IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 14, 2019, accepted December 30, 2019, date of publication January 10, 2020, date of current version January 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2965651

DA-Capnet: Dual Attention Deep Learning Based on U-Net for Nailfold Capillary Segmentation

YULI SUN HARIYANI[1,2], HEESANG EOM[1], AND CHEOLSOO PARK[1], (Member, IEEE)
[1]Department of Computer Engineering, Kwangwoon University, Seoul 01897, South Korea
[2]School of Applied Science, Telkom University, Bandung 40257, Indonesia

Corresponding author: Cheolsoo Park (parkcheolsoo@kw.ac.kr)

This work was supported in part by the Basic Science Research Programs through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT under Grant NRF-2017R1A5A1015596, in part by the Ministry of Education under Grant NRF-2017R1D1A1B03031485, and in part by the Research Grant of Kwangwoon University in 2019.

ABSTRACT Automatic nailfold capillary segmentation is a challenging task owing to noise and large variabilities in images caused by insufficient focusing and low visibility of the capillaries. This task can be useful to detect and estimate the severity of autoimmune diseases of connective tissues or learning the status of white blood cells based on the cells' blood flow on the nailfold capillary. Previous studies have addressed this task using manual, semi-automated, and automated segmentation method. However, further improvement is still required. With the recent progress of deep learning on medical imaging, we herein propose dual attention deep learning based on U-Net for nailfold capillary segmentation, named DA-CapNet. Our DA-CapNet improves the U-Net architecture by integrating a dual attention module that can capture a better representation of feature maps from input images. Furthermore, DA-CapNet is compared with three baselines: adaptive Gaussian algorithm, SegNet, the original U-Net. We experimentally demonstrate that our proposed method outperforms these baselines.

INDEX TERMS Nailfold capillary, segmentation, dual attention, deep learning

I. INTRODUCTION

Nailfold capillaroscopy is a non-invasive, inexpensive, and reproducible imaging technique to evaluate microcirculations under a nailfold, which is a small vessel under the nail. The nailfold capillary could provide the status of white blood cells based on the cells' blood flow in the microvessel using a light source of specific wavelength [1], [2]. This technique is typically used to monitor the microcirculations by analyzing the morphology of nailfold capillary such as shape, length, and width [3], [4]. The morphology feature could be used to detect and estimate the severity of autoimmune diseases of connective tissue such as systemic sclerosis. To observe the morphology of the nailfold capillary or to analyze white blood cells in the capillary, each of the capillary must be segmented as it is crucial for the analysis. However, nailfold capillary segmentation is a challenging task owing to sensitivity to external factors when capturing nailfold capillary images (e.g., air bubbles trapped in the oil, or reflection owing to light source attached to the microscopy); moreover, large variabilities in the images caused by insufficient focusing and low visibility of the capillaries render the segmentation of the entire capillary from the background difficult, as shown in Figure 1.

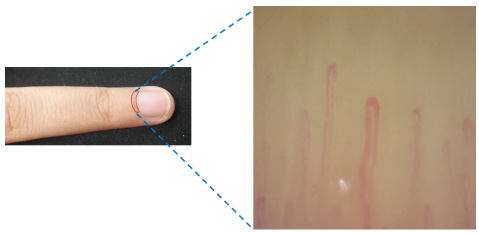


FIGURE 1. Nailfold capillary under the finger nail. The figure on the right shows the capillaries in red. As shown, the image lacks focus and contains light reflection as noise during the recording and scanning the finger nail.

In general, nailfold capillary segmentation has been addressed using three approaches: manual, semi-automated, and automated segmentation. Manual segmentation depends heavily upon human-recognizable features and requires experts to perform certain tasks, rendering it impractical for real applications [4]–[7]. Semi-automated segmentation

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

requires initial human intervention to mark the outer and inner parts of each capillary and requires considerable data analysis, which may cause bias and mistakes [1]. Isgro *et al.* [8] used an automated segmentation method by combining a local threshold and the Simultaneous Truth and Performance Level Estimation (STAPLE) algorithm to distinguish nailfold capillaries. However, the segmentation results were contaminated by noise and required post-processing such as morphological operations. To accelerate and facilitate nailfold capillary segmentation to benefit disease detection, it is necessary to develop an automated and precise method to segment nailfold capillary from images.

