AIC-UNet: Anatomy-informed Cascaded UNet for Robust Multi-Organ Segmentation

Young Seok Jeon[1*][0000−0002−4948−083X], Hongfei Yang[1*][0000−0002−8150−9364], Huazhu Fu[2][0000−0002−9702−5524], and Mengling Feng[1][0000−0002−5338−6248]

[1] National University of Singapore, Singapore
[2] Agency for Science, Technology and Research (A*STAR), Singapore

Abstract. Imposing key anatomical features, such as the number of organs, their shapes, sizes, and relative positions, is crucial for building a robust multi-organ segmentation model. Current attempts to incorporate anatomical features include broadening effective receptive fields (ERF) size with resource- and data-intensive modules such as self-attention or introducing organ-specific topology regularizers, which may not scale to multi-organ segmentation problems where inter-organ relation also plays a huge role. We introduce a new approach to impose anatomical constraints on any existing encoder-decoder segmentation model by conditioning model prediction with learnable anatomy prior. More specifically, given an abdominal scan, a part of the encoder spatially warps a learnable prior to align with the given input scan using thin plate spline (TPS) grid interpolation. The warped prior is then integrated during the decoding phase to guide the model for more anatomy-informed predictions. Code is available at https://anonymous.4open.science/r/AIC-UNet-7048.

Keywords: multi-organ segmentation · deep learning · anatomy-informed

1 Introduction

It is becoming increasingly common to encounter AI models with reported performance on par with or surpassing radiologists. However, it is highly unlikely that AI models will replace radiologists anytime soon. While the AI models perform well in most cases, they may still make anatomically flawed predictions that radiologists would never make. For instance, the AI may predict that the esophagus, a muscular tube that carries food and liquid from the throat to the stomach, is disjointed. Alternatively, the AI might mistakenly identify the tibia bone as a femur, since these two bones may look similar at a local level. These examples demonstrate that current AI models are unreliable in learning essential anatomical features.

What prevents current AI models from identifying crucial anatomical features even after being exposed to hundreds of thousands of instances during training?

[*] These authors contributed equally to this work

Currently, fully autonomous AI-based segmentation models are solely trained to detect organs based on the input scan [5, 4, 7]. Consequently, we expect the AIs to pick up these anatomical constraints by the AI itself in a fully data-driven way. However, not only are these anatomical features global features, which are significantly more challenging to learn than local features, but the model may not even perceive them as strict constraints. Therefore, learning or incorporating anatomical features is a great challenge to current AI models.

Several methods have been proposed to incorporate or better learn anatomical constraints. These methods can be broadly categorized into 1) broadening effective receptive fields (ERF) or 2) imposing topological constraints. Many works considered Graph Neural Networks (GNNs) [15, 9] and self-attention networks [17, 7], which are more suitable for discovering global features than standard Convolutional Neural Networks (CNNs). However, these networks typically require more training data for better generalization [6], which can be a challenging requirement in medical domains. Some works considered reformulating segmentation to mesh-defomration task [1, 10], naturally offering smoother contour prediction by learning to deform a template mesh. However, it struggles to represent intricate structures. Topology regularization techniques [14, 8], while effective for specific anatomical challenges, limit the technique's generalizability when the task scales to multi-organ segmentation where organs' relative location plays a greater role.

Our proposed model, the Anatomically Informed Cascaded UNet (AIC-UNet), is designed to incorporate global anatomical priors without making significant changes to the standard encoder-decoder segmentation network. The model incorporates anatomical features without relying on resource-intensive global context learners, or topology regularizer, which does not handle inter-organ relations. Instead, we introduce an extra learnable parameter called the "*prior*", which can be spatially deformed to match a patient's anatomy. The deformed prior acts as a soft constraint during prediction. More specifically, a portion of the image encoder output is used as the control points of thin plate spline (TPS) grid deformation. TPS deforms a learnable prior to match a patient's anatomy. The deformed prior is integrated during the decoding phase to guide the decoder for more anatomy-informed predictions. To further increase the deformation accuracy of intricate objects, we repeat the same process with cropped local patches, resulting in a global-local cascaded network.

