

SEMI-SUPERVISED CORONARY VESSELS SEGMENTATION FROM INVASIVE CORONARY ANGIOGRAPHY WITH CONNECTIVITY-PRESERVING LOSS FUNCTION

Haorui He^{*} Abhirup Banerjee^{†*} Marcel Beetz^{*} Robin P. Choudhury[†] Vicente Grau^{*}

^{*} Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK

[†] Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, UK

ABSTRACT

The segmentation of arteries in invasive coronary angiography is necessary to build quantitative models and eventually improve the diagnosis of cardiovascular diseases. Standard segmentation algorithms suffer due to the lack of fully annotated datasets and tend to return disconnected vessels. Thus, we explore a semi-supervised segmentation framework to address these issues. Specifically, we use a student model and a teacher model as the main framework with Nested U-Nets (UNet++) as their backbones. The student model learns by minimizing a segmentation loss between the output and the ground truth, and a consistency loss guided by the uncertainty information. Additionally, a special loss function based on elastic interaction is used to improve the connectivity of arterial branches. We demonstrate the effectiveness of our proposed techniques over 42 labeled and 60 unlabeled samples and find relative improvement of 5.59% for Dice score and 69.99% for Betti number compared to a U-Net.

Index Terms— Invasive Coronary Angiography, Coronary Arteries Segmentation, Semi-supervised Learning, Vessel Connectivity, Nested U-Net, Mean Teacher Model.

1. INTRODUCTION

Invasive Coronary Angiography (ICA) is the gold standard for assessment of coronary stenosis [1]. However, as ICA captures the projection of a three-dimensional (3D) moving structure onto a two-dimensional (2D) plane, it contains imaging artifacts such as overlapping organs and vessels, unpredictable motion patterns and unevenly distributed contrast [2]. This makes the segmentation of ICA particularly challenging, while necessary to obtain a full understanding of the complex coronary arterial structure in the presence of these artifacts.

The development of Convolutional Neural Networks has largely improved the quality of medical image segmentation especially with Fully Convolutional Networks (FCNs) and U-Nets [3, 4]. It has been shown to exceed human experts' performance in retinal vessel segmentation [5]. However, the direct application of existing retinal vessel segmentation methods for coronary vessel segmentation is infeasible, given the

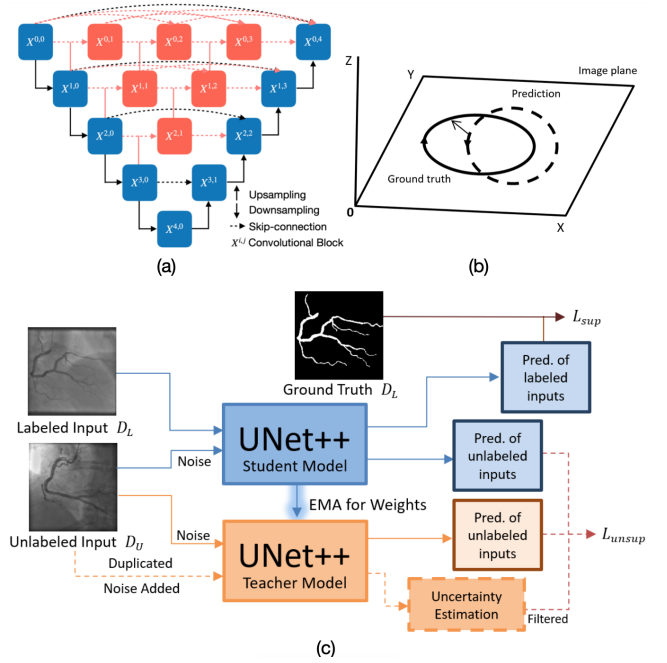


Fig. 1: (a) The structure of UNet++ with blue blocks and black arrows representing a conventional U-Net. (b) The image plane in 3D space. (c) The proposed framework for semi-supervised segmentation.

specific challenges of ICAs including large variations in vessel diameter, the presence of the catheter tube and motion artifacts. In order to address these challenges, patch-based approaches have been commonly used for multi-channel or multi-scale feature extraction in coronary arteries segmentation [6, 7]. However, these tend to result in disconnected vascular structures. More specialized networks were built to only segment the major vessels or localized stenosis [8, 9]. However, these methods all required a large amount of annotated ICAs, for which preparation is time-consuming and labor-intensive. Although there are a few semi-supervised studies to solve this problem, they cannot avoid transferring information from another widely explored imaging modality (cross-modality domain shift) to assist ICA segmentation, causing a decline in generalization of the network [10].