Recently, deep convolutional neural network (DCNN) methods have been proposed for semantic segmentation. Some of them are FCN [9], U-Net [10], DeconvNet [11], DeepLab [12], and SegNet [13]. Even though these methods have been effectively applied on medical image segmentation tasks, such as liver [14], pancreas [15], MRI [16], [17], and multiorgan [18], to the best of our knowledge, no DCNN method has been proposed for nailfold capillary segmentation. To perform nailfold capillary segmentation, two challenges must be addressed: (1) large variability in the images caused by insufficient focusing and low visibility of the capillaries complicating the segmentation of the entire capillary from the background (2) limited number of labeled training data, as previous studies on nailfold capillary used a private dataset. Therefore, DCNN methods that are proposed should be able to manage these two challenges. To address these issues and inspired by a large adaptation of U-Net and the attention mechanism [19], [20], we propose an improved U-Net using a dual attention module named DA-CapNet, which is designed to effectively extract contextual features of nailfold capillary from input images in an end-to-end manner. The integration of the dual attention module to U-Net allows for the optimization and performance improvement of its network.

In summary, our contributions are as follows:

1) We propose a dual attention module that combines the benefits of previous attention modules, squeeze-excitation (SE) [19] and convolution block attention module (CBAM) [20], which is then integrated into the U-Net architecture. Using the dual attention module can yield a better feature representation from the input image that has a large variability caused by insufficient focusing and low visibility of the capillaries.
2) We perform extensive experiments on a new collected nailfold capillary dataset and show that our method outperforms conventional methods and the original U-Net without an attention module.

The remainder of this paper is organized as follows: in Section II, related works are provided; in Section III, the proposed method and the baseline network used as the benchmark are described. The acquisition of input data, experimental settings, and results are addressed in Section IV. In Section V, the conclusions of this study are provided.

II. RELATED WORKS

A. DEEP-LEARNING-BASED SEGMENTATION

Deep learning has progressed rapidly and achieved state-of-the-art performance in many computer vision tasks, such as object detection, image classification, instance segmentation, semantic segmentation, image captioning, and object tracking [21]. Unlike traditional handcrafted methods, a data-driven deep-learning approach extracts discriminative features from the data itself, where these features contain different information at different levels of its networks [22], [23]. One of the most popular deep-learning models is based on a convolutional neural network (CNN), which could achieve a similar level performance to a human [24]. Therefore, it is not surprising that CNNs are currently often used in medical image processing, particularly for image segmentation tasks (e.g., pancreas [15], liver [14], MRI [16], [17], and multiorgan [18]). The first work involved U-Net, which produced the best accuracy and won the ISBI challenge 2015 for the segmentation of neuronal structures in electron microscopic stacks [25]. Other variants of CNN have achieved state-of-the-art performances on benchmark semantic segmentation tasks, such as SegNet [13], DeepLab [12], and DeconvNet [11]. Among these methods, U-Net is the most widely used architecture in medical image processing owing to its simplicity in building encoder and decoder paths with skip connections, affording efficient information flow and excellent performances as it is able to delineate complex-shaped objects well in biomedical images [25]. U-Net also resulted in an outstanding performance when the training process was combined with an excessive data augmentation and weighted loss [10]. Therefore, our study was motivated by U-Net with modifying and integrating several additional attention modules in the decoder path to learning better discriminative features from the input dataset than the original U-Net structure.

