



HAL
open science

Spatial Reasoning Loss for Weakly Supervised Segmentation of Skin Histological Images

Mateus Sangalli, Santiago Velasco-Forero, José Márcio Martins da Cruz, Virginie Flouret, Charlène Gayrard, Thérèse Baldeweck, Etienne Decencière

► **To cite this version:**

Mateus Sangalli, Santiago Velasco-Forero, José Márcio Martins da Cruz, Virginie Flouret, Charlène Gayrard, et al.. Spatial Reasoning Loss for Weakly Supervised Segmentation of Skin Histological Images. 21st IEEE International Symposium on Biomedical Imaging, The IEEE Signal Processing Society and the IEEE Engineering in Medicine and Biology Society, May 2024, Athens, Greece. hal-04485439

HAL Id: hal-04485439

<https://hal.science/hal-04485439v1>

Submitted on 1 Mar 2024

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.

Spatial Reasoning Loss for Weakly Supervised Segmentation of Skin Histological Images

Mateus Sangalli* Santiago Velasco-Forero*
José-Marcio Martins da Cruz* Virginie Flouret†
Charlène Gayrard† Thérèse Baldeweck†
Etienne Decencière *

Abstract

Biological images often follow some kind of geometrical structure. For example, histological images of reconstructed human skin follow a structural order with the stratum corneum as the outer layer and the living epidermis just below it. In this paper such spatial relationships are leveraged to define a loss function that penalizes structures that do not respect the given pattern as a form of weak supervision. The proposed loss function is based on fuzzy ontological spatial reasoning and morphological operators. The model is tested in a segmentation task on skin images, where a small number of labeled images and a large number of unlabeled ones are available. The proposed method leverages information in unlabeled images to improve segmentation results, compared with training only on the labeled data.¹

1 Introduction

Human visual reasoning, in particular *spatial reasoning*, has been extensively studied from a psychological and pedagogical perspective [1]. Trained humans can perform well on spatial reasoning tasks, mainly because they can solve them using spatial memory, logic, and imagination [2]. In other visual learning tasks, such as image classification, object detection, and segmentation, deep networks have been shown to outperform humans in memorizing indicative visual patterns from several image instances by means of supervised training approaches [3]. However, a distinguishing factor of human intelligence is the ability to learn new concepts from abstractions and a few examples, either by composing a new concept from primitives or by linking it to an existing concept [4]. Accordingly,

*Mines Paris, PSL University - Centre for Mathematical Morphology (CMM), Fontainebleau, France

†L'Oréal Research and Innovation, Aulnay-sous-Bois, France

¹Project page: https://github.com/mateussangalli/weakly_supervised_loss_fsr

we explore the learning of neural models from a few annotated data, including regularizations over spatial relationships in a semi-supervised manner.

Biological images are often subject to certain restrictions, for example, layers of skin exhibit a certain order, the outermost consists of dead cells, followed by the living epidermis, and so on. A human being trying to obtain a segmentation of a skin histological image can make use of that structural information to simplify the task. Our goal here is to design a loss function that allows a neural network to use the same information during its learning stage.

Most deep learning segmentation approaches focus are fully supervised, *i.e.* each pixel in the ground truth image is associated to a class. Weakly supervised semantic segmentation [5] refers to learning the segmentation when the available ground truth data is incomplete or uncertain. Recently, approaches based on geometric descriptors [6, 7] of the segmented regions have been proposed as a form of weak supervision. Our approach is also based on geometric information, but this information is in the form of spatial relations between the classes, e.g. class A is *above* class B. We model it using *fuzzy spatial reasoning*.

Fuzzy set theory [8] allows the representation of human knowledge in a semi-quantitative way, and makes for an ideal framework for spatial reasoning approaches [9]. That framework can be used to give a numerical value to expressions such as “*above*” or “*inside*”. The outputs of the softmax layer in a segmentation network are interpreted as fuzzy sets quantifying *how much* each pixel belongs to a each class. These sets can be used to optimize expressions that reflect human knowledge, such as “class A is not below class B”. Such loss function can extract knowledge from an image even in the absence of a ground truth.