To address these drawbacks, we propose an efficient framework (Fig. 1(c)) for whole coronary arterial tree seg-

mentation, inspired by the combination of elastic interaction energy loss [11] and semi-supervised Mean Teacher (MT) model [12] with UNet++ [13] (Fig. 1(a)) as its backbone.

2. METHOD

Our mean teacher framework consists of a student model and a teacher model. The student model is optimized on the combination of supervised and unsupervised losses. The parameters are then updated in the teacher model using exponential moving average. The framework extrapolates the information in unlabeled data by quantifying the consistency between predictions from the student model and the teacher model.

We denote x as the ICA image and y as the ground truth, where $x \in [0, 255]^{H \times W}$ and $y \in \{0, 1\}^{H \times W}$ with H and W representing the heights and widths, respectively. We denote the labeled set as $D_L = \{(x_n, y_n)\}_{n=1}^N$ and the unlabeled input as $D_U = \{x_m\}_{m=1}^M$, where N and M are the numbers of total labeled and unlabeled data, respectively. The segmentation by student model and teacher model are defined as $f_s(\cdot)$ and $f_t(\cdot)$, respectively. The weights of the network and the added noise are represented as θ and τ , respectively.

2.1. UNet++

The simplified structure of UNet++ is shown in Fig. 1(a). We have used VGG convolution blocks for $X^{i,j}$ (i, j representing the location of the block) with each of them formed by convolution, batch normalization and Rectified Linear Unit [14]. Each of the added red blocks on the conventional skip-connections behaves like the decoder blocks, which builds 3 sub-UNets to make the network denser. Random dropouts are added before $X^{4,0}$ and $X^{0,4}$ for the calculation of unsupervised loss.

2.2. Mean teacher model

In Fig. 1(c), the mean teacher model generates three outputs ($f_s(x_n), f_s(x_m), f_t(x_m)$) at each iteration. Firstly, we train the student model with labeled x_n and unlabeled x_m with Gaussian noise τ_s added to the inputs, giving us $f_s(x_n; \theta_s)$ and $f_s(x_m; \theta_s, \tau_s)$. However, only $f_s(x_n; \theta_s)$ is used to calculate the supervised loss. The same x_m but with different Gaussian noise is then fed into the teacher model under random dropout, producing $f_t(x_m; \theta_t, \tau_t)$. The consistency distance D is calculated as a mean squared error after softmax $S(\cdot)$ of both predictions:

$$D = \| S(f_s(x_m; \theta_s, \tau_s), f_t(x_m; \theta_t, \tau_t)) \|^2. \quad (1)$$

Secondly, we use the same x_m and replicate it K times to generate K sets of predictions from the teacher model with different Gaussian noise, as $\{f_{t,k}(x_m; \theta_t, \tau_{t,k})\}_{k=1}^K$. We obtain the softmax of these predictions and average along the total number of output channels to get a probability map μ . The

uncertainty u is calculated from predictive entropy as [15]

$$u = -\mu \cdot \log(\mu + \xi), \quad (2)$$

where ξ is a very small positive value. The uncertainty map u is at image level and has the same shape as the ground truth. Hence, it can be regarded as an “unlabeled” ground truth.

We adopt the idea of ensembling on weights from [16] to reduce the variance between multiple versions of the teacher model. The weights of the teacher model are updated as an exponential moving average (EMA) of the student weights. Mathematically, $\theta_t = \lambda \theta'_t + (1 - \lambda) \theta_s$, where λ is the hyperparameter that controls the rate of EMA decay, and θ'_t represents the weight of teacher model from last iteration.

