Adaptive Contour Noise Generation Network for Semi-supervised Medical Data Segmentation

Shunlong Ye
Dept. of Computer Science and Engineering
Nanjing University of Science and Technology
Nanjing, China
yeshunlong@njust.edu.com

***Abstract*—When training convolutional neural networks, noise distribution in data can affect segmentation accuracy. Therefore, many neural networks will strengthen the segmentation ability of the model by artificially adding noise. However, because the noise distribution of the input data is not known, the model training process can only generate random noise, e.g., gaussian noise, but cannot simulate real-world noise. In the medical imaging domain, more attention is paid to the image contours, which can also be used for noising simulation. Meanwhile, considering the existence of rich unlabeled data sets in medical images, this paper proposes a semi-supervised framework for uncertainty estimation by using Monte Carlo Dropout and generates adaptive noise using the uncertainty map. Specifically, the framework consists of three parts, including a noise generation model, a teacher model, and a student model. The noise generation part processes the data input to the other two parts and uses the contour feature identified by the posteriori uncertainty map to simulate priori noise to guide the process of training segmentation model. Meanwhile, the student model updates its parameters based on the teacher model output and the ground truth annotations by minimizing a joint loss. Segmentation experiments on left atrium 3D data show that our method has competitive results with other SOTA semi-supervised methods.**

***Index Terms*—semi-supervised learning, uncertainty estimation, contour noise generation, medical image segmentation**

I. Introduction

Automatic segmentation of medical images takes an important part in the analysis of many diseases nowadays. However, in medical imaging domain, delineating clear annotations from 3D medical images often requires cumbersome labeling, which means it's a costly process to acquire labeled data and requires the assistance of experienced experts. Considering there are plenty of unlabeled data, we think implementing a semi-supervised segmentation method can help to make the best use of the limited labeled data and the rich unlabeled data.

In the medical imaging community, many attempts have been made to utilize unlabeled data to make segmentation network achieves a higher performance. Adversarial learning has been adopted during the semi-supervised training process [1]–[3]. Zhang *et al.* [3] utilized unlabeled images by maximizing the consistency of segmentation results between unlabeled images and labeled images. Nie *et al.* [2] adopted adversarial networks to find trustworthy regions to crank up the segmentation performance. Wang *et al.* [4] overcame the effect of noisy pseudo labeling by incorporating Cycle-GAN's noise-robust learning method. Considering that self-ensembling methods [5], [6] have achieved good results on unlabeled data, Li *et al.* [7] utilized the $\prod$ model [5] to doing skin lesion segmentation task. Other approaches [8], [9] used weighted average consistency targets in the field of MR segmentation. Multi-task and multi-view learning methods have also received attention. Luo *et al.* [10] jointly predicted pixel-level segmentation maps and geometric-aware level set representations of targets using a dual-task deep network. Li *et al.* [11] jointly predicted semantic segmentation of object surfaces and signed distance maps (SDM) by developing a multi-task deep network. Xiao *et al.* [12] designed a model consisting of three deep convolutional streams to process different views. Considering that self-attention can better capture contextual relationships, Liu *et al.* [13] enhanced the efficiency of segmentation by introducing the attention mechanism. Luo *et al.* [14] improved the segmentation performance by combining CNN and Transformer. Yu *et al.* [15] presented a network with two identically structured models by drawing on the mean teacher idea. The model takes both roles of teacher and student. When it acts as a teacher, it generates student's learning goals. Meanwhile, it utilizes the targets generated by the other model as a student. Also, since the same architecture is adopted, the learning process does not require using a shared weight strategy. This overcomes the limitations of temporal ensembling, while the goal model has a better intermediate representation since weight averaging improves the output of all layers. The Monte Carlo Dropout method used to estimate the uncertainty of inference results was presented by Gal *et al.* [16]. Since the method was presented, several works have tried to utilize or improve its performance. Shamsi *et al.* [17] used the uncertainty in the loss to make the model more accurate and repeatable. Fabi *et al.* [18] found that it can also represent the uncertainty of intermediate features to obtain the feature map better. However, these methods can only improve the robustness of the model by generating random gaussian noise because the noise distribution in input data is unknown, and they cannot simulate real-world noise to guide the model training. Moreover, Monte Carlo Dropout utilized in these works was only used for uncertainty estimation in the posteriori process and was not considered a guide for priori noise generation.

