Blending variational approaches and deep learning to enforce prior constraints in medical image segmentation

Zoé Lambert, Caroline Petitjean, Carole Le Guyader

Less-Supervised Segmentation with CNNs

1st edition

Dedication

Contents

Biography

SIR STANLEY DAVIDSON (1894–1981)

One or more blank lines denote the end of a paragraph. The ends of words and sentences are marked by spaces. It doesn't matter how many spaces you type; one is as good as 100. The end of a line counts as a space.

Chapter 1

Blending variational approach and deep learning to enforce prior constraints in medical image segmentation

Zoé Lambert, Caroline Petitjean, Carole Le Guyader

ABSTRACT

Image segmentation plays a crucial role in computer vision, particularly in medical image analysis. While convolutional neural networks (CNNs) are the current state-of-the-art for segmentation, they may generate artifacts, such as isolated pixels or holes, when object contours are not well defined. Therefore, incorporating prior knowledge into the segmentation process is essential, whether through topological prescriptions (e.g., the number of connected components, the partial convexity of an object boundary) or geometrical constraints (e.g., penalizing volume through constraints), particularly when preserving contextual relationships between objects and achieving a segmentation that is homeomorphic to a known a priori is desired. In this paper, we propose two hybrid variational/deep learning frameworks that include geometric and topological constraints in the training of CNNs, in the form of two penalty terms in the loss function. The first model incorporates geometric constraints via a regularization based on weighted total variation, a volume/area penalty, and a Mumford-Shah like term. In the second model, we view the segmentation process as a registration task that pairs the ground truth and the image to be labelled, based on non-linear elasticity principles. Incompressibility conditions stated in terms of the Jacobian matrix determinant of the deformation ensures volume and topology preservation, without any material self-intersection. We evaluate the proposed losses in the challenging problem of thoracic organs at risk segmentation in Computed Tomographic scans, where some organs lack contrast, and their borders are ill-defined. Our ablation study demonstrates that our methods provide significant improvements over existing unconstrained approaches, in terms of both quantitative criteria such as the measurement of region overlap and qualitative assessment, particularly when dealing with unbalanced classes.

KEYWORDS

One, Two, Three

1.1 INTRODUCTION

Image segmentation is a critical step in medical image analysis. Today, deep convolutional neural networks (CNN), and especially the U-Net [1] architecture and its variant [2], are the state-of-the-art in automated medical image segmentation. The process consists in learning the weights of the network through the optimization of a differentiable loss function, such as the Dice loss or the cross-entropy, with a gradient descent technique. More accurate segmentations can be achieved, if some prior knowledge about the object to be segmented is incorporated in the process. Embedding prior knowledge on the object, such as shape, appearance or location, into the deep learning networks, can increase the network robustness and accuracy and even generate more anatomically plausible segmentations [3, 4]. In addition, in the medical domain, annotated images are costly to obtain, and weakly- or semi-supervised learning approaches are popular. In this regard, integrating prior knowledge can also help to mitigate the need for fully annotated images [5, 6, 7].

In this paper, we propose two loss functions, that enforce geometrical constraints for the former, and topological constraints for the latter, in the training of the CNN. Firstly, inspired by the work of Peng *et al.* [8], we build a loss that encodes several priors, the goal being that (i) the pixel intensity inside a delineated region should be homogeneous, (ii) the edges between the image region and the segmentation maps should be aligned through a weighted total variation term, (iii) the area of the segmented region should match the area of the ground truth. These priors are regularization terms or hard constraints in the loss function, added to a data-driven loss term, here the Dice loss. Our second objective is to introduce prior knowledge of a topological nature, for example, to preserve the number of connected components of the objects to be segmented when training convolutional neural networks. In this perspective, based on nonlinear elasticity concepts, we design a loss function in which segmentation is apprehended as a registration task between the predicted segmentation and the ground truth. In addition, the application of an incompressibility condition through a hard constraint ensure that the mapping between the segmentation and the ground truth is homeomorphic. In both cases, minimizing the designed loss function yields a non-smooth non-convex optimization problem. Auxiliary variables are introduced together with a splitting approach so as to fragment the initial problem into more easily resolvable sub-problems, yielding an easy implementable alternating algorithm since we can find explicit solutions for each involved variable.

In the following, we present the related works in prior based segmentation for medical imaging and introduce the two loss functions that we propose. Then in the experimental section, we assess these loss functions for the segmentation of thoracic organs at risk (OARs) in CT (computed tomography) images and propose ablation study to evaluate their added value.

1.2 RELATED WORKS

Before deep learning, variational approaches for image segmentation have made use of a variety of prior information, such as prior knowledge about the shape [9], topology preservation guarantee [10], prescription of the number of connected components/holes [11], (partial) convexity [12], or organ size range, and have proved that such priors help to achieve more accurate results [13]. Furthermore, the regularization terms themselves constitute a prior knowledge on the expected segmentations (smooth, piecewise constant, etc.). In segmentation, atlas registration techniques are very efficient to insert prior knowledge. This method consists in searching for the optimal deformation between an image to be segmented and an atlas image, to which its ground truth is associated. Once the transformation matching these two images is obtained, it is applied to the labelled atlas to obtain the final segmentation map of the image initially considered. If the transformation is a homeomorphism, it guarantees in particular the preservation of the topology and orientation of the atlas, and the obtained segmentation benefits from good properties. Therefore, a well-defined registration model and a correctly chosen atlas are crucial to achieve an accurate result. In [14], a method that relies on an atlas, in addition to other techniques, is used to segment 17 OARs throughout the body. Segmentation of thoracic OARs is achieved in [15] by combining a deformable multi-atlas registration with a local search relying on a level-set method.

Although the integration of prior knowledge is problematic with CNNs, especially because of the differentiability criterion required for backpropagation, it is becoming increasingly desirable. Thus several methods have emerged to solve this problem, often inspired by techniques from a variational framework.

A first way to incorporate prior information into the process is to use a CNN with traditional shape model and deformable model methods in a sequential prediction/correction type scheme. Thus, the CNN first provides a segmentation prediction that is then post-processed and corrected in order to comply with prescribed constraints. Conversely, the CNN can perform a pre-processing useful to the segmentation method applied in a second step. Bohlender *et al.*. [16] review many papers that use convolutional neural networks as a pre- or post-processing step. For example, Rupprecht *et al.* [17] use an active contour model [18] relying on contour detection, to provide coarse segmentations of the left ventricle that are then refined by a CNN. In contrast, Kamnitsas *et al.* [19] improve the brain lesion segmentation results obtained with a CNN by using conditional random fields (CRFs, *Conditional Random Fields*) [20]. CRFs are able to take into account the context of the image thanks to the use of global knowledge (contours, intensities, etc.).

