



# Adding geodesic information and stochastic patch-wise image prediction for small dataset learning

Adam Hammoumi, Maxime Moreaud, Christophe Ducottet, Sylvain Desroziers

## ► To cite this version:

Adam Hammoumi, Maxime Moreaud, Christophe Ducottet, Sylvain Desroziers. Adding geodesic information and stochastic patch-wise image prediction for small dataset learning. *Neurocomputing*, 2021, 456, pp.481-491. 10.1016/j.neucom.2021.01.108 . hal-02879709v3

**HAL Id: hal-02879709**

**<https://hal.science/hal-02879709v3>**

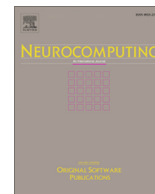
Submitted on 6 Sep 2021

**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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



# Adding geodesic information and stochastic patch-wise image prediction for small dataset learning

Adam Hammoumi<sup>a,\*</sup>, Maxime Moreaud<sup>a,b</sup>, Christophe Ducottet<sup>c</sup>, Sylvain Desroziers<sup>d</sup>

<sup>a</sup> IFP Energies nouvelles, Rond-point de l'échangeur de Solaize BP 3, 69360 Solaize, France

<sup>b</sup> MINES ParisTech, PSL-Research University, CMM, Fontainebleau, France

<sup>c</sup> Université de Lyon, UJM-Saint-Etienne, CNRS, IOGS, Laboratoire Hubert Curien UMR5516, F-42023 Saint-Etienne, France

<sup>d</sup> IFP Energies nouvelles, 1 et 4 avenue de Bois-Préau, 92852 Rueil-Malmaison, France

## ARTICLE INFO

### Article history:

Received 15 April 2020

Revised 8 January 2021

Accepted 9 January 2021

Available online 16 February 2021

### Keywords:

Image augmentation

Deep learning

Distance transform

Patch-wise segmentation

Stratified sampling

## ABSTRACT

Most recent methods of image augmentation and prediction are building upon the deep learning paradigm. A careful preparation of the image dataset and the choice of a suitable network architecture are crucial steps to assess the desired image features and, thence, achieve accurate predictions. We first propose to help the learning process by adding structural information with specific distance transform to the input image data. To handle cases with limited number of training samples, we propose a patch-based procedure with a stratified sampling method at inference. We validate our approaches on two image datasets, corresponding to two different tasks. The ability of our method to segment and predict images is investigated through the ISBI 2012 segmentation challenge dataset and generated electric field masks, respectively. The obtained results are evaluated using appropriate metrics: VRand for image segmentation and SSIM, UIQ and PSNR for image prediction. The proposed techniques demonstrate that the established framework is a reliable estimation method that could be used for a wide range of applications.

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Deep Convolutional Neural Networks (DCNNs) have demonstrated throughout the recent years their remarkable performance in handling a variety of problems in the fields of image processing and computer vision [1,2]. In particular, DCNNs are becoming a major tool for visual recognition modern tasks such as image classification, segmentation, semantic segmentation and so on.

Considering the global dimension that DCNNs have taken, predominant implementations often require large datasets, which is not always possible in many domains. Regarding that, the interest in small sample learning (SSL) is increasingly growing [3]. Novel network topologies and training methodologies are required to address this issue. In the literature, there are many approaches that attempt to face out the SSL paradigm. The data augmentation approach tries to compensate the lack of samples by applying adequate transformations to the original dataset [4]. Besides expanding the size of the initial dataset, distorted images improve the generalisability of the network. Another approach is knowledge transfer of fully trained networks that can be used to fit small data-

sets [5]. One could transfer knowledge from similar tasks or even from uncorrelated tasks [6]. Essentially, generic pattern matching tasks, created from unlabeled data, have been relevant for SSL [7]. To handle limited data, unsupervised pre-training strategies, consisting of making fine adjustments to the initialization parameters of the network, have shown promising results as well [8].

In this paper, we focus on specific DCNN architectures using only convolutional layers. These architectures referred to as fully connected neural networks (FCNNs) are mainly used for semantic segmentation or image prediction [9]. In this context, an important strategy which has been proposed to address SSL issue is to combine small neural net models and patch-based training [10–12]. First the input size of the model (ie. the number of input neurons) is reduced providing a global model with less parameters. And second, the training images are cut into small patches to feed the network. Besides requiring smaller architectures, the patch sampling strategy allows to enlarge the dataset by providing much more independent samples to train the network. At inference time, an arbitrary large target image can be processed by dividing it into patches and merging the corresponding inferred results to reconstruct a big prediction image. However, the reconstructed image is subject to patch border effects mainly due to the lack of translation invariance of the network. Several strategies have been proposed to overcome these effects [13] but there is no detailed

\* Corresponding author.

E-mail address: [adam.hammoumi@ifpen.fr](mailto:adam.hammoumi@ifpen.fr) (A. Hammoumi).

study about this problem. Another important issue of the patch-based strategy is related to spatial-awareness. Information outside the patch box is unexplored at training time due to the limited receptive field of the network. That is, the network may not learn spatial relationships larger than the patch size, which may be important for the considered task. In this work, we propose two contributions related to the patch-wise network scheme in the context of FCNN. First, we propose to reinforce the spatial-awareness of a patch-based training by adding geodesic information to the original images yielding richer spatial representations. Second, we propose a sampling strategy at inference time to overcome border effects, appearing when the assembled patches are added uniformly. The outline of this paper is organized as follows:

In Section 2, we discuss existing works about the distance transform in the context of semantic segmentation, as well as a method dealing with border effects and translation invariance of DCNNs. In Section 3, brief introduction to fully convolutional neural networks is given along with the used U-Net network architecture. We also describe some patch-based methods from the literature. We conclude this part by explaining our patch-based training procedure. In Section 4, we reveal our two contributions: adding geodesic information to the original training data, together with a stratified sampling strategy at inference time. Throughout Section 5, two datasets corresponding to two different tasks are introduced with their appropriate error metrics. In Section 6, we explain our method to quantify the shift invariance of the U-Net network, the corresponding results follow soon after. Henceforth, several comparisons are showed to demonstrate the effectiveness of our contributions. Section 7 concludes the paper.

## 2. Related work

A DCNN starts the process of features extraction according to a growing architecture from low level to high level features. Convolutional filters capture abstract features, often, not relevant for a human observer, they are recognized as crucial image characteristics. Seeking for more informative representations help extracting more features. A recent regression approach for semantic segmentation suggests that constraining the network to learn spatial information allows reducing blurry boundaries and ill segmented shapes in predictions [14]. The proposed technique is based on the distance transform. The latter yields a distance map, where, each pixel acquires a spatial awareness about its local proximity. In this configuration, less informative surfaces become more informative throughout all the image space. Without changing the network architecture, and by adding distance transformed labels as regression targets, the method proved to be an asset for the task of semantic segmentation. However, our method operates differently by combining the geodesic information, extracted from the distance map, with the original images. The addition of the distance map provides information about elements outside the scope of a patch, implicitly increasing the field of view of the network.

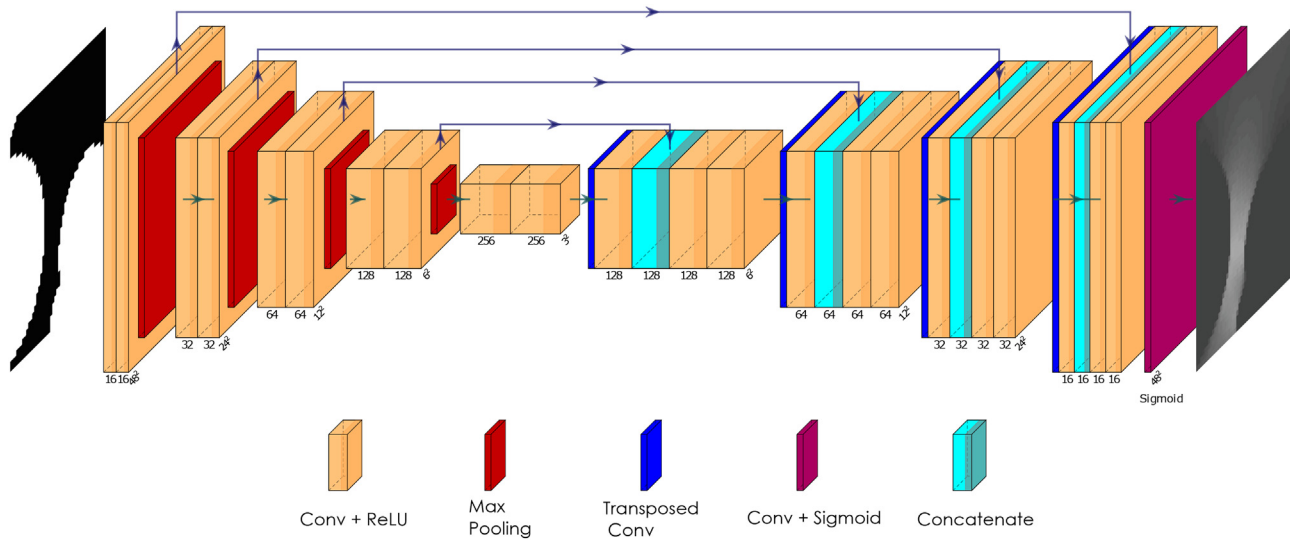
DCNNs are believed to be translation invariant [15] at some degree. Although convolutional layers have a property of equivariance to translation [16], it is not exactly the same for the complete network. This general consensus is supported by the fact that the networks have the inherent ability of learning arbitrary features: important ones, but also features as affine transformations that are irrelevant and must be discarded. Two main ideas in the literature try to address this issue. According to the first, the ability to learn translation invariance is due primarily to the networks architecture, in which the succession of convolution layers augments the receptive field of neurons [17], and to pooling layers that select a value from convolution layer output regardless of its position [18]. That is, the imperfect translation invariance is due to pooling

