Semi-Supervised Skin Lesion Segmentation Based on Pseudo-Labels

Bo Mu, JingXin Wei, and Yujun Zhang

Abstract—In recent years, deep learning has significantly advanced skin lesion segmentation. However, annotating medical image data is specialized and costly, while obtaining unlabeled medical data is easier. To address this challenge, we propose a semi-supervised learning strategy to improve segmentation accuracy by combining a small amount of annotated data with a larger volume of unlabeled data. Our approach employs a teacher-student model framework. In this framework, the teacher model generates pseudo-labels for the unlabeled data, and the student model is trained using both these pseudo-labels and the limited true labels. To improve the student model's learning capacity, we introduce auxiliary segmentation heads that provide joint guidance during training. We use the cross-entropy (CE) loss function to quantify the discrepancies between the segmentation outputs of the main head and auxiliary heads. Since pseudo-labels generated by the teacher model may contain noise, we developed a mechanism to identify and exclude uncertain regions in each unlabeled image. This reduces pseudo-label noise and mitigates its negative impact on the student model. Our method demonstrates significant improvements in skin lesion segmentation on the publicly available ISIC2018 dataset, achieving Dice coefficients of 87.84% and 88.73% with only 5% and 10% of the total annotated data, respectively, outperforming existing methods.

Index Terms—Medical Image Segmentation, Semi-Supervised Learning, Mean Teacher, Uncertainty Map.
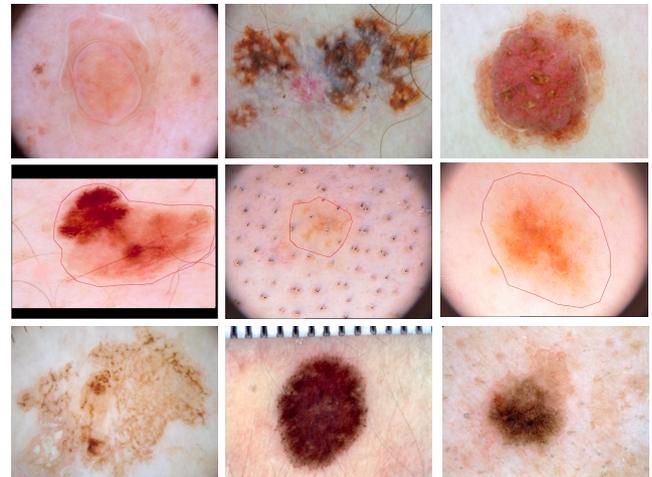
Fig. 1. Challenges in Melanoma Segmentation. The first row highlights artifacts and bubbles present in the images, which can complicate the segmentation process. The second row depicts cases where the annotations do not align with the actual boundaries of the lesions, adding to the difficulty in model training. Finally, the third row presents the correct annotations.

I. INTRODUCTION

SKIN diseases including basal cell carcinoma (BCC), melanoma, squamous cell carcinoma (SCC), and epithelial carcinoma, are highly prevalent worldwide. Among them, melanoma accounts for 75% of type of skin cancer [1]. Early symptoms of melanoma often appear as a small mole or spot, with changes like darkening or enlargement. Melanoma can be effectively treated in its early stages with minor surgical procedures [2]. Therefore, early detection is crucial. Diagnosing melanoma with high accuracy typically requires experienced physicians due to its complex presentation [3]. Recent advancements in artificial intelligence have significantly improved dermatological imaging, aiding physicians in diagnosis. This integration has enhanced diagnostic efficiency and reduced misdiagnosis rates.

Deep learning has revolutionized medical image segmentation, achieving remarkable success. In recent years, numerous methods based on deep learning have been proposed, leading to significant advancements in skin disease segmentation [4], [5], [6], [7]. Most existing methods utilize

Bo Mu is a postgraduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China (e-mail:1805248459@qq.com).

JingXin Wei is an undergraduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China (e-mail:2424322003@qq.com)

Yujun Zhang is a professor at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China (corresponding author, e-mail: m18804259885@163.com).

deep learning, with deep convolutional neural networks being the most prominent for skin cancer image segmentation. These methods primarily rely on fully supervised approaches, which perform best with sufficient annotated data [8]. However, generating reliable annotations in the medical field is highly time-consuming and labor-intensive. As a result, researchers have adopted alternative approaches for medical image segmentation, including weakly supervised [9], [10], [11], unsupervised [12], [13], [14], and semi-supervised learning [15], [16], [17].

Although deep learning methods have shown promise in skin disease image segmentation, challenges persist, particularly in melanoma segmentation owing to its complexity. Variability in lesion size, location, shape, and color, along with artifacts such as hair and blood vessels, often hinders accurate segmentation. Furthermore, as shown in Figure 1, discrepancies between true boundaries and annotations complicate model training. This issue is especially evident when labeled data is limited, as the model's generalization ability may be inadequate. Thus, the effective utilization of unlabeled data, reduction of pseudo-label noise, and improvement of segmentation performance remain critical challenges.

In this paper, we focus on semi-supervised learning for medical image segmentation, aiming to enable the model to learn effectively from a small set of annotated data and a large amount of unlabeled data. Semi-supervised learning in medical image segmentation is a fundamental yet challenging problem with critical implications for real-world clinical applications. Recently, significant progress has been

made in semi-supervised medical image segmentation, with many methods incorporating unlabeled data into the training process through unsupervised loss functions. Among them, the Mean Teacher model [18] has achieved notable success by enforcing consistency between the predictions of a student model and a teacher model using perturbed inputs. Building on this idea, researchers have proposed other semi-supervised learning algorithms based on consistency learning [19], [20]. Nevertheless, these consistency-based methods may not fully exploit the information available from unlabeled images, limiting their performance.

