



**HAL**  
open science

# Multi-Layer Joint Segmentation Using MRF and Graph Cuts

Nicolas Lermé, Sylvie Le Hégarat-Masclé, François Malgouyres, Marie Lachaize

► **To cite this version:**

Nicolas Lermé, Sylvie Le Hégarat-Masclé, François Malgouyres, Marie Lachaize. Multi-Layer Joint Segmentation Using MRF and Graph Cuts. *Journal of Mathematical Imaging and Vision*, In press. hal-02125044v2

**HAL Id: hal-02125044**

**<https://hal.science/hal-02125044v2>**

Submitted on 17 Dec 2019 (v2), last revised 25 Sep 2020 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Multi-Layer Joint Segmentation Using MRF and Graph Cuts

Nicolas Lermé · Sylvie Le Hégarat-Masclé · François Malgouyres · Marie Lachaize

Received: date / Accepted: date

**Abstract** The problem of jointly segmenting objects, according to a set of labels (of cardinality  $L$ ), from a set of images (of cardinality  $K$ ) to produce  $K$  individual segmentations plus one joint segmentation, can be cast as a Markov Random Field model. Coupling terms in the considered energy function enforce the consistency between the individual segmentations and the joint segmentation. However, neither optimality on the minimizer (at least for particular cases), nor the sensitivity of the parameters, nor the robustness of this approach against standard ones have been clearly discussed before. This paper focuses on the case where  $L > 1$ ,  $K > 1$  and the segmentation problem is handled using graph cuts. Noticeably, some properties of the considered energy function are demonstrated, such as global optimality when  $L = 2$  and  $K > 1$ , the link

with majority voting and the link with naive Bayes segmentation. Experiments on synthetic and real images depict superior segmentation performance and better robustness against noisy observations.

**Keywords** Multiple images · Markov random field · segmentation · graph cuts · hyperspectral images

## 1 Introduction

Image segmentation is a fundamental and longstanding task in computer vision. For more than 20 years, it has been an active and competitive research topic, fueled by constant advances in technology (in particular, computing power and memory storage), mathematical modeling, machine learning, numerical optimization (see [2] and the references therein).

Considering several images at once for segmentation allows for an increase of the information while rising new questions on its exploitation. Applications of this problem include (but are not limited to) video object-based segmentation [20], interactive image segmentation [17,35], watermark removal [13] and 3D reconstruction [3].

This problem has been extensively tackled by a variety of approaches, depending on the adopted hypotheses. All these approaches basically share the same idea, namely taking advantage of the redundancy of information provided by the multiple input images to achieve better segmentation results. In the following, we briefly review these approaches ordering them with respect to their assumptions.

When no assumption is made neither on the localization nor on the geometry of the objects present in the imaged scene, *cosegmentation* [33] has been proposed as a solution to the aforementioned problem.

---

N. Lermé  
SATIE laboratory, Université Paris-Sud,  
Université Paris-Saclay, 91405 Orsay, France  
E-mail: nicolas.lerme@u-psud.fr

S. Le Hégarat-Masclé  
SATIE laboratory, Université Paris-Sud,  
Université Paris-Saclay, 91405 Orsay, France  
E-mail: sylvie.le-hegarat@u-psud.fr

F. Malgouyres  
Institut de Mathématiques de Toulouse  
UMR5219 Université de Toulouse, CNRS UPS IMT,  
F-31062 Toulouse Cedex 9, France  
and  
Institut de Recherche Technologique Saint-Exupéry,  
E-mail: francois.malgouyres@math.univ-toulouse.fr

M. Lachaize  
Veolia Recherche & Innovation,  
291 av. Dreyfous Ducas, 78520 Limay, France  
E-mail: marie.lachaize@veolia.com

The authors show that jointly segmenting an image pair with two labels can yield better results with respect to segmenting them independently. Using graph cuts (GC), they minimize an energy combining standard smoothness prior terms plus a term penalizing the deviation of foreground histograms, hence forcing the histogram statistics to be similar. The algorithm outputs two (non-necessarily coregistered) segmentations but fails when objects are dissimilar. The limits of this approach have been later pushed away using deep learning by considering more than two images and more than two labels for segmenting objects with a large variability in terms of scale, appearance, pose, viewpoint and background [27].

When the coregistration is possible but not performed, the problem consists in jointly segmenting and registering images. Exploiting the variational framework with active contours, [38] proposed to minimize iteratively an energy defined in order to both find a closed curve of the objects to segment in each image and a mapping relating the two curves. Thus, the outputs of the algorithm consist of the two individual segmentations in their original geometry and the mapping relating them. We refer the reader to [26] for a more detailed state-of-the-art on similar approaches.

This paper focuses on the situation where the input images are either already coregistered (either from the acquisition system or from some preprocessing steps) or the geometric transforms between images is known. The interest of such a problem is that we can rely on a high level of redundancy of image data provided that it is correctly modeled. Basic data fusion techniques to solve this specific problem notably include (i) independent segmentation of the input images and combination of the resulting segmentations using more or less ad-hoc rules either at pixel [18] or superpixel [19] level (usually called *decision fusion*) and (ii) segmentation of the input images based on the product of their conditional likelihoods, called *Naive Bayes* (NB). Because of the independence assumption, these techniques miss taking into account the relationships between images and they lack of robustness against noise. To overcome these difficulties, some authors have proposed to segment jointly the input images, modeling their interactions. Such approaches provide individual segmentations but not necessarily a joint one. In our case, we aim at benefiting of the image redundancy not only to reduce the errors in individual segmentations but also (and mainly) to derive a joint segmentation. Such an objective is based on the assumption that most of the deviation between images is due to complementarity or noise. In contrast to the aforementioned cosegmentation methods, no appearance assumption is made neither on objects nor on

background, thus enabling us to use distinct segmentation models for the input images. In the remaining of this paper, we will refer to this problem by “Multi-Layer Joint Segmentation” (MLJS). Table 1 summarizes the identified existing strategies for jointly segmenting coregistered images. In what follows, we provide some insights about each strategy. In [21], the authors introduce a multi-layer and multi-label Markov Random Field (MRF) model combining texture and color features for segmenting a single color image. The MRF involves three layers: one layer per feature (corresponding to individual segmentations) and one layer for their indirect combination (corresponding to joint segmentation). This model includes standard spatial relationships between pixels per individual segmentation as well as relationships between pixels of individual segmentations and the joint segmentation. The intra-layer relationships ensure the regularity within each individual segmentation whereas the inter-layer ones penalize any deviation between each individual segmentation and the joint one. Nevertheless, the energy term enforcing the consistency between individual segmentations is not well defined for some pixel pairs. Furthermore, in the original work, this MRF is suboptimally solved using Iterated Conditional Modes (ICM), while the same MRF is optimally solved using graph cuts (GC) in [4] but only for two labels and for a single image pair.

The idea of [21] has been included in a number of subsequent papers [5, 4, 17, 13]. In [17], the authors derive binary segmentations from a pair of 3D medical images using GC. A disagreement cost, inversely proportional to the intensity difference between images, is assigned when individual segmentations differ. Although the underlying energy function is proved to be globally minimized, the segmentation model requires a user interaction and the approach is limited to a single image pair.

In [13], a complex iterative energy minimization procedure that aims at removing the watermarks from a given set of coregistered images, includes a two-label joint segmentation. Joint watermark segmentation is achieved using the Iteratively Reweighted-Least-Square (IRLS) algorithm. Nevertheless, the approach is presented for only two labels and optimality guarantee about the solution is not discussed.

Finally, a conditional mixed Markov model is used in [5] for detecting the relevant changes in a coregistered image pair. The underlying energy is iteratively minimized using a Modified Metropolis Relaxation Algorithm (MMRA) [22]. However, this algorithm can po-

---

<sup>1</sup> Optimal for two labels and within a constant factor of the global minimum beyond.

**Table 1** Comparison of existing strategies for jointly segmenting coregistered images.

	[21]	[4]	[17]	[13]	[5]	This paper
Number of images	2	2	2	$\geq 2$	2	$\geq 2$
Number of labels	$\geq 2$	2	2	2	2	$\geq 2$
Minimization algorithm	ICM	GC	GC	IRLS	MMRA	GC
Optimality guarantees	None	Optimal	Optimal	None	None	Near-optimal <sup>1</sup>

tentially converge slowly and no optimality guarantee about the solution is given by the authors.

In addition to optimality properties (at least for particular cases), neither the sensitivity of the involved parameters, nor the robustness against degraded observations, nor the link connecting MLJS to standard approaches (such as NB) have been clearly discussed before in [21, 4, 17, 13, 5].

In this study, as contributions, we show that MLJS problem can be cast as a MRF model with coupling terms enforcing the consistency between individual segmentations and joint segmentation and we propose a generic formulation so that the proposed energy function can be efficiently minimized using graph cuts. This choice is motivated by their well-established theoretical background [23] as well as the existence of generic and efficient minimization algorithms [8, 9]. Notably, we demonstrate that this function can be globally minimized for two labels, whatever the number of images. We also study the links connecting MLJS to standard approaches. Finally, the benefit of such an approach is shown both on simulated and real images using distinct segmentation models.

The remaining of this paper is organized as follows. In Section 2, the energy function handling more than one label and more than one image is presented, particular cases are addressed, and graph-cut optimization is specified. Numerical experiments depicting the superiority of the approach in terms of segmentation accuracy and robustness against noisy observations are provided in Section 3. Presented results have been achieved for two kinds of data: (i) simulated images and (ii) real hyperspectral data that are classically exploited for material classification. Finally, the contributions of this paper and future work are provided in Section 4.

## 2 MLJS framework

### 2.1 Problem formulation

As previously highlighted, the problem of jointly segmenting a set of images can be elegantly cast as a MRF model. Coupling terms in the considered energy function allow us to enforce the consistency between individual segmentations and joint segmentation. Examples

of the considered relationships involved in the MRF are illustrated in Figure 1. Let us introduce some notations before presenting the considered energy function.

For positive integers  $K > 1$  and  $L > 1$ , we define  $\mathcal{K} = \{1, \dots, K\}$  and  $\mathcal{L} = \{1, \dots, L\}$  as the finite sets of image indexes and labels, respectively. For convenience, we enlarge the set  $\mathcal{K}$  by defining the set  $\tilde{\mathcal{K}} = \mathcal{K} \cup \{c\}$ , where  $c$  denotes the index related to the joint segmentation. In the proposed modeling, different domains and neighborhood systems are considered. Their definitions are gathered in Table 2. Using these notations, in this study we assume  $\overline{\mathcal{N}}_{a,b} \cap \tilde{\mathcal{N}}_{a,b} = \emptyset$ .

We distinguish matching and nearby pixel pairs because they will be provided with different weight parameters in the minimized energy. For simplicity, all these neighborhoods are assumed to be reciprocal, i.e.

$$\forall (a, b) \in (\tilde{\mathcal{K}} \times \tilde{\mathcal{K}}), \forall (p, q) \in (\mathcal{P}_a \times \mathcal{P}_b), \\ (p, q) \in \mathcal{N}_{a,b} \Leftrightarrow (q, p) \in \mathcal{N}_{b,a},$$

and

$$\forall k \in \mathcal{K}, \forall (p, q) \in \mathcal{N}_k, (p, q) \in \mathcal{N}_k \Leftrightarrow (q, p) \in \mathcal{N}_k.$$

As soon as the above neighborhood systems can be well-defined, the input images have not to be physically coregistered. Even more, some images with different resolutions could be considered, provided that we are able to define these neighborhood systems. Additionally, the above definition of neighborhoods with nearby pixel pairs is general and can for instance link spatially distant pixels [10]. However, the specific case where grids satisfy some constraints defined just below will allow for additional properties.

#### Definition 1 (Grid constraints)

The grid  $\mathcal{P}_c$  is *under-constrained* if for all  $k \in \mathcal{K}$  and for all  $p \in \mathcal{P}_k$ , there exists a unique  $r \in \mathcal{P}_c$  such that  $(p, r) \in \overline{\mathcal{N}}_{k,c}$ .

When  $\mathcal{P}_c$  is under-constrained, one pixel  $r \in \mathcal{P}_c$  might be connected to several pixels in  $\mathcal{P}_k$ . Also, some pixels of  $\mathcal{P}_c$  might not be connected at all (thus the name “under-constrained”).

Additionally to pixel domains and neighborhood systems, let us introduce segmentation field notations:  $u^c \in \mathcal{L}^{\mathcal{P}_c}$  is the joint segmentation defined over domain  $\mathcal{P}_c$ ,

Notation/definition	Description
$\mathcal{K} = \{1, \dots, K\}$	Finite set of image indexes
$\mathcal{L} = \{1, \dots, L\}$	Finite set of labels
$\tilde{\mathcal{K}} = \mathcal{K} \cup \{c\}$	Extended finite set of image indexes
$d > 0$	Dimension of image domains
$\mathcal{P}_k \subset \mathbb{Z}^d$	Domain of the $k^{\text{th}}$ image or individual segmentation, $\forall k \in \mathcal{K}$
$\mathcal{P}_c \subset \mathbb{Z}^d$	Domain of the joint segmentation
$\tilde{\mathcal{P}} = \cup_{k \in \tilde{\mathcal{K}}} \mathcal{P}_k$	Union of all domains
$\mathcal{N}_k \subseteq (\mathcal{P}_k \times \mathcal{P}_k)$	Neighborhood system composed of pixel pairs of the $k^{\text{th}}$ individual segmentation <sup>2</sup> , $\forall k \in \mathcal{K}$
$\bar{\mathcal{N}}_{a,b} \subseteq (\mathcal{P}_a \times \mathcal{P}_b)$	Matching pixel pairs between the domains $\mathcal{P}_a$ and $\mathcal{P}_b$ , for any $(a, b) \in (\tilde{\mathcal{K}} \times \tilde{\mathcal{K}})$ , s.t. $a \neq b$
$\tilde{\mathcal{N}}_{a,b} \subseteq (\mathcal{P}_a \times \mathcal{P}_b)$	Nearby pixel pairs between the domains $\mathcal{P}_a$ and $\mathcal{P}_b$ , for any $(a, b) \in (\tilde{\mathcal{K}} \times \tilde{\mathcal{K}})$ , s.t. $a \neq b$
$\mathcal{N}_{a,b} = \bar{\mathcal{N}}_{a,b} \cup \tilde{\mathcal{N}}_{a,b}$	Matching and nearby pixel pairs between the domains $\mathcal{P}_a$ and $\mathcal{P}_b$ , for any $(a, b) \in (\tilde{\mathcal{K}} \times \tilde{\mathcal{K}})$ , s.t. $a \neq b$

**Table 2** Summary of notations/definitions used in the paper.

and for every  $k \in \mathcal{K}$ ,  $u^k \in \mathcal{L}^{\mathcal{P}_k}$  is the individual segmentation of the  $k^{\text{th}}$  image defined over domain  $\mathcal{P}_k$  and  $u_p^k$  denotes the label of the segmentation  $u^k$  at pixel  $p$ .