In summary, our main contribution is a loss function rooted in fuzzy spatial reasoning that leverages known spatial relations between objects. This weak supervision method has the benefit of determining constraints for the entire dataset rather than on an image-by-image basis. We test it in a segmentation task with few annotated images, but with many images following a certain spatial structure.

The rest of the paper is organized as follows: Section 2 gives a brief introduction on the computation of the spatial relations through fuzzy morphological operators. The spatial reasoning loss is introduced in Section 3; the proposed approach is tested in Section 4 in a framework where we have few annotated images and several non-annotated ones.

2 Morphological spatial relations

Spatial relations between objects have been modeled as morphological operations between fuzzy sets [9]. Fuzzy set theory generalizes sets by assigning a *degree of pertinence* to the set, instead of a binary inclusion. In this way, a fuzzy set can be represented as a function $\mu : \Omega \rightarrow [0, 1]$, where $\mu(p) = 0$ means that p is fully excluded from μ and $\mu(p) = 1$ means that it is fully included in it.

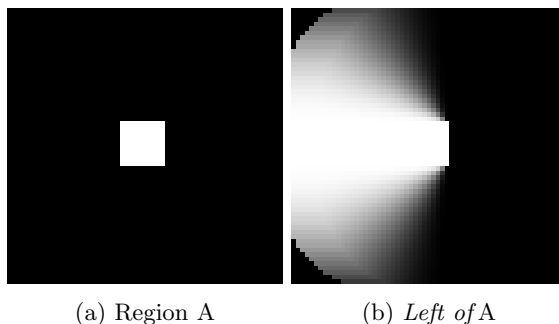


Figure 1: Illustration of the “left of and close to” operator.

If we think of Ω as space, *e.g.* the 2D grid of pixels of an image, then we can interpret fuzzy sets as uncertain regions in space and express meaningful geometric relationships such as “between”, “inside”, “above”, *etc* as operators applied to fuzzy sets. These relations can be described in terms of fuzzy morphological operators. Dilations and erosions are the basis for other morphological operators [10, 11] and have been used in Deep Learning to build robust features for high-dimensional signals [12], multi-scale representations [13] and others. We define the *fuzzy dilation* following [14] as

$$\delta_K(x)(p) = \sup_{q \in \Omega} \top(\mu(p), K(p - q)), \quad (1)$$

where $\top : [0, 1] \times [0, 1] \rightarrow [0, 1]$ denotes a *t*-norm [15], a generalization of the AND operator for fuzzy logic and $K : \Omega \rightarrow [0, 1]$ is called *structuring element*. Formally a *t*-norm is an associative, increasing and commutative operator with 1 as its neutral element.

The method uses a loss function based on directional relations, *e.g.* “to the left of” or “above”, such as the one illustrated in Fig. 1. To compute a directional relation with respect to a direction $\hat{v} \in \mathbb{R}^2$, with $\|\hat{v}\| = 1$, we use a directional structuring element [16] $K_{\hat{v}}(p) = \left(\max \left\{ \frac{\langle \hat{v}, p \rangle}{\|p\|}, 0 \right\} \right)^2$.

In practice, we also limit the distance between objects. That can also be described as a spatial relation “close to” and we implemented it by truncating the structuring element at distance R , *i.e.*

$$K_{\hat{v}, R}(p) = \begin{cases} \max \left\{ \frac{\langle \hat{v}, p \rangle}{\|p\|}, 0 \right\}^2, & 0 < \|p\| \leq R, \\ 0, & \|p\| > R, \\ 1, & p = 0. \end{cases} \quad (2)$$

We denote the dilation by $K_{\hat{v}, R}$ as $\delta^{\hat{v}, R} := \delta_{K_{\hat{v}, R}}$.

Finally, logical expressions can be written in terms of *t*-norms, *t*-conorms and other fuzzy operators of these regions, for example, given fuzzy sets A and B , the region “above A and not in B ” can be written as $\top(\delta^{(0,1), R}(A), (1 - B))$.

3 Method

3.1 Weakly supervised semantic segmentation

We explore a scenario akin to semi-supervised segmentation, characterized by a limited set of images featuring comprehensive labels. These labeled instances are denoted as image/label pairs $\{(x_n^l, y_n^l)\}_{n=1}^{N_l}$, and a large amount of unlabeled images is present $\{x_n^u\}_{n=1}^{N_u}$ [17]. We assume that all the images x have support on a region Ω .