Another solution consists in directly modifying the architecture of the neural network to take into account additional information directly in the learning

process [21, 22, 23]. This alternative may allow the variational and deep learning frameworks to be intertwined. In particular, Liu *et al.* develop a new softmax activation function in [24], based on a dual variable, to regularize the CNN outputs through the total variation or a volume constraint.

A third alternative for incorporating prior knowledge is to design a loss function that can integrate it. As an example in [25], Chen *et al.* propose a supervised loss function inspired by the Chan-Vese global energy, to integrate area and size information, with the aim of segmenting different parts of the heart. Kim and Ye define a new loss function based on the Mumford-Shah functional to guarantee the homogeneity of the light intensity of the pixels and to regularize the contour length. They apply it to the segmentation of brain and liver tumors. In addition, there are other types of loss functions that allow for the incorporation of spatial information or knowledge. El Jurdi *et al.* [26] present many of these and classify them according to the nature of the prior (shape, size, topology, or inter-regions of objects). For example, Ganaye *et al.* [27] introduce a cost function that penalizes anatomically incorrect adjacency relations in a brain MRI and whole-body CT image context. In the work of Clough *et al.* [28], prior knowledge of topology is incorporated into the objective function and experimented with, among other things, cardiac imaging. Kervadec *et al.* [29] establish an objective function to enforce size inequality constraints in the context of cardiac images. These penalty-based methods are very simple to use but do not guarantee constraint satisfaction. Thus, Dolz *et al.* [30] propose to constrain the loss function with a shape compactness information and solve the optimization problem via an alternating direction multiplier method (ADMM) algorithm with the goal of segmenting the aorta, esophagus, or right ventricle. Peng *et al.* [8] also use this algorithm to apply size constraints and contour length regularization in learning a CNN. In [31], the authors suggest instead intertwining the variational and deep learning frameworks. Specifically, ADMM algorithms are used as a substitute for gradient descent in the training step.

1.3 WEIGHTED TOTAL VARIATION-BASED AND AREA-CONSTRAINED LOSS FUNCTION

1.3.1 Model formulation

We consider a dataset of K 2D images to be segmented and associated with their ground truth that we denote by $\left\{y^k\right\}_{k=1,\cdots,K}$ with $y^k \in \{1, \cdots, L\}$ where L is the number of classes. Training a CNN will yield a segmentation function parameterised by θ such that its output is the foreground probability $(s^{l,k}(\theta))_{i,j}$ at each pixel (i, j) of the discrete image domain $\mathcal{G}$ that must be compared with the ground truth. Thus $y^{l,k} \in \{0, 1\}$ represents the one-hot encoding of the ground truth with $l \in \{1, \cdots, L\}$. The segmentation problem being separable with respect to variable k, we omit the dependency in k from now on. The

objective of the process is to obtain a probability map as close as possible to the ground truth through the backpropagation of the error of a carefully designed loss function.

Therefore, we formulate a new optimization problem based on the discrete version of the weighted total variation (TV_{g^l}) to promote contour alignment between the predicted segmentation and the ground truth subject to discrete constraints to increase their proximity in the learning phase. We denote by g^l the edge detector function applied to the norm of the ground truth gradient, of the form $g^l(\zeta) = \frac{1}{1+\beta|\zeta|^2}$, where β is a parameter tuning the contour detection level.

Regarding hard constraints, we restrict the obtained segmentation area, $\sum_{(i,j)\in\mathcal{G}} (s^l(\theta))_{i,j}$, to being equal to the area of the ground truth α^l, which is a prior information since it is calculated from the ground truth. Then we impose $(s^l(\theta))_{i,j}$ to be binary. Finally, to preserve the probability property and thus prevent a pixel from belonging to several classes, we impose that the sum of the l probabilities $(s^l(\theta))_{i,j}$ is equal to 1 at each pixel (i, j).

These elements are included in the loss function whose associated minimization problem is formulated as follows

$$\begin{aligned} &\inf_{\theta} \mathcal{L}(\theta) = \mathcal{F}_{DL}(s(\theta), y) + \sum_{l=1}^{L} TV_{g^l}(s^l(\theta)) \\ &\text{s.t } \ (s^l(\theta))_{i,j} \in \{0,1\}, \ \sum_{(i,j)\in\mathcal{G}} (s^l(\theta))_{i,j} = \alpha^l \text{ and } \sum_{l=1}^{L} (s^l(\theta))_{i,j} = 1, \end{aligned} \quad (1.1)$$

with $\mathcal{F}_{DL}$ representing a classical deep learning function for image segmentation such that Dice score or cross-entropy, that can be combined with a Mumford-Shah-like fidelity term [32, 33] that enforces intensity homogenization of the regions, defined as:

$$\sum_{l=1}^{L} \sum_{(i,j)\in\mathcal{G}} (y^l_{i,j} - l)^2 (s^l(\theta))_{i,j}.$$

In order to solve the problem and inspired by the work of [8], we introduce an auxiliary variable u to simulate $s(\theta)$ and the minimization problem can be equivalently rephrased as:

$$\begin{aligned} &\inf_{\theta, u=(u^l)} \mathcal{L}(\theta, u) = \mathcal{F}_{DL}(s(\theta), y) + \sum_{l=1}^{L} TV_{g^l}(u^l) \\ &\text{s.t } u^l = s^l(\theta),\ u^l_{i,j} \in \{0,1\}, \ \sum_{(i,j)\in\mathcal{G}} u^l_{i,j} = \alpha^l \text{ and } \sum_{l=1}^{L} u^l_{i,j} = 1. \end{aligned} \quad (1.2)$$

This problem is solved using an augmented Lagrangian method, more precisely the scaled form [34], stated as:

$$\begin{aligned} \max_{w=(w^l)} \min_{\theta, u=(u^l)} \quad & \mathcal{L}(\theta, u, w) = \mathcal{F}_{DL}(s(\theta), y) + \sum_{l=1}^{L} TV_{g^l}(u^l) + \frac{\mu}{2} \|s^l(\theta) - u^l + w^l\|^2 \\ \text{s.t} \quad & u^l_{i,j} \in \{0,1\}, \sum_{(i,j)\in\mathcal{G}} u^l_{i,j} = \alpha^l \text{ and } \sum_{l=1}^{L} u^l_{i,j} = 1, \end{aligned} \tag{1.3}$$

where $\mu > 0$ is the augmented Lagrangian parameter. Now the next section describes how to find the three parameters θ, u and w.