B. CAPILLARY SEGMENTATION

Most studies regarding nailfold capillary have focused on pattern description and classification; studies regarding nailfold capillary segmentation are scarce. Paradowski *et al.* [26] proposed an automated technique to detect an avascular area from a nailfold capillary image. Jones *et al.* [27] proposed a classification method based on morphological features and established a high correlation between morphology and disease. Nivedha *et al.* [28] and Suma *et al.* [29] employed a nonlinear support vector machine and fuzzy logic kernels for nailfold capillary image classification to identify healthy, hypertensive, and diabetic patients. In general, nailfold capillary classification is used to identify diseases such as scleroderma, Raynaud's phenomenon, systemic lupus erythematosus, and progress systemic sclerosis. Additionally, Tama *et al.* [30] used a binarization algorithm to conduct nailfold capillary segmentation. They specifically used images from recorded nailfold microcirculation videos as input. Bourquard *et al.* [1] utilized a semi-manual method

to segment capillaries; however, the expert should have marked the outer and inner parts of each capillary repetitively that may lead to concentration loss and, thus, mislabeling. Semi-manual methods are also highly dependent on the physician's experience. Isgro *et al.* [8] used a combination of local threshold and the STAPLE algorithm to distinguish nailfolds. However, the segmentation results were still contaminated with noise and required post-processing, such as morphological operations. Suma *et al.* [31] classify nailfold capillary to diagnose vascular dysfunction in the Indian Population using several machine learning algorithms such as Logistic Regression, CNN, and Random Forest.

C. ATTENTION MECHANISM

Attention mechanism is critical in human perception. Humans do not proceed a whole scene at once but sequentially exploit a partial glance of a scene and carefully focus only on salient portions to visually capture a better understanding. Hence, several studies have been performed to insert the attention module on the CNN architecture; this has been proven to improve the performances in large-scale image classification tasks [19], [20], [32], [33]. Wang *et al.* [32] proposed the residual attention network, which employs an attention module based on an encoder–decoder structure. By obtained refined feature maps, the network could not only perform better but was also more robust to noisy input images compared with the original CNN. Hu *et al.* [19] exploited an inter-channel relationship for an attention module named the squeeze-and-excitation (SE) module. Global pooling feature maps are used to compute a channel-wise attention part. Unlike SE, Woo *et al.* [20] proposed the convolutional block attention module (CBAM) that exploits both spatial and channel-wise attention. Despite these improvements, only a handful of studies used attention mechanisms in medical image classification tasks. In this study, to learn a better representation of feature maps compared with using the conventional method, we developed a new attention module named the dual attention module by utilizing the advantages of SE and CBAM.

III. PROPOSED METHOD

In this section, the proposed method including the baseline network U-Net is described: the dual attention mechanism and dual attention U-Net for nailfold capillary segmentation, named DA-CapNet.

A. U-NET

We first describe U-Net [10], a baseline network for our proposed method. U-Net is a convolutional-based deep-learning architecture that yields high performance due to its network combined with the training strategy using excessive data augmentation. Owing to its capability, U-Net has become a popular deep-learning method in medical image segmentation. It is named U-net owing to its symmetric U-shape, which comprises encoder and decoder paths. The encoder is similar to the typical CNN; it downsamples the dimension of spatial information progressively while simultaneously increasing the number of channels per layer. Meanwhile, every step in the decoder upsamples the feature maps followed by a convolution layer, thus increasing the spatial information dimension of the output feature maps. To produce better localization, U-Net uses skip connections at every level of the decoder by concatenating the output of the upsampling layers with the feature maps of the encoder at the same level.

B. DUAL ATTENTION MODULE

The attention mechanism proposed herein is defined as a dual attention module inspired by the SE [19] and CBAM modules [20]. The benefits of SE and CBAM are combined to better capture contextual information for segmentation compared with the original U-Net. Details of the dual attention module are shown in Fig.2b. The first building block is SE, which comprises squeeze and excitation processes. In the squeeze process, local descriptors that statistically convey the entire image using global average pooling (GAP) is produced. Meanwhile, the excitation operator maps those descriptors to a set of channel weights. Here, SE assigns adaptive weights to each channel when creating the output feature maps to focus on feature maps that provide a meaningful impact based on the GAP output. The second building block is CBAM, which highlights significant features in addition to channels and spatial dimensions. Here, CBAM is used to sequentially highlight channel and attention modules as it can learn what object to focus and where the important position and location from the input image. Therefore, both attention modules in CBAM are able to learn efficiently which information or path of the image is to be highlighted or suppressed.

