Progressive Weighted Self-Training Ensemble for Multi-Type Skin Lesion Semantic Segmentation

CHEOLWON LEE, (Member, IEEE), SANGWOOK YOO, SEMIN KIM, AND JONGHA LEE
AI Research and Development Center, Lulu Lab Inc., Seoul 06054, South Korea
Corresponding author: Jongha Lee (jongha.lee@lulu-lab.com)

This work was supported in part by the Establishment and Demonstration of Smart City Platform to Support Untact Public Services Based on Artificial Intelligence (AI) Kiosk under Grant S3037924, in part by the “High-performance computing (HPC) Support” Project of the Ministry of Science and Information and Communications Technology (ICT), and in part by National IT Industry Promotion Agency (NIPA).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Severance Hospital Human Research Protection Center under Application No. 4-2020-1208.

ABSTRACT In this study, we propose the Progressive Weighted Self-training Ensemble (PWStE) method that reinforces efficiency of labeled data for multi-type skin lesion semantic segmentation. The generation of multi-type skin lesion labeled data is extremely expensive as it should only be performed by dermatologists due to the small pixel variations and irregularly shaped lesion characteristics. For the reason, the reality is that labeled data for skin lesion segmentation model training is absolutely insufficient. The core idea of the proposed PWStE method is to minimize the transfer of uncertainty in the training phase of general SSL by progressively using the pseudo-labeled data referenced in training. The PWStE uses procedures such as Progressive Selector, Ensemble, and Pseudo Labeler designed using conventional Semi-Supervised Learning (SSL) concepts to more accurately generate detailed features of skin lesions from unlabeled data to pseudo-labeled data. We performed ensembles using a combination of models (U-Net, FPN, LinkerNet, PSPNet) and backbones (ResNet50, EfficientNet-b3, InceptionV3, DenseNet121, SE-ResNet101, SE-ResNeXt101). Validation was performed on our Multi-Type Skin Lesion Label Database (MSLD) dataset compared to conventional SSL methods. The experiments have shown that the model trained with PWStE shows similar results to the model of trained the entire label data using the Supervised Learning (SL) method, even with 30% less label data. These results show that our proposed PWStE can increase the efficiency of the given labeled data even in the multi-type skin lesion field.

INDEX TERMS Deep learning, ensemble, multi-type skin lesion, semi-supervised learning, semantic segmentation.

I. INTRODUCTION

Semantic segmentation is one of the fundamental topics in computer vision to perform image pixel-wise classification for biomedical image analysis tasks. Recently, encoder-decoder based Deep Convolutional Neural Network (DCNN) models such as U-Net [1], FPN [2], LinkerNet [3], PSPNet [4] have been predominantly used in the biomedical field and achieved state-of-the-art using Supervised Learning (SL). As the value of data is highlighted with the development of technologies such as those mentioned above, many hospitals are accumulating raw data. However, the performance of these techniques cannot be fully used due to the lack of labeled data in skin lesion segmentation domain. The biggest reason is that it is expensive to create skin lesion segmentation labeled data from raw data. In general, skin lesions have features that are difficult for annotation tasks such as small pixel variations and irregular shapes. And several types of skin lesions with these characteristics may overlap or exist nearby. Even the cost of a dermatologist to distinguish lesions is not negligible.

Semi-supervised Learning (SSL) is one alternative to solve the drought of labeled data in the sea of raw data. The main strategy of SSL is to perform training with
using labelled as well as unlabelled data [5]. The SSL research trend is that consistency-based methods perform better than pseudo-labeling methods in multi-class problems. On the other hand, pseudo-labeling methods show superior performance in the multi-label problem. The multi-class problem is a task of classifying multiple classes containing one class in one image. The multi-label problem is a more realistic and difficult task to classify multiple classes in an image as a multi-type skin segmentation problem. However, pseudo-labeling method does not work well with the multi-type skin lesion semantic segmentation task. The multi-type skin lesions contain factors that make them difficult to distinguish, such as small pixel variations from lesion severity, overlapping regions, and irregular shapes. For these factors, pseudo-labeled data contain errors, as shown in Figure 1. These errors are amplified as it iteratively trains with the SSL method. The SSL does not necessarily improve the performance of SL for this reason, even if unlabelled data is added in the form of pseudo-labels; performance is sometimes compromised.

