



**HAL**  
open science

## Improving The Automatic Segmentation Of Elongated Organs Using Geometrical Priors

Rebeca Vétel, Alexandre Bône, Marie-Pierre Vullierme, Marc-Michel Rohé,  
Pietro Gori, Isabelle Bloch

► **To cite this version:**

Rebeca Vétel, Alexandre Bône, Marie-Pierre Vullierme, Marc-Michel Rohé, Pietro Gori, et al.. Improving The Automatic Segmentation Of Elongated Organs Using Geometrical Priors. IEEE International Symposium on Biomedical Imaging (ISBI 2022 ), Mar 2022, Kolkata, India. hal-03628860

**HAL Id: hal-03628860**

**<https://telecom-paris.hal.science/hal-03628860>**

Submitted on 3 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# IMPROVING THE AUTOMATIC SEGMENTATION OF ELONGATED ORGANS USING GEOMETRICAL PRIORS

Rebeca Vétil<sup>1,2</sup>  
Marc-Michel Rohé<sup>1</sup>

Alexandre Bône<sup>1</sup>  
Pietro Gori<sup>2</sup>

Marie-Pierre Vullierme<sup>3</sup>  
Isabelle Bloch<sup>2,4</sup>

<sup>1</sup> Guerbet Research, Villepinte, France

<sup>2</sup> LTCI, Télécom Paris, Institut Polytechnique de Paris, France

<sup>3</sup> Department of Radiology, Hospital of Annecy-Genevois, Université de Paris, France

<sup>4</sup> Sorbonne Université, CNRS, LIP6, Paris, France

## ABSTRACT

Deep neural networks are widely used for automated organ segmentation as they achieve promising results for clinical applications. Some organs are more challenging to delineate than others, for instance due to low contrast at their boundaries. In this paper, we propose to improve the segmentation of elongated organs thanks to Geometrical Priors that can be introduced during training, using a local Tversky loss function, or at post-processing, using local thresholds. Both strategies do not introduce additional training parameters and can be easily applied to any existing network. The proposed method is evaluated on the challenging problem of pancreas segmentation. Results show that Geometrical Priors allow us to correct the systematic under-segmentation pattern of a state-of-the-art method, while preserving the overall segmentation quality.

**Index Terms**— Geometrical Prior, Medical Image Segmentation, Deep Learning, Pancreas, Tversky Loss.

## 1. INTRODUCTION

Automated organ segmentation is an important methodological step for the development of automated decision support tools in medical analysis. Yet, some organs are more difficult to segment than others because of ambiguous contours or elongated shapes.

Several methods have been proposed, and deep Convolutional Neural Networks (CNN) are increasingly recognized as the reference method as they achieve the best results in terms of Dice scores [1]. Most of the recently proposed CNNs derive from the U-Net [2], an encoder-decoder architecture which can operate on 2D and 3D images. Lately, in [3], the authors introduced the “no-new U-Net”, abbreviated nnU-Net, where the focus is put on the pre-processing, training and post-processing steps rather than on the network backbone. It was ranked first in a series of biomedical segmentation challenges, among which the Medical Segmentation Decathlon<sup>1</sup>.

While some automatic segmentation algorithms reach human-level performances [4], some organs, such as the pancreas or the colon, still constitute an open technical challenge for state-of-the-art models like the nnU-Net, as performances remain lower compared to those obtained on other abdominal organs [5]. These organs exhibit elongated shape with low-contrast at their boundaries, making the delineation of extremities difficult. Consequently, algorithms may miss the extremities of the organ. From a clinical point of view, this error can be critical if the automatic segmentation task is part of a