1.3.2 Updating of parameters θ, u and w

An efficient method to solve this optimization problem is the ADMM algorithm [34] which consists in updating one of the variables while the others are considered fixed. Algorithm 1 shows the principle of this method. We now detail how to update the three parameters θ, u and w.
First, to update the network parameters θ, we simply use a mini-batch gradient descent technique on the loss (the functional to be minimized with respect to θ being smooth (non-convex) here) that reduces to :

$$\bar{\mathcal{L}}(\theta) = \mathcal{F}_{DL}(s(\theta), y) + \frac{\mu}{2} \sum_{l=1}^{L} \|s^l(\theta) - u^l + w^l\|^2. \tag{1.4}$$

The gradient of this loss function is given by

$$\nabla_\theta \bar{\mathcal{L}}(\theta) = \nabla_\theta \mathcal{F}_{DL}(s(\theta), y) + \mu \sum_{l=1}^{L} \left(s^l(\theta) - u^l + w^l\right) \nabla_\theta s^l(\theta).$$

The parameters are then updated by taking a step opposite to the gradient, *i.e.*, $\theta_{n+1} = \theta_n - \eta \nabla_\theta \bar{\mathcal{L}}(\theta_n)$, where η is the learning rate and n represents the set of all successive steps of the mini-batch stochastic gradient descent technique performed during this epoch.

About the update of $u = (u^l)$, we again seek to decompose the original difficult problem into more tractable problems. Thus we introduce a variable $v = (v^l)$ related to the area constraint and such that v^l is close to u^l. This soft coupling between u^l and v^l is achieved through an L^2-penalization. The subproblem is thus stated as

$$\min_{\substack{u=(u^l)\in\mathcal{C}_1 \\ v=(v^l)\in\mathcal{C}_2}} \mathcal{J}(u, v) = \sum_{l=1}^{L} \left(TV_{g^l}(u^l) + \frac{\mu}{2} \|s^l(\theta) - u^l + w^l\|^2 + \frac{1}{2} \|u^l - v^l\|^2 \right), \quad (\mathcal{P})$$

with C_1 and C_2 the closed convex sets defined by

$$C_1 = \left\{ u = (u^l) \,|\forall l \in \{1, \cdots, L\}, \forall (i,j) \in \mathcal{G}, u^l_{i,j} \in [0,1] \text{ and } \sum_{l=1}^{L} u^l_{i,j} = 1 \right\}$$

and

$$C_2 = \left\{ v = (v^l) \,|\, \forall l \in \{1, \cdots, L\}, \forall (i,j) \in \mathcal{G}, \sum_{(i,j)\in\mathcal{G}} v^l_{i,j} = \alpha^l \right\}.$$

To avoid the non-differentiability issue inherent to the TV_g we now consider its dual formulation which allows us to rewrite this term as:

$$\sum_{l=1}^{L} TV_{g^l}(u^l) = \max_{p=(p^l)} \langle \nabla u, p \rangle - i_{\mathcal{B}}(p),$$

with p being the dual variable and $\mathcal{B}$ the convex set given by

$$\mathcal{B} = \left\{ p = (p^l) \,|\, \forall l \in \{1, \cdots, L\}, \forall (i,j) \in \mathcal{G}, |p^l_{i,j}| \leq g^l_{i,j} \right\}.$$

These elements allow us to rephrase our optimization problem $(\mathcal{P})$ as a min-max one as follows

$$\min_{\substack{u=(u^l)\in C_1 \\ v=(v^l)\in C_2}} \max_{p=(p^l)\in\mathcal{B}} \mathcal{L}(u,v,p) := \langle \nabla u, p \rangle + \frac{\mu}{2} \|u - s(\theta) - w\|^2 + \frac{1}{2} \|u - v\|^2. \tag{$\bar{\mathcal{P}}$}$$

Our problem $(\bar{\mathcal{P}})$ falls within the general framework of convex optimization problems with known saddle-point structure addressed by Chambolle and Pock in [35] and phrased as

$$\min_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} \mathcal{L}(x,y) = \langle Kx, y \rangle + f(x) + g(x) - h^*(y). \tag{CP}$$

Actually, problem $(\bar{\mathcal{P}})$ is a special instance of (CP) with $x = (u, v)$, $y = p$, $K = \begin{pmatrix} \nabla & 0 \end{pmatrix}$, $f(u,v) = \frac{1}{2}\|u - v\|^2$, $g(u,v) = \frac{\mu}{2}\|u - s(\theta) - w\|^2 + i_{C_1}(u) + i_{C_2}(v)$, and $h^* = i_{\mathcal{B}}$, where i_C stands for the indicator function of a subset C. As such, this problem can be solved by [35, Algorithm 1] which is based on the proximal operators [36].
In a nutshell, our method to solve the subproblem related to the variable u, referred to as PD in Algorithm 1, is based on a Primal-Dual formulation and gives explicit solutions. The whole method and solutions are detailed in [37] along with theoretical results.

Finally, the Lagrangian variable w is simply updated thanks to a gradient ascent technique, *i.e.*, $w_{n+1} = w_n + \mu(s(\theta_{n+1}) - u_{n+1})$.

Before presenting the experimental results obtained with this loss function, we introduce a second one focused on topological aspects.