In this study, we propose the Progressive Weighted Self-training Ensemble (PWStE) method that can reduce the labeling cost through SSL on multi-type skin lesion data. The proposed idea consists of a progressive selector, self-ensemble model, and Pseudo Labeler. The progressive Selector suppresses error amplification by ratio scaling the labeled, unlabeled, and pseudo-labeled data through progressively increasing the amount of pseudo-labeled data used for training. The self-ensemble model is a strategy to create an encoder-decoder structure DCNN by combining various models and backbones, improves pixel-wise classification accuracy and, provides tolerance for uncertain pseudo-labeled data. Finally, the pseudo labeler transforms self-ensemble model inference results of unlabeled data requested by the progressive selector into a pseudo-label data.

We experiment using well-known semantic segmentation models and backbones to evaluate the performance of the proposed PWStE method, the models are U-Net [1], FPN [2], LinkerNet [3], and PSPNet [4], the backbones are Resnet50 [6], Efficientnetb3 [7], Inceptionv3 [8], Densenet121 [9], SE-ResNet101 [10], SE-ResNeXt101 [10], Efficientnetb7 [10].

Experimental results show that the proposed method achieved similar performance using a smaller dataset than the SL method.

The overall contributions of our study can be summarized as follows:

1) We designed the Progressive Weighted Self-training Ensemble (PWStE) method to utilize the pseudo-labeling SSL method in the multi-type skin lesion semantic segmentation field.

2) We proposed a simple weighting method called progressive selector that improve fine feature segmentation performance through reducing the error amplification of pseudo-label semi-supervised learning.

3) The proposed method can be easily attached to conventional SOTA backbones and architectures.

II. RELATED WORK
A. PREVIOUS WEIGHTING METHODS
SSL trains a model to cover unlabeled data based on available labeled data. At this time, the conventional method is to reduce uncertainty by weighting how much confidence to use the features extracted from this unlabeled data. Consistency based method directly weights supervised loss and consistency loss based on consistency relationship found in the data amplified by augmentation [11], [12], [13]. The pseudo-labeling method applies weights to each of the labeled data and the generated pseudo-labeled data at the stage of retraining [5], [14]. For example, the training dataset consists of labeled data $x_L$ and $y_L$, unlabeled data $x_U$, and $y_U$. The $n$ and $m$ the number of instances of labeled and unlabeled data, respectively. The $w$ is a weight parameter according to uncertainty for pseudo-labeled data, is adjusted to a value between 0 to 1. The dataset $D$ of typical SSL is expressed by the following formula:

$$D = \{(x_{ln}^m, y_{ln}^m(1-w))\} + \{(x_{un}^n, y_{un}^nw)\}$$

(1)

This approach is very common and works slightly well. The generated pseudo-label data has a lower impact than the label data according to the weight $w$. However, since this method uses all of the unlabeled data at once regardless of the ratio of the labeled data, it may cause imbalance of the sample. In this paper, we argue that a factor of performance degradation is the amplification of error due to such sample imbalance and low variance between lesions in multi-type skin lesion segmentation.

B. SELF-ENSEMBLE MODEL
The ensemble method generally yields better predictions for multiple neural networks than for single networks. In SSL, the ensemble method has been approached in various ways. [5] reduces the error rate by using different augmentation and dropout regularization of multiple sub-networks. Reference [14] uses a multi-resolution ensemble of stacked dilated U-Net for improving segmentation accuracy. Reference [15] adopt a deformable partbased model.
to capture a stable global structure and salient objects. Reference [16] proposed a iterative robust semi-supervised based imputation. Reference [17] proposed a method to ensemble SL and Unsupervised Learning (UL) based on fuzzy information. In such SSL, the ensemble technique can improve the noise and uncertainty about the ground truth, and it also makes the generalization performance robust.