To address this, we propose UAMT (Uncertainty-guided Auxiliary Mean Teacher), an end-to-end semi-supervised segmentation framework that extends the Mean Teacher model.

In summary, our work achieves the following:

- Innovative Framework Design: We introduce UAMT, a semi-supervised segmentation model that builds on the Mean Teacher approach by adding a dual-headed student network. This design provides more robust guidance during training, enhancing the model's segmentation capabilities.
- Effective Use of Uncertainty Maps: By incorporating uncertainty maps, we identify and exclude unreliable regions in the unlabeled data, reducing noise in pseudo-labels. This allows the model to focus on more reliable areas, particularly edge information, which is crucial for accurate medical image segmentation.
- Enhanced Consistency Learning: We improve the learning process by applying a cross-entropy loss to quantify and minimize inconsistency between the two segmentation heads, ensuring better utilization of both labeled and pseudo-labeled data.
- Finally, our method is extensively validated on the publicly available ISIC2018 dataset, where it significantly outperforms other state-of-the-art semi-supervised segmentation methods, demonstrating its ability to effectively leverage unlabeled data.

II. Related work

A. Medical Image Segmentation

Medical image segmentation plays a crucial role in identifying distinct tissues or organs within images, facilitating both diagnosis and research. Initially, segmentation techniques were primarily based on traditional methods. These algorithms can be broadly categorized into two main approaches: threshold-based segmentation and edge detection-based segmentation.

The threshold-based segmentation algorithm [23], [24] is a relatively straightforward method that operates by selecting one or more thresholds to classify pixels in an image into different regions based on brightness or other features. However, its effectiveness may decrease in complex images characterized by significant noise or variability in illumination.

On the other hand, edge detection-based segmentation [25] focuses on identifying regions within the image where there are substantial changes in the intensity of the pixels. By detecting these intensity variations, the algorithm delineates edges, thus enabling the segmentation of the image into distinct areas. Although this method is well-suited for images with pronounced edges and high contrast, its performance can be compromised in the presence of blurred edges or considerable noise interference.

With the advent of technological advancements, deep learning-based image segmentation methods have emerged as the predominant focus in medical image segmentation research. These approaches leverage the capabilities of neural networks to learn complex patterns and features from data, significantly improving segmentation accuracy and robustness compared to traditional techniques.

In 2015, Ronneberger et al. introduced a distinctive U-shaped architecture known as U-Net [26], which revolutionized medical image segmentation. This model marked a significant advancement in the performance of deep learning-based segmentation techniques. The U-Net architecture effectively combines the strengths of convolutional neural networks (CNNs) and fully convolutional networks (FCNs). A key feature of U-Net is its designed skip connections, which facilitate the integration of low-level and high-level features extracted during the learning process, leading to substantial improvements in segmentation accuracy. Following the success of U-Net, numerous extensions and variants have been proposed to further improve segmentation performance. Among these, Attention U-Net [27], MultiResUNet [28], UNet++ [29], and TransUNet [31]—which incorporates Transformer [30] architecture—have introduced significant innovations.

Attention U-Net enhances the original U-Net by introducing an attention mechanism that reweights the output features of the encoder before concatenating them with the corresponding features in the decoder. This mechanism allows the model to focus on more relevant features, improving segmentation performance. MultiResUNet addresses the limitations of simple skip connections in U-Net, which directly fuse low-level features from the encoder with high-level features from the decoder. The authors argue that this direct fusion can create a semantic gap, potentially impairing the prediction process. To mitigate this gap, MultiResUNet introduces residual connections and additional convolutional layers within the skip connections, thereby enhancing model performance. UNet++ further refines segmentation capabilities by incorporating dense convolutional blocks, allowing for a more diverse fusion of image features from different depths. Additionally, TransUNet integrates a hybrid encoder that combines CNNs with Transformer architecture, resulting in a novel design that leverages the advantages of both structures to improve segmentation outcomes.

B. Semi-supervised Medical Image Segmentation

Despite the success of U-Net and its various extensions, all of these models rely on fully supervised learning, which requires large quantities of pixel-level annotated data. In medical imaging, such annotations must be provided by experts, leading to considerable time and resource costs. To address these limitations, semi-supervised learning methods have gained increasing attention, offering the potential to improve segmentation performance by utilizing a large amount of unlabeled data alongside a smaller, labeled dataset.

Several existing semi-supervised medical image segmentation methods have explored the generation of pseudo-

labels for unlabeled data to enhance the learning process. One of the earliest contributions was made by Lee et al. in 2013 [32], who introduced a semi-supervised approach based on deep learning that created pseudo-labels by using the predictions of fully supervised models. However, this strategy often resulted in noisy labels, which could degrade model performance.

In 2017, Tarvainen et al. [18] proposed the Mean Teacher model to mitigate this issue. They identified that relying heavily on the teacher model's predictions could introduce errors and negatively impact the student model's learning. To address this, the teacher model was updated by applying the exponential moving average of the student model's weights, leading to a more stable and improved learning process. Yu et al. [19] further extended this concept by introducing an uncertainty-aware mechanism, allowing the student model to gradually focus on more reliable pseudo-labels. This strategy enhanced the robustness of the learning process. More recently, in 2023, Zhang et al. [33] incorporated uncertainty quantification into the semi-supervised learning process, guiding model consistency learning. By leveraging Monte Carlo Dropout, they were able to estimate uncertainty in segmentation outputs, which improved the quality of pseudo-labels. Shen et al. [34] also advanced the field in the same year by proposing the UCMT (Uncertainty-guided Collaborative Mean Teacher) model. This approach introduced the Collaborative Mean Teacher (CMT) framework, which encouraged model diversity by training multiple sub-networks collaboratively. Additionally, the Uncertainty-guided Region Mixture (UMIX) technique was employed to adjust input images based on the CMT's uncertainty map, aiming to refine the generation of pseudo-labels.