We proceed by training a neural network using a composite loss function represented as follows:

$$\mathcal{L} = \mathcal{L}_l + \lambda \mathcal{L}_u, \quad (3)$$

Here, \mathcal{L}_l is associated with the labeled data, and \mathcal{L}_u with the unlabeled data. The term \mathcal{L}_l is defined as the cross-entropy loss between softmax probabilities and their corresponding targets. Importantly, the \mathcal{L}_u component of the loss is crafted considering established spatial relationships between classes, serving as a form of weak supervision. This allows us to leverage known information to enhance the learning process. In the subsequent sections, we will introduce a specific formulation for \mathcal{L}_u based on *a priori* spatial relations inherent to each class.

In the rest of this section, the loss \mathcal{L}_u will be constructed based on known spatial relationships between classes and thus constitutes a form of weak supervision.

3.2 Proposed unlabeled loss

Let us denote by μ_c a fuzzy set corresponding to the degree of pertinence to class c . We define it to be equal to the softmax probabilities i.e. $\mu_c(p)$ is equal to the predicted probability of point p being of class c .

In order to obtain a loss function, we assume that certain transitions between classes cannot occur. Let us assume that, when traversing the image going in the direction \hat{v} , it is impossible to transition from class c to class c' . In terms of the previously mentioned spatial relations, we can write such restriction as “ p is in the direction \hat{v} of class c' implies p does not belong to c ” or, by equivalency “ p is not in the direction \hat{v} of class c' or p does not belong to c ”. The negation of that expression is “ p is class c' and p is in the direction \hat{v} of class c ”, which gives us an expression to minimize, namely the fuzzy set $\top(\delta^{\hat{v}, R}(\mu_c), \mu_{c'})$. This gives us a loss term:

$$\mathcal{L}_{\hat{v}}^{c, c'} = \frac{1}{|\Omega|} \sum_{p \in \Omega} \top(\delta^{\hat{v}, R}(\mu_c)(p), \mu_{c'}(p)). \quad (4)$$

And for multiple constraints $(c_1, c'_1), \dots, (c_K, c'_K)$, the loss function (3) can be defined as

$$\mathcal{L}_u = \sum_{k=1}^K \mathcal{L}_{\hat{v}}^{c_k, c'_k}. \quad (5)$$

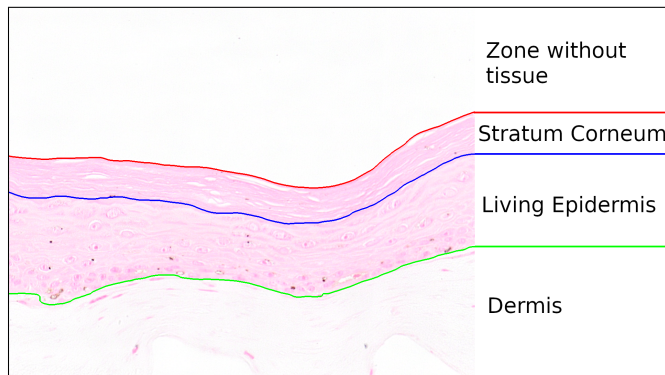


Figure 2: Illustration of a skin histological image with the boundary between the regions of interest highlighted.

We use two t -norms in this expression. To compute the dilation (1) we use a Lukasiewicz t -norm $\top_L(a, b) = \max\{a + b - 1, 0\}$ to more closely resemble the classical grayscale dilations. However, the AND operator in (4) uses the product t -norm $\top_p(a, b) = ab$.

3.3 Unlabeled loss for skin image segmentation

Here, the proposed loss terms are utilized for segmenting histological skin tissue images, specifically focusing on delineating the *Stratum Corneum* (SC) and the *Living Epidermis* (LED). The remaining areas, such as the Dermis and the region above SC (lacking tissue), are considered background (BG) for segmentation, as they lie outside the region of interest. Refer to Fig. 2 for an image illustrating the boundaries between these regions.

This task has some geometric priors that we can model in terms of transitions between those regions: all images follow the top-to-bottom order BG, SC, LED, BG. In particular this tells us of transitions between these labels that are illegal: BG directly above LED, SC below LED, BG directly below SC and their reciprocals.

