms significantly, especially in difficult images.
2. **A graph-based colour lines model for image analysis**, D. Duque-Arias, S. Velasco-Forero, J.-E. Deschaud, F. Goulette and B. Marcotegui, International Conference on Image Analysis and Processing, Springer, Cham, pp. 181-191, 2019.
   This paper addresses the problem of obtaining a concise description of spectral representation for colour images. The proposed method is a graph-based formulation of the well-known colour Lines model. It generalises the lines to piecewise lines, being able to fit more complex structures. We illustrate the goodness of the proposed method by measuring the quality

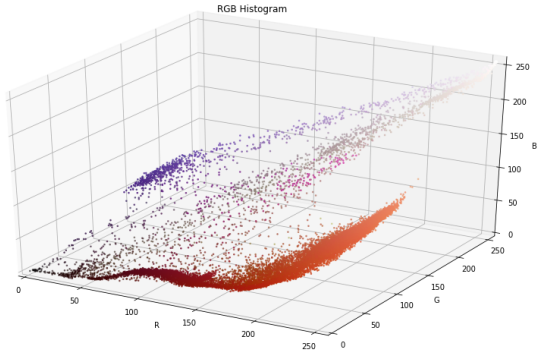
Figure 6.1 Examples of co-segmentation results by method in (54)

of the simplified representations in images and videos. The quality of video sequences reconstructed by means of proposed colour lines extracted from the first frame demonstrates the robustness of our representation. Our formalism allows us to address applications such as image segmentation, shadow correction among others.

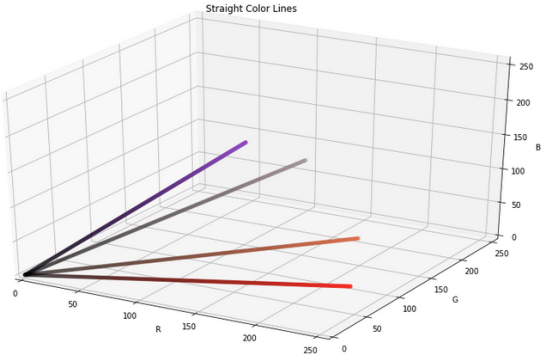
6.1.1 Astronomy

1. **Deep learning for galaxy surface brightness profile fitting**, D. Tuccillo, M. Huertas-Company, E. Decenciére, S. Velasco-Forero, H. Domínguez-Sánchez and P. Dimauro, Monthly Notices of the Royal Astronomical Society, Dec., 2017
   Numerous ongoing and future large area surveys (*e.g.* DES, EUCLID, LSST, WFIRST), will increase by several orders of magnitude the volume of data that can be exploited for galaxy morphology studies. The full potential of these surveys can only be unlocked with the development of automated, fast and reliable analysis methods. In this paper, we present DeepLeGATo, a new method for two-dimensional photometric galaxy profile modelling, based on convolutional neural networks. Our code is trained and validated on analytic profiles (HST/CANDELS F160W filter) and it is able to retrieve the full set of parameters of one-component Sérsic models: total magnitude, effective radius, Sérsic index, axis ratio. We show detailed comparisons between our code and GALFIT. On simulated data, our method is more accurate than GALFIT and 3000 times faster on GPU (50 times when running on the same CPU). On real data, DeepLeGATo trained on simulations behaves similarly to GALFIT on isolated galaxies. With a fast domain adaptation step made with the $0.1-0.8$ per cent the size of the training set, our code is easily capable to reproduce the results obtained with GALFIT even on crowded regions. DeepLeGATo does not require any human intervention beyond the training step, rendering it much more automated than traditional profiling methods. The development of this method for more complex models (two-component galaxies, variable PSF, dense sky regions) could constitute a fundamental tool in the era of big data in astronomy.
2. **The strong gravitational lens finding challenge**, R. Benton Metcalf et al., Astronomy and Astrophysics, volume 625, 22 pages, May. 2019
   Large-scale imaging surveys will increase the number of galaxy-scale strong lensing candidates by maybe three orders of magnitudes beyond the number known today. Finding these rare objects will require picking them out of at least tens of millions of images, and deriving scientific results from them will require quantifying the efficiency and bias of any search

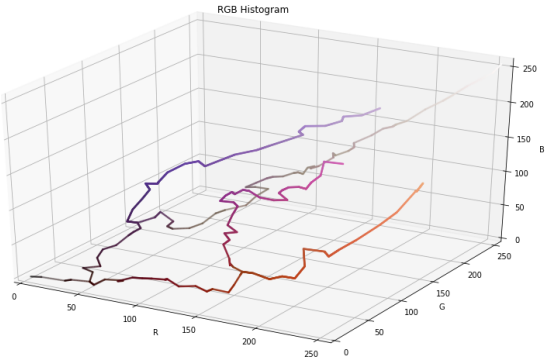
(a) Example image



(b) RGB (Spectral) representation of cropped section of **a)**



(c) Straight colour Lines (PSNR: 21.66 dB)



(d) Geodesic colour Lines (PSNR: 34.96 dB)

Figure 6.2 Comparison of colour Lines / Geodesic based colour lines proposed in (73). This model is another example of modelling spectral information that can be compared in the context of Fig. 1.2

method. To achieve these objectives automated methods must be developed. Because gravitational lenses are rare objects, reducing false positives will be particularly important. We present a description and results of an open gravitational lens finding challenge. Participants were asked to classify 100 000 candidate objects as to whether they were gravitational lenses or not with the goal of developing better automated methods for finding lenses in large data sets. A variety of methods were used including visual inspection, arc and ring finders, support vector machines (SVM) and convolutional neural networks (CNN). We find that many of the methods will be easily fast enough to analyse the anticipated data flow. In test data, several methods are able to identify upwards of half the lenses after applying some thresholds on the lens characteristics such as lensed image brightness, size or contrast with the lens galaxy without making a single false-positive identification. This is significantly better than direct inspection by humans was able to do. Having multi-band, ground based data is found to be better for this purpose than single-band space based data with lower noise and higher resolution, suggesting that multi-colour data is crucial. Multi-band space based data will be superior to ground based data. The most difficult challenge for a lens finder is differentiating between rare, irregular and ring-like face-on galaxies and true gravitational lenses. The degree to which the efficiency and biases of lens

finders can be quantified largely depends on the realism of the simulated data on which the finders are trained.

6.1.2 Counting Models

1. **On-the-go grapevine yield estimation using image analysis and Boolean model**, B. Millan, S. Velasco-Forero, A. Aquino and J. Tardaguila, Journal of Sensors, vol. 2018, 14 pages, Dec 2018.
   This paper describes a new methodology for noninvasive, objective, and automated assessment of yield in vineyards using image analysis and Boolean model. Image analysis, as an inexpensive and noninvasive procedure, has been studied for this purpose, but the effect of occlusions from the cluster or other organs of the vine has an impact that diminishes the quality of the results. To reduce the influence of the occlusions in the estimation, the number of berries was assessed using the Boolean model. To evaluate the methodology, three different datasets were studied: cluster images, manually acquired vine images, and vine images captured on-the-go using a quad. The proposed algorithm estimated the number of berries in cluster images with a root mean square error of 20 and a coefficient of determination (R^2) of 0.80. Vine images manually taken were evaluated, providing 310 grams of mean error and $R^2 = 0.81$. Finally, images captured using a quad equipped with artificial light and automatic camera triggering were also analysed. The estimation obtained applying the Boolean model had 610 grams of mean error per segment (three vines) and $R^2 = .78$. The reliability against occlusions and segmentation errors of the Boolean model makes it ideal for vineyard yield estimation. Its application greatly improved the results when compared to a simpler estimator based on the relationship between cluster area and weight.

6.1.3 Hyperspectral

1. **Classification of hyperspectral images by tensor modeling and additive morphological decomposition**, S. Velasco-Forero and J. Angulo, vol. 46, 2013, Pattern Recognition.
   Pixel-wise classification in high-dimensional multivariate images is investigated. The proposed method deals with the joint use of spectral and spatial information provided in hyperspectral images. Additive morphological decomposition (AMD) based on morphological operators is proposed. AMD defines a scale-space decomposition for multivariate images without any loss of information. AMD is modelled as a tensor structure and tensor principal components analysis is compared as dimensional reduction algorithm versus classic approach. Experimental comparison shows that the proposed algorithm can provide better performance for the pixel classification of hyperspectral image than many other well-known techniques

Figure 6.3 Example of grape segmentation obtained by the method proposed in (182).

6.2 3D Shapes

1. **SHREC'13 Track: Retrieval on textured 3D models**, A. Cerri et al., Eurographics Workshop on 3D Object Retrieval, pp. 73-80, 2013.
   Retrieval on Textured 3D Models, whose goal is to evaluate the performance of retrieval algorithms when models vary either by geometric shape or texture, or both. The collection to search in is made of 240 textured mesh models, divided into 10 classes. Each model has been used in turn as a query against the remaining part of the database. For a given query, the goal was to retrieve the most similar objects. The track saw six participants and the submission of eleven runs.
2. **SHREC-14 Track: Retrieval and classification on Textured 3D Models**, S. Biasotti et al., Eurographics Workshop on 3D Object Retrieval, 2014.
   This paper reports the results of the SHREC'14 track: Retrieval and classification on textured 3D models, whose goal is to evaluate the performance of retrieval algorithms when models vary either by geometric shape or texture, or both. The collection to search in is made of 572 textured mesh models, having a two-level classification based on geometry and texture. Together with the dataset, a training set of 96 models was provided. The track saw eight participants and the submission of 22 runs, to either the retrieval or the classification contest, or both. The evaluation results show a promising scenario about textured 3D retrieval methods, and reveal interesting insights in dealing with texture information in the CIELab rather than in the RGB colour space
3. **Retrieval and classification methods for textured 3D models: A comparative study**, S.Biasotti, M. Aono, A. Ben Hamza, V. Garro, A. Giachetti, D. Giorgi, A. Godil,

C. Li, C. Sanada, M. Spagnuolo, A. Tatsuma and S. Velasco-Forero, 1–25, August, 2015, The Visual Computer Journal.

This paper presents a comparative study of six methods for the retrieval and classification of textured 3D models, which have been selected as representative of the state of the art. To better analyse and control how methods deal with specific classes of geometric and texture deformations, we built a collection of 572 synthetic textured mesh models, in which each class includes multiple texture and geometric modifications of a small set of null models. Results show a challenging, yet lively, scenario and also reveal interesting insights in how to deal with texture information according to different approaches, possibly working in the CIELab as well as in modifications of the RGB colour space.

4. **SHREC'16 Retrieval of Human Subjects from Depth Sensor Data**, A. Giachetti et al., Eurographics Workshop on 3D Object Retrieval, 2016.

   In this paper we report the results of the SHREC 2016 contest on "Retrieval of human subjects from depth sensor data". The proposed task was created in order to verify the possibility of retrieving models of query human subjects from single shots of depth sensors, using shape information only. Depth acquisition of different subjects were realised under different illumination conditions, using different clothes and in three different poses. The resulting point clouds of the partial body shape acquisitions were segmented and coupled with the skeleton provided by the OpenNI software and provided to the participants together with derived triangulated meshes. No colour information was provided. Retrieval scores of the different methods proposed were estimated on the submitted dissimilarity matrices and the influence of the different acquisition conditions on the algorithms were also analysed. Results obtained by the participants and by the baseline methods demonstrated that the proposed task is, as expected, quite difficult, especially due the partiality of the shape information and the poor accuracy of the estimated skeleton, but give useful insights on potential strategies that can be applied in similar retrieval procedures and derived practical applications

5. **SHREC'17 Track: Retrieval of surfaces with similar relief patterns**, S Biasotti, et al., 10th Eurographics Workshop on 3D Object Retrieval, 2017.

   This paper presents the results of the SHREC'17 contest on retrieval of surfaces with similar relief patterns. The proposed task was created in order to verify the possibility of retrieving surface patches with a relief pattern similar to an example from a database of small surface elements. This task, related to many real world applications, requires an effective characterisation of local "texture" information not depending on patch size and bending. Retrieval performances of the proposed methods reveal that the problem is not quite easy to solve and, even if some of the proposed methods demonstrate promising results, further research is surely needed to find effective relief pattern characterisation techniques for practical applications.

6. **SHREC'18 track: Recognition of geometric patterns over 3D models**, S. Biasotti, et al., 11th Eurographics Workshop on 3D Object Retrieval, 2018.

   This track of the SHREC 2018 originally aimed at recognizing relief patterns over a set

of triangle meshes from laser scan acquisitions of archaeological fragments. This track approaches a lively and very challenging problem that remains open after the end of the track. In this report we discuss the challenges to face to successfully address geometric pattern recognition over surfaces; how the existing techniques can go further in this direction, what is currently missing and what is necessary to be further developed.