III. METHOD

A. Overall Architecture

As illustrated in Figure 2, we propose a two-stage end-to-end semi-supervised segmentation model that integrates both teacher and student models. In the first stage, the student model is trained on labeled data through supervised learning, while the teacher model generates pseudo-labels from the unlabeled data to guide the student model's learning process. To improve the segmentation performance of the student model, we incorporate additional segmentation heads, forming a dual-headed structure where both heads collaborate by sharing complementary information. In the second stage, we address the challenge of high-uncertainty regions within the pseudo-labels generated for the unlabeled data. These uncertain regions are identified and refined using a patching strategy to create new training samples. These refined samples are then fed back into both the teacher and student models for further training. Importantly, the teacher and student models share parameters across both stages, with the teacher model's parameters being updated using the exponential moving average (EMA) of the student model's parameters [18]. The entire training process is outlined in Algorithm 1 using pseudocode.

B. Design of the Dual-Headed Segmentation Network

In a segmentation network, the primary purpose of a segmentation head is to generate the final segmentation

---

Algorithm 1 : Pseudocode of UAMT

Input:
Labeled Data and Unlabeled Data, $\mathcal{D}_L = \{(X_i, Y_i)\}_{i=1}^N$, $\mathcal{D}_U = \{\{X_j\}_{j=1}^M\}$

Parameter:
$f_{\theta_s}(x)$ = student model with parameters θ_s
$f_{\theta_t}(x)$ = teacher model with parameters θ_t

Output:
1: **for** T in $[1, \text{numepochs}]$ **do**
2:      **for** each minibatch B **do**
3:          **Step 1: Uncertainty Estimation**
4:          $\hat{Y}_i^1, \hat{Y}_i^2 \leftarrow f_{\theta_s}(x_i), \hat{Y}_i^0 \leftarrow f_{\theta_t}(x_i)$
5:          $\hat{Y}_j^1, \hat{Y}_j^2 \leftarrow f_{\theta_s}(x_j), \hat{Y}_j^0 \leftarrow f_{\theta_t}(x_j)$
6:          $L_{\text{total}} \leftarrow L_{\text{sup}}(\hat{Y}_i^1, \hat{Y}_i^2) + \lambda L_U(\hat{Y}_j^0, \hat{Y}_j^1, \hat{Y}_j^2)$
7:          **Update** $f_{\theta_s}(x), f_{\theta_t}(x)$ using optimizer
8:          $x_i^u \leftarrow U(f_{\theta_s}(x_i))$
9:          **Step 2: Train with Uncertain Regions**
10:         $\hat{Y}_i^u \leftarrow U_{\text{mix}}(x_i^u)$
11:         $\hat{Y}_j^u \leftarrow U_{\text{mix}}(x_j^u)$
12:      **end for**
13: **end for**
14: **return** $f_{\theta_s}(x)$

---

output. It does this by transforming the shared feature maps into pixel-wise or region-wise classifications (for example, classifying each pixel in a medical image as a specific tissue type or lesion). Each segmentation head is responsible for converting the extracted feature maps into concrete segmentation predictions. By employing different network architectures or strategies, multiple segmentation heads can interpret the same feature maps in diverse ways, enhancing the model's performance and robustness.

To improve the segmentation capability of the student model, we introduce an auxiliary segmentation head, forming a dual-headed network structure. This design enables the model to produce two complementary segmentation outputs from the same input image. Specifically, one segmentation head is based on the Fully Convolutional Network (FCN) architecture, while the other leverages the Deeplabv3 design. By integrating these two architectures, the dual-headed network can harness their strengths, generating diverse yet complementary segmentation results that improve overall accuracy and robustness.

To ensure consistency between the outputs of the two segmentation heads, we utilize a cross-entropy (CE) loss function that measures the difference between their segmentation predictions, as shown in Equation (1):

$$L_{CE}(y_{\text{head1}}, y_{\text{head2}}) \tag{1}$$

Where y_{head1} and y_{head2} represent the outputs from the FCN and Deeplabv3 heads, respectively. By minimizing this loss, we encourage alignment between the two segmentation results, which enhances the stability, robustness, and overall effectiveness of the segmentation model.

C. Uncertainty Map

To tackle the unreliability of predictions in medical imaging caused by factors like blurriness, noise, and low
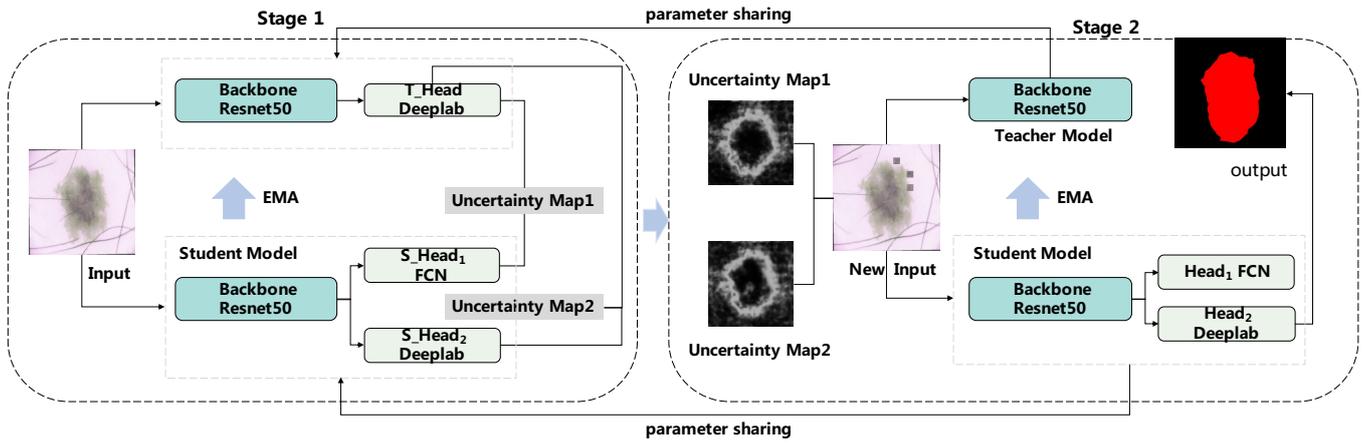
Fig. 2. The pipeline of our UMAT framework for semi-supervised segmentation. UMAT consists of two stages. In the first stage, an uncertainty map is generated based on the input data. In the second stage, this uncertainty map is used to create a new input, which is then fed into the second stage for further processing. Both stages share parameters.
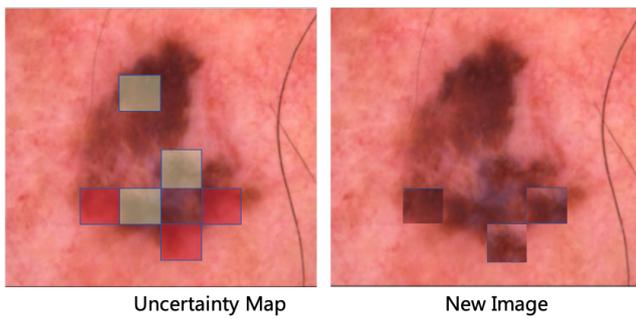