Algorithm 1 ADMM algorithm to update θ, u and w

Initialize θ_0 randomly and set $u_0 = w_0 = 0$
Fix $\mu > 0$
for $n = 0, 1, \cdots$ **do**
  Compute $s(\theta_n)$
  Set $\bar{\mathcal{L}}(\theta_n) = \mathcal{F}_{DL}(s(\theta_n), y) + \frac{\mu}{2}(s(\theta_n) - u_n + w_n)^2$
  $\theta_{n+1} = \theta_n - \eta \nabla_\theta \bar{\mathcal{L}}(\theta_n)$
  $u_{n+1} = PD(s(\theta_{n+1}), u_n, w_n)$
  $w_{n+1} = w_n + (s(\theta_{n+1}) - u_{n+1})$
end for

1.4 CONSTRAINED REGISTRATION-BASED LOSS FUNCTION

1.4.1 Model formulation

The second model for incorporating prior information into a supervised deep learning model is inspired by image registration. Indeed, here, in order to force the predicted segmentation and the ground truth to exhibit the same number of connected components, and thus increase the similarity between both, we introduce a transformation of the ground truth, which is sought as close as possible to the identity. Thus we aim to formulate a new loss function optimization problem that performs the segmentation and registration tasks jointly.
The envisaged functional consists of several terms. First, we construct a simple quadratic data fidelity term to quantify, and thus penalize the dissimilarity between the prediction and the ground truth transformation. Then we define a regularization term related to the deformation φ in order to monitor the smoothness and to control the changes in lengths. Finally, the introduction of a condition concerning the determinant $\det \nabla\varphi$ not only forces the prediction to exhibit the same number of connected components as y but also constrains the area and orientation to be preserved. This is formalized by the hard constraint $\det \nabla\varphi = 1$. In addition, these elements are obviously combined with a fitting term for segmentation. The minimization problem for the loss function thus reads:

$$\begin{aligned} \inf_{\theta,\varphi} \quad & \mathcal{L}(\theta, \varphi) = \mathcal{F}_{DL}(s(\theta), y) + \frac{\nu}{2} \|s(\theta) - y \circ \varphi\|^2 + \mu \|\nabla\varphi\|^4 \\ \text{s.t} \quad & \det \nabla\varphi = 1, \end{aligned} \tag{1.5}$$

where ν and μ are weighting terms and $\mathcal{F}_{DL}$, as before, a standard function for segmentation in deep learning.

The idea behind this formulation is that the topological constraints are implicitly applied by apprehending the segmentation assignment as a registration task between the considered image and its associated ground truth: the constraints prescribed on $\det \nabla\varphi$ are now a substitute for the explicit monitoring of topology.

1.4.2 Updating of parameters θ and φ

To solve this problem a splitting technique proves to be a suitable approach. Algorithm 2 shows the steps of the process, which consists in updating each variable while considering the other one fixed. In this way we obtain two subproblems related to the variables θ and φ that are easier to solve.
In the same way as for the first method, the network parameters θ are updated using a stochastic gradient descent technique on the loss function reduced to

$$\bar{\mathcal{L}}(\theta) = \mathcal{F}_{DL}(s(\theta), y) + \frac{\nu}{2}\|s(\theta) - y \circ \varphi\|^2. \tag{1.6}$$

The subproblem related to φ exhibits numerical difficulties due to non-linearity and non-convexity. To overcome this problem, the idea is to introduce auxiliary variables to separate again the initial problem into more easily solvable sub-problems. We therefore introduce two auxiliary variables V and W such that $V = \nabla\varphi$ and $W = \nabla\varphi$, the coupling occurring through an L^2-penalty. The hard constraint is maintained on W, while relaxed on V. This enables us to reformulate the problem with respect to the variables φ, V, and W, as follows

$$\begin{aligned}\min_{\varphi,V,W} \mathcal{J}(\varphi, V, W) = \mu\|V\|^4 + \frac{\nu}{2}\|s(\theta) - y \circ \varphi\|^2 + \frac{\mu\alpha}{2}\|\det V - 1\|^2 \\ + \frac{\gamma_1}{2}\|W - \nabla\varphi\|^2 + \frac{\gamma_2}{2}\|V - \nabla\varphi\|^2 \quad \text{s.t } \det W = 1,\end{aligned} \tag{1.7}$$

where α, γ_1 and γ_2 are positive parameters.

We then obtain three subproblems that are easily solved by using a splitting technique again. Indeed, closed form solutions can now be given for each of the three variables. We provide the main elements for each of them:

1. for V, the first-order optimal conditions give a 4-equation system that can be reduced to a 2 equation-optimality system by a collinearity property. Then by setting $\alpha = 8$, this system is decoupled and the problem amounts to solving cubic equations for which explicit solutions are available;
2. for W, we draw inspiration from the work of Glowinski [38]. The theory of Lagrange multipliers allows to obtain the explicit expression of the —unique in the general case —solution of the problem;
3. for φ, a Discrete Sinus Transform method is applied on the solutions of the Euler-Lagrange equations given by

$$-\nu(s(\theta) - y \circ \varphi)\nabla y(\varphi) + \gamma_1 \mathrm{div}\, W + \gamma_2 \mathrm{div}\, V - (\gamma_1 + \gamma_2)\Delta\varphi = 0.$$

We are now ready to study the segmentation results obtained with our two loss functions.

1.5 EXPERIMENTS AND RESULTS

When planning a radiotherapy treatment, it is necessary to segment the target tumor as well as adjacent healthy organs (so-called organs at risk). In the case

Algorithm 2 Splitting algorithm to update θ and φ

Initialize θ_0 randomly and set $\varphi = \mathrm{Id}$
Fix $\mu,\ \nu > 0$
for $n = 0, 1, \cdots$ **do**
  Compute $s(\theta_n)$
  $\varphi_n = \arg\min_{\varphi} \frac{\nu}{2} \|s(\theta_n) - y \circ \varphi\|^2 + \mu \|\nabla\varphi\|^4$ such that $\det \nabla\varphi = 1$
  Set $\bar{\mathcal{L}}(\theta_n) = \mathcal{F}_{DL}(s(\theta_n), y) + \frac{\nu}{2} \|s(\theta_n) - y \circ \varphi_n\|^2$
  $\theta_{n+1} = \theta_n - \eta \nabla_\theta \bar{\mathcal{L}}(\theta_n)$
end for

of lung cancer, the organs at risk are located in the thorax and include usually the esophagus, the heart, the trachea and the aorta. The segmentation of this 4 OARs is of particular interest to evaluate the proposed losses: the segmentation of the heart, aorta and especially the esophagus is complex due to the lack of contrast. In addition, the shape and position of the esophagus vary greatly from patient to patient. The aorta, due to its cane shape, has one or two connected components depending on the slice. The segmentation of the trachea is expected to be easier as it is filled with air and thus appears as clearly black on CT images.

1.5.1 Dataset

The SegTHOR dataset consists of 60 thoracic CT scans, acquired with or without intravenous contrast, of 60 patients diagnosed with lung cancer or Hodgkin's lymphoma and treated at the Henri Becquerel Center, Rouen, France. The segmentations of the esophagus, the heart, the trachea and the aorta were manually drawn by an expert. All scanner images are 512×512×(150 ~ 284) voxels in size. Indeed, the number of slices changes according to the patients. The in-plane resolution varies between 0.90 mm and 1.37 mm per pixel and the z-resolution fluctuates between 2 mm and 3.7 mm per pixel. Finally, the most common resolution is 0.98×0.98×2.5 mm^3.