2.3. Supervised loss function

The supervised loss function was previously proposed in [11]. It is inspired by the elastic energy of dislocations in crystals and uses level sets to express the topological changes. The energy system of the solid curve $\gamma(x(s), y(s), z(s))$ in Fig. 1(b) consists of three key functions defined as [17]:

$$\nabla \times \vec{w} = \delta(\gamma) \vec{\tau} \quad (3)$$

$$\vec{w}(x, y, z) = -\frac{1}{4\pi} \int_{\gamma} \frac{\vec{r} \times d\vec{l}}{|\vec{r}|^3} \quad \text{and} \quad E = \frac{1}{8\pi} \int_{\gamma} \int_{\gamma'} \frac{d\vec{l} \cdot d\vec{l}'}{|\vec{r}|} \quad (4)$$

where $\delta(\cdot)$ is a delta function in 2D, $\vec{\tau}$ is the unit tangent vector of γ , $d\vec{l} = \vec{\tau} \delta(\gamma) dx dy dz$ is an infinitely small part of the curve, $\vec{w} = (0, 0, w_3)$ is a vector perpendicular to the image plane ($w_1 = w_2 = 0$) and $\vec{r} = (x - x(s), y - y(s), z - z(s))$ is a vector between the point (x, y, z) and a point $(x(s), y(s), z(s))$ on $\gamma(s)$. We define the stationary boundary of the ground truth and the moving boundary of the prediction in Fig. 1(b) as γ_1 and γ_2 respectively. The γ in Eqs. (3)-(4) is substituted with $\gamma_1 \cup \gamma_2$ when applying this system on image segmentation where a collection of curves is involved. By expressing the moving γ_2 with the evolution equation of a level set ϕ and the velocity with $\vec{\tau}$ and \vec{w} , the velocity of the moving boundary can be expressed as:

$$v(x, y) = -\frac{1}{4\pi} \int_{\mathbb{R}^2} \frac{\vec{r} \cdot \nabla(G_s + \alpha H(\phi))}{|\vec{r}|^3} dx dy, \quad (5)$$

where G_s is γ_1 obtained by convolving the ground truth y_n with a 2D Gaussian function, and $H(\phi)$ is a regularized Heaviside function. Similarly, the total energy can be derived from Eq. (4) and we define $G_s + \alpha H(\phi) = T(x, y)$ for convenience:

$$E = \frac{1}{8\pi} \int_{\mathbb{R}^2} dx dy \int_{\mathbb{R}^2} \frac{\nabla T(x, y) \cdot \nabla T(x', y')}{|\vec{r}|} dx' dy'. \quad (6)$$

Eq. (6) is the supervised loss function L_{sup} for this energy system and Eq. (5) is the corresponding gradient of the loss function. To boost the efficiency during back propagation, Eqs. (5) and (6) can be further simplified by transforming them into 2D Fourier space.

2.4. Unsupervised loss function

For unsupervised loss, the idea is to filter and preserve the uncertainty map in Eq. (2) as smaller than a threshold T_{sig} to obtain a mask. We define T_{sig} as a sigmoid ramp-up function of the uncertainty: $T_{sig} = 0.5 * u_{max} (1 + e^{-5(1 - \frac{t_{now}}{t_{max}})^2})$, where u_{max} is the maximum uncertainty value, t_{now} is the current number of iterations and t_{max} is the maximum number of iterations in training. The loss value is calculated by summing up the confidence value within the mask for consistency distance (Eq. (1)) to divide by its total remaining uncertainty. Formally, the unsupervised loss function is summarized as:

$$L_{unsup} = \frac{\sum(u < T_{sig}) \cdot D}{2 \cdot \sum(u < T_{sig}) + \xi}, \quad (7)$$

where the summation is over all pixels. Therefore, the goal of this semi-supervised framework for ICA segmentation is to minimize the weighted sum of two loss functions:

$$\min_{\theta_s} [L_{sup}(f_s, y_n) + \lambda_{unsup} L_{unsup}(f_s, f_t)], \quad (8)$$

where λ_{unsup} is the hyper-parameter that controls the balance between supervised and unsupervised loss.

3. EXPERIMENTS AND RESULTS

3.1. Dataset and implementation details

The raw data are sequences of ICAs accompanied by synchronized electrocardiograms (ECGs) at multiple projection planes from 36 patients in the John Radcliffe Hospital, Oxford. After being anonymized in the hospital, each image is selected as the ICA end-diastolic frame based on ECG gating, with size of $H \times W = 512 \times 512$. The dataset consists of 118 ICA frames with 73 right coronary artery (RCA) images and 45 left anterior descending artery (LAD) images (14 patients provide both RCA and LAD images, 5 patients provide LAD only and 17 patients provide RCA only). Among them, 27 RCA images and 31 LAD images are randomly selected and manually annotated by experts to make up the labeled set D_L . The remaining 60 frames, corresponding to different subjects to those in the training set, are assigned to unlabeled set D_U . The labeled set D_L is further randomly divided into 3 groups for training, validation and testing containing 42 frames (18 RCAs and 24 LADs), 6 frames (3 RCAs and 3 LADs) and 10 frames (6 RCAs and 4 LADs), respectively.