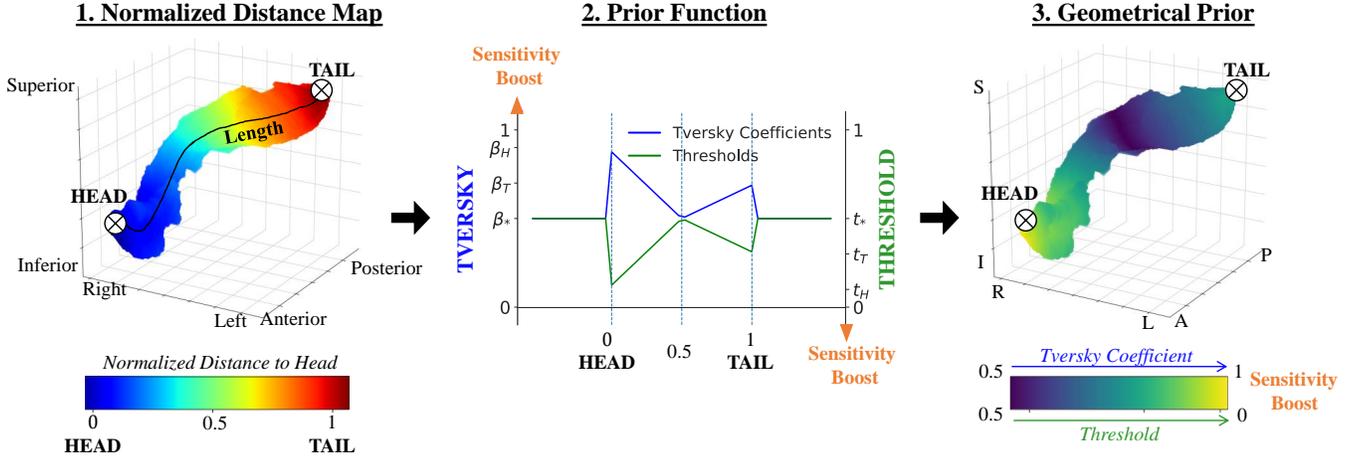
more complex tumor detection pipeline, since tumors at the extremities could be missed. A possible methodological approach to tackle this challenge is to perform multi-organ segmentation, delineating all the major structures close to the target one [6, 7]. The costs required to build such datasets motivated the study of semi-supervised approaches [8]. In particular, in [9], the segmentation of 13 abdominal structures was performed using a loss function which embedded anatomical priors computed on unlabeled data. More precisely, the outputs of the network were enforced to match a distribution of organ sizes, learned on a small fully labeled dataset. Consequently, the quality of the prior was highly dependent on the size of the fully labeled dataset, and the prior did not take the geometry of the organs into account. Such observations motivated us to investigate other anatomical priors, geometry-based, and whose definitions would not depend on labeled data.

In this work, we seek to improve the automatic segmentation of elongated organs. As a typical example, we illustrate our contributions on the pancreas segmentation problem. Besides having an elongated shape, the pancreas shows a strong ambiguity at its extremities which are highly intricate with the duodenum, the small intestine and the spleen. Such characteristics make segmentation of the pancreas challenging, even for experienced radiologists [1]. We evaluate our methods on two public datasets, namely The Cancer Imaging Archive Pancreas-CT (TCIA) [10] and the Medical Segmentation Decathlon (MSD) [11]. We review the performances of the state-of-the-art nnU-Net in segmenting the pancreas and verify that the nnU-Net significantly underestimates its extremities. To fix these systematic under-segmentation patterns, we propose two strategies that control the sensitivity of the predictions using geometrical priors which can be introduced either during training, using a local Tversky loss function [12], or during post-processing, using local thresholds.

## 2. METHOD

Consider an elongated shape organ  $O_{el}$ , whose extremities  $e_0, e_1$  are difficult to delineate. Let  $\mathcal{R} = \{r_i \in \{0, 1\}, i = 1 \dots N\}$  denote its segmentation of reference, with  $N$  being the number of voxels. Let  $\mathcal{P} = \{p_i \in [0, 1], i = 1 \dots N\}$  denote the probability output of a segmentation network. We define the Geometrical Prior  $\mathcal{G}_{prior}$  that gives for each voxel  $i$  the desired sensitivity boost  $\rho_i$ , *i.e.*,  $\mathcal{G}_{prior} = \{\rho_i \in [0, 1], i = 1 \dots N\}$ . Taking the illustrative example of pancreas segmentation, the extremities  $e_0$  and  $e_1$  will be referred to as the head ( $H$ ) and tail ( $T$ ) in the rest of this section.

<sup>1</sup><http://medicaldecathlon.com/results/>



**Fig. 1. Construction of the Geometrical Prior.** Starting from a binary segmentation, a normalized distance map is computed. Then, a prior function transforms the distance map into a Geometrical Prior by assigning to each voxel a sensitivity boost. Finally, the resulting Geometrical Prior can be translated either into local Tversky coefficients or into local threshold values.

