



HAL
open science

Binary Multi Channel Morphological Neural Network

Théodore Aouad, Hugues Talbot

► **To cite this version:**

Théodore Aouad, Hugues Talbot. Binary Multi Channel Morphological Neural Network. 2022. hal-03644498

HAL Id: hal-03644498

<https://hal.science/hal-03644498v1>

Preprint submitted on 19 Apr 2022

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.

Binary Multi Channel Morphological Neural Network

Theodore Aouad¹ and Hugues Talbot¹

CentraleSupélec, Université Paris-Saclay, Inria. Gif-sur-Yvette, France
`{firstname.lastname}@centralesupelec.fr`

Abstract. Neural networks and particularly Deep learning have been comparatively little studied from the theoretical point of view. Conversely, Mathematical Morphology is a discipline with solid theoretical foundations. We combine these domains to propose a new type of neural architecture that is theoretically more explainable. We introduce a Binary Morphological Neural Network (BiMoNN) built upon the convolutional neural network. We design it for learning morphological networks with binary inputs and outputs. We demonstrate an equivalence between BiMoNNs and morphological operators that we can use to binarize entire networks. These can learn classical morphological operators and show promising results on a medical imaging application.

Keywords: Binary morphology · deep learning · binarized neural networks

1 Introduction

While demonstrating considerable success in applications, there are few theoretical results in Deep Learning. Many elements are not well understood, with networks operating as black boxes, hindering critical applications such as medical imaging or robotics. Conversely, Mathematical Morphology (MM) is a computer vision discipline with solid theoretical foundations. In this article, we propose to combine ideas from the two domains to construct a new type of neural architecture that is theoretically more justified and explainable. As argued in [1] combining the two fields is promising, as MM can be used to construct simpler and more understandable deep neural networks. Using MM is more natural than convolutional neural networks (ConvNets) for some specific tasks. For example, ConvNets are not designed to deal with binary images, while MM is constructed on set-representation of images. Also, as stated in [9], designing a correct sequence of morphological operators can be complicated and time-consuming.

Past researchers have proposed to learn both operators and structuring elements, e.g using the max-plus definition of dilations and erosion [12,4]. Other approaches introduce differentiable approximations of the max and min operators, such as the adaptive morphological layer [14], the PConv layer [10], and the most recent \mathcal{LMorph} and \mathcal{SMorph} layers [9]. However, these methods deal with grey-scale morphology and grey-scale images.

We introduces the Binary Morphological Neural Network (BiMoNN), a learnable morphological network with binary inputs and output, inspired by the structure of ConvNets. We replace the convolution by elementary operations from the set of erosions, dilations, anti-erosions, and anti-dilations: we can learn the type of operation and its associated structuring element as well as any intersection/union of these elementary operations. we show equivalence properties between our neural network and morphological operators. They can be used both as explainable results or as a way to binarize the network: once these equivalences are respected, the morphological operation can replace the neural network. Booleans weights can replace real-valued weights, improving the inference’s computing efficiency [6,7,15]. We demonstrate how we managed to learn some classical morphological operators as well as an intermediate step of a challenging medical imaging problem. A preliminary version of our work is available at [2]. Our code is publicly available online at https://github.com/TheodoreAouad/Bimonn_DGMM2022.

2 Method

2.1 Notations

Let \mathbb{N}^* be the set of non zero natural integers. Let $d \in \mathbb{N}^*$ the dimension of the image (usually $d = 2$ or $d = 3$). If $a \leq b \in \mathbb{Z}$, we denote the discrete interval as $\llbracket a, b \rrbracket = [a, b] \cap \mathbb{Z}$. Let $N_I \in \mathbb{N}^*$ and $\Omega_I = \llbracket -N_I, N_I \rrbracket^d$ be the support of binary images $I \subset \Omega_I$. Let $\Omega_S = \llbracket -N_S, N_S \rrbracket^d$ be the support of structuring elements (short: SE) $S \subset \Omega_S$. If $X \subset \mathbb{Z}^d$, we denote $\check{X} = \{-x \mid x \in X\}$ the symmetric of X with respect to the origin. For a binary image $I \subset \Omega_I$, we define its complementation $I^C = \Omega_I \setminus I$. For a SE $S \subset \Omega_S$, if there is no ambiguity, we define its complementation as $S^C = \Omega_S \setminus S$. For a set $X \subset \mathbb{Z}^d$, we denote its indicator function $\mathbb{1}_X : \mathbb{Z}^d \mapsto \mathbb{R}$ such that $\mathbb{1}_X(x) = 1$ if $x \in X$, else $\mathbb{1}_X(x) = 0$. If $\mathbb{1}_I \in \mathbb{R}^{\Omega_I}$ and $\mathbb{1}_S \in \mathbb{R}^{\Omega_S}$ we denote the convolutional product $\mathbb{1}_I \circledast \mathbb{1}_S : k \in \Omega_I \mapsto \sum_{i \in \Omega_S \mid k-i \in \Omega_I} (\mathbb{1}_I(k-i) \mathbb{1}_S(i))$.

Definition 1. *Let $S \subset \Omega_S$ be a SE. Let $I \subset \Omega_I$ be a binary image. We define the following morphological operators.*

$$\text{the dilation by } S \quad \delta_S(I) = I \oplus S = \bigcup_{s \in S} (I + s) \quad (1)$$

$$\text{the erosion by } S \quad \varepsilon_S(I) = I \ominus \check{S} = \bigcap_{s \in S} (I - s) \quad (2)$$

$$\text{the anti-dilation by } S \quad \delta_S(I)^C \quad (3)$$

$$\text{the anti-erosion by } S \quad \varepsilon_S(I)^C \quad (4)$$

$$\text{the opening by } S \quad I \circ S = (I \ominus \check{S}) \oplus S \quad (5)$$

$$\text{the closing by } S \quad I \bullet S = (I \oplus S) \ominus \check{S} \quad (6)$$

$$\text{the black top-hat by } S \quad T_b(I, S) = (I \bullet S) \setminus I = (I \bullet S) \cap I^C \quad (7)$$

$$\text{the white top-hat by } S \quad T_w(I, S) = I \setminus (I \circ S) = I \cap (I \circ S)^C \quad (8)$$

While not exhaustive, this list is representative of many useful morphological operators. These operators are obtained through composition and intersection or union of dilations, erosions, anti-dilations, and anti-erosions. Therefore, it is enough to learn these elementary operations, their SEs, and their aggregation.