We view the outputs of the softmax probabilities as fuzzy sets, which allows us to apply the methods described so far to penalize wrong orderings between layers.

To penalize wrong orderings, we propose the restrictions²: “BG is not below SC”, “BG is not above LED”, “LED is not above SC”, “SC is not below LED”, “LED is not below BG”, and “SC is not above BG”. Computing the union of the associated fuzzy sets results in the highlighted regions in Fig. 3. The terms

²We write “above” and “below” in these restrictions instead of “above and close to” and “below and close to” for brevity

above originate the following fuzzy sets

$$\begin{aligned}
 L_1 &= \top(\mu_{\text{BG}}, \delta^{(0,-1),R}(\mu_{\text{SC}})) & L_2 &= \top(\mu_{\text{BG}}, \delta^{(0,1),R}(\mu_{\text{LED}})) \\
 L_3 &= \top(\mu_{\text{LED}}, \delta^{(0,-1),R}(\mu_{\text{BG}})) & L_4 &= \top(\mu_{\text{BG}}, \delta^{(0,1),R}(\mu_{\text{LED}})) \\
 L_5 &= \top(\mu_{\text{SC}}, \delta^{(0,-1),R}(\mu_{\text{LED}})) & L_6 &= \top(\mu_{\text{LED}}, \delta^{(0,1),R}(\mu_{\text{SC}}))
 \end{aligned} \tag{6}$$

And the unlabeled loss is $\mathcal{L}_u = \sum_{i=1}^6 \sum_{p \in \Omega} L_i(p)$.

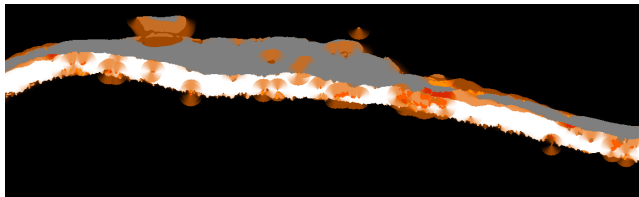


Figure 3: Regions penalized by the proposed unlabeled loss function \mathcal{L}_u . The structuring element used in (6) for the directional relations has radius $R = 20$ pixels.

Incorporating this regularization term holds significance as it enables the penalization of errors within an image, even when an accurate ground truth is unavailable. This flexibility empowers us to leverage images for which the labels are unknown, expanding the applicability of our approach.

4 Experiments

In this section, the proposed penalizations (6) are tested in the case where we have few fully annotated training images and many non-annotated training images.

4.1 Dataset

We use a dataset of images as described as Section 3.3. It's images have varying characteristics in both color and shape of the regions of interest, while respecting the top-to-bottom order discussed. It contains 254 images: 138 for training, 39 for validation and 77 for testing. They have a resolution of about $1.09\mu\text{m}/\text{pixel}$ and have variable height and variable width, but on average they have around 0.5 million pixels.

In order to fit within the semi / weakly supervised framework of Section 3, the training set contains few annotated images. The precise number of annotated images used are $N_l = 1, 3, 5$ and 10. We also compare all results with a fully labeled baseline, i.e. a model trained using all 138 of the training annotations.

4.2 Implementation details

We use the same architecture for all settings of the experiment: a U-Net consisting of four downsampling / upsampling blocks with two convolutions each (each followed by leaky ReLU and instance normalization) and doubling the number of channels at each downsampling step (starting at 32).

Batches are formed from cropped images of size 256×256 , where each training batch contains 96 unlabeled images and 32 labeled pairs. The networks are trained for 200 epochs using the Adam optimizer. The initial learning rate is set to 10^{-3} and decays exponentially with a factor of 0.977 after epoch 20. We use a weight decay of 10^{-4} . We augment the data with scale jittering in the interval $[\frac{1}{1.4}, 1.4]$, random horizontal flips and rotations by angles in $[-\frac{\pi}{8}, \frac{\pi}{8}]$.

To quantify the results, we use the mean *Intersection over Union* (IoU) between the predictions and the labels in the test set. We test methods using (3) with different values of λ ranging from zero to one. Regardless of the value of λ , at the beginning of the first epoch we initialize it to zero and increase it linearly until it reaches the desired value at epoch ten. We found that this λ schedule results in more stable improvements than initializing λ to the desired value.

