A geometrically-constrained deep network for CT image segmentation

Zoé Lambert, Carole Le Guyader, Caroline Petitjean

A GEOMETRICALLY-CONSTRAINED DEEP NETWORK FOR CT IMAGE SEGMENTATION

Zoé Lambert[⋆] *Carole Le Guyader*[⋆] *Caroline Petitjean*[†]

[⋆] Normandie Univ, INSA Rouen, LMI, 76000 Rouen, France
[†]Université de Rouen Normandie, LITIS, 76801 Saint-Etienne-du-Rouvray Cedex, France

ABSTRACT

Incorporating prior knowledge into a segmentation process, whether it be geometrical constraints such as volume penalisation, (partial) convexity enforcement, or topological prescriptions to preserve the contextual relations between objects, proves to improve accuracy in medical image segmentation, in particular when addressing the issue of weak boundary definition. Motivated by this observation, the proposed contribution aims to include geometrical constraints in the training of convolutional neural networks in the form of a penalty in the loss function. These geometrical constraints take several forms and encompass level curve alignment through the weighted total variation component, an area penalisation phrased as a hard constraint in the modelling, and an intensity homogeneity criterion based on a combination of the standard Dice loss with the piecewise constant Mumford-Shah model. The mathematical formulation yields a non-smooth non-convex optimisation problem, which rules out conventional smooth optimisation techniques and leads us to adopt a Lagrangian setting. The application falls within the scope of organ-at-risk segmentation in CT (Computed Tomography) images, in the context of radiotherapy planning. Experiments demonstrate that our method provides significant improvements over existing non-constrained approaches.

***Index Terms*—** image segmentation, CT scans, deep learning, proximal operator, ADMM algorithm, Douglas-Rachford algorithm

1. INTRODUCTION

Image segmentation is a critical step in image processing on the way to make image analysis automatic. It consists in identifying meaningful constituents of a given image (textures, edges, shapes, etc.) for quantitative analysis (see [1] for an exhaustive overview). It is at the crux of many medical image processing chains, among which radiotherapy planning. The goal of radiotherapy is to irradiate the tumour with ionising beams to prevent the proliferation of cancer cells, while preserving healthy tissues and surrounding organs called Organs at Risk (OAR). Thus delimiting the target tumour and OAR on CT images is the first step in treatment planning. This task performed manually by an expert turns out to be time-consuming and tedious. Therefore an automatic approach may be very useful to simplify the segmentation of OARs. In this regard, deep convolutional neural networks (CNN) are now the state-of-the-art in medical image segmentation. The process consists in learning the weights of the network through the optimisation of a differentiable loss function, such as the Dice loss or the cross-entropy, with a gradient descent. Nevertheless, incorporating a small amount of high-level information in the segmentation process (shape prior knowledge [2], topology preservation enforcement [3], prescription of the number of connected components/holes [4], (partial) convexity [5]) proves to achieve more accurate results. Motivated by this observation and inspired by the work of Peng et al. [6], we propose enforcing geometrical constraints in the training of the convolutional neural network by designing a suitable loss function which, in addition to intensity pairing, includes a criterion of edge alignment through a weighted total variation term, an area penalisation, and a component ensuring intensity homogeneity, yielding a non-smooth non-convex optimisation problem.

An auxiliary variable is introduced together with a splitting approach so as to separate the problem of non-convexity from the problem of non-smoothness, yielding an easy implementable alternating algorithm consisting of a smooth non-convex subproblem and a non-smooth convex one —the lack of smoothness stemming both from the applied hard constraints and the weighted total variation term —. This latter subproblem is solved using a proximal splitting method [7], namely Douglas-Rachford (DR) algorithm, the adjective *proximal* meaning that each non-smooth function is involved via its proximity operator, guaranteeing theoretical convergence.