layers [13]. Using translation sensitivity maps and radial translation-sensitivity to quantify shift invariance introduced in [13], it can be demonstrated that the use of appropriate input data along with data augmentation comes beyond the network architecture in terms of learning translation invariant representation. A careful examination of the extent to which the network architecture is shift invariant is a relevant information for data preparation. We propose in this article a throughout study of the translation invariance of the U-Net architecture using different metrics computed on the common area of translated inference patches. Additionally, we propose an adequate stochastic sampling strategy to overcome the lack of strict translation invariance.

## 3. Background

### 3.1. Fully Convolutional Neural Networks

While DCNNs for image classification are predicting a single class for a whole image, FCNNs can be used to make dense predictions. Given an input image of any size, a FCNN produces an output having the same spatial support (possibly re-sampled) and predicting a value associated to each input pixel (or each group of pixels). For instance, the tasks of image segmentation and prediction require a pixel-wise (or a patch-wise) labeling of the output image. Therefore, feature elements are distinguished from each other by unique labels obtained from a classification process, at the pixel or patch level. Usually in a convolutional neural network, the input image goes through the convolutional layers for features extraction and gets downsized by the pooling layers. Thereupon, the results of the convolution/pooling operations are fed to the fully connected layers (FC) to classify the image. At this stage, two situations arise, if the labeling process yields one class for the whole image, it becomes a classification task. If, on the other hand, the image is classified by a label map, where each sub-part of the image is uniquely labeled, the corresponding task is most likely to be image prediction or segmentation. To obtain a label map instead of a single value label, an up-sampling step is mandatory to calculate a pixel wise output. Our attention is drawn to FCNNs and dense prediction [19,9] since they recapture the spatial information lost during down-sampling operations by up-sampling or deconvolution. FCNN architecture transforms the size of the label map back to the size of the input image or a sub-sampled version through the up-sampling process. Hence, the output image have a pixel-to-pixel correspondence with the input image. U-Net is a popular FCNN. It has first appeared under the scope of biomedical images segmentation [20]. It consists of a contraction path made of consecutive  $(3 \times 3)$  convolutions followed by  $(2 \times 2)$  max pooling matrices. An expansion path, which is composed of consecutive  $(3 \times 3)$  convolutions and  $(2 \times 2)$  transposed convolution matrices. To regain spatial information lost during the contraction path, a concatenation procedure, that consists of transferring feature maps to the expansion path, through a layer-by-layer correspondence, is established. The final step is a  $(1 \times 1)$  convolution matrix along with a *Softmax* activation function. Our experimental setup, namely the used FCNN architecture, is more or less identical to the U-Net architecture as shown in Fig. 1. It consists of an alternating sequence of two convolutions per max pooling operations. The ReLU is the used activation for each convolution and it is preceded by a batch normalization operation. The max pooling downsizes the image by a factor of two and doubles the number of features maps (or, channels). The transposed convolution operation is performing both convolution and up-sampling (through a stride of  $(2, 2)$ , that has the effect of spacing out the input) while the concatenation operation collects information from features map of the contraction path. In this work, the U-Net architecture is cus-



**Fig. 1.** U-Net architecture. Left: contraction path; right: expansion path. After each set of operations, the size of the image and the number of channels is indicated. Operations are: convolution, transposed convolution, max pooling and concatenation. Input and output images are described later in the experimental results section.

tomized for use with a patch-wise procedure. The network is rather focused on image patterns occurring locally at the patch level. The resulting output is a prediction patch as well. The training is carried out by a set of sliding ( $48 \times 48$ ) windows cropped from training images. The output is obtained from a  $(1 \times 1)$  convolution followed by a Sigmoid activation function.

### 3.2. Patch procedure

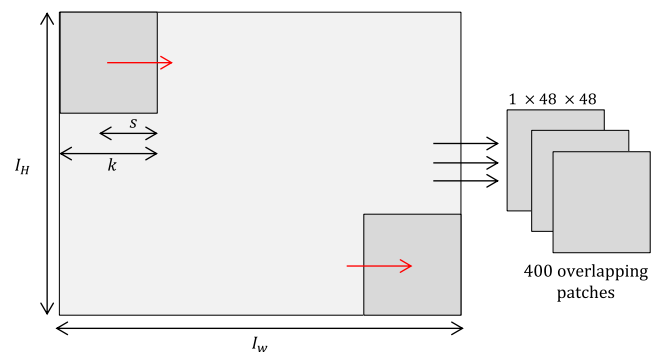
Training a FCNN at the pixel level of an image can be a challenging task in several cases. For example, in many medical applications [10], the training data is a set of high resolution images, which will require a very large memory footprint. The process of gathering a large set of training data in many domains is not always possible. An additional drawback to this approach arises, which is the risk to bias the training by forcing the network to only learn the most distinctive features from the whole image. Many attempts have been made to address this issue. In particular, evidence from [11] draw our attention to a patch-wise setup. In order to overcome the SSL issue, the latter suggests training the network on large set of patches instead of few original training images. By means of this scheme, the question of the use of context arises regarding the ability to learn compelling patterns. Basically, the size of the patches and the number of hidden layers of the FCNN control the field of view of the network and contribute explicitly in the learning of important features. For example, the authors in [12] postulate that a small patch size is not needed for their specific case, since there is little chance of finding relevant information in small image regions. Other techniques, such as [21] proposes to combine pixel-level and patch-level (ie. one label per patch) to improve segmentation accuracy. We build our method in compliance with the patch-based method proposed in [11]. Traditionally, patches are assembled in a mosaic way or by overlapping half patches. In our configuration, a patch is characterized by its size  $K \times K$  with a sliding step  $s$  over the image. That makes a total number  $(1 + \lfloor (I_W - K)/s \rfloor) \times (1 + \lfloor (I_H - K)/s \rfloor)$  of patches for an image of size  $I_W \times I_H$ . The following patch parameters are adopted: patch size  $K = 48$  with a sliding step  $s = 24$ .  $s$  is chosen in a manner to cause the overlap of patches. Suppositionally, significant information captured in-between patches can be extracted. This hypothesis was verified empirically by trying different sliding steps. For

instance, a training image of the size  $I_W = I_H = 512$ , contains  $L = 400$  overlapping patches. Fig. 2 illustrates this process.