1.5.2 Implementation

In all our experiments, the network architecture that we used is a simplified version of U-Net, called sU-Net, previously introduced in [39] and shown in Fig. 1.1, that has fewer dense connected layers than the original one. We evaluate the loss function $\mathcal{F}_{DL}$ as, on the one hand, the Dice loss only, and on the other hand, combined with the Mumford-Shah (MS) term. For computational reasons, images are normalized and cropped from the center to obtain images of size 304×240 pixels. As data augmentation has clearly been shown to boost the performance in this application [39], we take advantage of it in the follow-

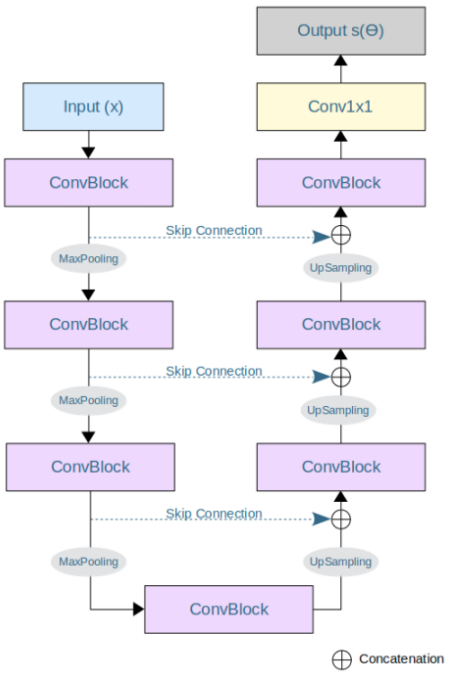

FIGURE 1.1 The network architecture of sU-Net, a simplified version of U-Net, where each ConvBlock consists of two convolution operations with a 3×3 kernel size followed by a ReLU activation function and a batch normalization. The spatial resolution is reduced by max-pooling and recovered by bilinear up-sampling.

ing. Thus data augmentation is implemented and used to artificially triple the number of training images. As described in [40], each image is modified by a random affine transform on the one hand, and is randomly deformed using a dense deformation field obtained through a $2 \times 2 \times 2$ grid of control-points and B-Spline interpolation on the other hand. The parameters of sU-Net are randomly initialized thanks to Glorot's initialization technique [41] and updated using SGD. The network hyperparameters are optimized following a grid search technique, in particular the initial learning rate is fixed at 10^{-3} and the batch size at 4 for all the experiments described. With the same procedure, the different parameters involved in each method are set as in Table 1.1, except parameters σ and τ of the Primal-Dual algorithm that are fixed to comply with the conditions stated in [37, Theorem 4].

At last, the whole training process takes place over a maximum of 150 epochs for the first method (Algorithm 1), while the second one (Algorithm 2) only requires a maximum of 100 epochs. The whole code is developed with Pytorch.

TABLE 1.1 Parameters for our two proposals fixed via grid search technique.

First method			Second method			
μ = 0.5	σ = 0.01	τ=0.4	γ_1 = 10000	γ_2 = 80000	μ = 125	ν = 2

1.5.3 Protocol and evaluation metrics

The dataset is split in two: 40 CT scans are used for training and the remaining 20 are kept for inference, following the standardized protocol given in the SegTHOR challenge. In this section, all segmentation results are evaluated in 3D per patient, using two well-known metrics, namely the Dice score and the Hausdorff distance, this latter measuring quantifying the ability of the model to remove outliers. These indicators should be assessed in the light of more qualitative criteria (spatial coherence/ regularity, compliance with prescribed geometrical constraints, segmentations more faithful to anatomical reality, fewer excrescences and false detections, etc.) since the Dice coefficient for instance, by its underlying averaging effect, tends to mask the visual disparities that can be observed between the constrained and unconstrained cases.

1.5.4 Results

In the results reported in Table 1.2, we present the automatic segmentations obtained in three cases: (i) with a classical unconstrained deep learning approach, *i.e.* with the same architecture sU-Net, in which weights are only updated by backpropagation of the gradient over a maximum of 200 epochs, out of any other optimization scheme ('no constraint' rows in Tab. 1.2); (ii) with our first proposal ('geom. const.') and (iii) with our second proposal ('topo. const.'). For each case, we also assess the contribution of the MS term, in the 'Dice' vs 'Dice + MS' rows in Tab. 1.2. In addition, the segmentation results of four different patients are shown in Figure 1.2.

The trachea (light green in Fig. 1.2) is easily distinguishable in the CT image, hence Dice values and Hausdorff distance are similar for all methods. Similarly, for the aorta (in yellow in Fig. 1.2), Dice values are fairly equivalent for all six methods whereas the Hausdorff distance proves to be lower thanks to the MS term or topological loss. The anomalies that these models are able to correct are illustrated in Fig. 1.2, first and fourth columns. For the last two OARs, the heart and the esophagus, the use of the MS term in combination with geometric or topological constraints improves the Dice values as well as the Hausdorff distance. In the following, we deepen our analysis on these 2 organs.

The heart (dark green in Fig. 1.2) is a large organ that encompasses many different tissues and structures, making its contours blurred and segmentation eventually difficult. This explains why the unconstrained method without the homogeneity term results in a largely over-segmented organ as shown in Fig

TABLE 1.2 Segmentation results (mean±stdev) without and with the proposed geometrical constraints. $\mathcal{F}_{DL}$ is the standard loss function for semantic segmentation. Best significant results are in bold.