More concretely, the CNN, based on U-Net [8], is trained thanks to the alternating direction method of multipliers (ADMM) algorithm [9] to separate the network parameter optimisation with a stochastic gradient descent (SGD) method, from the one of the auxiliary variable with Douglas-Rachford algorithm. This alternating scheme thus unifies the two formalisms deep-learning approaches (supervised setting)/variational models (unsupervised one) in a single framework, meaning that the processing chain is not a simple sequential link between a deep learning-based part that would provide an estimated segmentation, and a post-processing

step achieved in a variational setting. By sharing representations between tasks and carefully intertwining them, one can create a synergy, increase the accuracy of the outcomes, while achieving better generalisation capabilities. Another interesting point is that this approach reconciles the intrinsic discrete nature of segmentation —which consists in assigning a label to each image pixel —with the continuous dimension of variational methods. To put it in another way, the labels that are discrete in essence are approximated by continuous variables. To summarise, our contributions are: (i) the proposition of a mathematically well-motivated and computationally tractable method to train a constrained deep network, the novelties relying on an original equality-constrained optimisation problem, composed of a weighted total variation term, a piecewise constant Mumford-Shah (MS)-like term to enforce intensity homogeneity and an area penalisation; (ii) an efficient algorithm based on the introduction of an auxiliary variable and on a splitting into subproblems; (iii) extensive evaluations with respect to classical metrics in thoracic CT image segmentation. Even if the numerical experiments are restricted to the case of binary segmentation, the model can be straightforwardly extended to multiple classes, which is beyond the scope of this contribution.

2. MATHEMATICAL MODELLING

2.1. Weighted total variation-based and area-constrained loss function

We consider a dataset of K 2D images associated with their binary ground truth that we denote by $\left\{y^k\right\}_{k=1,\cdots,K}$. The problem being separable with respect to variable k, we omit the dependency in k from now on. Let us denote by $g : \mathbb{R}^+ \to \mathbb{R}^+$ an edge detector function satisfying $g(0) = 1$ with g strictly decreasing and $\lim_{r\to+\infty} g(r) = 0$. We apply the edge detector function to the norm of the ground truth gradient $g(|\nabla y|)$. For the sake of conciseness, we set $g := g(|\nabla y|)$. We then consider the generalisation of the notion of functions of bounded variation to the setting of BV-spaces associated with a weight w as introduced in [10, Definition 2] and denoted by $TV_w(u)$. A more intuitive characterisation of this quantity emerges in the case of a characteristic function. Indeed, if v is a characteristic function, $1_{\Omega_\mathcal{C}}$, of a closed set $\Omega_\mathcal{C} \subset \Omega$ with $\mathcal{C}$ the boundary of $\Omega_\mathcal{C}$, then

$$TV_w(v = 1_{\Omega_\mathcal{C}}) = \int_\mathcal{C} w\, ds,$$

the term $\int_\mathcal{C} w\, ds$ constituting a new definition of the curve length, balanced by the weight w. Equipped with this material, we now depict the proposed optimisation problem.

Training the CNN will yield a segmentation function parameterised by θ such that its output is the foreground probability $s(\theta)_{i,j}$ at each pixel (i, j) of the discrete image domain $\mathcal{G}$. To achieve this probability assignment, we formulate an optimisation problem based on the *(discrete version of the)* weighted total variation (TV_g) in order to enforce level curve agreement/boundary alignment between the predicted segmentation and the ground truth, and subject to discrete constraints including area penalisation through the hard constraint in α. More precisely, we restrict the obtained segmentation area, $\sum_{(i,j)\in\mathcal{G}} (s(\theta))_{i,j}$, to being equal to the area of the ground truth α, which is assessed from the associated ground truth image. We also impose $s(\theta)_{i,j}$ to be binary. The resulting loss function minimisation problem reads as follows:

$$\begin{aligned} &\inf_\theta \mathcal{L}(\theta) = \mathcal{F}(s(\theta), y) + TV_g(s(\theta)) \\ &\text{s.t} \qquad (s(\theta))_{i,j} \in \{0, 1\} \text{ and } \sum_{(i,j)\in\mathcal{G}} (s(\theta))_{i,j} = \alpha, \end{aligned} \tag{1}$$

where $\mathcal{F}$ is a standard loss function in image segmentation that can be Dice, cross-entropy, a Mumford-Shah-like fidelity term [11, 12], or even a combination of these.
Introducing an auxiliary variable u—the underlying aim being to split the original problem into more easily solvable problems—, the minimisation problem can be equivalently rephrased as:

$$\begin{aligned} &\inf_{\theta,u} \quad \mathcal{L}(\theta, u) = \mathcal{F}(s(\theta), y) + TV_g(u) \\ &\text{s.t} \quad u = s(\theta), u_{i,j} \in \{0, 1\}, \sum_{(i,j)\in\mathcal{G}} u_{i,j} = \alpha. \end{aligned} \tag{2}$$