## 4. Adding geodesic information and stratified patch sampling

### 4.1. Distance transform

Within the framework of SSL, we propose a methodology based on the distance transform to enhance the amount of accessible information in input data. Flat surfaces contained in images are less informative than textured surfaces. The lack of information in these regions will cancel the effect of the convolution matrices. The features outside the size of the convolution matrix will not be extracted as well. The distance map computation is a commonly used technique in several image processing tasks such as connected components labeling [22], Skeletonization [23], Voronoi diagrams and so on. For binary images, the distance map can be computed in the following way: we consider the two dimensional metric space  $E = \mathbb{R}^2$ . Let  $I : \psi \rightarrow \{0, 1\}$  be a binary image and  $\psi \subset E$  the support of  $I$ . The latter can be divided into background and foreground subspace. We let the set of foreground elements  $\omega = \{x \in \psi : I(x) = 1\}$  be the reference set of features. A distance map is an image transform that substitutes the value of each element in  $\psi$  by its distance from the closest feature object of  $\omega$ . The operator of the distance transform writes:



**Fig. 2.** Illustration of the patch extraction process. Parameters:  $I_H = I_W = 512$ ,  $K = 48$  and  $s = 24$ .

$$DT^d(x) = \min_{\{y|I(y)=0\}} d(x,y) \quad x,y \in \psi \quad (1)$$

An overall formulation of the distance transform that extends to grayscale and color images may be found in [24]. In general, the distance between two points  $x$  and  $y$  is expressed by:

$$d(x,y) = \inf_{\Gamma \in P_{x,y}} \int_0^{l(\Gamma)} \sqrt{1 + \gamma^2 (\nabla I(s) \cdot \Gamma'(s)^2)} ds \quad (2)$$

where  $\Gamma$  is a path parameterized by its arc length  $s \in [0, l(\Gamma)]$  and  $P_{x,y}$  is the set of all differentiable paths. The geodesic factor  $\gamma$  measures the contribution of the image gradient  $\nabla I(s)$  and spatial distances.  $\Gamma'(s) = \partial \Gamma(s) / \partial s$  is the unit vector tangent to the direction of the path. Notice that the binary image distance transform is a special case of Eq. (2), where the image has scalar values  $\{0, 1\}$  and  $\gamma = 0$ . In this case, Eq. (2) simplifies to the euclidean length of path  $\Gamma$ . Our strategy to extract a maximum amount of information from the image consists of probing both the background and foreground space. The related distance map from the foreground writes:

$$DT_c^d(x) = \min_{\{y|I(y)=\max(I)\}} d(x,y) \quad x,y \in \psi \quad (3)$$

Our main assumption is that distance maps can favorably enhance the spatial information contained in the image. Thus, we propose to combine both the original image and distance maps and use this enhanced image for training the network. Denoting  $I_e$  the enhanced image, we have:

$$I_e(x) = I(x) + \alpha DT^d(x) - \beta DT_c^d(x) \quad \alpha, \beta \text{ constants} \quad (4)$$

It is possible to compute the distance map for a grayscale image using the approach from [25]. It corresponds to a distance transform starting from lowest to highest grayscale intensities:

$$\overline{DT}^d(x) = \sum_i d(x, F_i) w_i \mid F_i = \{x; I(x) \geq i\}, w_i = 1 \quad (5)$$

Similarly, an extended symmetric distance map starting from highest to lowest grayscale intensities can be defined:

$$\overline{DT}_c^d(x) = \sum_i d(x, G_i) w_i \mid G_i = \{x; I(x) < i\}, w_i = 1 \quad (6)$$

The resulting enhanced image writes:

$$I_e(x) = I(x) + \alpha \overline{DT}^d(x) - \beta \overline{DT}_c^d(x) \quad \alpha, \beta \text{ constants} \quad (7)$$

The former distance has an important time complexity, which leads us to define an approximated distance map for grayscale images that can be deduced from a functional projected distance map  $d^\perp$  [26]. The latter is independent of grayscale scaling. A formulation of the distance in the background space is achieved by using the set of pixels of low intensities:

$$\hat{DT}^d(x) = \min_{\{y|I(y)=\min(I)\}} d^\perp(x,y) \quad x,y \in \psi \quad (8)$$

where  $d^\perp$  is the projected distance of the one developed in Eq. (2). Likewise, its symmetric distance map is given by:

$$\hat{DT}_c^d(x) = \min_{\{y|I(y)=\max(I)\}} d^\perp(x,y) \quad x,y \in \psi \quad (9)$$

The enhanced image writes:

$$I_e(x) = I(x) + \alpha \widehat{DT}^d(x) - \beta \widehat{DT}_c^d(x) \quad \alpha, \beta \text{ constants} \quad (10)$$

Distance transform can be computed with a two pass raster scanning algorithm which is well established in the literature [27]. The utility of the distance transform is stressed by the added information, emphasized in Eqs. (4), (7) or (10). Fig. 3 illustrates this process for binary and grayscale images. For our illustrations and results parts,  $\alpha$  and  $\beta$  are taken equal to 1.

## 4.2. Stratified sampling of patches

On the basis of individual patches extracted from the input image, the estimated image is fully assembled by adding the corresponding predicted patches. It was found that an exact convolution strategy (one patch for each pixel) leads to fuzzy and low quality results. This problem is tackled in the experimental results section through the shift invariance analysis. However, a major shortfall needs to be considered when using patches. When the latter are regularly distributed and summed to produce the whole prediction image, an edge effect at the border of each patch may appears, as shown in Fig. 9. Thereby, an adequate sampling strategy is required to reduce this effect. Consider the set of patches contained in one image  $\chi = \{\chi_{ij}\}$ . Each patch occupies a total area of  $A = W \times H$  where  $W$  and  $H$  are the width and the height of each patch respectively. An uniform sampling strategy consists of cropping patches as fragments of the original image following a scanning strategy from top to bottom and from left to right. Mathematically, this sampling strategy boils down to:

$$\chi_{ij} = \{x,y \mid x \in [i, W+i], y \in [j, H+j]\}, \text{ with} \quad (11)$$

$$\begin{aligned} i &= 0, s', (2 \times s'), (3 \times s'), \dots, (I_w - W + s') \\ j &= 0, s', (2 \times s'), (3 \times s'), \dots, (I_h - H + s') \end{aligned} \quad (12)$$

Eq. (11) corresponds to a formulation of the patches. The indexation strategy expressed in Eq. (12) yields a uniform sampling. To remove the edge effect at the borders, patches must be drawn stochastically. We propose using a stratified sampling strategy [28]. It consists of a uniform density of values drawn in the interval  $[-s', s']$  denoted  $\mathbf{U}(-s', s')$ . Indexing the patches with random coordinates will guarantee the generation of fresh ones every time. The new indexation strategy can be phrased in terms of:

$$\begin{aligned} i_r &= \frac{s'}{2} + \mathbf{U}(-s', s'), \frac{3s'}{2} + \mathbf{U}(-s', s'), \frac{5s'}{2} + \mathbf{U}(-s', s'), \dots \\ j_r &= \frac{s'}{2} + \mathbf{U}(-s', s'), \frac{3s'}{2} + \mathbf{U}(-s', s'), \frac{5s'}{2} + \mathbf{U}(-s', s'), \dots \end{aligned} \quad (13)$$

Consider  $N$  random samples of the whole set of patches contained in one image. The latter are yielding  $N$  predictions. The final result should be obtained by averaging over the  $N$  samples. We write:

$$I_f = 1/N \times \left( \sum_{v=1}^N \sum_{i_r, j_r} \hat{\chi}_{i_r, j_r} \right) \quad (14)$$

where  $\hat{\chi}$  refers to inferred patches.  $i_r$  and  $j_r$  have random values according to Eq. (13) and are subject to constant change. Thus, the distribution of patches in each image is unique.

