IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 7, 2021, accepted July 17, 2021, date of publication July 26, 2021, date of current version July 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3099936

A Simple Generic Method for Effective Boundary Extraction in Medical Image Segmentation

MINKI KIM AND BYOUNG-DAI LEE
Division of AI Computer Science and Engineering, Kyonngi University, Suwon 16227, Republic of Korea
Corresponding author: Byoung-Dai Lee (blee@kgu.ac.kr)

This work was supported by Kyonggi University Research Grant 2020.

ABSTRACT Accurate identification of the shape and position of organs and abnormal objects (e.g., tumors) in medical images plays an important role in surgical planning as well as in the diagnosis and prognosis of diseases. However, this is difficult to achieve from two-dimensional medical images as these images present inaccurate and ambiguous organ boundaries. Further, traditional image processing-based boundary detection methods such as the Canny edge detector and Sobel operator exhibit poor boundary detection performance for images with substantial noise. Recently, the use of deep learning has resulted in improvements in semantic segmentation in medical images. In this paper, we propose a generic boundary-aware loss function to facilitate the effective discernment of the boundaries of organs and abnormal objects in medical images. Specifically, the proposed loss function introduces a boundary area and assigns higher weights to the loss of pixels located in the boundary area than to those in the non-boundary areas, thereby promoting effective learning in the boundary area. The results of experiments conducted using public medical datasets comprising colon polyp, skin lesion, and chest X-ray data indicate that the standard loss functions, such as cross-entropy loss and Dice loss, combined with the proposed boundary-aware loss function, achieve comparable or better performance than those without the boundary-aware loss function.

INDEX TERMS Boundary extraction, deep learning, loss function, medical image analysis, segmentation.

I. INTRODUCTION

Precisely distinguishing the boundary of each region of organs and abnormal objects of interest is essential when the segmentation results of such entities are used for diagnosing diseases. Various recent studies have combined deep learning with medical images for tasks such as detecting cardiac hypertrophy by locating the heart and lungs in X-ray images [1] and identifying brain tumors in many clinically produced magnetic resonance imaging (MRI) images [2]. However, two-dimensional (2D) medical images, such as X-ray images, show inaccurate and ambiguous boundaries of organs because three-dimensional (3D) objects are represented in 2D in such images. Thus, it is difficult to accurately distinguish boundaries in medical images without expert anatomical knowledge [3].

Traditional image processing methods for boundary detection, such as the Canny edge detector [4] and Sobel operator [5], exhibit poor boundary detection performance for images with substantial noise owing to their relatively simple algorithm processing method [6]. Recently, deep learning-based image segmentation has been applied to detect the boundaries of organs and specific objects of interest with encouraging results [7]. In particular, a convolutional neural network (CNN) has been used to extract the features of input images and is frequently used to detect boundaries in medical images [8]. The three necessary components for obtaining valid and generalized results from deep learning are as follows: a deep neural network structure, a rich dataset to train the network with, and a loss function to optimize the network parameters. Many studies have been conducted on neural network structures because they can be designed according to the characteristics of the dataset and the learning objective. In addition, the use of feature maps extracted from the learning process is also being actively researched.

A loss function is a crucial component in deep learning because it calculates the difference between the ground truth and the predicted value in the deep learning process. Representative loss functions that are currently used to segment medical images include the mean square error (MSE) loss [9], cross-entropy (CE) loss [10], and Dice loss [11]. The MSE loss is mainly used for linear regression problems.

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

It calculates the loss value by squaring the difference between the value predicted by the CNN and the ground truth and using it for learning. CE loss is used in multi-class division problems. Here, a loss function is obtained between the class predicted by the CNN at the highest probability and the class of the ground truth, and is used for learning. Dice loss uses the Dice coefficient between the value predicted by the CNN and the ground truth for learning. Currently, novel loss functions based on the differential framework or distance maps are also being researched in addition to these traditional loss functions [12]. A feature common to many of these loss functions is that pixels that are close to the border as well as those that are not, are considered equally important.

This paper proposes a loss function to facilitate effective discernment of the boundaries of organs and abnormal objects (e.g., tumor) in 2D images. The proposed loss function introduces a boundary area and assigns higher weights to the loss of pixels located in the boundary area than to those in the non-boundary areas, thereby promoting effective learning in the boundary area. The efficacy of the proposed loss function was evaluated using three medical image datasets: polyp dataset, skin lesion dataset, and chest X-ray dataset.

The remainder of this paper is organized as follows. Section II presents an overview of the studies related to the proposed loss function. Section III explains the proposed loss function in detail. Section IV describes the datasets used in the experiments. Section V validates the proposed loss function by applying it to each dataset described above and compares its performance. Finally, Section VI analyzes the performance of the proposed loss function and outlines future research directions.

II. RELATED WORK

Effective extraction of the boundary area of organs and other specific objects of interest is a critical part of image segmentation. Traditional image processing techniques such as edge detection algorithms have been used to extract boundaries in input images. In addition, adaptive histogram equalization and contrast-limited adaptive histogram equalization have been used to identify the distribution of light and dark parts in images and readjust them for effective boundary distinction.

With the progressively increasing number of studies focusing on deep learning, efforts to effectively detect boundary areas are being made, such as improvements in the neural network structure and changes to the loss function. Various network structures apart from U-Net [13], which is predominantly used in the medical image segmentation field, are being researched. DoubleUNet [14], which is a combination of two U-Nets, has effectively extracted features using atrous spatial pyramid pooling [15] instead of an existing pooling method. Furthermore, experiments using the 2015 MICCAI sub-challenge on automatic polyp detection dataset [16] and the 2018 Data Science Bowl challenge dataset [17] showed improved segmentation performance for DoubleUNet compared to other segmentation networks. PraNet [18] used an inverted image of the ground truth in the learning process as the input of an attention module, realizing an effective medical image segmentation performance. The self-attention technique, which is mainly used in recurrent neural networks (RNNs), is applied to CNNs; this technique plays a role in emphasizing the important parts in the learning process. The self-attention technique has been applied to the neural network structure in various ways, and has shown improved performance compared to existing neural networks [19].

