SEMI-SUPERVISED SEMANTIC SEGMENTATION VIA ENTROPY MINIMIZATION

Jiawei Wu[1,2], *Haoyi Fan*[3], *Xiaoqing Zhang*[2], *Shouying Lin*[1,*], *Zuoyong Li*[2,*]

[1]College of Mecanical and Electrical Engineering,
Fujian Agriculture and Forestry University, Fuzhou 350002, China
[2]Fujian Provincial Key Laboratory of Information Processing and Intelligent Control,
College of Computer and Control Engineering, Minjiang University, Fuzhou 350121, China
[3]School of Computer Science and Technology,
Harbin University of Science and Technology, Harbin 150080, China
{*wpwjw123, fzulzytdq*}*@126.com*.

ABSTRACT

In this paper, we propose a novel entropy minimization based semi-supervised method for semantic segmentation. Entropy minimization has proven to be an effective semi-supervised method for realizing the cluster assumption, where the decision boundary should lie in low-density regions. Inspired by the existing consistency training semi-supervised segmentation networks with encoder-decoder architecture, we found that there tend to be more large gradient values at the object edges than other positions in the feature map of the encoder, and therefore propose a feature gradient map regularization to enlarge inter-class distance in the feature space for low-entropy of segmentation prediction. Additionally, we introduce an adaptive sharpening scheme with aleatoric uncertainty, and a class consistency constraint regularization, to alleviate the interference of noise with pseudo labels. Extensive experiments on PASCAL VOC, PASCAL-Context, and Leukocyte datasets show that the proposed method achieves state-of-the-art semi-supervised semantic segmentation performance without almost additional calculations and network structures.

Index Terms— Semi-supervised learning, Semantic segmentation, Entropy minimization

1. INTRODUCTION

In recent years, with the development of deep supervised learning, various computer vision tasks have made significant progress and made impressive results. However, training a deep neural network requires a large amount of labeled data which acquisition is often time-consuming and expensive. Especially in semantic segmentation tasks, which require a large number of pixel-level labels, and labeling cost is 15 times and 60 times larger than that of region-level and image-level labels, respectively [1]. The cost of medical image segmentation is even more evident due to the need of professional annotations. Therefore, a growing attention is focused on weakly-supervised methods [2, 3] and semi-supervised methods [4, 5, 6, 7] .

In this work, we focus mainly on semi-supervised semantic segmentation which assumes that there is a large amount of unlabeled data and limited labeled data within the same distribution. The current dominant semi-supervised segmentation methods can be divided into Generative Adversarial Networks (GANs) based [4, 5] and consistency training based [6, 7, 8, 9] . The methods based on GANs extend the generic GAN framework to pixel-level predictions to try to make fake unlabeled data fool the discriminator. The consistency training methods expect the output of the network to be smooth under different perturbations. These methods have shown effectiveness in semi-supervised semantic segmentation. Nevertheless, these approaches suffer from some limitations. The methods based on adversarial training exploit the unlabeled data, but require careful design of additional network structures and can be hard to train. The methods of consistency training require additional calculations for multiple forward of every perturbation, and disturbances implicitly enhance data, which can be unfair to a fully-supervised network without data augmentation.

To address these limitations, we proposed a simple entropy minimization based method for semi-supervised semantic segmentation. With almost no additional network structure and calculations, we explore a series of regularizations to make the decision boundary in a low-density region. We visualized the gradient values of encoder feature maps to analyze the consistency training method [8], observed that gradient information at the edges will be larger than the baseline i.e. fully supervised network. Therefore, one possible reason for the effectiveness of consistency training methods is that it can increase the ability of discriminating the edges in high-dimensional space, which means it can enlarge the inter-class distance. We further observed that the gradient of low layers and high layers have local edges consistency, hence we propose feature gradient map regularization (FGMR) which uses gradient maps of low encoder layers to compensate the encoding ability of deep encoder layers. Meanwhile, inspired by Kendall et al. [10], the outputs of segmentation network are changed to mean and variance to obtain aleatoric uncertainty, which can measure noise of samples. Then, aleatoric uncertainty is used to generate pseudo labels for unlabeled data by the proposed adaptive sharpening. To further decrease the interference of noisy pseudo labels, we constrain the consistency of low-confidence predictions between segmentation and classification, in which two predictions are in the same network. Additionally, classification can refine the segmentation during inference.