7. **SHREC'18 track: Retrieval of gray patterns depicted on 3D models**, E. Moscoso Thompson et al, 11th Eurographics Workshop on 3D Object Retrieval, 2018.
   This paper presents the results of the SHREC'18 track: Retrieval of grey patterns depicted on 3D models. The task proposed in the contest challenges the possibility of retrieving surfaces with the same texture pattern of a given query model. This task, which can be seen as a simplified version of many real world applications, requires a characterisation of the surfaces based on local features, rather than considering the surface size and/or bending. All runs submitted to this track are based on feature vectors. The retrieval performances of the runs submitted for evaluation reveal that texture pattern retrieval is a challenging issue. Indeed, a good balance between the size of the pattern and the dimension of the region around a vertex used to locally analyse the colour evolution is crucial for pattern description.
8. **SHREC'20 track: Retrieval of digital surfaces with similar geometric reliefs**, E. Moscoso Thompson et al., Computers and Graphics, pp. 199-218, 2020
   This paper presents the methods that have participated in the SHREC'20 contest on retrieval of surface patches with similar geometric reliefs and the analysis of their performance over the benchmark created for this challenge. The goal of the context is to verify the possibility of retrieving 3D models only based on the reliefs that are present on their surface and to compare methods that are suitable for this task. This problem is related to many real world applications, such as the classification of cultural heritage goods or the analysis of different materials. To address this challenge, it is necessary to characterise the local "geometric pattern" information, possibly forgetting model size and bending. Seven groups participated in this contest and twenty runs were submitted for evaluation. The performances of the methods reveal that good results are achieved with a number of techniques that use different approaches.

6.3 3D Point Cloud/LIDAR

1. **SHREC'17 Track: Point-Cloud Shape Retrieval of Non-Rigid Toys**, FA Limberger, et al., 10th Eurographics Workshop on 3D Object Retrieval, 2017.
   In this paper, we present the results of the SHREC'17 Track: Point-Cloud Shape Retrieval of Non-Rigid Toys. The aim of this track is to create a fair benchmark to evaluate the performance of methods on the non-rigid point-cloud shape retrieval problem. The database used in this task contains 100 3D point-cloud models which are classified into 10 different categories. All point clouds were generated by scanning each one of the models in their final poses using a 3D scanner, *i.e.*, all models have been articulated before scanned. The retrieval performance is evaluated using seven commonly-used statistics (PR-plot, NN, FT,

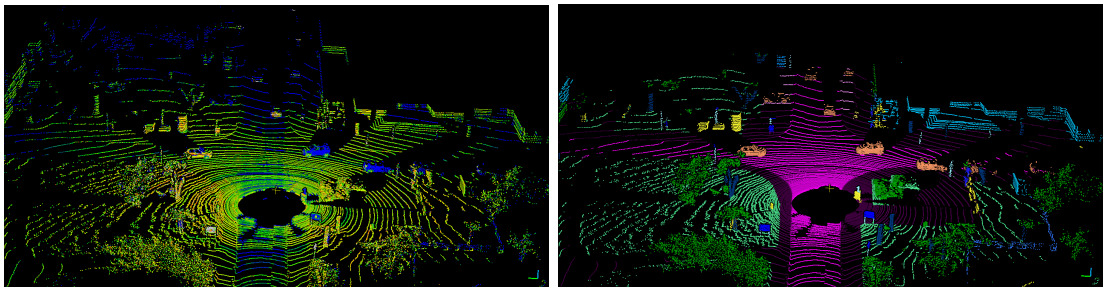


Figure 6.4 Example of a point of view of data from a Light Detection and Ranging scanner with it respective ground truth.(142).

ST, E-measure, DCG, mAP). In total, there are eight groups and 31 submissions taking part in this contest. The evaluation results shown by this work suggest that researchers are in the right direction towards shape descriptors which can capture the main characteristics of 3D models, however, more tests still need to be made, since this is the first time we compare non-rigid signatures for point-cloud shape retrieval.

2. **SHREC 2020 Track: 3D Point Cloud Semantic Segmentation for Street Scenes**, T. Ku et al, Computer and Graphics, pp. 13-24, 2020
   Scene understanding of large-scale 3D point clouds of an outer space is still a challenging task. Compared with simulated 3D point clouds, the raw data from LiDAR scanners consist of tremendous points returned from all possible reflective objects and they are usually non-uniformly distributed. Therefore, its cost-effective to develop a solution for learning from raw large-scale 3D point clouds. In this track, we provide large-scale 3D point clouds of street scenes for the semantic segmentation task. The data set consists of 80 samples with 60 for training and 20 for testing. Each sample with over two million points represents a street scene and includes a couple of objects. There are five meaningful classes: building, car, ground, pole and vegetation. We aim at localising and segmenting semantic objects from these large-scale 3D point clouds. Four groups contributed their results with different methods. The results show that learning-based methods are the trend and one of them achieves the best performance on both Overall Accuracy and mean Intersection over Union. Next to the learning-based methods, the combination of hand-crafted detectors are also reliable and rank second among comparison algorithms.

3. **Road segmentation on low resolution LIDAR point clouds for autonomous vehicles**, L. Gigli et al, ISPRS2020
   Point cloud datasets for perception tasks in the context of autonomous driving often rely on high resolution 64-layer Light Detection and Ranging scanners. They are expensive to deploy on real-world autonomous driving sensor architectures which usually employ 16/32 layer LIDARs. We evaluate the effect of subsampling image based representations of dense point clouds on the accuracy of the road segmentation task. In our experiments the low resolution 16/32 layer LIDAR point clouds are simulated by subsampling the original 64 layer data, for subsequent transformation into a feature map in the Bird-Eye-View (BEV) and SphericalView (SV) representations of the point cloud. We introduce the usage of the

local normal vector with the LIDAR's spherical coordinates as an input channel to existing LoDNN architectures. We demonstrate that this local normal feature in conjunction with classical features not only improves performance for binary road segmentation on full resolution point clouds, but it also reduces the negative impact on the accuracy when subsampling dense point clouds as compared to the usage of classical features alone. We assess our method with several experiments on two datasets: KITTI Road-segmentation benchmark and the recently released Semantic KITTI dataset.

4. **Dartboard based ground detection on 3D point cloud**, L. Gigli et al, ISPRS2022
   3D laser scanners acquire 3D point clouds of real environments. The process consists in sampling the scene with laser beams rotating around an axis. By construction, the point density decreases with the distance to the scanner. This density heterogeneity is a major issue, in particular for mobile systems in the context of autonomous driving, as usually a single scan is processed simultaneously (instead of mapping applications that can integrate several scans, reducing the density heterogeneity). We propose a dartboard grid with cell size increasing radially in order to adapt the grid size to the point density. The effectiveness of this strategy is demonstrated by means of a ground detection task, a fundamental step in many workflows of analysis of 3D point clouds.

5. **Paris-CARLA-3D: A real and synthetic outdoor point cloud dataset for challenging tasks in 3D mapping**, J.E. Deschaud et al., Remote Sensing, 2021
   Paris-CARLA-3D is a dataset of several dense colored point clouds of outdoor environments built by a mobile LiDAR and camera system. The data are composed of two sets with synthetic data from the open source CARLA simulator (700 million points) and real data acquired in the city of Paris (60 million points), hence the name Paris-CARLA-3D. One of the advantages of this dataset is to have simulated the same LiDAR and camera platform in the open source CARLA simulator as the one used to produce the real data. In addition, manual annotation of the classes using the semantic tags of CARLA was performed on the real data, allowing the testing of transfer methods from the synthetic to the real data. The objective of this dataset is to provide a challenging dataset to evaluate and improve methods on difficult vision tasks for the 3D mapping of outdoor environments: semantic segmentation, instance segmentation, and scene completion. For each task, we describe the evaluation protocol as well as the experiments carried out to establish a baseline.

References

[1] Y. I. Abramovich and O. Besson, *Regularized Covariance Matrix Estimation in Complex Elliptically Symmetric Distributions Using the Expected Likelihood Approach- Part 1: The Over-Sampled Case*, IEEE Transactions on Signal Processing, 61 (2013), pp. 5807–5818.

[2] Y. I. Abramovich and N. K. Spencer, *Diagonally loaded normalised sample matrix inversion (lnsmi) for outlier-resistant adaptive filtering*, in ICASSP 2007, vol. 3, IEEE, 2007, pp. 1111–1105.

[3] A. Adadi and M. Berrada, *Peeking inside the black-box: a survey on explainable artificial intelligence (xai)*, IEEE access, 6 (2018), pp. 52138–52160.

[4] L. Alvarez and L. Mazorra, *Signal and image restoration using shock filters and anisotropic diffusion*, SIAM Journal of Numerical Analysis, 31 (1994), pp. 590–605.

[5] J. An and S. Cho, *Variational autoencoder based anomaly detection using reconstruction probability*, Special Lecture on IE, 2 (2015), pp. 1–18.

[6] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, no. 118, Cambridge University Press, 2010.

[7] T. W. Anderson, *An introduction to multivariate statistical analysis*, Wiley, 1958.

[8] J. Angulo, *Morphological colour operators in totally ordered lattices based on distances: Application to image filtering, enhancement and analysis*, Computer Vision and Image Understanding, 107 (2007), pp. 56–73.

[9] ——, *Morphological bilateral filtering and spatially-variant adaptive structuring functions*, in ISMM'11, vol. 6671 of LNCS, Springer, 2011, pp. 212–223.

[10] ——, *Lipschitz Regularization of Images supported on Surfaces using Riemannian Morphological Operators*. Preprint, Nov. 2014.

[11] J. Angulo and D. Jeulin, *Stochastic watershed segmentation.*, in ISMM (1), 2007, pp. 265–276.

[12] J. Angulo and S. Velasco-Forero, *Structurally adaptive math. morph. based on nonlinear scale-space decomp.*, Image Analysis & Stereology, 30 (2011), pp. 111–122.

[13] ——, *Riemannian mathematical morphology*, Pattern Recognition Letters, 47 (2014), pp. 93–101.

[14] E. Aptoula and S.Lefèvre, *A comparative study on multivariate mathematical morphology*, Pattern Recognition, 40 (2007), pp. 2914–2929.

[15] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, *Contour detection and hierarchical image segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33 (2011), pp. 898–916.

[16] M. Bai and R. Urtasun, *Deep watershed transform for instance segmentation*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 5221–5229.

[17] G. J. F. Banon and J. Barrera, *Minimal representations for translation-invariant set mappings by mathematical morphology*, SIAM Journal on Applied Mathematics, 51 (1991), pp. 1782–1798.

[18] A. Baraldi and F. Parmiggiani, *Single linkage region growing algorithms based on the vector degree of match*, IEEE Transactions on Geoscience and Remote Sensing, 34 (1996), pp. 137–148.

[19] V. Barnett, *The ordering of multivariate data (with discussion)*, Journal of the Royal Statistical Society Series A, 139 (1976), pp. 318–354.

[20] P. L. Bartlett and M. H. Wegkamp, *Classification with a reject option using a hinge loss.*, Journal of Machine Learning Research, 9 (2008).

[21] M. Bauw, S. Velasco-Forero, J. Angulo, C. Adnet, and O. Airiau, *Deep random projection outlyingness for unsupervised anomaly detection*, arXiv preprint arXiv:2106.15307, (2021).

[22] ——, *Near out-of-distribution detection for low-resolution radar micro-doppler signatures*, arXiv preprint arXiv:2205.07869, (2022).

[23] G. Bécigneul and O.-E. Ganea, *Riemannian adaptive optimization methods*, arXiv preprint arXiv:1810.00760, (2018).

[24] J. Bedi and D. Toshniwal, *Empirical mode decomposition based deep learning for electricity demand forecasting*, Ieee Access, 6 (2018), pp. 49144–49156.

[25] A. Ben-David and C. E. Davidson, *Eigenvalue estimation of hyperspectral wishart covariance matrices from limited number of samples*, IEEE Transactions on Geoscience and Remote Sensing, 50 (2012), pp. 4384–4396.

[26] A. Ben-David and J. Marks, *Geodesic paths for time-dependent covariance matrices in a Riemannian manifold*, IEEE Transactions on Geoscience and Remote Sensing Letters, 11 (2014), pp. 1499–1503.

[27] S. Ben-David and M. Lindenbaum, *Learning distributions by their density levels: A paradigm for learning without a teacher*, Journal of Computer and System Sciences, 55 (1997), pp. 171–182.

[28] K. P. Bennett and E. J. Bredensteiner, *Duality and geometry in svm classifiers*, in In Proc. 17th International Conf. on Machine Learning, Morgan Kaufmann, 2000, pp. 57–64.

[29] G. Benton, M. Finzi, P. Izmailov, and A. G. Wilson, *Learning invariances in neural networks*, arXiv preprint arXiv:2010.11882, (2020).