2.2 Binary Structuring Element Neuron

First, we design a single neuron that replaces the convolution operation: it can learn the dilation, erosion, anti-dilation, anti-erosion, and their SEs.

Definition

Proposition 1 shows that we can express a dilation and erosion exactly using a thresholded convolution.

Proposition 1 (Morphological operators from convolution). *Let $S \subset \Omega_S$ be a binary SE and $X \subset \Omega_I$ be a binary image.*

$$X \oplus S = \left(\mathbb{1}_X \otimes \mathbb{1}_S \geq 1 \right) = \left(\mathbb{1}_X \otimes \mathbb{1}_S - 1 \geq 0 \right) \quad (9)$$

$$X \ominus S = \left(\mathbb{1}_X \otimes \mathbb{1}_S \geq |S| \right) = \left(\mathbb{1}_X \otimes \mathbb{1}_S - |S| \geq 0 \right) \quad (10)$$

We stress that the dilations and erosions only differ by a scalar, 1 for the dilation and $|S|$ for the erosion. Given S , we can learn the operation using only this scalar. Taking inspiration from these expressions, we relax S into real weights $W \in \mathbb{R}^{\Omega_S}$. We denote the softplus function by $f^+ : x \in \mathbb{R} \mapsto \ln(1 + \exp(x))$. Let $\xi : \mathbb{R} \mapsto [0, 1]$ be a smooth increasing function such that $\xi(x) \rightarrow_{x \rightarrow -\infty} 0$ and $\xi(x) \rightarrow_{x \rightarrow +\infty} 1$. In practice, $\xi(x) = \frac{1}{2} \tanh(x) + \frac{1}{2}$.

Definition 2 (BiSE neuron). *Let $W \in \mathbb{R}^{\Omega_S}$ be a weight matrix, $b \in \mathbb{R}$ a bias and $p \in \mathbb{R}$ a scaling number. We define a **BiSE (Binary Structuring Element) neuron** as follow:*

$$\epsilon^{W,b,p} : x \in [0, 1]^{\mathbb{Z}^d} \mapsto \xi(p[x \otimes f^+(W) - f^+(b)]) \in [0, 1]^{\mathbb{Z}^d} \quad (11)$$

This expression approximates the thresholded convolution of proposition 1. Indeed, we apply the convolution and subtract a bias. If this expression is negative, we want the result to be 0, else we want the result to be 1. Therefore, we multiply it by a scaling factor and threshold it. The higher $|p|$ is, the closer the output is to 0 or 1. We enforce the weights and bias to be positive: we will explain that later.

The BiSE neuron is a convolution layer with the weights and bias forced to be positive, with a smooth thresholding function as activation. In practice, all the operations are differentiable with real inputs and outputs, not binary. Thus, we define almost binary images, with pixels value either close to 0 or close to 1.

Definition 3 (Almost Binary Image). *We say an image $I \in [0, 1]^{\Omega_S}$ is **almost binary** if there exists $u < v \in [0, 1]$ such that $I(\Omega_S) \cap]u, v[= \emptyset$. We denote this set $\mathcal{I}(u, v)$.*

Morphological equivalence

The following Theorem states the conditions ensuring that the BiSE neuron is equivalent to a morphological operator. We denote $\epsilon^{W,b,+\infty} : x \in [0, 1]^{\Omega_I} \mapsto (x \otimes W > b) \in \Omega_I$ and $\epsilon^{W,b,-\infty} : x \in [0, 1]^{\Omega_I} \mapsto (x \otimes W < b) \in \Omega_I$.

Theorem 1 (Dilation / Erosion Equivalence). *Let $W \in \mathbb{R}^{\Omega_S}$ be a set of weights, $b \in \mathbb{R}$ a bias and $p \in \mathbb{R}$ a scaling factor. Let $S \subset \Omega_S$ be a candidate SE. Given an almost binary input in $\mathcal{I}(u, v)$*

- $\epsilon_{W,b,+\infty}$ is a dilation by S if and only if $\epsilon_{W,b,-\infty}$ is an anti-dilation by S if and only if

$$\sum_{i \in S^c, w_i \geq 0} w_i + u \sum_{i \in S, w_i \geq 0} w_i \leq b < v \min_{i \in S} w_i + \sum_{i \in \Omega, w_i \leq 0} w_i \quad (12)$$

- $\epsilon_{W,b,+\infty}$ is an erosion by S if and only if $\epsilon_{W,b,-\infty}$ is an anti-erosion by S if and only if

$$\sum_{i \in \Omega, w_i \geq 0} w_i - (1 - u) \min_{i \in S} w_i \leq b < v \sum_{i \in S, w_i \geq 0} w_i + \sum_{i \in \Omega, w_i \leq 0} w_i \quad (13)$$

If one of these expressions is fulfilled, we say that the BiSE neuron is **activated**. Then we have $\forall i \in S, w_i \geq 0$ and $b \geq 0$, and the output of the BiSE $\epsilon_{W,b,p}$ is almost binary.

Equations 12 and 13 check if a candidate $S \subset \Omega_S$ activates the BiSE or not. If the BiSE is activated, given an almost binary input, the output is almost binary. Therefore, applying a BiSE on an almost binary image is the same as applying the morphological operator to the binary image. Finally, we can replace the entire BiSE neuron with the associated morphological operation in the activation case. The sign of p determines if complementation is applied or not.