Finally, we denote by  $(v^k)_{k \in \mathcal{K}}$  the finite set of images where any image  $v^k$  is defined over domain  $\mathcal{P}_k$ . Without loss of generality, we consider that all images have the same number of channels (denoted by the positive integer  $M > 0$ ). For every  $k \in \mathcal{K}$  and every  $p \in \mathcal{P}_k$ , we denote by  $v_p^k \in \mathbb{R}^M$  the values taken by the image  $v^k$  at the pixel  $p$ .

## 2.2 Proposed energy function

Given images  $(v^k)_{k \in \mathcal{K}}$  and neighborhood systems as defined in the previous section, we propose to minimize the following energy function with respect to  $(u^k)_{k \in \tilde{\mathcal{K}}} \in \mathcal{L}^{\tilde{\mathcal{P}}}$ :

$$E((u^k)_{k \in \tilde{\mathcal{K}}}) = \sum_{k \in \mathcal{K}} (E^k(u^k) + E^{k,c}(u^k, u^c)), \quad (1)$$

where

$$E^k(u^k) = \alpha_k \sum_{p \in \mathcal{P}_k} E_p^k(u_p^k) + \beta_k \sum_{(p,q) \in \mathcal{N}_k} E_{p,q}^k(u_p^k, u_q^k), \quad (2)$$

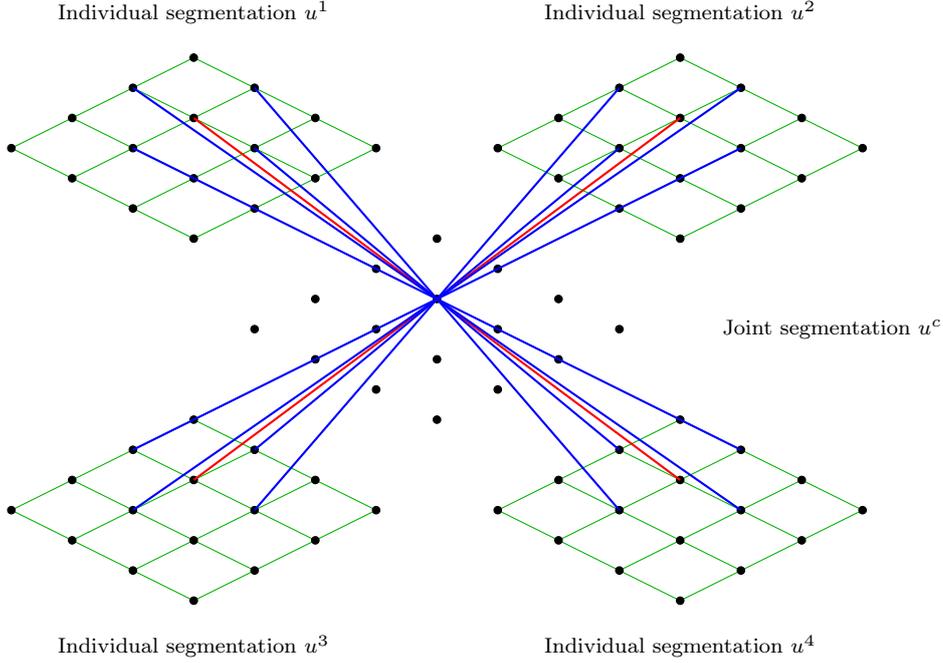
is the energy related to the  $k^{\text{th}}$  image, composed of data fidelity terms and spatial regularization terms whose relative weight is balanced by the weighting parameters  $(\alpha_k)_{k \in \mathcal{K}} \in \mathbb{R}_{\geq 0}^K$  and  $(\beta_k)_{k \in \mathcal{K}} \in \mathbb{R}_{\geq 0}^K$ . Data fidelity terms  $E_p^k$  usually depend on  $(v^k)_{k \in \mathcal{K}}$  and on the way the classes are modeled. The regularization terms  $E_{p,q}^k$  typically enforce some prior on the shape and contrast of the objects. Examples of these terms are provided in Section 3 and derived from [7].

For any  $k \in \mathcal{K}$ , the coupling terms

$$\begin{aligned} E^{k,c}(u^k, u^c) = & \gamma_k \left( \sum_{(p,r) \in \bar{\mathcal{N}}_{k,c}} E_{p,r}^{k,c}(u_p^k, u_r^c) \right. \\ & \left. + \sum_{(r,p) \in \bar{\mathcal{N}}_{c,k}} E_{r,p}^{c,k}(u_r^c, u_p^k) \right) \\ & + \eta_k \left( \sum_{(p,r) \in \tilde{\mathcal{N}}_{k,c}} E_{p,r}^{k,c}(u_p^k, u_r^c) \right. \\ & \left. + \sum_{(r,p) \in \tilde{\mathcal{N}}_{c,k}} E_{r,p}^{c,k}(u_r^c, u_p^k) \right), \end{aligned} \quad (3)$$

penalize discrepancies between individual segmentation  $u^k$  and joint segmentation  $u^c$ . The weighting parameters  $(\gamma_k)_{k \in \mathcal{K}} \in \mathbb{R}_{\geq 0}^K$  and  $(\eta_k)_{k \in \mathcal{K}} \in \mathbb{R}_{\geq 0}^K$  balance the coherence between individual segmentation  $u^k$  and joint segmentation  $u^c$ . In the case of a supervised approach, all weighting parameters are tuned by the user. Notice that, according to this model, no coupling terms linking individual segmentations  $(u^k)_{k \in \mathcal{K}}$  are considered although previously defined neighborhood systems allow us to involve such terms in Eq. (1):  $\forall (a, b) \in (\mathcal{K} \times \mathcal{K}), \mathcal{N}_{a,b} = \emptyset$ , as illustrated in Figure 1. Indeed, the most general formulation is at the expense of a high computational complexity. On the other hand, in the proposed formulation, the joint segmentation has a central role and coupling terms  $E^{k,c}(u^k, u^c)$  not only allow for a comparison between joint and individual segmentations but also an indirect comparison between individual segmentations.

Notice also that, in the same spirit of removing non-essential relationships, no spatial relationships are considered between pixels of the joint segmentation  $u^c$  since spatial regularization in  $u^c$  will be the natural consequence of spatial regularization in individual segmentations, that is controlled through  $E_{p,q}^k$ .



**Fig. 1** Example of MRF structure in the MLJS framework for  $K = 4$  images with spatial relationships in individual segmentations (green) as well as matched (red) and nearby (blue) relationships between individual segmentations and joint segmentation.

The spatial regularization terms in Eq. (2) are defined as

$$E_{p,q}^k(u_p^k, u_q^k) = w_{p,q}^k [u_p^k > u_q^k], \quad \forall k \in \mathcal{K}, \forall (p, q) \in \mathcal{N}_k, \quad (4)$$

where  $w_{p,q}^k \in \mathbb{R}_{>0}$  are given coefficients that do not depend on segmentations  $(u^k)_{k \in \tilde{\mathcal{K}}}$  and  $[\cdot]$  denotes the Iverson bracket, returning 1 if its argument is true and 0 otherwise. If we denote by  $\Omega_l^k \subseteq \mathcal{P}_k$  the set of pixels in the segmentation  $u^k$  labeled by  $l$  for any  $l \in \mathcal{L}$  and any  $k \in \mathcal{K}$ , then  $E_{p,q}^k$  is positive along the boundary of the region  $\Omega_l^k$  and null elsewhere. Notice that the above definition in Eq. (4) is general and allows us to have  $w_{p,q}^k \neq w_{q,p}^k$ . For the particular case where  $w_{p,q}^k = w_{q,p}^k$  for any  $k \in \mathcal{K}$  and for any pixel pair  $(p, q) \in \mathcal{N}_k$ , let us remark that

$$\begin{aligned} E_{p,q}^k(u_p^k, u_q^k) + E_{q,p}^k(u_q^k, u_p^k) &= w_{p,q}^k ([u_p^k > u_q^k] \\ &\quad + [u_q^k > u_p^k]) \\ &= w_{p,q}^k [u_p^k \neq u_q^k], \end{aligned} \quad (5)$$

Similarly, the coupling terms appearing in Eq. (3) can be written  $\forall k \in \mathcal{K}$

$$\begin{cases} E_{p,r}^{k,c}(u_p^k, u_r^c) = \Phi_{p,r}^{k,c} [u_p^k > u_r^c], & \forall (p, r) \in \mathcal{N}_{k,c}, \\ E_{r,p}^{c,k}(u_r^c, u_p^k) = \Phi_{r,p}^{c,k} [u_r^c > u_p^k], & \forall (r, p) \in \mathcal{N}_{c,k}, \end{cases} \quad (6)$$

where  $\Phi_{p,r}^{k,c} \in \mathbb{R}_{>0}$  are given coefficients that do not depend on segmentations  $(u^k)_{k \in \tilde{\mathcal{K}}}$ . Like for Eq. (5), for

any  $k \in \mathcal{K}$  and any pixel pair  $(p, r) \in \mathcal{N}_{k,c}$ , when  $\Phi_{p,r}^{k,c} = \Phi_{r,p}^{c,k}$ , one can sum Eq. (6) for opposite pixel pairs:

$$E_{p,r}^{k,c}(u_p^k, u_r^c) + E_{r,p}^{c,k}(u_r^c, u_p^k) = [u_p^k \neq u_r^c] \Phi_{p,r}^{k,c}. \quad (7)$$

Then, when Eq. (7) holds, Eq. (3) can be written:

$$\begin{aligned} E^{k,c}(u^k, u^c) &= \gamma_k \sum_{(p,r) \in \tilde{\mathcal{N}}_{k,c}} [u_p^k \neq u_r^c] \Phi_{p,r}^{k,c} \\ &\quad + \eta_k \sum_{(p,r) \in \tilde{\mathcal{N}}_{k,c}} [u_p^k \neq u_r^c] \Phi_{p,r}^{k,c}. \end{aligned} \quad (8)$$

When  $L = 2$ , we remind that an energy function is said to be submodular if all its pairwise terms verify a submodularity condition. For the spatial regularization terms (see Eq. (4)), this condition writes:

$$E_{p,q}^k(0, 0) + E_{p,q}^k(1, 1) \leq E_{p,q}^k(1, 0) + E_{p,q}^k(0, 1), \quad \forall k \in \mathcal{K}, \forall (p, q) \in \mathcal{N}_k. \quad (9)$$

Moreover, when all coupling terms are submodular, the energy  $E$  can be minimized exactly using graph cuts.

When  $L > 2$ , large scale problem cannot be minimized exactly. However, when the pairwise terms of the energy function, such as those defined in Eq. (4), satisfy

$$\begin{cases} E_{p,q}^k(l', l) + E_{p,q}^k(l, l'') \geq E_{p,q}^k(l', l''), & \forall l, l', l'' \in \mathcal{L}, \\ E_{p,q}^k(l, l') = 0 \Leftrightarrow l = l', & \forall l, l' \in \mathcal{L}, \\ E_{p,q}^k(l, l') = E_{p,q}^k(l', l) \geq 0, & \forall l, l' \in \mathcal{L}, \end{cases}$$

$$(10) \quad \text{For all } (l, l') \in \mathcal{L}^2$$

it still remains possible to get an approximate minimizer of the energy function Eq. (1) using the  $\alpha$ -expansion algorithm.

These numerical aspects and the details of the graph construction are described in Section 2.5.

### 2.3 Particular cases related to classic approaches

There are three cases where the proposed approach boils down to common segmentation strategies, namely: independent segmentations, NB segmentation (see Section 1) and majority decision [30].

The first case occurs when the weighting parameters  $(\gamma_k)_{k \in \mathcal{K}}$  and  $(\eta_k)_{k \in \mathcal{K}}$  are both null. In that situation, the coupling terms vanish and the value of  $E$  is independent of the joint segmentation. Minimizing Eq. (1) is thus equivalent to minimize Eq. (2) in  $u^k$ , independently for all  $k \in \mathcal{K}$ .

The second case establishes a link between our approach and NB segmentation. NB segmentation is a generic approach referring to the fact that the different layers of a multi-layer data set used to estimate a joint segmentation (and only it, i.e. no individual segmentations) are assumed independent conditionally to joint segmentation. In accordance with the notations defined in Section 2.2, it means that the functional to minimize with respect to  $u^c \in \mathcal{L}^{\mathcal{P}_c}$  can be written

$$\tilde{E}(u^c) = \sum_{r \in \mathcal{P}_c} \sum_{k \in \mathcal{K}} \left( \alpha_k \tilde{E}_r^k(u_r^c) \right) + \sum_{(r,s) \in \mathcal{N}_c} \tilde{E}_{r,s}^c(u_r^c, u_s^c), \quad (11)$$

where  $\mathcal{N}_c \subseteq (\mathcal{P}_c \times \mathcal{P}_c)$  is a neighborhood system,  $\tilde{E}_r^k(u_r^c)$  only depends on  $v_p^k$  (observations at pixels  $p$ , with the pixel pair  $(p, r) \in \mathcal{N}_{k,c}$ , of layer  $k$ ) and  $u_r^c$ . In the case of the assumed weighted Potts model (see e.g.[7,31]),  $\tilde{E}_{r,s}^c(u_r^c, u_s^c) = x_{r,s}[u_r^c \neq u_s^c]$  where  $x_{r,s} \in \mathbb{R}_{>0}$  are given coefficients.

In the following, having formalized the idea of a *consensual* solution (Definition 2), then Proposition 1 states the conditions under which our model boils down to NB one.