Given an intermediate feature maps F as the input, our dual attention module produces refined feature maps as F', where $F \in \mathbb{R}^{H\times W\times C}$ and $F' \in \mathbb{R}^{H\times W\times C}$. The overall dual attention process can be summarized as follows:

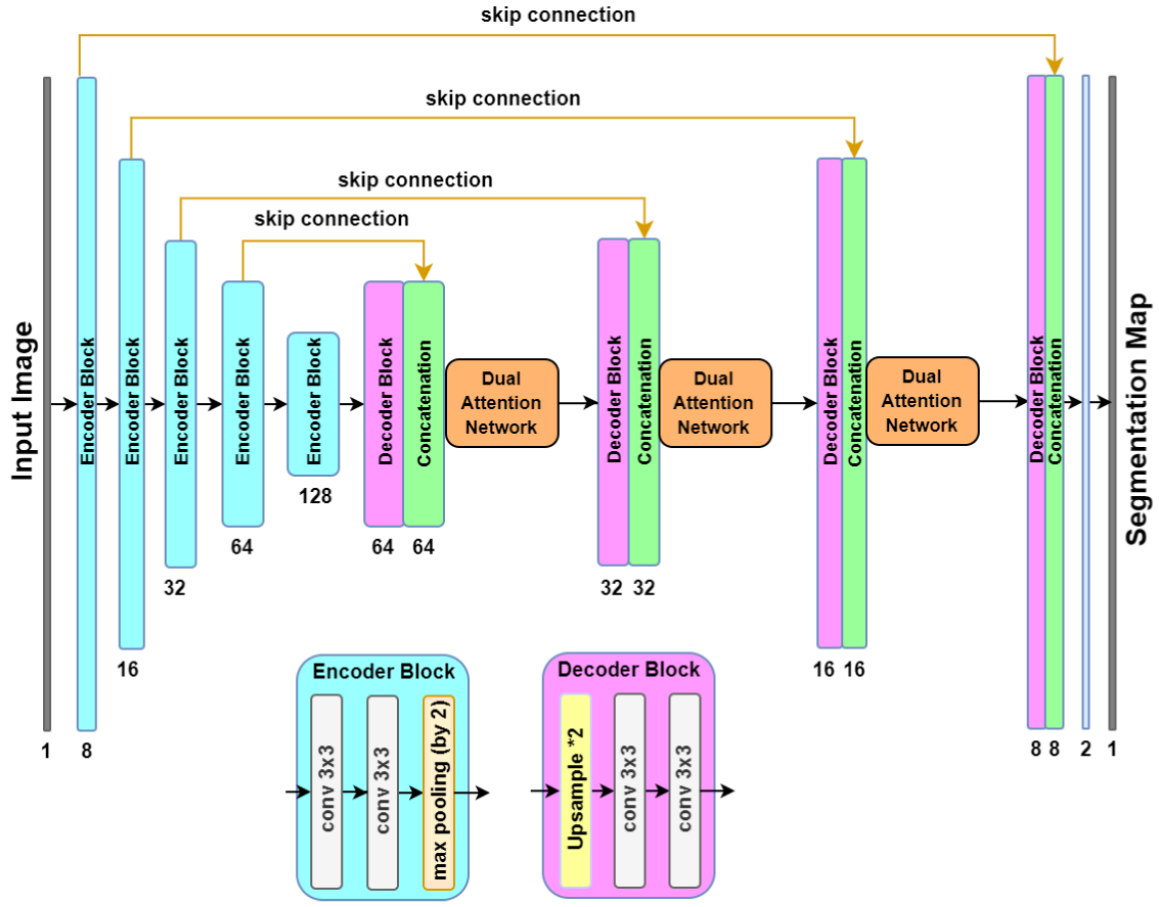
$$F' = F_{SE} \oplus F_{CBAM}, \tag{1}$$

where F_{SE} and F_{CBAM} are the output feature maps of the SE and CBAM modules, respectively; F' is the element-wise addition of these two output feature maps. The process of F_{SE} is formulated as follows:

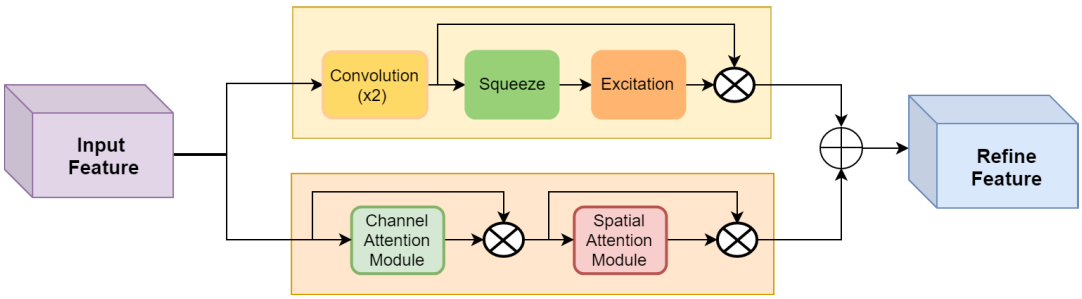
$$F_{SE} = Scaling(M_e(M_s(conv(F)))) \otimes conv(F), \tag{2}$$
$$M_s = GAP(conv(F)), \tag{3}$$
$$M_e = \sigma(MLP(M_s)), \tag{4}$$

where M_s, M_e, GAP, and MLP denote the feature maps of the squeeze module and excitation module, global average pooling, and multilayer perceptron, respectively. Furthermore, we produce F_{CBAM} by sequentially calculating the channel feature maps and spatial feature maps. The process is calculated as follows:

$$F_{CBAM} = M_{sp}(M_c(F) \otimes F) \otimes (M_c(F) \otimes F), \tag{5}$$
$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))), \tag{6}$$
$$M_{sp} = \sigma(conv([AvgPool(F'); MaxPool(F')])), \tag{7}$$



(a) Network architecture of DA-CapNet.



(b) Dual attention module.

FIGURE 2. Overview of the proposed network for nailfold capillary segmentation. (a) Network architecture of DA-CapNet. The network was modified from the original U-Net which consists of smaller number of output channels in its layer. In decoder part, the network employed several dual-attention modules. (b) Dual attention module. The attention mechanism proposed herein is defined as a dual attention module inspired by the SE [19] and CBAM modules [20].

where $M_c(F)$ and M_{sp} are the output feature maps of the channel and spatial attention modules, respectively.

C. DA-CAPNET

The proposed attention mechanism is incorporated with the U-Net architecture to exploit local information. The complete architecture of the dual-attention-based U-Net proposed herein is shown in Fig. 2a. The network was modified from the original U-Net, which is smaller compared with the original U-Net. It consists of an encoder path on the left side, bottleneck layer in the middle, and decoder path on the right side. In the encoder path, the DA-CapNet receives a single channel input of size 256 × 256. The network consists of four encoder blocks, in which each block extracts the feature maps using two convolutional layers with a pooling layer downsampling the feature maps and then passes them down

(a) (b) (c)

FIGURE 3. The procedure of data acquisition process. (a) The image of nailfold capillary area (indicated by blue area). (b) Recording capillary using dino-lite microscope. (c) An image of nailfold capillary.

to the next block. Each convolutional layer has a filter size of 3×3 followed by the Rectified Linear Unit (ReLU) activation function and each the pooling layer has the size of 2×2. In the fourth encoder block, a dropout layer is added for regularization to avoid overfitting. We set 8, 16, 32, and 64 as the numbers of output channel for the first, second, third, and fourth block, respectively. In between the encoder and decoder parts, one block serves as a bottleneck layer that has two 3×3 convolutional layers followed by a dropout layer. The dimension of the output channel in this block is set to 128. The decoder part consists of four decoder block. Each decoder block has an upsampling layer with stride 2 to produce a segmentation mask of the same size as the input image. Each layer in the decoder is concatenated with the corresponding crop features from the respective block in the encoder to obtain more contextual information and reconstruct a precise location of the segmentation map. The concatenating process called skip connection was followed by convolutional layers comprising 3×3 kernels.