The proposed method is simple, efficient and flexible. With extensive experiments, which demonstrate the effectiveness of our approach on leukocyte segmentation dataset and a range of popular semantic segmentation datasets: PASCAL VOC 2012 [11] and PASCAL-Context [12]. We obtain competitive results with almost no additional calculations and careful network design.

Specifically, our contributions are as follows:

*Corresponding authors: Shouying Lin, Zuoyong Li.

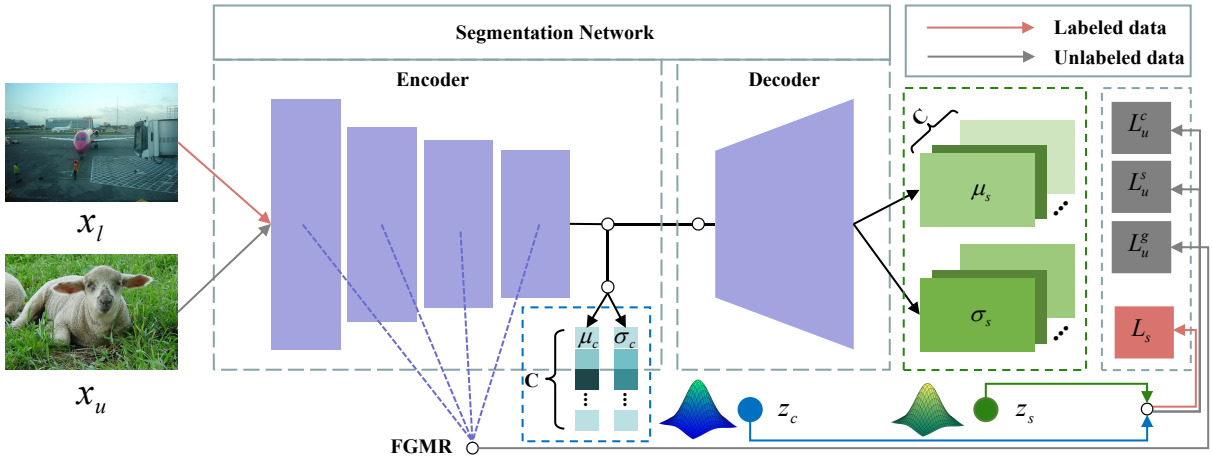


Fig. 1: Illustration of proposed approach. For one training iteration, sampling a batch of labeled data (x_l, y_s, y_c) together with a batch of unlabeled images x_u. All samples are passed through a segmentation network to obtain classification predictions and segmentation predictions. Supervised loss term L_s is calculated for labeled data, and L_u^g, L_u^c and L_u^s loss terms are calculated for unlabeled data.

- We first observe that consistency training essentially enhances the ability of discriminating image edges, and propose a feature gradient map regularization (FGMR) that uses gradient maps of low layers in the encoder to enhance the encoding ability of deep layers in the encoder.
- We propose an adaptive sharpening scheme, which keeps the decision boundary of unlabeled data in a low-density region.
- We propose a low-confidence prediction consistency regularization for low-level and high-level semantic information.
- Extensive experiments validate the superiority of the proposed method over the state-of-the-art methods.

2. RELATED WORK

2.1. Semi-Supervised Learning

Recently, people have made many efforts in semi-supervised field to overcome the shortcomings, such as consistency training methods [6, 7, 8, 13], graph based methods [14], and entropy minimization methods [15, 16, 17, 18]. In this paper, we focus mainly on entropy minimization methods. Minimize entropy, which is based on a rule: the decision boundary should not cross high-density region of marginal distribution. The specific method is to force the classifier to make low-entropy predictions on unlabeled data. For example, Grandvalet et al. [15] proposed a loss term which minimized the entropy of $p_{model}(y|x;\theta)$ for unlabeled data explicitly. Lee [17] does entropy minimization implicitly by constructing pseudo labels which are regard as training targets in a standard cross-entropy loss.