Method	$\mathcal{F}_{DL}$	Metrics	Aorta	Heart	Trachea	Esophagus
sU-Net (no constraint).	Dice	DSC	94.17 ± 2.28	85.44 ± 5.26	90.46 ± 2.51	81.14 ± 4.96
		HD	16.25 ± 12.65	116.14 ± 39.19	19.59 ± 8.28	49.25 ± 32.29
	Dice + MS	DSC	94.39 ± 1.85	91.97 ± 3.36	90.73 ± 2.44	82.76 ± 4.45
		HD	10.53 ± 6.66	28.57 ± 18.11	19.06 ± 9.38	35.01 ± 24.64
Proposed: sU-Net + geom. const.	Dice	DSC	94.39 ± 1.93	91.40 ± 3.34	90.53 ± 2.39	82.02 ± 4.71
		HD	16.64 ± 12.09	36.90 ± 17.25	18.96 ± 8.60	34.72 ± 24.87
	Dice + MS	DSC	**94.75 ± 1.53**	**93.04 ± 3.00**	91.05 ±2.55	82.91 ± 4.73
		HD	9.73 ± 5.54	20.47 ± 10.52	18.05 ± 9.63	23.97 ± 17.69
Proposed: sU-Net + topo. const.	Dice	DSC	94.31 ± 2.29	91.94 ± 4.32	90.98 ± 2.41	82.61 ± 4.73
		HD	9.98 ± 5.94	27.47 ± 23.65	18.77 ± 7.84	31.03 ± 27.48
	Dice + MS	DSC	94.40 ± 2.11	92.88 ± 3.28	91.43 ±2.25	82.79 ± 5.28
		HD	8.98 ± 4.67	**19.53 ± 8.61**	**16.03 ± 6.63**	**14.47 ± 6.66**

.1.2-(c). Adding the MS term or constraints and spatial regularization clearly improves the results quantitatively and qualitatively (see rows d,e and g in Fig. 1.2). The combination of these two elements corrects persistent outliers (row f and h). Indeed, an average Dice score around 93% is achieved, *i.e.* 1 to 2 pp higher than the three cases with MS term and over 7 pp higher than the standard unconstrained method. The result is even more obvious for the average Hausdorff distance where it drops close to 20mm, *i.e.* about 6 times less than the unconstrained methods without MS and is practically halved compared to classical approach with the MS term or the constrained method without the MS term. These analyses are supported by a paired Wilcoxon signed rank-test which shows that the above-mentioned results (in bold in Tab 1.2) are significantly different between unconstrained and constrained approaches without and with the MS term (p-value < 0.05). This reflects the qualitative and fine properties that can be observed in the constrained case as opposed to the unconstrained case, namely segmentation more faithful to anatomical reality, fewer excrescences and false detection.

As previously explained, the esophagus (blue in Fig. 1.2) is a small organ that is particularly difficult to segment due to the lack of contrast, in addition to its high inter- and intra-patient variability. The methods tend to under-segment it but also to segment isolated pixels. The area penalization, but even more so the incompressibility condition on the determinant, help constrained methods to counteract this effect as well as the homogenization of the regions ensured by the MS term, see for instance Fig. 1.2 (columns 1 and 3), which results in a lower average Hausdorff distance.

Overall, our two constraint-inclusive methods combined with the Dice and

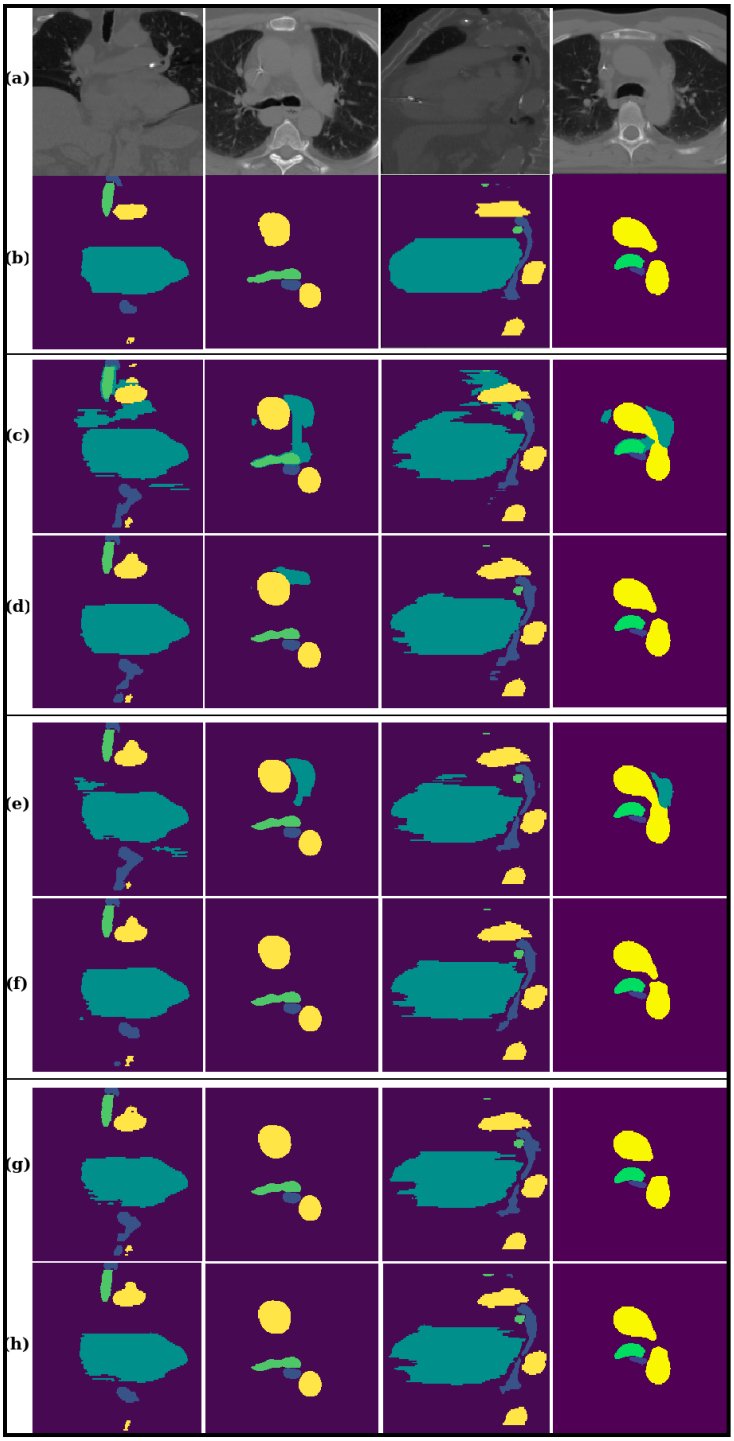

FIGURE 1.2 CT-scans (a) with associated GT (b) and segmentation results on four patients of the aorta (yellow), the trachea (light green), the heart (dark green) and the esophagus (blue) obtained with the classical DL approach (c-d) and our proposals (e-h), without and with the Mumford-Shah term.

MS terms improve the segmentation results. The first method gives slightly better Dice scores while the second one reduces the Hausdorff distances a bit more.