The contributions of this paper are summarized as follows:

- We propose improving the multi-organ segmentation model's robustness by conditioning its prediction with a deformed anatomical prior.
- We propose a global-local cascaded deformation approach to increase the deformation accuracy of intricate objects.
- We propose an activation maximization technique to learn a generic prior instead of using a fixed anatomy template.

2 Prior Works

Existing methods for enhancing anatomical feature learning focus on broadening ERF, reformulating segmentation to mesh deformation, or imposing topological constraints with regularizers, each with its own drawbacks.

Broadening ERF GNNs [15] and self-attention networks [17] are models that can attain larger ERF than standard CNNs. Therefore, these models are more suitable for learning distant dependencies within the data, making them good candidates for learning anatomical feature [3, 7, 13]. However, in practice, these models may struggle to effectively learn anatomical priors due to the limited data available to supervise the learning of long-range dependencies.

Mesh Deformation Mesh-deformation [10, 1], which naturally offers smoother contour predictions compared to conventional pixel prediction, encounters difficulties representing intricate structures. One potential solution is integrating mesh-based segmentation with pixel-based methods. However, this approach poses challenges due to the differing nature of object representation between the two methods.

Topology regularization Different shapes of organs can be described by their topological features, e.g. number of objects and number of cavities for 3D volumes. There are studies [18, 8] that use topological constraints to regularize network predictions. Topology-based techniques are often custom-tailored to specific anatomical challenges, diminishing their generalizability. Furthermore, these methods tend to prioritize penalizing local concepts, such as the number of holes or local connectedness, without addressing broader considerations like organ shapes and their relations.

3 Method

Network Overview At its core, AIC-UNet (depicted in Figure 1) is a cascaded network that requires both the global $\mathbf{X}_g \in \mathbb{R}^{H\times W\times D}$ and local view $\mathbf{X}_l \in \mathbb{R}^{H'\times W'\times D'}$ to produce a comprehensive local multi-organ prediction $\hat{\mathbf{Y}}_l \in [0,1]^{C\times H'\times W'\times D'}$ where C signifies the number of organs.

Initially, the model takes a down-scaled global view $\mathbf{X}_g$ as input to generate an initial, rough global prediction $\hat{\mathbf{Y}}_g \in [0,1]^{C\times H\times W\times D}$. In addition to the standard encoder-decoder structure of a segmentation model, AIC-UNet incorporates a free-parameter $\mathbf{Pr}_g \in \mathbb{R}^{C\times H\times W\times D}$ alongside 3 computation blocks: PriorEncoder_g, Deform_g, and $\{\text{SE}_g^{(i)}\}$. $\mathbf{Pr}_g$ is optimized to represent a generic anatomy. $\mathbf{Pr}_g$ is spatially deformed by a deformation module Deform_g to produce $\hat{\mathbf{Pr}}_g$ which closely follows the anatomy of $\mathbf{X}_g$. The degree of deformation is modulated by a concatenated feature extracted from $\text{Encoder}_g^{(4)}$ and PriorEncoder_g.