In this paper, as shown in Fig.1, we illustrate a new network using the uncertainty-aware approach and semi-supervised

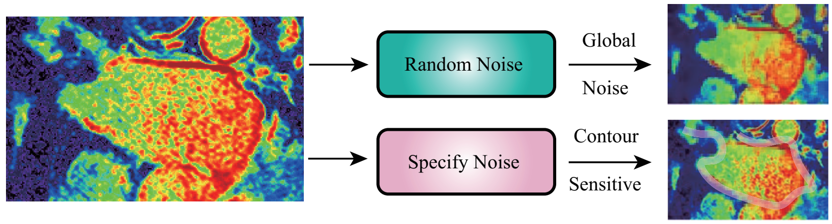


Fig. 1. Comparison of methods for artificially adding noise to strengthen the segmentation ability of the model. Conventional methods can only generate random gaussian noise because the noise distribution in input data is unknown and cannot simulate real-world noise. Our method can simulate the noise associated with image contours during the priori process.

framework. The framework generates adaptive noise by estimating model uncertainty using Monte Carlo Dropout. Our approach achieves higher segmentation performance by drawing on the same mindset as the mean teacher spirit [6]. The framework consists of three parts, including a noise generation model, a teacher model, and a student model, where the noise generation model simulates the priori noise containing contour feature determined by the uncertainty of the inference process. Our method is evaluated on MICCAI 2018 Atrial Segmentation Challenge dataset[1]. The results show that our semi-supervised method has competitive results with other SOTA semi-supervised segmentation methods.

II. Method

Fig.2 illustrates our proposed segmentation framework based on adaptive noise generation. The input image is first fed to a noise generation model to generate adaptive noise containing contour features. Generated noise will be combined with all unlabeled data and passed to the teacher model for inference. The inference results are again passed through the noise generation model, and the uncertainty of the inference process is estimated. All the image patches in the input data will be considered as containing noise and fed to the student model for prediction. The parameters of the final noise generation model are updated by the uncertainty map calculated using the Monte Carlo Dropout method. Meanwhile, after receiving inputs containing contour noise, the student model updates its parameters based on the teacher model output and the ground truth annotations by minimizing a joint loss, which consists of consistency loss on the unlabeled data compared with the target of the teacher model and segmentation loss of the supervised process. In this regard, uncertainty is estimated using Monte Carlo sampling, and generate the contour feature based on the uncertainty of the network prediction and used as the ground truth of the noise generation model. Guided by the estimated uncertainty, unreliable predictions can be filtered out, and noise with more prominent contour features can be generated. As a result, the weighted parameters of the student model are further optimized. In turn, we update the weighted parameters of the teacher model by adopting the *exponential moving average* method, improving robustness while encouraging the teacher model to perform a more accurate segmentation. In this process, noise generation, student, and teacher models adopt the same backbone, which is modified from VNet, and the basic content remains the same as traditional VNet, with down-sampling and up-sampling modules in successive convolutional and deconvolutional blocks and skip connections between each block. Adaptive noise or prediction is finally generated.

A. Semi-Supervised Learning

We formally present the semi-supervised segmentation learning process. Assume the input dataset is a combination of labeled data and unlabeled data with size N and M, respectively. input $x_i \in \mathbb{R}^{H\times W\times D}$ and target annotation $y_i \in \{0,1\}^{H\times W\times D}$. Our segmentation framework is updated by minimizing the following joint loss function.

$$Loss = \sum_{i=1}^{M} L_c(f(x_i + noise); y_i) + \lambda \sum_{i=1}^{N} L_s(f(x_i); y_i) \quad (1)$$

Here unsupervised consistency loss is L_c, which is used to evaluate the consistency of the predictions of the teacher model and the student model given the identical data under added noise. Supervised loss is L_s, which is used to measure the quality of the model's prediction on labeled data.

B. Noise Ground Truth Generating

More attention paid to image contours in the medical image can boost segmentation performance so that contour-weighted generation noise can be considered in the a priori process. Monte Carlo Dropout estimation method is used to obtain the approximate contour of the image at the input data stage to calculate a posteriori uncertainty map and use it as the target when training the noise generation model.