#### Definition 2

A solution  $(u^k)_{k \in \tilde{\mathcal{K}}}$  of Eq. (1) is said *consensual* if it verifies  $\forall r \in \mathcal{P}_c, \forall k \in \mathcal{K}$

$$u_p^k = u_r^c, \quad \forall p \in \mathcal{P}_k \text{ s.t. } (p, r) \in \overline{\mathcal{N}}_{k,c}.$$

In order to state the proposition, we define the following quantities, for  $k \in \mathcal{K}, (p, r) \in \overline{\mathcal{N}}_{k,c}$ :

$$R_{p,r}^k(l, l') = \frac{\alpha_k E_p^k(l) + \beta_k \sum_{q:(p,q) \in \mathcal{N}_k} E_{p,q}^k(l, l')}{\Phi_{p,r}^{k,c}},$$

$$m_k = \min_{\substack{(p,r) \in \overline{\mathcal{N}}_{k,c} \\ (l,l') \in \mathcal{L}^2}} R_{p,r}^k(l, l'),$$

and

$$M_k = \max_{\substack{(p,r) \in \overline{\mathcal{N}}_{k,c} \\ (l,l') \in \mathcal{L}^2}} \left( R_{p,r}^k(l, l') + \frac{\beta_k \sum_{q:(p,q) \in \mathcal{N}_k} E_{q,p}^k(l', l)}{\Phi_{p,r}^{k,c}} \right).$$

#### Proposition 1 (NB case)

For any set of images such that  $\mathcal{P}_c$  is under-constrained (see Definition 1), if we assume that for all  $k \in \mathcal{K}, \forall (p, r) \in \overline{\mathcal{N}}_{k,c}, \Phi_{p,r}^{k,c} = \Phi_{r,p}^{k,c} \in \mathbb{R}_{>0}$ , when

$$\begin{cases} \eta_k = 0 \\ \gamma_k > M_k - m_k \end{cases},$$

any minimizer of Eq. (1) is consensual (see Definition 2).

Additionally, for any consensual minimizer  $(u^k)_{k \in \tilde{\mathcal{K}}}$  of Eq. (1), the segmentation  $u^c$  solves an NB segmentation model, as defined in Eq. (11).

*Proof.* Since  $\eta_k = 0, \forall k \in \mathcal{K}$  and  $\Phi_{p,r}^{k,c} = \Phi_{r,p}^{k,c}, \forall k \in \mathcal{K}, \forall (p, r) \in \overline{\mathcal{N}}_{k,c}$ , using Eq. (8), we can rewrite Eq. (1) as

$$E((u^k)_{k \in \tilde{\mathcal{K}}}) = \sum_{k \in \mathcal{K}} \left( E^k(u^k) + \sum_{(p,r) \in \overline{\mathcal{N}}_{k,c}} \gamma_k \Phi_{p,r}^{k,c} [u_p^k \neq u_r^c] \right),$$

and obtain

$$E((u^k)_{k \in \tilde{\mathcal{K}}}) = \sum_{k \in \mathcal{K}} \sum_{p \in \mathcal{P}_k} \left( \alpha_k E_p^k(u_p^k) + \beta_k \sum_{q:(p,q) \in \mathcal{N}_k} E_{p,q}^k(u_p^k, u_q^k) + \gamma_k \Phi_{p,r}^{k,c} [u_p^k \neq u_r^c] \right). \quad (12)$$

Consider any minimizer  $(u^k)_{k \in \tilde{\mathcal{K}}}$  of Eq. (12). We now assume that  $(u^k)_{k \in \tilde{\mathcal{K}}}$  is not consensual and will show that this hypothesis leads to an impossible statement. In order to do so, we consider  $(\tilde{u}^k)_{k \in \tilde{\mathcal{K}}}$  the segmentations defined<sup>3</sup>, for  $\forall k \in \mathcal{K}, \forall r \in \mathcal{P}_r$ , by

$$\tilde{u}_p^k = u_r^c, \quad \forall p \in \mathcal{P}_k \text{ s.t. } (p, r) \in \overline{\mathcal{N}}_{k,c},$$

and

$$\tilde{u}_r^c = u_r^c, \quad \forall r \in \mathcal{P}_c.$$

We easily establish that  $\tilde{u}$  is consensual. In order to establish that  $E$  is smaller at  $\tilde{u}$  than at  $u$  we distinguish three configurations contributing to Eq. (12):

<sup>3</sup> Notice that the definition of  $\tilde{u}_p^k$  is correct because  $\mathcal{P}_c$  is under-constrained.

- Terms of the first kind: When  $k \in \mathcal{K}$  and  $p \in \mathcal{P}_k$  are such that  $u_p^k \neq \tilde{u}_p^k$ .
- Terms of the second kind: When  $k \in \mathcal{K}$  and  $p \in \mathcal{P}_k$  are such that  $u_p^k = \tilde{u}_p^k = u_r^c$  but there exists  $q$  such that  $(q, p) \in \mathcal{N}_k$  and  $u_q^k \neq \tilde{u}_q^k$ .
- Terms of the third kind: When  $k \in \mathcal{K}$  and  $p \in \mathcal{P}_k$  are such that  $u_p^k = \tilde{u}_p^k = u_r^c$  and for all  $q$  such that  $(q, p) \in \mathcal{N}_k$ ,  $u_q^k = \tilde{u}_q^k$ .

These configurations cover all the possible situations. Notice that the first configuration appears at least once, since  $u$  is not consensual.

- Assume  $k \in \mathcal{K}$  and  $p \in \mathcal{P}_k$  are such that  $u_p^k \neq \tilde{u}_p^k$ . Since, the grid  $\mathcal{P}_c$  is under-constrained, there is a unique  $r \in \mathcal{P}_c$  such that  $(p, r) \in \mathcal{N}_{k,c}$  and we have

$$u_p^k \neq \tilde{u}_p^k = \tilde{u}_r^c = u_r^c.$$

As a consequence, using the definition of  $M_k$  and since  $[\tilde{u}_p^k \neq \tilde{u}_r^c] = 0$ ,

$$\begin{aligned} \alpha_k E_p^k(\tilde{u}_p^k) + \beta_k \sum_{q:(p,q) \in \mathcal{N}_k} E_{p,q}^k(\tilde{u}_p^k, \tilde{u}_q^k) \\ + \gamma_k \Phi_{p,r}^{k,c}[\tilde{u}_p^k \neq \tilde{u}_r^c] \\ + \beta_k \sum_{\substack{q:(q,p) \in \mathcal{N}_k \\ u_q^k = \tilde{u}_q^k}} E_{q,p}^k(\tilde{u}_q^k, \tilde{u}_p^k) \leq M_k \Phi_{p,r}^{k,c}, \end{aligned} \quad (13)$$

where the last term in the sum on the left of the inequality is here to take into account that, when computing  $E((\tilde{u}^k)_{k \in \tilde{\mathcal{K}}})$ , a term  $E_{q,p}^k(\tilde{u}_q^k, \tilde{u}_p^k)$  is present when adding the pixel  $q$ , and  $q$  might be of the second kind. Because of the assumption on  $\gamma_k$ , we have

$$\begin{aligned} M_k \Phi_{p,r}^{k,c} &< m_k \Phi_{p,r}^{k,c} + \gamma_k \Phi_{p,r}^{k,c} \\ &\leq \alpha_k E_p^k(u_p^k) + \beta_k \sum_{q:(p,q) \in \mathcal{N}_k} E_{p,q}^k(u_p^k, u_q^k) \\ &\quad + \gamma_k \Phi_{p,r}^{k,c}[u_p^k \neq u_r^c]. \end{aligned}$$

- Assume  $k \in \mathcal{K}$  and  $p \in \mathcal{P}_k$  are such that  $u_p^k = \tilde{u}_p^k = u_r^c$  but there exists  $q$  such that  $(q, p) \in \mathcal{N}_k$  and  $u_q^k \neq \tilde{u}_q^k$ . In that case, every term  $\beta_k E_{p,q}^k(\tilde{u}_p^k, \tilde{u}_q^k)$  is compensated by the last term in sum in the left of the inequality Eq. (13), in the corresponding term of the first kind<sup>4</sup>. The terms  $\beta_k E_{p,q}^k(u_p^k, u_q^k)$  are non-negative and can be added on the right of the inequality. The other terms are identical for  $\tilde{u}$  and  $u$ .
- Assume  $k \in \mathcal{K}$  and  $p \in \mathcal{P}_k$  are such that  $u_p^k = \tilde{u}_p^k = u_r^c$  and for all  $q$  such that  $(q, p) \in \mathcal{N}_k$ ,  $u_q^k = \tilde{u}_q^k$ . All terms are identical for  $u$  and  $\tilde{u}$ .

<sup>4</sup> In Eq. (13), the name of the index are inverted and the term has the form  $\beta_k E_{q,p}^k(\tilde{u}_q^k, \tilde{u}_p^k)$ .

We finally obtain, by summing over all the  $k, p$

$$E((\tilde{u}^k)_{k \in \tilde{\mathcal{K}}}) < E((u^k)_{k \in \tilde{\mathcal{K}}}),$$

that is in contradiction with the assumption that  $u$  is a minimizer of Eq. (12). Therefore, the assumption that  $(u^k)_{k \in \tilde{\mathcal{K}}}$  is not consensual is not valid. In order to prove the second statement, we consider a consensual minimizer  $(u^k)_{k \in \tilde{\mathcal{K}}}$  of

$$\begin{aligned} E((u^k)_{k \in \tilde{\mathcal{K}}}) = \sum_{k \in \mathcal{K}} \left( \alpha_k \sum_{p \in \mathcal{P}_k} E_p^k(u_p^k) \right. \\ \left. + \beta_k \sum_{(p,q) \in \mathcal{N}_k} E_{p,q}^k(u_p^k, u_q^k) \right). \end{aligned}$$

Since the solution is consensual and grids are under-constrained,  $\forall p \in \mathcal{P}_k$ , there exists a unique  $r$  such that  $(p, r) \in \mathcal{N}_{k,c}$ ,  $u_p^k = u_r^c$ . As a consequence,  $u_p^k$  in the sum on  $p \in \mathcal{P}_k$  can be replaced by  $u_r^c$  in a sum on  $r \in \mathcal{P}_c$  and we have

$$\sum_{k \in \mathcal{K}} \alpha_k \sum_{p \in \mathcal{P}_k} E_p^k(u_p^k) = \sum_{r \in \mathcal{P}_c} \sum_{k \in \mathcal{K}} \alpha_k \sum_{p:(r,p) \in \mathcal{N}_{k,c}} E_p^k(u_r^c).$$

The latter term has the right format, once we let

$$\tilde{E}_r^k(l) = \sum_{p:(r,p) \in \mathcal{N}_{k,c}} E_p^k(l), \quad \forall l \in \mathcal{L}.$$

Similarly, we have

$$\begin{aligned} \sum_{k \in \mathcal{K}} \beta_k \sum_{(p,q) \in \mathcal{N}_k} E_{p,q}^k(u_p^k, u_q^k) \\ = \sum_{(r,s) \in \mathcal{N}_c} \sum_{k \in \mathcal{K}} \beta_k \sum_{\substack{(p,q) \in \mathcal{N}_k \\ (p,r) \in \mathcal{N}_{k,c} \\ (q,s) \in \mathcal{N}_{k,c}}} E_{p,q}^k(u_r^c, u_s^c), \end{aligned}$$

where we have defined

$$\mathcal{N}_c = \{(r, s) \in \mathcal{P}_c^2 \mid \text{there exists } k \in \mathcal{K} \text{ and } (p, q) \in \mathcal{N}_k \text{ such that } (p, r) \in \mathcal{N}_{k,c} \text{ and } (q, s) \in \mathcal{N}_{k,c}\}.$$

Again, the binary term has the right format once we let

$$\tilde{E}_{r,s}^k(l, l') = \sum_{k \in \mathcal{K}} \beta_k \sum_{\substack{(p,q) \in \mathcal{N}_k \\ (p,r) \in \mathcal{N}_{k,c} \\ (q,s) \in \mathcal{N}_{k,c}}} E_{p,q}^k(l, l'), \quad \forall l, l' \in \mathcal{L}.$$

□

The third case establishes a link between our approach and majority decision. In order to state it, we define for all  $k \in \mathcal{K}$  and all  $r \in \mathcal{P}_c$

$$f_r^k(l, u^k) = \#\{p \in \mathcal{P}_k \mid (r, p) \in \mathcal{N}_{k,c}, u_p^k = l\}$$

which counts the neighbors of  $r$  having the label  $l$  in the individual segmentation  $u^k$  ( $\sharp$  denotes the cardinality of a set) and

$$f_r(l, (u^k)_{k \in \mathcal{K}}) = \sum_{k \in \mathcal{K}} f_r^k(l, u^k)$$

which counts the neighbors of  $r$  having the label  $l$  among all individual segmentations  $(u^k)_{k \in \mathcal{K}}$ . We also remind that the majority decision is defined for all  $r \in \mathcal{P}_c$  by

$$\operatorname{argmax}_{l \in \mathcal{L}} f_r(l, (\hat{u}^k)_{k \in \mathcal{K}}).$$

**Proposition 2** (Majority decision case)

For any set of images associated to neighborhood systems  $(\mathcal{N}_k)_{k \in \mathcal{K}}$  and  $(\mathcal{N}_{k,j})_{(k,j) \in (\tilde{\mathcal{K}} \times \tilde{\mathcal{K}})}$ , let us assume that

$$\begin{cases} \Phi_{p,r}^{k,c} = \Phi_{r,p}^{c,k} = C' \in \mathbb{R}_{>0}, \forall k \in \mathcal{K}, \forall (p,r) \in \mathcal{N}_{k,c} \\ \gamma_k = \eta_k = C' \in \mathbb{R}_{>0}, \forall k \in \mathcal{K}. \end{cases}$$

Then, any minimizer  $(\hat{u}^k)_{k \in \tilde{\mathcal{K}}}$  of Eq. (1) verifies

$$\hat{u}_r^c \in \operatorname{argmax}_{l \in \mathcal{L}} f_r(l, (\hat{u}^k)_{k \in \mathcal{K}}), \quad \forall r \in \mathcal{P}_c.$$

*Proof.* Since  $\Phi_{p,r}^{k,c} = \Phi_{r,p}^{c,k} = C' \in \mathbb{R}_{>0}, \forall k \in \mathcal{K}, \forall (p,r) \in \mathcal{N}_{k,c}$  and  $\gamma_k = \eta_k = C' \in \mathbb{R}_{>0}, \forall k \in \mathcal{K}$  and using Eq. (8), Eq. (1) can be written

$$\begin{aligned} E((u^k)_{k \in \tilde{\mathcal{K}}}) &= \sum_{k \in \mathcal{K}} \left( E^k(u^k) + C' \sum_{(p,r) \in \tilde{\mathcal{N}}_{k,c}} [u_p^k \neq u_r^c] C \right. \\ &\quad \left. + C' \sum_{(p,r) \in \tilde{\mathcal{N}}_{k,c}} [u_p^k \neq u_r^c] C \right) \\ &= \sum_{k \in \mathcal{K}} E^k(u^k) \\ &\quad + CC' \sum_{k \in \mathcal{K}} \sum_{(p,r) \in \mathcal{N}_{k,c}} [u_p^k \neq u_r^c]. \end{aligned} \quad (14)$$