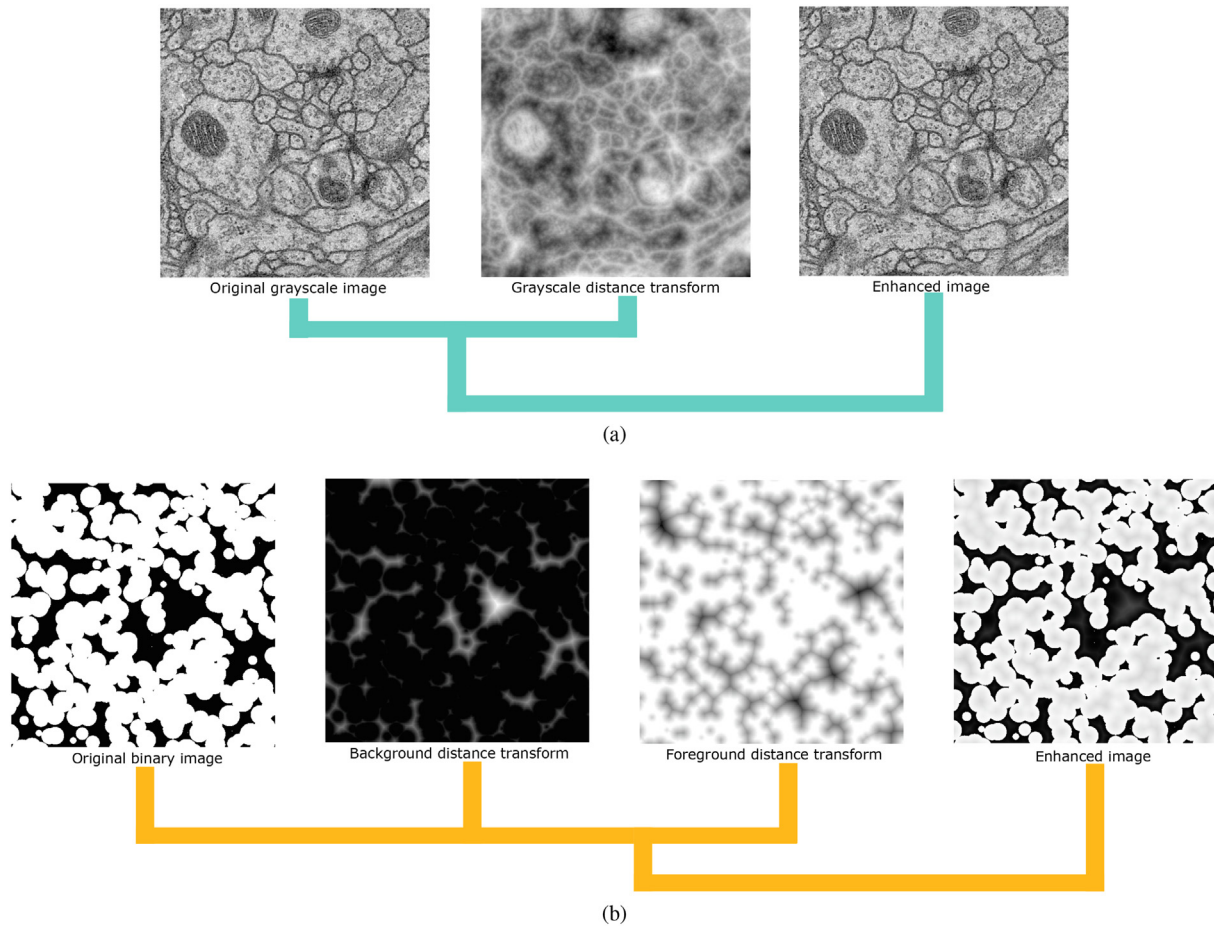
## 5. Experiments

We evaluate our distance transform-based enhancement and patch sampling strategy for two datasets. One is a state-of-the-art example on 2D segmentation of electron microscopic (EM) images of the brain, while the other is about predicting the electric field from binary images of heterogeneous materials. Each dataset requires a different task. In particular, image segmentation and prediction. Results are evaluated using adequate scoring metrics, namely: **VRand** for image segmentation and **PSNR**, **UIQ**, **SSIM** for image prediction.

### 5.1. 2D segmentation of EM images of the brain

The training dataset is a part of the public ISBI 2012 EM segmentation challenge [29]. The aim is to precisely segment an EM image, where, pixels inside a cell area have value 1 and pixels at the boundaries between neurite cross sections have value 0. A set of 30 training images ( $512 \times 512$ ) pixels along with their





**Fig. 3.** Distance transform application to grayscale (a) and binary (b) images [25]. From left to right: the process of extracting the distance map from the original image. The combination of the two yields an enhanced input image.

ground truth annotations are made public for participants. Another set of 30 validation images of the same resolution is available. Its corresponding ground truth data is kept secret by the organizers. The latter is used to evaluate the performance of the proposed algorithms. Comparisons between some of the existing state-of-the-art methods, such as, the original U-Net network instead of the patch-wise procedure, also, with and without data augmentation (mainly, affine transformations), are investigated and compared to our approach. To extract training, validation and inference images, the original training dataset is randomly shuffled into  $k = 5$  equal sized samples. Each subsample contains 26 training images, 3 validation images and 1 inference image. The outcome estimation is the average of the 5 predicted images. Thus, with the chosen  $K$  and  $s$  parameters (see Section 3.2), one network is trained on a dataset of 11466 patches and is validated on 1323 patches. The *sigmoid* activation function  $f(x) = 1/(1 + e^{-x})$  is computed over the final feature map. The network is trained using the Adam [30] optimizer and *binary cross entropy loss function* defined as:  $L(y, y') = -1/N \times (\sum_{i=1}^N (y \log(y'_i)) + (1 - y) \log(1 - y'_i))$   $y$  and  $y'$  being ground truth and predicted patches, respectively. It was noticed that the stability of accuracy and loss values requires a number of epochs  $\geq 25$ . For a batch size of 4, 30 epochs are performed. Afterwards, the measures are done for the 5 networks that were trained on random combinations of the data above. As we are not using the validation set of the competition, our results cannot be compared with the results from other participants. We have thus reproduced some of the state-of-the-art methods and networks, and evaluated them in our configuration. Our contributions

are focused on demonstrating the effect of adding the distance transform to the initial image, and using a stochastic patch-wise procedure. Detailed comparisons showing improvements are presented in the results section.

## 5.2. Electric field estimation

Knowing properties of components and spatial distribution of heterogeneous media, an efficient way to solve the problem of homogenization of physical properties uses numerical solutions of the corresponding partial differential solutions before estimating the effective properties by spatial averaging of the solution. The case of dielectric permittivity conducts to the prediction of electric field by solving the Gauss equation of electrostatics from Maxwell's equations. Iterative Fourier Transform numerical scheme can compute this field [31]. Here we explore the possibility of estimating this field directly from the media using deep learning. The heterogeneous media are modeled by Cox Boolean random models of spheres [32,33] allowing to generate realistic multi-scale microstructures. A specific algorithm described in [34] uses an original construction method which allows to run wide simulations with the least computational cost. We follow the guidelines of the latter algorithm to generate our training images. In this framework, a multi-scale microstructure is modeled by volume fractions that define aggregates ( $V_{v,inc}$  of inclusion areas), grains inside and outside the inclusion areas ( $V_v$  and  $V_{v,out}$ , respectively). The training dataset is made of  $500^2$  pixel images. The parameters  $R = 20$  (radius of spheres),  $V_{v,inc} = 0.4$ ,  $V_v = 0.6$  and  $V_{v,out} = 0.7$

are fixed for the whole image set. Based on the foregoing microstructures and phase dielectric properties, we use a numerical scheme to estimate the electric field. This method lies on several works, namely [35–37,31]. Labeled images are representations of electric field response  $E(x)$  estimated inside and outside microstructures. For dielectric constants of the phases of the binary microstructure, 0.1 and 100 (no imaginary part) are used for the black and white pixels respectively. The resulting electric field response module is converted to 8bit format (256 values) by uniform sampling. The training is performed on a dataset of 4800 training and 800 validation patches. Only binary and grayscale images are used, colored illustrations are shown solely for clarity purposes. The experimental setup is akin to the one described for the EM images of the brain. The network reaches its optimal performance after one epoch and a batch size 4. Dataset generations and homogenization codes can be found in the open access software “plug im!” [38] (see Figs. 4 and 5).

### 5.3. Evaluation metrics

#### 5.3.1. Image segmentation

**Foreground-restricted Rand scoring:**  $V_{\alpha}^{Rand}$  The boundary maps assessment is done on the basis of the official metric of the ISBI 2012 challenge. Such a boundary detection problem is sensitive to split and merge errors, where one feature element is incorrectly split into two segments, and where two distinct feature elements are incorrectly merged into one segment [29]. The  $V_{\alpha}^{Rand}$  score combines the two errors:

$$V_{\alpha}^{Rand} = \frac{\sum_{ij} p_{ij}^2}{\alpha \sum_k s_k^2 + (1 - \alpha) \sum_k t_k^2} \quad (15)$$

where  $\sum_{ij} p_{ij}^2$  is the probability that two random elements belong to the same segment of the predicted segmentation  $S$  and to the same segment of the groundtruth segmentation  $T$ . A segment by definition is a connected component. Belonging to an object  $A$  in this context means having the same label as  $A$ . The  $\alpha$  measures the importance of both the merge and split errors. In fact, they can be derived from the equation above. We have:  $V_{\alpha=0}^{Rand} = V_{split}^{Rand}$  and  $V_{\alpha=1}^{Rand} = V_{merge}^{Rand}$ . Finally,  $\sum_k s_k^2$  and  $\sum_k t_k^2$  are appropriate normalization. The used script allowing the computation of  $V_{\alpha}^{Rand}$  is the one that is proposed in the website of the challenge [29].