This section introduces the Geometrical Priors and how to use them in practice.

**Geometrical Priors.** Geometrical Priors assign a sensitivity boost to each voxel of a segmentation mask. As illustrated in Figure 1, the construction of one Geometrical Prior  $\mathcal{G}_{prior}$  follows three steps. First, a Euclidean distance map assigns to each voxel its distance to the head. The distance map is normalized so that distance values are 0 and 1 at the head and the tail, respectively. Secondly, a prior function  $f_{prior}$  assigns to each voxel  $i$  its sensitivity boost  $\rho_i$  according to its distance to the head. Function  $f_{prior}$  is designed to give regions prone to under-segmentation a higher sensitivity boost, so as to favor their segmentation. For the pancreas,  $f_{prior}$  is built as a piece-wise linear function that emphasizes the extremities. It is defined by three parameters:  $\rho_H, \rho_T, \rho_*$ , corresponding to the sensitivity boost in the head, the tail and the background, respectively. Finally, the resulting Geometrical Prior can be used to define either a local Tversky loss function or local thresholds.

**Local Tversky loss function.** A first way to use Geometrical Priors is during training, through a local version of the Tversky loss function defined in [12]:

$$T(\mathcal{P}, \mathcal{R}, \mathcal{G}_{prior}^{Tversky}) = \frac{\sum_{i=1}^N p_i^0 r_i^0}{\sum_{i=1}^N p_i^0 r_i^0 + \sum_{i=1}^N \alpha_i p_i^0 r_i^1 + \sum_{i=1}^N \beta_i p_i^1 r_i^0}$$

where  $p_i^0$  (respectively,  $p_i^1$ ) is the predicted probability of voxel  $i$  belonging to  $O_{el}$  (respectively, background). Same notations go for  $r_i \in \mathcal{R}$ . The parameters  $(\alpha_i, \beta_i)$  control the trade-off between the specificity and the sensitivity for voxel  $i$ . In particular, higher values of  $\beta_i$  penalize more under-segmentation mistakes. While this loss function is used in [12] with global hyper-parameters  $(\alpha, \beta)$ , our method uses local coefficients, defined by the Geometrical Prior:  $\mathcal{G}_{prior}^{Tversky} = \{(\alpha_i, \beta_i) \in \mathbb{R}^2, i = 1 \dots N\}$ . In practice,  $\alpha_i = 1 - \beta_i$ , and  $\beta_i \in ]0, 1[$ . The  $\mathcal{G}_{prior}^{Tversky}$  values are computed for each patient once, before training, using their reference segmentation mask.

**Local thresholds.** The second way to use Geometrical Priors is in post-processing, using them to define local thresholds  $t_i$ , such as

$\mathcal{G}_{prior}^{threshold} = \{t_i \in ]0, 1[, i = 1 \dots N\}$ . The local thresholds are applied on the probability output  $\mathcal{P}$ , *i.e.*, voxel  $i$  is classified as belonging to the pancreas if  $p_i \geq t_i$ . In particular, areas with high sensitivity boost will exhibit lower thresholds. This time,  $\mathcal{G}_{prior}^{threshold}$  is computed from the predicted segmentation obtained with a global threshold  $t = 0.5$  where the head and tail are differentiated using the orientation of the images during acquisition. The final result is denoted by  $\mathcal{P}_{bin}$ .

Both proposed methods were based on the same rationale: lower  $t_i$ , just as larger  $\beta_i$ , would boost the sensitivity in desired areas, while keeping other areas unaffected. For the sake of simplicity, the sensitivity boost  $\rho$  will be used to refer to local Tversky coefficients, as  $\rho_i = \beta_i$ , or to local thresholds, as  $\rho_i = 1 - t_i$ , depending on the context.

### 3. EXPERIMENTS

**Dataset.** Two public datasets were used: (i) the TCIA dataset [10], containing 80 healthy subjects, (ii) the training set of the MSD pancreas dataset [11], containing 281 pathological subjects. Both datasets provided 3D portal CT-scans along with the reference segmentation masks of the pancreas and pancreatic tumors for pathological cases. To harmonize the datasets, MSD tumor masks were merged into the pancreas. The TCIA scans, with higher z-resolution than the MSD scans, were resampled to halve their z-resolution. After manual review by a radiologist with 25 years of expertise in abdominal imaging, 20 cases were put aside because their reference segmentation exhibited under-segmentation of the tail or the head. The remaining 341 cases were split into train/validation/test sets, with an equal TCIA/MSD 75:25 ratio across the sets.