Since  $CC' > 0$ , for any minimizer  $(\hat{u}^k)_{k \in \mathcal{K}}$  and any  $r \in \mathcal{P}_c$ ,  $\hat{u}_r^c$  minimizes in Eq. (14) (with respect to  $l \in \mathcal{L}$ )

$$\begin{aligned} e_r(l) &= \sum_{k \in \mathcal{K}} \sum_{(p,r) \in \mathcal{N}_{k,c}} [\hat{u}_p^k \neq l] = \sum_{k \in \mathcal{K}} \sum_{\substack{l' \in \mathcal{L} \\ l' \neq l}} f_r^k(l', \hat{u}^k), \\ &= C_r - f_r(l', (\hat{u}^k)_{k \in \mathcal{K}}), \end{aligned}$$

where

$$\begin{aligned} C_r &= \cup_{k \in \mathcal{K}} \{p \in \mathcal{P}_k \mid (p,r) \in \mathcal{N}_{k,c}\}, \\ C_r &= \sharp \mathcal{C}_r = \sum_{k \in \mathcal{K}} \sum_{l \in \mathcal{L}} f_r(l, (\hat{u}^k)_{k \in \mathcal{K}}), \end{aligned} \quad (15)$$

counts all the neighbors of  $r$ . Therefore,  $\hat{u}_r^c$  maximizes  $f_r$ .  $\square$

For any pixel  $r \in \mathcal{P}_c$ , the majority decision is taken among the pixels  $\mathcal{C}_r$  (see Eq. (15)). Then, it may occur that the number  $\sharp \mathcal{C}_r$  of pixels involved in the majority rule is larger than  $K$ . Finally notice that, to be a majority label, a label  $l \in \mathcal{L}$  shall satisfy (see Proposition 2):  $f_r(l, (u^k)_{k \in \mathcal{K}}) \geq \frac{\sharp \mathcal{C}_r}{2}$  if only two labels are actually present in  $\mathcal{C}_r$  or  $\geq \max\{\frac{\sharp \mathcal{C}_r}{L}, 1\}$  if labels are uniformly distributed in  $\mathcal{C}_r$ .

In this section, we have established some links with classic approaches. Beyond the illustration of the behavior of the proposed energy function (in some borderline cases), it allows us to test the use of some heuristics. Specifically, when  $L > 2$ , we use the  $\alpha$ -expansion heuristic to compute an approximate minimizer of Eq. (1) (see Section 2.5). The experiments of Section 3 performed in this third particular case allowed us to check that, in practice, each pixel of the resulting joint segmentation obtained by  $\alpha$ -expansion algorithm was actually assigned to a majority label (as predicted by Proposition 2).

## 2.4 Strategy for defining weighting coefficients

In previous section, we consider cases at the margin with particular values assigned to weighting parameters  $\gamma_k, \eta_k$  and coefficients  $\Phi_{p,r}^{k,c}$ . In this section, we consider standard cases of the energy function Eq. (1) and we discuss about the coefficients  $w_{p,q}^k$  and  $\Phi_{p,r}^{k,c}$ .

Classically,  $w_{p,q}^k$  coefficients help to better align the boundary of a segmentation onto boundaries of the objects of interest. Let us consider a pair of spatially neighboring pixels  $(p,q) \in \mathcal{N}_k$ , for some  $k \in \mathcal{K}$ , where labels  $u_p^k$  and  $u_q^k$  are expected to be different. When image intensities at  $p$  and  $q$  are dissimilar,  $w_{p,q}^k$  coefficients typically assign a small penalty. Conversely, when image intensities are similar, the cost for having  $u_p^k$  different from  $u_q^k$  is large. Examples of such coefficients include gradient magnitude, Laplacian zero-crossing, etc [29].

In the same spirit,  $\Phi_{p,r}^{k,c}$  coefficients aim to enforce coherence of all segmentations, for pixels that are in different images but at the matched locations ( $\tilde{\mathcal{N}}$  links) or at a nearby locations ( $\tilde{\mathcal{N}}$  links). In our case, the coupling between individual segmentations is only done indirectly through the joint segmentation. Thus,  $\Phi_{p,r}^{k,c}$  coefficients cannot be defined from the same dissimilarity measurements as the ones used for  $w_{p,q}^k$  coefficients since no observations are associated to the joint segmentation  $u^c$ .

Let us first consider the simplest case where the number of images  $K = 2$ ,  $\tilde{\mathcal{N}}_{a,b} \neq \emptyset$  and  $\tilde{\mathcal{N}}_{a,b} = \emptyset$ . For any triplet of matching pixels  $(r,p,q) \in (\mathcal{P}_c \times \mathcal{P}_1 \times \mathcal{P}_2)$ ,

$(r, p) \in \bar{\mathcal{N}}_{c,1}$ ,  $(r, q) \in \bar{\mathcal{N}}_{c,2}$ , it seems reasonable that, when image intensities/features are similar in pixels  $p$  and  $q$ ,  $u_p^1$ ,  $u_q^2$  and  $u_r^c$  are expected to be equal and thus we will choose a large penalty when it is not the case. Conversely, the observation of dissimilar image intensities/features (in  $p$  and  $q$ ) can explain a difference in individual segmentation labels  $u_p^1$  and  $u_q^2$ , which cannot thus both be consistent with the joint segmentation  $u_r^c$ . Thus,  $\Phi_{p,r}^{1,c}$  and  $\Phi_{q,r}^{2,c}$  positive coefficients can be driven by  $v_p^1$  and  $v_q^2$  similarity.

In a more general way, i.e. for  $K \geq 2$ , we propose to define  $\Phi_{p,r}^{k,c}$  coefficients based on the (dis)similarity of the observation in image  $v^k$  at pixel  $p$  with respect to observations in other images  $(v^j)_{j \in (\mathcal{K} \setminus \{k\})}$ , restricted to pixels  $q$  matched to pixel  $r$  ( $q \in \mathcal{P}_j$ ,  $(r, q) \in \bar{\mathcal{N}}_{c,j}$ ).

Let us define  $\varphi_{p,q,r}^{k,j,c} \in \mathbb{R}_{>0}$  a coefficient that measures the similarity between observations in pixels  $p \in \mathcal{P}_k$  and  $q \in \mathcal{P}_j$  where  $p$  and  $q$  are indirectly linked through  $r \in \mathcal{P}_c$ . To define  $\Phi_{p,r}^{k,c}$ , the simplest strategy is to sum  $\varphi_{p,q,r}^{k,j,c}$  coefficients over all pixels  $q$  with respect to  $r$  neighborhood relationships:

$$\begin{cases} \forall k \in \mathcal{K}, \forall (p, r) \in \mathcal{N}_{k,c}, \Phi_{p,r}^{k,c} = \sum_{\substack{j \in \mathcal{K}, j \neq k \\ (r, q) \in \bar{\mathcal{N}}_{c,j}}} \varphi_{p,q,r}^{k,j,c}, \\ \forall k \in \mathcal{K}, \forall (p, r) \notin \mathcal{N}_{k,c}, \Phi_{p,r}^{k,c} = 0. \end{cases}$$

Notice that the latter expression allows us to have  $\varphi_{p,q,r}^{k,j,c} \neq \varphi_{q,p,r}^{j,k,c}$ . The averaging strategy (of  $\varphi_{p,q,r}^{k,j,c}$  coefficients) is the one retained for the experiments presented in Section 3. However, we underline that other combination rules of coefficients  $\varphi_{p,q,r}^{k,j,c}$  could be chosen accordingly to the application.

## 2.5 Energy minimization

Graph cuts are a combinatorial optimization problem of graph theory and refer to minimum-cut (or, by duality, maximum-flow). Introduced in the 80's as a new energy minimization tool for binary image denoising, it has received more attention a decade later thanks to better computational resources, the incoming of a fast maximum-flow algorithm [8] and key theoretical results [23].

Noticeably, the latter paper shows that the pairwise terms of an energy function need to be submodular to be minimized exactly. The same condition is also sufficient to minimize this function using graph cuts. The interest of such an approach is that a minimum-cut can be computed rapidly using a fast algorithm [8], with a near-linear empirical complexity of the number of nodes. When  $L = 2$  and the submodularity condition Eq. (9) holds for any pairwise term involved in Eq. (1), the graph cut algorithm provides an exact minimizer of Eq. (1).

Let us first consider the case where  $L = 2$  and then the case where  $L > 2$ . For doing so, let us consider a directed and capacitated graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  with a set of nodes defined as

$$\mathcal{V} = \tilde{\mathcal{P}} \cup \{s, t\},$$

where  $s$  and  $t$  are particular nodes usually named the source and the sink respectively, and a set of edges defined as

$$\begin{aligned} \mathcal{E} = (\cup_{k \in \mathcal{K}} (\{s\} \times \mathcal{P}_k) \cup (\mathcal{P}_k \times \{t\}) \\ \cup \mathcal{N}_k \cup \mathcal{N}_{k,c} \cup \mathcal{N}_{c,k}) \subset (\mathcal{V} \times \mathcal{V}). \end{aligned}$$

Edges connected to  $s$  or  $t$  are usually named t-links while edges in  $\mathcal{N}_k$ ,  $\mathcal{N}_{k,c}$  and  $\mathcal{N}_{c,k}$ , for every  $k \in \mathcal{K}$ , are named n-links. For any edge  $(p, q) \in \mathcal{E}$ , we assign to it a non-negative capacity as follows

$$\begin{cases} c(s, p) = \alpha_k E_p^k(1), \forall k \in \mathcal{K}, \forall p \in \mathcal{P}_k, \\ c(p, t) = \alpha_k E_p^k(0), \forall k \in \mathcal{K}, \forall p \in \mathcal{P}_k, \\ c(p, q) = \beta_k w_{p,q}^k, \forall k \in \mathcal{K}, \forall (p, q) \in \mathcal{N}_k, \\ c(p, r) = \gamma_k \Phi_{p,r}^{k,c}, \forall k \in \mathcal{K}, \forall (p, r) \in \bar{\mathcal{N}}_{k,c}, \\ c(r, p) = \gamma_k \Phi_{r,p}^{c,k}, \forall k \in \mathcal{K}, \forall (r, p) \in \bar{\mathcal{N}}_{c,k}, \\ c(p, r) = \eta_k \Phi_{p,r}^{k,c}, \forall k \in \mathcal{K}, \forall (p, r) \in \tilde{\mathcal{N}}_{k,c}, \\ c(r, p) = \eta_k \Phi_{r,p}^{c,k}, \forall k \in \mathcal{K}, \forall (r, p) \in \tilde{\mathcal{N}}_{c,k}. \end{cases} \quad (16)$$

Let us remind that an s-t cut  $\mathcal{C}$  of the graph  $\mathcal{G}$  is a partition of the set of nodes  $\mathcal{V}$  into two disjoint sets  $\mathcal{S}$  and  $\mathcal{T}$ , such that  $s \in \mathcal{S}$  and  $t \in \mathcal{T}$ . The value of any s-t cut  $\mathcal{C}$  is given by

$$\text{val}_{\mathcal{G}}(\mathcal{C}) = \sum_{(p,q) \in (\mathcal{S} \times \mathcal{T})} c(p, q).$$

For an s-t cut  $\mathcal{C}$ , we also define

$$x_p^k(\mathcal{C}) = \begin{cases} 0 & \text{if } p \in \mathcal{T} \\ 1 & \text{if } p \in \mathcal{S} \end{cases}, \quad \forall p \in \mathcal{P}_k, \quad \forall k \in \tilde{\mathcal{K}}.$$

From the latter expression, one can remark that, for every  $k \in \tilde{\mathcal{K}}$ , there is a one-to-one correspondence between the set of s-t cuts and the elements of  $\{0, 1\}^{\mathcal{P}_k}$ . For every  $k \in \mathcal{K}$  and every  $p \in \mathcal{P}_k$ , the t-links involving  $p$  contribute to  $\text{val}_{\mathcal{G}}(\mathcal{C})$  with

$$x_p^k(\mathcal{C}) E_p^k(0) + (1 - x_p^k(\mathcal{C})) E_p^k(1). \quad (17)$$

Furthermore, for every  $k \in \mathcal{K}$  and any n-link  $(p, q) \in \mathcal{N}_k$ , we have

$$(p, q) \in (\mathcal{S} \times \mathcal{T}) \quad \text{if and only if} \quad [x_p^k(\mathcal{C}) > x_q^k(\mathcal{C})] = 1. \quad (18)$$

Similarly, for every  $k \in \tilde{\mathcal{K}}$  and any n-link  $(p, r) \in \mathcal{N}_{k,c}$ , we have

$$\begin{cases} (p, r) \in (\mathcal{S} \times \mathcal{T}) & \text{if and only if } [x_p^k(\mathcal{C}) > x_r^c(\mathcal{C})] = 1, \\ (r, p) \in (\mathcal{S} \times \mathcal{T}) & \text{if and only if } [x_r^c(\mathcal{C}) > x_p^k(\mathcal{C})] = 1. \end{cases}$$

(19)

Given the capacities of the graph  $\mathcal{G}$  in Eq. (16) and by summing Eq. (17), Eq. (18) and Eq. (19), it now becomes straightforward to check that, for any  $s$ - $t$  cut  $\mathcal{C}$  in the graph  $\mathcal{G}$ , we have

$$\begin{aligned}
\text{val}_{\mathcal{G}}(\mathcal{C}) &= \sum_{k \in \mathcal{K}} \left( \alpha_k \sum_{p \in \mathcal{P}_k} x_p^k(\mathcal{C}) E_p^k(0) + (1 - x_p^k(\mathcal{C})) E_p^k(1) \right. \\
&+ \beta_k \sum_{(p,q) \in \mathcal{N}_k} w_{p,q}^k [x_p^k(\mathcal{C}) > x_q^k(\mathcal{C})] \\
&+ \gamma_k \sum_{(p,r) \in \tilde{\mathcal{N}}_{k,c}} \Phi_{p,r}^{k,c} [x_p^k(\mathcal{C}) > x_r^c(\mathcal{C})] + \Phi_{r,p}^{c,k} [x_r^c(\mathcal{C}) > x_p^k(\mathcal{C})] \\
&\left. + \eta_k \sum_{(p,r) \in \tilde{\mathcal{N}}_{k,c}} \Phi_{p,r}^{k,c} [x_p^k(\mathcal{C}) > x_r^c(\mathcal{C})] + \Phi_{r,p}^{c,k} [x_r^c(\mathcal{C}) > x_p^k(\mathcal{C})] \right) \\
&= \sum_{k \in \mathcal{K}} \left( \sum_{p \in \mathcal{P}_k} \alpha_k E_p^k(x_p^k(\mathcal{C})) + \sum_{(p,q) \in \mathcal{N}_k} \beta_k E_{p,q}^k(x_p^k(\mathcal{C}), x_q^k(\mathcal{C})) \right. \\
&+ \gamma_k \left( \sum_{(p,r) \in \tilde{\mathcal{N}}_{k,c}} E_{p,r}^{k,c}(x_p^k(\mathcal{C}), x_r^c(\mathcal{C})) + \sum_{(r,p) \in \tilde{\mathcal{N}}_{c,k}} E_{r,p}^{c,k}(x_r^c(\mathcal{C}), x_p^k(\mathcal{C})) \right) \\
&\left. + \eta_k \left( \sum_{(p,r) \in \tilde{\mathcal{N}}_{k,c}} E_{p,r}^{k,c}(x_p^k(\mathcal{C}), x_r^c(\mathcal{C})) + \sum_{(r,p) \in \tilde{\mathcal{N}}_{c,k}} E_{r,p}^{c,k}(x_r^c(\mathcal{C}), x_p^k(\mathcal{C})) \right) \right) \\
&= E((x^k(\mathcal{C}))_{k \in \tilde{\mathcal{K}}}).
\end{aligned}$$

Thus, if  $\mathcal{C}^*$  is a  $s$ - $t$  minimum-cut in the graph  $\mathcal{G}$ , then  $(x^k(\mathcal{C}^*))_{k \in \tilde{\mathcal{K}}}$  minimizes Eq. (1). When  $L > 2$ , as all pairwise terms satisfy the axioms of a metric in the labels space  $\mathcal{L}$  (see Eq. (10)), a minimizer within a constant factor of the global one can thus be efficiently computed using graph cuts as well [9].