[30] O. Besson and Y. I. Abramovich, *Regularized Covariance Matrix Estimation in Complex Elliptically Symmetric Distributions Using the Expected Likelihood Approach– Part 2: The Under-Sampled Case*, IEEE Transactions on Signal Processing, 61 (2013), pp. 5819–5829.

[31] S. Beucher, *Segmentation d'images et morphologie mathématique*, PhD thesis, Ecole Nationale Supérieure des Mines de Paris, 1990.

[32] S. Beucher, J. M. Blosseville, and F. Lenoir, *Traffic spatial measurements using video image processing*, in Proc. Intelligent Robots and Computer Vision, SPIE, 1988.

[33] S. Blusseau, B. Pochon, S. Velasco-Forero, J. Angulo, and I. Bloch, *Approximating morphological operators with part-based representations learned by asymmetric auto-encoders*, Mathematical Morphology-Theory and Applications, (2020).

[34] S. Blusseau, S. Velasco-Forero, J. Angulo, and I. Bloch, *Adaptive anisotropic morphological filtering based on co-circularity of local orientations*, Image Processing On Line, 12 (2022), pp. 111–141.

[35] ——, *Morphological adjunctions represented by matrices in max-plus algebra for signal and image processing*, in International Conference on Discrete Geometry and Mathematical Morphology, Springer, 2022, pp. 206–218.

[36] S. Bohlender, I. Oksuz, and A. Mukhopadhyay, *A survey on shape-constraint deep learning for medical image segmentation*, arXiv preprint arXiv:2101.07721, (2021).

[37] C. Boncelet, *Image noise models*, in Handbook of image and video processing, A. C. Bovik, ed., New York: Academic, 2000, pp. 325–335.

[38] D. Borghys, V. Kåsen, I.and Achard, and C. Perneel, *Hyperspectral anomaly detection: comparative evaluation in scenes with diverse complexity*, Journal of Electrical and Computer Engineering, 2012 (2012), p. 5.

[39] J. Bosworth and S. T. Acton, *The morphological lomo filter for multiscale image processing*, in Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348), vol. 4, IEEE, 1999, pp. 157–161.

[40] Y.-L. Boureau, J. Ponce, and Y. LeCun, *A theoretical analysis of feature pooling in visual recognition*, in ICML 2010, 2010, pp. 111–118.

[41] N. Boutry, T. Géraud, and L. Najman, *An equivalence relation between morphological dynamics and persistent homology in n-d*, in International Conference on Discrete Geometry and Mathematical Morphology, Springer, 2021, pp. 525–537.

[42] Y. Boykov and G. Funka-Lea, *Graph cuts and efficient nd image segmentation*, International journal of computer vision, 70 (2006), pp. 109–131.

[43] J. Bruna and S. Mallat, *Invariant scattering convolution networks*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35 (2013), pp. 1872–1886.

[44] B. Burgeth and A. Kleefeld, *An approach to color-morphology based on Einstein addition and Loewner order*, Pattern Recognition Letters, 47 (2014), pp. 29–39.

[45] G. Cao, L. R. Bachega, and C. A. Bouman, *The Sparse Matrix Transform for Covariance Estimation and Analysis of High Dimensional Signals*, IEEE Transactions on Image Processing, 20 (2011), pp. 625–640.

[46] G. Cao and C. A. Bouman, *Covariance estimation for high dimensional data vectors using the sparse matrix transform*, Advances in Neural Information Processing Systems, 21 (2009), pp. 225–232.

[47] M. Carlsson, *On convex envelopes and regularization of non-convex functionals without moving global minima*, Journal of Optimization Theory and Applications, 183 (2019), pp. 66–84.

[48] V. Caselles and P. Monasse, *Geometric description of images as topographic maps*, Springer, 2009.

[49] I. Chami, A. Gu, V. Chatziafratis, and C. Ré, *From trees to continuous embeddings and back: Hyperbolic hierarchical clustering*, Advances in Neural Information Processing Systems, 33 (2020).

[50] C.-I. Chang and S.-S. Chiang, *Anomaly detection and classification for hyperspectral imagery*, IEEE Transactions on Geoscience and Remote Sensing, 40 (2002), pp. 1314–1325.

[51] V. Charisopoulos and P. Maragos, *Morphological perceptrons: geometry and training algorithms*, in International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing, 2017, pp. 3–15.

[52] J. Chen, X. Wang, and C. Planiden, *A proximal average for prox-bounded functions*, SIAM Journal on Optimization, 30 (2020), pp. 1366–1390.

[53] M. Chen, S. Pang, T. Cham, and A. Goh, *Visual tracking with generative template model based on Riemannian manifold of covariances*, in Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on, IEEE, 2011, pp. 1–8.

[54] W. Chen, J. Wilson, S. Tyree, K. Weinberger, and Y. Chen, *Compressing neural networks with the hashing trick*, in International Conference on Machine Learning, 2015, pp. 2285–2294.

[55] Y. Chen, A. Wiesel, and A. O. Hero, *Shrinkage estimation of high dimensional covariance matrices*, in International Conference on Acoustics, Speech and Signal Processing, IEEE, 2009, pp. 2937–2940.

[56] ——, *Robust shrinkage estimation of high-dimensional covariance matrices*, Signal Processing, IEEE Transactions on, 59 (2011), pp. 4097–4107.

[57] E. Chevallier and J. Angulo, *The Irregularity Issue of Total Orders on Metric Spaces and Its Consequences for Mathematical Morphology*, Journal of Mathematical Imaging and Vision, 54 (2016), pp. 344–357.

[58] E. C. Chi and K. Lange, *Stable estimation of a covariance matrix guided by nuclear norm penalties*, Computational statistics & data analysis, 80 (2014), pp. 117–128.

[59] G. Chierchia and B. Perret, *Ultrametric fitting by gradient descent*, in Advances in neural information processing systems, vol. 32, 2019, pp. 3181–3192.

[60] Y.-M. Chung, S. Day, and C.-S. Hu, *A multi-parameter persistence framework for mathematical morphology*, Scientific reports, 12 (2022), pp. 1–25.

[61] J. R. Clough, N. Byrne, I. Oksuz, V. A. Zimmer, J. A. Schnabel, and A. P. King, *A topological loss function for deep-learning based image segmentation using persistent homology*, arXiv preprint arXiv:1910.01877, (2019).

[62] J. Cousty, G. Bertrand, L. Najman, and M. Couprie, *Watershed cuts: Minimum spanning forests and the drop of water principle*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31 (2009), pp. 1362–1374.

[63] J. Cousty, L. Najman, Y. Kenmochi, and S. Guimarães, *Hierarchical segmentations with graphs: quasi-flat zones, minimum spanning trees, and saliency maps*, Journal of Mathematical Imaging and Vision, 60 (2018), pp. 479–502.

[64] N. Cristianini and J. Shawe-Taylor, *An Introduction to support vector machines and other kernel based learning methods*, Cambridge University Press, 2000.

[65] O. Cuisenaire, *Locally adaptable mathematical morphology using distance transformations*, Pattern Recognition, 39 (2006), pp. 405–416.

[66] C. E. Davidson and A. Ben-David, *On the use of covariance and correlation matrices in hyperspectral detection*, in Applied Imagery Pattern Recognition Workshop (AIPR), IEEE, 2011, pp. 1–6.

[67] J. L. Davidson and G. X. Ritter, *Theory of morphological neural networks*, in Digital Optical Computing II, vol. 1215, 1990, pp. 378–389.

[68] J. Debayle and J.-C. Pinoli, *Spatially adaptive morphological image filtering using intrinsic structuring elements*, Image Analysis and Stereology, 39 (2005), pp. 145–158.

[69] E. Decencière, S. Velasco-Forero, F. Min, J. Chen, H. Burdin, G. Gauthier, B. Laÿ, T. Bornschloegl, and T. Baldeweck, *Dealing with topological information within a fully convolutional neural network*, in International Conference on Advanced Concepts for Intelligent Vision Systems, Springer, 2018, pp. 462–471.

[70] E. H. S. Diop, R. Alexandre, and L. Moisan, *Intrinsic nonlinear multiscale image decomposition: A 2D empirical mode decomposition-like tool*, Computer Vision and Image Understanding, 116 (2012), pp. 102–119. Virtual Representations and Modeling of Large-scale Environments (VRML).

[71] E.-H. S. Diop, R. Alexandre, and V. Perrier, *A PDE model for 2D intrinsic mode functions*, IEEE ICIP, (2009).

[72] D. Donoho and M. Gasko, *Breakdown properties of location estimates based on halfspace depth and projected outlyingness*, The Annals of Statistics, 20 (1992), pp. 1803–1827.

[73] D. Duque-Arias, S. Velasco-Forero, J.-E. Deschaud, F. Goulette, and B. Marcotegui, *A graph-based color lines model for image analysis*, in International Conference on Image Analysis and Processing, Springer, 2019, pp. 181–191.

[74] R. Durbin and D. E. Rumelhart, *Product units: A computationally powerful and biologically plausible extension to backpropagation networks*, Neural computation, 1 (1989), pp. 133–142.

[75] A. Edelman and N. R. Rao, *Random matrix theory*, Acta Numerica, 14 (2005), pp. 233–297.

[76] B. Efron and C. Morris, *Data analysis using Steinś estimator and its generalizations*, Journal of the American Statistical Association, 70 (1975), pp. 311–319.

[77] ——, *Multivariate empirical bayes and estimation of covariance matrices*, The Annals of Statistics, (1976), pp. 22–32.

[78] K.-T. Fang, S. Kotz, and K. W. Ng, *Symmetric multivariate and related distributions*, no. 36 in Monographs on statistics and applied probability, Chapman Hall, London, 1990.

[79] A. Fehri, S. Velasco-Forero, and F. Meyer, *Automatic selection of stochastic watershed hierarchies*, in 24th European Signal Processing Conference (EUSIPCO), IEEE, 2016, pp. 1877–1881.

[80] ——, *Prior-based hierarchical segmentation highlighting structures of interest*, in International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing, Springer, 2017, pp. 146–158.

[81] ——, *Combinatorial space of watershed hierarchies for image characterization*, Pattern Recognition Letters, 129 (2020), pp. 41–47.

[82] G. Frahm, *Generalized elliptical distributions: theory and applications*, PhD thesis, Universität zu Köln, 2004.

[83] G. Franchi, A. Fehri, and A. Yao, *Deep morphological networks*, Pattern Recognition, 102 (2020), p. 107246.

[84] J. Frankle, D. J. Schwab, and A. S. Morcos, *Training batchnorm and only batchnorm: On the expressive power of random features in CNNs*, ICLR, (2021).

[85] J. Frontera-Pons, M. Mahot, J. Ovarlez, and F. Pascal, *Robust Detection using M- estimators for Hyperspectral Imaging*, in IEEE Workshop on Hyperspectral Image and Signal Processing, 2012.

[86] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel, *Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness*, arXiv preprint arXiv:1811.12231, (2018).

[87] Z. Ghafoori and C. Leckie, *Deep multi-sphere support vector data description*, in Proceedings of the 2020 SIAM International Conference on Data Mining, SIAM, 2020, pp. 109–117.

[88] L. Gigli, B. Marcotegui, and S. Velasco-Forero, *End-to-End similarity learning and hierarchical clustering for unfixed size datasets*, in International Conference on Geometric Science of Information, Springer, 2021, pp. 596–604.

[89] L. Gigli, S. Velasco-Forero, and B. Marcotegui, *On minimum spanning tree streaming for hierarchical segmentation*, Pattern Recognition Letters, 138 (2020), pp. 155–162.

[90] G. Gilboa, N. A. Sochen, and Y. Y. Zeevi, *Regularized shock filters and complex diffusion*, in Proceeding of the 7th European Conference on Computer Vision-Part I, ECCV '02, London, UK, 2002, Springer-Verlag, pp. 399–413.

[91] N. Gillis, *Nonnegative matrix factorization*, SIAM, 2020.

[92] X. Glorot and Y. Bengio, *Understanding the difficulty of training deep feedforward neural networks*, in Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010, pp. 249–256.

[93] X. Glorot, A. Bordes, and Y. Bengio, *Deep sparse rectifier neural networks*, in 14th International Conference on Art. Intell. and Statistics, 2011, pp. 315–323.

[94] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, MIT press, 2016.

[95] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, *Maxout networks*, arXiv preprint arXiv:1302.4389, (2013).

[96] J. Goutsias and H. Heijmans, *Mathematical Morphology*, IOS Press, 2000.

[97] J. Goutsias, H. Heijmans, and K.Sivakumar, *Morphological operators for image sequences*, Comput. Vis. Image Underst., 62 (1995), pp. 326–346.

[98] J. Gower and G. Ross, *Minimum spanning trees and single linkage cluster analysis*, Applied statistics, 18 (1969), pp. 54–64.

[99] J. Grazzini and P. Soille, *Edge preserving smoothing using a similarity measure in adaptive geodesic neighbourhoods*, Pattern Recognition, 42 (2009), pp. 2036–2316.