#### 5.3.2. Image prediction

For the evaluation of the predicted electric field image, we use some of the most common reference image quality measures, in particular the ones that are based on different measuring

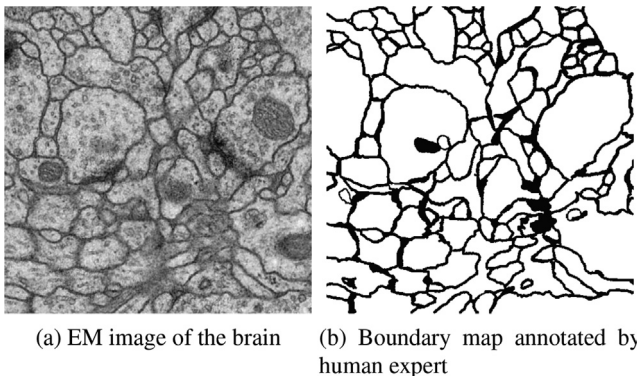


Fig. 4. The ISBI 2012 segmentation challenge training dataset.

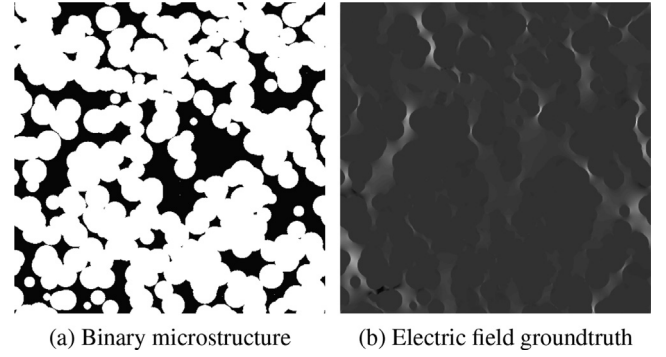


Fig. 5. The dataset utilized for the image prediction task. (a) Is a binary microstructure obtained from a boolean model of spheres and (b) its corresponding electric field image. Field intensity is proportional to grayscale pixel values.

approaches. The goal is to measure dissimilarities between two images. Hereafter in this part, we use the formulation  $I_{\star} = \{I_{\star}(i,j); \forall i = 1, \dots, W, \forall j = 1, \dots, H\}$ , with  $\star = \{o, d\}$  for both original (ground truth) and distorted (predicted) image.

**Peak signal-to-noise ratio: PSNR** Related to the mean squared error (MSE) is firstly deployed. PSNR is based upon an explicit numerical criterion which is the comparison between pixel values. Let  $I_o$  be the original image and  $I_d$  the distorted image. To perform a comparison between these images, the PSNR metric writes:

$$\text{PSNR}(I_o, I_d) = 10 \times \log_{10} \left[ \frac{(2^d - 1)^2}{\text{MSE}(I_o, I_d)} \right], \text{ with} \quad (16)$$

$$\text{MSE}(I_o, I_d) = \frac{1}{W \times H} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} (I_o(i,j) - I_d(i,j))^2 \quad (17)$$

$2^d - 1$  denotes the maximum possible value that a pixel can have. For  $d = 8$  byte coded image, the maximum value is 255. Eq. (17) measures the value differences between corresponding pixels of each image. PSNR is expressed in decibels which is a logarithmic unit. From Eq. (16), we can see that higher PSNR value is an indicator of highly similar images.

**Universal Image Quality: UIQ** Is an important tool to measure dissimilarities between two images in terms of their statistical properties [39]. The UIQ index writes:

$$\text{UIQ}(I_o, I_d) = \frac{\sigma_{I_o I_d}}{\sigma_{I_o} \sigma_{I_d}} \times \frac{2 \bar{I}_o \bar{I}_d}{\bar{I}_o^2 + \bar{I}_d^2} \times \frac{2 \sigma_{I_o} \sigma_{I_d}}{\sigma_{I_o}^2 + \sigma_{I_d}^2} \quad (18)$$

where  $\bar{I}$  and  $\sigma^2$  denote mean and variance value, respectively. Eq. (18) is an expression of the UIQ index as a product of three factors: loss of correlation (measures linear correlation), luminescence and contrast distortion. UIQ range is  $[-1, 1]$  so that the index of very similar images approaches 1.

**Structural similarity index measure: SSIM** Is an adaptation of the human visual system (HVS) that aims to assess the structural information of an image [40]. The SSIM equation writes:

$$\text{SSIM}(I_o, I_d) = A(I_o, I_d) \times B(I_o, I_d) \times C(I_o, I_d), \text{ where} \quad (19)$$

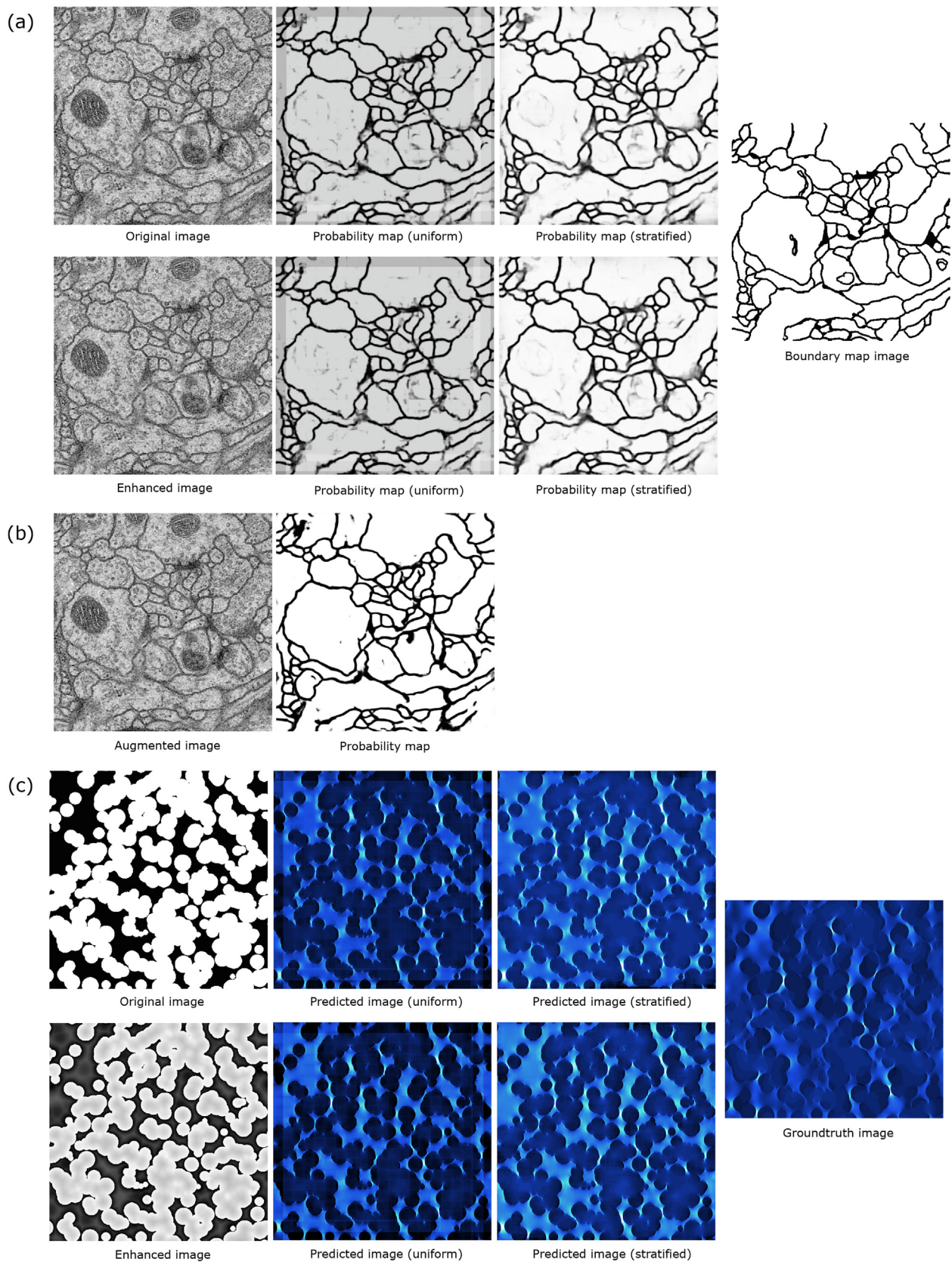
$$A(I_o, I_d) = (2 \bar{I}_o \bar{I}_d + C_1) / (\bar{I}_o^2 + \bar{I}_d^2 + C_1)$$

$$B(I_o, I_d) = (2 \sigma_{I_o} \sigma_{I_d} + C_2) / (\sigma_{I_o}^2 + \sigma_{I_d}^2 + C_2)$$