The traditional medical image segmentation approach uses basic loss functions such as MSE, CE, and Dice, which are also used in image classification and object detection. Many loss functions specializing in boundary distinctions have recently been researched [12]. For example, Kervadec *et al.* [20] effectively distinguished the boundaries of brain tumors by calculating the difference between the area predicted by a neural network and the ground truth area using the differential framework. The HausDorff loss [21] has been applied to effectively distinguish boundaries by minimizing the HausDorff distance between the area predicted by a neural network and the ground truth area. Caliva *et al.* [22] used a loss function with a distance map for a dataset with data imbalance and effectively distinguished a 3D knee joint boundary. The focal loss function [23] resolves the data imbalance problem, and when the prediction by the neural network is incorrect, it facilitates the accurate training of the neural network by assigning weights to the loss values. The boundary-attention semantic loss [24] is similar to the proposed loss function in that the loss of pixels located on the boundaries is weighted. However, the proposed loss function has a more general form as it allows both the boundaries and the neighboring areas to be considered in learning definitive boundary features.

III. BOUNDARY-AWARE LOSS FUNCTION

This paper proposes a boundary-aware (BA) loss function that can effectively distinguish boundaries. The proposed function calculates the loss values for boundary and non-boundary areas separately to perform learning with a focus on the boundary area. It sets weights to the loss values of the boundary area so that learning can be performed accurately when the neural network makes an incorrect judgment in a boundary area. A boundary area refers to the inner and outer areas starting at a boundary and extending to a distance of r pixels from the boundary, as shown in Fig. 1. The boundary areas of individual training images are pre-computed in the training process using the range of the boundary area and corresponding ground truths. Therefore, when calculating the loss value of a model prediction, the deep learning model determines whether each pixel is included in the boundary area and calculates the loss values separately for the boundary and non-boundary areas.

Equation (1) is the general form of the BA loss function, where L_B denotes the loss value for the boundary area,

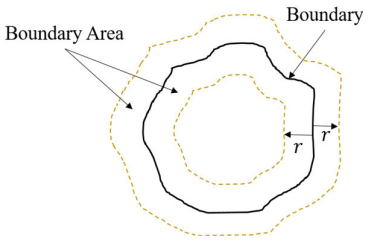

FIGURE 1. An example of a boundary area (Thick black line: boundary, brown dotted line: boundary of the boundary area, r: range of the boundary area).

L_{notB} denotes the loss value for the non-boundary area, and p_1 and p_2 denote the weights given to each area. Each weight emphasizes the boundary and non-boundary areas in the learning process. Equation (2) shows an example of implementing the proposed BA loss function by applying it to the MSE loss function.

$$L_{BA} = p_1 \times L_B + p_2 \times L_{notB}, \quad p_1 > p_2 > 0 \quad (1)$$

$$L_{MSE+BA} = \frac{1}{n}\left(p_1 \sum_{i}^{n} b(i)\,(y_i - t_i)^2 + p_2 \sum_{j}^{n} (1 - b\,(j))\,\left(y_j - t_j\right)^2\right) \quad (2)$$

The MSE loss function calculates the loss by squaring the difference between the ground truth (t_i) and the predicted value (y_i). In Equation (2), n denotes the total number of pixels in the input image. The $b(i)$ is an indicator function to determine if a given pixel i belongs to the boundary area. That is,

$$b\,(i) = \begin{cases} 1, & \text{if the pixel } i \text{ belongs to the boundary area,} \\ 0, & \text{otherwise} \end{cases}$$

If k denotes the number of pixels located within the boundary area, then the BA loss function calculates the loss values of k pixels in the boundary area and $(n-k)$ pixels in the non-boundary area based on the MSE loss function, and weights are given for p_1 and p_2. Finally, the loss values calculated using the BA loss function are added to the two loss values (boundary area and non-boundary area loss values), which are divided by n to obtain the average loss value.

As shown in Equation (3), the BA loss function is sufficiently general that it can be used in combination with various other standard loss functions, in addition to the MSE loss function. For example, it can be combined with the CE loss (e.g., L_{CE+BA}) and Dice loss (e.g., $L_{Dice+BA}$) functions, which are frequently used in image segmentation.

$$L_{CE+BA} = -\frac{1}{n}\left(p_1 \sum_{i}^{n} b(i) y_i \log t_i + p_2 \sum_{j}^{n} (1 - b(j)) y_j \log t_j\right)$$

$$L_{Dice+BA} = \frac{1}{n}\left(p_1 \sum_{i}^{k} b(i)\left(1 - 2 \times \frac{t_i \cap y_i}{t_i + y_i}\right) + p_2 \sum_{j}^{n} (1 - b(j))\left(1 - 2 \times \frac{t_j \cap y_j}{t_j + y_j}\right)\right) \quad (3)$$

Boundary (BD) loss, a relatively new type of loss function, aims to minimize the distance between the ground truth and predicted segmentation. In particular, for addressing highly unbalanced segmentation, the loss function uses integrals over the boundary between regions of interest, rather than using unbalanced integrals over the regions [12], [20]. As with other standard loss functions, the BA loss function can be easily combined with the BD loss function (e.g., L_{BD+BA})—see (4).

$$L_{BD+BA} = -\frac{1}{n}\left(p_1 \int b(q)\Phi_G\,(q)\,s_\theta\,(q)\,dq + p_2 \int (1 - b\,(q))\Phi_G\,(q)\,s_\theta\,(q)\,dq\right) \quad (4)$$

In Equation (4), $\Phi_G(\cdot)$ is the level set representation of the boundary that evaluates the distance between a pixel on a model prediction and the nearest pixel in the boundary area of the target object, whereas $s_\theta\,(\cdot)$ is the softmax probability outputs of the deep learning model. More details can be found in [20].

IV. DATASETS

A. POLYP DATASET

For the polyp datasets, the CVC-ClinicDB [25] and ETIS [26] datasets were used. CVC-ClinicDB is composed of 612 frames extracted from 29 colonoscopy images and videos. These frames contain several examples of polyps. The database was used in the training stages of the MICCAI2016 sub-challenge for automatic polyp detection based on colonoscopy. Similarly, the ETIS database consists of 196 frames extracted from colonoscopy videos.

B. SKIN LESION DATASET

For the skin lesion datasets, open PH2 [27] and ISBI [28] were used. The PH2 dataset, provided by a research team at Proto University in Portugal, is composed of 200 images, including 80 atypical and typical moles each, and 40 melanomas. The ISBI dataset was created for the detection of melanoma and is composed of 900 images in total. Because the skin lesion dataset is composed of skin images captured with a microscope, the skin areas have a circular shape, whereas the other parts are shown in black.