Theorem 1 justifies a posteriori our enforcing of the weights and bias to be positive in the BiSE definition 2: the goal is to achieve activation, which is only possible for positive parameters.

The BiSE can only be activated by at most one $S \subset \Omega_S$ simultaneously. We can recover this S with proposition 2.

Proposition 2 (Linear Check).

Let us assume the BiSE is activated for almost binary images in $\mathcal{I}(u, v)$. Let $b \in \mathbb{R}$ be the BiSE bias, let $W \in \mathbb{R}^{\Omega_S}$ be the weights. Then there exists $\tau \in \mathbb{R}$ such that $S = \{i \in \Omega \mid W(i) \geq \tau\}$.

- If the BiSE is a dilation / anti-dilation

$$\tau = \frac{1}{v} \left(b - \sum_{i \in S^c, w_i \leq 0} w_i \right) \quad (14)$$

– If the BiSE is an erosion / anti-erosion

$$\tau = \frac{1}{1-u} \left(\sum_{i \in \Omega, w_i \geq 0} w_i - b \right) \quad (15)$$

We can check the two SEs associated with each threshold to recover the activation operation. If none works, then by contraposition, the BiSE is not activated. This is done in $\mathcal{O}(|\Omega|)$ operations.

Binarization

Formalism The purpose of binarization is to replace the real parameters (W, b, p) with binary parameters to improve inference efficiency [15]. If the activation inequalities of Theorem 1 are respected, we saw that we could replace the BiSE with the morphological operation. That is a form of binarization: floating point numbers are no longer required. However, these inequalities are not necessarily respected. How to binarize the BiSE neuron in this case? We want to find the closest morphological operation. Let $(W^*, b^*, p^*) \in \mathbb{R}^{|\Omega_S|} \times \mathbb{R}_+ \times \mathbb{R}$ be the learned parameters. Let $A : (S, op) \in \mathcal{P}(\Omega_S) \times \{\text{dilation, erosion}\} \mapsto A(S, op) = \{(W, b) \in \mathbb{R}^{|\Omega_S|+1} \mid f_S^{op}(W) \leq b < g_S^{op}(W)\}$, with f_S^{op} and g_S^{op} the bounds defined in Theorem 1. We want to define a dissimilarity function $d : (S, op), (W, b) \in (\mathcal{P}(\Omega_S) \times \{\text{dilation, erosion}\}) \times (\mathbb{R}^{|\Omega_S|} \times \mathbb{R}) \mapsto d(A(S, op), (W, b))$, and we want to find

$$\underset{(S, op) \in \mathcal{P}(\Omega_S) \times \{\text{dilation, erosion}\}}{\operatorname{argmin}} d(A(S, op), (W^*, b^*)) \quad (16)$$

Choice of dissimilarity A first choice is to use:

$$d(A(S, op), (W, b)) = \max \left(0, f_S^{op}(W) - b, b - g_S^{op}(W) \right) \quad (17)$$

The search for $(S, op) \in \mathcal{P}(\Omega_S) \times \{\text{dilation, erosion}\}$ has exponential complexity $\mathcal{O}(2^{|\Omega_S|})$. To simplify this search, we take inspiration from proposition 2: if the BiSE is activated, we know that $S = \{i \mid w_i \geq \tau\}$ for a certain τ . Therefore, we reduce the search to all S of this thresholded form, with $\tau \in \{w_i \mid i \in \Omega_S\}$. This reduces the complexity to $\mathcal{O}(|\Omega_S|)$.

2.3 Binary Structuring Element Layer

One strength of ConvNets is their ability to learn multiple filters per layer, which we also want to ensure. In a convolutional layer, each final channel is a sum of one filter per input channel. In our case, the final result is either a union or an intersection of the elementary operators (dilation, erosion, anti-dilation, anti-erosion). Therefore, we want a layer that can learn the union or intersection of any combination of inputs. Let us consider n binary images $x_1, \dots, x_n \subset \Omega$. Let $\mathcal{C} \subset \llbracket 1, n \rrbracket$. Then the intersection and union are given as:

$$\mathbb{1}_{\bigcap_{i \in \mathcal{C}} x_i} = \left(\sum_{i \in \mathcal{C}} \mathbb{1}_{x_i} \geq |\mathcal{C}| \right) \quad (18)$$

$$\mathbb{1}_{\bigcup_{i \in \mathcal{C}} x_i} = \left(\sum_{i \in \mathcal{C}} \mathbb{1}_{x_i} \geq 1 \right) \quad (19)$$

As with the BiSE, we can use a single scalar to discriminate between the union or the intersection. To learn the set \mathcal{C} , we can use a parameter β_i for each image. The following definition ensues:

Definition 4 (LUI). Let $\beta = (\beta_1, \dots, \beta_c) \in \mathbb{R}^c$. Let $b \in \mathbb{R}$ be a bias and $p \in \mathbb{R}$ a scaling factor. We define the LUI (Layer Intersection Union) as a thresholded linear combination:

$$LUI^{\beta, b, p} : x \in (\mathbb{Z}^d)^c \mapsto \xi \left(p \left(\sum_{i=1}^c f^+(\beta_i) x_i - f^+(b) \right) \right) \in \mathbb{Z}^d \quad (20)$$

A LUI layer can learn any intersection or union of any number of almost binary inputs. It is a particular case of BiSE layer. In this case, the SE support would be $\Omega_{l_{ui}} = \{0\}^d \times \llbracket 1, n \rrbracket$. Learning the intersection is equivalent to learning an erosion, and learning the union is equivalent to learning a dilation. Then, we can deduce the following Theorem, by using the BiSE Theorem 1. We denote by $\mathcal{I} = \times_{k=1}^n \mathcal{I}(u_k, v_k)$ the set of images with n almost binary channels.