III. METHODOLOGY

The goal of Progressive Weighted Self-training Ensemble (PWStE) is to reduce the cost of labeled data for multi-type skin lesion segmentation with features such as small pixel variations and irregular shapes by using SSL theory. More specifically, it is to use unlabeled data to generate pseudo-labeled data meaningful for training through self-training. PWStE consists of Progressive Selector, Self-ensemble, and Pseudo Labeler, the overview can be shown in Figure 2.

### A. PROGRESSIVE SELECTOR

Our proposed Progressive Selector is one way to effectively respond to the sample imbalance problem. In SSL, the sample balance problem is a balance problem between labeled data and unlabeled data. The training weight is strongly influenced by the sample. Therefore, methods to solve the imbalance problem have a significant impact on the performance of deep learning models. In general, SSL consists of a mechanism that first, infers unlabeled data with a model that has learned labeled data, and second, re-learns the inferred results from the data. For examples, [5], [11], [12], [13], [14], and [18] are designed in such a way that the sample balance greatly affects the learning result as shown in Equation 1. Of course, the uncertainty and noise of inaccurate unlabeled data are controlled through weights in the augmentation stage and loss stage, but the uncertainty increases with the number of samples of unlabeled data. Especially in situations where all classes must be performed on a pixel basis, such as semantic segmentation, the number of samples increases extremely, and uncertainty is amplified by repetition of training. We propose the simple but effective method of progressively increasing the proportion of pseudo-labeled data by diluting the weights for each sample. The unlabeled data amount is progressively increased at every training phase. Although this simplistic method must perform heavier operations, can generate more accurate pseudo-labeled data because it leverages the features of the unlabeled data as well as the labeled data for training. The error amplification can be suppressed by dividing the pseudo-labeled data into mini-batches when classifying multi-type skin lesions for segmentation according to the consistency regularization theory. Expressed as the following formula:

$$D_{PWStE} = \left\{ (x^l_n, y^l_n) \right\} + \left\{ (x^u_m, y^u_m) \right\}_{i=0}^{ref}$$

The $x^l$ is the image of the labeled data, $y^l$ is the annotation information of the labeled data. And $n$ is the number of samples in the labeled data. The $x^u$ is an image of unlabeled data, $y^u$ is annotation information of unlabeled data. The $m$ is the number of unlabeled data, $w$ is the weight that adjusts the number of unlabeled samples. We used $w$ as the ratio of using unlabeled data, increasing it according to the loss and average pixel accuracy, using an initial value of 0.01. This method minimizes the affection of initial non-optimized weights by providing a small unlabeled data set at the early stages of training.
B. SELF-ENSEMBLE

The Self-Ensemble is to get meaningful pseudo-label data through the bagging style ensemble of the encoder-decoder combination. The proposed self-ensemble focuses on using ground truth by generating inference results from various perspectives through different combinations of models and backbones, as shown in Figure 3. We believe that the effect of using various feature filters can be replaced by SOTA backbones such as Resnet50 [6], EfficientNetB3 [7], Inceptionv3 [8], DenseNet121 [9], SE-ResNet101 [10], and SE-ResNeXt101 [10]. Also, we think that resolution-related viewpoints can be secured through various segmentation models such as U-Net [1], FPN [2], LinkerNet [3], and PSPNet [4]. And each generated viewpoint is evaluated through Intersection over Union (IoU) metric per class from inference result.