Similarly, the proposed method enforces an entropy minimization of predictions so that the decision boundary located in a low-density region.

2.2. Semi-Supervised Semantic Segmentation

A large number of the approaches based on consistency training achieved significant results on semi-supervised semantic segmentation tasks. Gerda et al. [7] propose an approach that learns consistency under transformations on both labeled and unlabeled data, in addition to supervised learning from labeled data. Chen et al. [8] introduced an auxiliary network to restrict the consistency of the reconstruction between labeled data and unlabeled data.

The Generative Adversarial Networks plays a vital role in computer vision and image processing, especially in semi-supervised semantic segmentation. Souly et al. [4] use the GAN to generate additional images to enhance the features learned by the segmentation network, and additional class-conditional images are extended to their semi-supervised method. Hung et al. [5] use an FCN-based discriminator which yields a dense probabilistic map for each pixel, which provides additional supervision signal make it can discover the trust-worthy regions. Different from the previous work, Mittal et al. [6] fused the original image and segmentation mask as the input of discriminator, and used feature matching loss [19] to improve semi-supervised semantic segmentation.

However, the consistency training methods need to perform forward pass two or more times or additional auxiliary network for computing the consistency loss. The GANs based methods are difficult to train and required careful design of the discriminator. These methods all require additional calculations or extra network structure, which increases training costs. In comparison, the proposed method exploits the unlabeled examples by enforcing an entropy minimization. We introduce a series of regularizations to semi-supervised semantic segmentation field, and show that the proposed method outperforms previous semi-supervised semantic segmentation methods with least additional costs.

3. METHOD

In this section, we first introduce the method overview of the proposed semi-supervised semantic segmentation approach. Then, we describe the details of the components in framework.

3.1. Method Overview

Fig. 1 shows an overview of the proposed method which can apply to any segmentation network. It only needs to make general minor changes to the existing segmentation network without careful

design. Suppose the input image size is $H \times W$, and the number of classes is C. The specific steps are to change the outputs to the mean $\mu_s \in \mathbb{R}^{H\times W\times C}$ and variance $\sigma_s^2 \in \mathbb{R}^{H\times W\times C}$ of the segmentation results. Similarly, output the mean $\mu_c \in \mathbb{R}^C$ and variance $\sigma_c^2 \in \mathbb{R}^C$ of the classification results in the last layer of encoder. In addition to the above-mentioned minor changes to the network, the other steps of this method are reflected in loss functions, which can be divided into supervised loss and unsupervised loss functions:

$$L = L_s + \lambda L_u \tag{1}$$

where L_s is the supervised loss, L_u is the unsupervised loss, and λ is a hyperparameter that adjusts the balance between supervised loss and unsupervised loss.

For labeled data $x_l \in \mathbb{R}^{H\times W\times 3}$ and the corresponding segmentation label $y_s \in \mathbb{R}^{H\times W\times C}$ and classification label $y_c \in \mathbb{R}^C$, x_l is sent to the network to obtain $< \mu_s, \sigma_s^2 >$ and $< \mu_c, \sigma_c^2 >$. Then, similar to previous work [20], sampling a data $\epsilon_s \setminus \epsilon_c$ from $\mathcal{N}(0, I)$, and convert it to $z_s = \epsilon_s \times \sigma_s + \mu_s \setminus z_c = \epsilon_c \times \sigma_c + \mu_c$ so that obtaining a sample in $\mathcal{N}(\mu_s, \sigma_s^2) \setminus \mathcal{N}(\mu_c, \sigma_c^2)$. The most commonly used cross-entropy loss is used to supervise both the segmentation results z_s and the classification results z_c by y_s and y_c, respectively. The supervised loss can be described as:

$$L_s = \sum_{H,W,C} \mathcal{H}(y_s^{H,W,C}, \alpha_s(z_s^{H,W,C})) + \sum_{C} \mathcal{H}(y_c^C, \alpha_c(z_c^C)) \tag{2}$$

where $\mathcal{H}(\cdot,\cdot)$ is cross-entropy loss function, and $\alpha(\cdot)$ is the activation function of last layer.

