Semi-supervised Skin Lesion Segmentation via Transformation Consistent Self-ensembling Model

Xiaomeng Li[1]
xmli@cse.cuhk.edu.hk

Lequan Yu[1]
lqyu@cse.cuhk.edu.hk

Hao Chen[12]
hchen@cse.cuhk.edu.hk

Chi-Wing Fu[1]
cwfu@cse.cuhk.edu.hk

Pheng-Ann Heng[1]
pheng@cse.cuhk.edu.hk

[1] Department of Computer Science and Engineering
The Chinese University of Hong Kong
Hong Kong

[2] Imsight Medical Technology, Inc.

Abstract

Automatic skin lesion segmentation on dermoscopic images is an essential component in computer-aided diagnosis of melanoma. Recently, many fully supervised deep learning based methods have been proposed for automatic skin lesion segmentation. However, these approaches require massive pixel-wise annotation from experienced dermatologists, which is very costly and time-consuming. In this paper, we present a novel semi-supervised method for skin lesion segmentation, where the network is optimized by the weighted combination of a common supervised loss for labeled inputs only and a regularization loss for both labeled and unlabeled data. To utilize the unlabeled data, our method encourages the consistent predictions of the network-in-training for the same input under different regularizations. Aiming for the semi-supervised segmentation problem, we enhance the effect of regularization for pixel-level predictions by introducing a transformation, including rotation and flipping, consistent scheme in our self-ensembling model. With only 300 labeled training samples, our method sets a new record on the benchmark of the International Skin Imaging Collaboration (ISIC) 2017 skin lesion segmentation challenge. Such a result clearly surpasses fully-supervised state-of-the-arts that are trained with 2000 labeled data.

1 Introduction

Skin cancer is currently one of the fastest growing cancers worldwide, and melanoma is the most deadly form of skin cancer, leading to an estimated 9,730 deaths in the United States in 2017 [24]. To improve the diagnostic performance of melanoma, dermoscopy has been proposed as a noninvasive imaging technique to enhance the visual effect of pigmented skin
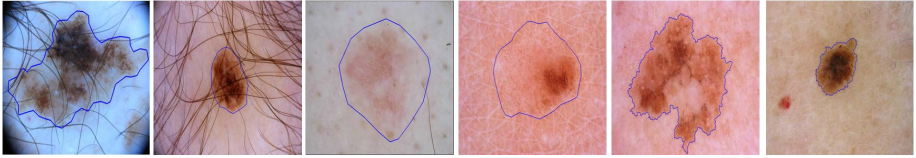
Figure 1: Skin lesion cases with artifacts (the left two); examples of ambiguous (the middle two) and clear-cut (the right two) labels.

lesions. However, recognizing malignant melanoma by visual interpretation alone is time-consuming and error-prone to inter- and intra-observer variabilities. To assist dermatologists in the diagnosis, an automatic melanoma segmentation method is highly demanded in the clinical practice.

Automatic melanoma segmentation is a very challenging task due to large variations in lesion size, location, shape and color over different patients and the presence of artifacts such as hairs and veins; see Figure 1. Traditional segmentation methods are mainly based on clustering, intensity thresholding, region growing, and deformable models. These methods, however, rely on hand-crafted features, and have limited feature representation capability. Recently, convolutional neural networks (CNNs) have been widely used and achieved remarkable success in a variety of vision recognition tasks. Many researchers advanced the skin lesion segmentation and showed decent results [2, 6, 17, 28]. For example, Yuan *et al.* [28] proposed a deep convolutional neural network (DCNN), trained it with multiple color spaces, and achieved the best performance in the ISIC 2017 skin lesion segmentation challenge.

All the above methods, however, are based on fully supervised learning, which requires a large amount of annotated images to train the network for accuracy and robustness. Such pixel-level annotation is laborious and difficult to obtain, especially for melanoma in the dermoscopic images, which rely heavily on experienced dermatologists. Moreover, the limited amount of labeled data with pixel-wise annotations also restricts the performance of deep networks. Lastly, there exists some cases that display ambiguous melanocytic or borderline features of melanoma. These cases are inherently difficult to have an accurate annotation from the dermoscopic diagnosis [27]; see again Figure 1. Previous supervised learning based methods do not have specific schemes to deal with these ambiguous annotations, which may degrade the performance on those dermoscopic images with clear-cut lesions. To alleviate the above issues, we address the skin lesion segmentation problem via semi-supervised learning, which leverages both a limited amount of labeled and an arbitrary amount of unlabeled data. As a by-product, our semi-supervised method is robust and has a potential to be tolerant to ambiguous labels; see experiments in Section 4.2. There are some semi-supervised approaches for dermoscopy images and other medical image processing [1, 11, 14, 18]. However, they either suffer from limited representation capacity of hand-crafted features or may easily get into local minimum.