Monte Carlo Dropout method for estimating model uncertainty. In the concrete implementation of the method, we propagate the model T times and add noise generated by the noise generation model during each propagation. Assume the probability of the c-th class after the k-th propagation is p_k^c. Then the overall uncertainty u can be obtained following the below equation.

$$\mu_i = \frac{1}{T} \sum_k p_k^c \quad (2)$$

$$u = -\sum \mu_i log \mu_i \quad (3)$$

[1]http://atriaseg2018.cardiacatlas.org/

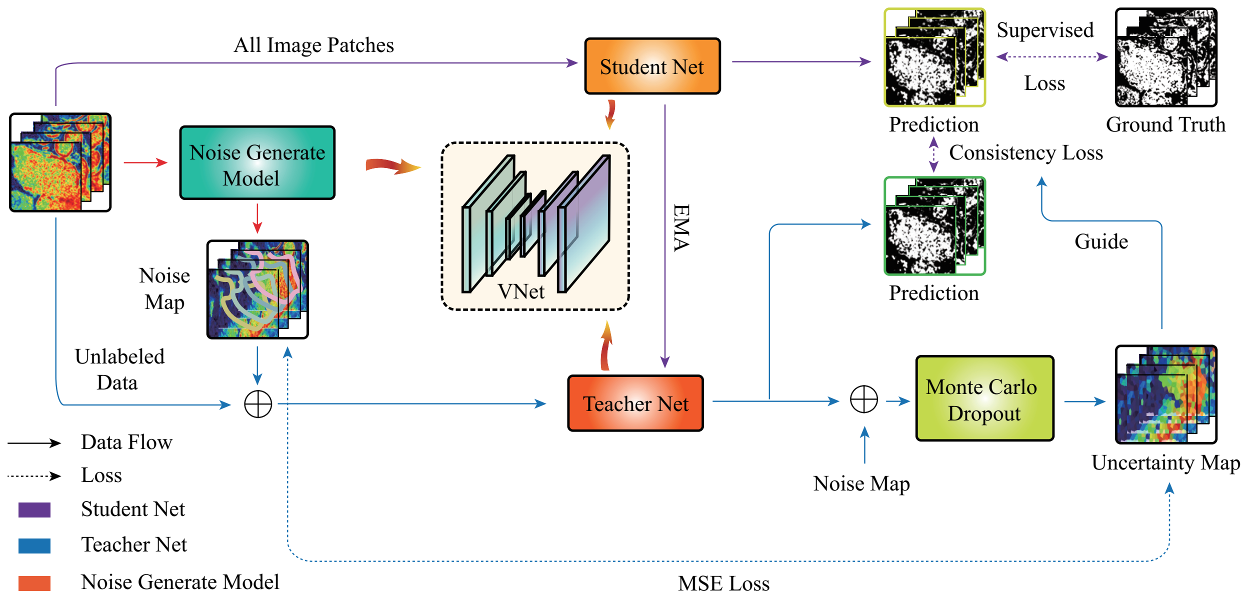


Fig. 2. Our pipeline of proposed segmentation framework based on adaptive noise generation. Noise generation model adds adaptive noise to simulate the contour feature in input data to enhance the model performance. The student model updates its weighted parameters by minimizing joint loss including consistency loss on unlabeled data and supervised loss on labeled data. Meanwhile the teacher model help the student model to make more trustworthy prediction by estimating the uncertainty of the target. EMA is short for *exponential moving average*.

Uncertainty map guides the consistency loss. The consistency loss will only be calculated for voxels below a given threshold on the uncertainty map when the student model updates its parameters. The calculation equation is given by the following.

$$L_c = \frac{\sum_i \parallel (u_i \leq H) MSELoss(f_t, f_s)}{\sum_i \parallel (u_i \leq H)} \tag{4}$$

Here $I(\parallel)$ denotes the indicator function. f_t and f_s denote the predicted values of the teacher and student models at voxel i, respectively. the uncertainty u at voxel i is u_i, and a threshold value H is set to select the most reliable target.