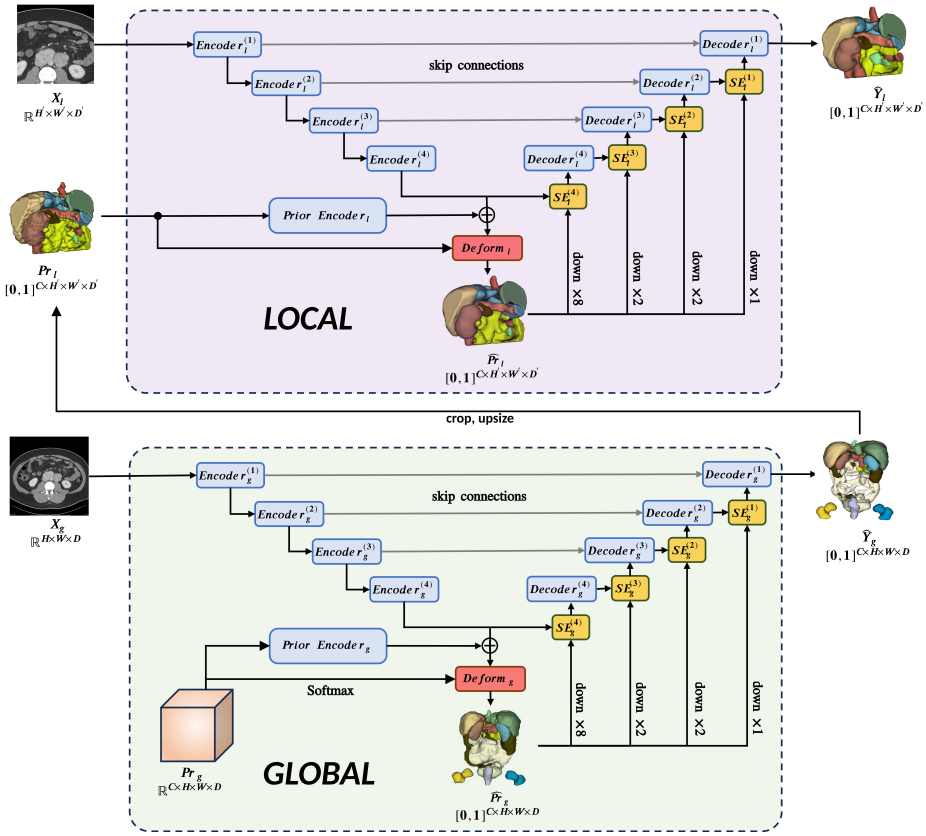


Fig. 1: Overview of AIC-UNet. AIC-UNet is a cascaded network combining global and local views for comprehensive multi-organ prediction. Initial input $\mathbf{X}_g$ yields rough global prediction $\hat{\mathbf{Y}}g$, enhanced by $\hat{\mathbf{Pr}}_g$, a spatially deformed anatomy from a learnable prior $\mathbf{Pr}_g$ via Deform_g. This process repeats in the local segment of the model for further enhancements, taking local view $\mathbf{X}_l$ and local prior $\mathbf{Pr}_l$.

This process is replicated in the local segment of the model, taking local view $\mathbf{X}_l$ and local prior $\mathbf{Pr}_l$, a cropped and rescaled global prediction $\hat{\mathbf{Y}}_g$, as an input prior. The local model serves to refine the deformed global prior, yielding a deformed local prior $\hat{\mathbf{Pr}}_l$.

Deformation block As shown in Figure 2, Deform block takes 2 inputs: estimated source control points $\{\mathbf{p}^{(i)}_{source}\}_{i=1}^{N}$ and prior $\mathbf{Pr}$, producing a deformed prior $\hat{\mathbf{Pr}}$. We use TPS deformation [2], which allows non-

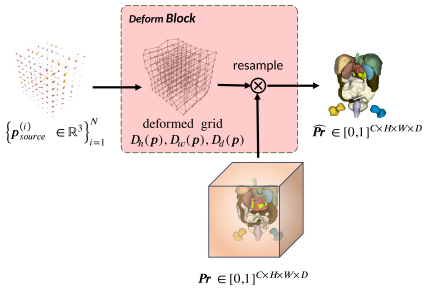


Fig. 2: Deform block deforms a learnable prior anatomy $\mathbf{Pr}$ to a patient-specific anatomy $\hat{\mathbf{Pr}}$.

linear deformation with sparse control points, to deform the generic anatomy $\mathbf{Pr}$.