Although the memory usage only scales linearly with the number of images  $K$ , situations where the graph  $\mathcal{G}$  do not fit in memory can be encountered (especially when  $K$  and/or the size of images  $(v^k)_{k \in \mathcal{K}}$  become large). Since segmentations remain weakly connected between each other, dual decomposition can be used to overcome this problem [36]. Optimality on the solution is still guaranteed for  $L = 2$  and a local minimum is guaranteed within the same known factor as previously for  $L > 2$ .

### 3 Experimental results

The purpose of this section is to illustrate the performance of the MLJS framework against the NB approach on 2D (hence  $d = 2$ ) simulated data (with  $L = 2$ ) and real data (with  $L > 2$ ). More particularly, two variants of the proposed approach are considered, differing by the considered neighborhood systems for the coupling

terms. For the first variant (denoted by MLJS-M), only matched pixel pairs between segmentations  $(u^k)_{k \in \mathcal{K}}$  are considered (weighting parameters  $(\eta_k)_{k \in \mathcal{K}}$  are null, or  $\forall k \in \mathcal{K}, \tilde{\mathcal{N}}_{k,c} = \emptyset$ ). For the second variant (denoted by MLJS-N), neighborhood systems  $(\mathcal{N}_{k,c})_{k \in \mathcal{K}}$  include matched pairs as well as 8-nearest pixel pairs between the joint segmentation  $u^c$  and the individual segmentations  $(u^k)_{k \in \mathcal{K}}$  (weighting parameters  $(\gamma_k)_{k \in \mathcal{K}}, (\eta_k)_{k \in \mathcal{K}}$  are positive). 8-nearest pixel pairs are used for modeling spatial relationships in individual segmentations  $(u^k)_{k \in \mathcal{K}}$ . For any  $k \in \mathcal{K}$  and any pixel pair  $(p, q) \in \mathcal{N}_k$ , coefficients  $w_{p,q}^k$  will be equal for opposite pixel pairs for simplicity, i.e.  $w_{p,q}^k = w_{q,p}^k$ .

The performance of any segmentation strategy producing a segmentation  $SG : \mathcal{P}_c \rightarrow \mathcal{L}$ , is evaluated with respect to an available ground truth  $GT : \mathcal{P}_c \rightarrow \mathcal{L}$  using two common metrics, i.e. accuracy (in %) and F-measure (in %)

$$ACC(SG, GT) = \frac{100}{\#\mathcal{P}_c} \sum_{p \in \mathcal{P}_c} [SG(p) = GT(p)],$$

and F-measure (in %)

$$\begin{aligned}
FM(SG, GT) &= \\
&100 \sum_{l \in \tilde{\mathcal{L}}} \left( \frac{2 \sum_{p \in \mathcal{P}_c} [SG(p) = l] \times [GT(p) = l]}{\sum_{p \in \mathcal{P}_c} [SG(p) = l] + [GT(p) = l]} \right),
\end{aligned}$$

where  $\tilde{\mathcal{L}}$  denotes the set of available labels in the ground truth  $GT$ . The larger the two above metrics are, the better the considered segmentation strategy is.

For each segmentation strategy, multiple values of the weighting parameters involved in the energy function Eq. (1) are explored using an exhaustive grid search strategy. The values of the weighting parameters maximizing either  $ACC$  or  $FM$  are then retained. Nevertheless, for being tractable in a reasonable amount of time, these parameters are constrained as follows (when available):  $\beta_k = \beta \in \mathbb{R}_{\geq 0}$  and  $\gamma_k = \eta_k = \gamma \in \mathbb{R}_{\geq 0}$ ,  $\forall k \in \mathcal{K}$ . Hence, the number of weighting parameters is drastically reduced but this simplification remains reasonable with regards to the input data considered in the subsequent sections <sup>5</sup>.

#### 3.1 Toy example

##### 3.1.1 Simulated data

The set  $(v^k)_{k \in \mathcal{K}}$  is composed of 8 coregistered images of two types. The first type consists of replicated images

<sup>5</sup> Notice that this paper focuses on the supervised approach. Estimation of hyperparameters will be the subject of a forthcoming paper.

from a  $64 \times 64$  low-contrasted simulated image composed of homogeneous areas and thin structures. The second type consists of distinct realizations of white Gaussian noise of standard deviation  $\sigma \in ]0, 1]$ , added to the simulated image. The ratio of degraded images with respect to the number of images  $K$  is controlled by a parameter denoted by  $\rho \in [0, 1]$ . Also, the intensities of all images lie in the unit interval  $[0, 1]$ . The simulated image as well as degraded versions of it are depicted in Figure 2.

### 3.1.2 Modeling

To minimize the energy Eq. (1) presented in Section 2, we first need to instantiate appropriately the data terms and the regularization terms (see Eq. (6) and Eq. (4)). Given the input images, using a Chan-Vese segmentation model [11] where image intensities are assumed to be Gaussian-distributed in any region, is consistent with our simulations. For every  $k \in \mathcal{K}$  and every  $l \in \mathcal{L}$ , the data term of this model measures the discrepancy of region  $\Omega_l^k$  with

$$E_p^k(u_p^k) = \frac{1}{2\sigma^2} \|v_p^k - \mu_{k,u_p^k}\|_2^2,$$

where  $\mu_{k,l} \in \mathbb{R}^M$  denotes mean intensity of region  $\Omega_l^k$ , while the spatial regularization terms measure the length of the boundary of  $\Omega_l^k$  with

$$E_{p,q}^k(u_p^k, u_q^k) = \frac{1}{\|\nu \odot (p - q)\|_2} H(u_p^k - u_q^k),$$

where  $\nu \in \mathbb{R}_{>0}^2$  denotes the image resolution and  $\odot$  denotes the Hadamard product between two elements of  $\mathbb{R}^2$ . By identification with the proposed model, given the parameter  $\nu$ , we thus have  $w_{p,q}^k = \frac{1}{\|\nu \odot (p - q)\|_2}$ . Moreover, for any triplet  $(p, q, r) \in (\mathcal{P}_k \times \mathcal{P}_j \times \mathcal{P}_c)$  such that  $(p, r) \in \mathcal{N}_{k,c}$  and  $(q, r) \in \mathcal{N}_{j,c}$  and  $(p, q)$  is a matching pixel pair, the coefficients involved in Eq. (6) are defined as

$$\varphi_{p,q,r}^{k,j,c} = \frac{1}{1 + \|\nu \odot (p - r)\|_2} \exp\left(-\frac{\|v_p^k - v_q^j\|_2^2}{2\sigma_{k,j}^2}\right), \quad (20)$$

where  $\sigma_{k,j} \in \mathbb{R}_{>0}$  denotes the standard deviation of the difference of intensities of the images  $v^k$  and  $v^j$ . In the latter equation, notice that the denominator remains positive when the pixels  $p$  and  $r$  share the same spatial location.

### 3.1.3 Numerical experiments

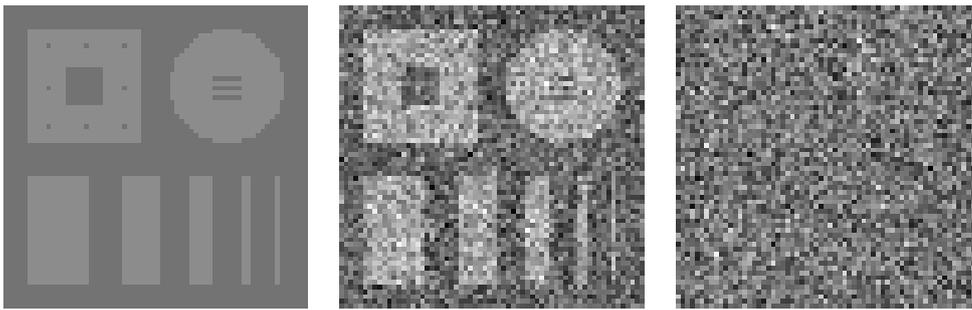
Given a noise level  $\sigma$  and a ratio of degraded images  $\rho$ , each segmentation strategy is run with weighting parameters taking values in the following ranges:  $\beta \in$

$\{2^{-13}, 2^{-12}, \dots, 2^0\}$  and  $\gamma \in \{2.5^{-13}, 2.5^{-12}, \dots, 2.5^0\}$ . This process is repeated 10 times with different noise realizations. Performance metrics are then averaged over these runs.

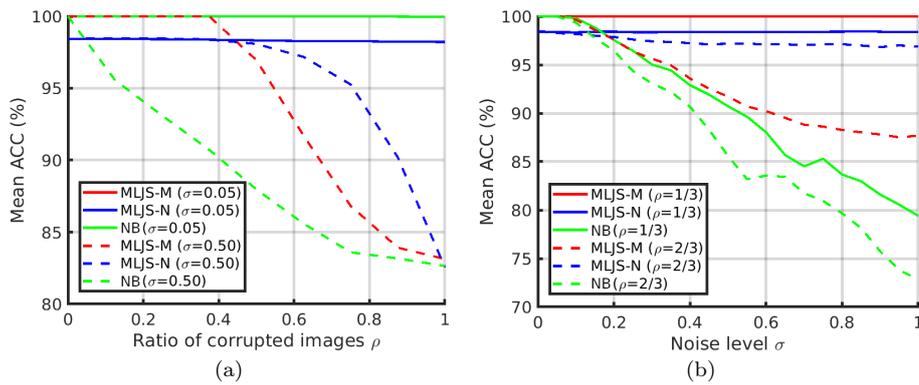
The results are presented in Figure 3, 4 and 5.

Quantitative results are only provided in terms of *ACC* since they are very well correlated with *FM*. Note that in Figure 3(a), since the achieved mean accuracy of MLJS-M and MLJS-N is the same, the solid green line and solid red line overlap. According to Figure 3, when the noise level  $\sigma$  remains small with respect to the image dynamic (i.e. up to  $\sigma \simeq 0.1$ ), whatever the ratio of degraded images  $\rho$ , all segmentation strategies perform equally with an *ACC* of about 100%. Notice that MLJS-N perform slightly worse in such situation on thin structures since they are more impacted by the spatial regularization. As the noise level  $\sigma$  increases, the performance of all segmentation strategies diminishes in a way that is even more pronounced when the ratio of degraded images  $\rho$  becomes large. However, the performance of MLJS-M and MLJS-N decreases less rapidly than NB. In these experiments, whatever the proportion of degraded images and whatever the noise level, MLJS-M outperforms NB (see Section 2.3). Using the proposed approach, the objects are well recovered while less than half of the images are degraded. In such a case, unlike NB, MLJS-M and MLJS-N can indeed still rely on non-degraded images to produce the desired segmentation. Furthermore, MLJS-N appears to be more robust than MLJS-M. When all images are degraded, all segmentation strategies however perform equally. The aforementioned observations are confirmed by the segmentation results depicted in Figure 4. In this figure, for  $\sigma = 0.5$ , false positives (cyan) and false negatives (magenta) of resulting segmentations with respect to ground truth are superimposed to the simulated image for a moderate ( $\rho = 1/3$ ) and strong ( $\rho = 2/3$ ) proportion of degraded images.

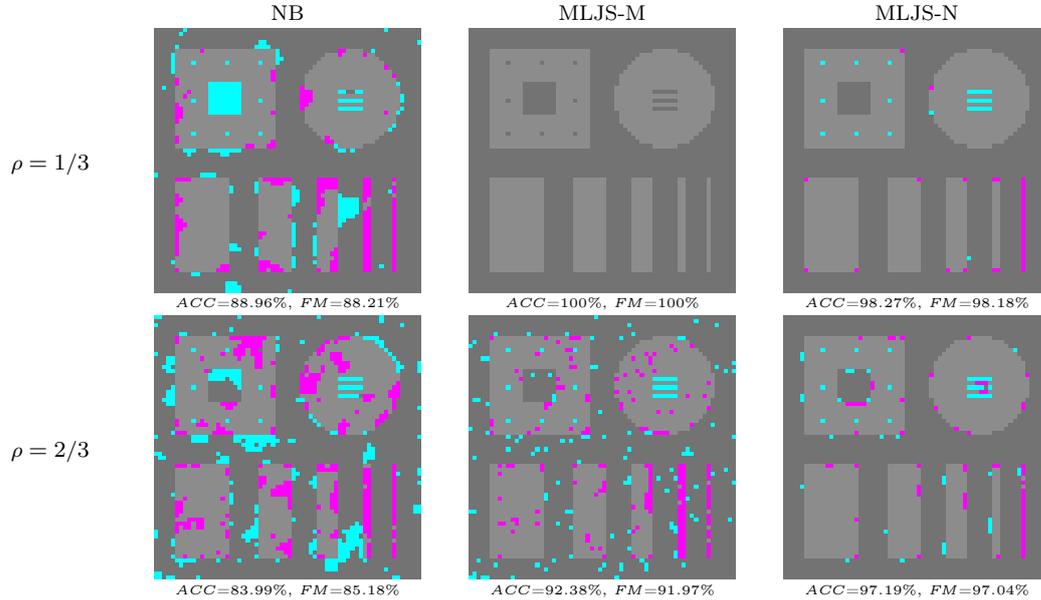
Finally, Figure 5 depicts the sensitivity of the weighting parameters  $\beta$  and  $\gamma$  of all segmentation strategies for  $\rho = 0.5$ . One can observe that the size of the intervals of the weighting parameters  $\beta$  and  $\gamma$  offering good segmentation performance shrinks as the noise level increases. Nevertheless, the location of the maxima of *ACC* remains stable. As expected, the segmentation performance diminishes for extreme values of the weighting parameters.