It is solved using an augmented Lagrangian method, more precisely the scaled form [13], stated as:

$$\begin{aligned} &\max_w \min_{\theta,u} \quad \mathcal{L}(\theta, u, w) = \mathcal{F}(s(\theta),y)+TV_g(u)+\frac{\mu}{2}\|s(\theta)-u+w\|^2 \\ &\text{s.t} \qquad\quad u_{i,j} \in \{0, 1\}, \sum_{(i,j)\in\mathcal{G}} u_{i,j} = \alpha. \end{aligned} \tag{3}$$

2.2. Updating of parameters θ, u and w

An efficient method to solve this optimisation problem is the ADMM algorithm 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 minimised with respect to θ being smooth (non-convex) here) that reduces to :

$$\mathcal{F}(s(\theta), y) + \frac{\mu}{2}\|s(\theta) - u + w\|^2. \tag{4}$$

Then, to update the variable u, we first slightly relax the binary constraint and convert it into $\forall(i, j) \in \mathcal{G}$, $u_{i,j} \in [0, 1]$,

Algorithm 1 ADMM Algorithm

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

yielding a convex non-smooth optimisation problem in u:

$$\begin{aligned} \min_u \quad & TV_g(u) + \frac{\mu}{2} \|s(\theta) - u + w\|^2 \\ \text{s.t} \quad & u_{i,j} \in [0,1], \sum_{(i,j)\in\mathcal{G}} u_{i,j} = \alpha. \end{aligned} \tag{5}$$

We introduce auxiliary variables z and v, the latter one being related to the area constraint, such that $z = \nabla u$ and $v = u$, along with the convex sets

$$\begin{aligned} \mathcal{C}_1 &= \{u | \forall (i,j) \in \mathcal{G},\, u_{i,j} \in [0,1]\}, \\ \mathcal{C}_2 &= \left\{u, z | \forall (i,j) \in \mathcal{G},\, z_{i,j} = (z^1_{i,j}, z^2_{i,j})^T = (\nabla u)_{i,j}\right\}, \\ \mathcal{C}_3 &= \{u, v | \forall (i,j) \in \mathcal{G},\, u_{i,j} = v_{i,j}\}, \\ \mathcal{C}_4 &= \left\{ v | \sum_{(i,j)\in\mathcal{G}} v_{i,j} = \alpha \right\}. \end{aligned}$$

We recall that given a non-empty convex subset $\mathcal{C}$ of $\mathbb{R}^N$, the indicator function of $\mathcal{C}$ is $i_\mathcal{C} : x \mapsto \begin{cases} 0 & \text{if} \quad x \in \mathcal{C} \\ +\infty & \text{if} \quad x \notin \mathcal{C} \end{cases}$. Also, if $\mathcal{C}$ is closed and convex, the projection of $x \in \mathbb{R}^N$ onto $\mathcal{C}$ is the unique point $P_\mathcal{C}(x) \in \mathcal{C}$ such that $d_\mathcal{C}(x) = \|x - P_\mathcal{C}(x)\|$. At last, we remind the reader with the definition of the proximity operator.