Fig. 3. Visualization of Uncertainty Map. In the left image, green areas represent regions with high confidence, while red areas indicate regions with low confidence.

contrast, we leverage a teacher-student model framework incorporating unlabeled data. Initially, the teacher model generates pseudo-labels by making predictions on unlabeled data. However, due to data quality challenges, these pseudo-labels may not always be reliable.

To enhance reliability, we introduce an uncertainty map that quantifies the model's confidence in its predictions by integrating outputs from both the teacher and student models. Specifically, we define the uncertainty map X_i^u as a mixture probability distribution given by:

$$X_i^u = U_{\text{mix}}(f_\theta(x_i), f_{\theta_1}(x_i)) = \sum_c p_c \log(p_c) \quad (2)$$

$$p_c = \frac{1}{2}(\text{softmax}(f_\theta(x_i)) + \text{softmax}(f_{\theta_1}(x_i))) \quad (3)$$

where $f_\theta(x_i)$ and $f_{\theta 1}(x_i)$ are the outputs of the teacher and student models, respectively, and p_c represents the averaged softmax probabilities of both model outputs.

With the uncertainty map established, we utilize the confidence information to integrate new data. We categorize the top K regions with high confidence as "trusted" regions and the top K regions with low confidence as "uncertain" regions. A patching approach is then employed, replacing data from uncertain regions with information from trusted regions to create new image data, as illustrated in Figure 3.

This process ultimately yields new image data, which is then input into the model for further training, thereby enhancing the overall two-stage model's robustness.

D. Loss Function

The loss function for our method is composed of a supervised loss L_s and an unsupervised loss L_u, defined as follows:

$$\text{Loss} = \lambda L_s + (1 - \lambda)L_u \quad (4)$$

In this equation, λ is a regularization parameter that balances the contributions of the supervised and unsupervised losses. To adapt to different training stages, we gradually increase λ using a Gaussian ramp function. This approach smoothens the transition in the weighting of losses, mitigating training instability caused by abrupt changes, as described by:

$$\lambda = \lambda_m \times \exp\left[-5\left(1 - \frac{t}{m}\right)^2\right] \quad (5)$$

Here, λ_m represents the maximum value of the scaling factor, t denotes the current training iteration, and m indicates the maximum number of iterations.

The supervised loss function is defined as follows:

$$\begin{aligned} L_s =& L_{\text{Dice}}(S1_{\text{head\_pred}}, y) \\ &+ L_{\text{CE}}(S1_{\text{head\_pred}}, y) \\ &+ L_{\text{Dice}}(S2_{\text{head\_pred}}, y) \\ &+ L_{\text{CE}}(S2_{\text{head\_pred}}, y) \end{aligned} \quad (6)$$

In this formulation, y represents the ground truth labels, $S1_{\text{head\_pred}}$ corresponds to the segmentation results produced by the student model's FCN head, and $S2_{\text{head\_pred}}$ corresponds to the results generated by the Deeplab head.

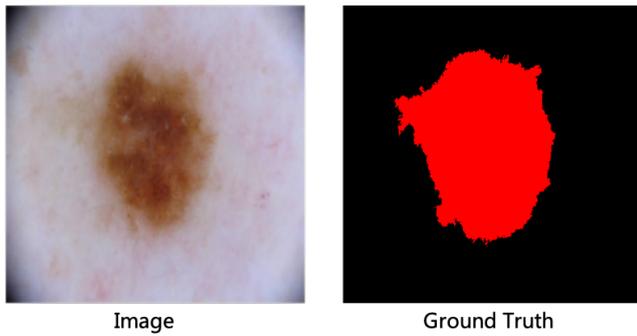The semi-supervised loss function is articulated as follows:

Fig. 4. ISIC2018 dataset, each entry comprises a pair of images. The left image represents the original dermoscopic image of the skin lesion under examination, while the right image provides the ground truth segmentation annotation.

with an NVIDIA GeForce GTX V100-32GB GPU and four Intel Xeon Processor CPUs operating at 2593.904 MHz. For the training process, we resized the input images to 256 x 256 pixels before feeding them into our semi-supervised segmentation model. The segmentation model utilized is DeepLabv3 [36], with ResNet50 serving as the backbone network [37]. We employed the AdamW optimizer to train the network, using a batch size of 16. The training was conducted over a total of 25 epochs to ensure adequate learning and convergence of the model.