**Fig. 2** Simulated image (left) degraded by an additive white Gaussian noise of standard deviation  $\sigma = 0.05$  (middle) and  $\sigma = 0.5$  (right).



**Fig. 3** Mean *ACC* against ratio of degraded images  $\rho \in [0, 1]$  (a) and amount of noise  $\sigma \in [0, 1]$  (b). Red solid line is occluded by green solid line since the two curves overlap.



**Fig. 4** False positives (cyan) and false negatives (magenta) with respect to ground truth superimposed to the simulated image for  $\sigma = 0.5$ . Mean *ACC* and *FM* are provided below each image.

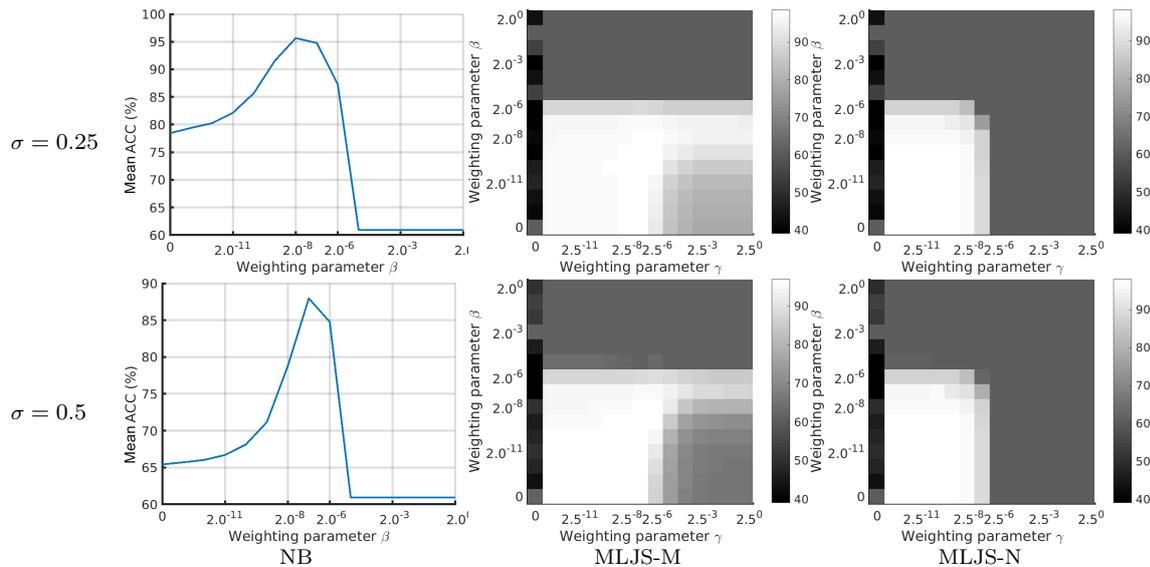


Fig. 5 Sensitivity of weighting parameters  $\beta$  and  $\gamma$  on a toy example for  $\rho = 0.5$ .

## 3.2 Materials classification

### 3.2.1 Context

The purpose of this section is to illustrate the performance of the MLJS framework on real data, i.e. for identifying the material of different objects. This problem is challenging for two reasons: (i) we cannot rely on the shape, the location and the orientation of the objects since they are independent of object material; (ii) we can neither rely on the appearance since, in some cases, objects made of different materials may be very close in terms of appearance.

Nowadays, like for many problems in computer vision, learning-based approaches, such as Deep Neural Networks (DNN) overcome classical approaches. The performance of deep networks type methods comes from huge databases used to train them. In the case of color images, a known example is the database acquired by Google to operate their TensorFlow network [1], the image database ImageNet [14] or more specifically, the recent plastic waste image database [6]. However, even if results are promising, in case of material classification, DNN are faced to an intrinsic limitation, namely the low discriminating power of color. Even more, when during the learning step, for a given material a color is not represented but it appears during the operational step (e.g., a new color for water plastic bottle for a special event), the classifier will fail.

Hyperspectral imaging in the near infra red field brings a suitable solution for this problem since it can provide relevant information about the chemical components of materials. Using DNN on hyperspectral data

does not yet appear as an efficient solution nowadays for at least two reasons. Firstly, hyperspectral data are not so common than visible data, especially when we take into account the sensor dependency; then, even if in the field of remote sensing, some free hyperspectral databases have appeared, in the field of materials, only very small databases exist, each one being related to a specific sensor. Secondly, in such a few-shot learning, the huge size of the input data jeopardizes the training efficiency.

Recently, material type recognition was cast as a multiclass classification problem in an evidential unified framework. The multiclass problem is first decomposed in binary subproblems (dichotomizers, e.g. binary Support Vector Machines) and then partial outputs combined together [25]. One-versus-all (OVA) and one-versus-one are particular cases of such a strategy also known as ECOC (Error Correcting Output Codes) e.g. [15,24]. However, such approaches are blind and do not benefit of the spatial neighborhood information which may provide a very useful prior as in most image processing problems. Thus, we aim to take advantage of such information using the proposed approach.

### 3.2.2 Data

The Headwall hyperspectral sensor used for our tests acquires images where every pixel stores the spectrum sampled on exactly 275 wavelengths between 900nm and 2500nm. Now, raw measurements in these spectrums are usually not considered as they are directly. Classical preprocessing involves spectrum filtering and derivation at different orders, e.g. based on the Savitsky-

Golay filter [34,37]. This filter fits a low degree polynomial on data within a sliding window with a fixed size. It allows us to smooth the data and to compute the derivatives from the fitted polynomials. Then, because of both some processing time considerations and data redundancy control, dimensionality reduction is applied, for instance using a Principal Component Analysis (PCA), e.g. [12].

Specifically, in addition to the performance assessment of the proposed approach, we are interested by evaluating if the whole spectrum is useful or if the information contained in the half spectrum is sufficient. Indeed, acquiring information far from the visible field is increasingly expensive. Establishing if fewer wavelengths are sufficient could lead to the use of cheaper sensors. For doing so, we split the spectrum into two parts: from 900nm to 1700nm (denoted by  $fh$ ) and from 1700nm to 2500nm (denoted by  $sh$ ). This split roughly corresponds to classical sensor ranges and allows us to obtain two equal parts of the spectrum. Then, Savitsky-Golay’s filter having been applied to each half spectrum part, a PCA is applied to each output (derivative) of Savitsky-Golay’s filter, with an adaptive number of kept components preserving 99% of the information, namely between 3 to about 20 in our experiments. In the following, the input data resulting of PCA are denoted  $D_k^s$ , where the subscript  $k \in \{0, 1, 2\}$  denotes the derivation order and the superscript  $s \in \{fh, sh\}$  denotes the considered spectrum part. The absence of superscript means that the whole spectrum is used.

In this study, we consider  $L = 9$  classes of materials, namely 7 different kinds of polymer (corresponding to different plastics present in packaging: water, milk, shampoo, plastic film, etc.) and 2 fibrous classes (paper and cardboard). From hyperspectral images, we extract a data subset, specifically used for SVM training. From 1000 samples per class using Gaussian kernels, it allows for the estimation of each dichotomizer parameters, determined by 5 fold cross validation and grid search. In this work, we focus on the OVA strategy because of its linear cost (versus the number of classes) while providing good performance when correctly trained [32]. For  $L$  labels,  $L$  dichotomizers are trained to distinguish any class against the other ones (hence the strategy name). In our case, we thus trained  $K \times L$  dichotomizers so that for a given  $k \in \mathcal{K}$ , for every pixel  $p \in \mathcal{P}_k$ , the dichotomizer trained to distinguish class  $l$  in  $(v^k)_{k \in \mathcal{K}}$  data provides as raw output a score (denoted by  $s_p^k(l) \in \mathbb{R}$ ) that may be converted into a soft output with possible probabilistic interpretation. Among the different ways to estimate likelihoods with this strategy, the most intuitive one is to map the score  $s_p^k(l)$  to the unit interval

**Table 3** Boards (left) and classes (right) in hyperspectral cosegmentation.

Board	Image size	Classes
1	141 × 96	2, 8, 9
2	141 × 96	2, 8, 9
3	141 × 96	4, 5, 9
4	141 × 96	4, 5, 9
5	141 × 96	1, 3, 7, 9
6	141 × 96	1, 3, 6, 7, 9
7	203 × 96	4, 5, 8, 9

Class	Material
1	Paper
2	Cardboard
3	Elastomers
4	Polyethylene Terephthalate
5	Opaque Polyethylene Terephthalate
6	Polythene
7	Polypropylene
8	Polystyrene
9	Conveyor belt

with (see [28])

$$v_p^k(l) = \frac{g_{\theta_l^k}(s_p^k(l))}{\sum_{m \in \mathcal{L}} g_{\theta_m^k}(s_p^k(m))} \in ]0, 1[, \quad (21)$$

where  $v_p^k(l)$  denotes the  $l^{th}$  component of a  $M$ -dimensional vector storing the normalized class-likelihood or *mapped scores*, the denominator is a normalization factor and  $g_{\theta_j^i}$  is the sigmoid function defined by

$$g_{\theta_j^i}(x) = \frac{1}{1 + \exp((x, 1)^T \theta_j^i)} \in ]0, 1[,$$

where  $\theta_j^i \in (\mathbb{R}_{<0} \times \mathbb{R})$  are regression coefficients for the  $i^{th}$  image and the  $j^{th}$  label. Regression coefficients  $\theta_j^i$  are estimated on calibration sets. Each calibration set includes 200 samples: 100 labelled  $j$  according  $GT$  and 100 labelled “not  $j$ ”. The estimation of regression coefficients is performed by minimizing the resulting likelihood function with respect to  $\theta_j^i$  (that is a convex function), using the BFGS algorithm [16]. In Eq. (21), we intentionally use the notation  $v_p^k$  since normalized likelihood images are precisely our input images for MLJS.

The set of samples used for the experiments presented in next sections have been collected in the Veolia laboratories using a starter-kit hyperspectral sensor configured for lab experiments with halogen lamps and a 30cm large linear stage. They consist of 7 specimen boards with small material samples with an horizontal and a vertical resolution of 1.25 and 1.6 pixels, respectively. Hence, the image resolution is set as  $\nu = (1.25, 1.6)^T$ . The 7<sup>th</sup> board of this dataset presents real objects stacked on top of each others to provide more realistic conditions. Characteristics of the boards are summarized in Table 3.

### 3.2.3 Modeling

To minimize the energy Eq. (1) presented in Section 2, we need to instantiate the data terms and the regularization terms (see Eq. (3) and Eq. (4)). For every  $k \in \mathcal{K}$  and every pixel  $p \in \mathcal{P}_k$ , we propose to define the data terms as

$$E_p^k(u_p^k) = -\log(v_p^k(u_p^k)), \\ = \log\left(1 + \exp\left(\left(s_p^k(u_p^k), 1\right)^T \theta_{u_p^k}^k\right)\right).$$

For every  $k \in \mathcal{K}$  and any pixel pair  $(p, q) \in \mathcal{N}_k$ , we propose to define the spatial regularization terms with

$$E_{p,q}^k(u_p^k, u_q^k) = \frac{1}{\|\nu \odot (p - q)\|_2} \\ \times \exp\left(-\frac{\|v_p^k - v_q^k\|_2^2}{2\sigma_k^2}\right) H(u_p^k - u_q^k),$$

where we remind that  $\nu \in \mathbb{R}_{>0}^2$  denotes the image resolution of any image and  $\odot$  denotes the Hadamard product between two vectors of  $\mathbb{R}^2$ . Taking into account the score vector distance in the weighting coefficient (in addition to the spatial distance already used in Chan-Vese model) allows us to detect automatically object boundaries and not to penalize a solution with different labels assigned to neighboring pixels in such situation.

Finally, the coefficients involved in coupling terms are the same as those defined by Eq. (20) (see Section 3.1.2).

### 3.2.4 Numerical experiments

The two variants of the proposed approach are compared to NB as well as to independent segmentations of coregistered images (obtained by canceling coupling terms with  $\beta = 0$ ). Different datasets derived from the spectrum are investigated: the first half of the spectrum, the second half of the spectrum, the concatenation of split spectrums and the non-split spectrum. The robustness of the considered segmentation approaches is also studied when mapped scores are degraded with an impulsive noise. In this case, the metrics *ACC* and *FM* are averaged over 3 noise realizations.

For NB, the weighting parameter  $\beta$  takes values in the following range:  $\{1.5^{-16}, 1.5^{-15}, \dots, 1.5^{17}\}$ . For MLJS-M, the same interval is used for both the weighting parameters  $\beta$  and  $\gamma$ . For MLJS-N, the range of values of these two parameters need to be slightly adapted by using  $\{1.6^{-23}, 1.6^{-22}, \dots, 1.6^{10}\}$ .

The results of the experiments are presented in Figure 6, 7, 8, 9 and Table 5. For each split spectrum, *ACC* and *FM* measurements are compared between the considered segmentation approaches in the Table 5. For 5

images out of 7, MLJS outperforms other approaches. Both the mean and the overall *ACC* and *FM* are also in favor of MLJS. Best results (with respect to *ACC*) for boards 5 and 7 are shown in Figure 8 for NB, MLJS-M, MLJS-N and best independent segmentation (here  $D_1^{sh}$ ). Although MLJS offers a slightly lower performance for the board 7 compared to  $D_1^{sh}$  for one of the two segmentation metrics, it clearly outperforms NB for this complex board. In both boards, MLJS-M and MLJS-N perform equally well.