Definition 1 *(Taken from [7, Definition 10.1]) Let f be a lower semicontinuous convex function from $\mathbb{R}^N$ to $]-\infty, +\infty]$ such that dom $f \neq \emptyset$. For every $x \in \mathbb{R}^N$, the minimisation problem $\min_{y\in\mathbb{R}^N} f(y) + \frac{1}{2}\|x - y\|^2$ admits a unique solution denoted $prox_f(x)$. The operator $prox_f : \mathbb{R}^N \to \mathbb{R}^N$ thus defined is the proximity operator of f.*

Equipped with these elements, we aim to solve, using Douglas-Rachford algorithm, the equivalent problem

$$\begin{aligned} \min_{u,z,v} & \sum_{(i,j)\in\mathcal{G}} g_{i,j} \|z_{i,j}\| + \frac{\mu}{2}\|u - s(\theta) - w\|^2 + i_{\mathcal{C}_1}(u) \\ & + i_{\mathcal{C}_3}(u,v) + i_{\mathcal{C}_2}(u,z) + i_{\mathcal{C}_4}(v). \end{aligned} \tag{6}$$

The Douglas-Rachford algorithm is an iterative scheme, presented in Algorithm 2 in its general form, to minimise the sum of convex functions. Solutions are obtained thanks

Algorithm 2 Douglas-Rachford Algorithm ([7, Algorithm 10.15]): generic problem $\min_{x\in\mathbb{R}^N} f_1(x) + f_2(x)$ with f_1 and f_2 lower semicontinuous convex functions from $\mathbb{R}^N$ to $]-\infty, +\infty]$.

Fix $\varepsilon \in]0, 1[$, $\gamma > 0$, $y_0 \in \mathbb{R}^N$
for $n = 0, 1, \cdots$ **do**
  $x_n = \text{prox}_{\gamma f_2} y_n$
  $\lambda_n \in [\varepsilon, 2 - \varepsilon]$
  $y_{n+1} = y_n + \lambda_n \left(\text{prox}_{\gamma f_1}(2x_n - y_n) - x_n\right)$.
end for

to proximal operators that can be seen as a natural extension of the notion of projection operator onto a convex set ([7]). In compliance with the above notations, we define $f_1(u,v,z) = g(u,v) + h(z)$ such that $g(u,v) = \frac{\mu}{2}\|u - s(\theta) - w\|^2 + i_{\mathcal{C}_1}(u) + i_{\mathcal{C}_3}(u,v)$ and $h(z) = \sum_{(i,j)\in\mathcal{G}} g_{i,j}\|z_{i,j}\|$. The proximal operator of f_1 is given by $\text{prox}_{\gamma f_1}(u,v,z) = (\text{prox}_{\gamma g}(u,v), \text{prox}_{\gamma h}(z))^T$, where $\text{prox}_{\gamma h}(z_{i,j}) = \begin{cases} \left(1 - \frac{\gamma g_{i,j}}{\|z_{i,j}\|}\right) z_{i,j} & \text{if} \quad \|z_{i,j}\| \geq \gamma g_{ij} \\ 0 & \text{otherwise} \end{cases}$, and $\text{prox}_{\gamma g}(u,v) = (P_{[0,1]}(x), P_{[0,1]}(x))^T$, with $x = \frac{\mu\gamma}{\mu\gamma+2}(s(\theta) + w) + \frac{1}{\mu\gamma+2}u + \frac{1}{\mu\gamma+2}v$ and $P_{[0,1]}(x) = \min(\max(x,0),1)$. Next, $f_2(u,z,v) = i_{\mathcal{C}_2}(u,z) + i_{\mathcal{C}_4}(v)$ and one has $\text{prox}_{\gamma f_2}(u,z,v) = (P_{\mathcal{C}_2}(u,z), P_{\mathcal{C}_4}(v))$ with $P_{\mathcal{C}_4}(v) = v + \frac{\alpha - \sum_{(i,j)\in\mathcal{G}} v_{i,j}}{MN} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ and $P_{\mathcal{C}_2}(u,z) = (\tilde{u}, \nabla\tilde{u})$, where, setting $z = (z_1, z_2)^T$, $\tilde{u} = (I + \nabla_x^T\nabla_x + \nabla_y^T\nabla_y)^{-1}(u + \nabla_x^T z_1 + \nabla_y^T z_2)$.

Finally, the Lagrangian variable w is updated thanks to a gradient ascent technique.