TPS deformation works as follows, given a sequence of the predefined dense target control points $\{\mathbf{p}^{(i)}_{target} \in \mathbb{R}^3\}_{i=1}^{M}$ and sparse source control points $\{\mathbf{p}^{(i)}_{source} \in \mathbb{R}^3\}_{i=1}^{N}$, $N << M$, a unique grid deformation function $\mathcal{D}$ is determined by matching the control points $\mathbf{p}^{(i)}_{target} \mapsto \mathbf{p}^{(i)}_{source}$ with minimal bending energy. The matching of the h, w, and d coordinates of the control points, together with regularization conditions, give three sets of coefficients $(a^{(1)}, \cdots, a^{(N+4)})$, $(b^{(1)}, \cdots, b^{(N+4)})$, and $(c^{(1)}, \cdots, c^{(N+4)})$, such that a general target point $\mathbf{p} = (h, w, d)$ is mapped to $(\mathcal{D}_h(\mathbf{p}), \mathcal{D}_w(\mathbf{p}), \mathcal{D}_d(\mathbf{p}))$ with

$$\mathcal{D}_h(\mathbf{p}) = a^{(N+1)} + a^{(N+2)}h + a^{(N+3)}w + a^{(N+4)}d + \sum_{i=1}^{N} a^{(i)} U(|\mathbf{p} - \mathbf{p}^{(i)}_{source}|), \tag{1a}$$

$$\mathcal{D}_w(\mathbf{p}) = b^{(N+1)} + b^{(N+2)}h + b^{(N+3)}w + b^{(N+4)}d + \sum_{i=1}^{N} b^{(i)} U(|\mathbf{p} - \mathbf{p}^{(i)}_{source}|), \tag{1b}$$

$$\mathcal{D}_d(\mathbf{p}) = c^{(N+1)} + c^{(N+2)}h + c^{(N+3)}w + c^{(N+4)}d + \sum_{i=1}^{N} c^{(i)} U(|\mathbf{p} - \mathbf{p}^{(i)}_{source}|), \tag{1c}$$

where $U(r) = r^2 \log r^2$ is a kernel function. Further details on optimizing the coefficients of TPS are given in Appendix.

Learnable Prior Having a realistic organ anatomy as a global prior greatly benefits the accuracy of the subsequently deformed global and local priors. Unlike other atlas-based segmentation methods [10, 1], which assigns an arbitrary ground truth anatomy from a training set or simple structures like a sphere, AIC-UNet learns to find the optimal global prior during training. This is achieved by turning the global prior $\mathbf{Pr}_g$ as a free parameter with size $C \times H \times W \times D$, which matches the global view's spatial dimension. We apply Softmax along the channel dimension C to limit the range between $[0, 1]$.

We further explain the optimization trick used to accelerate the prior learning in Loss Function and Optimization subsection.

Aggregation of Prior The deformed prior $\hat{\mathbf{Pr}}$ is combined with the outputs from decoder blocks using Squeeze Excitation (SE) attention modules. A convolution is applied to the attention-modulated feature to match the desired channel size of the subsequent decoder block.

Loss Function and Optimization AIC-UNet is trained to minimize Dice, the combination of Soft-Dice and Cross Entropy loss computed both at a global and local level, as well regularization terms to stabilize the source control point estimations. Loss is defined as :

$$\begin{aligned} &\text{Dice}(\mathbf{Y}_g, \hat{\mathbf{Y}}_g) + \text{Dice}(\mathbf{Y}_g, \hat{\mathbf{Pr}}_g) + \lambda_g \sum_i ||\mathbf{p}^{(i)}_{\text{source},g}||_2 + \\ &\text{Dice}(\mathbf{Y}_l, \hat{\mathbf{Y}}_l) + \text{Dice}(\mathbf{Y}_l, \hat{\mathbf{Pr}}_l) + \lambda_l \sum_i ||\mathbf{p}^{(i)}_{\text{source},l}||_2, \end{aligned} \tag{2}$$

where $\mathbf{Y}_g$ and $\mathbf{Y}_l$ are the global and local ground truth anatomies. λ_g and λ_l are coefficients to control the degree of L_2 regularization.

Directly optimizing global prior $\mathbf{Pr}_g$ along with the other model parameters leads to slow convergence. We hypothesize that this is due to a correlation between the prior and TPS control points. For instance, if the predicted anatomy is smaller than the ground truth, the error can be reduced in two ways: 1) shrinking the source control points in TPS deformation or 2) enlarging the global prior. This correlation may confuse optimization priority. We prevent confusion by alternating the optimization of the model parameters and global prior.