As in Fig. 2a, several dual attention modules are added in the decoder part. Each module obtains the input feature from its previous decoder blocks. The details of dual attention modules are shown in Fig. 2b. The upper part of the dual attention module contains a convolution layer and squeeze and excitation blocks. The output of the convolution layer is multiplied with the output feature from squeeze and excitation processes. The lower part of the dual attention module consists of channel and spatial attention blocks in series, whereby performance based on [20] is expected. Those two parts are processed in parallel and added up at the end of the module to obtain refined features.

D. OPTIMIZATION

The weights are learned by minimizing the loss function in Eq. 8. Nailfold capillary image segmentation is considered as a pixel classification problem. In medical image segmentation, an imbalanced sample number of classes between background and foreground images often cause the learning process to be trapped in the local minima of the loss function, resulting in a network whose predictions are heavily biased toward the background. According to Jaeger *et al.* [34], a loss function based on jointly learning the binary cross-entropy (BCE) and dice coefficient are employed. We directly use the predicted probabilities instead of thresholding and converting them into a binary mask. The dice coefficient could effectively manage the imbalanced classes between the background and foreground in pixel-wise segmentation. Our loss function L_{total} is formulated as follows:

$$L_{total} = 0.5 * BCE(y, \hat{y}) - DiceCoefficient(y, \hat{y}), \quad (8)$$

$$BCE(y, \hat{y}) = -(y \log(\hat{y}) + (1-y) log(1-\hat{y})), \quad (9)$$

$$DiceCoefficient(y, \hat{y}) = \frac{2 \sum_{i=1}^{N} y_i \hat{y}_i}{\sum_{i=1}^{N} y_i^2 + \sum_{i=1}^{N} \hat{y}_i^2}, \quad (10)$$

where y is the ground truth label, $\hat{y}$ prediction label, and N number of pixels.

IV. EXPERIMENT SETTINGS

This section first describes the dataset for the experiments, pre-processing, and data augmentation to enlarge our dataset, experimental setting, and evaluation metrics. Subsequently, we compare our proposed method with the baseline method, which is a handcrafted method using the adaptive Gaussian threshold. We compare our proposed method with U-Net using different attention modules.

A. DATASET

The dataset was recorded from seven healthy participants of age 20–35. Permission to perform experiments involving human participants was obtained from the Korean IRB (P01-201903-11-002). The overall procedure of data acquisition is described in Fig. 3. The image of the nailfold capillary area (indicated by blue area) is captured using dino-lite microscope. Capillaroscopy videos were recorded from the third and fourth finger of the participants from their non-dominant hand using a handheld digital microscope designed for nail microcirculation analysis with a 500x magnification

rate. Following [3], the recording process was conducted as follows:

1) Participants were seated at in a room (of temperature 20–25 degrees Celsius) and were given time to relax and adapt to the new environment.
2) The fingers to be recorded were cleaned.
3) On each observed finger, a drop of vegetable oil was deposited to render the skin more transparent such that capillaries become more visible.
4) The video of each finger was recorded within 1–2 min. The target area of the nailfold capillary was marked in blue, as shown in Fig. 3(a).

Using this procedure, we build out dataset by extracting images from the videos. A total of 40 images, 30 for training, and 10 for testing were prepared. The ground truth was created manually by a human expert using annotation tools. Three salient capillaries were selected and annotated in each image.