In this paper, we present a novel semi-supervised learning method for skin lesion segmentation. The whole framework is trained with a weighted combination of the supervised loss and the unsupervised loss. To utilize the unlabeled data, our self-ensembling method encourages the consistent prediction of the network for the same input data under different regularizations (*e.g.*, randomized Gaussian noise, network dropout and randomized data transformation). In particular, we design our method to account for the challenging semi-supervised segmentation task, in which pixel-level classification is required to be predicted.

We observe that in the segmentation problem, if one transforms (*e.g.*, rotate) the input image, the expected prediction should be transformed in the same manner. Actually, when the inputs of CNNs are rotated, the corresponding network predictions would not rotated in the same way [25]. In this regard, we take advantages of this property by introducing a transformation (*i.e.*, rotation, flipping) consistent scheme at the input and output space of our network. Specifically, we design the unsupervised/regularization loss by minimizing the differences between the network predictions under different transformations of the same input.

In summary, our work has the following achievements:

- We present a novel semi-supervised learning method for the practical biomedical image segmentation problem by taking advantage of a large amount of unlabeled data, which largely reduces annotation efforts for the dermatologists.

- To better utilize the unlabeled data for segmentation tasks, we propose a transformation consistent scheme in self-ensembling model and demonstrate the effectiveness for semi-supervised learning.

- We establish a new record with only 300 labeled data on the benchmark of ISIC 2017 skin lesion segmentation challenge, which excels the state-of-the-arts that are based on fully supervised learning with 2000 labeled data.

2 Related Work

Skin lesion segmentation. Early approaches on skin lesion segmentation mainly focused on thresholding [9], iterative/statistical region merging [13] and machine learning related methods [11, 20]. Recently, many researchers employed deep learning based methods for skin lesion segmentation. For example, Yu *et al.* [27] explored the network depth property and developed a deep residual network with more than 50 layers for automatic skin lesion segmentation, where several residual blocks were stacked together to increase the network representative capability. Bi *et al.* [4] proposed a multi-stage approach to segment skin lesion by combining the results from multiple cascaded fully convolutional networks. Yuan *et al.* [29] proposed a 19-layer deep convolutional neural network and trained it in an end-to-end manner for skin lesion segmentation. However, these approaches are based on fully supervised learning, requiring massive pixel-wise annotations from experienced dermatologists to create a training dataset.

Transformation equivariant representation. There is a body of related literature on equivariance representations, where the transformation equivariance is encoded to the network to explore the network equivariance property [7, 8, 25]. For example, Cohen and Welling [7] proposed group equivariant neural network to improve the network generalization, where equivariance to 90°-rotations and dihedral flips is encoded by copying the transformed filters at different rotation-flip combinations. Concurrently, Dieleman et al. [8] designed four different equivariance to preserve feature map transformations by rotating feature maps instead of filters. Recently, Worrall et al. [25] restricted the filters to circular harmonics to achieve continuous 360°-rotations equivariance. However, these works aim to encode equivariance to the network to improve the generalization capability of the network, while our method targets to better utilize the unlabeled data in the semi-supervised learning.