**Baseline.** The pre-processing, network architecture, optimization and post-processing hyperparameters were automatically selected for our dataset using the nnU-Net self-configuring procedure detailed in [3]. All default parameters from public implementation<sup>2</sup> were kept without modification in a baseline experiment.

**Settings.** To evaluate the effect of our Geometrical Prior methods, experiments were run with different combinations of sensitivity

<sup>2</sup>github.com/MIC-DKFZ/nnUNet

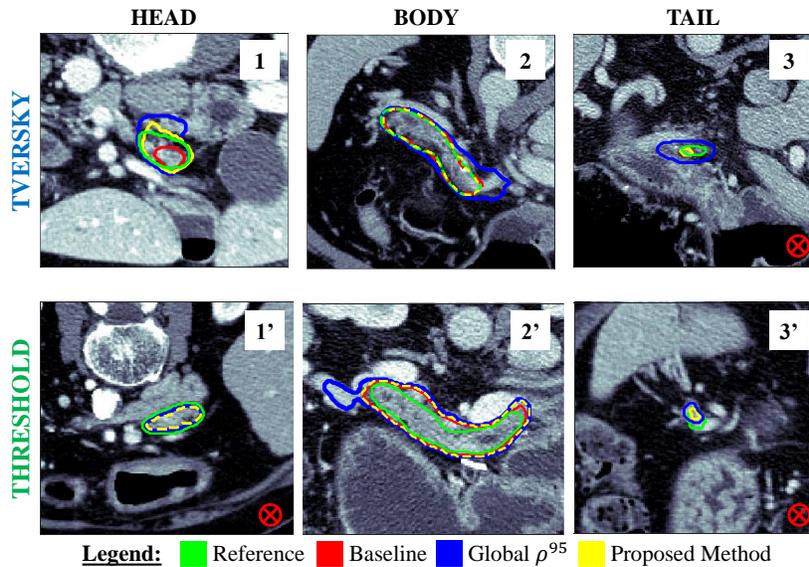
boosts  $(\rho_H, \rho_T, \rho_*)$ , where each combination represented a specific prior function. Values for  $\rho_x$  varied in  $[0.5, 0.7, 0.95]$ , referred to as  $[\rho_x^5, \rho_x^7, \rho_x^{95}]$ . Values smaller than 0.5 were not explored. In addition, the proposed methods were compared against the original Tversky loss function [12] with global coefficients.

**Performance measures.** The quality of the segmentation at the extremities was quantitatively evaluated with two measures. The first was the error on the length of the organ, expressed in *mm*. To compute the length, the segmentation mask was first converted into a voxel adjacency graph. In particular,  $n_T$  was the node corresponding to the tail, estimated as the closest point to the Left-Posterior-Superior corner in the abdomen, as shown in Figure 1. Finally, the length was estimated as the eccentricity of  $n_T$ , *i.e.*, the maximum graph distance from  $n_T$  to other nodes. This also provided the coordinates of the other extremity, the head. The signed length error

was then defined as  $Err\ Length = Length(\mathcal{P}_{bin}) - Length(\mathcal{R})$ , where  $\mathcal{P}_{bin}$  and  $\mathcal{R}$  refer to the predicted and reference segmentation masks, respectively. On the pancreas example, where extremities tend to be missed, we could estimate that negative length errors referred to under-segmentation whereas positive errors to over-segmentation. The second measure was the Dice score at the extremities, defined as the mean of the Dice scores computed on 5% of the length of the pancreas, from each extremity. Finally, in order to ensure that the overall segmentation quality was not deteriorated, the difference of the global Dice score with respect to the baseline was also reported.