4 Dataset and Experimental Setup

Dataset We use the freely available Whole abdominal ORgan Dataset (WORD) [12]. WORD consists of 150 anonymized CT scans. Each scan contains 159–330 slices with 512×512 pixel and an in-plane resolution of 0.976 mm $\times$ 0.976 mm. Slice spacing varies between 2.5 and 3.0 mm. WORD provides annotations of 16 organs, including the liver, spleen, kidneys, and various digestive organs. For training, validation, and testing purposes, the dataset is randomly split into 100 scans, 20 scans, and 30 scans, respectively.

Experimental Setup The pixel intensity is truncated between $[-250, 500]$ and spacing normalized to $[1.5, 1.5, 2.0]$. Axial direction (d-dimension) is zero-padded to have an identical volume size. Global view and global mask are down-sampled by a factor of $[3, 3, 2]$. The dimension of local view is set to $[128, 128, 128]$. We use AdamW [11] optimizer with linear warmup cosine annealing. Maximum learning rate and weight-decay is set to 3e$-$4 and 1e$-$5. For the optimization of the prior, the learning rate is set to 1e$-$3. Every 500 iterations, we conduct training for the prior over a span of 100 iterations. The model is trained for 350 epochs, each epoch with 200 iterations. Batch size is set to 2. We use 1e$-$8 for both $\lambda_{l,g}$ in the loss (2) to regularize TPS control points learning.

Baseline Methods We compare with two baseline methods. The first is a standard UNet that forms the backbone of our local segmentation network. The

Table 1: Comparison of Proposed and baseline models on WORD dataset.

	95% Haus ↓			NSD ↑			Dice ↑		
	AIC-UNet	**CUNet**	**UNet**	**AIC-UNet**	**CUNet**	**UNet**	**AIC-UNet**	**CUNet**	**UNet**
liver	2.408	**2.149**	**2.337**	95.888	**96.322**	**95.904**	96.274	**96.404**	**96.333**
spleen	**1.462**	1.678	**1.626**	**98.924**	**98.949**	98.722	95.463	**95.494**	**95.469**
left kidney	**1.562**	**1.452**	1.567	98.313	**98.499**	**98.39**	95.17	**95.236**	**95.238**
right kidney	**1.367**	7.233	**1.47**	**98.856**	98.047	**98.719**	**95.337**	95.058	**95.405**
stomach	**7.047**	**8.2**	8.456	**91.663**	**91.751**	90.939	**91.673**	**91.529**	91.154
gallbladder	**4.183**	**4.176**	5.069	**91.3**	**91.719**	89.737	**80.592**	**80.532**	79.698
esophagus	**3.617**	**3.461**	3.931	90.924	**91.491**	**91.369**	76.556	**77.432**	**77.065**
pancreas	5.18	**4.786**	**4.823**	90.357	**90.571**	**91.368**	83.265	**83.411**	**83.978**
duodenum	**13.433**	**12.954**	13.774	**76.68**	**76.137**	75.899	**68.602**	**67.617**	67.353
colon	8.834	**8.35**	**8.004**	88.029	**88.28**	**88.803**	85.251	**85.33**	**85.901**
intestine	**4.327**	4.511	**4.206**	91.917	**92.177**	**92.08**	**87.777**	**87.889**	88.093
adrenal	**4.097**	**4.026**	4.223	**88.99**	**89.381**	88.667	**69.72**	**70.42**	69.187
rectum	**7.212**	20.257	**10.494**	**84.987**	83.881	**84.352**	**81.745**	80.926	**81.307**
bladder	**2.792**	3.057	**2.704**	**93.992**	93.828	**93.963**	**90.981**	90.911	**91.191**
head of femur left	**11.74**	**3.795**	17.85	**94.181**	92.811	**93.467**	**91.978**	91.781	**91.825**
head of femur right	**3.553**	**10.493**	18.591	**94.024**	**93.043**	92.964	**92.242**	**91.792**	91.43
mean	**5.176**	**6.286**	6.82	**91.814**	**91.68**	91.584	**86.414**	**86.36**	86.289

second is a cascaded UNet (CUNet), which has an identical structure to our AIC-UNet except that it does not have the common prior $\mathbf{Pr}_g$. For CUNet, we use self-attention on the global segmentation network and attend features from the local segmentation network with cropped predictions from the global network.