[100] M. Grimaud, *New measure of contrast: the dynamics*, in Proceeding of SPIE, International Society for Optics and Photonics, 1992, pp. 292–305.

[101] L. Gueguen and P. Soille, *Frequent and dependent connectivities*, in Proc. of 10th Int. Symp. on Mathematical Morphology, vol. 6671 of Lecture Notes in Computer Science, Springer-Verlag, 2011, pp. 120–131.

[102] F. Guichard, P. Maragos, and J.-M. Morel, *Partial differential equations for morphological operators*, in Space, Structure and Randomness, Springer, 2005, pp. 369–390.

[103] L. Guigues, J. P. Cocquerez, and H. Le Men, *Scale-sets image analysis*, International Journal of Computer Vision, 68 (2006), pp. 289–317.

[104] H. K. Hartline and F. Ratliff, *Inhibitory interaction of receptor units in the eye of limulus*, The Journal of General Physiology, 40 (1957), pp. 357–376.

[105] K. He, X. Zhang, S. Ren, and J. Sun, *Delving deep into rectifiers: Surpassing human-level performance on imagenet classification*, in IEEE ICCV, 2015, pp. 1026–1034.

[106] H. Heijmans and J. Goutsias, *Nonlinear multiresolution signal decomposition schemes. II. morphological wavelets*, IEEE Transactions on Image Processing, 9 (2000), pp. 1897–1913.

[107] H. Heijmans and P. Maragos, *Lattice calculus of the morphological slope transform*, Signal Processing, 59 (1997), pp. 17–42.

[108] H. Heijmans and C. Ronse, *The algebraic basis of mathematical morphology I. dilations and erosions*, Computer Vision, Graphics, and Image Processing, 50 (1990), pp. 245–295.

[109] H. Heijmans and R. van den Boomgaard, *Algebraic framework for linear and morphological scale-spaces*, Journal of Visual Communication and Image Representation, 13 (2002), pp. 269–301.

[110] H. J. A. M. Heijmans, *Morphological Image Operators*, vol. 10, Academic Press, 1994.

[111] ——, *Composing morphological filters*, IEEE Transactions on Image Processing, 6 (1997), pp. 713–724.

[112] H. J. A. M. Heijmans, P. Nacken, A. Toet, and L. Vincent, *Graph morphology*, Journal of Visual Communication and Image Representation, 3 (1992), pp. 24–38.

[113] D. Hendrycks, M. Mazeika, and T. Dietterich, *Deep anomaly detection with outlier exposure*, Proceedings of the International Conference on Learning Representations, (2019).

[114] K. Hermann, T. Chen, and S. Kornblith, *The origins and prevalence of texture bias in convolutional neural networks*, Advances in Neural Information Processing Systems, 33 (2020), pp. 19000–19015.

[115] R. Hermary, G. Tochon, É. Puybareau, A. Kirszenberg, and J. Angulo, *Learning grayscale mathematical morphology with smooth morphological layers*, Journal of Mathematical Imaging and Vision, (2022).

[116] G. Hernández, E. Zamora, H. Sossa, G. Téllez, and F. Furlán, *Hybrid neural networks for big data classification*, Neurocomputing, 390 (2020), pp. 327–340.

[117] J. P. Hoffbeck and D. A. Landgrebe, *Covariance matrix estimation and classification with limited training data*, IEEE Transactions on Image Processing, 18 (1996), pp. 763–767.

[118] K. Hornik, M. Stinchcombe, and H. White, *Multilayer feedforward networks are universal approximators*, Neural networks, 2 (1989), pp. 359–366.

[119] P. O. Hoyer, *Non-negative matrix factorization with sparseness constraints*, CoRR, cs.LG/0408058 (2004).

[120] X. Hu, F. Li, D. Samaras, and C. Chen, *Topology-preserving deep image segmentation*, Advances in Neural Information Processing Systems, 32 (2019).

[121] N. Huang, S. Zheng, S. Long, M. Wu, H. Shih, Q. Zheng, N.-C. Yen, C. Tung, and H. Liu, *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*, The Royal Society, 454 (1998), pp. 903–995.

[122] H. Hwang and R. A. Haddad, *Adaptive median filters: new algorithms and results*, IEEE Transactions on Image Processing, 4 (1995), pp. 499–502.

[123] M. A. Islam, B. Murray, A. Buck, D. T. Anderson, G. J. Scott, M. Popescu, and J. Keller, *Extending the morphological hit-or-miss transform to deep neural networks*, IEEE Transactions on NNs and Learning Systems, 32 (2020), pp. 4826–4838.

[124] P. T. Jackway, *Morphological scale-spaces*, in Advances in Imaging and Electron Physics, P. W. Hawkes, ed., vol. 99, Elsevier, 1997, pp. 1–64.

[125] P. T. Jackway and M. Deriche, *Scale-space properties of the multiscale morphological dilation-erosion*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 18 (1996), pp. 38–51.

[126] D. Jeulin, *Dead leaves models: From space tessellation to random functions*, in Advances in Theory and Applications of Random Sets, World Scientific Publishing, 1997, pp. 137–156.

[127] Z. Ji, Y. Shen, C. Ma, and M. Gao, *Scribble-based hierarchical weakly supervised learning for brain tumor segmentation*, in International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 175–183.

[128] Z. Jia, X. Huang, I. Eric, C. Chang, and Y. Xu, *Constrained deep weak supervision for histopathology image segmentation*, IEEE transactions on medical imaging, 36 (2017), pp. 2376–2388.

[129] X.-B. Jin, N.-X. Yang, X.-Y. Wang, Y.-T. Bai, T.-L. Su, and J.-L. Kong, *Deep hybrid model based on emd with classification by frequency characteristics for long-term air quality prediction*, Mathematics, 8 (2020), p. 214.

[130] P.-L. L. J.M. Lasry, *A remark on regularization in Hilbert spaces*, Israel Journal of Mathematics, 55 (1986), pp. 257–266.

[131] T. Jochems, *Morphologie Mathématique Appliquée au Contrôle Industriel de Pièces Coulées*, PhD thesis, Ecole National Supérieure des Mines de Paris, 1997.

[132] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, 1986.

[133] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*, vol. 344, John Wiley & Sons, 2009.

[134] H. Kervadec, J. Dolz, M. Tang, E. Granger, Y. Boykov, and I. B. Ayed, *Constrained-CNN losses for weakly supervised segmentation*, Medical image analysis, 54 (2019), pp. 88–99.

[135] H. Kervadec, J. Dolz, J. Yuan, C. Desrosiers, E. Granger, and I. B. Ayed, *Constrained deep networks: Lagrangian optimization via log-barrier extensions*, arXiv preprint arXiv:1904.04205, (2019).

[136] R. Keshet and H. J. A. M. Heijmans, *Adjunctions in pyramids, curve evolution and scale-spaces*, Int. J. Comput. Vision, 52 (2003), pp. 139–151.

[137] B. R. Kiran and J. Serra, *Global–local optimizations by hierarchical cuts and climbing energies*, Pattern Recognition, 47 (2014), pp. 12–24.

[138] S. Kotz, N. Johnson, and C. Read, *Encyclopedia of Statistical Sciences: Multivariate analysis to plackett and burman designs*, Encyclopedia of Statistical Sciences, Wiley, 1982.

[139] H. P. Kramer and J. B. Bruckner, *Iterations of a non-linear transformation for enhancement of digital images*, Pattern Recognition, 7 (1975), pp. 53–58.

[140] F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz, *The spectral image processing system (SIPS)—Interactive visualization and analysis of imaging spectrometer data*, Remote Sensing of Environment, 44 (1993), pp. 145–163.

[141] J. Kruskal, *On the shortest spanning subtree of a graph and the traveling salesman problem*, Proc. Am. Math. Soc., 7 (1956), pp. 48–50.

[142] T. Ku, R. C. Veltkamp, B. Boom, D. Duque-Arias, S. Velasco-Forero, J.-E. Deschaud, F. Goulette, B. Marcotegui, S. Ortega, A. Trujillo, et al., *SHREC 2020: 3D point cloud semantic segmentation for street scenes*, Computers & Graphics, 93 (2020), pp. 13–24.

[143] K. L. Lange, R. J. A. Little, and J. M. G. Taylor, *Robust statistical modeling using the t distribution*, Journal of the American Statistical Association, 84 (1989), pp. 881–896.

[144] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, nature, 521 (2015), pp. 436–444.

[145] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, *Gradient-based learning applied to document recognition*, Proceedings of the IEEE, 86 (1998), pp. 2278–2324.

[146] O. Ledoit and M. Wolf, *Honey, I shrunk the sample covariance matrix*, UPF Economics and Business Working Paper, (2003).

[147] ——, *A well-conditioned estimator for large-dimensional covariance matrices*, Journal of multivariate analysis, 88 (2004), pp. 365–411.

[148] R. Lerallut, E. Decencière, and F. Meyer, *Image filtering using morphological amoebas*, Image and Vision Computing, 4 (2007), pp. 395–404.

[149] O. Lezoray, C. Charrier, and A. Elmoataz, *Learning complete lattices for manifold mathematical morphology*, in Proc. of the ISMM, 2009, pp. 1–4.

[150] C. Liu and D. B. Rubin, *ML estimation of the t distribution using EM and its extensions, ECM and ECME*, Statistica Sinica, 5 (1995), pp. 19–39.

[151] D. Looney and D. P. Mandic, *A machine learning enhanced empirical mode decomposition*, in IEEE ICASSP, IEEE, 2008, pp. 1897–1900.

[152] R. Lukac, B. Smolka, K. Martin, K. Plataniotis, and A. Venetsanopoulos, *Vector filtering for color imaging*, Signal Processing Magazine, IEEE, 22 (2005), pp. 74–86.

[153] D. Lunga and O. Ersoy, *Unsupervised classification of hyperspectral images on spherical manifolds*, in ICDM, 2011, pp. 134–146.

[154] A. L. Maas, A. Y. Hannun, A. Y. Ng, et al., *Rectifier nonlinearities improve neural network acoustic models*, in Proc. ICML, vol. 30, 2013, p. 3.

[155] L. v. d. Maaten and G. Hinton, *Visualizing data using t-SNE*, Journal of machine learning research, 9 (2008), pp. 2579–2605.

[156] P. C. Mahalanobis, *On the generalized distance in statistics*, 1936.

[157] J. Mairal, F. R. Bach, and J. Ponce, *Sparse modeling for image and vision processing*, CoRR, abs/1411.3230 (2014).

[158] S. Mallat, *A wavelet tour of signal processing*, Elsevier, 1999.

[159] ——, *Group invariant scattering*, Communications on Pure and Applied Mathematics, 65 (2012), pp. 1331–1398.

[160] D. Manolakis, D. Marden, and G. A. Shaw, *Hyperspectral image processing for automatic target detection applications*, Lincoln Laboratory Journal, 14 (2003), pp. 79–116.

[161] P. Maragos, *A representation theory for morphological image and signal processing*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11 (1989), pp. 586–599.

[162] ——, *PDEs for morphological scale-spaces and eikonal applications*, The Image and Video Processing Handbook, (2005), pp. 587–612.

[163] ——, *Representations for morphological image operators and analogies with linear operators*, Advances in imaging and electron physics, 177 (2013), pp. 45–187.

[164] P. Maragos, J. F. Kaiser, and T. F. Quatieri, *Energy separation in signal modulations with application to speech analysis*, IEEE Transactions on Signal Processing, 41 (1993), pp. 3024–3051.

[165] P. Maragos and R. Schafer, *Morphological skeleton representation and coding of binary images*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 34 (1986), pp. 1228–1244.

[166] V. A. Marčenko and L. A. Pastur, *Distribution of eigenvalues for some sets of random matrices*, Sbornik: Mathematics, 1 (1967), pp. 457–483.

[167] D. Marin, M. Tang, I. B. Ayed, and Y. Boykov, *Beyond gradient descent for regularized segmentation losses*, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10187–10196.

[168] R. Maronna, *Robust M-estimators of multivariate location and scatter*, The Annals of Statistics, (1976), pp. 51–67.

[169] D. Martin, C. Fowlkes, D. Tal, and J. Malik, *A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics*, in Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, vol. 2, IEEE, 2001, pp. 416–423.

[170] G. Matheron, *Random sets and integral geometry*, Wiley New York, 1975.

[171] S. Matteoli, M. Diani, and G. Corsini, *Improved estimation of local background covariance matrix for anomaly detection in hyperspectral images*, Optical Engineering, 49 (2010), pp. 1–16.

[172] ——, *A tutorial overview of anomaly detection in hyperspectral images*, Aerospace and Electronic Systems Magazine, IEEE, 25 (2010), pp. 5–28.

[173] S. Matteoli, M. Diani, and J. Theiler, *An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7 (2014), pp. 2317–2336.