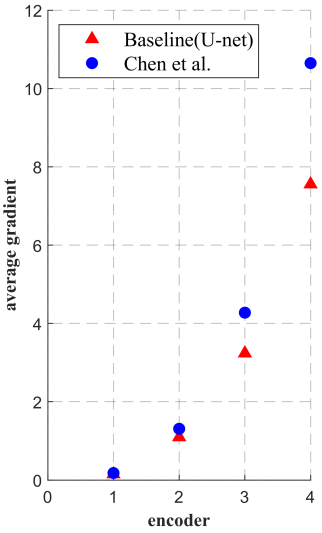
For unlabeled data x_u, we utilize feature gradient map regularization (FGMR) to enhance the edge gradient values of feature maps in encoder. Then we use variance as aleatoric uncertainty to search noise samples, which are used to guide adaptive sharpening to obtain pseudo-labels of unlabeled data, in which pseudo-labels which may bring noise are used to supervise unlabeled data. Even aleatoric uncertainty can filter some noise, most of the time the generated noise of pseudo-labels is more than the suppressed noise. To solve this problem, low-confidence categories in the classification results are further used to suppress the segmentation prediction of the corresponding categories to keep classes consistency. The adaptive sharpening loss and the class consistency loss will fight against each other to make the decision boundary in a low-density region to get the robust prediction result. The unsupervised loss function can be written as follows:

$$L_u = L_u^g + L_u^s + L_u^c \tag{3}$$

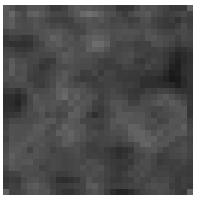
where L_u^g, L_u^s and L_u^c are the loss terms of FGMR, adaptive sharpening, and class consistency, respectively. The details of them will be introduced in Section 3.2, Section 3.3, and Section 3.4.

3.2. Feature Gradient Map Regularization

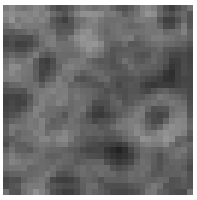
As shown in Fig. 2, the gradient statistics of different encoder layers describe that the ability of encoder to extract the edge information from the low layers to the high layers is gradually enhanced, and the average gradients of different encoder layers are significantly enhanced after the consistency training [8]. These results show that a great segmentation network expects to find out more edge information to improve segmentation predictions. Inspire of these observations, the key goal of semantic segmentation is how to improve the ability of encoder to discriminate the edges of objects. As shown in Fig. 2 (b) and Fig. 2 (c), the gradient information of edges in deep encoder is significantly enhanced after consistency training [8], which



(a) Encoder gradient statistics



(b) Encoder-4 (baseline [21])



(c) Encoder-4 (Chen et al. [8])

Fig. 2: Statistics and observations of the gradient of U-net [21] encoder layers on Leukocyte test dataset. (a) is the average gradient statistics of the encoder of Chen et al. [8] and baseline (U-net) [21]. (b) and (c) are the gradient maps of baseline(U-net) [21] and Chen et al. [8] on the fourth layers of encoder, respectively.

confirmed that the reason for the effectiveness of consistency training methods is that makes the encoder more discriminative. Therefore, integrating the progressive character of the gradient information of different encoder layers, and the goal of improving discriminative ability, the design of the gradient feature map regularization as follows:

$$L_u^g = \frac{1}{|S_e|} \sum_{e_j \in S_e} \frac{1}{|e_j|} \left\| \sum_{i=0}^{j} \nabla e_i - \nabla e_j \right\|^2 \tag{4}$$

Where $\nabla(\cdot)$ is a gradient operator and S_e denotes the encoder of the segmentation network. $\sum_{i=0}^{j} \nabla e_j$ is set to detach without backpropagation during training phase.