The proposed self-ensemble model is also considered to train while effectively controlling sample imbalance and class imbalance. The sample imbalance is solved through a structure that can learn in response to the progressive selector that gradually increases the ratio of unlabeled data from a very small number. The class imbalance is a foreground-background class balance problem that has traditionally been dealt with in semantic segmentation tasks. In general, since the number of pixels in the background is larger than that in the foreground, reducing the bias from the background can improve the classification accuracy of classes. So, we use a loss that combines bce loss [19] and dice loss [20] to equation our loss function $L$:

$$L_{bce} = - \sum_{i} (y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i))$$  \hspace{1cm} (3)  

$$L_{dice} = 1 - \frac{\sum_{i} y_i p_i}{\sum_{i} y_i + \sum_{i} p_i}$$  \hspace{1cm} (4)  

$$L = L_{bce}(1 - \lambda) + L_{dice}\lambda.$$  \hspace{1cm} (5)

In our loss $L$ of Equation 5 enabling high tolerance pixel classification performance for multi-type skin lesions with complex color and gradient distribution. The $\lambda$ is a hyper-parameter with a value between 0 and 1, the 0.3 gave the best results in our MSLD dataset.

The training is performed through iterations of training and pseudo-labeling procedures. Training procedure learns $D_{PTStE}$ passed from Progressive Selector with model combinations of classifier parameter $i$. The training data are labeled data with annotations and unlabeled data. After training, the training procedure transfers the weight and the IoU evaluation result of the weight to the pseudo-labeling procedure. The pseudo-labeling procedure infers unlabeled data without annotation in $D_{PTStE}$ using the received model information such as weight, IoU. Then, the prediction result and class IoU are computed as a dot product. This result is sent to the pseudo labeler, and annotation $y^u$ is generated based on the pixel with the highest confidence through argmax operation. Finally, when all unlabeled data is used for training, the pseudo-labeled data is rearranged using the final weight. The PWStE algorithm can be shown in Algorithm 1.

Algorithm 1 Progressive Weighted Self-Training Ensemble  

Input:  
- Labeled data set $x^l$, $y^l$  
- Unlabeled data set $x^u$  
- Progressive selector weight parameter $w$  
- Classifier parameter $i$  
- Loss weight parameter $\lambda$  
- Backbone and model combinations $F = \{f_0, f_1, \cdots, f_m\}$  
- Ensemble algorithm $Ens$

Process:  
1. $D_{0 \ldots n} \leftarrow$ Progressive Selector($x^l$, $y^l$, $x^u$, $w$)  
2. Loop ($x^l$, $y^l$, $x^u$) from $D_0$ to $D_n$
   - TrainingProcedure:  
     - Weight, ClassIoU $\leftarrow$ $Ens(x^l$, $y^l$, $F$, $i$, $\lambda$)  
     - Pseudo-labelingProcedure:  
       - PredictionResult $\leftarrow$ $Ens(x^u$, $F$, $i$, $\lambda$, Weight)  
       - WeightedPredictionMap $\leftarrow$ PredictionResult $\cdot$ ClassIoU  
       - $y^u \leftarrow$ PseudoLabeler(WeightedPredictionMap, $i$)  
     - Update pseudo-labeled data ($x^l \leftarrow x^u$, $y^l \leftarrow y^u$)  
3. FinalWeight $\leftarrow$ TrainingProcedure:$Ens(x^l$, $y^l$, $F$, $i$, $\lambda$)

Output: FinalWeight

C. PSEUDO LABELER

We define a Pseudo Labeler as a module that determines class from weight prediction maps of self-ensemble model. The pseudo labeler uses a strategy that collects the opinions of all active classifiers while trusting the map with the higher IoU per class. The prediction result can be applied by the dot product operation of the class IoU to each class channel. Expressed as the following formula:

$$D_{pseudo}(pm_n, p_i) = \text{argmax}(\frac{1}{n} \sum_{i=1}^{n} (pm_i^h \cdot w_c \cdot p_i^c))$$  \hspace{1cm} (6)