C. CHEST X-RAY DATASET

For the chest X-ray datasets, Montgomery [29] and Japanese Society of Radiological Technology (JSRT) [30] datasets were used. The Montgomery dataset, published by the Health Department of Montgomery County in the US, consists of 138 images, including 80 tubercular patient images and

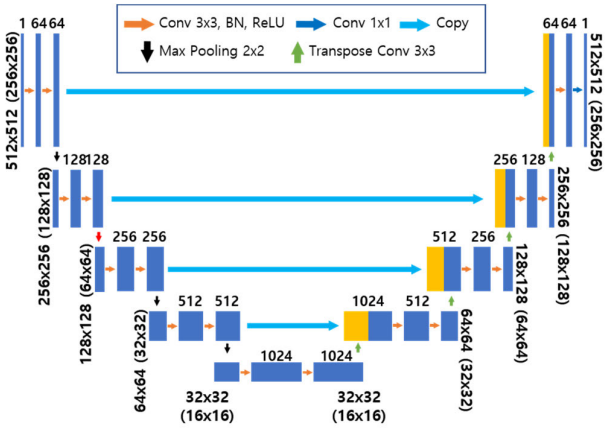

FIGURE 2. The segmentation deep learning model using the standard U-Net architecture. Numbers in parentheses represent the spatial resolutions of feature maps when images of 256 × 256 pixels are used as inputs.

58 normal images. The JSRT dataset is composed of 154 lung nodule patient images and 93 normal images.

V. EXPERIMENTAL EVALUATION

A. EXPERIMENTAL SETUP

To evaluate the effectiveness of the BA loss function, deep learning models for segmentation were built based on the standard U-Net architecture using ResNet-101 [31] as the backbone (see Fig. 2). U-Net is a representative encoder–decoder network that has proven to be efficient and stable in many biomedical image segmentation tasks such as brain image segmentation and liver image segmentation. The encoding part of U-Net is similar to that of the typical CNNs, where convolutional operations followed by a max-pool downsampling are applied to encode the input image into feature representation at multiple different levels. The decoding part consists of upsampling and concatenation followed by regular convolutional operations to semantically project the discriminative features learned by the encoder network onto pixel space to obtain a dense classification. The key aspect of U-Net is its skip connections between mirrored layers in both encoder and decoder networks for concatenating higher resolution feature maps from the encoder network with the upsampled feature map to transfer coarse global contextual information, thus enabling the recovery of local discriminative features attenuated during downsampling. For implementation, the convolutional block is composed of the repeated application of a 3×3 convolution with stride and padding of one, followed by batch normalization (BN) and a rectified linear unit (ReLU). For downsampling, a 2×2 maxpooling with stride of two is used, whereas a 3×3 transpose convolution with stride of two and no padding is used for upsampling at the decoder side.

All deep learning models in our study were implemented using the PyTorch [32] framework and ran on an NVIDIA TITAN-XP GPU. The whole network was trained end-to-end using a stochastic gradient descent optimizer [33] with a mini-batch size of eight and a momentum of 0.9. The base learning rate was set to 10^{-2} and then decreased by a factor of 10 when the validation set accuracy ceased to improve. The epoch size was set to 100. We applied early stopping to avoid overfitting. The boundary area (r) was set to two pixels, and the weights of the boundary area to five for p_1 and one for p_2. These values were empirically set. Our models were initialized with standard ResNet-101 pre-trained on the ImageNet [34] classification dataset. Eighty percent of the total data was used as training data while the remaining 20% was used as evaluation data. We conducted such split for all datasets. Prior to training and evaluation, colonoscopy skin images were resized to 256×256 pixels, whereas chest X-ray images were resized to 512×512 pixels.

For performance comparison, Dice coefficient, precision, and recall were used. The corresponding equations are presented as Equation (5), where TP, FP, and FN indicate the number of true positive, false positive, and false negative pixels, respectively.

$$\begin{aligned} Dice &= \frac{2TP}{TP + FP + FN} \\ Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \end{aligned} \tag{5}$$

To improve the validity of the proposed method, we applied a 5-fold cross validation on the dataset: for each fold, we measured the corresponding performance metrics. The final result was the averaged result incorporating the standard deviation of the five evaluations.

B. RESULTS BY DATASET

1) POLYP DATASET

Table 1 presents the results of the experiment conducted using the polyp dataset. The results show a poor segmentation performance in general because polyps are similar to wrinkles in the large intestine. However, when the proposed BA loss function was combined with the MSE, CE, and Dice loss functions, the Dice, precision, and recall values generally improved. The method of calculating the loss by dividing the boundary and the non-boundary areas, which is a characteristic of the BA loss function, can detect polyps effectively because of the difference created between them and the background. Figs. 3 shows an example of the polyp dataset, ground truths (GTs) indicating disease areas, and the segmentation results when each loss function was combined with BA loss. As shown in the figures, particularly when the CVC-ClinicDB dataset was used, the area predicted by the existing loss function is smaller than the area of the ground truth, or its shape is not similar to that of the ground truth. However, when the BA loss function was applied, the shape of the area predicted was similar to that of the ground truth, showing an improvement compared to when only the existing loss function was applied.

TABLE 1. Experimental results obtained using the polyp dataset.

Loss Function	CVC-ClinicDB dataset			ETIS dataset		
	Dice	Precision	Recall	Dice	Precision	Recall
MSE Loss	0.795±0.001	0.911±0.001	0.723±0.001	0.620±0.001	0.775±0.001	0.566±0.001
MSE+BA Loss	0.815±0.001	0.905±0.003	0.760±0.002	0.752±0.003	0.834±0.001	0.733±0.001
CE Loss	0.802±0.004	0.897±0.003	0.708±0.001	0.706±0.001	0.802±0.002	0.694±0.002
CE+BA Loss	0.816±0.003	0.896±0.001	0.711±0.003	0.728±0.001	0.838±0.001	0.736±0.001
Dice Loss	0.799±0.001	0.906±0.004	0.719±0.001	0.623±0.002	0.901±0.005	0.719±0.003
Dice+BA Loss	0.805±0.001	0.911±0.006	0.726±0.001	0.632±0.001	0.908±0.002	0.726±0.004
BD Loss	0.806±0.008	0.910±0.004	0.728±0.003	0.709±0.004	0.857±0.010	0.729±0.006
BD+BA Loss	0.834±0.002	0.907±0.009	0.751±0.005	0.730±0.002	0.872±0.005	0.731±0.006