Theorem 2 (LUI intersection / union equivalence).

Let $n \in \mathbb{N}^*$ and $\mathcal{C} \subset \llbracket 1, n \rrbracket$. Let $b \in \mathbb{R}$. Let $(u_i < v_i) \in [0, 1]^{2n}$. Let $\beta \in \mathbb{R}^n$.

– $LUI^{\beta, b, +\infty}$ is an intersection by \mathcal{C} if and only if

$$\sum_{k=1, \beta_k \geq 0}^n \beta_k - \min_{k \in \mathcal{C}} \left[(1 - u_k) \beta_k \right] \leq b < \sum_{k \in \mathcal{C}, \beta_k \geq 0} \beta_k v_k + \sum_{k=1, \beta_k \leq 0}^n \beta_k \quad (21)$$

– $LUI^{\beta, b, +\infty}$ is a union by \mathcal{C} if and only if

$$\sum_{k \in \mathcal{C}, \beta_k \geq 0} \beta_k u_k + \sum_{k \in \llbracket 1, n \rrbracket \setminus \mathcal{C}, \beta_k \geq 0} \beta_k \leq b < \min_{k \in \mathcal{C}} (\beta_k v_k) + \sum_{k=1, \beta_k \leq 0}^n \beta_k \quad (22)$$

If one of these inequalities is respected, we say that the LUI is **activated**. Then, we have $\forall k \in \mathcal{C}, \beta_k \geq 0$ and $b \geq 0$ and the output of the LUI is almost binary.

As for the BiSE, if the LUI is activated, the set \mathcal{C} can be found by thresholding the $(\beta_k)_k$ for a specific value. If the LUI is not activated, it is still possible to binarize with the same method described for the BiSE in section 2.2.

We combine the BiSE neurons and the LUI in order to be able to learn the morphological operators and aggregate them as unions or intersections.

Definition 5 (BiSEL). A *BiSEL (BiSE Layer)* is the combination of multiple *BiSE* and multiple *LUI*. Let $(\epsilon_{n,k})$ be $N * K$ *BiSE* and $(LUI_k)_k$ be K *LUI*. Then we define a *BiSEL* as:

$$\phi : x \in (\mathbb{Z}^d)^n \mapsto \left(LUI_k \left[(\epsilon_{n,k}(x_n))_n \right] \right)_k \in (\mathbb{Z}^d)^K \quad (23)$$

The *LUI* and *BiSE* are called the *BiSEL's elements*. If every element is activated, we say that the *BiSEL* is **totally activated**.

If the *BiSEL* is activated, given almost binary inputs, the outputs of the *BiSEL* are also almost binary.

The *BiSEL* follows the same logic as a convolutional layer. In *ConvNets*, at each layer, we have k filters. Each filter applies one convolution to each channel, and then we apply a linear combination to all convoluted channels. In *BiSEL*, we have K filters (the number of *LUI*). For each filter, we apply a morphological operation to a channel, and then we aggregate by taking the intersection or union of any channels. A schema of *BiSEL* can be found in figure 1.

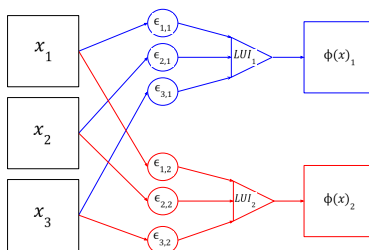


Fig. 1: Schema of *BiSEL*. Input x with 3 channels. Output $\phi(x)$ with 2 channels.

Binarizing a *BiSEL* is equivalent to binarizing each of its element. If an element is activated, it is easy to binarize it. For those who are not, we must approximate them.

The parameters are the *BiSE* weights $\{W_{n,k}\}_{n,k} \subset \mathbb{R}^{\Omega_S}$, the *BiSE* biases $\{b_{n,k}\}_{n,k} \subset \mathbb{R}$, the *LUI* parameters $\{\beta_{n,k}\}_{n,k} \subset \mathbb{R}$ and the *LUI* biases $\{b_k^l\}_k \subset \mathbb{R}$. There are $N \cdot K (|\Omega_S| + 2) + K$ parameters.

2.4 BiMoNN

The *BiSEL* is akin to a complete convolution layer with multiple filters. We can simply stack multiple *BiSELs* to create a network similar to a convolutional neural network.

Definition 6 (BiMoNN). Let ϕ_1, \dots, ϕ_L be L *BiSEL*. We define the *Binary Morphological Neural Network (BiMoNN)* as:

$$BiMoNN = \phi_L \circ \dots \circ \phi_1 \quad (24)$$

We say that the BiMoNN is **totally activated** if all the composing BiSEs are totally activated.

The BiMoNN can learn the composition of any intersection or union of dilation, erosion, anti-dilation, and anti-erosion. That includes but is not limited to opening, closing, black top-hat, and white top-hat.

There are two drawbacks to our current formulation. First, the input has the same dimension as the output. ConvNets can be combined in multi-scale architectures, which compresses the data. Currently, we cannot solve classification tasks with our method. Second, we can only learn predetermined composition of operators. However, some morphological operators like skeletonization or reconstruction need an undetermined number of iterations before convergence. Incorporating these features is left for further work.

Learning phase The BiMoNN is fully differentiable, and we learn its parameters with classical deep learning techniques. Let $\mathcal{L} : \Omega_I^2 \rightarrow \mathbb{R}$ be a differentiable loss function. Depending on our interpretation of the outside domain $\mathbb{Z}^d \setminus \Omega_I$, the effects on the border can vary. Thus, the loss is computed by avoiding the borders, whose size is half of the kernel size Ω_S in each direction. We try classical segmentation losses: the dice loss [11], the binary cross-entropy (BCE), and the mean squared error (MSE).