3.3. Adaptive Sharpening

The sharpening strategy proposed in Mixmatch [16] was applied to reduce the entropy of label distribution, which uses the common approach of adjusting the "temperature" of the categorical distribution [22]. The sharpening function is defined as follows:

$$Sharpen(p, T)_i = p_i^{\frac{1}{T}} / \sum_{j=1}^{C} p_j^{\frac{1}{T}} \tag{5}$$

where T is a hyperparameter. As $T \to 0$, the output of $Sharpen(p, T)$ will approach a Dirac ("one-hot") distribution. Since the results of $Sharpen$ as the targets for the unlabeled data, lowering temperature encourages model to produce lower-entropy predictions. However, the setting of T needs to be carefully designed and it is unreasonable to assign the same T to all samples.

Therefore, we propose adaptive sharpening to use the variance predicted by model as the aleatoric uncertainty to filter noise samples, and adaptively adjust T for each sample according to the confidence of the prediction, in which the lower confidence make the

greater sharpening of the sample:

$$\hat{p}'_i = p'_i + \sigma_i^2 \tag{6}$$

$$\hat{T}_i = \frac{\hat{p}'_i - min(\hat{p}')}{max(\hat{p}') - min(\hat{p}')} \tag{7}$$

where $p'_i = max(softmax(\mu_i))$. Eq. 7 and Eq. 5 can adaptively yield a pseudo-label for each sample, and then use MSE loss to optimize the unlabeled data:

$$L_u^s = \frac{1}{N}\sum_{i=0}^{N}(p_i - Sharpen(p_i, \hat{T}_i))^2 \tag{8}$$

The proposed adaptive sharpening makes model pay more attention to non-noise and hard unlabeled data, and less attention to noise and easy unlabeled data.

3.4. Class Consistency

Due to the strong sharpening for hard samples, it is likely to introduce additional noise to the network, additional noise smoothing strategy is required. Different from previous work [6], we expect classification and segmentation to be consistent on low-confidence predictions instead of high-confidence predictions. High-confidence classification results are likely to mislead the segmentation results due to the imbalance of the samples and the limited number of samples. But the network can always easily predict the correct categories with very low confidence. The loss function can be expressed as:

$$L_u^c = \frac{1}{|(p^c < \beta)|}\sum_{p_i^c \in (p^c < \beta)} p_i^c \times p_i^s \tag{9}$$

where $p^c = softmax(\mu_c)$, $p^s = softmax(\mu_s)$, and β is a threshold that determines the boundary of low-confidence consistency. In this paper, β is set to $1/C$.

4. EXPERIMENTS

4.1. Datasets and Evaluation Metrics

PASCAL VOC [11], PASCAL-Context [12], and a Leukocyte dataset are used in this work to evaluate the performance of the proposed method. The PASCAL VOC dataset consisting of 21 classes (with background included). Following the common practice [5, 6], we augment the training dataset with additional images from [24]. The augmented dataset consists of a total of 10582 training images and 1449 validation images. The PASCAL-Context dataset is a whole scene parsing dataset containing 4998 training and 5105 testing images with dense semantic labels. Following the previous work [6], we used semantic labels for 60 most frequent classes including the background class. The Leukocyte dataset which contains 3 classes was collected from a regular hospital. There are 500 images with the size 256×256 for training and 500 images within the same size for testing.

The mIoU, i.e. mean of class-wise intersection over union, is adopted as the metric to evaluate PASCAL VOC and PASCA-Context. And f1 score, recall, precision, and accuracy as evaluation metrics for Leukocyte dataset.

Table 1: Ablation study of the contribution for each loss term on the PASCAL VOC dataset with 1/8 labeled data.

Loss Terms	Labeled Data 1/8
CE only	62.0
CE + Sharpen	63.3
CE + adaptively sharpen (AS)	63.9
CE + class consistency (CC)	63.2
CE + AS + CC	65.5
CE + FGMR + AS + CC	**66.0**

Table 2: Comparison with the-state-of-the-art on PASCAL VOC dataset and PASCAL-Context dataset.