B. IMPLEMENTATION DETAILS

The original images contain channels of red, green, and blue. In this study, we used only the green channel, as it reveals more information related to the nailfold capillary compared with the red and blue channels [28]. The images were then resized to 256×256 to reduce the processing time. In the adaptive Gaussian threshold method, a median blurring filter of kernel size five was added to smooth the input image and reduce ambient noise. To generate more training data and avoid overfitting for further improving the performance, an augmentation technique was applied. Specifically, the resized images were rotated, shifted, and zoomed-in to produce 3000 training data with the respective ground truth. Our network architecture was trained with random samples of batch size 2. The training was conducted using the ADAM optimizer with an initial learning rate of 10^{-4} and a weight decay of 10^{-5}. The training process was implemented using the Keras library on Tensorflow, on a computer equipped with NVIDIA GeForce GTX 1080.

C. EVALUATION METRICS

As the segmented area is small compared to the background, a general accuracy formula cannot be used, which might provide a false interpretation owing to data imbalance. Therefore, an intersection over union (IoU) known as the Jaccard index, which is a statistic measure to verify the similarity and diversity of sample sets, was used. As shown in Eq. (11) IoU measures the similarity of the finite sample sets and is defined as the intersection size divided by the size of the sample sets union. IoU is useful when we have imbalanced numbers of pixels within an image, as it provides the same weight to all classes. The precision and recall in Eq. (12) and (13) were calculated as the evaluation metrics. Precision adequately describes the purity of positive detection relative to the ground truth. Precision calculates the number of pixels within an object are predicted as a matching ground truth annotation. Meanwhile, recall effectively described the completeness of positive predictions relative to the ground truth,

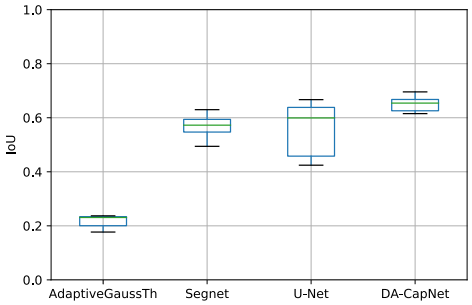


FIGURE 4. Quantitative segmentation performance is presented as box plots. The y axis denotes the IoU scores, while the x axis represents different segmentation methods. The median of DA-CapNet is the highest among those of other methods with small variance.

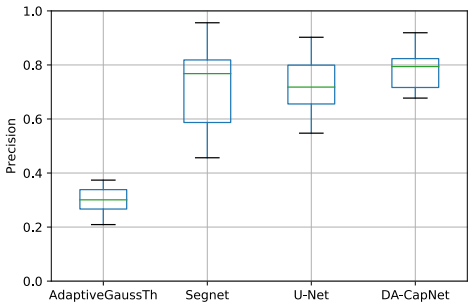


FIGURE 5. Quantitative segmentation performance is presented as box plots. The y axis indicates the precision scores, while the x axis represents different segmentation methods. The median of DA-CapNet is the highest among those of other methods.

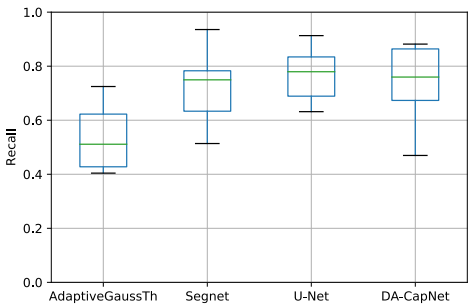


FIGURE 6. Quantitative segmentation performance is presented as box plots. The y axis indicates the recall scores, while the x axis shows the segmentation methods. The median of DA-CapNet is the highest among those of the other methods.

$$IoU\ (A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}, \quad (11)$$

$$Precision = \frac{TP}{TP + FP}, \quad (12)$$

$$Recall = \frac{TP}{TP + FN}, \quad (13)$$

where TP is true positive, FP is false positive, and FN is false negative.