$$\begin{aligned}
L_u =\ &L_{\text{CE}}(S1_{\text{head\_pred}}, T_{\text{pseudo}}) \\
&+ L_{\text{Dice}}(S1_{\text{head\_pred}}, T_{\text{pseudo}}) \\
&+ L_{\text{CE}}(S2_{\text{head\_pred}}, T_{\text{pseudo}}) \\
&+ L_{\text{Dice}}(S2_{\text{head\_pred}}, T_{\text{pseudo}}) \\
&+ L_{\text{CE}}(S1_{\text{head\_pred}}, S2_{\text{head\_pred}}) \\
&+ L_{\text{Dice}}(S1_{\text{head\_pred}}, S2_{\text{head\_pred}})
\end{aligned} \tag{7}$$

In this case, T_{pseudo} represents the pseudo-labels generated by the teacher model.

Both the supervised and unsupervised loss functions incorporate the Dice loss and cross-entropy (CE) loss. The Dice loss evaluates the overlap between the model's segmentation results and the ground truth labels, while the CE loss assesses the disparity between the predicted class probabilities for each pixel and the actual class labels.

IV. Results and Discussion

A. Dataset and Evaluation Metric

In this study, we focus on skin lesion segmentation using the publicly available ISIC2018 dataset [22]. All experiments were conducted on this dataset, which consists of a total of 2,594 images. For our analysis, we randomly divided the dataset into training and testing subsets, with the training set containing 1,815 images and the testing set comprising 779 images, each associated with its corresponding ground truth label.

The segmentation task involves two label classes: the lesion region and the non-lesion region, as illustrated in Figure 4. To train the semi-supervised segmentation algorithm, we selected 5% and 10% of the training images as labeled data, while the remaining images served as the unlabeled dataset.

For evaluation, we utilize the Dice similarity coefficient (Dice) metric to quantitatively assess the segmentation performance of our model. The Dice coefficient measures the overlap between the predicted regions and the ground truth annotations, with higher values indicating improved model performance. This metric is crucial for understanding how effectively our segmentation approach delineates the lesion regions from the non-lesion regions in the images.

B. Experimental Setup

Our proposed method was implemented using Python 3.8 and PyTorch 1.11.0 [35], running on a machine equipped

C. Comparison Experiments with Other Semi-Supervised Segmentation Methods

We evaluated the performance of our proposed semi-supervised method on the ISIC2018 dataset, comparing it with several relevant semi-supervised segmentation methods, including Mean Teacher [18], Cross-consistency Training [38], Cross Pseudo Supervision [39], and Uncertainty-guided methods [40].

The ISIC2018 dataset was divided into labeled subsets of 5% and 10%, with the results presented in Table I and Table II, respectively. Table I displays the experimental results using a 5% label proportion, while Table II showcases the outcomes with a 10% label proportion.

From Table I, it is evident that our method achieves a Dice coefficient of 87.84%. This represents improvements of 1.17% over Mean Teacher, 3.87% over Cross-consistency Training, 1.03% over Cross Pseudo Supervision, and 15.17% over Uncertainty-guided methods. Similarly, Table II indicates that our method enhances the Dice metric by 1.49%, 2.3%, 1.03%, and 9.25% when compared to the same four semi-supervised methods using a 10% label proportion.

By synthesizing the experimental results from both tables, we observe that our proposed method demonstrates greater overlap and similarity between the segmentation outcomes and the ground truth annotations. It excels in the Dice evaluation metric, outperforming other semi-supervised methods. This strong performance underscores the effectiveness of our approach, reinforcing its superiority in skin lesion segmentation tasks.

TABLE I
A COMPARISON EXPERIMENT WAS CONDUCTED UNDER THE TRAINING OF 5% LABELED DATA WITH OTHER SEMI-SUPERVISED METHODS.