3. EXPERIMENTS

3.1. 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. 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. While the original SegTHOR dataset contains the segmentation of 4 OAR, we only segmented the aorta, since we are addressing binary segmentation. The aorta is a challenging organ to segment: surrounded by a low contrast environment, it has a particular cane shape and can be constituted of 1 or 2 connected components depending on the

slice. The dataset is split in two: 40 CT scans are used for training and the remaining 20 are kept for inference.

3.2. Implementation

In all our experiments, the network architecture that we used is a simplified version of U-Net, called sU-Net, previously introduced in [14], that has fewer dense connected layers than the original one. Various loss functions $\mathcal{F}$ are evaluated, such as Dice loss, cross-entropy (CE), Mumford-Shah (MS), or even a combination of these. Images are normalised and cropped from the centre to obtain images of size 256×152 pixels. Data augmentation techniques are used to artificially triple the number of training images. The parameters of sU-Net are updated using SGD with an initial learning rate of 1e-3 and a batch size of 6. The parameter μ is fixed to 0.5. As for the parameters γ and λ of Douglas-Rachford algorithm, they are both set to 1. The code is developed with Pytorch.

4. RESULTS

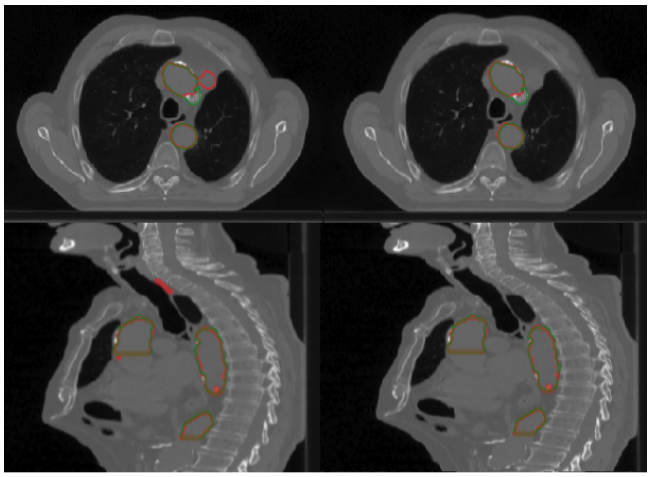
Segmentation results are evaluated in 3D patientwise, using two well-known metrics, namely the Dice score and the Hausdorff distance, and reported in Tab. 1. We also present results obtained with an unconstrained deep network, *i.e.* with the same architecture sU-Net, but in which weights are only updated by backpropagation of the gradient ('no constraint' row in Tab. 1). It is important to specify that no post-processing is applied so that the results can be analysed impartially. Dice values are similar for all methods, whereas the average Hausdorff distance proves to be lower in constrained methods, with the use of the MS term: indeed a paired Wilcoxon signed rank-test shows that HD values are significantly different between unconstrained and constrained approaches (p-value < 0.05). A segmentation result is shown in Fig. 1, first row, where a large group of missegmented pixels can be observed in the unconstrained case. This reflects the qualitative and fine properties that can be observed in the constrained case as opposed to the unconstrained case, namely segmentations more faithful to anatomical reality, fewer excrescences and false detections —additional results can be visualised at `https://github.com/zoelambert/GEOMETRICALLY-CONSTRAINED-DEEP-NETWORK`—. While the Hausdorff metric highlights the fact that non-constrained methods detect a high number of false positives unlike our constrained model, the Dice coefficient, by its underlying averaging effect, tends to conceal these spurious errors as shown in Fig. 1, second row.

5. CONCLUSION

In this paper, we have presented a new loss function which incorporates geometrical constraints during training of 2D-CNN. This function is built on a Dice term encouraging intensity pairing, a weighted total variation term inducing edge alignment, a piecewise-constant Mumford-Shah (MS)-like term enforcing intensity homogeneity, and an area penalisation. The resulting minimisation problem is split into two sub-problems so as to be solved using the ADMM algorithm. The optimisation of the first sub-problem is based on SGD, while that of the second one is performed using Douglas-Rachford algorithm. The introduction of this new loss with a very basic CNN improves the segmentation of the aorta of the SegTHOR dataset compared to that solely based on Dice loss minimisation, both from a qualitative (segmentations more faithful to anatomical reality) and quantitative viewpoint (especially for the Hausdorff distance metric). To confirm this analysis, the method will be evaluated on other datasets, extended to multiple classes, in weakly and semi-supervised settings.