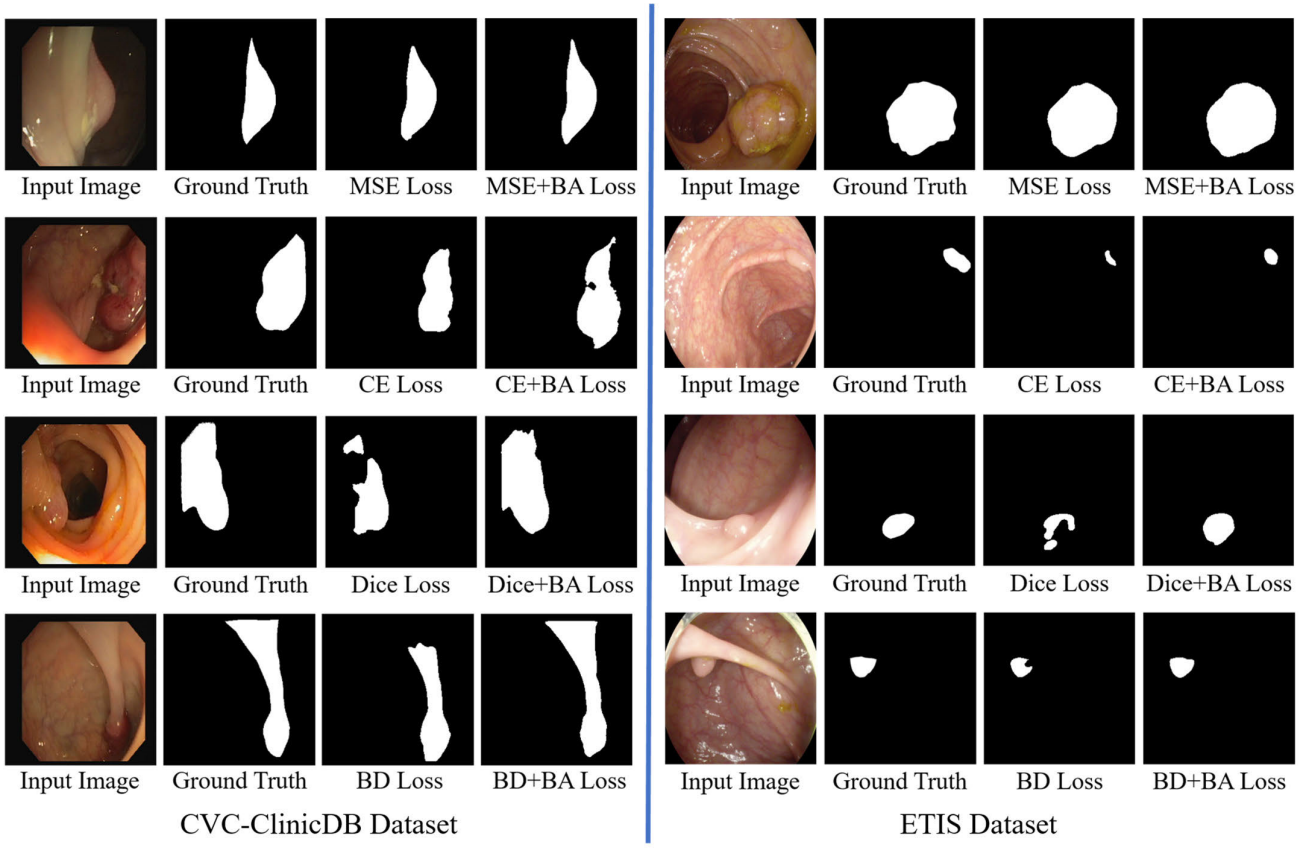

FIGURE 3. Polyp segmentation results obtained using the CVC-ClinicDB and ETIS datasets. For each dataset, the first column shows input images from the dataset and the second column includes the segmentation masks used as the ground truth. The third and fourth columns show the segmentation results for different loss functions.

2) SKIN LESION DATASET

Table 2 shows the experimental results obtained when the skin lesion dataset was used. It was observed that a relatively high segmentation performance was achieved compared to those with the polyp datasets because the differences in pixel values between the area affected by disease and the background are clear, which enables accurate prediction of the inner part of lesions. The Dice, precision, and recall values in Table 2 show a slight improvement in performance, except the MSE loss function for the ISBI dataset. However, many of the skin lesion images contain relatively coarse boundaries, as shown in Fig. 4, making it difficult for a neural network to accurately predict their boundaries. As a result, when the area predicted by the existing loss function is compared with that obtained using the same loss function combined with the BA loss function, it was observed that accurate prediction of the boundary was difficult to achieve and only a rough boundary was obtained in both cases.

3) CHEST X-RAY DATASET

Table 3 shows the result of the experiment conducted using the chest X-ray dataset. The results show high segmentation

TABLE 2. Experimental results obtained using the skin lesion dataset.

Loss Function	PH2 dataset			ISBI dataset		
	Dice	Precision	Recall	Dice	Precision	Recall
MSE Loss	0.934±0.003	0.949±0.001	0.930±0.002	0.891±0.001	0.893±0.001	0.952±0.002
MSE+BA Loss	0.934±0.000	0.948±0.001	0.930±0.001	0.843±0.000	0.860±0.001	0.859±0.001
CE Loss	0.857±0.000	0.923±0.001	0.898±0.000	0.801±0.002	0.858±0.002	0.838±0.003
CE+BA Loss	0.827±0.001	0.958±0.000	0.874±0.000	0.786±0.002	0.899±0.001	0.828±0.002
Dice Loss	0.940±0.000	0.964±0.000	0.924±0.001	0.884±0.000	0.894±0.001	0.903±0.000
Dice+BA Loss	0.952±0.000	0.961±0.000	0.934±0.000	0.895±0.000	0.904±0.000	0.913±0.001
BD Loss	0.945±0.003	0.955±0.004	0.902±0.008	0.903±0.002	0.899±0.004	0.926±0.005
BD+BA Loss	0.945±0.003	0.955±0.002	0.916±0.004	0.915±0.004	0.895±0.005	0.936±0.003