The $pm$ is the prediction map composed of $h$, $w$, and $c$ axis, $h$ is height, $w$ is width, $c$ is class. The $pi$ is the predicted Class IoU constructed from class-specific IoU values. The $n$ is the number of ensemble model. For example, if the $n^{th}$ class IoU is 0.53, only 53% of prediction result of $n^{th}$ class is referenced for sharpening. The same operation is performed on the results of all classifiers. Then, the results of activated classifiers on which the IoU dot product operation
IV. EXPERIMENTS

A. DATASETS

We are building a Multi-Type Skin Lesion Label Database (MSLD). Out of a total of 27,790 frontal face images, 1,225 were annotated while the remainder was outsourced. MSLD consists of 19 classes and we used 11 classes in our experiments: Hair, NormalSkin, BlockDot, Freckle, Pih, Flush, Dark Circle, Papule, Pustule, AgeSpot, and Melasma. The resolution is 3288 x 2512 at 24-bit. The images are captured at autofocus and auto-white balance in the distance of 20 mm. Annotation of each image was performed by a dermatologist. In this experiment, multi-type skin information in the MSDL dataset image is anonymized to protect privacy. So the experiment results are compared through masks and part of images. Note that we do not disclose our MSLD dataset samples. We used 1,000 annotated images for training and 225 images for validation in all experiments.

B. EVALUATION OF PROGRESSIVE SELECTOR

We evaluated the performance of the progressive selector on the MSLD dataset. The evaluation was compared with progressive as Equation 2, non-progressive, and previous weight methods as Equation 1. The progressive method trains by progressively increasing the amount of data used for learning from 1 percent according to the progressive selector. The Non-progressive methods learn by increasing a fixed amount by 10% of the data per training. The Previous weight method uses all of the data for training and weights the data itself, experiments give weights from 1 to 10 and present the best results.

The model and backbone were used as U-Net and EfficientNet-b3, ensemble is not used. The progressive method resulted in 8 and 1% IoU reduction in normal skin and hair regions, respectively. However, It show better results than previous weight method for target classes, such as Melasma 6%, Pih 14%, Freckle 3%, Papule 2%, and Dark Circle 10%, as shown in Figure 6. Additionally, we conducted an experiment comparing progressive and non-progressive methods to prove the amplification of errors triggered by the sample imbalance. This result showed that the progressive method performed better in skin lesion segmentation than the non-progressive method, the non-progressive method showed better results than the previous weight method. Finally, the sample imbalance problem is actually affecting the accuracy of the model, which means that it can be improved through controlling the ratio of labeled data to unlabeled data.

C. EVALUATION OF PROGRESSIVE WEIGHTED SELF-TRAINING ENSEMBLE

We could not find a public data set containing multi-type skin lesion to evaluate the performance of the proposed PWStE. The MSLD dataset is difficult to disclose due to the user’s privacy policy. The training model of this experiment was implemented through modification of Segmentation Models [21]. The random seed of all cases were fixed 110 value for consistency. We used U-Net, FPN, LinkNet, PSPNet models and Resnet50, EfficientNet-b3, InceptionV3, DenseNet121, SE-ResNet101, SE-ResNeXt101 backbones.

We use the pixel accuracy and mean Intersection over Union (mIoU) metrics for evaluation. The pixel accuracy an extension of the commonly known the accuracy metric in pixel units as shown in Equation 7:

\[
\text{pixelAccuracy} = \frac{\sum_{i=1}^{N} PR_{ii}}{\sum_{i=1}^{N} GT_{i}}
\]

The \( N \) represents the number of classes, the sigma \( GT_{i} \) of the denominator term means the total of pixels of \( i \)th class. The molecular term \( PR_{ii} \) means the sum of pixels in \( i \)th class prediction result among which groundtruth is \( i \)th.
TABLE 1. The mIoU performance of combinations in Multi-Type Skin Lesion Label Database (MSLD) dataset. It shows the individual performance of each model and combination of backbones, and shows also the performance according to the amount of training dataset. The 10% means that the model trained only 100 samples in 1000 samples. The 100% results mean the supervised learning performance of all 1000 samples each combination.