Semi-supervised segmentation for medical images. Semi-supervised approaches have been applied in various medical imaging tasks. Portela *et al.* [19] employed Gaussian Mixture Model (GMM) to automatically segment brain MR images. For retinal vessel segmenta-
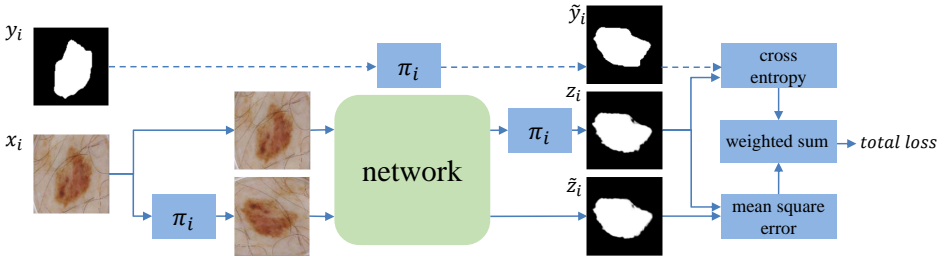
Figure 2: The pipeline of our proposed transformation consistent self-ensembling model for semi-supervised skin lesion segmentation. The total loss is weighted combination of the cross-entropy loss on labeled data, and mean square error loss on both labeled and unlabeled data. The model encourages the network to be transformation consistent to utilize the unlabeled data. Note that π_i remains the same in each training pass but are changed at different passes.

tion, You et al. [26] combined radial projection and semi-supervised learning. Gu et al. [10] proposed a semi-supervised method to segment vessel by constructing forest oriented super pixels. While Sedai et al. [23] introduced a variational autoencoder for optic cup segmentation in retinal fundus images. There are also some semi-supervised works for diagnosing in dermoscopy images [14, 18]. For example, Jaisakthi et al. [14] proposed a semi-supervised skin lesion segmentation method using K-means clustering on color features. However, these methods are based on hand-crafted features, which suffer from limited representation capacity. Recently, as the surprising results achieved by CNNs in the supervised learning, semi-supervised approaches with CNNs started to attract attentions in the medical imaging field. For example, Bai et al. [1] proposed a semi-supervised fully convolutional neural network (FCN) to segment cardiac from MR images, where the network parameters and the segmentations for unlabeled data were alternatively updated. However, this method was trained offline and may easily get into local minimum.

3 Method

3.1 Overview

We first describe the overview of our semi-supervised segmentation method. The training set consists N inputs in total, including M labeled inputs and $N-M$ unlabeled inputs. Let $\mathcal{L} = \{(x_i, y_i)\}_{i=1}^{M}$ be the labeled set and $\mathcal{U} = \{x_i\}_{i=M+1}^{N}$ be the unlabeled set, where $x_i \in \mathbb{R}^{H \times W \times 3}$ is the input image and $y_i \in \{0,1\}^{H \times W}$ is the ground-truth label. Our proposed semi-supervised segmentation method can be formulated to learn the network parameters θ by optimizing:

$$\min_{\theta} \sum_{i=1}^{M} l(f(x_i; \theta), y_i) + \lambda R(\theta, \mathcal{L}, \mathcal{U}), \tag{1}$$

where $f(\cdot)$ represents the network mapping, l is the supervised loss function and R is the regularization (unsupervised) loss. The first component is designed for supervised training, optimized by the cross-entropy loss and used for evaluating the correctness of network output on labeled inputs only. While the second regularization component is designed for unsupervised training by regularizing the network output on both labeled and unlabeled inputs. λ is
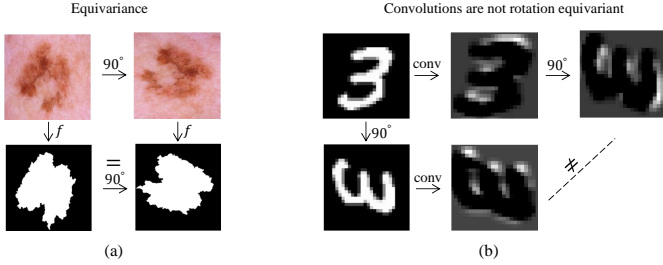
Figure 3: (a) Segmentation is desired to be rotation equivariant. If we rotate the input image, the excepted segmentation mask should have same rotation as the original segmentation mask. (b) Convolutions are not rotation equivariant in general.

a weighting factor that controls how strong the regularization is.