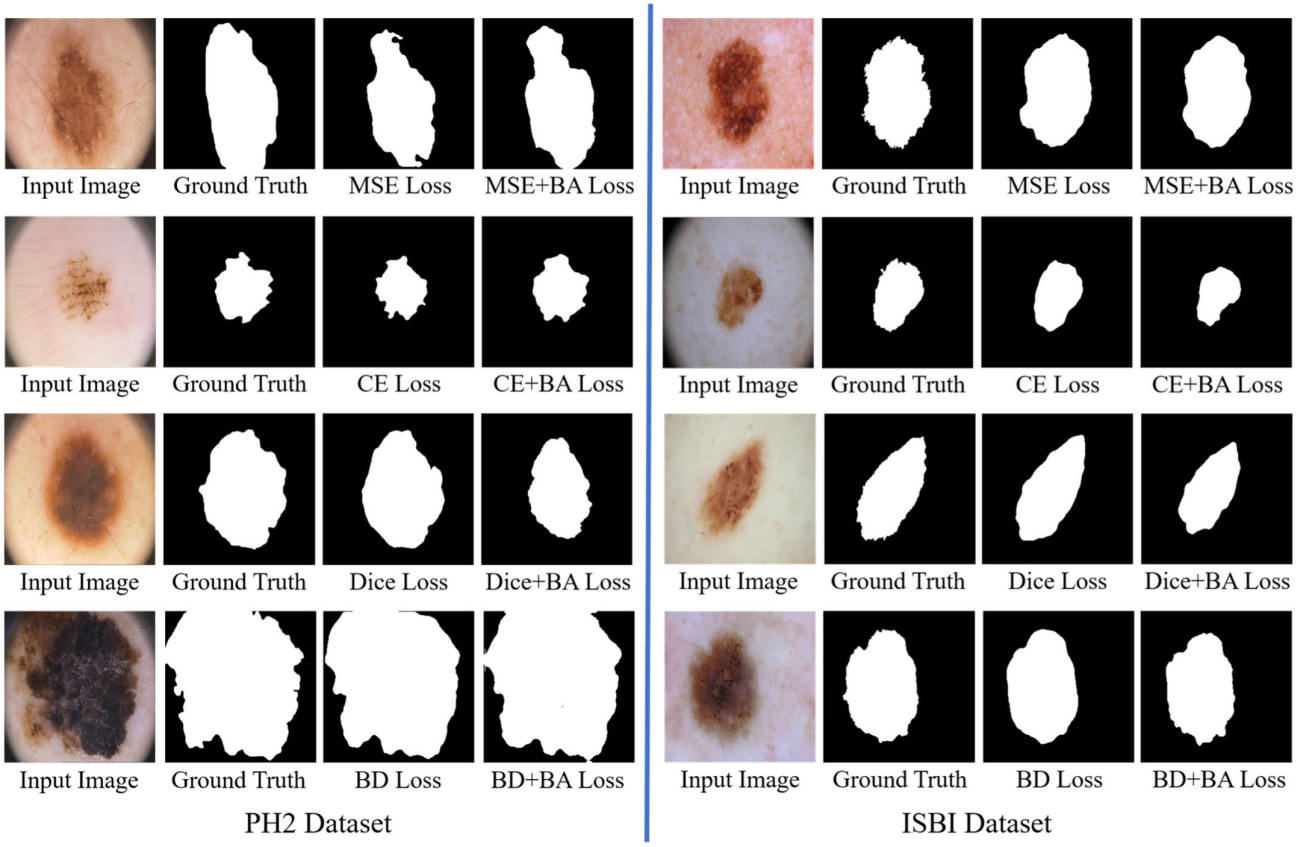

FIGURE 4. Skin lesion segmentation results obtained using the PH2 and ISBI datasets. For each dataset, the first column shows input images from the dataset and the second column includes the segmentation masks used as the ground truth. The third and fourth columns show the segmentation results for different loss functions.

performance for every loss function because it has a relatively clear boundary area compared to the colon polyp dataset or skin lesion dataset, which makes it relatively easy to distinguish the boundary area from the background. According to the experimental results shown in Table 3, although the existing loss functions combined with the BA loss function had slightly superior performance over the corresponding loss functions without the BA loss function, both types of loss functions already showed a very high level of segmentation accuracy and the differences between them were marginal. As shown in Fig. 5, both types of loss functions were able to accurately predict the boundaries of the left and right side of the thorax. However, they differ in segmentation results for the areas where the heart and the lungs are in contact and/or the diaphragmatic region, where the boundaries were predicted less accurately. We acknowledge that because of the diversity of the shape of the heart and the lungs, the segmentation accuracy was not improved considerably by the BA loss function alone, but it can improve if data on various lung areas are provided.

C. SENSITIVITY TO THE CONFIGURABLE PARAMETER

The range of the boundary area is a hyperparameter that controls where in the input image the deep learning models

TABLE 3. Experimental results obtained using the chest X-ray dataset.

Loss Function	MONTGOMERY dataset			JSRT dataset		
	Dice	Precision	Recall	Dice	Precision	Recall
MSE Loss	0.977±0.003	0.983±0.000	0.970±0.001	0.972±0.000	0.972±0.001	0.959±0.000
MSE+BA Loss	0.976±0.002	0.982±0.000	0.965±0.003	0.973±0.002	0.968±0.002	0.958±0.000
CE Loss	0.954±0.000	0.971±0.000	0.953±0.000	0.958±0.001	0.968±0.001	0.952±0.000
CE+BA Loss	0.962±0.000	0.975±0.001	0.959±0.000	0.954±0.002	0.966±0.000	0.945±0.000
Dice Loss	0.976±0.001	0.984±0.001	0.969±0.002	0.966±0.001	0.966±0.001	0.933±0.001
Dice+BA Loss	0.997±0.000	0.987±0.001	0.973±0.000	0.968±0.000	0.968±0.001	0.967±0.000
BD Loss	0.979±0.003	0.979±0.002	0.968±0.002	0.970±0.001	0.969±0.002	0.958±0.003
BD+BA Loss	0.977±0.003	0.980±0.002	0.970±0.004	0.971±0.001	0.970±0.002	0.968±0.001