Then, given a dataset of N labeled images (X, Y) , we minimize the error $\frac{1}{N} \sum_{i=1}^N \mathcal{L}(\text{BiMoNN}(x_i), y_i)$ using Adam [8]. We compute the loss gradient with the backpropagation algorithm [13]. In practice, we replace $f^+(b)$ by $f^+(b) + 0.5$ to help with training for all BiSE and LUI operators. The convolution weights W follow the kaiming uniform initialization [5], and the biases begin at $f^+(2) + 0.5 = 0.63$. We initialize The BiSE and LUI scaling factors p at 0. That may not be optimal, as it could bias the BiSE towards some operations and Sels. Could we find other unbiased methods toward a particular operation or Sel ? This is left for further research.

Binarization Once the training is done, we can binarize each element independently to binarize the whole BiMoNN. Theorems 1 and 2 give activation conditions regarding a specific set of almost binary images $\mathcal{I}(u, v)$. Given a layer l , the range $[u_l, v_l]$ depends on the activation of the layer $l-1$ and only exists if all the layers $\phi_1, \dots, \phi_{l-1}$ are totally activated. This ideal situation does not occur in most cases. In practice, we binarize the network sequentially layer by layer, from the first to the last, and compute the almost binary range $[u_l, v_l]$ along the way. If a layer ϕ_l is not totally activated, we use the approximation described in 2.2: then, the next range becomes $[u_{l+1}, v_{l+1}] = [0, 1]$.

3 Experiments

3.1 Classical morphological operators

We will be using the three datasets described in [2]: Diskorect, a set of custom-generated images composed of ellipses and rectangles that create as many dif-

ferent situations as possible; thresholded MNIST upsized to 50×50 , and its complementation that we call inverted MNIST. Examples are shown in figure 3.

Experiment description We learn the following operators: dilation, erosion, opening, closing, black top-hat, and white top-hat. The 3 SEs used can be seen on table 2. For erosion and dilation on MNIST, we use smaller SEs of size 5×5 . Else for all other operations, SEs are of size 7×7 . The experimental protocol is as follows: let γ be one of these morphological operators. The input-output couples are $\{(x_i, \gamma(x_i)) \mid x_i \in \text{dataset}\}$. The hyperparameters are the loss used, the learning rate, and the batch size. For each result, see the code for the hyperparameters used. We measure the DICE score [3] between the predictions and targets.

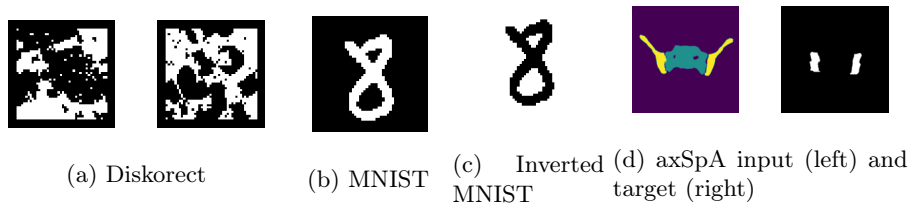


Fig. 2: Target SEs

3.2 A real example: defining regions of interest

Axial Spondyloarthritis (axSpA) is the most common auto-immune rheumatic disease. It is painful and debilitating if untreated or incorrectly diagnosed. The disease can be detected using MRIs: the inflammation is located on the sacroiliac joint. Given the sacrum and the iliac shapes, the task is to detect this joint region. We call this task axSpA.

The dataset is composed of around 1100 images with categorical images of size about 512×512 with two values: one for the iliac and one for the sacrum. We transform them into binary images by separating each categorical value into one channel: final images have 2 channels. The target is defined as the joint region. Let $S_a = \llbracket -20, 20 \rrbracket \times \{0\}$. The target operation is $\gamma_a : X \in \{0, 1\}^{512 \times 512 \times 2} = (X_1 \oplus S_a) \cap (X_2 \oplus S_a)$, with X_i the i -channel. The resulting input-output couples are $\{(x_i, \gamma_a(x_i)) \mid x_i \in \text{axSpA dataset}\}$ (see figure 3d).



(a) Diskorect

(b) MNIST

(c) Inverted
MNIST

(d) axSpA input (left) and
target (right)

Fig. 3: Datasets example

4 Discussion

Results on all operations are depicted in table 1.

Table 1: DICE Results on morphological operations. \mathbb{R} means real weights, \mathbb{B} means after binarization (or approximate). **Bold** means DICE > 0.8.

Dataset Operation \ Sel	Diskorect			MNIST			Inverted MNIST			
	Disk	Stick	Cross	Disk	Stick	Cross	Disk	Stick	Cross	
Dilation \oplus	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00
Erosion \ominus	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00
Opening $T_w(\cdot)$	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 0.99	\mathbb{R} 1.00 \mathbb{B} 0.99	\mathbb{R} 0.99 \mathbb{B} 0.93	\mathbb{R} 1.00 \mathbb{B} 0.98	\mathbb{R} 0.99 \mathbb{B} 0.90	\mathbb{R} 1.00 \mathbb{B} 0.90
Closing $T_w(\cdot)$	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 0.92 \mathbb{B} 0.87	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 0.66	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 0.91	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 1.00 \mathbb{B} 1.00
White tophat $T_w(\cdot)$	\mathbb{R} 0.82 \mathbb{B} 0.28	\mathbb{R} 0.97 \mathbb{B} 0.33	\mathbb{R} 0.81 \mathbb{B} 0.80	\mathbb{R} 0.00 \mathbb{B} 0.40	\mathbb{R} 0.06 \mathbb{B} 0.32	\mathbb{R} 0.78 \mathbb{B} 0.65	\mathbb{R} 0.00 \mathbb{B} 0.12	\mathbb{R} 0.13 \mathbb{B} 0.14	\mathbb{R} 0.22 \mathbb{B} 0.21	\mathbb{R} 0.22 \mathbb{B} 0.21
Black tophat $T_b(\cdot)$	\mathbb{R} 1.00 \mathbb{B} 0.31	\mathbb{R} 1.00 \mathbb{B} 0.82	\mathbb{R} 1.00 \mathbb{B} 1.00	\mathbb{R} 0.00 \mathbb{B} 0.00	\mathbb{R} 0.38 \mathbb{B} 0.00	\mathbb{R} 0.00 \mathbb{B} 0.00	\mathbb{R} 0.98 \mathbb{B} 0.76	\mathbb{R} 1.00 \mathbb{B} 0.29	\mathbb{R} 1.00 \mathbb{B} 0.26	\mathbb{R} 1.00 \mathbb{B} 0.26