5 Experimental Results

Segmentation Metrics We measure segmentation performance by three metrics: the dice score (Dice), the normalized surface dice (NSD) [16], and the 95% Hausdorff distance. Our AIC-UNet achieved best mean results on all three metrics.

Visualization of Deformed Prior Figure 3 shows the global prior anatomy and patient-specific deformed anatomy $\mathbf{Pr}_g$ learned by the TPS deform block. The figure depicts that the learned global prior closely aligns with our understanding of a generic organ anatomy. Additionally, the prior anatomy is successfully deformed into different patient-specific anatomies. For example, the deformed anatomy of the subject in the middle accurately represents the subject's relatively shorter torso and a smaller waist-to-hip ratio.

Qualitative Comparison The attention mechanism on the learned common prior in our AIC-UNet can promote anatomically accurate segmentation. This is supported by results in Figure 4. In Figures 4a–4b, UNet wrongly segmented bones near knees as femur heads, while AIC-UNet gives more accurate segmentation. In the case of gallbladder segmentation, as shown in Figures 4c–4d, UNet prediction has an extra component, and it deviates from the ground truth position (colored as a transparent red), whereas, AIC-UNet correctly identifies the number of components and their general position.

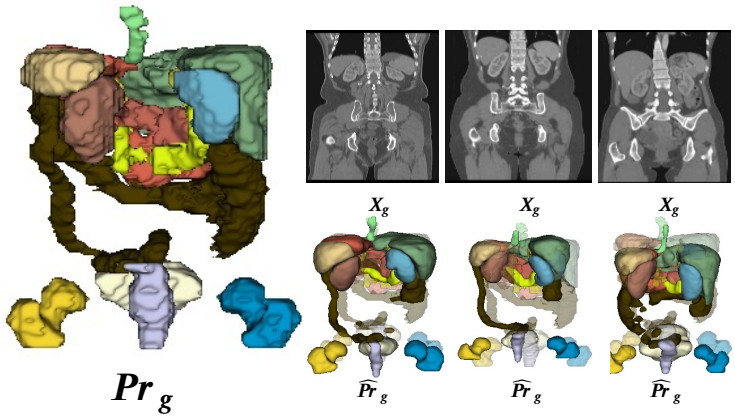
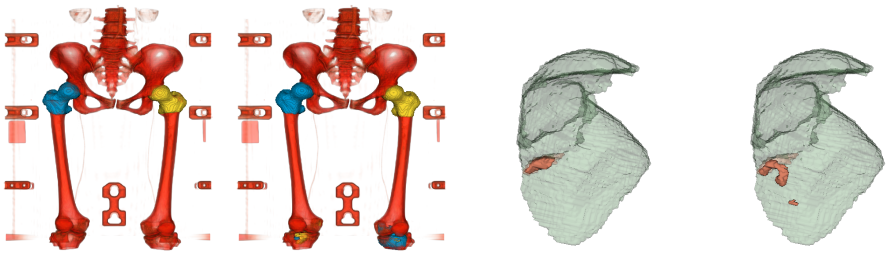


Fig. 3: Deformation of global prior $\mathbf{Pr}_g$ into different patient-specific anatomies $\hat{\mathbf{Pr}}_g$. The intestine is occluded for better visualization. The transparent anatomy in each $\hat{\mathbf{Pr}}_g$ are $\mathbf{Pr}_g$.