| Model         | Backbone 10%  | Backbone 20%  | Amount of Train Dataset (mIoU)  |
|---------------|---------------|---------------|---------------------------------|
|               | 10%           | 20%           | 30%                            | 40% | 50% | 70% | 100% |
| Unet          | Resnet50      | 27.53         | 35.58                          | 37.88 | 41.12 | 43.45 | 47.36 | 57.21 |
|               | EfficientNetB3| 31.76         | 41.89                          | 42.74 | 44.1  | 47.19 | 49.41 | 61.35 |
|               | InceptionV3   | 28.53         | 37.91                          | 38.64 | 39.29 | 41.83 | 45.8  | 56.14 |
|               | Densnet121    | 25.74         | 35.53                          | 37.49 | 38.44 | 40.92 | 44.64 | 53.16 |
|               | SE-ResNet101  | 27.17         | 35.86                          | 38.37 | 39.37 | 43.64 | 53.1  | 57.42 |
|               | SE-ResNetX101 | 26.48         | 34.22                          | 41.49 | 44.66 | 45.02 | 53.72 | 55.32 |
| FPN           | Resnet50      | 28.86         | 37.33                          | **40.18** | 41.31 | 43.3  | 50.08 | 54.93 |
|               | EfficientNetB3| 32.15         | 36.85                          | 40.02 | 40.13 | **45.45** | 49.11 | 56.15 |
|               | InceptionV3   | 29.37         | 37.02                          | 39.16 | 41.93 | 41.39 | 46.06 | 54.11 |
|               | Densnet121    | 28.61         | 35.07                          | 38.01 | 40.64 | 42.33 | 46.49 | 56.1  |
|               | SE-ResNet101  | 27.14         | **38.47**                      | 39.66 | 42.76 | 43.76 | 48.27 | 54.08 |
|               | SE-ResNetX101 | 27.86         | 37.4                           | 39.1  | **43.83** | 45.28 | 48.43 | 53.73 |
| LinkNet       | Resnet50      | 26.05         | **39.61**                      | 38.98 | 42.16 | 43.01 | 46.27 | **58.84** |
|               | EfficientNetB3| 29.61         | 36.49                          | **39.81** | **42.59** | 43.67 | 48.76 | 57.98 |
|               | InceptionV3   | 26.8          | 33.43                          | 38.8  | 41.63 | **44.73** | 48.5  | 52.75 |
|               | Densnet121    | 25.73         | 31.96                          | 37.42 | 41.77 | 41.02 | 47.77 | 56.81 |
|               | SE-ResNet101  | 31.28         | 36.05                          | 38.38 | 42.45 | 43.04 | 48.81 | 53.17 |
|               | SE-ResNetX101 | **34.71**     | 36.28                          | 39.4  | 42.7  | 42.56 | **49.66** | 57.2  |
| PSPNet        | Resnet50      | 29.31         | 36.28                          | 40.86 | 42.39 | 43.5  | 47.23 | 54.18 |
|               | EfficientNetB3| 25.93         | 35.48                          | **41.23** | **43.68** | 45.02 | **49.68** | 57.88 |
|               | InceptionV3   | **31.25**     | 36.08                          | 40.78 | 42.88 | 44.56 | 49.66 | 55.44 |
|               | Densnet121    | 29.21         | 35.03                          | 38.47 | 41.19 | 43.7  | 46.13 | 55.95 |
|               | SE-ResNet101  | 30.36         | 36.4                           | 39.39 | 42.73 | **45.72** | 49.34 | 53.42 |
|               | SE-ResNetX101 | 25.72         | **37.04**                      | 39.55 | 41.9  | 44.51 | 48.29 | 57.64 |