Method	$\mathcal{F}$	Dice score %	HD (mm)
sU-Net	Dice	93.44 ± 2.07	51.68 ± 19.38
(no constraint)	CE	93.66 ± 1.87	54.69 ± 21.06
Proposed:	Dice	93.85 ± 1.64	50.67 ± 12.46
sU-Net	Dice + MS	93.71 ± 1.67	$\mathbf{43.72 \pm 21.48}$
+geom. const.	CE + MS	93.63 ± 2.36	$\mathbf{42.77 \pm 19.62}$

Table 1. Comparison of segmentation results (mean$\pm$stdev) without and with the proposed geometrical constraints, under various loss functions. Best significative results are in bold.



(a) No constraint (b) Geometric constraints

Fig. 1. Segmentation results on two patients of the aorta (red) and ground truth (green) with (a) Dice loss and no constraint and (b) Dice Loss, MS term and geometric constraints, with sU-Net. For the second row, the Dice coefficients are both equal to 94%, while the Hausdorff distance is 28.86 mm for the unconstrained case (a) against 2.82 mm for the constrained one (b).

6. COMPLIANCE WITH ETHICAL STANDARDS

This SegTHOR data acquisition was conducted retrospectively using human subject data, made available in open access by the LITIS and the Centre Henri Becquerel (CHB) [15]. The protocol was reviewed and approved by the institution (CHB) board.

7. ACKNOWLEDGEMENTS

This project was co-financed by the European Union with the European regional development fund (ERDF, 18P03390/18E0 1750/18P02733) and by the Haute-Normandie Régional Council via the M2SINUM project. The authors would like to thank the CRIANN (Centre Régional Informatique et d'Applications Numériques de Normandie, France) for providing computational resources.

8. REFERENCES

[1] H. Zhu, F. Meng, J. Cai, and S. Lu, "Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation," *J. Vis. Commun. Image Represent.*, vol. 34, pp. 12–27, 2016.

[2] O. Alexandrov and F. Santosa, "A topology-preserving level set method for shape optimization," *J. Comput. Phys.*, vol. 204, no. 1, pp. 121 – 130, 2005.

[3] C. Le Guyader and L. A. Vese, "Self-Repelling Snakes for Topology-Preserving Segmentation Models," *IEEE Trans. Image Process.*, vol. 17, no. 5, pp. 767–779, 2008.

[4] F. Ségonne, "Active Contours Under Topology Control—Genus Preserving Level Sets," *Int. J. Comput. Vis.*, vol. 79, no. 2, pp. 107–117, 2008.

[5] C. Y. Siu, H. L. Chan, and L. M. Lui, "Image segmentation with partial convexity prior using discrete conformality structures," *SIAM J. Imaging Sci.*, vol. to appear, 2020.

[6] J. Peng, H. Kervadec, J. Dolz, I. Ben Ayed, M. Pedersoli, and C. Desrosiers, "Discretely-constrained deep network for weakly supervised segmentation," *Neural Netw*, vol. 130, pp. 297 – 308, 2020.

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

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

[9] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comput. Math. with Appl.*, vol. 2, no. 1, pp. 17 – 40, 1976.

[10] A. Baldi, "Weighted BV functions," *Houston J. Math.*, vol. 27, no. 3, pp. 683–705, 2001.

[11] D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems," *Comm. Pure Appl. Math.*, vol. 42, no. 5, pp. 577–685, 1989.

[12] B. Kim and J. C. Ye, "Mumford-Shah Loss Functional for Image Segmentation With Deep Learning," *IEEE Trans. Image Process.*, vol. 29, pp. 1856–1866, 2020.

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

[14] Z. Lambert, C. Petitjean, B. Dubray, and S. Ruan, "SegTHOR: Segmentation of Thoracic Organs at Risk in CT images," in *IEEE IPTA*, 2020.

[15] *SegTHOR: an ISBI 2019 challenge*, `https://competitions.codalab.org/competitions/21145`.