6 Conclusion and Future Research

We propose AIC-UNet, an encoder-decoder segmentation model that takes advantage of anatomical information by using prior deformation. To improve this model, future research could focus on two areas. Firstly, by developing a more effective target control point selection strategy to enhance the TPS deformation performance. Secondly, by designing a more powerful feature aggregation module that can integrate the information from the deformed prior to decoder blocks.



(a) AIC-UNet (b) UNet (c) AIC-UNet (d) UNet

Fig. 4: Qualitative comparisons of AIC-UNet and UNet on femur heads (yellow and blue) and gallbladder segmentation. Ground truth liver and gallbladder are superimposed with transparency.

References

[1] Fabian Bongratz, Anne-Marie Rickmann, and Christian Wachinger. "Abdominal organ segmentation via deep diffeomorphic mesh deformations". In: *Scientific Reports* 13.1 (2023), p. 18270.

[2] Fred L. Bookstein. "Principal warps: Thin-plate splines and the decomposition of deformations". In: *IEEE Transactions on pattern analysis and machine intelligence* 11.6 (1989), pp. 567–585.

[3] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. "Transunet: Transformers make strong encoders for medical image segmentation". In: *arXiv preprint arXiv:2102.04306* (2021).

[4] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. "Encoder-decoder with atrous separable convolution for semantic image segmentation". In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 801–818.

[5] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. "3D U-Net: learning dense volumetric segmentation from sparse annotation". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*. Springer. 2016, pp. 424–432.

[6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. "An image is worth 16x16 words: Transformers for image recognition at scale". In: *arXiv preprint arXiv:2010.11929* (2020).

[7] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R Roth, and Daguang Xu. "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images". In: *International MICCAI Brainlesion Workshop*. Springer. 2021, pp. 272–284.

[8] Xiaoling Hu, Fuxin Li, Dimitris Samaras, and Chao Chen. "Topology-preserving deep image segmentation". In: *Advances in neural information processing systems* 32 (2019).

[9] Juntao Jiang, Xiyu Chen, Guanzhong Tian, and Yong Liu. "ViG-UNet: Vision Graph Neural Networks for Medical Image Segmentation". In: *arXiv preprint arXiv:2306.04905* (2023).

[10] Fanwei Kong, Nathan Wilson, and Shawn Shadden. "A deep-learning approach for direct whole-heart mesh reconstruction". In: *Medical image analysis* 74 (2021), p. 102222.

[11] Ilya Loshchilov and Frank Hutter. "Decoupled weight decay regularization". In: *arXiv preprint arXiv:1711.05101* (2017).

[12] Xiangde Luo, Wenjun Liao, Jianghong Xiao, Jieneng Chen, Tao Song, Xiaofan Zhang, Kang Li, Dimitris N Metaxas, Guotai Wang, and Shaoting Zhang. "Word: A large scale dataset, benchmark and clinical applicable

study for abdominal organ segmentation from ct image". In: *arXiv preprint arXiv:2111.02403* (2021).

[13] Olivier Petit, Nicolas Thome, Clement Rambour, Loic Themyr, Toby Collins, and Luc Soler. "U-net transformer: Self and cross attention for medical image segmentation". In: *Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12*. Springer. 2021, pp. 267–276.

[14] Ainkaran Santhirasekaram, Mathias Winkler, Andrea Rockall, and Ben Glocker. "Topology Preserving Compositionality for Robust Medical Image Segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 543–552.

[15] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. "The graph neural network model". In: *IEEE transactions on neural networks* 20.1 (2008), pp. 61–80.

[16] Silvia Seidlitz, Jan Sellner, Jan Odenthal, Berkin Özdemir, Alexander Studier-Fischer, Samuel Knödler, Leonardo Ayala, Tim J Adler, Hannes G Kenngott, Minu Tizabi, et al. "Robust deep learning-based semantic organ segmentation in hyperspectral images". In: *Medical Image Analysis* 80 (2022), p. 102488.

[17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).

[18] Xiao Zhang, Jingyang Zhang, Lei Ma, Peng Xue, Yan Hu, Dijia Wu, Yiqiang Zhan, Jun Feng, and Dinggang Shen. "Progressive deep segmentation of coronary artery via hierarchical topology learning". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2022, pp. 391–400.