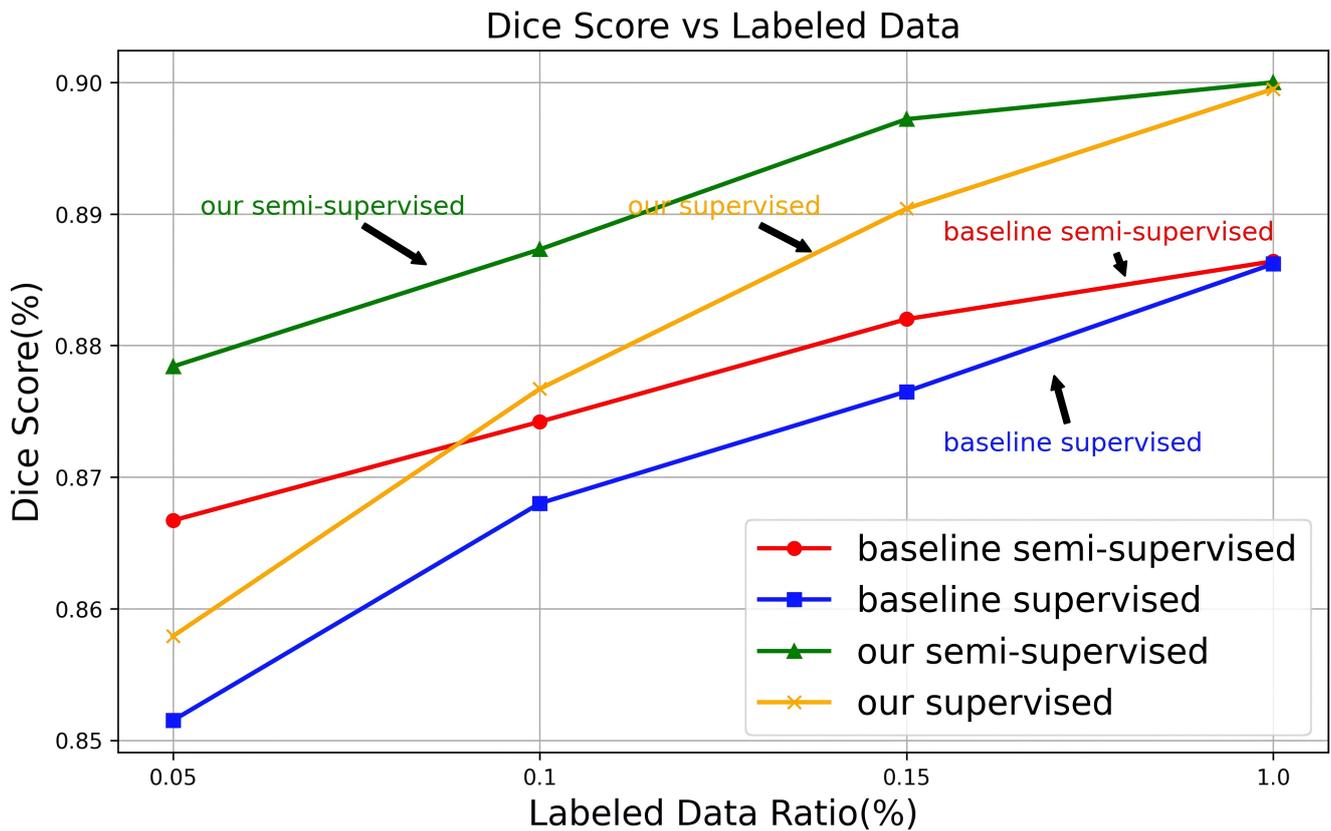| Methods	Labeled images	Unlabeled images	Dice(%)
Mean Teacher	91(5%)	1724	86.67
Cross-consistency training	91(5%)	1724	83.97
Cross pseudo supervision	91(5%)	1724	86.81
Uncertainty-guided	91(5%)	1724	72.67
UAMT (Ours)	91(5%)	1724	**87.84**

Fig. 5. Comparison of Segmentation Performance in Supervised and Semi-Supervised Settings.

TABLE II
A COMPARISON EXPERIMENT WAS CONDUCTED UNDER THE TRAINING
OF 10% LABELED DATA WITH OTHER SEMI-SUPERVISED METHODS.

Methods	Labeled images	Unlabeled images	Dice(%)
Mean Teacher	181(10%)	1634	87.42
Cross-consistency training	181(10%)	1634	86.43
Cross pseudo supervision	181(10%)	1634	87.70
Uncertainty-guided	181(10%)	1634	79.48
UAMT (Ours)	181(10%)	1634	**88.73**

D. Comparison of Supervised and Semi-Supervised Learning Performance

It is well-known that if we do not leverage additional unlabeled images, semi-supervised methods essentially operate as supervised methods. For example, in a supervised setting, we might train using only a small portion of the labeled data (such as 5%). In contrast, in the semi-supervised setting, we still use the same 5% of labeled data, but we also incorporate the remaining 95% of unlabeled data into the training process. This approach enables the model to benefit from both labeled and unlabeled data, allowing for better generalization. To evaluate the effectiveness of our proposed method, we compared its performance under both supervised and semi-supervised settings against a baseline method (Mean Teacher) and also assessed the performance of the baseline in both supervised and semi-supervised settings.

For this comparison, we conducted experiments with varying amounts of labeled data, focusing on the performance of our UAMT method, its fully supervised variant, as well as the baseline method (Mean Teacher) in both supervised and semi-supervised settings. The results, shown in Figure 5, clearly demonstrate that our semi-supervised UAMT approach outperforms both the supervised version of our method and the semi-supervised baseline (MT). These findings indicate that our method leverages the available unlabeled data more effectively, resulting in enhanced segmentation performance, even when the labeled data constitutes only a small fraction of the overall dataset. This underscores the potential of semi-supervised learning techniques in improving medical image segmentation, especially when labeled data is scarce.

TABLE III
ABLATION STUDIES ON DESIGNED MODULES.

Aux-head	Uncertainty-map	5% labeled	10% labeled
		86.67%	87.42%
✓		87.66%(+0.99%)	88.33%(+0.91%)
	✓	87.66%(+0.65%)	88.35%(+0.93%)
✓	✓	87.84%(+1.17%)	88.73%(+1.31%)

E. Ablation Experiments

To evaluate the contribution of each module in our proposed semi-supervised method, we conducted a series of ablation experiments. These experiments incrementally added different components to the baseline method, allowing us to assess their individual and combined impacts on model performance. We designed four experimental setups to validate the effectiveness of each module. Starting with the baseline method, we progressively introduced auxiliary
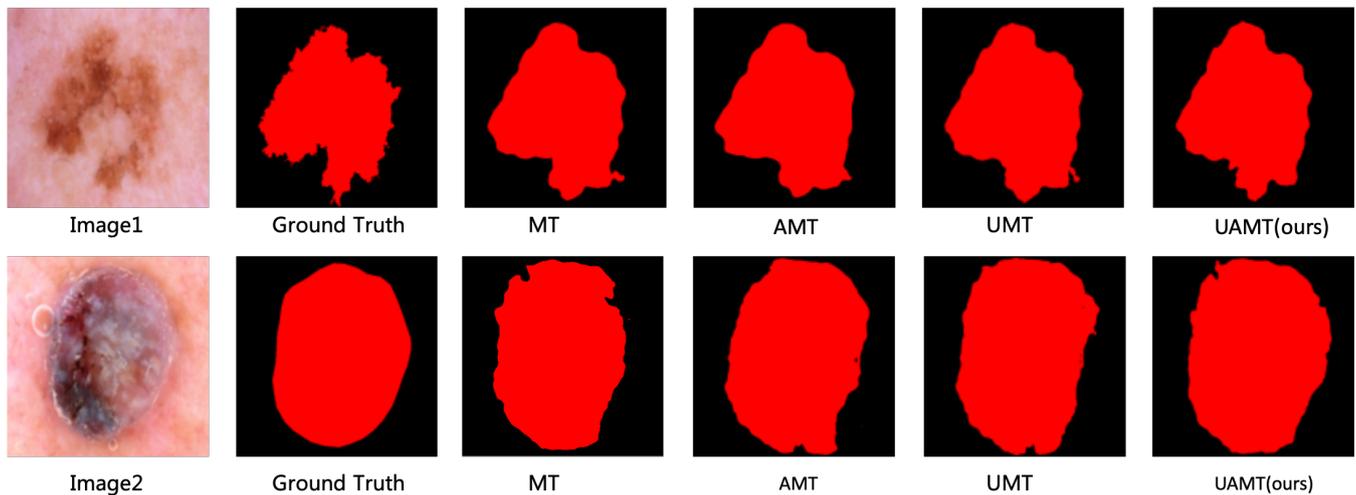
Fig. 6. The 2D visualization comparison between the baseline results, ground truth annotations, and the segmentation results obtained using different modules proposed in this paper.