[174] R. Menon, P. Gerstoft, and W. Hodgkiss, *Asymptotic eigenvalue density of noise covariance matrices*, IEEE Transactions on Signal Processing, 60 (2012), pp. 3415–3424.

[175] F. Meyer, *Minimum spanning forests for morphological segmentation*, in Mathematical morphology and its applications to image processing, Springer, 1994, pp. 77–84.

[176] ——, *The levelings*, in ISMM '98: Proceedings of the fourth international symposium on Mathematical morphology and its applications to image and signal processing, Norwell, MA, USA, 1998, Kluwer Academic Publishers, pp. 199–206.

[177] ——, *Stochastic watershed hierarchies*, in ICAPR, 2015, pp. 1–8.

[178] ——, *The waterfall hierarchy on weighted graphs*, in International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing, Springer, 2015, pp. 325–336.

[179] F. Meyer and P. Maragos, *Morphological scale-space representation with levelings*, Lecture Notes in Computer Science, 1682 (1999), pp. 187–198.

[180] ——, *Nonlinear scale-space representation with morphological levelings*, Journal of Visual Communication and Image Representation, 11 (2000), pp. 245–265.

[181] F. Meyer and J. Serra, *Contrasts and activity lattice*, Signal Processing, 16 (1989), pp. 303–317.

[182] B. Millan, S. Velasco-Forero, A. Aquino, and J. Tardaguila, *On-the-go grapevine yield estimation using image analysis and boolean model*, Journal of Sensors, 2018 (2018).

[183] T. Miller, *Explanation in artificial intelligence: Insights from the social sciences*, Artificial intelligence, 267 (2019), pp. 1–38.

[184] P. Monasse and F. Guichard, *Fast computation of a contrast-invariant image representation*, IEEE Transactions on Image Processing, 9 (2000), pp. 860–872.

[185] R. Mondal, M. S. Dey, and B. Chanda, *Image restoration by learning morphological opening-closing network*, Mathematical Morphology-Theory and Applications, 4 (2020), pp. 87–107.

[186] T. K. Moon, *The expectation-maximization algorithm*, Signal processing magazine, 13 (1996), pp. 47–60.

[187] V. Morard, E. Decencière, and P. Dokládal, *Region growing structuring elements and new operators based on their shape*, in Signal and Image Proc., ACTA Press, 2011.

[188] J. J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bulletin de la Société mathématique de France, 93 (1965), pp. 273–299.

[189] ——, *Inf-convolution, sous-additivité, convexité des fonctions numériques*, Journal de Mathématiques Pures et Appliquées, (1970).

[190] K. Muller, S. Mika, G. Ritsch, K. Tsuda, and B. Scholkopf, *An introduction to kernel-based learning algorithms*, IEEE Transactions on Neural Networks, 12 (2001), pp. 181–201.

[191] R. Nadakuditi and J. W. Silverstein, *Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples*, IEEE Journal of Selected Topics in Signal Processing, 4 (2010), pp. 468–480.

[192] S. Nadarajah and S. Kotz, *Estimation methods for the multivariate t distribution*, Acta Applicandae Mathematicae, 102 (2008), pp. 99–118.

[193] M. Nagao, T. Matsuyama, and Y. Ikeda, *Region extraction and shape analysis in aerial photographs*, Computer Graphics and Image Processing, 10 (1979), pp. 195–223.

[194] V. Nair and G. E. Hinton, *Rectified linear units improve restricted boltzmann machines*, in ICML, 2010.

[195] L. Najman and J. Cousty, *A graph-based mathematical morphology reader*, Pattern Recognition Letters, 47 (2014), pp. 3–17.

[196] L. Najman, J. Cousty, and B. Perret, *Playing with kruskal: algorithms for morphological trees in edge-weighted graphs*, in International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing, Springer, 2013, pp. 135–146.

[197] L. Najman and H. Talbot, *Mathematical morphology: from theory to applications*, ISTE-Wiley, June 2010.

[198] M. Nakashizuka, *Image regularization with higher-order morphological gradients*, in 2015 23rd European Signal Processing Conference (EUSIPCO), IEEE, 2015, pp. 1820–1824.

[199] Z. Niu, G. Zhong, and H. Yu, *A review on the attention mechanism of deep learning*, Neurocomputing, 452 (2021), pp. 48–62.

[200] S. J. Nowlan and G. E. Hinton, *Simplifying neural networks by soft weight-sharing*, Neural computation, 4 (1992), pp. 473–493.

[201] S. Osher and L. Rudin, *Shocks and other nonlinear filtering applied to image processing*, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 1567, Dec. 1991, pp. 414–431.

[202] G. Ouzounis and P. Soille, *Pattern spectra from partition pyramids and hierarchies*, in Proc. of 10th Int. Symp. on Mathematical Morphology, vol. 6671 of Lecture Notes in Computer Science, Springer-Verlag, 2011, pp. 108–119.

[203] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, *Deep learning for anomaly detection: A review*, ACM Computing Surveys (CSUR), 54 (2021), pp. 1–38.

[204] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, *Deep learning classifiers for hyperspectral imaging: A review*, ISPRS Journal of Photogrammetry and Remote Sensing, 158 (2019), pp. 279–317.

[205] N. Parikh and S. Boyd, *Proximal algorithms*, Foundations and Trends in optimization, 1 (2014), pp. 127–239.

[206] V. Penaud-Polge, S. Velasco-Forero, and J. Angulo, *Fully trainable gaussian derivative convolutional layer*, in 2022 IEEE International Conference on Image Processing (ICIP), IEEE, 2022, pp. 2421–2425.

[207] X. Pennec, *Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements*, Journal of Mathematical Imaging and Vision, 25 (2006), pp. 127–154.

[208] B. Perret, G. Chierchia, J. Cousty, S. F. G. aes, Y. Kenmochi, and L. Najman, *Higra: Hierarchical graph analysis*, SoftwareX, 10 (2019), pp. 1–6.

[209] B. Perret and J. Cousty, *Component tree loss function: Definition and optimization*, arXiv preprint arXiv:2101.08063, (2021).

[210] L. F. Pessoa and P. Maragos, *MRL-filters: A general class of nonlinear systems and their optimal design for image processing*, IEEE Transactions on Image Processing, 7 (1998), pp. 966–978.

[211] ——, *Neural networks with hybrid morphological/rank/linear nodes: a unifying framework with applications to handwritten character recognition*, Pattern Recognition, 33 (2000), pp. 945–960.

[212] G. Peyré, *Manifold models for signals and images*, Computer Vision and Image Understanding, 113 (2009), pp. 249–260.

[213] G. Peyré and M. Cuturi, *Computational optimal transport*, Foundations and Trends in Machine Learning, 11 (2019), pp. 1–257.

[214] X. Pitkow, *Exact feature probabilities in images with occlusion*, Journal of vision, 10 (2010), pp. 42–42.

[215] X. Qiu, Y. Ren, P. N. Suganthan, and G. A. Amaratunga, *Empirical mode decomposition based ensemble deep learning for load demand time series forecasting*, Applied Soft Computing, 54 (2017), pp. 246–255.

[216] H. Qu, P. Wu, Q. Huang, J. Yi, Z. Yan, K. Li, G. M. Riedlinger, S. De, S. Zhang, and D. N. Metaxas, *Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images*, IEEE transactions on medical imaging, 39 (2020), pp. 3655–3666.

[217] H. Ravishankar, R. Venkataramani, S. Thiruvenkadam, P. Sudhakar, and V. Vaidya, *Learning and incorporating shape models for semantic segmentation*, in International conference on medical image computing and computer-assisted intervention, Springer, 2017, pp. 203–211.

[218] I. S. Reed and X. Yu, *Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution*, IEEE Trans. Acoust. Speech, Signal Process., 38 (1990), pp. 1760–1770.

[219] G. X. Ritter and P. Sussner, *An introduction to morphological neural networks*, in 13th International Conf. on Pattern Recognition, vol. 4, IEEE, 1996, pp. 709–717.

[220] J.-F. Rivest, P. Soille, and S. Beucher, *Morphological gradients*, Journal of Electronic Imaging, 2 (1993), pp. 326–336.

[221] D. A. Roberts, S. Yaida, and B. Hanin, *The Principles of Deep Learning Theory: An Effective Theory Approach to Understanding Neural Networks*, Cambridge University Press, 2022.

[222] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, vol. 317, Springer Science & Business Media, 2009.

[223] J. B. T. M. Roerdink, *Adaptivity and group invariance in mathematical morphology*, in International Conference of Image Processing, IEEE, 2009, pp. 2253–2256.

[224] ——. personal communication, 2012.

[225] R. Rojas, *The backpropagation algorithm*, in Neural networks, Springer, 1996, pp. 149–182.

[226] C. Ronse, *Partial partitions, partial connections and connective segmentation*, Journal of Mathematical Imaging and Vision, 32 (2008), pp. 97–125.

[227] ——, *Idempotent block splitting on partial partitions, i: Isotone operators*, Order, 28 (2011), pp. 273–306.

[228] C. Ronse and J. Serra, *Fondements algébriques de la morphologie*, in Morphologie Mathématique 1: Approches Déterministes, L. Najman and H. Talbot, eds., Lavoisier, Paris, 2008, ch. 2, pp. 49–96.

[229] F. Rosenblatt, *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*, tech. rep., Cornell Aeronautical Lab Inc Buffalo NY, 1961.

[230] Y. Rubner, C. Tomasi, and L. J. Guibas, *Earth mover's distance as a metric for image retrieval*, International Journal of Computer Vision, 40 (2000), pp. 99–121.

[231] L. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. G. Dietterich, and K.-R. Müller, *A unifying review of deep and shallow anomaly detection*, Proceedings of the IEEE, 109 (2021), pp. 756–795.

[232] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, *Deep one-class classification*, in International conference on machine learning, PMLR, 2018, pp. 4393–4402.

[233] L. Ruff, R. A. Vandermeulen, N. Görnitz, A. Binder, E. Müller, K.-R. Müller, and M. Kloft, *Deep semi-supervised anomaly detection*, in International Conference on Learning Representations, 2020.

[234] L. K. S. Bartels and S. Scholtes, *Continuous selections of linear functions and nonsmooth critical point theory*, Nonlinear Analysis, Theory and Applications, 24 (1995), pp. 385–407.

[235] P. Salembier, A. Oliveras, and L. Garrido, *Antiextensive connected operators for image and sequence processing*, IEEE Transactions on Image Processing, 7 (1998), pp. 555–570.

[236] P. Salembier and J. Serra, *Morphological multiscale image segmentation*, in Visual Communications and Image Processing'92, vol. 1818, International Society for Optics and Photonics, 1992, pp. 620–631.

[237] ——, *Flat zones filtering, connected operators, and filters by reconstruction*, IEEE Transactions on Image Processing, 4 (1995), pp. 1153–1160.

[238] M. Sangalli, S. Blusseau, S. Velasco-Forero, and J. Angulo, *Scale equivariant neural networks with morphological scale-spaces*, in International Conference on Discrete Geometry and Mathematical Morphology, Springer, 2021, pp. 483–495.

[239] ——, *Differential invariants for SE(2)-equivariant networks*, in 2022 IEEE International Conference on Image Processing (ICIP), IEEE, 2022, pp. 2216–2220.

[240] ——, *Moving frame net: SE(3)-equivariant network for volumes*, in Symmetry and Geometry in Neural Representations, NeurReps, 2022.

[241] ——, *Scale equivariant u-net*, in British Machine Vision Conference, 2022.

[242] A. Schaum, *Hyperspectral anomaly detection beyond RX*, in Defense and Security Symposium, International Society for Optics and Photonics, 2007, pp. 656502–656502.

[243] J. Schavemaker, M. J. T. Reinders, J. J. Gerbrands, and E. Backer, *Image sharpening by morphological filtering*, Pattern Recognition, 33 (2000), pp. 997–1012.

[244] M. Schmidt and J. Weickert, *Morphological counterparts of linear shift-invariant scale-spaces*, Journal of Mathematical Imaging and Vision, 56 (2016), pp. 352–366.

[245] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, *Grad-cam: Visual explanations from deep networks via gradient-based localization*, in Proceedings of the IEEE international conference on computer vision, 2017, pp. 618–626.

[246] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, Inc., Orlando, FL, USA, 1983.

[247] ——, *Image Analysis and Mathematical Morphology, Vol. 2: Theoretical Advances*, Academic Press, 1st ed., Feb. 1988.

[248] ——, *Mathematical morphology for Boolean lattices*, in Image Analysis and Mathematical Morphology. Volume 2: Theoretical Advances, J. Serra, ed., Academic Press, 1988, ch. 2, pp. 37–58.

[249] ——, *The "false colour" problem*, in ISMM '09: Proceedings of the 9th International Symposium on Mathematical Morphology and Its Application to Signal and Image Processing, Berlin, Heidelberg, 2009, Springer-Verlag, pp. 13–23.