PASCAL VOC	Labeled Data			
Methods	1/20	1/8	1/3	Time
Baseline (DeepLabv2) [23]	56.8	62.0	66.8	-
Hung et al. [5]	59.1	64.3	67.4	0.63h
Mittal et al. [6] (s4GAN)	60.9	65.4	68.1	0.69h
Ours	**61.2**	**66.0**	**68.4**	**0.52h**

PASCAL-Context	Labeled Data		
Methods	1/8	1/3	Time
Baseline (DeepLabv2) [23]	32.1	35.4	-
Hung et al. [5]	32.8	34.8	0.25h
Mittal et al. [6] (s4GAN)	32.4	37.1	0.27h
Ours	**33.8**	**37.3**	**0.20h**

4.2. Implementation Details

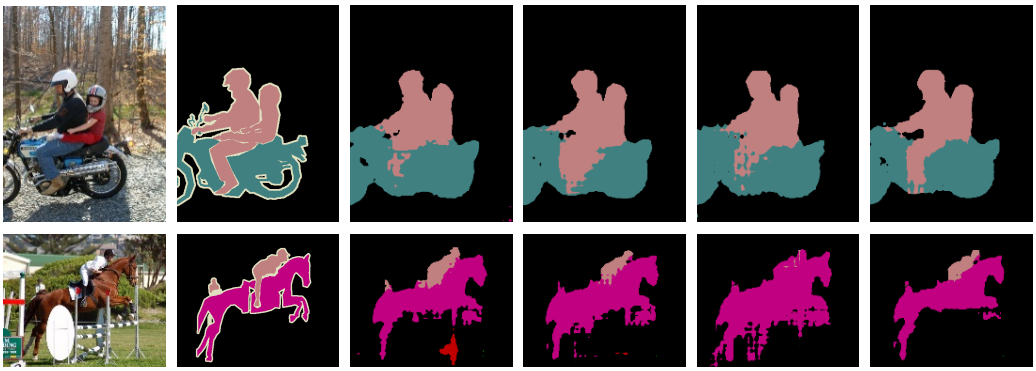
Motivated by previous work [5, 6], we use DeepLabv2 [23] as the baseline network for PASCAL VOC and PASCAL-Context datasets. The U-net [21] , which is widely used in medical image segmentation, is used as the baseline network for Leukocyte dataset.

Similar to [6], this work uses the poly-learning policy for the segmentation network, where the base learning rate is multiplied by a factor of $(1-\frac{iter}{max\_iter})^{pow}$ in every iteration. In our setup, $pow = 0.9$. Following the learning scheme in [6], the segmentation network is optimized using the SGD optimizer with a base learning rate of 1e-2 for Leukocyte dataset and 2.5e-4 for others, momentum 0.9, weight decay 5e-4, batch size 8 and 40K iterations. Hyper-parameter λ is set to 0.1. Our implementation is based on the open source framework Pytorch. All the experiments were run on two Nvidia Tesla P100 GPUs.

4.3. Results

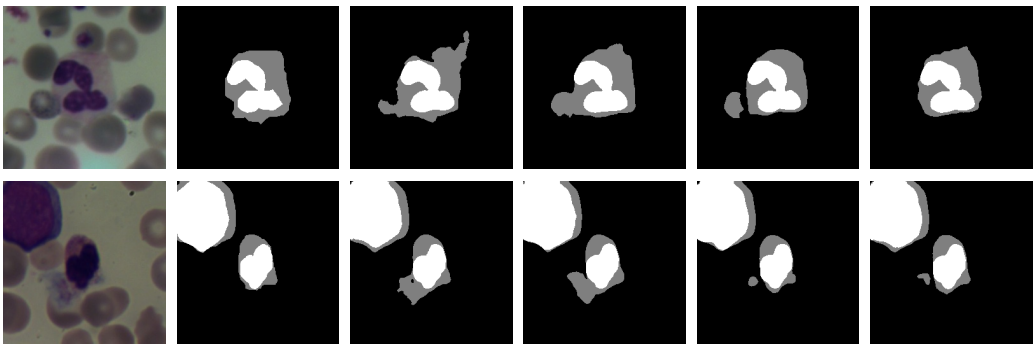
Ablation Studies. The method consists of three loss terms. We thus explored the effectiveness of each loss term and their combinations. We measure the effect of common sharpening and adaptive sharpening individually. After that, we experiment with class consistency constraints and combine with adaptive sharpening. It can be seen from the Table 1 that all of loss terms effectively improve the performance. Finally, adding feature gradient map regularization to explore the effectiveness of the complete method. As shown in Table 1, each loss term can improve the performance of network. We use the complete method as the method for all experiments.