TABLE 2. The mIoU and Pixel Accuracy performance based on the number of ensemble models at 50% and 70% of data volume. The $i$ is the number of classifiers used in the ensemble, which is the average of 10 times of a random combination.

| $i$ | 50% | 70% |
|-----|-----|-----|
|     | mIoU | Pixel Accuracy | mIoU | Pixel Accuracy |
| 2   | 48.31 | 73.38 | 53.29 | 82.06 |
| 3   | 49.87 | 74.14 | 54.42 | 83.37 |
| 5   | 50.24 | 76.51 | 56.31 | 85.54 |
| 6   | 51.1  | 77.32 | 58.61 | 88.19 |
| 7   | **51.16** | **77.92** | **58.97** | **88.22** |
| 8   | 51.03 | 77.81 | 58.16 | 88.47 |

TABLE 3. The comparison of mIoU performance between the PWE and the previous SSL ensemble methods. 50% means the ratio of labeled data and unlabeled data, is labeled data ratio in 1000 images.

| Method                     | 50% | 70% |
|----------------------------|-----|-----|
| SL: U-Net - EfficientNet-B3 | 47.19 | 49.41 |
| SL: U-Net - SE-ResNetX101   | 45.02 | 53.72 |
| PWStE + Classifier $i = 2$  | 48.31 | 53.29 |
| $\Pi$-model                | **52.33** | **55.19** |
| Multi-resolution Model      | 51.97 | 57.26 |
| PWStE + Classifier $i = 5$  | 51.1 | 58.61 |
| PWStE + Classifier $i = 7$  | 51.16 | **58.97** |
Accuracy 88.22 that is 5.25% better than mIoU 53.72 of best combination in Table 1 70%. We argued that using multiple viewpoints and feature filters across ensembles would help generate pseudo-labeled data for multi-type skin lesion semantic segmentation. And this experiment shows that different combinations of ensemble structures lead to mIoU enhancement. Of course, simply stacking ensembles infinitely doesn’t make things better. In the MSLD dataset, we confirmed that saturation occurs in the 5 to 8 section of ͞. Finally, we compare the performance of supervised learning, previous Semi-Supervised Learning and PWStE method in Table 3. For comparison of the results of supervised learning, we selected the U-Net+EffectiveNet-B3 combination and U-Net+SE-ResNeXt101 combination with the best results at 50% and 70% in Table 1, and 70% of SL used only 50% and 70% of the total data. The comparison of the previous SSL uses the result obtained by implementing ͞-model [5] and Muli-resolution model [14] and trained it with our data. The proposed PWStE is compared with the result based on the number of combinations being ensembled. Here, the previous SSL and PWStE methods used the remaining unlabeled additionally for learning, unlike SL. The SL method and the SSL method showed better results by an average of 4.86% at 50% case and 5.08% at 70% case, the SSL methods include previous SSL and PWStE. The PWStE + Classifier ͞ = 7 shows the best result with mIoU 58.97 in 70% of cases compared to ͞-model 55.19 and Muli-resolution model 57.26. Nevertheless, in 50% cases, the results are slightly lower than the previous SSL methods compared. This is because self-ensembles need enough data to work well. In more than 70% of cases, the results were better than the performance of 100% SL and previous SSL.

V. CONCLUSION

We proposed a Progressive Weighted Self-training Ensemble (PWStE) method that can reduce the labeling cost by utilizing unlabeled data in multi-skin lesion semantic segmentation data, including the small pixel variations and irregularly shaped lesion. The progressive selector attenuated the amplification of errors in an environment where lesions overlap, and the ensemble model reduced the error of pseudo-labeled data. We verified through experiments that PWStE obtained similar results to that of using 100% of labeled data in the supervised learning method even when using 30% less labeled data. This means that it is more accurate to generate pseudo-labeled data by ensembles the inferred results of multiple models without sample and class imbalance. In fact, most of the deep learning deployment environments have unlabeled data processing issues, and we expect that our PWStE will help solve these issues, especially in the multi-type skin lesion field with large annotation cost. However, the computation time increases by about five times compared to previous methods. In the future, we plan to design a lightweight model considering the time complexity of PWStE. In addition, there is still a problem that SSL does not exceed the performance of SL when the same size of data is used.

REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham, Switzerland: Springer, 2015, pp. 234–241.
[2] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2881–2890.
[3] A. Chaurasia and E. Culurciello, “LinkNet: Exploiting encoder representations for efficient semantic segmentation,” in Proc. IEEE Int. Conf. Image Process. (VCIP), Dec. 2017, pp. 1–4.
[4] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2117–2125.
[5] S. Laine and T. Aila, “Temporal ensembling for semi-supervised learning,” 2016, arXiv:1610.02242.
[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.
[7] M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. Int. Conf. Mach. Learn., 2019, pp. 6105–6114.
[8] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2818–2826.
[9] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 4700–4708.
[10] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7132–7141.
[11] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. Raffel, “MixMatch: A holistic approach to semi-supervised learning,” 2019, arXiv:1905.02249.
C. Lee et al.: PWStE for Multi-Type Skin Lesion Semantic Segmentation

SANGWOOK YOO received the B.S. degree in computer science from Sogang University, South Korea, in 2005, and the M.S. and Ph.D. degrees in computer science from the Korea Advanced Institute of Science and Technology (KAIST), in 2008 and 2013, respectively. He was a Senior Engineer at the Health and Medical Equipment (HME) Business Division, Samsung Electronics. He co-founded Lulu Lab Inc., in 2017, where he is currently the CTO. His recent research interests include skin analysis using deep learning algorithms, optics for skin analysis, and image analysis-based disease prediction.

SEMIN KIM received the M.E. degree in computer engineering from Kyungpook National University, Daegu, South Korea, in 2008, and the Ph.D. degree in information and communication engineering from the Korea Advanced Institute of Science and Technology, Daejeon, in 2014. He was a Staff at the Korea Institute for Advancement of Technology, Seoul, in 2008. He worked at Samsung Electronics for medical image recognition based on deep learning, from 2014 to 2019. He worked at Hyundai Mobis for autonomous driving using road image recognitions, from 2019 to 2021. Currently, he is analyzing face skin diseases based on deep learning with Lulu Lab Inc., Seoul. His main research interests include image recognitions based on deep learning and implementation of image inference systems.

JONGHA LEE received the B.S. degree in electronics from Kyungpook National University, Daegu, South Korea, in 1999, and the M.S. degree in electrical engineering from Seoul National University, Seoul, South Korea, in 2001. He joined as a Researcher at the Samsung Advanced Institute of Technology (SAIT), where he became a Senior Researcher, in 2007. During the SAIT Member, his research topics were face detection and recognition, until 2008, and medical image processing, until 2014. In 2014, he moved to Samsung Electronics, where he was a Senior Engineer dedicated to medical image processing for computer aided diagnosis (CAD) and radiography systems. Since March, 2021, he has been the Director with the AI Research and Development Center, Lulu-Lab Inc. His current research interests include machine learning, computer vision, and pattern recognition for skin disease diagnosis and its applications.

**CHEOULWON LEE (Member, IEEE)** was born in Chungju, South Korea, in 1987. He received the B.S., M.S., and Ph.D. degrees in computer engineering from Konkuk University, South Korea, in 2012, 2015, and 2020, respectively. In 2020, he joined the Department of Smart Farm Application, Electronics and Telecommunications Research Institute (ETRI). Since November 2021, he has been with the AI Research and Development Center, Lulu Lab Inc., where he is a Researcher. His current research interests include facial lesion diagnosis using vision deep learning, semi-supervised learning, image processing technology, and robotics for autonomous driving.

**P. Yakubovskiy. (2019).** Segmentation Models. [Online]. Available: https://github.com/qubvel/segmentation_models