[250] R. C. Sharpley and V. Vatchev, *Analysis of the intrinsic mode functions*, Constructive Approximation, 24 (2006), pp. 17–47.

[251] P. Soille, *Morphological Image Analysis*, Springer-Verlag, 2003.

[252] ——, *Constrained connectivity for hierarchical image partitioning and simplification*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30 (2008), pp. 1132–1145.

[253] ——, *Preventing chaining through transitions while favouring it within homogeneous regions*, in Proc. of 10th Int. Symp. on Mathematical Morphology, vol. 6671 of Lecture Notes in Computer Science, Springer-Verlag, 2011, pp. 96–107.

[254] P. Soille and H. Talbot, *Directional morphological filtering*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (2001), pp. 1313–1329.

[255] A. Stallone, A. Cicone, and M. Materassi, *New insights and best practices for the successful use of empirical mode decomposition, iterative filtering and derived algorithms*, Scientific reports, 10 (2020), pp. 1–15.

[256] C. Stein, *Estimation of a covariance matrix*, Rietz Lecture, (1975).

[257] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, *Anomaly detection from hyperspectral imagery*, Signal Processing Magazine, IEEE, 19 (2002), pp. 58–69.

[258] I. Steinwart, D. Hush, and C. Scovel, *A classification framework for anomaly detection.*, Journal of Machine Learning Research, 6 (2005).

[259] R. Sullivan, *Life: 100 photographs that changed the world*, Time-Life Books, 2003.

[260] Y. Sun, P. Babu, and D. Palomar, *Regularized Tyler's scatter estimator: Existence, uniqueness, and algorithms*, IEEE Transactions on Signal Processing, 62 (2014), pp. 5143–5156.

[261] P. Sussner and I. Campiotti, *Extreme learning machine for a new hybrid morphological/linear perceptron*, Neural Networks, 123 (2020), pp. 288–298.

[262] A. Taguchi and T. Matsumoto, *Removal of impulse noise from highly corrupted images by using noise position information and directional information of image*, in Proc. SPIE 4304, Nonlinear Image Processing and Pattern Analysis XII, 2001, pp. 188–196.

[263] H. Talbot, C. Evans, and R. Jones, *Complete ordering and multivariate mathematical morphology*, in ISMM '98: Proceedings of the fourth international symposium on Mathematical morphology and its applications to image and signal processing, Norwell, MA, USA, 1998, Kluwer Academic Publishers, pp. 27–34.

[264] K. Tanaka, *Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities.*, Cerebral Cortex, 13 1 (2003), pp. 90–9.

[265] D. M. J. Tax and R. P. W. Duin, *Support Vector Data Description*, Machine Learning, 54 (2004), pp. 45–66.

[266] J. Theiler, *By definition undefined: Adventures in anomaly (and anomalous change) detection*, in IEEE Workshop on Hyperspectral Image and Signal Processing, 2014.

[267] J. Theiler, G. Cao, L. Bachega, and C. Bouman, *Sparse matrix transform for hyperspectral image processing*, IEEE Journal of Selected Topics in Signal Processing, 5 (2011), pp. 424–437.

[268] F. J. Theis, K. Stadlthanner, and T. Tanaka, *First results on uniqueness of sparse non-negative matrix factorization*, in 13th IEEE European Signal Processing Conference, 2005, pp. 1–4.

[269] D. E. Tyler, *A distribution-free M-estimator of multivariate scatter*, The Annals of Statistics, 15 (1987), pp. 234–251.

[270] C. Vachier and F. Meyer, *Extinction value: a new measurement of persistence*, in IEEE Workshop on Non Linear Signal/Image Processing, 1995, pp. 254–257.

[271] M. E. Valle, *Reduced dilation-erosion perceptron for binary classification*, Mathematics, 8 (2020), p. 512.

[272] R. Van Den Boomgaard, L. Dorst, S. Makram-Ebeid, and J. Schavemaker, *Quadratic structuring functions in mathematical morphology*, in Mathematical morphology and its applications to image and signal processing, Springer, 1996, pp. 147–154.

[273] R. Van den Boomgaard and A. Smeulders, *The morphological structure of images: The differential equations of morphological scale-space*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 16 (1994), pp. 1101–1113.

[274] L. van Vliet, I. Young, and G. Beckers, *A nonlinear Laplace operator as edge detector in noisy images*, Computer Vision, Graphics, and Image Processing, 45 (1989), pp. 167–195.

[275] L. J. van Vliet, I. T. Young, and G. L. Beckers, *A nonlinear Laplace operator as edge detector in noisy images*, Computer Vision, Graphics, and Image Processing, 45 (1989), pp. 167 – 195.

[276] S. Velasco-Forero and J. Angulo, *Mathematical morphology for vector images using statistical depth*, in Mathematical Morphology and Its Applications to Image and Signal Processing, vol. 6671 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2011, pp. 355–366.

[277] ——, *Supervised ordering in R^n: Application to morphological processing of hyperspectral images*, IEEE Transactions on Image Processing, 20 (2011), pp. 3301–3308.

[278] ——, *Classification of hyperspectral images by tensor modeling and additive morphological decomposition*, Pattern Recognition, 46 (2012), pp. 566–577.

[279] ——, *Random projection depth for multivariate mathematical morphology*, IEEE Journal of Selected Topics in Signal Processing, 6 (2012), pp. 753–763.

[280] ——, *On nonlocal mathematical morphology*, in International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing, Springer, 2013, pp. 219–230.

[281] ——, *Vector Ordering and Multispectral Morphological Image Processing*, in Advances in Low-Level Color Image Processing, M. E. Celebi and B. Smolka, eds., Dordrecht, 2014, Springer Netherlands, pp. 223–239.

[282] ——, *Non-Negative Sparse Mathematical Morphology*, vol. 202, Elsevier Inc.Academic Press, 2017, ch. 1.

[283] ——, *Morphoactivation: Generalizing relu activation function by mathematical morphology*, in DGMM, 2022.

[284] S. Velasco-Forero, J. Angulo, and P. Soille, *Conditional toggle mappings: principles and applications*, Journal of mathematical imaging and vision, 48 (2014), pp. 544–565.

[285] S. Velasco-Forero, M. Chen, A. Goh, and S. K. Pang, *A comparative analysis of covariance matrix estimation in anomaly detection*, in IEEE Workshop on Hyperspectral Image and Signal Processing, 2014.

[286] S. Velasco-Forero, M. Chen, A. Goh, and S. K. Pang, *Comparative analysis of covariance matrix estimation for anomaly detection in hyperspectral images*, IEEE Journal of Selected Topics in Signal Processing, 9 (2015), pp. 1061–1073.

[287] S. Velasco-Forero, R. Pagès, and J. Angulo, *Learnable empirical mode decomposition based on mathematical morphology*, SIAM Journal on Imaging Sciences, 15 (2022), pp. 23–44.

[288] S. Velasco-Forero, A. Rhim, and J. Angulo, *Fixed point layers for geodesic morphological operations*, in British Machine Vision Conference, 2022.

[289] C. Villani, *Optimal Transport*, vol. 338 of Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[290] S. Wang, *General constructive representations for continuous piecewise-linear functions*, IEEE Transactions on Circuits and Systems I: Regular Papers, 51 (2004), pp. 1889–1896.

[291] T. Wang and P. Isola, *Understanding contrastive representation learning through alignment and uniformity on the hypersphere*, in International Conference on Machine Learning, PMLR, 2020, pp. 9929–9939.

[292] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, *Dynamic graph cnn for learning on point clouds*, Acm Transactions On Graphics (tog), 38 (2019), pp. 1–12.

[293] M. Welk, J. Weickert, and I. Galć, *Theoretical foundations for spatially discrete 1-d shock filtering*, Image and Vision Computing, 25 (2007), pp. 455–463.

[294] A. Wiesel, *Unified framework to regularized covariance estimation in scaled gaussian models*, IEEE Transactions on Signal Processing, 60 (2012), pp. 29–38.

[295] S. Wolf, L. Schott, U. Kothe, and F. Hamprecht, *Learned watershed: end-to-end learning of seeded segmentation*, in IEEE International Conference on Computer Vision, 2017, pp. 2011–2019.

[296] J.-H. Won, J. Lim, S.-J. Kim, and B. Rajaratnam, *Condition-number-regularized covariance estimation*, Journal of the Royal Statistical Society: Series B, 75 (2013), pp. 427–450.

[297] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, *CBAM: Convolutional block attention module*, in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3–19.

[298] D. Worrall and M. Welling, *Deep scale-spaces: Equivariance over scale*, Advances in Neural Information Processing Systems, 32 (2019).

[299] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, *Sparse Representation for Computer Vision and Pattern Recognition*, Proceedings of the IEEE, 98 (2010), pp. 1031–1044.

[300] H. Xu, G. Zhu, H. Peng, and D. Wang, *Adaptive fuzzy switching filter for images corrupted by impulse noise*, Pattern Recogn. Lett., 25 (2004), pp. 1657–1663.

[301] Z. Yan and X. S. Zhou, *How intelligent are convolutional neural networks?*, arXiv preprint arXiv:1709.06126, (2017).

[302] J. Yang, J. Wright, T. S. Huang, and Y. Ma, *Image super-resolution via sparse representation*, IEEE Transactions on Image Processing, 19 (2010), pp. 2861–2873.

[303] F. Zanoguera and F. Meyer, *On the implementation of non-separable vector levelings*, in Proceedings of VIth Int. Symp. on Mathematical Morphology, Sydney, Australia, 2002, Commonwealth Scientific and Industrial Research Organisation, pp. 369–377.

[304] S. Zhang and M. Karim, *A new impulse detector for switching median filters*, IEEE Signal Processing Letters, 9 (2002), pp. 360–363.

[305] Y. Zhang, S. Blusseau, S. Velasco-Forero, I. Bloch, and J. Angulo, *Max-plus operators applied to filter selection and model pruning in neural networks*, in International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing, Springer, 2019, pp. 310–322.

[306] C. Zhou and R. C. Paffenroth, *Anomaly detection with robust deep autoencoders*, in Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, 2017, pp. 665–674.

[307] Y. Zhou, Z. Li, S. Bai, C. Wang, X. Chen, M. Han, E. Fishman, and A. L. Yuille, *Prior-aware neural network for partially-supervised multi-organ segmentation*, in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 10672–10681.

[308] J. Zhu, S. Rosset, R. Tibshirani, and T. J. Hastie, *L1-norm support vector machines*, in Advances in NIPS, 2004, pp. 49–56.

A Curriculum Vitae

A.1 Personal information

Name: Santiago VELASCO-FORERO
Address: Résidence du Parc du Château de Courcelle
91190 Gif-surYvette
Married, Two Children
Languages: English, French, Spanish

A.2 Research experience

(Oct. 2014 - Currently) Chargé de recherche, Center for Mathematical Morphology, MINES ParisTech, PSL Research University, Paris, France.
Scientific advisor for nine Ph.D. students, three post-doctoral researchers and 13 master students.
Development of several industrial collaborations including Thales, Safran, L'Oreal.
Co-responsible of Deep Learning course for Mines Paris-Tech.
Author of a Tensorflow library on Deep Learning via mathematical morphology operators

(Jun. 2013 - Sep. - 2014) Research Fellow, Department of Mathematics, National University of Singapore, Singapore.

(Sept. 2012 - Jun. 2013) PostDoc, ITWM Fraunhofer Institute, Kaiserslautern, Germany.

(Oct. 2009 - Ago. 2012) R&D Engineer/Ph.D. Student, MINES ParisTech, Centre of Mathematical Morphology, Fontainebleau, France.

(Jan. 2011 - Mar. 2011) R&D Engineer/Visitor, Institute for the Protection and Security of the Citizen, Joint Research Centre, European Commission, Ispra, Italy.

(Feb. 2009 - Sep. 2009) R&D Engineer, ARMINES, France.

(Jan. 2007- Dec. 2008) R&D Engineer, Center for Subsurface Sensing and Imaging Systems, Puerto Rico, USA.

(Jun. 2005 - Jun. 2006) Research Assistant, Computational and Statistical Learning Group, Puerto Rico, USA.

A.3 Education

(Sept. 2009 - Jun. 2012)
MINES ParisTech, *Ph.D.* in Image Processing (European label).
Title: "Topics in Mathematical Morphology to Multivariate Image Analysis".
(Jun. 2003 - Jun. 2005)
University of Puerto Rico, *M.Sc.* in Mathematics (Machine Learning).
(Jan. 1997 - Mar. 2002)
National University of Colombia, *B.Sc.* in Statistics.