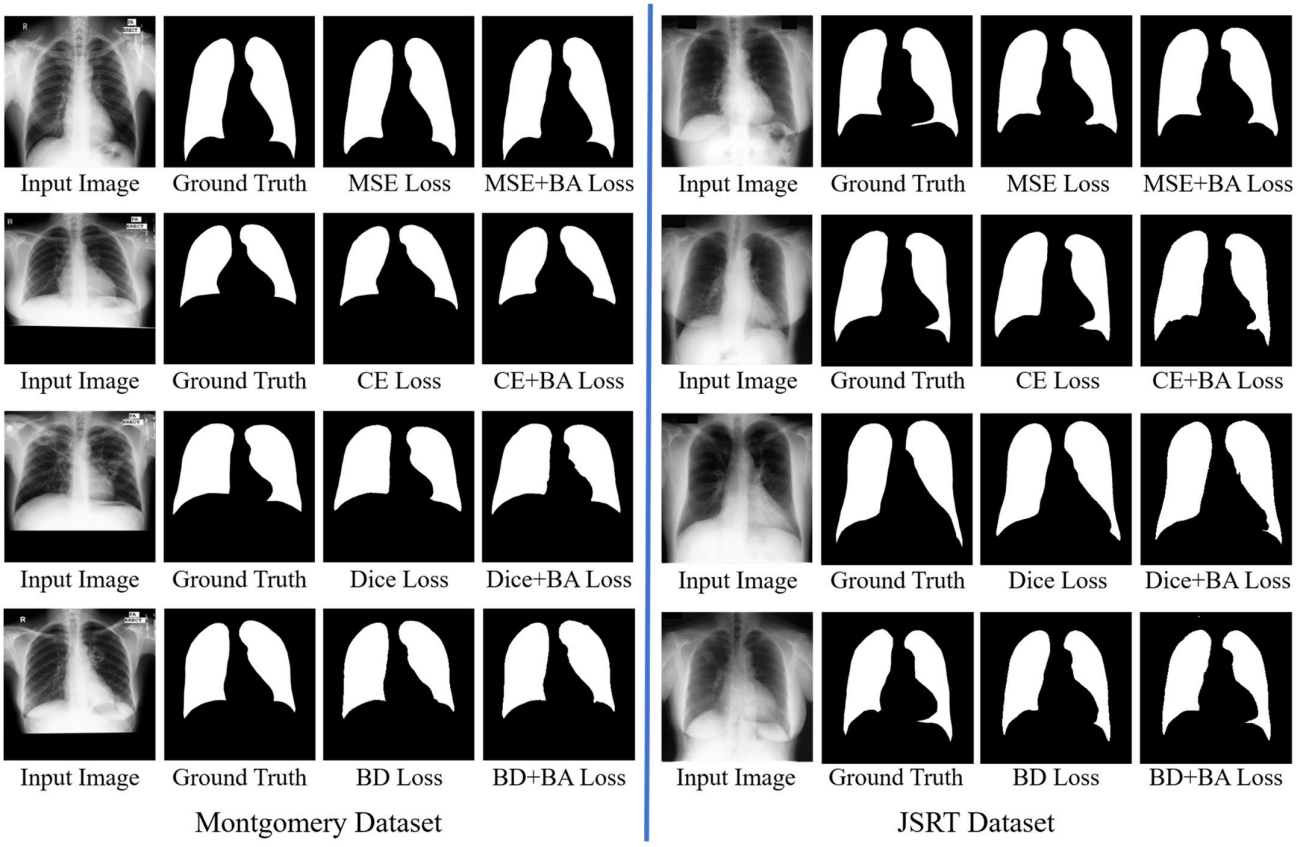

FIGURE 5. Lung segmentation results obtained using the Montgomery and JSRT datasets. For each dataset, the first column shows input images from the dataset and the second column includes the segmentation masks used as the ground truth. The third and fourth columns show the segmentation results for different loss functions.

should focus more; therefore, it can affect the overall performance of the deep learning models using the loss function. To explore the behavior of the BA loss function according to the hyperparameter, we conducted several experiments while changing the values of the range of the boundary area. For the experiments, the range of the boundary area was set to 1, 5, and 10. Note that the setting of various boundary area ranges is only for comparing the performance changes according to the range of the boundary area, and not for finding the optimal range of the boundary area in each dataset. Fig. 6 shows the Dice, precision, and recall performances for various combinations of the datasets (CVC-ClinicDB, ETIS, PH2, ISBI, Montgomery, and JSRT) and loss functions (MSE, CE, and Dice) used in this experiment. The experimental results show performance variations because each dataset has very different characteristics. In general, as shown in the figure for each dataset, if the range of the boundary is too small, there are fewer factors affecting the learning, making it difficult to achieve performance improvement. Conversely, if the range is too large, performance improvement is also difficult to achieve because the meaning of the boundary area disappears.

The other hyperparameters are the weights assigned to the loss of pixels for the boundary area and the

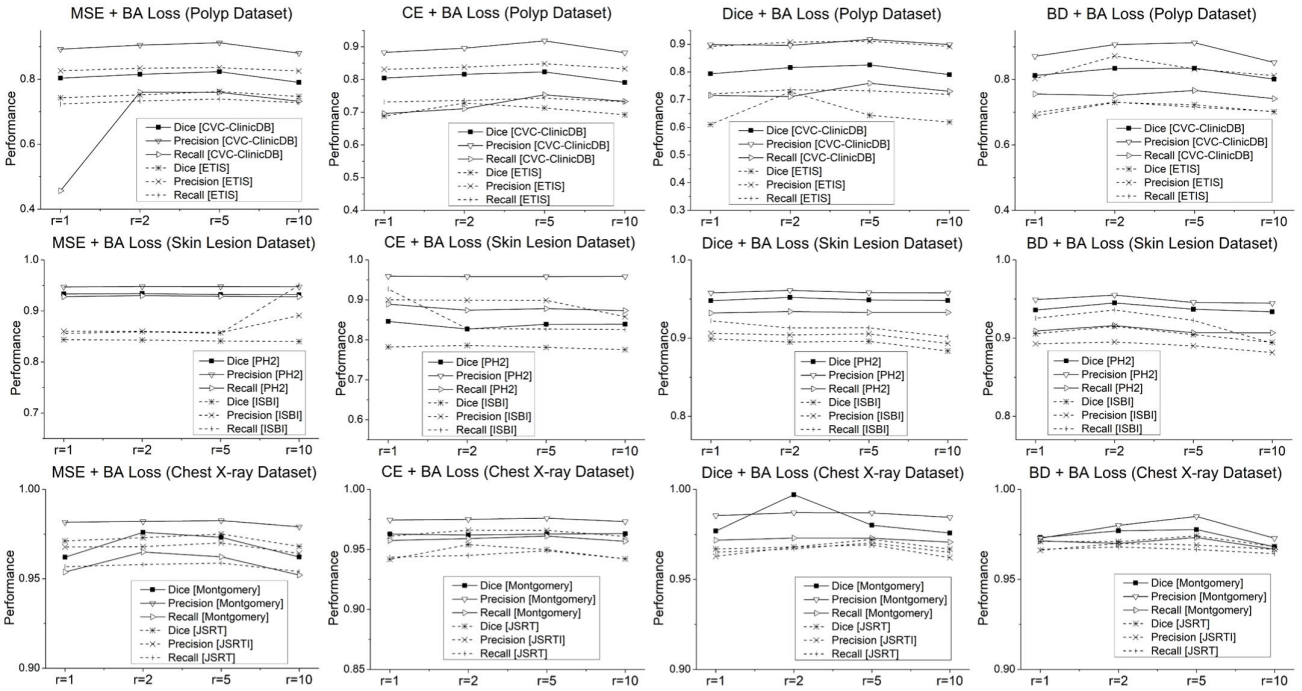

FIGURE 6. Sensitivity to the range of the boundary area.