segmentation heads, incorporated the uncertainty map, and finally combined both the uncertainty map and auxiliary segmentation heads.

The ablation experiments were carried out on the ISIC2018 dataset with label proportions of 5% and 10%. In Table III, we established Mean Teacher (MT) as the baseline method (Experiment 1). We then conducted subsequent experiments based on this baseline: Experiment 2 introduced auxiliary segmentation heads (AMT), Experiment 3 implemented the Uncertainty-map (UMT) training approach, and Experiment 4 represents our complete proposed method, integrating all modules into the baseline.

The results from Experiments 2 and 3 show that the inclusion of either the auxiliary segmentation head or the utilization of the Uncertainty-map method, when applied individually, enhances model performance for dataset segmentation. Notably, Experiment 4, which combines both modules, yields a significant performance increase on the ISIC2018 dataset. By integrating all methods into the baseline approach, our proposed method demonstrates improvements of 1.17% and 1.31% in the Dice evaluation metric for labeled proportions of 5% and 10%, respectively, compared to the baseline.

To further illustrate the efficacy of our approach, we randomly selected segmentation results of skin lesions from two patients for visual analysis. Figure 6 provides a 2D visual comparison of segmentation results obtained using the different modules proposed in this paper against the baseline method and the corresponding ground truth annotations. The figure clearly demonstrates that our method significantly enhances segmentation performance, particularly in boundary regions, when compared to the baseline method.

V. CONCLUSIONS

In this paper, we present a pseudo-label-based semi-supervised method for skin lesion segmentation using the ISIC2018 dataset. By integrating both labeled and unlabeled data, we establish a robust semi-supervised learning framework. To enhance the model's segmentation performance, we introduce auxiliary segmentation heads within the student model and apply the cross-entropy (CE) loss function between these heads. Additionally, we leverage uncertainty

maps to refine the quality of pseudo-labels for unlabeled data. Through comprehensive comparative experiments and ablation studies, we validate the contributions of each module to segmentation performance. The experimental results indicate that our proposed semi-supervised model outperforms existing semi-supervised methods on the ISIC2018 dataset. In future work, we plan to extend this approach to other medical image segmentation tasks, further refine the model architecture, and explore additional strategies for leveraging unlabeled data to further improve model performance.

REFERENCES

[1] Verma, R., Anand, S., Vaja, C., Bade, R., Shah, A., & Gaikwad, K. (2016). Metastatic malignant melanoma: A case study. *International Journal of Scientific Study*, **4**(1), 188–190.

[2] Celebi, M. E., Kingravi, H. A., Uddin, B., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V., & Moss, R. H. (2007). A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics*, **31**(3), 362–373.

[3] Whited, J. D., & Grichnik, J. M. (1998). Does this patient have a mole or a melanoma? *New England Journal of Medicine*, **338**(1), 100–107.

[4] Thomas, S. M., Lefevre, J. G., Baxter, G., & Hamilton, N. A. (2021). Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer. *Medical Image Analysis*, **68**, 101915.

[5] Reis, H. C., Turk, V., Khoshelham, K., & Kaya, S. (2022). InSiNet: A Deep Convolutional Approach to Skin Cancer Detection and Segmentation. *Medical & Biological Engineering & Computing*, **50**(1), 1–20.

[6] Kaur, R., GholamHosseini, H., Sinha, R., & Lindén, M. A. (2022). Automatic Lesion Segmentation Using Atrous Convolutional Deep Neural Networks in Dermoscopic Skin Cancer Images. *BMC Medical Imaging*, **22**(1), 103.

[7] Bibi, A., Khan, M. A., Javed, M. Y., Tariq, U., Kang, B. G., Nam, Y., ... & Sakr, R. H. (2022). Skin Lesion Segmentation and Classification Using Conventional and Deep Learning-Based Framework. *Computational Materials Science*, **198**, 2477–2495.

[8] Li, H., Pan, Y., Zhao, J., & Zhang, L. (2021). Skin Disease Diagnosis with Deep Learning: A Review. *Neurocomputing*, **464**, 364–393.

[9] Yang, G., Wang, C., Yang, J., Chen, Y., Tang, L., Shao, P., ... & Luo, L. (2020). Weakly-Supervised Convolutional Neural Networks of Renal Tumor Segmentation in Abdominal CTA Images. *BMC Medical Imaging*, **20**, 1–12.