A.4 Professional Service

A.4.1 Reviewer

- **IA Conferences**: IJCAI 2021/2022 WCCI 2022 ECML-PKDD 2022
- **Image Processing and Signal Processing journals**: Journal of Mathematical Imaging and Vision, IEEE Transactions on Image Processing, IEEE Selected Topics in Signal Processing, IEEE Transactions on Signal Processing, IEEE Signal Processing Letters, Digital Signal Processing, Image Analysis and Stereology, Information Fussion
- **Remote Sensing journals** IEEE-Transactions on Geoscience and Remote Sensing IEEE-Special Topics in Remote Sensing International Journal of Remote Sensing
- **Pattern Recognition journals** IEEE Transactions on Pattern Analysis and Machine Intelligence IEEE Transactions on Neural Networks. and Learning Systems Neurocomputing Pattern Recognition Pattern Recognition Letters Information Sciences Applied Mathematical Modelling

A.4.2 Participation on Ph.D. thesis jury

- Francisco José Peñaranda Gómez, *Application of artificial vision algorithms to images of microscopy and spectroscopy for the improvement of cancer diagnosis*, Universitat Politècnica de València, 2018.
- Juan Manuel Ponce Real, *Computer Vision in Oliviculture. Contributions to the post-harvest estimation of individual fruit features, early in-the-field yield prediction, and individual tree characterisation from aerial imagery, by means of image analysis*, University of Huelva, 2020

- Maria Ximena Bastidas Rodriguez, *A textural deep neural network architecture for mechanical failure analysis*, National University of Colombia, 2020.
- José Gabriel García Pardo, *Machine learning strategies for diagnostic imaging support on histopathology and optical coherence tomography*, Universitat Politécnica de València, 2022.

A.5 Publications

A.5.1 Patents

1. Method for Automatically Reconstituting the Reinforcing Architecture of a Composite Material
   WO2021116602 - 17/06/2021
2. Confidential / Submitted on 1 mars 2022 / Number of Application: FR2201789

A.5.2 Thesis

Topics in mathematical morphology for multivariate images, 14 June 2012
École nationale supérieure des mines de Paris (Mines ParisTech)
Advisor: Jesús ANGULO, Chargé de recherche, CMM, Mines ParisTech
Jury:

- Dominique JEULIN, Professeur, CMM-MS, Mines ParisTech (Président)
- Jos B.T.M. ROERDINK, Professeur, University of Groningen (Rapporteur)
- Pierre SOILLE, Directeur de recherche, Joint Research Centre of the European Commission (Rapporteur)
- Jean SERRA, Professeur émérite, ESIEE, Université Paris-Est (Examinateur)
- Jón Atli BENEDIKTSSON, Professeur, University of Iceland (Examinateur)
- Fernand MEYER, Directeur de recherche, CMM, Mines ParisTech (Examinateur)

A.5.3 Proceedings Editor

E1: Mathematical Morphology and Its Applications to Signal and Image Processing 13th International Symposium, ISMM 2017, Fontainebleau, France, May 15–17, 2017, Proceedings,

A.5.4 International Journal

J1: Learnable Empirical Mode Decomposition based on Mathematical Morphology, S. Velasco-Forero, R. Pages and J. Angulo, *SIAM Journal on Imaging Sciences*, 15 (1), 23-44, 2022.

J2: Real-time classification of aircraft manoeuvers, S. Jouaber et al., *Journal of Signal Processing Systems*, 2022.

J3: Instance segmentation of 3D woven fabric from tomography images by Deep Learning and morphological pseudo-labeling, S. Blusseau et al, *Composites Part B: Engineering*, 2022.

J4: Irregularity Index for Vector-Valued Morphological Operators, M. Valle, M. Granero, S. Francisco and S. Velasco-Forero, *Journal of Mathematical Imaging and Vision*, 1-17, 2022.

J5: Adaptive anisotropic morphological filtering based on co-circularity of local orientations, S. Blusseau, S. Velasco-Forero, J.Angulo and I. Bloch. *Image Processing On Line*,12, 111-141, 2022.

J6: Paris-CARLA-3D: A real and synthetic outdoor point cloud dataset for challenging tasks in 3D mapping, J-E Deschaud et al. *Remote Sensing*, Volume 12, 22, 4713, 2021.

J7: On minimum spanning tree streaming for hierarchical segmentation, L. Gigli, S. Velasco-Forero and B. Marcotegui, *Pattern Recognition Letters* 138, 155–62, 2020.

J8: SHREC'20 Track: 3D Point Cloud Semantic Segmentation for Street Scenes, T. Ku et al., *Computer and Graphics*, Volume 93, , 13–24, December 2020.

J9: SHREC'20 track: Retrieval of digital surfaces with similar geometric reliefs, E. Moscoso-Thompson et al., *Computers and Graphics*, Volume 91, 199–218, October 2020

J10: Approximating morphological operators with part-based representations learned by asymmetric auto-encoders, S. Blusseau et al., *Mathematical Morphology-Theory and Applications*, 4: 64-86, 2020.

J11: Combinatorial space of watershed hierarchies for image characterization, A. Fehri, S. Velasco-Forero and F. Meyer, *Pattern Recognition Letters*, Volume 129, 41–47, January 2020

J12: Prior-based Hierarchical Segmentation Highlighting Structures of Interest, A. Fehri, S. Velasco-Forero and F. Meyer, *Mathematical Morphology-Theory and Applications*, Volume 3, 29–44, October 2019.

J13: The strong gravitational lens finding challenge, R.B. Metcalf et al., *Astronomy and Astrophysics*, Volume 625, May 2019

J14: On-the-go grapevine yield estimation using image analysis and Boolean model, B. Millan, S. Velasco-Forero, A. Aquino and J. Tardaquila, *Journal of Sensors*, 0–14, December 2018

J15: Manipulating the alpha level cannot cure significance testing, D. Trafimow et al., *Frontiers in Psychology*, May 2018

J16: Deep learning for galaxy surface brightness profile fitting, D. Tuccillo, M. Huertas-Company, E. Decenciére, S. Velasco-Forero, H. Domínguez-Sánchez and P. Dimauro, *Monthly Notices of the Royal Astronomical Society*, Dec., 2017

J17: Non-Negative Sparse Mathematical Morphology, J. Angulo and S. Velasco-Forero, *Advances in Imaging and Electron Physics* Volume 202, 1-37, 2017.

J18: Retrieval and classification methods for textured 3D models: A comparative study, S.Biasotti, et al., *The Visual Computer Journal*, August, 1–25, 2015.

J19: Comparative Analysis of Covariance Matrix Estimation for Anomaly Detection in Hyperspectral Images, S. Velasco-Forero, M.Chen, A. Goh and S.K. Pang, *IEEE Journal of Selected Topics in Signal Processing*, Volume 9, Nro. 6, 1061–1073, Sept 2015.

J20: Riemannian mathematical morphology, J. Angulo and S. Velasco-Forero, *Pattern Recognition Letters*, Volume 47, Nro. 1, 93–101, October 2014.

J21: Conditional toggle mappings: principles and applications, S. Velasco-Forero, J. Angulo and P. Soille, *Journal of Mathematical Imaging and Vision*, Volume 48, Issue 3, pp. 544-565, March 2014.

J22: Local mutual information for dissimilarity based image segmentation, L. Gueguen, S. Velasco-Forero and P. Soille, *Journal of Mathematical Imaging and Vision*, Volume 48, Issue 3, pp 625-644, March 2014.

J23: Classification of hyperspectral images by tensor modeling and additive morphological decomposition, S. Velasco-Forero and J. Angulo, *Pattern Recognition*, vol. 46, num. 2, Feb. 2013,

J24: Random projection depth for multivariate mathematical morphology, S. Velasco-Forero and J. Angulo, *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, num. 7, Oct. 2012, .

J25: Supervised ordering in $\mathbb{R}^p$: Application to morphological processing of hyperspectral images, S. Velasco-Forero and J. Angulo, *IEEE Transactions on Image Processing*, vol. 20, num. 11, Oct. 2011.

J26: Structurally adaptive mathematical morphology,*Image Analysis and Stereology*, J. Angulo and S. Velasco-Forero, num. 30, 101-112, 2011, .

J27: Improving hyperspectral image classification using spatial preprocessing, *IEEE Geoscience and Remote Sensing Letters*, S. Velasco-Forero and V. Manian, vol.6, num. 2, pp. 297–301, 2008.

A.5.5 Chapter Books

CB1: Morphological processing of univariate Gaussian distribution valued images based on Poincaré upper-half plane representation, J. Angulo and S. Velasco-Forero, Geometric Theory of Information Signals and Communication Technology, pp. 331-366, 2014.

CB2: Vector Ordering and Multispectral Morphological Image Processing, J. Angulo and S. Velasco-Forero, Advances in Low-Level Color Image Processing, Signals and Communication Technology, pp. 223-239, 2014.

A.5.6 International conference with lecture committee

IC1: Moving Frame Net: SE(3)-Equivariant Networks for Volumes, M. Sangalli et al., *Workshop Symmetry and Geometry in Neural Representation, NIPS*, 2022

IC2: Scale Equivariant U-Net, M. Sangalli et al., *BMVC*, 2022

IC3: Fixed Point Layers for Geodesic Morphological Operations. S. Velasco-Forero et al., *BMVC*, 2022

IC4: Near out-of-distribution detection for low-resolution radar micro-Doppler signatures, M. Bauw et al, *ECML*, 2022

IC5: Fully Trainable Gaussian Derivative Convolutional Layer, V. Penaud-Polge et al., *ICIP*, 2022

IC6: Differential Invariants for SE(2)-Equivariant Networks, M. Sangalli et al., *ICIP*, 2022

IC7: MorphoActivation: Generalizing ReLU activation function by mathematical morphology, S. Velasco-Forero and J. Angulo, *DGMM*, 2022

IC8: Morphological adjunctions represented by matrices in max-plus algebra for signal and image processing, S. Blusseau et al., *DGMM*, 2022

IC9: DarthBoard based ground detection on 3D Point Cloud, Gigli L.,Marcotegui B., Velasco-Forero S.,*ISPRS*, 2022.

IC10: ICLR 2021 Challenge for Computational Geometry and Topology: Design and Results, N. Miolane, et al. , *ICLR*, 2021.

IC11: End-to-End Similarity Learning and Hierarchical Clustering for Unfixed Size Datasets. Gigli L., Marcotegui B., Velasco-Forero S., In: Nielsen F., Barbaresco F. (eds) *Geometric Science of Information 2021.* Lecture Notes in Computer Science, vol 12829. Springer, Cham.

IC12: Scale Equivariant Neural Networks with Morphological Scale-Spaces, M. Sangalli, S Blusseau, S Velasco-Forero, J Angulo, *DGMM*, 2021.

IC13: Measuring the Irregularity of Vector-valued Morphological Operators using Wasserstein Metric, ME Valle, S Francisco, MA Granero, S Velasco-Forero, *DGMM*, 2021.

IC14: NNAKF: A neural network adapted Kalman filter for target tracking, S Jouaber, S Bonnabel, S Velasco-Forero, M Pilté, *IEEE International Conference on Acoustics Speech and Signal Processing* 2021.

IC15: On power Jaccard losses for semantic segmentation, D. Duque-Arias, S.Velasco-Forero, et al., *VISAPP* 2021.

IC16: From unsupervised to semi-supervised anomaly detection methods for High Resolution Range Profiles, M. Bauw, S. Velasco-Forero et al. *IEEE Radar Conference* 2020.

IC17: Road segmentation on low resolution lidar point clouds for autonomous vehicles, L. Gigli, B. Marcotegui and S. Velasco-Forero *ISPRS*, 2020.

IC18: A graph-based color lines model for image analysis, D. Duque-Arias, S. Velasco-Forero, J.-E. Deschaud, F. Goulette, B. Marcotegui, *CIAP*, 2019.

IC19: Part-based approximations for morphological operators using asymmetric auto-encoders, B. Ponchon, S. Velasco-Forero, S. Blusseau, J. Angulo, and I. Bloch, *ISMM*, 2019.

IC20: Max-plus operators applied to filter selection and model pruning in neural networks, Y Zhang, S Blusseau, S. Velasco-Forero, I. Bloch, J. Angulo, *ISMM*, 2019.

IC21: A New Color Augmentation Method for Deep Learning Segmentation of Histological Images, Y. Xiao, E. Decenciére, S. Velasco-Forero et al., *ISBI*, 2019.

IC22: Tropical and morphological operators for signals on graphs, S. Blusseau et al., *ICIP*, 2018.

IC23: Dealing with Topological Information within a Fully Convolutional Neural Network, E. Decenciere et al., *ACIVS*, 2018

IC24: SHREC'18 track: Recognition of geometric patterns over 3D models, *11th Eurographics Workshop on 3D Object Retrieval*, 2018.

IC25: SHREC'18 track: Retrieval of gray patterns depicted on 3D models, *11th Eurographics Workshop on 3D Object Retrieval*, 2018.

IC26: SHREC'17 Track: Point-Cloud Shape Retrieval of Non-Rigid Toys, FA Limberger, et al., *10th Eurographics Workshop on 3D Object Retrieval*, 2017.