Self-ensembling methods [15, 21] demonstrate great promise in semi-supervised learning. The essential to this success relies on the key *smoothness* assumption; that is, data points close to each other are likely to have the same label. In our work, we inherit the spirit of these methods and design the regularization term as a consistency loss to encourage smooth predictions for the same data under different regularization or perturbations (*e.g.*, Gaussian noise, network dropout, and randomized data transformation). Then the regularization loss R can be described as:

$$R(\theta, \mathcal{L}, \mathcal{U}) = \sum_{i=1}^{N} \mathbb{E}_{\xi', \xi} \left\| f(x_i; \theta, \xi') - f(x_i; \theta, \xi) \right\|^2, \qquad (2)$$

where ξ and ξ' denote different regularization or perturbations of input data. In the following subsection, we will introduce how to effectively design the randomized data transformation regularization for the segmentation problem.

3.2 Transformation Consistent Self-ensembling Model

Most regularization and pertrbations are easily designed for classification problem, while we are confronted with a more challenging and practical skin lesion segmentation problem. One important difference is that the classification problem is transformation *invariant* while the segmentation task is desired to be transformation *equivariant*. Specifically, in the classification problem, we are only interested in the presence or absence of an object in the whole image, the classification result should remain the same, no matter what the data transformation (*i.e.*, translation, rotation, and flipping) are applied to the input image. While in the segmentation task, if we rotate the input image, the expected segmentation mask should have the same rotation with original mask, although the corresponding pixel-wise predictions are same; see Figure 3(a). However, in general, convolutions are not transformation (*i.e.*, flipping, rotation) equivariant[1], meaning that if one rotates or flips the CNN input, then the feature maps do not necessarily rotate in a meaningful or easy to predict manner [25], as shown in Figure 3(b). Therefore, the convolutional network consisting of a series of convolutions is also not transformation equivariant. Formally, every transformation $\pi \in \Pi$ of input

---

[1]Convolution is translation equivariant, and we focus on flipping and rotation transformation in this work.

$\mathbf{x}$ associates with a transformation $\psi \in \Psi$ of the outputs; that is $\psi[f(\mathbf{x})] = f(\pi[\mathbf{x}])$, but in general $\pi \neq \psi$.

This property limits the unsupervised regularization effect of randomized data transformation for the segmentation problem [15]. To enhance the regularization and more effectively utilize unlabeled data in our segmentation task, we introduce a transformation consistent scheme in the unsupervised regularization term. Specifically, this transformation consistent scheme is embedded into the framework by approximating ψ to π at the input and output space. The pipeline of our proposed transformation consistent self-ensembling

---

Algorithm 1 Model pseudocode.

Input: $x_i \in \mathcal{L} + \mathcal{U}, y_i \in \mathcal{L}$
    $\lambda(t)$ = unsupervised weight function
    $f_\theta(x)$ = neural network with trainable parameters θ
    $\pi_i(x)$ = transformation operation
    for t in $[1, numepochs]$ **do**
        for each minibatch B **do**
            randomly update $\pi_i(x)$
            $z_{i \in B} \leftarrow \pi_i(f_\theta(x_{i \in B}))$
            $\tilde{z}_{i \in B} \leftarrow f_\theta(\pi_i(x_{i \in B}))$
            $loss \leftarrow -\frac{1}{|B \cap \mathcal{L}|} \sum_{i \in (B \cap \mathcal{L})} \log z_i[\pi_i(y_i)] +$
            $\lambda(t) \frac{1}{|B|} \sum_{i \in B} \|z_i - \tilde{z}_i\|^2$
            update θ using optimizer
        end for
    end for
    return θ;

---

model is shown in Figure 2, and the pseudocode is in Algorithm 1. Each input x_i is fed into the network for twice evaluation under transformation consistent scheme and other different perturbations (*e.g.*, Gaussian noise and network dropout) to acquire two outputs z_i and $\tilde{z}_i$. Specifically, the transformation consistent scheme consists of triple π_i operations; see Figure 2. For one same training sample x_i, in the first evaluation, the operation π_i is applied to the input image while in the second evaluation, the operation π_i is applied on the prediction map. Through minimizing the difference between z_i and $\tilde{z}_i$ with a mean square error loss function, we can regularize the network to be transformation consistent and further increase the network generalization capacity. Note that this regularization loss is applied for both labeled and unlabeled inputs. For those labeled inputs $x_i \in \mathcal{L}$, we also apply the same operation π_i to y_i and use the standard cross-entropy loss to evaluate the correctness of network output. Finally, the network is trained by minimizing the weighted combination of unsupervised regularization loss and supervised cross-entropy loss. Note that we employed the same data augmentation in the training procedure of all the experiments for fair comparison. However, our method is different from traditional data augmentation. Specifically, our method utilized the unlabeled data by minimizing network output difference under the transformed inputs, while obeying the *smoothness* assumption.