4.1 Sequential operations

For dilation and erosion, the SEs are perfectly learned. The perfect DICE (=1) is reached in almost all cases. Even when the BiSEs are not activated, the approximate binarizations yield good results. Performance can decrease (see disk erosion on inverted MNIST), or increase (see disk erosion on MNIST).

On opening and closing, we reach perfect dice for most cases. All SEs can be seen on the code repository. We manage to recover the target SEs for most operations except on the opening on the inverted MNIST. We want to test the following hypothesis \mathcal{H} : dual operators learn similarly on dual datasets. On Diskorect, for any image X in this dataset, then X^C is also in it. Therefore \mathcal{H} implies that opening and closing should have similar results. Already the hypothesis is rejected if we look at the stick opening vs. the stick closing: one succeeds and the other fails. On MNIST, \mathcal{H} implies that the closing on MNIST should behave the same as the opening on inverted MNIST and vice versa. That is not observed and refutes further the hypothesis. Note that the model learns two anti-dilations for the closing, which is equivalent to learning dilation followed by erosion.

We compare our results to \mathcal{LMorph} and \mathcal{SMorph} [9]. They reach almost perfect DICE for erosion and dilation, but recovering the SEs from the learned weights is not straightforward. Moreover, on the opening and closing, the \mathcal{LMorph} presents stability issues, and the \mathcal{SMorph} fails to converge correctly to the proper operations. Those results are not surprising as these networks were designed for grey-scale morphology.

4.2 Parallel operations

In this section, we study the model’s ability to learn the complementation and intersection with top-hat transform. We manage to learn the black top-hat for the Diskrect and inverted MNIST. For the cross, the binarization is perfect (see figure 4). The upper branch learns the complementation of the identity, and the bottom branch learns the closing with two sequential anti-dilations. The white top-hat only worked on Diskrect. The binarization decreases performance.

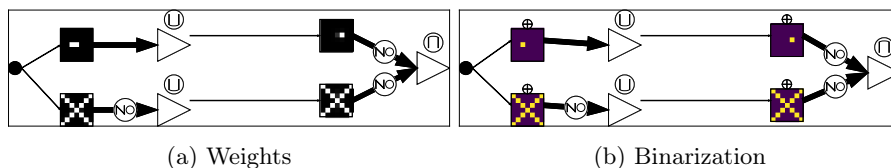


Fig. 4: Learned model for black top-hat with cross

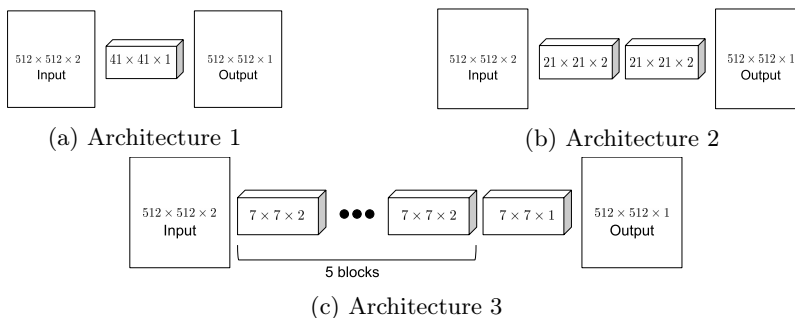


Fig. 5: Architectures used for axSpA

4.3 AxSpA regions of interest

We attempt to learn the joint region. We want to learn the intersection of two dilations of size 41×41 . We try three different architectures that can theoretically learn this operation (see figure 5). They go from wide and shallow to deeper and narrower. Architectures 1 and 2 learn the operation well, reaching an excellent DICE (1 and 0.9). Architecture 1’s weights are close to the target, and the binarization yields good results (DICE=0.98). On the other hand, architecture 2 learns wider dilations, and its binarization is quite bad (DICE=0.7). Note that there are multiple ways for this architecture to learn the target. Finally, the deeper network failed to learn anything. That may be due to the initialization:

the output is constant and stuck in a zero-grad zone, which is a case of vanishing gradient. This application is hopeful in the study of axSpA. This model could be used with classical convolutional neural networks to improve MRI analysis by exploiting the ROI's detection using segmentation results.

5 Conclusion

We create a neural network designed to work on binary elements. It comes with explainable results that can be used to binarize the network, improving its embedded applicability for inference. We manage to learn the erosion, dilation, opening, and closing of small structuring elements. We can learn sequential operations and more complex ones that need the intersection of multiple operators. Even with bigger structuring elements, we learn appropriate behaviors for concrete applications in medical imaging. However, there are still unanswered questions. Why does the training differ for dual operations? How can we avoid vanishing gradient in deeper networks? How can we reduce the dimension to create a classification model? Can we redesign the model to take into account undetermined number of iterations? How can we binarize more efficiently to preserve the performance, especially if there are multiple channels?