C. Technique Details

We randomly crop one million sub-volumes as network inputs. Standard data enhancement techniques are used to improve segmentation performance, including random flips and random rotations. SGD optimizer is used to update the model parameters. Specifically, we adopt β as the *exponential moving average* decay to control the update rate and modify the teacher model's weighted parameters. A gaussian heating function changing depending on time is used following the process of [5], [6] to solve the different contributions of unsupervised consistency loss and supervised loss. $T = 8$ is set to trade off the quality of uncertainty estimation against training efficiency. This strategy is also used to update the uncertainty threshold H when calculating the consistency loss. Other important hyperparameters are listed as follows. Batch size is set to 32. The maximum number of training iterations is set to 6000, where every compared method convergences on this state. Learning rate is set to 0.01 initially and updated every 2500 iterations using this formula $lr = 0.01 * 0.1^{\frac{iteration\ number}{2500}}$.

III. Experiments

A. Dataset

Our method is evaluated on the dataset of the 2018 Atrial Segmentation Challenge, which contains 100 MR imaging scans and annotated ground truth for learning. We divide these 100 scans into 80 for training and 20 for evaluation. All scans are pre-processed using the standard data processing techniques including normalizing, unit variance, and center cropping.

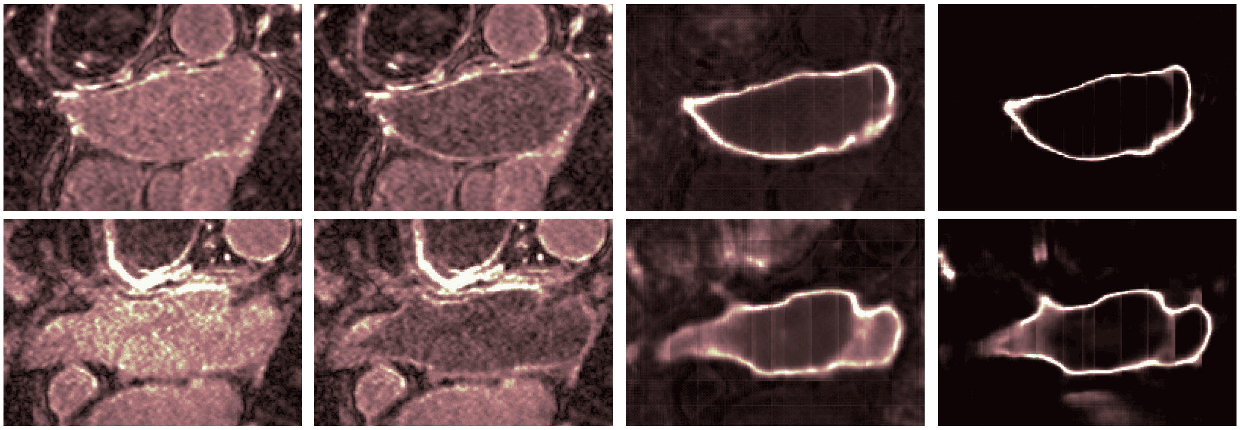
B. Evaluation of Our Method

Fig.3 shows the segmentation examples of our semi-supervised approach, the visualization results of the noise map and uncertainty map in two test samples. It can be noticed that the a priori noise map in Fig.3 (c) can simulate the Fig.3 (d) posteriori uncertainty map well.

We evaluate our method using four metrics: Dice, average surface distance (ASD), Jaccard, and 95% Hausdorff Distance (95HD). Among the training samples, we divided labeled data and unlabeled data with a ratio of 1:5, which means most of the data are unlabeled data. Table I shows the results of our method compared with other methods. Our method has the best result for Dice, and ASD, Jaccard, 95HD perform equally well.

IV. Conclusion

We propose a new semi-supervised framework by generating adaptive noise through uncertainty estimation for segmentation in medical imaging domain. By utilizing the posteriori uncertainty map, we simulate the contour feature during the priori process to guide the model training. Meanwhile, the



(a) Image (b) Our Method (c) Noise Map (d) Uncertainty Map

Fig. 3. Visualization of results, including segmentation masks, adaptive noise map, and uncertainty map. The color brightness indicates the prediction probability. Our method can segment the target better, while the adaptive noise map can represent the contour feature well.

TABLE I
RESULTS OF OUR METHOD COMPARED WITH OTHER METHODS.