3.3 Training and Inference Procedures

In the above transformation consistent scheme, we apply four kinds of rotation operations to the input with angles of $\gamma \cdot 90°$, where $\gamma \in \{0, 1, 2, 3\}$. We also apply a horizontal flipping
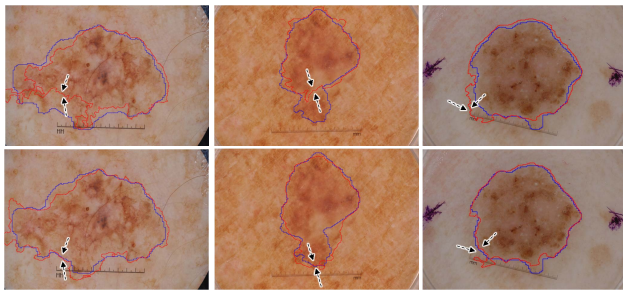
Figure 4: Segmentation results of supervised-only learning (top) and our method (bottom) on the validation dataset. The blue and red contours denote the ground truth and our result, respectively.

operation. In total, eight possible transformation operations are obtained, and we randomly choose one operation in each training pass. We avoid the other angles for implementation simplification, but the proposed framework can be generalized to other angles in the future work. We employ the 2D DenseUNet-167 architecture in [16] as our network backbone. The dropout layer is applied after each convolutional layer in the encoding and decoding parts excepting for the last convolutional layer. We use the standard data augmentation techniques on-the-fly to avoid overfitting. The data augmentation includes randomly flipping, rotating as well as scaling with a random scale factor from 0.9 to 1.1. Note that all the experiments employed data augmentation for fair comparison. The model was implemented using *Keras* package [5], and was trained with stochastic gradient descent (SGD) algorithm (momentum is 0.9 and minibatch size is 10). The initial learning rate was 0.01 and decayed according to the equation $lr = lr * (1 - iterations/total\_iterations)^{0.9}$. In the inference phase, we remove the transformation operations in the network and do one single test with original input for fair comparison. After getting the probability map from the network, we first apply thresholding with 0.5 to get the binary segmentation result, and then use morphology operation, *i.e.*, filling holes to get the final skin lesion segmentation result.

4 Experiments and Results

4.1 Dataset and Evaluation Metrics

We evaluate our method on the dataset of 2017 ISIC skin lesion segmentation challenge [6], which includes a training set with 2000 annotated dermoscopic images, a validation set with 150 images, and a testing set with 600 images. Five evaluation metrics are calculated in the challenge to evaluate the segmentation performance, including Jaccard index (JA), dice coefficient (DI), pixel-wise accuracy (AC), sensitivity (SE) and specificity (SP). Note that the final rank is determined according to JA in the 2017 ISIC skin lesion segmentation challenge.

4.2 Analysis of Our Method

Quantitative and visual results with 50 labeled data. In this part, we report the performance of our method trained with only 50 randomly selected labeled images and 1950 unlabeled images. Table 1 shows the experiments run with supervised-only method (the first

one), supervised with regularization (the second one) and our semi-supervised method (the third one) on the validation dataset. The supervised-only experiment is trained with the same network backbone, but only optimized by the standard cross-entropy loss on the 50 labeled images. It is obvious that compared with the supervised-only method, our semi-supervised method can achieve higher performance on all the evaluation metrics, with 2.46%, 2.64%, and 3.60% improvements on JA, DI and SE, respectively. These prominent improvements on JA, DI and SE indicate that in most cases the false negative regions shrink while true positive regions expand to fit the true boundary of the lesion. Comparing with the segmentation ground truth (blue contour), we can see the semi-supervised method can expand the segmented region to fit the ground truth lesion; see the left two examples in Figure 4. Notably, our method would not simply amplify the segmentation result in all cases, it would also reasonably shrink the segmentations; see the examples in Figure 4. These observations manifest that our semi-supervised learning method can improve the network generalization capability, compared with the supervised-only method. It is worth mentioning that our method can also improve the supervised training; see Supervised with regularization in Table 1.