4.3 Results

Fig. 4 helps elucidate the interest of using the spatial reasoning loss term from Section 3 by illustrating the correlation between small values of the proposed loss term and elevated IoU, reinforcing the relevance of the employed regularization strategy.

Table 1 highlights the efficacy of the proposed method, showing notable improvements compared to training exclusively with labeled data (i.e. with $\lambda = 0$) when the quantity of labeled data is small. Notably, the achieved results closely approach the fully labeled baseline. A qualitative analysis also helps to illustrate an increased regularity in predictions from models trained using the spatial reasoning loss, as shown in Fig. 5.

Table 1: Mean IoU values on the test set for different values of λ , with $N_l = 1, 3, 5$ or 10 labeled training images. We also show the results for the fully labeled baseline i.e. using $N_l = 138$. Values are averaged from five random repetitions of the experiment.

		Number of labeled images (N_l)			
		1	3	5	10
Weakly supervised	$\lambda = 0$	0.687	0.837	0.891	0.900
	$\lambda = 0.001$	0.696	0.839	0.888	0.910
	$\lambda = 0.01$	0.706	0.814	0.895	0.906
	$\lambda = 0.1$	0.761	0.852	0.891	0.907
	$\lambda = 1$	0.757	0.839	0.895	0.906
Fully supervised		0.914			

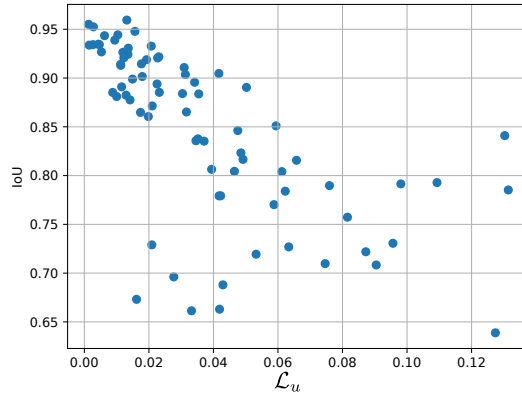


Figure 4: Relation between the value of the spatial reasoning loss and the final IoU measure for every image on the test dataset. The model used to produce this figure was trained only on *one* labeled image, with $\lambda = 0$. Segmentations with lower \mathcal{L}_u values tend to generate higher IoU scores.

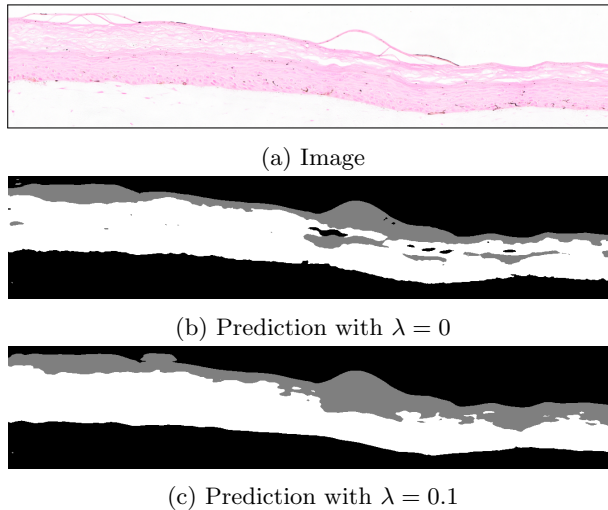


Figure 5: Worst image from the test set in terms of IoU of the prediction of a model trained with $\lambda = 0$ and its prediction by a model with $\lambda = 0.1$. Models trained with $\lambda > 0$ tend to produce outputs that better respect the order constraints.

5 Conclusions

In this paper we proposed a loss for weakly supervised semantic segmentation based on spatial relationships between objects. It penalizes transitions from

certain pairs of classes and is based on a fuzzy spatial reasoning framework.

In the task of segmentation of skin layers in histological images, the method penalizes segmentations that deviate from the specified order of SC, LED and BG. Consequently, this loss is applied to extract information from unlabeled data during training. The results show that this loss can induce a significant improvement in the prediction quality.

It would be interesting to apply this method with other constraints, *e.g.* using morphological operators to optimize topological constraints, like a class being inside another.