PASCAL VOC and Context Datasets. Table 2 shows the evaluation results on the PASCAL VOC and the PASCAL-Context datasets without pretraining on the Microsoft COCO [1] dataset. The proposed method can improve baseline 2.4% to 7.7% under



(a) Original (b) Ground truth (c) Baseline [23] (d) Hung et al. [5] (e) Mittal et al. [6] (f) Ours

Fig. 3: Qualitative results on the PASCAL VOC dataset using 1/8 labeled samples.



(a) Original (b) Ground truth (c) Baseline [21] (d) Chen et al. [8] (e) Yu et al. [9] (f) Ours

Fig. 4: Qualitative results on the Leukocyte dataset using 1/10 labeled samples.

different data splits by utilizing unlabeled samples. The methods of [5] and [6] are the previous start-of-the-art method and the current start-of-the-art method in semi-supervised semantic segmentation, respectively, in which the method of Hung et al. [5] is based on GAN and Mittal et al. [6] introduced consistency training in [5]. Under the same experimental settings, the proposed method obtains the best results in the PASCAL VOC dataset with 1/3, 1/8, and 1/20 labeled data and the PASCAL-Context dataset with 1/3 and 1/8 labeled data. The last column of Table 2 is the comparisons of the average training time per epoch, the proposed method takes the shortest time compared to other comparison semi-supervised semantic segmentation methods. Fig. 3 shows the qualitative results on the PASCAL VOC dataset using 1/8 labeled samples.

Leukocyte Dataset. In order to further prove that the proposed method has good universality. We test on 1/10 labeled data on the Leukocyte dataset without data augmentation. Table 3 shows the proposed method yields an improvement over the baseline of 2.23%, 1.67%, 2.46%, and 0.95% for the f1 score, recall, precision, and accuracy, respectively. Meanwhile, the results of comparing with current state-of-the-art semi-supervised medical semantic segmentation methods [8, 9] describe the proposed method reach the new start-of-the-art with the least cost, in which [8] and [9] are the semi-supervised medical segmentation methods based on consistency training proposed by Chen et al. [8] and Yu et al. [9], respectively. Fig. 4 shows the qualitative results on the Leukocyte dataset using 1/10 labeled samples, for images of white blood cells whose cytoplasm is close to the background, the proposed method can effectively segment the cytoplasm.

Table 3: Semi-supervised comparisons on the Leukocyte dataset using 1/10 labeled samples.

	F1	Recall	Precision	Acc	Time
Baseline (Unet) [21]	90.56	91.98	90.49	97.21	-
Chen et al. [8]	92.21	93.11	92.19	97.88	0.021h
Yu et al. [9]	92.75	93.34	**92.95**	98.01	0.020h
Ours	**92.79**	**93.65**	**92.95**	**98.16**	**0.018h**

Table 4: Space complexity comparisons on the PASCAL VOC dataset.

	Extra Parameters
Baseline (DeepLabv2) [23]	-
Hung et al. [5]	2.78M
Mittal et al. [6] (s4GAN)	2.78M
Ours	**1.16M**

Space Complexity. As shown in the space complexity comparisons on the PASCAL VOC dataset in Table 4, compared with the baseline parameters, the proposed method only adds 1.16M extra parameters, while Huang et al. [5] method and Mittal et al. [6] method add 2.78M extra parameters. The proposed method requires the least

amount of extra parameters in the comparison methods, in which the extra parameters of the proposed method are less than about one-half parameters of the comparison methods.