The training of the network is performed on an NVIDIA V100 Tensor Core GPU. The weight of the unsupervised loss function in Eq. (8) is specified as $\lambda_{unsup} = 0.1 \cdot e^{-5(1 - \frac{t_{now}}{t_{max}})^2}$ similar to the ramp-up for T_{sig} . At the beginning of training, λ_{unsup} is very small; then it gradually increases. This forces the network to depend on supervised loss at an early stage of training, as the teacher model is not ready yet and only generates arbitrary predictions. We perform a random grid search

and set hyper-parameter λ for EMA to 0.99 and α for Heaviside function to 0.3. The last hyper-parameter K for uncertainty estimation is set to be 10 limited by the GPU memory, and the dropout is only activated when training. We optimize the network using stochastic gradient descent with weight decay 10^{-6} and momentum 0.7. The initial learning rate is set to be 10^{-5} and controlled by a cosine annealing scheduler to gradually decrease to its min value 10^{-8} with respect to the numbers of epochs. The training lasts for 10000 iterations for the assured convergence of the loss function. During training, the batch size is 10 consisting of 4 labeled and 6 unlabeled images. All samples are augmented on-the-fly with standard random rotation, flipping, changing saturation and changing contrast. The convolution operation size is set to be 3×3 . The depth of UNet++ is kept at 4 after a pruning study of the network.

For evaluation, we use Dice score, recall, over segmentation $H(f|y_n)$, under segmentation $H(y_n|f)$, and Betti number error [18]. The Betti number error is added specifically to evaluate the effect of the elastic energy loss function. For images, it directly compares the topology (the number of connected components) between them. To verify the effectiveness of the proposed methods, 8 types of networks are included in the experiments covering training using Dice loss versus elastic energy loss (L_{sup}), U-Net vs UNet++, and supervised model vs mean teacher (MT) model. The quantitative results are presented in Table 1. The result from Vanilla U-Net (Dice) and MT-UNet++ L_{sup} (proposed method) are shown in Fig. 2(a).

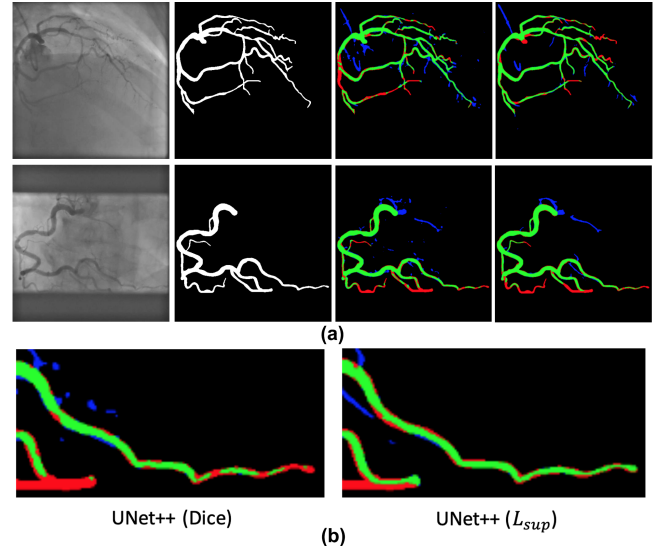


Fig. 2: (a) An example of the LAD (top) and RCA (bottom) with the original image, ground truth, results of U-Net with Dice loss and MT-UNet++ with energy loss L_{sup} , placed from left to right. True positive, false positive, and false negative are in green, blue, and red, respectively. (b) Zoomed view of distal part of a RCA segmented using the Dice loss and L_{sup} , respectively, in UNet++ framework.

Table 1: Comparison between the proposed method and state-of-the-art approaches with increasing framework complexity. All metrics are calculated after applying a mean threshold on predictions to identify the intensity of background for binarization.