References

1. Angulo, J.: Some open questions on morphological operators and representations in the deep learning era. In: International Conference on Discrete Geometry and Mathematical Morphology. pp. 3–19. Springer (2021)
2. Aouad, T., Talbot, H.: Binary morphological neural network. arXiv preprint arXiv:2203.12337 (2022)
3. Dice, L.R.: Measures of the amount of ecologic association between species. *Ecology* **26**(3), 297–302 (1945)
4. Franchi, G., Fehri, A., Yao, A.: Deep morphological networks. *Pattern Recognition* **102**, 107246 (2020)
5. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 249–256. JMLR Workshop and Conference Proceedings (2010)
6. Hubara, I., Courbariaux, M., Soudry, D., El-Yaniv, R., Bengio, Y.: Binarized neural networks. *Advances in neural information processing systems* **29** (2016)
7. Kim, M., Smaragdis, P.: Bitwise neural networks. arXiv preprint arXiv:1601.06071 (2016)
8. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2017)
9. Kirszenberg, A., Tochon, G., Puybureau, É., Angulo, J.: Going beyond p-convolutions to learn grayscale morphological operators. In: International Conference on Discrete Geometry and Mathematical Morphology. pp. 470–482. Springer (2021)
10. Masci, J., Angulo, J., Schmidhuber, J.: A learning framework for morphological operators using counter-harmonic mean. In: International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing. pp. 329–340. Springer (2013)

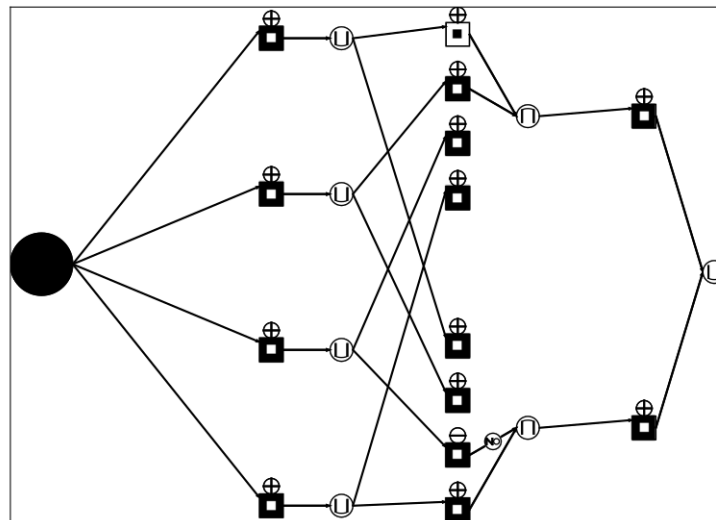
11. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation (2016)
12. Mondal, R., Dey, M.S., Chanda, B.: Image restoration by learning morphological opening-closing network. *Mathematical Morphology-Theory and Applications* **4**(1), 87–107 (2020)
13. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning Representations by Back-propagating Errors. *Nature* **323**(6088), 533–536 (1986). <https://doi.org/10.1038/323533a0>, <http://www.nature.com/articles/323533a0>
14. Shen, Y., Zhong, X., Shih, F.Y.: Deep morphological neural networks. arXiv preprint arXiv:1909.01532 (2019)
15. Simons, T., Lee, D.J.: A review of binarized neural networks. *Electronics* **8**(6), 661 (2019)

Table 2: Results on erosion and dilation. DICE error ($1 - \text{DICE}$) is presented for each case: \mathbb{R} before binarization, \mathbb{B} after binarization. \checkmark indicates that the neuron is activated.

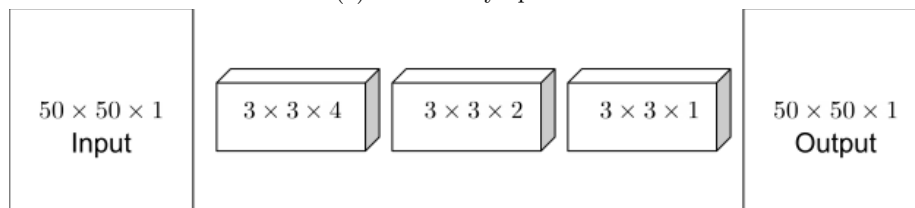
Dataset	Diskrect			MNIST			Inverted MNIST		
Operation	Disk	Stick	Cross	Disk	Stick	Cross	Disk	Stick	Cross
Target									
Dilation \oplus	 \mathbb{R} 0.000 \mathbb{B} 0.000	 \mathbb{R} 0.000 \mathbb{B} 0.000	 \mathbb{R} 0.000 \mathbb{B} 0.000	 \mathbb{R} 0.000 \mathbb{B} 0.000	 \mathbb{R} 0.000 \mathbb{B} 0.000	 \mathbb{R} 0.000 \mathbb{B} 0.000	 \mathbb{R} 0.000 \mathbb{B} 0.000	 \mathbb{R} 0.000 \mathbb{B} 0.000	 \mathbb{R} 0.000 \mathbb{B} 0.000
Erosion \ominus	 \mathbb{R} 0.000 \mathbb{B} 0.000	 \mathbb{R} 0.000 \mathbb{B} 0.000	 \mathbb{R} 0.000 \mathbb{B} 0.000	 \mathbb{R} 0.001 \mathbb{B} 0.000	 \mathbb{R} 0.000 \mathbb{B} 0.000	 \mathbb{R} 0.000 \mathbb{B} 0.000	 \mathbb{R} 0.000 \mathbb{B} 0.004	 \mathbb{R} 0.001 \mathbb{B} 0.001	 \mathbb{R} 0.000 \mathbb{B} 0.000

Table 3: Results on opening and closing. DICE error ($1 - \text{DICE}$) is presented for each case: \mathbb{R} before binarization, \mathbb{B} after binarization. \checkmark indicates that the neuron is activated.