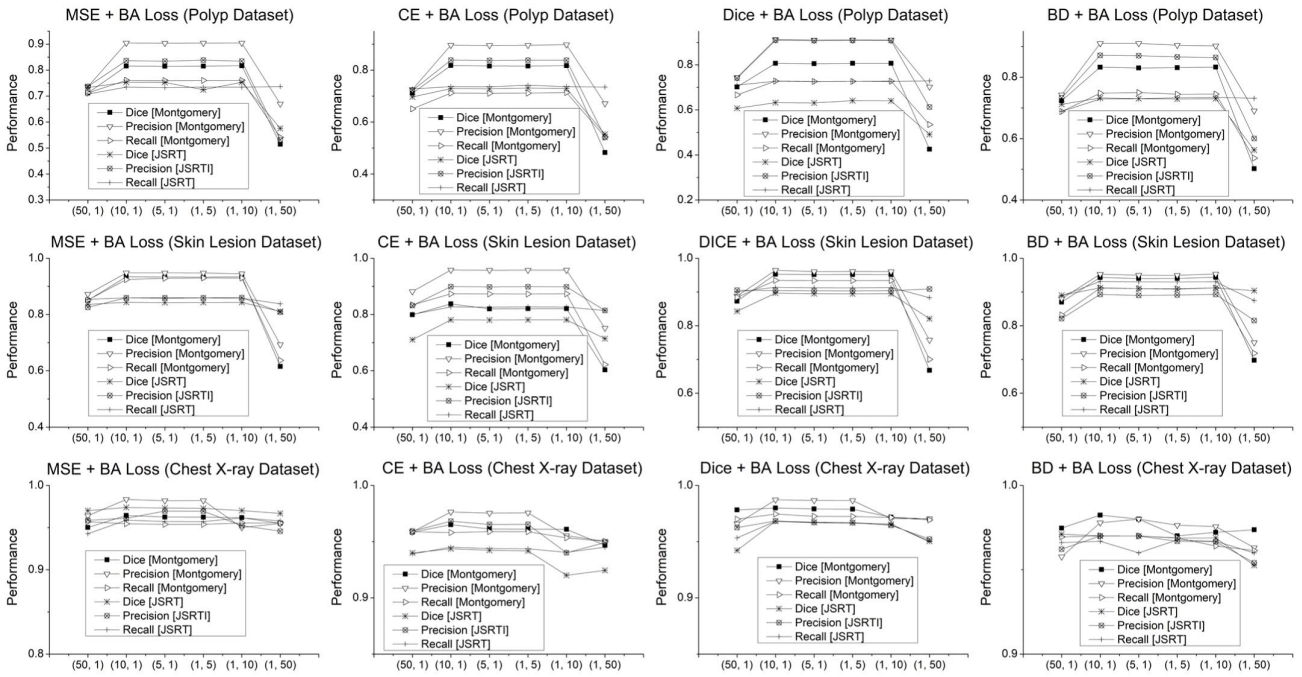

FIGURE 7. Sensitivity to the weights. X-axis represents the weights pair (p_1, p_2).

non-boundary area. For the experiments, with a fixed boundary range of two, we used six different weights pairs for (p_1, p_2) from (50, 1) to (1, 50). As shown in Fig. 7, the individual loss functions exhibited different performance patterns depending on the datasets used. However, in general, the larger the difference between two weights, the less likely it is to achieve better performance (see Fig. 8). In particular, when a higher weight is assigned to the loss of pixels in the boundary area, the boundary of the object of interest can be more accurately predicted. However, the downside of this

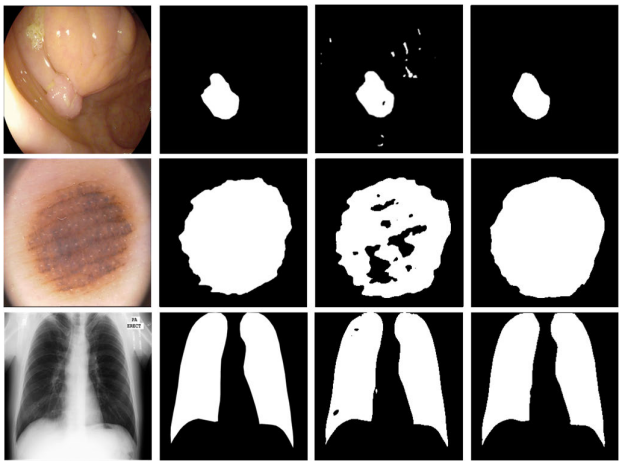

FIGURE 8. Examples of segmentation using different weight parameters. The first column shows input images from different datasets and the second column includes the segmentation masks used as the ground truth. The third and fourth columns show segmentation results for the $L_{Dice+BA}$ loss with ($p_1 = 50$, $p_2 = 1$) and ($p_1 = 1$, $p_2 = 50$), respectively.

is less-accurate predictions for the inner part of the target object to segment, resulting in holes in the inner part of the segmentation result. On the other hand, with a higher weight for the non-boundary area, the performance for the boundary extraction can degrade.

VI. CONCLUSION

This paper proposed a loss function for effective distinction of the boundary areas in medical images. The proposed loss function calculates the loss values separately for the boundary and non-boundary areas. Furthermore, it assigns different weights to the loss values of the two separated areas to promote effective learning of the neural network in the boundary area.

To verify the efficacy of the proposed loss function, experiments were conducted using three types of datasets: colon polyp dataset, skin lesion dataset, and chest X-ray dataset. The experimental results obtained using these public medical datasets showed that the existing loss functions, such as cross entropy loss and Dice loss, combined with the proposed boundary-aware loss function achieved comparable or better performance than those without the boundary-aware loss function depending on the dataset used.