## 4. RESULTS

In total, 10 experiments, in addition to the baseline, were conducted. In a first set of experiments A-E, the baseline loss function was re-



**Fig. 2.** Segmentation results obtained at the head, body and tail of the pancreas. The first row shows local Tversky loss function experiments, where blue and yellow correspond to experiment A and D, respectively. The second row shows local thresholds experiments, where blue and yellow correspond to experiment A' and E', respectively. D and E' were selected as our *Proposed Method*, according to Table 1.  $\otimes$  indicates void prediction for the baseline.

| PRIOR     |                                  | GLOBAL      |             |                     | LOCAL                                |                                   |                                   |
|-----------|----------------------------------|-------------|-------------|---------------------|--------------------------------------|-----------------------------------|-----------------------------------|
|           |                                  | $\rho^{95}$ | $\rho^7$    |                     | $\rho_H^{95}, \rho_T^{95}, \rho_*^5$ | $\rho_H^{95}, \rho_T^7, \rho_*^5$ | $\rho_H^{95}, \rho_T^7, \rho_*^7$ |
| TVERSKY   | <b>Experiment</b>                | Baseline    | A           | B                   | C                                    | D                                 | E                                 |
|           | <i>Length Error (mm)</i>         | -3.2 (6.4)  | 1.2 (7.8)   | -2.0 (6.5)          | 2.3 (7.6)                            | <b>-0.4 (8.2)</b>                 | 0.2 (6.8)                         |
|           | <i>Dice at Extremities (%)</i>   | 70.0 (3.4)  | 67.5 (2.1)  | 71.0 (3.4)          | 68.0 (1.7)                           | <b>72.5 (2.8)*</b>                | 70.5 (3.2)                        |
|           | <i>ΔDice (%) w.r.t. Baseline</i> | -           | -4*         | 0.2                 | -1.3*                                | 0.1                               | -2*                               |
| THRESHOLD | <b>Experiment</b>                | Baseline    | A'          | B'                  | C'                                   | D'                                | E'                                |
|           | <i>Length Error (mm)</i>         | -3.2 (6.4)  | -0.3 (6.3)  | -1.8 (7.2)          | -1.4 (7.6)                           | -1.6 (7.3)                        | <b>-1.0 (7.7)</b>                 |
|           | <i>Dice at Extremities (%)</i>   | 70.0 (3.4)  | 72.0 (2.3)* | <b>75.0 (2.95)*</b> | 74.0 (2.5)*                          | 74.0 (2.7)*                       | 74.0 (2.6)*                       |
|           | <i>ΔDice (%) w.r.t. Baseline</i> | -           | -2*         | 0.3                 | 0.1                                  | -0.1                              | 0.2                               |

**Table 1.** Evaluation measures obtained on the test set (65 cases). As the data are not normally distributed, we report the median instead of the mean. The interquartile range is indicated in brackets. Median Dice score for the Baseline is 88.0 (5.0). Significant differences with baseline are indicated by \*, except for the length error for which all the results are significant. Statistical significance is obtained with paired t-tests at  $p = 0.05$  level. For each evaluation measure, the best result, in **bold**, is selected among the experiments that do not deteriorate significantly the Dice score w.r.t. the baseline.

placed by a local Tversky loss function. In a second set of experiments A'-E', the baseline post-processing procedure was changed for a local thresholding approach, while the baseline loss function was kept. Details and qualitative results are reported in Table 1.

**Baseline results.** Baseline results were characterized by a great under-segmentation as they exhibited a negative median length error of  $-3.2\text{mm}$ . All the experiments A-E and A'-E' managed to resolve this under-segmentation issue, as the median error was significantly shifted towards positive values.

**Local Tversky.** Experiments A-E, using global or local Tversky loss function, produced significant effects on the length error. Yet, (A) showed that global  $\beta^{95}$  decreased the global Dice score, meaning that the overall segmentation was deteriorated. In addition, (A) also reported a drop of the Dice score at the extremities. On the other hand, global  $\beta^7$ , as in (B), did not affect the quality of the segmentation but produced only slight improvement on the length error. Both (A, B) illustrated the power but also the limitations of the global Tversky loss function, that would prevent its use with strong coefficients. These drawback effects were mitigated by Geometrical Priors: C-E did not result in a significant drop in the global Dice score, while they strongly corrected the length error. This mitigating capability was further illustrated in the upper row of Figure 2: at the extremities, (D) (yellow) behaved as (A), trained with a strong global Tversky loss coefficient (blue), and outperformed the baseline (red). By contrast, near the body of the pancreas, where the sensitivity boost was lower, (D) mimicked the baseline and avoided the substantial over-segmentation mistakes caused by (A). In addition to reducing the length error, (D) also significantly increased the Dice score at the extremities, suggesting the extremities were not only better detected, but also better delineated.