Framework	Loss	Numbers of samples		Metrics				
		Labeled	Unlabeled	Dice (%)	Recall (%)	Over (%)	Under (%)	Betti
U-Net	Dice	42	0	77.34±2.42	79.05±5.95	14.48±3.77	13.93±3.30	22.76±10.72
	L_{sup}	42	0	79.45±2.01	80.12±5.97	13.15±3.64	11.29±2.38	8.25±4.57
UNet++	Dice	42	0	78.84±1.83	80.23±5.67	14.38±3.68	14.15±3.14	27.89±12.72
	L_{sup}	42	0	80.86±3.63	81.27±6.33	11.01±2.99	11.67±2.43	7.28±4.22
MT-UNet	Dice	42	60	76.81±2.24	81.89±5.27	14.83±2.86	12.03±2.77	30.8±12.27
	L_{sup}	42	60	78.59±2.64	81.48±5.42	13.40±2.03	11.64±2.42	7.15±4.81
MT-UNet++	Dice	42	60	79.52±1.97	82.06±5.16	13.09±2.54	11.20±2.49	20.48±6.68
	L_{sup}	42	60	81.66±1.89	83.89±4.81	11.19±2.33	10.62±1.84	6.83±3.87
SS-CADA [10]		60	92	78.4±4.60	83.27±3.95	N/A	N/A	N/A

3.2. Experimental results

The performance of our method surpasses those compared in terms of Dice score, recall, under segmentation and Betti number error. The extra information brought by the unlabeled data is clearly useful with UNet++ as the backbone. UNet++ combined with elastic loss L_{sup} creates the best delineation of vessel boundaries with minimum under segmentation achieved. The improvement by introducing the elastic loss L_{sup} on disconnected vessels is obvious from the Betti number error and can be seen clearly in Fig. 2(b) where the distal parts of the RCA are connected. To be noticed that the proposed method can capture even smaller vessels quite accurately, which have not been annotated by the expert in the ground truth (marked as blue false positives in Fig. 2). Our work also compares favourably with the previous state-of-the-art method Semi-Supervised Cross-Anatomy Domain Adaptation (SS-CADA) [10], as shown in Table 1. It should be noted that the results are not perfectly comparable since both number and type of training samples used are different.

3.3. Ablation study

In this study we focus on the MT-UNet++ (L_{sup}) and UNet++ (L_{sup}) to fully explore the effect of using semi-supervision on different numbers of labeled data N . Both frameworks are trained with $N = \{4, 6, 8, \dots, 40\}$ and the same set-ups. In the results presented in Fig. 3, mean teacher demonstrates strong advantages in Dice and under segmentation at small numbers of labeled data. The gap in Dice drops after around 16 labeled samples, but the recall continues to show an advantage. Only minor differences can be observed in over segmentation and no improvements can be found with respect to Betti number error. Thus, we observe the largest performance boost at smaller numbers of labeled samples, as expected.

4. CONCLUSIONS

In this paper, we have proposed a new semi-supervised framework for coronary artery segmentation in ICA. We have em-

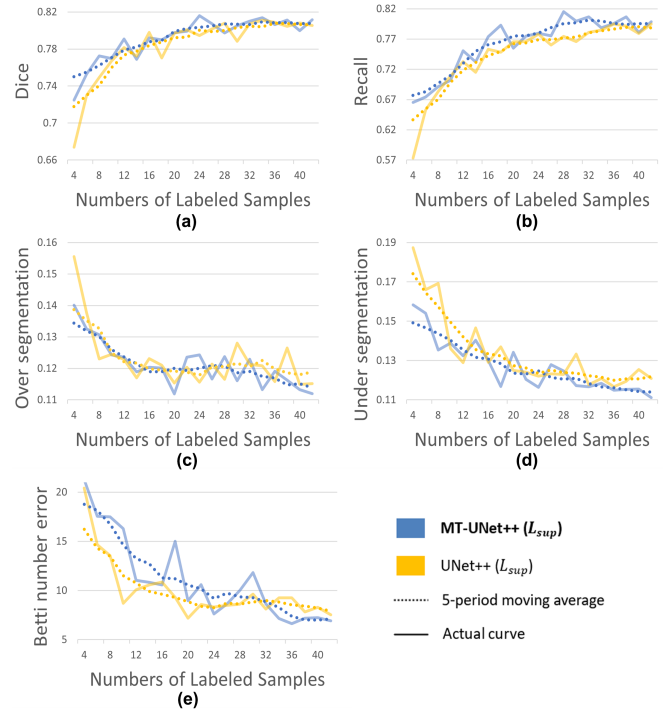


Fig. 3: The performance of MT-UNet++ (L_{sup}) and UNet++ (L_{sup}) over varying numbers of labeled samples.

bedded UNet++ with connectivity preserving loss function into a mean teacher model. Our method reduces the problem of disconnected vessels. Also importantly, it exploits the useful information in the unlabeled data under perturbation. UNet++ further boosts the performance of the framework. The proposed approach yields the Dice score of 81.66%, recall of 83.89% and Betti number error of 6.83, while the ablation study shows the largest improvement occurring at the number of labeled samples smaller than 16.