6 Acknowledgements

This work was granted access to the HPC resources of IDRIS under the allocation 2023-AD011012212R2 made by GENCI.

7 Compliance with Ethical Standards

Cells in reconstructed skins were isolated from skin explant. Normal human skin biopsies were obtained from mammary surgical residues, after written informed consent from the donors according to the principles expressed in the Declaration of Helsinki and in article L.1243–4 of the French Public Health Code. Given their special nature, surgical residues are subject to specific legislation included in the French Code of Public Health (anonymity, gratuity, sanitary/safety rules...). This legislation does not require prior authorization by an ethics committee for sampling or use of surgical waste.

References

- [1] Sherry Hsi, Marcia C Linn, and John E Bell, “The role of spatial reasoning in engineering and the design of spatial instruction,” *Journal of engineering education*, vol. 86, no. 2, pp. 151–158, 1997.
- [2] Harrison J Kell, David Lubinski, Camilla P Benbow, and James H Steiger, “Creativity and technical innovation: Spatial ability’s unique role,” *Psychological science*, vol. 24, no. 9, pp. 1831–1836, 2013.
- [3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *Proceedings of the IEEE CVPR*, 2015, pp. 1026–1034.
- [4] György Gergely, Harold Bekkering, and Ildikó Király, “Rational imitation in preverbal infants,” *Nature*, vol. 415, no. 6873, pp. 755–755, 2002.
- [5] George Papandreou, Liang-Chieh Chen, Kevin P Murphy, and Alan L Yuille, “Weakly-and semi-supervised learning of a deep convolutional

- network for semantic image segmentation,” in *Proceedings of the IEEE CVPR*, 2015, pp. 1742–1750.
- [6] Hoel Kervadec, Houda Bahig, Laurent Letourneau-Guillon, Jose Dolz, and Ismail Ben Ayed, “Beyond pixel-wise supervision for segmentation: A few global shape descriptors might be surprisingly good!,” in *Medical Imaging with Deep Learning*. PMLR, 2021, pp. 354–368.
 - [7] Hoel Kervadec, Jose Dolz, Meng Tang, Eric Granger, Yuri Boykov, and Ismail Ben Ayed, “Constrained-CNN losses for weakly supervised segmentation,” *Medical image analysis*, vol. 54, pp. 88–99, 2019.
 - [8] Lotfi A Zadeh, “Fuzzy sets,” *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.
 - [9] Isabelle Bloch, “Spatial reasoning under imprecision using fuzzy set theory, formal logics and mathematical morphology,” *International Journal of Approximate Reasoning*, vol. 41, no. 2, pp. 77–95, 2006.
 - [10] Laurent Najman and Hugues Talbot, *Mathematical morphology: from theory to applications*, John Wiley & Sons, 2013.
 - [11] Jean Serra, “Image analysis and mathematical morphology,” 1982.
 - [12] Santiago Velasco-Forero and Jesus Angulo, “Learnable empirical mode decomposition based on mathematical morphology,” *SIAM Journal on Imaging Sciences*, vol. 15, no. 1, pp. 23–44, 2022.
 - [13] Mateus Sangalli, Samy Blusseau, Santiago Velasco-Forero, and Jesus Angulo, “Scale equivariant neural networks with morphological scale-spaces,” in *International Conference on Discrete Geometry and Mathematical Morphology*. Springer, 2021, pp. 483–495.
 - [14] Isabelle Bloch and Anca Ralescu, “Fuzzy mathematical morphology,” in *Fuzzy Sets Methods in Image Processing and Understanding: Medical Imaging Applications*, pp. 85–119. Springer, 2022.
 - [15] Ronald R Yager, “Connectives and quantifiers in fuzzy sets,” *Fuzzy sets and systems*, vol. 40, no. 1, pp. 39–75, 1991.
 - [16] Isabelle Bloch and Anca Ralescu, “Spatial relations,” in *Fuzzy Sets Methods in Image Processing and Understanding: Medical Imaging Applications*, pp. 137–201. Springer, 2023.
 - [17] Yassine Ouali, Céline Hudelot, and Myriam Tami, “Semi-supervised semantic segmentation with cross-consistency training,” in *Proceedings of the IEEE CVPR*, 2020, pp. 12674–12684.