Usage of the proposed losses in less supervised setting. As prior losses have an interest in a weakly or unsupervised learning setting, let us examine how the proposed loss terms may be used in such settings. Regarding the first geometrical loss, the equality constraint on the surface of the objects of interest to be segmented can no longer hold ; instead, it is necessary to take into account inequality constraints that arise from a statistical study conducted in advance on the objects [29]. The Mumford-Shah based intensity homogeneity term relies on the ground truth labels, and as such cannot be used in a weakly or unsupervised learning setting. The weighted total variation also depends on the gradient of the ground truth. However, an alternative may be to apply the edge detector to the input image and to use it in the weighted total variation, as a substitute for the gradient of the ground truth - even though it is expected to be noisier.

The second topological loss is versatile enough to be used in the inference phase, and by extension in a weakly or semi-supervised learning framework. To do so, we could build an average template to register on the predicted segmentation to obtain the final result. In this case, since we are only interested in preserving the contextual relationship between the objects, the equality constraint on the determinant must be relaxed to the inequality constraint $\det \nabla\varphi > 0$, still preserving the topology.

However, a challenge arises due to the fact that topology and area vary based on the slice. For example, the shape of the aorta can result in one or two connected components, and as a result, it is necessary to develop a strategy to fix the inequality constraint on the area or to correctly match each 2D image with the appropriate template. To overcome this challenge, models could be expanded to 3D. The primary scientific challenge lies in the numerical implementation, especially for the second loss, which becomes more intricate due to the more complex expression of the energy density that involves the cofactor matrix of the Jacobian matrix $\nabla\varphi$ for controlling area variations.
An alternative approach would be to directly integrate these different elements as a connected layer into the deep network. In [24] the total variation and the constraints are integrated into the softmax function. The registration task is embedded into the network architecture in [42, 43] through the use of connected layers called diffeomorphic, allowing the authors to systematically guarantee that the targeted topology is preserved. In these works, the authors optimize the deformation parameters through a spatial transformation network [44, 45].

1.6 CONCLUSION

This paper introduces two novel loss functions that use geometric and topological information to train 2D CNNs. The first loss function incorporates

geometric constraints by combining a Dice term that encourages intensity pairing, a weighted total variation term that induces edge alignment, a piecewise-constant Mumford-Shah-like term that enforces intensity homogeneity, and an area penalty. The second loss function combines Dice and Mumford-Shah terms with a functional that models a registration problem. The registration serves as a prior to guide the segmentation towards more realistic results and ensures that the predicted segmentation is homeomorphic to the ground truth. The determinant of the Jacobian matrix is constrained to preserve topology and volume, while points that cause topological errors such as related components, isolated pixels, and holes, are penalized during the learning process using an L^2 coupling. The registration allows for large deformations while considering the contextual relationships between objects in the image.

In both scenarios, we divide the optimization problem into two subproblems to be solved using a splitting strategy. The first subproblem entails optimizing the network parameters θ using a gradient descent, and the second subproblem resolution depends on the type of loss. For the geometric loss, we address the second subproblem using the Primal-Dual algorithm. For the topological loss, the second subproblem involves a constrained registration problem, which requires an efficient solution strategy based on nonlinear elasticity principles and auxiliary variables.

We evaluated the effectiveness of both loss functions on the SegTHOR dataset for organ at risk segmentation. Our results indicate improved accuracy compared to a conventional DL approach in the multi-class case of thoracic OAR, with higher Dice scores and lower Hausdorff distances. These hybrid variational-DL frameworks can be applied directly to all neural networks and applications.

1.7 ACKNOWLEDGEMENTS

This project was co-financed by the European Union with the European regional development fund (ERDF, 18P03390/18E01750/18P02733), by the Haute-Normandie Régional Council via the M2SINUM project and by the French Research National Agency ANR via AAP CE23 MEDISEG ANR project. The authors would like to thank the CRIANN (Centre Régional Informatique et d'Applications Numériques de Normandie, France) for providing computational resources.

[1] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in: MICCAI, 2015, pp. 234–241.

[2] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, S... 11045 (2018) 3–11.

[3] L. Liu, J. M. Wolterink, C. Brune, R. N. J. Veldhuis, Anatomy-aided deep learning for medical image segmentation: a review, Physics in Medicine Biology 66 (11) (2021) 11TR01. doi:10.1088/1361-6560/abfbf4.
URL https://dx.doi.org/10.1088/1361-6560/abfbf4

[4] A. Larrazabal, C. Martínez, B. Glocker, E. Ferrante, Post-dae: Anatomically plausible segmentation via post-processing with denoising autoencoders (2020).

[5] R. El Jurdi, C. Petitjean, P. Honeine, F. Abdallah, Organ Segmentation in CT Images With Weak Annotations: A Preliminary Study, in: 27th GRETSI Symposium on Signal and Image Processing, Lille, France, 2019.

[6] R. El Jurdi, C. Petitjean, P. Honeine, F. Abdallah, BB-UNet: U-Net with Bounding Box Prior, IEEE Journal of Selected Topics in Signal Processing 14 (6) (2020) 1189 – 1198.

[7] H. Kervadec, J. Dolz, S. Wang, E. Granger, I. ben Ayed, Bounding boxes for weakly supervised segmentation: Global constraints get close to full supervision, in: Medical Imaging with Deep Learning, 2020.
URL https://openreview.net/forum?id=VOQMC3rZtL

[8] J. Peng, H. Kervadec, J. Dolz, I. Ben Ayed, M. Pedersoli, C. Desrosiers, Discretely-constrained deep network for weakly supervised segmentation, Neural Netw. 130 (2020) 297 – 308.

[9] O. Alexandrov, F. Santosa, A topology-preserving level set method for shape optimization, J. Comput. Phys. 204 (1) (2005) 121 – 130.

[10] C. Le Guyader, L. A. Vese, Self-Repelling Snakes for Topology-Preserving Segmentation Models, IEEE Trans. Image Process. 17 (5) (2008) 767–779.

[11] F. Ségonne, Active contours under topology control—genus preserving level sets, Int. J. Comput. Vision 79 (2) (2008) 107–117.

[12] C. Y. Siu, H. L. Chan, L. M. Lui, Image Segmentation with Partial Convexity Shape Prior Using Discrete Conformality Structures, SIAM J. Imaging Sci. 13 (4) (2020) 2105–2139.