In Figure 6, mean *ACC* is provided with respect to noise level for different spectrum data either half spectrum or whole spectrum considered at once or split. Consistently with other works (see [25]), it can be observed that the derivatives of order 1 provides best results, compared to those of order 0 and 2, and remains very competitive against NB. All segmentation approaches perform equally in the absence of noise, whatever which part(s) of the spectrum is handled. However, MLJS-M and MLJS-N appear to be more robust against noise when the spectrum is split.

A rather interesting point is the difference of behavior between non-split spectrum (Figure 6(c)) and split spectrum (Figure 6(d)). Indeed, conversely to half spectrum cases, the whole information is considered in both cases but processed differently. It points out that increasing artificially the number of images  $K$  in the proposed approach allows for an increase of the data redundancy and thus better filtering of noise.

Best achieved results (with respect to *ACC*) for boards 2 and 7 are shown in Figure 9 for NB, MLJS-M, MLJS-N and best independent segmentation (here  $D_1^{sh}$ ). These results definitely confirm that MLJS-M, and especially MLJS-N, are more robust against noise than NB.

Additionally, the segmentation performance (in terms of *ACC*) of the weighting parameters  $\beta$  and  $\gamma$  is provided with and without noise for NB, MLJS-M and MLJS-N in Figure 7. As the noise density increases, it can be noticed that the weighting parameters remain stable.

### 3.3 Computational aspects

The running time of our approach is affected by the max-flow/min-cut algorithm and increases with the number of labels  $L$ , the number of images  $K$ , the density of neighborhood systems  $(\mathcal{N}_k)_{k \in \mathcal{K}}$  and  $(\mathcal{N}_{k,j})_{(k,j) \in (\bar{\mathcal{K}} \times \bar{\mathcal{K}})}$  as well as the amount of regularization (controlled by the weighting parameters  $(\beta_k)_{k \in \mathcal{K}}$ ,  $(\gamma_k)_{k \in \mathcal{K}}$  and  $(\eta_k)_{k \in \mathcal{K}}$ ). We remind that the theoretical complexity of the max-flow/min-cut algorithm [8] used is  $O(mn^2 \text{val}_{\mathcal{G}}(C^*))$ , where

	MLJS-M	MLJS-N
Non-split spectrum ( $K = 3$ )	$2.55 \pm 0.55$	$6.56 \pm 2.73$
Split spectrum ( $K = 6$ )	$6.86 \pm 1.48$	$18.91 \pm 3.66$

**Table 4** Comparison of running times (in secs) between MLJS-M and MLJS-N for  $L = 9$  using non-degraded non-split spectrum and split spectrum. The couple of values of the weighting parameters  $(\beta_k)_{k \in \mathcal{K}}$  and  $(\gamma_k)_{k \in \mathcal{K}}$  maximizing overall ACC is selected (see Table 5).

$n$  and  $m$  denote respectively the number of nodes and edges in the graph  $\mathcal{G}$  and  $\text{val}_{\mathcal{G}}(\mathcal{C}^*)$  denotes the value of the min-cut  $\mathcal{C}^*$ . In the proposed approach,  $n = \#P_c \times (K + 1)$  (assuming a constant number of pixels  $\#P_c$  per image) and  $m = \#P_c \times K(a + b + 2)$  with  $a$  and  $b$  two constant values. Specifically, for each node (pixel in a given image),  $a$  denotes its number of n-links (spatial neighbors) while  $b$  denotes the number of links towards the consensus image (e.g. in Figure 1,  $a = 4$  and  $b = 5$ ).

In addition to this worst case theoretical complexity evaluation, we conduct an empirical study. In Table 4, we provide running times from of a under-optimized implementation of our approach. Our experiments were run under MATLAB<sup>6</sup> on an Intel i7-6900K CPU @ 3.20GHz with 16 cores and 64GB of RAM.

These times (in secs) are compared between MLJS-M and MLJS-N using non-degraded non-split spectrum ( $K = 3$  images) and a split spectrum ( $K = 6$  images) still for  $L = 9$  (see Section 3.2.2). These measurements are averaged over 10 runs and the couple of values of the weighting parameters  $(\beta_k)_{k \in \mathcal{K}}$  and  $(\gamma_k)_{k \in \mathcal{K}}$  maximizing overall ACC is retained (see Table 5). For both MLJS-M and MLJS-N, as the number of images doubles (from  $K = 3$  to  $K = 6$  in Table 5), the observed running times are multiplied by a factor between 2 and 3. For a fixed number of images  $K$ , as the number of considered relationships surrounding consensus segmentation  $u^c$  grows from MLJS-M to MLJS-N (from  $b = 1$  to  $b = 9$ ), the observed running time is again multiplied by a factor between 2 and 3.

## 4 Conclusion and future work

In this paper, we have considered the problem of jointly segmenting objects of interest from a set of images. Our contributions are the following. Firstly, we propose a rather generic model (called MLJS) so that both it generalizes previous models proposed for 2 labels and/or two images, and it allows us to recover particular cases rather intuitive and simple. In particular, we demonstrate the links connecting MLJS to standard

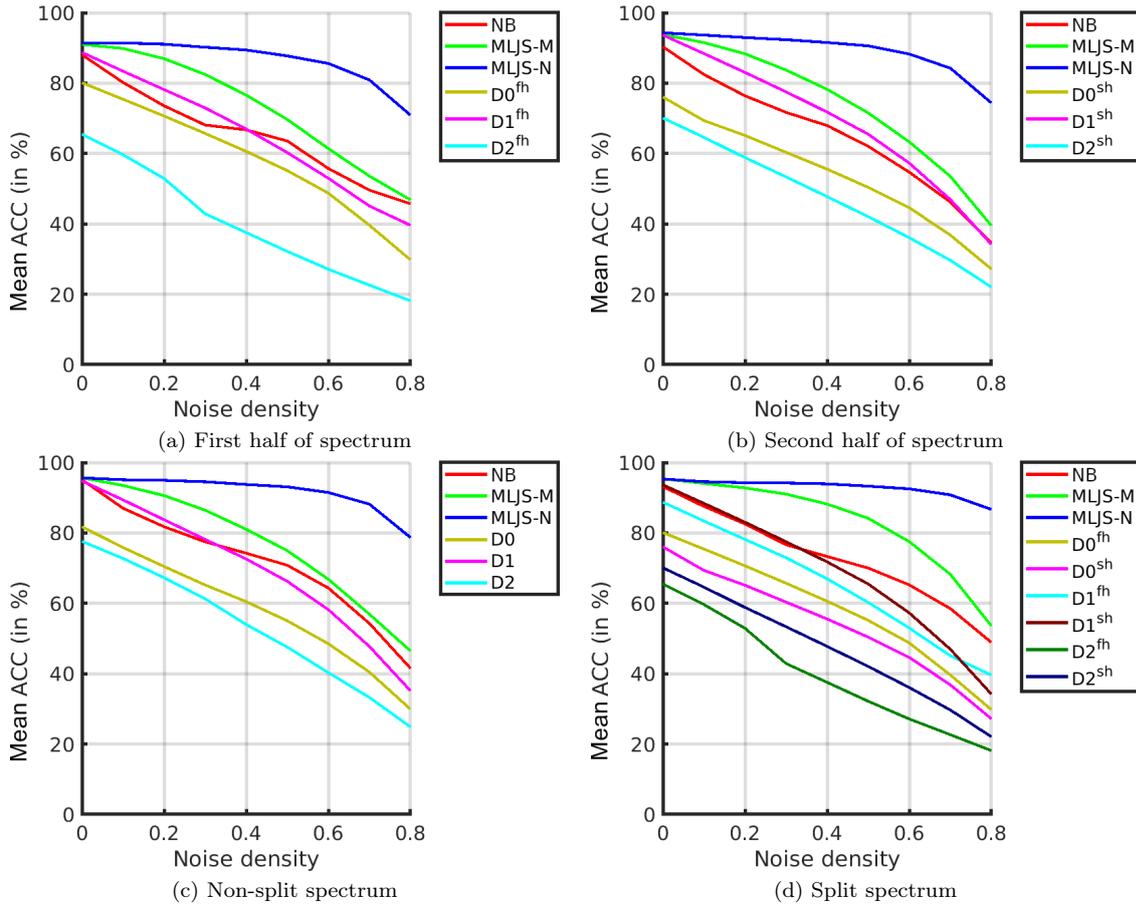
approaches (such as NB, majority decision and independent segmentations). To illustrate the genericity of the approach, we instantiate the energy function of MLJS into two different ways in the our experiments. Secondly, we demonstrate that this energy function can be globally minimized for two labels, whatever the number of images, and we experiment that graph-cut approach also allows for efficient minimization in the case of more than two labels. Thirdly, we conduct experiments to study the sensitivity of the results versus the MLJS parameters, and the robustness against degraded observations. The benefit of our approach has then been demonstrated both on simulated and real images using distinct segmentation models.

For future work, we plan to extend our approach in two directions. Firstly, we will consider Bayesian estimation techniques in order to jointly estimate the model parameters (class features and hyperparameters) in addition to the segmentations (individual and joint ones). Secondly, we will consider the case where coregistration is possible but unknown. We aim at estimating the mappings between images during the energy minimization since geometric consistency is a requirement to segmentation regularization.

## References

1. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. Tensorflow: A system for large-scale machine learning. In *Symposium on Operating Systems Design and Implementation (SOSDI)*, pages 265–283, 2016.
2. G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, volume 147 of *Applied Mathematical Sciences*. Springer US, 2002.
3. D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen. Interactively co-segmenting topically related images with intelligent scribble guidance. *International Journal of Computer Vision*, 93(3):273–292, 2011.
4. C. Benedek, M. Shadaydeh, Z. Kato, T. Sziranyi, and J. Zerubia. Multi-layer MRF models for change detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 107:22–37, 2015.
5. C. Benedek and T. Sziranyi. Change detection in optical aerial images by a multilayer conditional mixed Markov model. *IEEE Transactions on Geoscience and Remote Sensing*, 47(10):3416–3430, 2009.
6. J. Bobulski and J. Piatkowski. PET waste classification method and plastic waste DataBase-WaDaBa. In *International Conference on Image Processing and Communications (ICIC)*, pages 57–64, 2017.
7. Y. Boykov and G. Funka-Lea. Graph cuts and efficient N-D image segmentation. *International Journal of Computer Vision*, 70(2):109–131, 2006.
8. Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 26(9):1124–1137, 2004.

<sup>6</sup> The implementation used for the  $\alpha$ -expansion algorithm is publicly available at <https://github.com/akanazawa/MRF/tree/master/utis/gco-v3.0/matlab>.

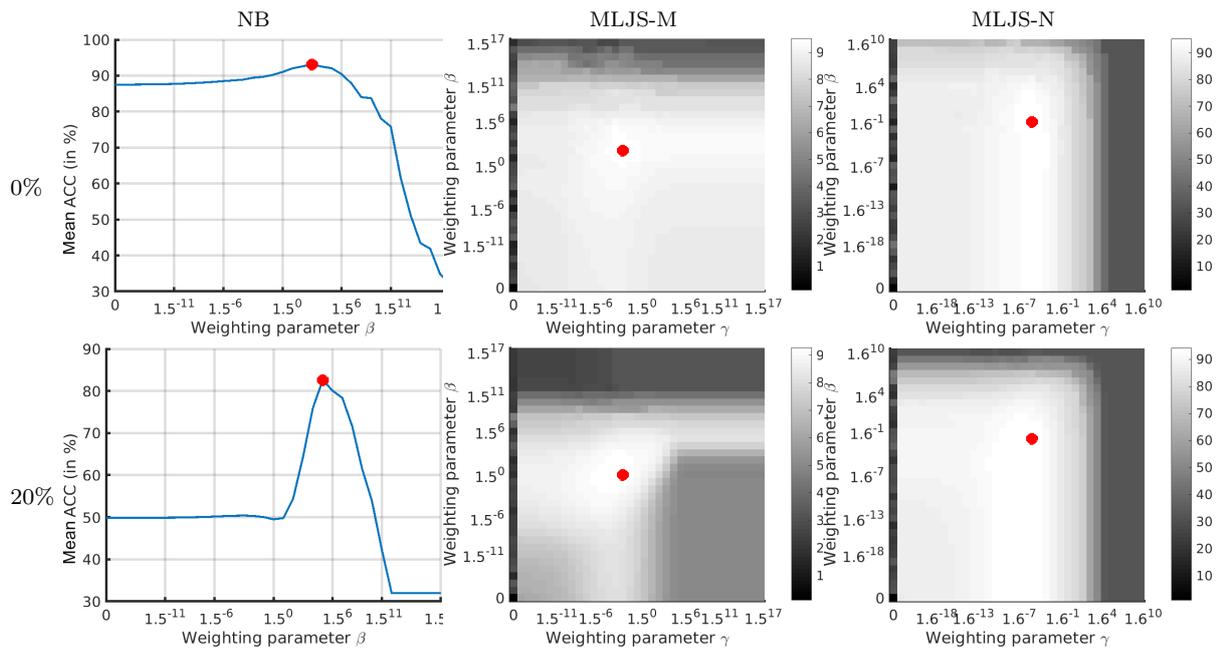


**Fig. 6** Mean ACC against corrupted mapped scores with impulsive noise for different spectrum data.

**Table 5** Comparison (in %) of ACC (top) and FM (bottom) between (NB), the variants of the MLJS framework (MLJS-M and MLJS-N) and independent segmentations using a split spectrum without noise. Best measurements are emphasized in bold.