5. COMPLIANCE WITH ETHICAL STANDARDS

This is a numerical simulation study for which no ethical approval was required.

6. ACKNOWLEDGMENTS

The authors express no conflict of interest. The work of A. Banerjee was supported by the British Heart Foundation (BHF) Project under Grant HSR01230. The work of M. Beetz was supported by the Stiftung der Deutschen Wirtschaft (Foundation of German Business). The work of V. Grau was supported by the CompBioMed 2 Centre of Excellence in Computational Biomedicine (European Commission Horizon 2020 research and innovation programme, grant agreement No. 823712).

7. REFERENCES

- [1] Viktor Kočka, “The coronary angiography – An old-timer in great shape,” *Cor et vasa*, vol. 57, no. 6, pp. e419–e424, 2015.
- [2] Abhirup Banerjee, Francesca Galassi, Ernesto Zacur, Giovanni Luigi De Maria, Robin P. Choudhury, and Vicente Grau, “Point-cloud method for automated 3D coronary tree reconstruction from multiple non-simultaneous angiographic projections,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 1278–1290, 2020.
- [3] Jonathan Long, Evan Shelhamer, and Trevor Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2015, pp. 3431–3440.
- [4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention (MICCAI)*. Springer, 2015, pp. 234–241.
- [5] Kevis-Kokitsi Maninis, Jordi Pont-Tuset, Pablo Arbeláez, and Luc Van Gool, “Deep retinal image understanding,” in *International conference on medical image computing and computer-assisted intervention (MICCAI)*. Springer, 2016, pp. 140–148.
- [6] Ebrahim Nasr-Esfahani, Nader Karimi, Mohammad H Jafari, et al., “Segmentation of vessels in angiograms using convolutional neural networks,” *Biomedical Signal Processing and Control*, vol. 40, pp. 240–251, 2018.
- [7] Jingfan Fan, Jian Yang, Yachen Wang, et al., “Multi-channel fully convolutional network for coronary artery segmentation in x-ray angiograms,” *IEEE Access*, vol. 6, pp. 44635–44643, 2018.
- [8] Su Yang, Jihoon Kweon, Jae-Hyung Roh, et al., “Deep learning segmentation of major vessels in x-ray coronary angiography,” *Scientific Reports*, vol. 9, pp. 16897, 2019.
- [9] Benjamin Au, Uri Shaham, Sanket Dhruva, et al., “Automated characterization of stenosis in invasive coronary angiography images with convolutional neural networks,” *arXiv preprint arXiv:1807.10597*, 2018.
- [10] Jingyang Zhang, Ran Gu, Guotai Wang, Hongzhi Xie, and Lixu Gu, “SS-CADA: A semi-supervised cross-anatomy domain adaptation for coronary artery segmentation,” *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 1227–1231, 2021.
- [11] Yuan Lan, Yang Xiang, and Luchan Zhang, “An elastic interaction-based loss function for medical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2020, pp. 755–764.
- [12] Lequan Yu, Shujun Wang, Xiaomeng Li, et al., “Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2019, pp. 605–613.
- [13] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, et al., “Unet++: A nested u-net architecture for medical image segmentation,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3–11. Springer, 2018.
- [14] Karen Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [15] Alex Kendall and Yarin Gal, “What uncertainties do we need in Bayesian deep learning for computer vision?,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)*, 2017, pp. 5580–5590.
- [16] Antti Tarvainen and Harri Valpola, “Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)*, 2017, pp. 1195–1204.
- [17] Yang Xiang, Albert C.S. Chung, and Jian Ye, “An active contour model for image segmentation based on elastic interaction,” *Journal of computational physics*, vol. 219, no. 1, pp. 455–476, 2006.
- [18] Xiaoling Hu, Li Fuxin, Dimitris Samaras, and Chao Chen, “Topology-preserving deep image segmentation,” *arXiv preprint arXiv:1906.05404*, 2019.