Model	50 labeled, 1950 unlabeled data		
	JA	DI	SE
Supervised-only	72.85	81.15	82.77
Supervised with regularization	73.25	81.60	83.30
Our Method	**75.31**	**83.79**	**86.37**
Our Method-A	74.59	83.27	82.77
Our Method-B	74.21	82.68	83.15

Table 1: Comparison of supervised learning and semi-supervised learning on the validation dataset.
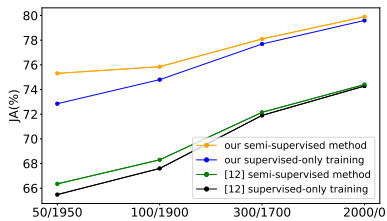

Figure 5: Results with different number of labeled/unlabeled data.

Effectiveness of transformation consistent scheme. To show the effectiveness of the transformation consistent regularization scheme, we perform ablation analysis on our method. We compare our method with the most common perturbations regularization, i.e., Gaussian noise and network dropout. Table 1 shows the experiment results, where "Our Method-A" refers to semi-supervised learning with Gaussian noise and dropout regularization, "Our Method-B" denotes to semi-supervised learning with transformation consistent regularization, and "Our Method" refers to the experiment with all of these regularizations. Note that these experiments are performed on the same training data with 50 labels. As shown in Table 1, both kinds of regularizations can independently contribute to the performance gains in semi-supervised learning. The performance improvement with transformation consistent regularization is very competitive, compared with the performance increment with Gaussian noise and dropout regularizations. We also see that these two regularizations are complementary. When the two kinds of regularizations are employed, the performance can be further boosted.

Results under different number of labeled data. Figure 5 demonstrates the network performance under different number of labeled images. We draw the JA score of our semi-supervised method (trained with labeled data and unlabeled data) and supervised-only training (trained only with labeled data). We can observe that the semi-supervised method consistently performs better than supervised-only in different labeled/unlabeled data settings, which demonstrates that our method effectively utilizes unlabeled data and is beneficial to the performance gains. Note that in all semi-supervised learning experiments, we train the

Table 2: Comparison on the test dataset in the ISIC 2017 skin lesion segmentation challenge.

Team	JA	DI	AC	SE	SP
Our Semi-supervised Method	**0.798**	**0.874**	**0.943**	**0.879**	0.953
Our Baseline	0.772	0.853	0.936	0.837	0.969
Yuan and Lo [23]	0.765	0.849	0.934	0.825	0.975
Berseth [0]	0.762	0.847	0.932	0.820	0.978
Bi et al. [0]	0.760	0.844	0.934	0.802	**0.985**
RECOD	0.754	0.839	0.931	0.817	0.970
Jer	0.752	0.837	0.930	0.813	0.976

Table 3: JA performance of different semi-supervised methods on the validation dataset.

Method	50 labeled	50 labeled and 1950 unlabeled	Improvement
Our	0.7285	0.7531	0.0246
Bai et al. [0]	0.7285	0.7440	0.0155
Huang et al. [1]	0.6548	0.6635	0.0087

network with 2000 images in total, including labeled images and unlabeled images. As expected, the performance of supervised-only training increases when more labeled training images are available; see the blue line in Figure 5. At the same time, the segmentation performance of semi-supervised learning can also be increased with more labeled training images; see the orange line in Figure 5. However, as we add more labeled samples, the difference in segmentation accuracy between semi-supervised and supervised-only becomes smaller. The observation conforms with our expectation that our method leverages the distribution information from the unlabeled dataset. As the labeled dataset is small, our method can gain a large improvement, since the regularization loss can effectively leverage the data distribution information from the unlabeled data. Comparatively, as the labeled data increases, the improvements become limited. This is because both labeled and unlabeled images are randomly selected from the same dataset. When we have more labeled images, our regularization term can benefit from less additional distribution information from the unlabeled data. In the clinical practice, our approach is highly promising when a large number of unlabeled data from different protocols are acquired every day.