There are several limitations in this study. The BA loss function assigned different weights to the boundary area and non-boundary area to force the deep learning models to learn from poorly segmented pixels in the boundary area. Therefore, if the boundaries are not definitive, as in the case of the skin lesions, the performance improvement may be marginal. The performance of the deep learning models using the BA loss function varied according to the weights and the range of the boundary area. Optimal values for these can vary depending on the applications, and they need to be determined empirically. The public medical datasets used for the experiments contain a relatively small number of samples compared to large-scale image datasets such as ImageNet. Therefore, experiments using a large number of training samples with various lesions may be necessary to further investigate the performance of the BA loss function. In future, we will investigate the possibility of extending the BA loss function to address the abovementioned limitations.

REFERENCES

[1] Z. Li, Z. Hou, Z. Hao, Y. An, S. Liang, B. Lu, and C. Chen, "Automatic cardiothoracic ratio calculation with deep learning," *IEEE Access*, vol. 7, pp. 37749–37756, 2019.

[2] A. Ari and D. Hanbay, "Deep learning based brain tumor classification and detection system," *Turkish J. Elect. Eng. Comput. Sci.*, vol. 26, no. 5, pp. 2275–2286, 2018.

[3] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," 2020, *arXiv:2001.05566*. [Online]. Available: http://arxiv.org/abs/2001.05566

[4] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[5] I. E. Sobel, "Camera methods and machine perception," Ph.D. dissertation, Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, 1970.

[6] R. B. Hegde, K. Prasad, H. Hebbar, and B. M. K. Singh, "Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images," *Biocybern. Biomed. Eng.*, vol. 39, no. 2, pp. 382–392, Apr. 2019.

[7] M. Lai, "Deep learning for medical image segmentation," 2015, *arXiv:1505.02000*. [Online]. Available: http://arxiv.org/abs/1505.02000

[8] B. Kayalibay, G. Jensen, and P. van der Smagt, "CNN-based segmentation of medical imaging data," 2017, *arXiv:1701.03056*. [Online]. Available: http://arxiv.org/abs/1701.03056

[9] C. Sammut and G. I. Webb, Eds. *Encyclopedia of Machine Learning*. Boston, MA, USA: Springer, 2011, doi: 10.1007/978-0-387-30164-8_528.

[10] S. Mannor, D. Peleg, and R. Rubinstein, "The cross entropy method for classification," in *Proc. 22nd Int. Conf. Mach. Learn. (ICML)*, 2005, pp. 561–568.

[11] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. Tempany, M. R. Kaus, S. J. Haker, W. M. Wells III, F. A. Jolesz, and R. Kikinis, "Statistical validation of image segmentation quality based on a spatial overlap index," *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, Feb. 2004.

[12] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel, "Loss Odyssey in medical image segmentation," *Med. Image Anal.*, vol. 71, Jul. 2021, Art. no. 102035.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Intervent*, 2015, pp. 234–241.

[14] D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen, and H. D. Johansen, "DoubleU-Net: A deep convolutional neural network for medical image segmentation," in *Proc. IEEE 33rd Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jul. 2020.

[15] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: http://arxiv.org/abs/1706.05587

[16] (Jun. 23, 2021). *2015 MICCAI Challenge*. [Online]. Available: http://www.imagenglab.com/wiki/mediawiki/index.php?title=2015_MICCAI_Challenge

[17] (Jun. 23, 2021). *2018 Data Science Bowl*. [Online]. Available: https://www.kaggle.com/c/data-science-bowl-2018

[18] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "PraNet: Parallel reverse attention network for polyp segmentation," 2020, *arXiv:2006.11392*. [Online]. Available: http://arxiv.org/abs/2006.11392

[19] Y. Wu, Y. Ma, J. Liu, J. Du, and L. Xing, "Self-attention convolutional neural network for improved MR image reconstruction," *Inf. Sci.*, vol. 490, pp. 317–328, Jul. 2019.

[20] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. B. Ayed, "Boundary loss for highly unbalanced segmentation," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101851.

[21] D. Karimi and S. E. Salcudean, "Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks," 2019, *arXiv:1904.10030*. [Online]. Available: http://arxiv.org/abs/1904.10030

[22] F. Caliva, C. Iriondo, A. M. Martinez, S. Majumdar, and V. Pedoia, "Distance map loss penalty term for semantic segmentation," 2019, *arXiv:1908.03679*. [Online]. Available: http://arxiv.org/abs/1908.03679

[23] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
[24] Y. Liu, H. Shi, H. Shen, Y. Si, X. Wang, and T. Mei, "A new dataset and boundary-attention semantic segmentation for face parsing," in *Proc. 34th AAAI Conf. Artif. Intell. (AAAI)*, 2020, pp. 11637–11644.
[25] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Comput. Med. Imag. Graph.*, vol. 43, pp. 99–111, Jul. 2015.
[26] J. Silva, A. Histace, O. Romain, X. Dray, and B. Granado, "Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 9, no. 2, pp. 283–293, Mar. 2014.
[27] (Jun. 23, 2021). *PH2 Dataset*. [Online]. Available: https://www.kaggle.com/sergeygoldyaev/ph2dataset
[28] (Jun. 23, 2021). *ISIC Challenge*. [Online]. Available: https://challenge.isic-archive.com/data
[29] S. Jaeger, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quant. Imag. Med. Surg.*, vol. 4, no. 6, pp. 475–477, Dec. 2014.
[30] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-I. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of Radiologists' detection of pulmonary nodules," *Amer. J. Roentgenol.*, vol. 174, no. 1, pp. 71–74, Jan. 2000.
[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Las Vegas, NV, USA, Jun. 2015, pp. 770–778.
[32] (Jun. 23, 2021). *PyTorch*. [Online]. Available: https://pytorch.org
[33] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*. [Online]. Available: http://arxiv.org/abs/1609.04747
[34] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, and A. C. Berg, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

MINKI KIM received the B.S. degree from the Department of Computer Science, Kyonggi University, South Korea, in 2020, where he is currently pursuing the M.S. degree with the Department of Computer Science. His current research interests include computer vision, deep learning, and medical image analysis.

BYOUNG-DAI LEE received the B.S. and M.S. degrees in computer science from Yonsei University, South Korea, in 1996 and 1998, respectively, and the Ph.D. degree in computer science and engineering from the University of Minnesota, Minneapolis, USA, in 2003. From 2003 to 2010, he worked with Samsung Electronics, Company Ltd., as a Senior Engineer. He is currently a Full Professor with the Division of Computer Science and Engineering, Kyonggi University, South Korea. His research interests include machine learning, deep learning, and medical image analysis.