5. CONCLUSION

In this paper, we presented an entropy minimization method for semi-supervised semantic segmentation, which was simple, efficient, and easily expandable. The proposed method designs three additional unsupervised loss function terms to optimize the network collaboratively and make the decision boundary lie in a low-density region. Extensive experiments on PASCAL VOC, PASCAL-Context, and Leukocyte datasets show that the proposed method is effective, without almost additional calculations and network structures. In the future, we will explore the fusion of the proposed method and the existing semi-supervised methods.

6. REFERENCES

[1] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick, "Microsoft coco: Common objects in context," in *European Conference on Computer Vision*. Springer, 2014, pp. 740–755.

[2] Yunchao Wei, Huaxin Xiao, Honghui Shi, Zequn Jie, Jiashi Feng, and Thomas S Huang, "Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7268–7277.

[3] Jiwoon Ahn, Sunghyun Cho, and Suha Kwak, "Weakly supervised learning of instance segmentation with inter-pixel relations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2209–2218.

[4] Nasim Souly, Concetto Spampinato, and Mubarak Shah, "Semi supervised semantic segmentation using generative adversarial network," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 5688–5696.

[5] Wei Chih Hung, Yi Hsuan Tsai, Yan Ting Liou, Yen Yu Lin, and Ming Hsuan Yang, "Adversarial learning for semi-supervised semantic segmentation," in *29th British Machine Vision Conference, BMVC 2018*, 2019.

[6] Sudhanshu Mittal, Maxim Tatarchenko, and Thomas Brox, "Semi-supervised semantic segmentation with high-and low-level consistency," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[7] Gerda Bortsova, Florian Dubost, Laurens Hogeweg, Ioannis Katramados, and Marleen de Bruijne, "Semi-supervised medical image segmentation via learning consistency under transformations," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 810–818.

[8] Shuai Chen, Gerda Bortsova, Antonio García-Uceda Juárez, Gijs van Tulder, and Marleen de Bruijne, "Multi-task attention-based semi-supervised learning for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 457–465.

[9] Lequan Yu, Shujun Wang, Xiaomeng Li, Chi-Wing Fu, and Pheng-Ann Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 605–613.

[10] Alex Kendall and Yarin Gal, "What uncertainties do we need in bayesian deep learning for computer vision?," in *Advances in Neural Information Processing Systems*, 2017, pp. 5574–5584.

[11] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[12] Roozbeh Mottaghi, Xianjie Chen, Xiaobai Liu, Nam-Gyu Cho, Seong-Whan Lee, Sanja Fidler, Raquel Urtasun, and Alan Yuille, "The role of context for object detection and semantic segmentation in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 891–898.

[13] Antti Tarvainen and Harri Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Advances in Neural Information Processing Systems*, 2017, pp. 1195–1204.

[14] Thomas N Kipf and Max Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[15] Yves Grandvalet and Yoshua Bengio, "Semi-supervised learning by entropy minimization," in *Advances in Neural Information Processing Systems*, 2005, pp. 529–536.

[16] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel, "Mixmatch: A holistic approach to semi-supervised learning," in *Advances in Neural Information Processing Systems*, 2019, pp. 5049–5059.

[17] Dong-Hyun Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on challenges in representation learning, ICML*, 2013, vol. 3.

[18] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez, "Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2517–2526.

[19] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen, "Improved techniques for training gans," *Advances in Neural Information Processing Systems*, vol. 29, pp. 2234–2242, 2016.

[20] Diederik P Kingma and Max Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[21] 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*. Springer, 2015, pp. 234–241.

[22] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio, *Deep learning*, vol. 1, MIT press Cambridge, 2016.

[23] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[24] Bharath Hariharan, Pablo Arbeláez, Lubomir Bourdev, Subhransu Maji, and Jitendra Malik, "Semantic contours from inverse detectors," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 991–998.