$$C(I_o, I_d) = (\sigma_{I_o I_d} + C_3) / (\sigma_{I_o} \sigma_{I_d} + C_3)$$

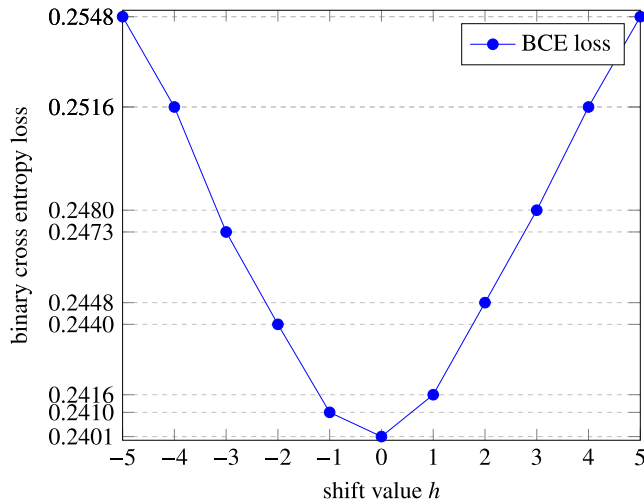
$C_1 = (K_1 \times L)^2$ ,  $C_2 = (K_2 \times L)^2$  and  $C_3 = C_2/2$ .  $L$  is the image dynamic,  $K_1$ ,  $K_2$  are constants and  $\bar{I}$  and  $\sigma^2$  are the mean and variance values, respectively. SSIM aims to identify the perceptual sim-





**Fig. 6.** (a), (c) Patch-wise U-Net results for two tasks, image prediction and image segmentation. Predictions images/probability maps are given in terms of two patch sampling strategies, uniform and stratified. Results are given for the original image and the enhanced image. (b) U-Net results for image segmentation. Probability map for augmented image dataset, by affine transformations, is illustrated.





**Fig. 7.** Evolution of the binary cross entropy loss in terms of shift transformations applied to patches of the validation dataset.

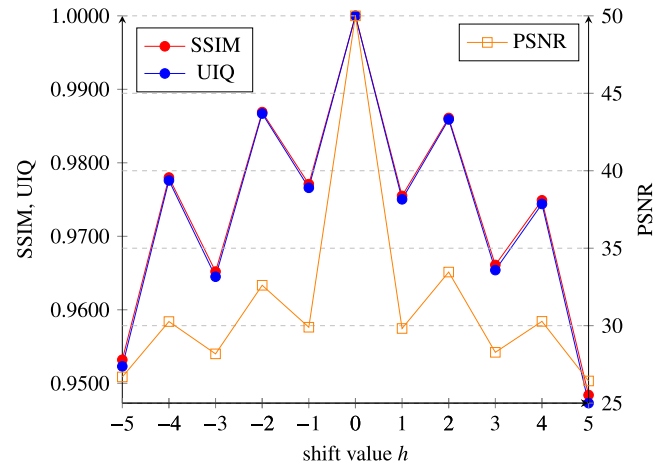
ilarity between two images through the evaluation of luminance (A-Eq. (19)), contrast (B-Eq. (19)) and image structure (C-Eq. (19)). In the presented application,  $K_1$  and  $K_2$  are taken equal to 0.01 and 0.03, respectively.

## 6. Results

In this section, experiments regarding the design of the used network are conducted. The ability of preserving network's invariance under shift transformation is investigated to demonstrate the use of the patch procedure. Afterwards, a sequence of comparisons between predicted results and ground truth data is presented.

### 6.1. Shift invariance

On the electric field dataset, we conduct shift transformations from  $-5$  to  $+5$  pixels (in the x's axis) applied to all patches of validation images. Given a fully trained U-Net on the original dataset, we firstly compare the *binary cross entropy* value for each shift value. The goal is to evaluate the ability of the network to predict translated images from non translated training data. Fig. 7 shows that the minimum error value coincides with the prediction of a non translated image, and, thence, the network smoothly loses accuracy proportionally to increasing shift values. An alternative approach would consist of feeding a trained network a prediction image  $I_p$  that yields a prediction field  $F_p$ . We select a patch  $X_p$  of the size  $W \times H$  centered at position  $(x, y)$ , the network result is an estimated patch field  $f_p$  at position  $(x, y)$ . Considering shift values  $h$ , patches  $X_p - h$  of the same size and their respective estimations  $f_p - h$  are assessed. The intersection area of these patches is of the size  $(W - 2h \times H)$  centered at  $(x, y)$ . For  $h = [-5, 5]$ , we compare the dissimilarities through the PSNR, UIQ and SSIM mean values of three different patches between the common area of each shifted patch. Table 1 and Figs. 7, 8 manifest these differences and quantify the shift "invariance" of the U-Net architecture. That



**Fig. 8.** Dissimilarities given by PSNR, SSIM and UIQ mean values between common area of original and shifted patches.

**Table 2**

$V^{Rand}$  evaluation of probability maps of the ISBI 2012 segmentation challenge dataset for two networks: Patch-wise U-Net and U-Net, using uniform and stratified sampling strategies. Original, enhanced and augmented images by affine transformations are used as datasets.

Network	Data	Sampling	$V^{Rand}$ % (Thinned)
Patch-wise U-Net	original	uniform	94, 1 $\pm$ 3, 1
		stratified	97, 3 $\pm$ 2, 2
	enhanced	uniform	95, 2 $\pm$ 2, 3
		stratified	<b>97, 9 <math>\pm</math> 1, 5</b>
U-Net	augmented	–	96, 7 $\pm$ 2, 0

suggests that the network, in our specific case, is not suited for a pixel-wise approach. Rather, a patch based procedure is stabilizing the network by forcing it to learn features from the whole patch dumping non-important features as the shift transformation. The use of U-Net as a predictor for central pixel of patch (pixel-by-pixel approach) leads to fuzzy results due to the averages of patch estimates that are not completely similar.

### 6.2. EM segmentation challenge ISBI 2012

Building upon the experimental setup described in Section 5, two experiments are conducted:

**Patch-wise U-Net** The first aim is to investigate the effect of training the network on the enhanced image dataset. Unlike the binary microstructures case, where there is a plenty of untextured, void areas that made the contribution of the distance map information, intuitively useful. This application is based on grayscale and highly textured images. Besides, the difference between the original and the transformed images is not clear from a visual point of view as shown in Fig. 6. Therefore, the small change in the pixel values of the image is expected to have a proportional improvement on the quality of the resulting probability maps. Which is not big, but sufficient to prove the interest of adding the distance

**Table 1**

PSNR, UIQ, SSIM mean values and BCE values evaluated between common area of original and shifted patches.

Shift values	-5	-4	-3	-2	-1	0	1	2	3	4	5
PSNR	26.6912	30.2654	28.1801	32.6074	29.8986	50	29.8177	33.4603	28.2784	30.2763	26.4302
UIQ	0.9523	0.9776	0.9645	0.9867	0.9766	1	0.97503	0.9859	0.9654	0.9744	0.9473
SSIM	0.9532	0.9780	0.9652	0.9869	0.9771	1	0.9755	0.9861	0.9661	0.9749	0.9484
BCE	0.2547	0.2516	0.2472	0.2440	0.2409	0.2401	0.2415	0.2448	0.2480	0.2516	0.2547

**Table 3**

PSNR, UIQ and SSIM values comparison of predicted images in Fig. 6 stemming from original and enhanced training datasets.

Images	Sampling	PSNR	UIQ	SSIM
original	uniform	34.5997	0.6891	0.8305
	stratified	32.1624	0.6443	0.7820
enhanced	uniform	35.7380	0.7030	0.8803
	stratified	<b>37.9949</b>	<b>0.7627</b>	<b>0.9026</b>
border Mask	uniform	30.8096	0.6890	0.7777
interior Mask	uniform	34.0679	0.6892	0.8137

transform. Such results will allow to show the contribution of the distance transform, even for a challenging dataset in terms of textural complexity. The sample is prepared differently, with and without the distance map. Correspondingly, probability maps are evaluated and compared to the ground truth segmentation images using the  $V^{Rand}$  metric. From the results shown in Table 2, more specifically by comparing the uniform sampling strategy results, it is clear that the  $V^{Rand}$  error has improved by training the network on the enhanced image dataset instead of the original images.