Dataset	Diskreect			MNIST			Inverted MNIST		
Operation	Disk	Stick	Cross	Disk	Stick	Cross	Disk	Stick	Cross
Target									
Opening \circ									
	\mathbb{R} 0.000	\mathbb{R} 0.000	\mathbb{R} 0.000	\mathbb{R} 0.000	\mathbb{R} 0.000	\mathbb{R} 0.000	\mathbb{R} 0.006	\mathbb{R} 0.001	\mathbb{R} 0.014
Closing \bullet									
	\mathbb{R} 0.000	\mathbb{R} 0.077	\mathbb{R} 0.000	\mathbb{R} 0.000	\mathbb{R} 0.000	\mathbb{R} 0.000	\mathbb{R} 0.000	\mathbb{R} 0.000	\mathbb{R} 0.000
	\mathbb{B} 0.000	\mathbb{B} 0.128	\mathbb{B} 0.000	\mathbb{B} 0.000	\mathbb{B} 0.336	\mathbb{B} 0.000	\mathbb{B} 0.000	\mathbb{B} 0.092	\mathbb{B} 0.001



(a) Elementary operators



(b) Corresponding architecture

Fig. 6: Bernoulli Denoising Set Up

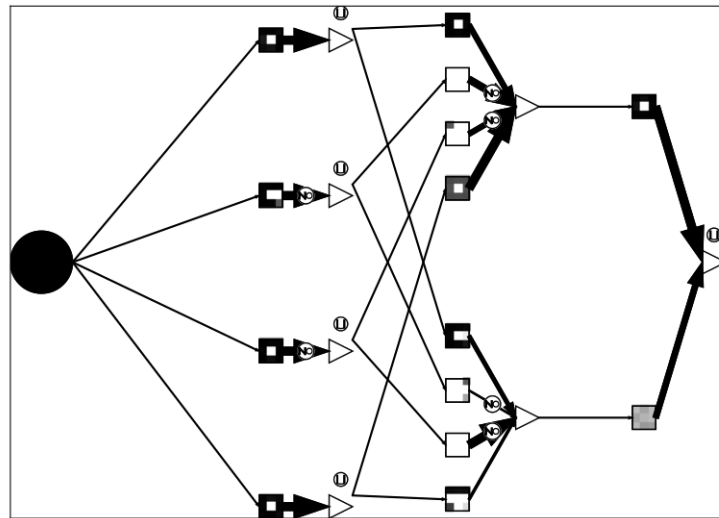


Fig. 7: Bernoulli Denoising Results, weights learned

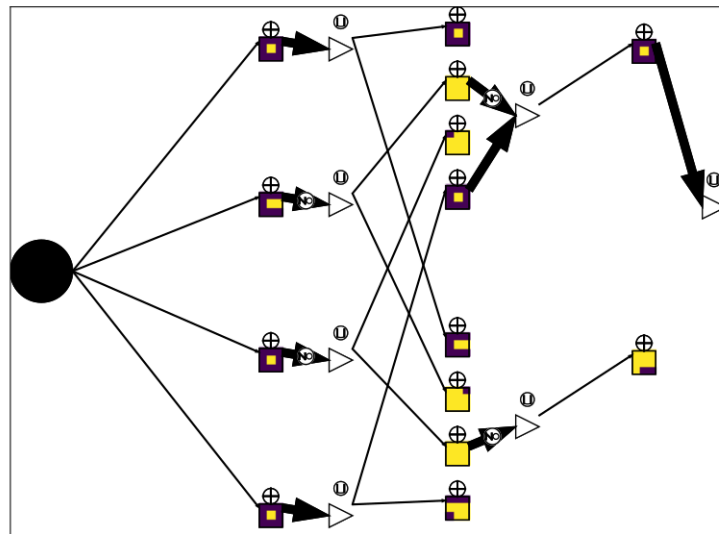


Fig. 8: Bernoulli Denoising Results, binarization

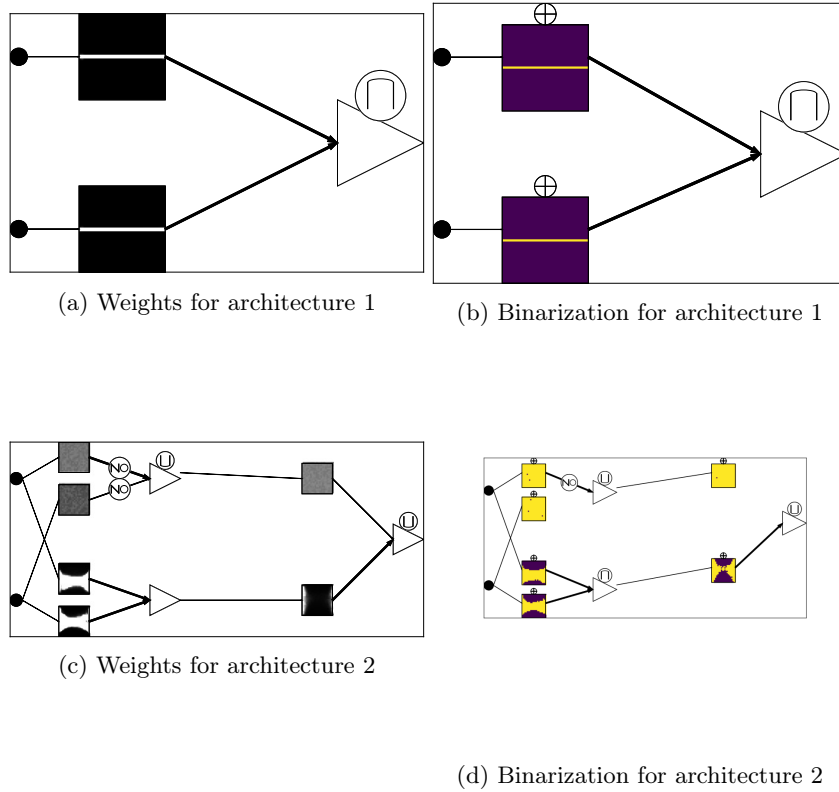


Fig. 9: Resulting Sels on axSpA results