Appendix

Solution of Thin Plate Spline Coefficients Given the estimated source control points $\{\mathbf{p}^{(i)}_{source} \in \mathbb{R}^3\}_{i=1}^N$, 3D-TPS maps a general target point $\mathbf{p} = (h, w, d)$ to $(\mathcal{D}_h(\mathbf{p}), \mathcal{D}_w(\mathbf{p}), \mathcal{D}_d(\mathbf{p}))$ with

$$\mathcal{D}_h(\mathbf{p}) = a^{(N+1)} + a^{(N+2)}h + a^{(N+3)}w + a^{(N+4)}d + \sum_{i=1}^{N} a^{(i)} U(|\mathbf{p} - \mathbf{p}^{(i)}_{source}|), \tag{3a}$$

$$\mathcal{D}_w(\mathbf{p}) = b^{(N+1)} + b^{(N+2)}h + b^{(N+3)}w + b^{(N+4)}d + \sum_{i=1}^{N} b^{(i)} U(|\mathbf{p} - \mathbf{p}^{(i)}_{source}|), \tag{3b}$$

$$\mathcal{D}_d(\mathbf{p}) = c^{(N+1)} + c^{(N+2)}h + c^{(N+3)}w + c^{(N+4)}d + \sum_{i=1}^{N} c^{(i)} U(|\mathbf{p} - \mathbf{p}^{(i)}_{source}|), \tag{3c}$$

where $U(r) = r^2 \log r^2$ is a kernel function, $(a^{(1)}, \cdots, a^{(N+4)})$, $(b^{(1)}, \cdots, b^{(N+4)})$, and $(c^{(1)}, \cdots, c^{(N+4)})$ are coefficients to be optimized. Here we explain how the three sets of coefficients are calculated. We use the h-coordinate coefficients as an example, and the calculation of the w and d coordinate coefficients are done in a similar manner.

The thin plate function (3a) has $N + 4$ coefficients to be computed. Though the function is highly non-linear with the kernel function U, the function is linear with respect to the coefficients. Hence, the coefficients have a closed-form solution.

Let $\mathbf{v} = (h^{(1)}, \cdots, h^{(N)}|0, 0, 0, 0)^T$, where $h^{(i)}$ is the h-coordinate of the i-th source control point. Also, define matrices

$$\mathcal{K} = \begin{bmatrix} 0 & U_{12} & \cdots & U_{1K} \\ U_{21} & 0 & \cdots & U_{2K} \\ \cdots & \cdots & \cdots & \cdots \\ U_{K1} & U_{K2} & \cdots & 0 \end{bmatrix}, N \times N; \quad \mathcal{P} = \begin{bmatrix} 1 & h^{(1)} & w^{(1)} & d^{(1)} \\ 1 & h^{(2)} & w^{(2)} & d^{(2)} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & h^{(N)} & w^{(N)} & d^{(N)} \end{bmatrix}, N \times 4; \tag{4}$$

and

$$\mathcal{M} = \begin{bmatrix} \mathcal{K} & \mathcal{P} \\ \mathcal{P}^T & O \end{bmatrix}, (N+4) \times (N+4) \tag{5}$$

where $U_{i,j} = U(|\mathbf{p}^{(i)}_{source} - \mathbf{p}^{(j)}_{source}|)$, $h^{(i)}$, $w^{(i)}$, and $d^{(i)}$ are the h-, w-, and d-coordinates of source control point $\mathbf{p}^{(i)}_{source}$, and O is a zero matrix of size 4×4. Then the coefficients $\mathbf{a} = (a^{(1)}, \cdots, a^{(N+4)})$ are given by

$$\mathbf{a} = \mathcal{M}^{-1}\mathbf{v}. \tag{6}$$

The additional last four rows of $\mathcal{M}$ guarantee that the coefficients $a^{(i)}$ sum to zero and that their cross-products with the points $\mathbf{p}^{(i)}_{source}$ are likewise zero.