IC27: SHREC'17 Track: Retrieval of surfaces with similar relief patterns, S Biasotti, et al., *10th Eurographics Workshop on 3D Object Retrieval*, 2017.

IC28: Morphological Semigroups and Scale-Spaces on Ultrametric Spaces, J. Angulo and S. Velasco-Forero, *ISMM*, 2017

IC29: Nonlinear Operators on Graphs via Stacks, S. Velasco-Forero and J. Angulo, *Geometric Science of Information*, 2016.

IC30: Automatic selection of Stochastic Watershed Hierarchies, A. Fehri, S. Velasco-Forero and F. Meyer, *EUSIPCO*, 2016.

IC31: Shrec'16 Retrieval of Human Subjects from Depth Sensor Data, A. Giachetti et al., *Eurographics Workshop on 3D Object Retrieval*, 2016.

IC32: A bayesian approach to linear unmixing in the presence of highly mixed spectra. B. Figliuzzi, et al, *International Conference on Advanced Concepts for Intelligent Vision Systems*, 2016.

IC33: Objects co-segmentation: Propagated from simpler images M. Chen, S. Velasco-Forero, I. Tsang and T.J. Cham, *IEEE International Conference on Acoustics Speech and Signal Processing*, 2015

IC34: Inner-Cheeger Opening and Applications S. Velasco-Forero, *Mathematical Morphology and Its Applications to Signal and Image Processing*, LNCS, vol. 9082, 2015.

IC35: Robust Anomaly Detection in Hyperspectral imaging J. Frontera-Pons et al., *IEEE International Geoscience and Remote Sensing Symposium* (IGARSS), 2014.

IC36: SHREC-14 Track: Retrieval and classification on Textured 3D Models, S. Biasotti et al., *Eurographics Workshop on 3D Object Retrieval*, 2014.

IC37: On nonlocal mathematical morphology S. Velasco-Forero and J. Angulo, *Mathematical Morphology and Its Applications to Signal and Image Processing*, LNCS, vol. 7883, pp. 219-230, 2013.

IC38: Stochastic morphological filtering and Bellman-Maslov chains J. Angulo and S. Velasco-Forero, *Mathematical Morphology and Its Applications to Signal and Image Processing*, LNCS, vol. 7883, pp. 171-182, 2013.

IC39: Supervised morphology for tensor structure-valued images based on symmetric divergence kernels S. Velasco-Forero and J. Angulo, in *Geometric Science of Information*, vol. 8085, pp. 543-550, 2013.

IC40: SHREC'13 Track: Retrieval on textured 3D models, A. Cerri et al., *Eurographics Workshop on 3D Object Retrieval*, pp. 73-80, 2013.

IC41: Edge Extraction by statistical dependence analysis: Application to multi-angular Worldview-2 series, L. Gueguen, S. Velasco-Forero and P. Soille, *IGARSS International Geoscience and Remote Sensing Symposium*, pp. 3447-3450, 2012.

IC42: Robust RX Anomaly Detector without covariance matrix estimation, S.Velasco-Forero and J. Angulo, 2012 *4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, June, pp. 1-4, 2012.

IC43: Hit-or-Miss Transform in Multivariate Images S.Velasco-Forero and J.Angulo in *Advanced Concepts for Intelligent Vision Systems*, vol. 6474, 2010, pp. 452-463, LNCS, 2010.

IC44: Statistical shape modeling using morphological representations, S. Velasco-Forero and Jesús Angulo, In : 2010 *20th International Conference on Pattern Recognition*. IEEE, 2010. p. 3537-3540.

IC45: Semi-supervised hyperspectral image segmentation using regionalized stochastic watershed. Jesús Angulo. Semi-supervised hyperspectral image segmentation using regionalized stochastic watershed. *In : Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVI*. International Society for Optics and Photonics, 2010. p. 76951F

IC46: Multiscale stochastic watershed for unsupervised hyperspectral image segmentation , J. Angulo, S. Velasco-Forero and J. Chanussot, In : 2009 *IEEE international geoscience and remote sensing symposium*. IEEE, 2009. p. III-93-III-96.

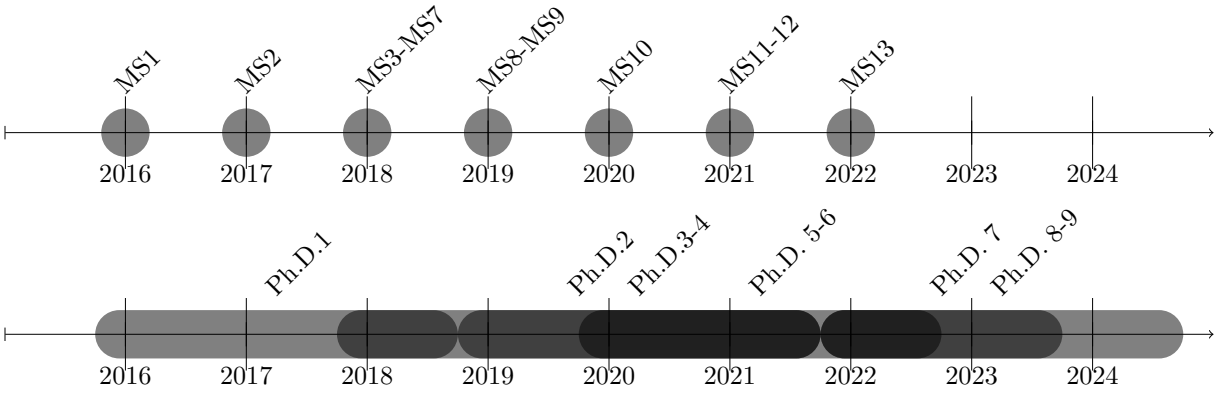
IC47: Morphological image distances for hyperspectral dimensionality exploration using Kernel-PCA and ISOMAP, S Velasco-Forero, J Angulo, J. Chanussot, In : 2009 *IEEE international geoscience and remote sensing symposium*. IEEE, 2009. III-109-III-112

B Teaching Activities and Industrial Collaborations

B.1 Teaching Responsibilities

- PreCalculus, Basic Statistics, University of Puerto Rico, 2014-2017.
- Statistics in Big Data, Department of Statistics, (50h) National University of Colombia, 2018, Summer Course.
- Statistics in Big Data, Department of Mathematics, (50h) Military University of Colombia, 2019, Summer Course.
- Deep Learning for Image Analysis with E. Decencière and T. Walter, (20h) (2019, 2020, 2021,2022), MINES ParisTech.
- Deep Learning for Image Analysis with E. Decencière and T. Walter, (20h) (2021,2022), Master 2 IASD, PSL Dauphine
- Module of Machine Learning of Geosciences (4h) (2020,2021) (C-A. Azencott), École Normale Supérieure d'Ulm.
- Module of Mathematical Morphology (2h) (2018,2021) (B.Marcotegui), École de Mines de Paris.
- ECSIA mini-course (2h): Mathematical Morphology meets Deep Learning with Samy Blusseau and Mateus Sangalli, 2021

B.2 Advisor Activities



B.2.1 Master Students

MS1: Jairo Ivan Peña (2016) from National University of Colombia.

MS2: Umang Aggarwal (2017) was an intern at Mines ParisTech/École Normale Supérieure collaborative project between CMM and EREEP, with E. Decencière.

MS3: Sami Jouaber (2018) from MINES ParisTech was an intern in a PGMO Project in collaboration with S. Bonnabel and J. Angulo.

MS4: Yang Xiao (2018) from University Paris-Sud was an intern at CMM/MINES ParisTech in collaboration with E. Decencière (Sponsored by L'Oréal).

MS5: Mael Abgrall (2018) was an intern at CMM/MINES ParisTech in collaboration with E. Decencière (Sponsored by L'Oréal).

MS6: Yunxiang Zhang (2018) from École Polytechnique was an intern in a IMT Project in collaboration with I. Bloch, S. Blusseau, and J. Angulo .

MS7: Bastien Ponchon (2018) from ENS Paris-Saclay was an intern in a IMT Project in collaboration with I. Bloch, S. Blusseau, and J. Angulo.

MS8: Tristan Lazard (2019) from ENS/University PSL was an intern in collaboration with S. Blusseau, and E. Decencière (Sponsored by L'Oréal).

MS9: Zyad Haddad (2019) from ESIEE was an intern in collaboration with S. Blusseau on Deep Learning for Material Science (Sponsored by Safran).

MS10: Romain Pagès (2020) from École Centrale de Lyon was an intern in a PGMO Project in collaboration with J. Angulo.

MS11: Josselin Lefèvre (2021) from ESIEE was an intern in a Mines/Safran Project collaboration with S. Blusseau.

MS12: Ayoub Rhim (2021) from École des Ponts was intern in a PGMO Project in collaboration with J. Angulo.

MS13: Pierre Onghena (2022) from Maastricht University is intern in collaboration with The Cross Product

B.2.2 Ph.D. Students

Alumni

PH1: Amin FEHRI (2016-2018) (Coadvisor with F. MEYER)
Thesis title: "Image Characterization by Morphological Hierarchical Representations".
Jury: I. BLOCH (Telecom Paristech), H. SAHLI (Vrije Universiteit Brussel), T. GERAUD (Epita), C. GOMILA (Technicolor)
Current Position: Research scientist at Cairn Biosciences.

PH2: Leonardo GIGLI (2018-2021) (Coadvisor with B. MARCOTEGUI)
Thesis title: "Contributions to graph-based hierarchical analysis for images and 3D point clouds".
Jury: A. PLAZA (University of Extremadura), P. SOILLE (JRC), S. LEFÈVRE(Université Bretagne Sud), B. PERRET (ESIEE Paris), R. KIRAN (NavyaTech) J. ANGULO(Mines ParisTech)
Current Position: R&D Engineer at The Cross Product.

PH3: David DUQUE (2019-2021) (Coadvisor with B. MARCOTEGUI and J-E. DESCHAUD)
Thesis title: 3D urban scene understanding by analysis of LiDAR, color and hyperspectral data
Jury: P. MONASSE (École des Ponts, ParisTech) , P. CHECCHIN(Université Blaise Pascal) , A. NÜCHTER(Unversity of Würzburg) , S. LEFÈVRE (Université Bretagne Sud)
Current Position: R&D Engineer at The Cross Product.

PH4: Sami JOUABER (2019-2022) (Coadvisor with S. BONNABEL)
Thesis title: Utilisation de méthodes d'intelligence artificielle pour de nouveaux défis en poursuite radar
Jury: A. GIREMUS (Université de Bordeaux) , F. BOUCHARA(Université de Toulon) , F. SEPTIER (Université Bretagne Sud) , M. PILTÉ (Thales LAS)
Current Position: PostDoctoral Researcher at CEA.

Current

PH5: Martin BAUW (Coadvisor J. ANGULO) (2020-2022)

PH6: Mateus SANGALI (Coadvisors J. ANGULO and S. BLUSSEAU) (2020-2022)

PH7: Valentin PENAUD POLGE (Coadvisor J. ANGULO) (2021-2023)

PH8: João CASAGRANDE BERTOLDO (Coadvisors E. DECENCIÈRE and J. ANGULO) (2022-2024)

PH9: Stéfan BAUDIER (Coadvisor J. ANGULO) (2022-2024)

B.2.3 Post-doctoral researchers

- 2016: Diego TUCCILLO, postdoc in a collaborative project with E. DECENCIÈRE (Mines Paris) and M. HUERTAS-COMPANY (Paris Observatory).

- 2017: Samy BLUSSEAU, postdoctoral researcher in CMM/Telecom ParisTech collaborative project with J. ANGULO (Mines Paris), I. BLOCH (Telecom Paristech).
- 2022: Yuriy SINCHUK, postdoctoral researcher in a collaborative project with S. BLUSSEAU and Safran Aircraft Engines.

B.3 Industrial Collaborations

- 2017-2022: L'Oréal, Industrial project on the use of deep learning methods and mathematical morphology for histology image analysis.
- 2018-2022: Thales. Industrial/academic collaboration with the support of the DGA for the use of AI methods in radar technology.
- 2019-2022: Safran Aircraft Engines: Industrial collaboration for the development of useful methods for the analysis of tomography of materials.
- 2022: The Cross Product: Development of hierarchical segmentation methods for point clouds.

B.4 Past projects

- COMSYS (2015-2016): Estimation de matrices de COvariance et Matrices de densité pour des SYStèmes complexes (CARNOT MINES)
- M3S (2015-2016) Molecular detection with Multimodal microscopy scanner. M3S relies on a previous proof of concept and aims at retrieving specific fingerprints for simplifying Chronic Lymphocyte Leukaemia (CLL) as well as Malaria diagnosis and prognosis.
- REPLICA: (2019-2021) aims to complement existing simulation platforms with new bricks in order to bridge the gap expected in autonomous vehicle testing.