[13] M. S. Nosrati, G. Hamarneh, Incorporating Prior Knowledge in Medical Image Segmentation: a Survey, CoRR abs/1607.01092 (2016). arXiv:1607.01092.
URL http://arxiv.org/abs/1607.01092

[14] M. Han, Y. Ma, J.and Li, M. Li, Y. Song, Q. Li, Segmentation of organs at risk in CT volumes of head, thorax, abdomen, and pelvis, in: SPIE Medical Imaging 2015: Image Processing, Vol. 9413, 2015, p. 94133J.

[15] E. Schreibmann, D. M. Marcus, T. Fox, Multiatlas segmentation of thoracic and abdominal anatomy with level set-based local search, J. Appl. Clin. Med. Phys. 15 (4) (2014) 22–38.

[16] S. Bohlender, I. Oksuz, A. Mukhopadhyay, A Survey on Shape-Constraint Deep Learning for Medical Image Segmentation (2021).

[17] C. Rupprecht, E. Huaroc, M. Baust, N. Navab, Deep active contours, CoRR abs/1607.05074 (2016). arXiv:1607.05074.
URL http://arxiv.org/abs/1607.05074

[18] M. Kass, A. P. Witkin, D. Terzopoulos, Snakes: Active Contour Models., J. Comput. Vis. 1 (4) (1988) 321–331.

[19] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon,

D. Rueckert, B. Glocker, Efficient Multi-scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation, Med. Image Anal. 36 (2017) 61–78.

[20] J. D. Lafferty, A. McCallum, F. C. N. Pereira, Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence data, in: Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2001, p. 282–289.

[21] R. El Jurdi, C. Petitjean, P. Honeine, F. Abdallah, Bb-unet: U-net with bounding box prior, IEEE J. Sel. Top. Signal Process. 14 (6) (2020) 1189–1198.

[22] P.-A. Ganaye, M. Sdika, H. Benoit-Cattin, Towards integrating spatial localization in convolutional neural networks for brain image segmentation, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 621–625. doi:10.1109/ISBI.2018.8363652.

[23] R. Trullo, C. Petitjean, B. Dubray, S. Ruan, Multi-Organ Segmentation using Distance-Aware Adversarial Networks, J. Med. Imaging 6 (1) (2019) 014001.

[24] J. Liu, X. Wang, X.-C. Tai, Deep Convolutional Neural Networks with Spatial Regularization, Volume and Star-shape Priori for Image Segmentation (2020). arXiv:2002.03989.

[25] X. Chen, B. M. Williams, S. R. Vallabhaneni, G. Czanner, R. Williams, Y. Zheng, Learning Active Contour Models for Medical Image Segmentation, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 11624–11632.

[26] R. El Jurdi, C. Petitjean, P. Honeine, V. Cheplygina, F. Abdallah, High-level prior-based loss functions for medical image segmentation: A survey, arXiv preprint arXiv:2011.08018 (2020).

[27] P.-A. Ganaye, M. Sdika, B. Triggs, H. Benoit-Cattin, Removing Segmentation Inconsistencies with Semi-Supervised Non-adjacency Constraint, Med. Image Anal. 58 (2019) 101551.

[28] J. Clough, N. Byrne, I. Oksuz, V. A. Zimmer, J. A. Schna-bel, A. King, A Topological Loss Function for Deep-Learning based Image Segmentation using Persistent Homology, IEEE Trans. Pattern Anal. Mach. Intell. (2020).

[29] H. Kervadec, J. Dolz, M. Tang, E. Granger, Y. Boykov, I. B. Ayed, Constrained-CNN Losses for Weakly Supervised Segmentation, Med. Image Anal. 54 (2019) 88–99.

[30] J. Dolz, I. B. Ayed, C. Desrosiers, Unbiased Shape Compactness for Segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 755–763.

[31] G. Taylor, R. Burmeister, Z. Xu, B. Singh, A. Patel, T. Goldstein, Training Neural Networks Without Gradients: A Scalable ADMM Approach, in: M. F. Balcan, K. Q. Weinberger (Eds.), Proceedings of The 33rd International Conference on Machine Learning, Vol. 48 of Proceedings of Machine Learning Research, PMLR, 2016, pp. 2722–2731.

[32] D. Mumford, J. Shah, Optimal approximation by piecewise smooth functions and associated variational problems, Comm. Pure Appl. Math. 42 (5) (1989) 577–685.

[33] B. Kim, J. C. Ye, Mumford–Shah Loss Functional for Image Segmentation with Deep Learning, IEEE Trans. Image Process. 29 (2020) 1856–1866.

[34] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, Found. Trends Mach. Learn 3 (2011) 1–122.

[35] A. Chambolle, T. Pock, On the ergodic convergence rates of a first-order primal-dual algorithm, Math. Program. 159 (1) (2016) 253–287.

[36] P. L. Combettes, J.-C. Pesquet, Proximal Splitting Methods in Signal Processing, Springer New York, New York, NY, 2011, pp. 185–212.

[37] Z. Lambert, C. Le Guyader, C. Petitjean, Enforcing Geometrical Priors in Deep Networks for Semantic Segmentation Applied to Radiotherapy Planning, J. Math. Imaging Vision (2022)

1–24.

[38] R. Glowinski, P. Le Tallec, Numerical solution of problems in incompressible finite elasticity by augmented Lagrangian methods. I. Two-dimensional and axisymmetric problems, SIAM Journal on Applied Mathematics 42 (2) (1982) 400–429.

[39] Z. Lambert, C. Petitjean, B. Dubray, S. Ruan, SegTHOR: Segmentation of Thoracic Organs at Risk in CT images, in: 2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA), 2020, pp. 1–6.

[40] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), IEEE, 2016, pp. 565–571.

[41] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.

[42] A. V. Dalca, G. Balakrishnan, J. Guttag, M. R. Sabuncu, Unsupervised learning for fast probabilistic diffeomorphic registration, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 729–738.

[43] M. K. Wyburd, N. K. Dinsdale, A. I. Namburete, M. Jenkinson, TEDS-Net: Enforcing diffeomorphisms in spatial transformers to guarantee topology preservation in segmentations, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 250–260.

[44] M. Jaderberg, K. Simonyan, A. Zisserman, et al., Spatial transformer networks, Advances in neural information processing systems 28 (2015).

[45] M. C. H. Lee, K. Petersen, N. Pawlowski, B. Glocker, M. Schaap, TETRIS: Template transformer networks for image segmentation with shape priors, IEEE Trans. Med. Imaging 38 (11) (2019) 2596–2606.