Robustness analysis. As we mentioned above, our semi-supervised method can improve the robustness of the network due to the regularization effect of the unsupervised loss. From the comparison between the semi-supervised method and supervised method trained with 2000 labeled images in Figure 5, we can see that our method can increase the JA performance when all labels are used (from 79.60% to 80.02%). Note that the unsupervised loss was employed on all the input data and both experiments used the same data augmentation. Therefore, the improvement indicates that the unsupervised loss can provide a strong regularization to the labeled data, which would be useful in the case that the ground truth is not accurate due to the ambiguous lesions; see Figure 1. In other words, the consistency requirement in the regularization term can encourage the network to learn more robust features and has a potential to be tolerant to ambiguous labels.

4.3 Comparison with Other Methods

We compare our method with state-of-the-art methods submitted to the ISIC 2017 skin lesion segmentation challenge. There are totally 21 submissions and the top results are listed in Table 2. We trained two models: semi-supervised learning with 300 labeled data and 1700

unlabeled data and supervised-only network with 300 labeled data. We refer the last experiment as our baseline model. As shown in Table 2, our semi-supervised method achieved the best performance on the benchmark, outperforming the state-of-the-art method [28] with 3.3% improvement on JA (from 76.5% to 79.8%). The performance gains on DI and SE are consistent with that on JA, with 2.5% and 5.4% improvement, respectively. Our baseline model with 300 labeled data also excels the other methods due to the state-of-the-art network architecture. Based on this architecture, our semi-supervised learning method further makes significant improvements, which demonstrate the effectiveness of the overall semi-supervised learning method.

We also compare our method with the latest semi-supervised segmentation method [1] in the medical imaging community and an adversarial learning based semi-supervised method [12]. We conduct experiments with the setting of 50 labeled images and 1950 unlabeled images. Table 3 shows the JA performance of different methods. As shown in Table 3, our proposed method achieves 2.46% JA improvement by utilizing unlabeled data. However, [1] and [12] can only enhance 1.55% and 0.87% improvement on JA, respectively. Due to the different network backbone (DenseUNet-167 in [1] and FC-ResNet101 in [12]), the performance with 50 labeled data is different. Figure 5 also shows the performance improvement of semi-supervised learning scheme of our method and [12] under the setting of 100, 300 and 2000 labeled data. We can see that the improvement of [12] is inferior than our method in all labeled/unlabeled settings, which also validates the effectiveness of our method.

5 Conclusion

In this paper, we present a novel semi-supervised learning method for skin lesion segmentation. Specifically, we introduce a novel transformation consistent self-ensembling model for the segmentation task, which enhances the regularization effects to utilize the unlabeled data. Comprehensive experimental analysis on the ISIC 2017 skin lesion segmentation challenge dataset demonstrates the effectiveness of our semi-supervised learning and the robustness of our method. Our method is general enough and can be extended to other semi-supervised learning problems in medical imaging field. In the future, we will explore more regularization forms and ensembling techniques for better leveraging of the unlabeled data.

6 Acknowledgments

References

[1] Wenjia Bai, Ozan Oktay, and Matthew et al. Sinclair. Semi-supervised learning for network-based cardiac mr image segmentation. In *MICCAI*, pages 253–260. Springer, 2017.

[2] Matt Berseth. Isic 2017-skin lesion analysis towards melanoma detection. *arXiv preprint arXiv:1703.00523*, 2017.

[3] Lei Bi, Jinman Kim, Euijoon Ahn, and Dagan Feng. Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks. *arXiv preprint arXiv:1703.04197*, 2017.

[4] Lei Bi, Jinman Kim, and Euijoon et al. Ahn. Dermoscopic image segmentation via multistage fully convolutional networks. *IEEE Transactions on Biomedical Engineering*, 64(9):2065–2074, 2017.

[5] François Chollet. Keras. https://github.com/fchollet/keras, 2015.

[6] Noel CF Codella, David Gutman, and M Emre et al. Celebi. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1710.05006*, 2017.

[7] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International Conference on Machine Learning*, pages 2990–2999, 2016.

[8] Sander Dieleman, Jeffrey De Fauw, and Koray Kavukcuoglu. Exploiting cyclic symmetry in convolutional neural networks. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 1889–1898, 2016.

[9] M Emre Celebi, Quan Wen, Sae Hwang, Hitoshi Iyatomi, and Gerald Schaefer. Lesion border detection in dermoscopy images using ensembles of thresholding methods. *Skin Research and Technology*, 19(1), 2013.