Method	Dice[%]	ASD[voxel]	Jaccard[%]	95HD[voxel]
Self-training [19] 2017	86.92	2.21	77.28	9.19
DAN [3] 2017	87.52	2.42	78.29	9.01
ASDNet [2] 2018	87.90	2.08	78.85	9.24
TCSE [7] 2018	88.15	2.44	79.20	9.57
UA-MT [15] 2019	88.83	3.12	80.13	**7.32**
DTC [14] 2021	88.52	2.08	79.74	8.54
Attention V-Net [13] 2022	89.08	**1.94**	**82.48**	8.22
Our method	**89.25**	2.29	80.71	7.49

mean teacher structure helps to make the best use of the rich unlabeled data. We demonstrate the effectiveness of our method by comparing it with other SOTA semi-supervised methods. Future work includes analyzing the impact of other noise generation backbones and applying our framework to other semi-supervised segmentation tasks in the medical imaging domain.

REFERENCES

[1] Dong N, Kampffmeyer M, Liang X, Wang Z, Dai W and Xing E 2018 Unsupervised domain adaptation for automatic estimation of cardiothoracic ratio *International conference on medical image computing and computer-assisted intervention* (Springer) pp 544–552

[2] Nie D, Gao Y, Wang L and Shen D 2018 Asdnet: attention based semi-supervised deep networks for medical image segmentation *International conference on medical image computing and computer-assisted intervention* (Springer) pp 370–378

[3] Zhang Y, Yang L, Chen J, Fredericksen M, Hughes D P and Chen D Z 2017 Deep adversarial networks for biomedical image segmentation utilizing unannotated images *International conference on medical image computing and computer-assisted intervention* (Springer) pp 408–416

[4] Wang L, Guo D, Wang G and Zhang S 2020 *IEEE Transactions on Medical Imaging* **40** 2795–2807

[5] Laine S and Aila T 2016 *arXiv preprint arXiv:1610.02242*

[6] Tarvainen A and Valpola H 2017 *Advances in neural information processing systems* **30**

[7] Li X, Yu L, Chen H, Fu C W and Heng P A 2018 *arXiv preprint arXiv:1808.03887*

[8] Cui W, Liu Y, Li Y, Guo M, Li Y, Li X, Wang T, Zeng X and Ye C 2019 Semi-supervised brain lesion segmentation with an adapted mean teacher model *International Conference on Information Processing in Medical Imaging* (Springer) pp 554–565

[9] Perone C S and Cohen-Adad J 2018 Deep semi-supervised segmentation with weight-averaged consistency targets *Deep learning in medical image analysis and multimodal learning for clinical decision support* (Springer) pp 12–19

[10] Luo X, Chen J, Song T and Wang G 2021 Semi-supervised medical image segmentation through dual-task consistency *Proceedings of the AAAI Conference on Artificial Intelligence* vol 35 pp 8801–8809

[11] Li S, Zhang C and He X 2020 Shape-aware semi-supervised 3d semantic segmentation for medical images *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer) pp 552–561

[12] Xiao J, Si D, Wu Y, Li M, Yin J and Ding H 2020 Multi-view learning for 3d lge-mri left atrial cavity segmentation *Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Pattern Recognition* pp 84–87

[13] Liu X, Yin R and Yin J 2022 *Applied Sciences* **12** 3764

[14] Luo X, Hu M, Song T, Wang G and Zhang S 2021 *arXiv preprint arXiv:2112.04894*

[15] Yu L, Wang S, Li X, Fu C W and Heng P A 2019 Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer) pp 605–613

[16] Gal Y and Ghahramani Z 2016 Dropout as a bayesian approximation: Representing model uncertainty in deep learning *international conference on machine learning* (PMLR) pp 1050–1059

[17] Shamsi A, Asgharnezhad H, Abdar M, Tajally A, Khosravi A, Nahavandi S and Leung H 2021 *arXiv preprint arXiv:2110.03260*

[18] Fabi K and Schneider J 2020 *arXiv preprint arXiv:2008.01468*

[19] Bai W, Oktay O, Sinclair M, Suzuki H, Rajchl M, Tarroni G, Glocker B, King A, Matthews P M and Rueckert D 2017 Semi-supervised learning for network-based cardiac mr image segmentation *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer) pp 253–260