**U-Net** At this stage, a comparison with a state-of-the-art data augmentation method is achieved. It was found that random elastic deformations are very efficient to overcome over-fitting in the case of small data samples. Examples are illustrated here [20] and here [41] for the ISBI 2012 segmentation challenge dataset. In our experiment, the following transformations are used: rotation, width and height shift, shear, zoom and horizontal flip. Slight changes induced by the former transformations makes the model more robust and generalizable. The use of these operations creates artificially new images. We started with 26 samples from the original challenge dataset. With data augmentation, the model reaches its optimal performance for 1000 steps per epoch. Each operation is defined on a range of values. The number of the possibly generated samples is limited by the allowed number of pixels. For instance, height and width shifts operations are restricted by the dimensions of the patch ( $W \times H$ ). For a height shift range  $h_s = 0.05$ , the number of the generated samples is equal to  $GS_i = h_s \times 2 \times H$ , considering the two translation directions, top to bottom and vice versa. Similar arguments can be advanced to compute the number of generated images from the other operations. The product  $\prod_i GS_i \times N_{samples}$  yields the total number of samples. Results shown in Table 2 reveals that the patch-wise U-Net with the stratified sampling strategy performs better on the used dataset compared to the U-Net with data augmentation.

### 6.3. Electric field estimation

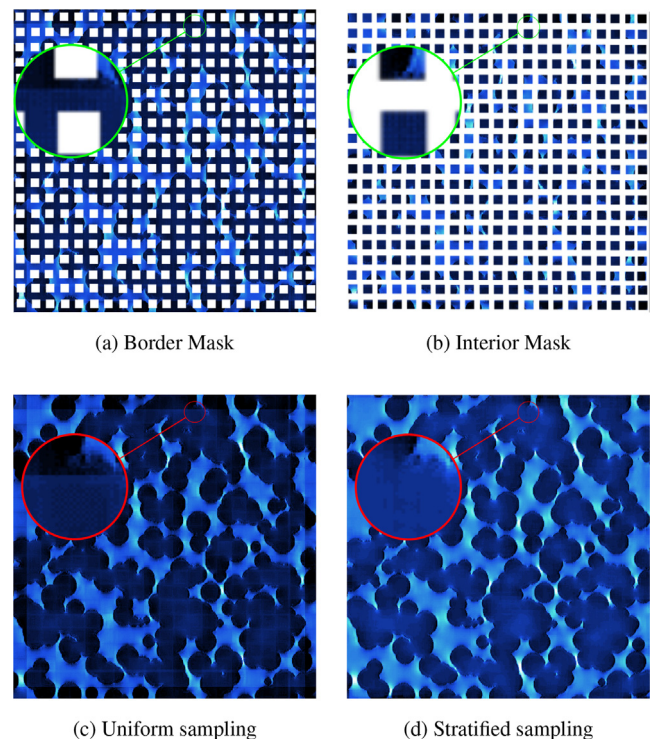
For this dataset, we investigate the effect of the enhanced images to the task of image prediction. As pointed out in the

method section, the distance map enriches the information to be extracted. A comparison of the PSNR, UIQ and SSIM indices is performed for two types of training data. Primarily, the network is trained on the original dataset. Afterwards, another training is performed on the enhanced image dataset. A visual comparison is shown in Fig. 6 and the values of the evaluation metrics are presented in Table 3. A concluding remark is that the PSNR, UIQ and SSIM indices have improved remarkably in the second case.

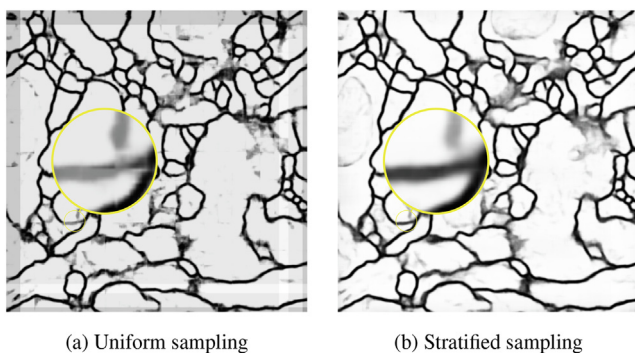
### 6.4. Effects of stratified sampling

The network processes patch by patch and yields accordingly a prediction for each one separately. To bring together the patches, a sampling strategy needs to be defined. Uniform and stratified sampling methods are investigated. In contrast to the first one, which consists of distributing the predicted patches uniformly over all the image size, the latter allocates random positions to each patch based on uniform sampled points. Thus, for each stratification iteration, novel patches and their respective predictions are created. Visually, stratified sampling reduces considerably the edge effect on the borders of patches, related results are shown in Figs. 9 and 10.

**EM segmentation challenge ISBI 2012** Here, segmentation results are compared to ground truth boundary maps. A part from



**Fig. 10.** Border area mask (a) and inside area mask (b). Prediction result with uniform sampling (c) and stratified sampling (d).



**Fig. 9.** Segmentation results with uniform sampling (a) and stratified sampling (b).

the edge effect, which is the main drawback of the uniform sampling strategy, the predicted membranes obtained from the stratification strategy are clearer and more precise as shown in Fig. 9. From a quantitative point of view, the  $V^{Rand}$  values related to the stratified sampling, shown in Fig. 9, exceed the ones obtained from the uniform sampling.

**Electric field estimation** A comparison between ground truth image and predicted results in terms of evaluation indices is shown in Table 3. For the enhanced images, stratified sampling has proven its ability to improve results in terms of scoring metrics. On the other hand, comparison between measures from the original data, subject to the two sampling strategies, doesn't confirm this postulate. We explain this by the fact that the evaluation metrics are averaged over the entire image and do not reflect the visual feeling. To verify this, two masks isolating both the edge borders and the inside area of patches are considered. A local similarity comparison between former masks applied to Fig. 10c and to ground truth image has been conducted. As expected, PSNR, UIQ and SSIM indices for uniform sampling evaluated at the edge borders area are lower than prior evaluated indices in the case of stratified sampling. Fig. 10 illustrates this operation and related results are shown in Table 3.

#### CRedit authorship contribution statement

**Adam Hammoumi:** Writing - original draft, Data curation, Investigation, Software. **Maxime Moreaud:** Methodology, Project administration, Supervision, Conceptualization, Software, Writing - review & editing. **Christophe Ducottet:** Methodology, Supervision, Conceptualization, Writing - review & editing. **Sylvain Desroziers:** Validation, Methodology, Investigation, Writing - review & editing.

## 7. Conclusion

A small sample learning methodology was proposed. It was shown that the distance transform improves the spatial-awareness of the network by providing spatial information to the original images. A patch-based procedure was applied to overcome the limited number of samples along with a stratified sampling method to remove border edge effect. The methodology was validated on two tasks: image segmentation and prediction. Tasks are portrayed by different types of dataset.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgment

The authors would like to thank Mohamed El Khamlichi for his contributions in a prior version of the used U-Net patch-based code.