**Local thresholds.** Experiments A'-E', using local thresholds, yielded slighter improvements on the length error, with the best improvement in (A') coming at the expense of the global Dice score. Thus, best results were achieved by the experiment (E'). We hypothesized that the difference in results between the two methods was due to the distribution of the probabilities outputted by the network, where values were strongly pushed towards 0 or 1. Thus, regions missed by the baseline model tended to exhibit probabilities less than 0.05, and were therefore difficult to be recovered by the post-processing method. In this regard, the local threshold technique was less efficient than the local Tversky loss function, which corrected the under-segmentation from the training. Yet, the local thresholding approach produced systematic and significant increase of the Dice score at the extremities, as well as qualitative improvements illustrated in Figure 2. In particular, Figures 2.1' and 2.2' show total failures of the baseline that were successfully recovered by (E').

## 5. DISCUSSION AND CONCLUSION

We proposed two methods to boost the segmentation sensitivity in extremities of elongated organs, which state-of-the-art algorithms tend to miss. Our methods rely on the computation of Geometrical Priors that assign to each voxel a sensitivity boost. This boost can be used during training, as a local Tversky coefficient, or at post-processing, as a local threshold. Quantitative results on the pancreas segmentation problem demonstrated that both proposed techniques managed to significantly increase the segmentation sensitivity at the extremities of the organ. This was supported by the improvement of both the length error and the Dice score at the extremities, which were achieved without deteriorating the overall segmentation. The strengths of our work are threefold: first, the proposed methods are

interpretable thanks to the prior function that reflects the anatomical complexity of the organ. Secondly, the general formulation of Geometrical Priors allows our methods to be applied to any organ with regions that are difficult to segment. Last, our methods can be easily applied to any existing network. In particular, the local thresholds strategy does not require re-training the model. The main limitation of our work is the introduction of hyper-parameters which, although guided by anatomical knowledge, can make the search time-consuming. Thus, future work may explore the use of learnable prior functions, whose parameters would be dynamically learned during training or post-processing.

**Compliance with ethical standards.** This research study was conducted retrospectively using human subject data made available in open access [10, 11]. Ethical approval was not required as confirmed by the license attached with the open access data.

**Acknowledgments.** This work was partly funded by a CIFRE grant from ANRT # 2020/1448.

**Conflict of interest.** R.V., A.B. and M.R. are employed by Guerbet.

## 6. REFERENCES

- [1] H. Kumar, S.V DeSouza, and M.S Petrov, "Automated pancreas segmentation from computed tomography and magnetic resonance images: a systematic review," *Comput Methods Programs Biomed*, vol. 178, pp. 319–328, 2019.
- [2] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.
- [3] F. Isensee et al., "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nat. Methods*, vol. 18, no. 2, pp. 203–211, 2021.
- [4] W. Bai et al., "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *J Cardiovasc Magn Reson*, vol. 20, no. 1, pp. 65, 2018.
- [5] X. Yao, Y. Song, and Z. Liu, "Advances on pancreas segmentation: a review," *Multimed. Tools. Appl.*, vol. 79, no. 9, pp. 6799–6821, 2020.
- [6] L. et al Chu, "Application of deep learning to pancreatic cancer detection: lessons learned from our initial experience," *JACR*, vol. 16, no. 9, 2019.
- [7] Y. Wang et al., "Abdominal multi-organ segmentation with organ-attention networks and statistical fusion," *MedIA*, vol. 55, pp. 88–102, 2019.
- [8] Y. Zhou et al., "Semi-supervised 3D abdominal multi-organ segmentation via deep multi-planar co-training," in *WACV*, 2019, pp. 121–140.
- [9] Y. Zhou et al., "Prior-aware neural network for partially-supervised multi-organ segmentation," in *CVPR*, 2019, pp. 10672–10681.
- [10] H. Roth et al., "Data from pancreas-CT. The Cancer Imaging Archive," 2016.
- [11] A.L. Simpson et al., "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," *arXiv:1902.09063*, 2019.
- [12] S.S.M Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep networks," in *MICCAI*, 2017, pp. 379–387.