[10] Chen, Z., Chen, Z., Liu, J., Zheng, Q., Zhu, Y., Zuo, Y., ... & Li, Y. (2020). Weakly Supervised Histopathology Image Segmentation with Sparse Point Annotations. *IEEE Journal of Biomedical and Health Informatics*, **25**, 1673–1685.

[11] Lerousseau, M., Vakalopoulou, M., Classe, M., Adam, J., Battistella, E., Carré, A., ... & Paragios, N. (2020). Weakly Supervised Multiple Instance Learning for Histopathological Tumor Segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V* (Vol. 23, pp. 470–479).

[12] Chen, C., Dou, Q., Chen, H., Qin, J., & Heng, P. A. (2020). Unsupervised Bidirectional Cross-Modality Adaptation via Deeply Synergistic Image and Feature Alignment for Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, **39**, 2494–2505.

[13] Xie, Q., Li, Y., He, N., Ning, M., Ma, K., Wang, G., ... & Zheng, Y. (2022). Unsupervised Domain Adaptation for Medical Image Segmentation by Disentanglement Learning and Self-Training. *IEEE Transactions on Medical Imaging*.

[14] Feng, W., Ju, L., Wang, L., Song, K., Zhao, X., & Ge, Z. (2023, June). Unsupervised Domain Adaptation for Medical Image Segmentation by Selective Entropy Constraints and Adaptive Semantic Alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 1, pp. 623–631).

[15] Luo, X., Chen, J., Song, T., & Wang, G. (2021, May). Semi-Supervised Medical Image Segmentation Through Dual-Task Consistency. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 10, pp. 8801–8809).

[16] Chaitanya, K., Karani, N., Baumgartner, C. F., Erdil, E., Becker, A., Donati, O., & Konukoglu, E. (2021). Semi-Supervised Task-Driven Data Augmentation for Medical Image Segmentation. *Medical Image Analysis*, **68**, 101934.

[17] Wu, Y., Ge, Z., Zhang, D., Xu, M., Zhang, L., Xia, Y., & Cai, J. (2022). Mutual Consistency Learning for Semi-Supervised Medical Image Segmentation. *Medical Image Analysis*, **81**, 102530.

[18] Tarvainen, A., & Valpola, H. (2017). Mean Teachers Are Better Role Models: Weight-Averaged Consistency Targets Improve Semi-Supervised Deep Learning Results. In *Advances in Neural Information Processing Systems 30*, pp. 68–101915.

[19] Yu, L., Wang, S., Li, X., Fu, C. W., & Heng, P. A. (2019). Uncertainty-Aware Self-Ensembling Model for Semi-Supervised 3D Left Atrium Segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II* (Vol. 22, pp. 605–613).

[20] Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., ... & Li, C. L. (2020). FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. In *Advances in Neural Information Processing Systems* (Vol. 33, pp. 596–608).

[21] Zhao, X., Qi, Z., Wang, S., Wang, Q., Wu, X., Mao, Y., & Zhang, L. (2023). RCPS: Rectified Contrastive Pseudo Supervision for Semi-Supervised Medical Image Segmentation. *IEEE Journal of Biomedical and Health Informatics*.

[22] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 Dataset, a Large Collection of Multi-Source Dermoscopic Images of Common Pigmented Skin Lesions. *Scientific Data*, **5**(1), 180161. doi:10.1038/sdata.2018.161.

[23] Huang, M., Yu, W., & Zhu, D. (2012). An Improved Image Segmentation Algorithm Based on the Otsu Method. In *2012 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing* (pp. 135–139).

[24] Upadhyay, P., & Chhabra, J. K. (2020). Kapur's Entropy Based Optimal Multilevel Image Segmentation Using Crow Search Algorithm. *Applied Soft Computing*, **97**, 105522.

[25] Xu, H., Xu, X., & Zuo, Y. (2019). Applying Morphology to Improve Canny Operator's Image Segmentation Method. *The Journal of Engineering*, **2019**(23), 8816–8819.

[26] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* (pp. 234–241).

[27] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention U-Net: Learning Where to Look for the Pancreas. *arXiv preprint arXiv:1804.03999*.

[28] Ibtehaz, N., & Rahman, M. S. A. (2020). MultiResUNet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation. *Neural Networks*, **121**, 74–87.

[29] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings* (Vol. 4, pp. 3–11).

[30] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention Is All You Need. In *Advances in Neural Information Processing Systems* (Vol. 30).

[31] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., ... & Zhou, Y. (2021). TransUnet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv preprint arXiv:2102.04306*.

[32] Lee, D. H. P. (2013). Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. In *Workshop on Challenges in Representation Learning, ICML* (Vol. 3, No. 2, p. 896).

[33] Zhang, Y., Jiao, R., Liao, Q., Li, D., & Zhang, J. (2023). Uncertainty-Guided Mutual Consistency Learning for Semi-Supervised Medical Image Segmentation. *Artificial Intelligence in Medicine*, **138**, 102476.

[34] Shen, Z., Cao, P., Yang, H., Liu, X., Yang, J., & Zaiane, O. R. (2023). Co-Training with High-Confidence Pseudo Labels for Semi-Supervised Medical Image Segmentation. *arXiv preprint arXiv:2301.04465*.

[35] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems* (Vol. 32).

[36] Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv preprint arXiv:1706.05587*.

[37] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778).

[38] Chen, X., Yuan, Y., Zeng, G., & Wang, J. (2021). Semi-Supervised Semantic Segmentation with Cross Pseudo Supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2613–2622).

[39] Ouali, Y., Hudelot, C., & Tami, M. (2020). Semi-Supervised Semantic Segmentation with Cross-Consistency Training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12674–12684).

[40] Wang, T., Lu, J., Lai, Z., Wen, J., & Kong, H. (2022). Uncertainty-Guided Pixel Contrastive Learning for Semi-Supervised Medical Image Segmentation. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22* (pp. 1444–1450).