[10] Lin Gu, Yinqiang Zheng, Ryoma Bise, Imari Sato, Nobuaki Imanishi, and Sadakazu Aiso. Semi-supervised learning for biomedical image segmentation via forest oriented super pixels (voxels). In *MICCAI*, pages 702–710. Springer, 2017.

[11] Yingding He and Fengying Xie. Automatic skin lesion segmentation based on texture analysis and supervised learning. In *Asian Conference on Computer Vision*, pages 330–341. Springer, 2012.

[12] W.-C. Hung, Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, and M.-H. Yang. Adversarial learning for semi-supervised semantic segmentation. *arXiv preprint arXiv:1802.07934*, 2018.

[13] Hitoshi Iyatomi, Hiroshi Oka, and Masataka et al. Saito. Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system. *Melanoma research*, 16 (2):183–190, 2006.

[14] SM Jaisakthi, Aravindan Chandrabose, and P Mirunalini. Automatic skin lesion segmentation using semi-supervised learning technique. *arXiv preprint arXiv:1703.04301*, 2017.

[15] Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*, 2016.

[16] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng. H-denseunet: Hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE Transactions on Medical Imaging*, 2018.

[17] Xiaomeng Li, Lequan Yu, Chi-Wing Fu, and Pheng-Ann Heng. Deeply supervised rotation equivariant network for lesion segmentation in dermoscopy images. *arXiv preprint arXiv:1807.02804*, 2018.

[18] Ammara Masood, Adel Al-Jumaily, and Khairul Anam. Self-supervised learning model for skin cancer diagnosis. In *Neural Engineering (NER), 2015 7th International IEEE/EMBS Conference on*, pages 1012–1015. IEEE, 2015.

[19] Nara M Portela, George DC Cavalcanti, and Tsang Ing Ren. Semi-supervised clustering for mr brain image segmentation. *Expert Systems with Applications*, 41(4):1492–1497, 2014.

[20] Amir Reza Sadri, Maryam Zekri, Saeed Sadri, Niloofar Gheissari, Mojgan Mokhtari, and Farzaneh Kolahdouzan. Segmentation of dermoscopy images using wavelet networks. *IEEE Transactions on Biomedical Engineering*, 60(4):1134–1141, 2013.

[21] Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. In *NIPS*, pages 1163–1171, 2016.

[22] Richard A Scolyer, Rajmohan Murali, Stanley W McCarthy, and John F Thompson. Histologically ambiguous ("borderline") primary cutaneous melanocytic tumors: approaches to patient management including the roles of molecular testing and sentinel lymph node biopsy. *Archives of pathology & laboratory medicine*, 134(12):1770–1777, 2010.

[23] Suman Sedai, Dwarikanath Mahapatra, and Sajini et al. Hewavitharanage. Semi-supervised segmentation of optic cup in retinal fundus images using variational autoencoder. In *MICCAI*, pages 75–82. Springer, 2017.

[24] Rebecca L. Siegel, Kimberly D. Miller, and Ahmedin Jemal. Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians*, 67(1):7–30, 2017. ISSN 1542-4863. doi: 10. 3322/caac.21387. URL http://dx.doi.org/10.3322/caac.21387.

[25] Daniel E Worrall, Stephan J Garbin, Daniyar Turmukhambetov, and Gabriel J Brostow. Harmonic networks: Deep translation and rotation equivariance. In *CVPR*, volume 2, 2017.

[26] Xinge You, Qinmu Peng, Yuan Yuan, Yiu-ming Cheung, and Jiajia Lei. Segmentation of retinal blood vessels using the radial projection and semi-supervised approach. *Pattern Recognition*, 44(10-11):2314–2324, 2011.

[27] Lequan Yu, Hao Chen, Qi Dou, Jing Qin, and Pheng-Ann Heng. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Transactions on Medical Imaging*, 36(4):994–1004, 2017.

[28] Yading Yuan and Yeh-Chi Lo. Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks. *IEEE Journal of Biomedical and Health Informatics*, 2017.

[29] Yading Yuan, Ming Chao, and Yeh-Chi Lo. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. *IEEE Transactions on Medical Imaging*, 36(9):1876–1886, 2017.