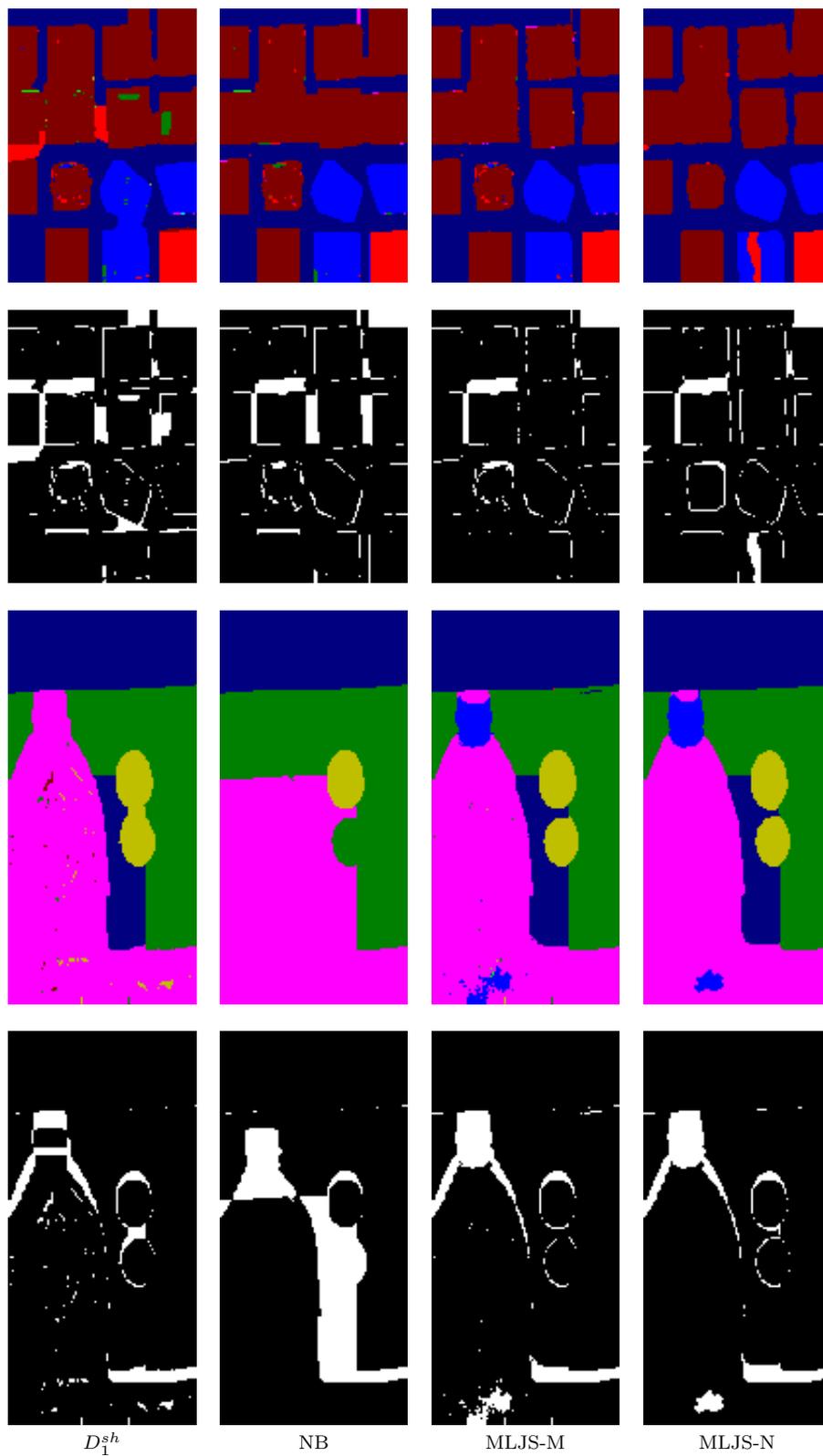
	$D_0^{fh}$	$D_0^{sh}$	$D_1^{fh}$	$D_1^{sh}$	$D_2^{fh}$	$D_2^{sh}$	NB	MLJS-M	MLJS-N
1	94.92	83.16	99.34	99.05	79.06	88.11	99.34	<b>99.36</b>	99.11
2	90.90	90.21	99.22	99.48	67.16	75.90	99.65	<b>99.72</b>	99.33
3	96.03	87.80	91.94	95.66	50.66	80.63	97.02	97.14	<b>97.27</b>
4	92.46	93.78	88.22	92.50	62.88	67.50	96.53	<b>97.04</b>	96.73
5	83.75	70.38	75.67	88.20	73.24	61.24	91.11	<b>93.97</b>	92.54
6	<b>97.18</b>	93.31	84.12	93.14	78.58	64.86	95.49	95.87	94.58
7	30.43	30.99	90.76	<b>94.13</b>	57.44	59.73	85.19	92.57	93.17
Mean	83.67	78.52	89.90	94.59	67.00	71.14	94.90	<b>96.52</b>	96.10
Overall	80.28	76.05	88.74	93.66	65.51	70.08	93.16	<b>95.40</b>	95.33

	$D_0^{fh}$	$D_0^{sh}$	$D_1^{fh}$	$D_1^{sh}$	$D_2^{fh}$	$D_2^{sh}$	NB	MLJS-M	MLJS-N
1	97.11	75.11	99.33	99.05	70.21	83.49	99.34	<b>99.39</b>	99.12
2	91.37	91.57	99.28	99.46	64.06	68.02	99.67	<b>99.74</b>	99.27
3	95.15	85.56	85.73	95.22	46.77	70.64	96.77	<b>96.87</b>	96.82
4	94.64	94.31	89.98	95.01	50.86	63.04	96.84	<b>97.23</b>	97.04
5	90.83	81.01	75.25	87.03	59.53	60.40	93.07	94.74	<b>94.75</b>
6	<b>96.66</b>	89.91	79.93	89.35	55.22	52.21	91.78	92.45	91.97
7	29.80	43.50	89.47	91.24	55.04	65.01	79.21	92.65	<b>92.92</b>
Mean	85.08	80.14	88.42	93.77	57.38	66.12	93.81	<b>96.15</b>	95.98
Overall	82.46	76.59	87.24	93.17	54.06	63.12	91.71	<b>94.12</b>	94.04

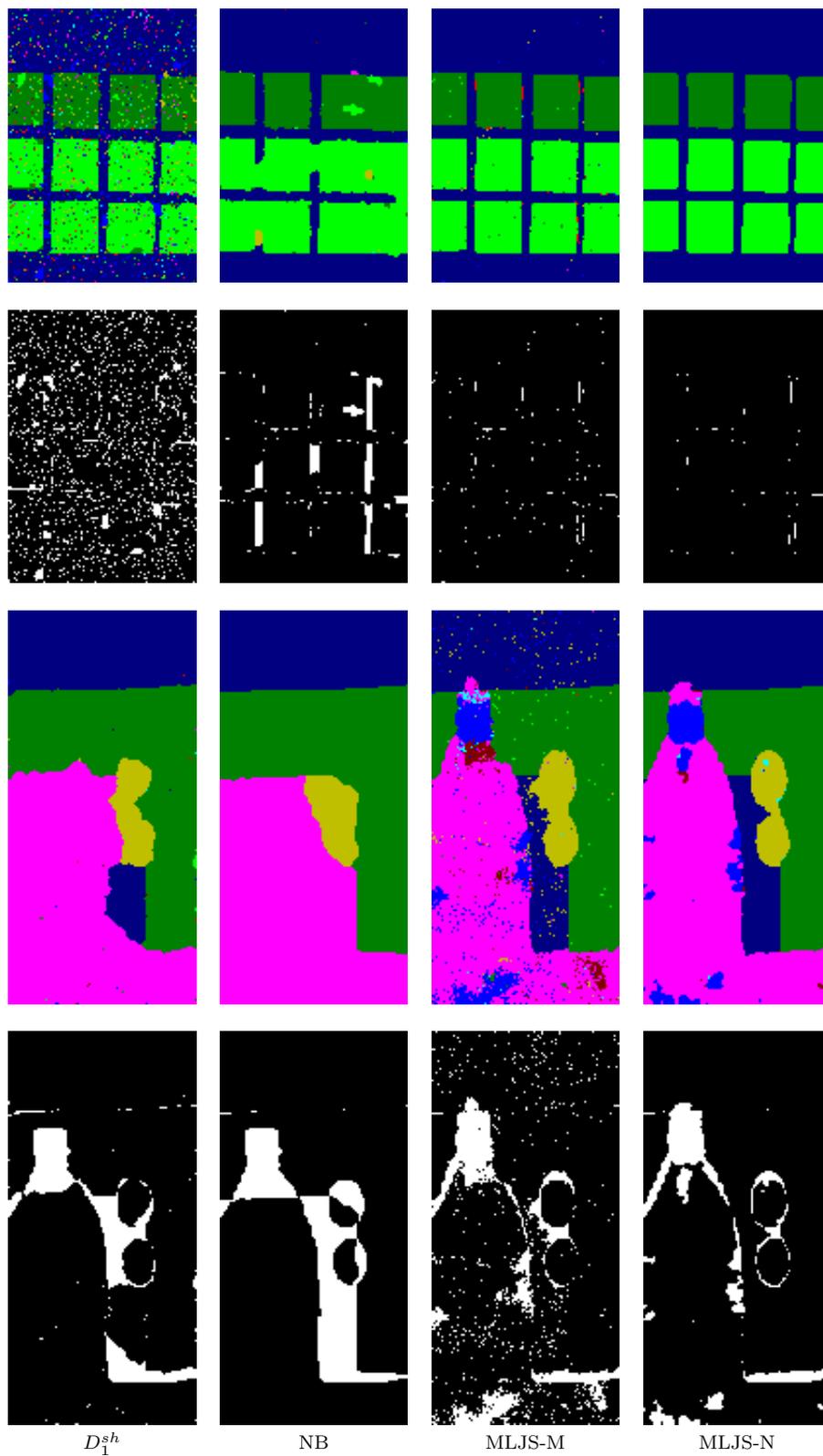


**Fig. 7** Comparison of sensitivity of weighting parameters  $\beta$  and  $\gamma$  between NB and MLJS framework using split spectrum. Location of best measurements is indicated by a red point.

9. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 23(11), 2001.
10. X. Bresson and F.T. Chan. Non-local unsupervised variational image segmentation models. Technical report, UCLA CAM, 2008.
11. T.F. Chan and L.A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(1):266–277, 2001.
12. G. Chen and S.-E. Qian. Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage. *IEEE Transactions on Geoscience and Remote Sensing*, 49(3):973–980, 2011.
13. T. Dekel, M. Rubinstein, C. Liu, and W.T. Freeman. On the effectiveness of visible watermarks. In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6864–6872, 2017.
14. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255, 2009.
15. S. Escalera, O. Pujol, and P. Radeva. On the decoding process in ternary error-correcting output codes. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 32(1):120–134, 2010.
16. R. Fletcher. *Practical methods of optimization; (2nd Ed.)*. Wiley-Interscience, 1987.
17. D. Han, J. Bayouth, Q. Song, A. Taurani, J. Buatti, M. Sonka, and X. Wu. Globally optimal tumor segmentation in PET-CT images: A graph-based co-segmentation method. In *Proceedings of International Conference on Information Processing in Medical Imaging (IPMI)*, pages 245–256, 2011.
18. B. Johnston, M.S. Atkins, B. Mackiewicz, and M. Anderson. Segmentation of multiple sclerosis lesions in intensity corrected multispectral MRI. *IEEE Transactions on Medical Imaging*, 15(2):154–169, 1996.
19. O.O. Karadag and F.T. Yarman Vural. Image segmentation by fusion of low level and domain specific information via Markov random fields. *Pattern Recognition Letters*, 46:75–82, 2014.
20. Z. Kato and T.-C. Pong. A Markov random field image segmentation model for color textured images. *Image and Vision Computing*, 24(10):1103 – 1114, 2006.
21. Z. Kato, T.-C. Pong, and S.G. Qiang. Multicue MRF image segmentation: Combining texture and color features. In *Proceedings of International Conference on Pattern Recognition (ICPR)*, volume 1, pages 660–663, 2002.
22. Z. Kato, J. Zerubia, and M. M. Berthod. Satellite image classification using a modified Metropolis dynamics. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 573–576, 1992.
23. V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 26(2):147–159, 2004.
24. B.-C. Kuo, H.-H. Ho, C.-H. Li, C.-C. Hung, and J.-S. Taur. A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(1):317–326, 2014.
25. M. Lachaize, S. Le Hégarat-Masclé, A. Aldea, E. Maitrot, and R. Reynaud. Evidential framework for Error Correcting Output Code classification. *Engineering Applications of Artificial Intelligence*, 73:10–21, 2018.
26. C. Le Guyader and L.A. Vese. A combined segmentation and registration framework with a nonlinear elasticity smoother. *Computer Vision and Image Understanding*, 115(12):1689–1709, 2011.
27. W. Li, O.H. Jafari, and C. Rother. Deep object co-segmentation. preprint, 2018.
28. J. Milgram, M. Cheriet, and R. Sabourin. “One Against One” or “One Against All”: Which one is better for handwriting recognition with SVMs? In *International Workshop on Frontiers in Handwriting Recognition (IWFHR)*, 2006.
29. E.N. Mortensen and W.A. Barrett. Interactive segmentation with intelligent scissors. *Graphical Models and Image Processing*, 60(5):349–384, 1998.



**Fig. 8** Best achieved results without noise (w.r.t.  $ACC$ ) for board 5 (upper half) and board 7 (lower half). Segmentations are shown in first and second rows, respectively. Error images w.r.t. ground truth are shown in second and fourth, respectively. Results of leftmost column are the best ones obtained among independent segmentations.



**Fig. 9** Best achieved results (w.r.t.  $ACC$ ) with mapped scores degraded by 20% of impulsive noise for board 2 (upper half) and board 7 (lower half). Segmentations are shown in first and second rows, respectively. Error images w.r.t. ground truth are shown in second and fourth, respectively. Results of leftmost column are the best ones obtained among independent segmentations.

30. E. Morvant, A. Habrard, and S. Ayache. Majority vote of diverse classifiers for late fusion. In *International Workshop on Structural Syntactic, and Statistical Pattern Recognition (IWSSSPR)*, pages 153–162, 2014.
31. R. B. Potts. Some generalized order-disorder transformations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 48(1):106–109, 1952.
32. R. Rifkin and A. Klautau. In defense of one-vs-all classification. *Journal of Machine Learning Research*, 5(Jan):101–141, 2004.
33. C. Rother, T. Minka, A. Blake, and V. Kolmogorov. Cosegmentation of image pairs by histogram matching—incorporating a global constraint into MRFs. In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 993–1000, 2006.
34. A. Savitzky and M.J.E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639, 1964.
35. A. Shen, H. Wang, W. Xi, X. Deng, J. Chen, and Y. Zhang. Multi-phase simultaneous segmentation of tumor in lung 4D-CT data with context information. *PLoS One*, 12(6), 2016.
36. P. Strandmark and F. Kahl. Parallel and distributed graph cuts by dual decomposition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2085–2092, 2010.
37. C. Vaiphasa. Consideration of smoothing techniques for hyperspectral remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(2):91–99, 2006.
38. A. Yezzi, L. Zollei, and T. Kapur. A variational framework for joint segmentation and registration. In *Proceedings of IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*, pages 44–51, 2001.