## References

- [1] Y. Le Cun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444, <https://doi.org/10.1038/nature14539>.
- [2] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M.S. Lew, Deep learning for visual understanding: A review, *Neurocomputing* 187 (2016) 27–48, recent Developments on Deep Big Vision. doi: 10.1016/j.neucom.2015.09.116.
- [3] J. Shu, Z. Xu, D. Meng, Small sample learning in big data era, *CoRR abs/1808.04572* (2018). arXiv:1808.04572.
- [4] T.D. Kulkarni, W.F. Whitney, P. Kohli, J. Tenenbaum, Deep convolutional inverse graphics network, in: C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, Garnett (Eds.), *Advances in Neural Information Processing Systems* 28, Curran Associates Inc., 2015, pp. 2539–2547.
- [5] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks? in: *Proceedings of the 27th International Conference on Neural Information Processing Systems – Volume 2, NIPS'14*, MIT Press, Cambridge, MA, USA, 2014, pp. 3320–3328.
- [6] D.C. Cireşan, U. Meier, J. Schmidhuber, Transfer learning for latin and chinese characters with deep neural networks, in: *The 2012 International Joint Conference on Neural Networks (IJCNN)*, 2012, pp. 1–6, <https://doi.org/10.1109/IJCNN.2012.6252544>.
- [7] A. Ahmed, K. Yu, W. Xu, Y. Gong, E. Xing, Training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks, in: D. Forsyth, P. Torr, A. Zisserman (Eds.), *Computer Vision – ECCV 2008*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 69–82.
- [8] R. Wagner, M. Thom, R. Schweiger, G. Palm, A. Rothermel, Learning convolutional neural networks from few samples, in: *The 2013 International Joint Conference on Neural Networks (IJCNN)*, 2013, pp. 1–7, <https://doi.org/10.1109/IJCNN.2013.6706969>.
- [9] F. Lateef, Y. Ruichek, Survey on semantic segmentation using deep learning techniques, *Neurocomputing* 338 (2019) 321–348, <https://doi.org/10.1016/j.neucom.2019.02.003>.
- [10] K. Nazeri, A. Aminpour, M. Ebrahimi, Two-stage convolutional neural network for breast cancer histology image classification, in: A. Campilho, F. Karay, B. ter Haar Romeny (Eds.), *Image Analysis and Recognition*, Springer International Publishing, 2018.
- [11] D.C. Cireşan, A. Giusti, L.M. Gambardella, J. Schmidhuber, Deep neural networks segment neuronal membranes in electron microscopy images, in: *Proceedings of the 25th International Conference on Neural Information Processing Systems – Volume 2, NIPS'12*, Curran Associates Inc., Red Hook, NY, USA, 2012, pp. 2843–2851.
- [12] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, A. Campilho, Classification of breast cancer histology images using convolutional neural networks, *PLOS One* 12 (6) (2017) 1–14, <https://doi.org/10.1371/journal.pone.0177544>.
- [13] E. Kauderer-Abrams, Quantifying translation-invariance in convolutional neural networks, *CoRR abs/1801.01450* (2018). arXiv:1801.01450.
- [14] N. Audebert, A. Boulch, B.L. Saux, S. Lefèvre, Distance transform regression for spatially-aware deep semantic segmentation, *Computer Vision and Image Understanding* 189 (2019) 102809. doi:https://doi.org/10.1016/j.cviu.2019.102809.
- [15] Y. Le Cun, Learning invariant feature hierarchies, in: A. Fusiello, V. Murino, R. Cucchiara (Eds.), *Computer Vision – ECCV 2012. Workshops and Demonstrations*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 496–505.
- [16] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016, <http://www.deeplearningbook.org>.
- [17] R. Gens, P.M. Domingos, Deep symmetry networks, in: Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, K.Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* 27, Curran Associates Inc., 2014, pp. 2537–2545.
- [18] M. Jaderberg, K. Simonyan, A. Zisserman, k. kavukcuoglu, Spatial transformer networks, in: C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, R. Garnett (Eds.), *Advances in Neural Information Processing Systems* 28, Curran Associates Inc., 2015, pp. 2017–2025.
- [19] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, F.E. Alsaadi, A survey of deep neural network architectures and their applications, *Neurocomputing* 234 (2017) 11–26, <https://doi.org/10.1016/j.neucom.2016.12.038>.
- [20] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer International Publishing, Cham, 2015, pp. 234–241.
- [21] T. Wang, Z. Ji, Q. Sun, S. Han, Combining pixel-level and patch-level information for segmentation, *Neurocomputing* 158 (2015) 13–25, <https://doi.org/10.1016/j.neucom.2015.02.010>.
- [22] L. He, Y. Chao, K. Suzuki, A run-based two-scan labeling algorithm, in: M. Kamel, A. Campilho (Eds.), *Image Analysis and Recognition*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 131–142.
- [23] C. Pudney, Distance-based skeletonization of 3d images, in: *Proceedings of Digital Processing Applications (TENCON '96)*, vol. 1, 1996, pp. 209–214.
- [24] A. Criminisi, T. Sharp, C. Rother, P. Perez, Geodesic image and video editing, *ACM Transactions on Graphics* 29 (5) (Nov. 2010). doi:10.1145/1857907.1857910.
- [25] I.S. Molchanov, P. Terán, Distance transforms for real-valued functions, *Journal of Mathematical Analysis and Applications* 278 (2) (2003) 472–484, [https://doi.org/10.1016/S0022-247X\(02\)00719-9](https://doi.org/10.1016/S0022-247X(02)00719-9).
- [26] J. Chanot, Efficient morphological characterization of materials using distance transforms, PhD thesis, 2019.
- [27] P.J. Toivanen, New geodesic distance transforms for gray-scale images, *Pattern Recognition Letters* 17 (5) (1996) 437–450, [https://doi.org/10.1016/0167-8655\(96\)00010-4](https://doi.org/10.1016/0167-8655(96)00010-4).
- [28] J. Neyman, On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection 97 (4) (1934) 558–625. doi:10.2307/2342192.
- [29] I. Arganda-Carreras, S.C. Turaga, D.R. Berger, D. Cireşan, A. Giusti, L.M. Gambardella, J. Schmidhuber, D. Laptev, S. Dwivedi, J.M. Buhmann, T. Liu, M. Seyedhosseini, T. Tasdizen, L. Kametsky, R. Burget, V. Uher, X. Tan, C. Sun, T.D.



- Pham, E. Bas, M.G. Uzunbas, A. Cardona, J. Schindelin, H.S. Seung, Crowdsourcing the creation of image segmentation algorithms for connectomics, *Frontiers in Neuroanatomy* 9 (2015) 142, <https://doi.org/10.3389/fnana.2015.00142>.
- [30] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization (2014). arXiv:1412.6980..
- [31] D.J. Eyre, G.W. Milton, A fast numerical scheme for computing the response of composites using grid refinement, *The European Physical Journal – Applied Physics* 6 (1) (1999) 41–47, <https://doi.org/10.1051/epjap:1999150>.
- [32] G. Matheron, *Random Sets and Integral Geometry*, G. Matheron, Wiley, New York, 1974.
- [33] M. Moreaud, J. Chaniot, T. Fournel, J.M. Becker, L. Sorbier, Multi-scale stochastic morphological models for 3d complex microstructures, 2018 17th Workshop on Information Optics (WIO), 2018, pp. 1–3..
- [34] D. Jeulin, M. Moreaud, Multi-scale simulation of random spheres aggregates – application to nanocomposites, in: *Proc. 9th European Congress on Stereology and Image Analysis* 1, 2005, pp. 341–348.
- [35] H. Moulinec, P. Suquet, A fft-based numerical method for computing the mechanical properties of composites from images of their microstructures, in: R. Pyrz (Ed.), *IUTAM Symposium on Microstructure-Property Interactions in Composite Materials*, Springer Netherlands, Dordrecht, 1995, pp. 235–246.
- [36] H. Moulinec, P. Suquet, A numerical method for computing the overall response of nonlinear composites with complex microstructure, *Computer Methods in Applied Mechanics and Engineering* 157 (1) (1998) 69–94.
- [37] D. Jeulin, M. Moreaud, Statistical representative volume element for predicting the dielectric permittivity of random media, in: D. Jeulin, S. Forest (Eds.), *11th International Symposium on Continuum Models and Discrete Systems CMDS 11*, Sciences de la matière, Presses des mines, Paris, France, 2007, pp. 429–436, ISBN: 078-2-35671-000-0.
- [38] “plug im!” an open access and customizable software for signal and image processing (2020). <https://www.plugim.fr>.
- [39] Zhou Wang, A.C. Bovik, A universal image quality index, *IEEE Signal Processing Letters* 9 (3) (2002) 81–84.
- [40] Z. Wang, A.C. Bovik, H.R. Sheikh, Structural similarity based image quality assessment, 2004.
- [41] W. Shen, B. Wang, Y. Jiang, Y. Wang, A. Yuille, Multi-stage multi-recursive-input fully convolutional networks for neuronal boundary detection (2017) 2410–2419. doi:10.1109/ICCV.2017.262.



**Adam Hammoumi** received his engineering degree from National School of Arts and Crafts Meknes, Morocco, in 2019. Currently, he is pursuing the Ph.D. degree from Paris-Saclay university in Image processing and Material sciences at IFP Energies nouvelles (France). His research interests include image processing, machine learning and numerical models for microstructures of materials.



**Maxime Moreaud** received his PhD degree in Mathematical Morphology from Ecole des Mines de Paris (France) in 2006. He is currently project leader for the Scientific Division of IFP Energies nouvelles, a major research player in the fields of energy, transport and the environment. He is associate researcher at Center of Mathematical Morphology research laboratory from Mines ParisTech and PSL Research University. His research interests concern image processing and materials science.



**Christophe Ducottet** received his PhD degree in 1994. He is currently full Professor at Saint-Etienne University with teaching activities at Telecom Saint-Etienne and research at Hubert Curien Laboratory. His research interests concern image processing and computer vision: feature extraction, image representation, visual tracking and machine learning.



**Sylvain Desroziers** received his PhD in Applied Mathematics from Pierre and Marie Curie University (UPMC), France, in 2006. He is a research engineer at IFPEN in the Applied Mathematics and Data Science laboratory. Its main areas of interest are high-performance computing and statistical deep learning for industrial applications.