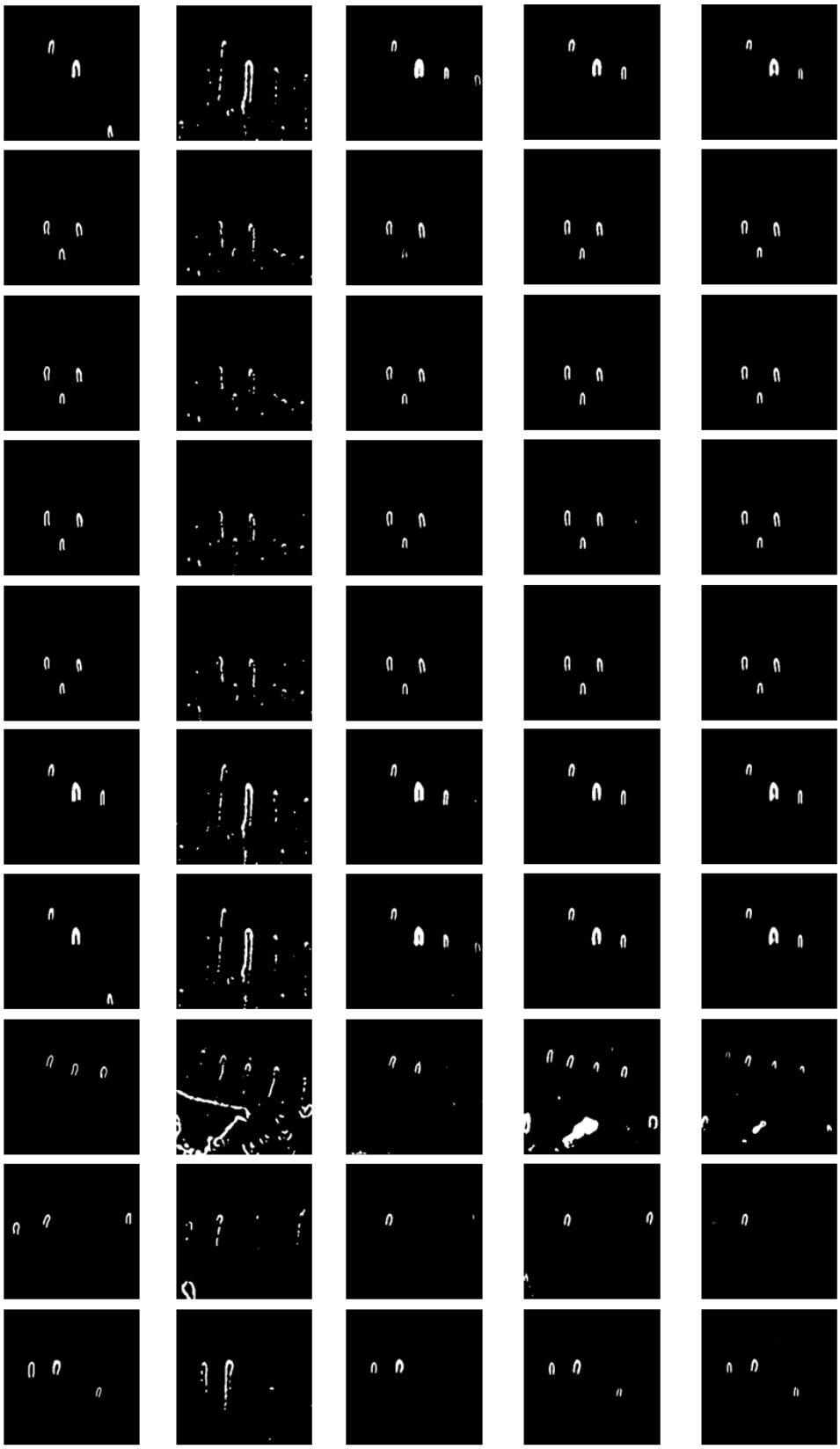


FIGURE 7. Quantitative results. We compare segmented results using DA-CapNet against other method using image 1–5 in test data. (a) Ground truth. (b) Adaptive gaussian threshold. (c) SegNet. (d) U-Net (e) DA-CapNet.

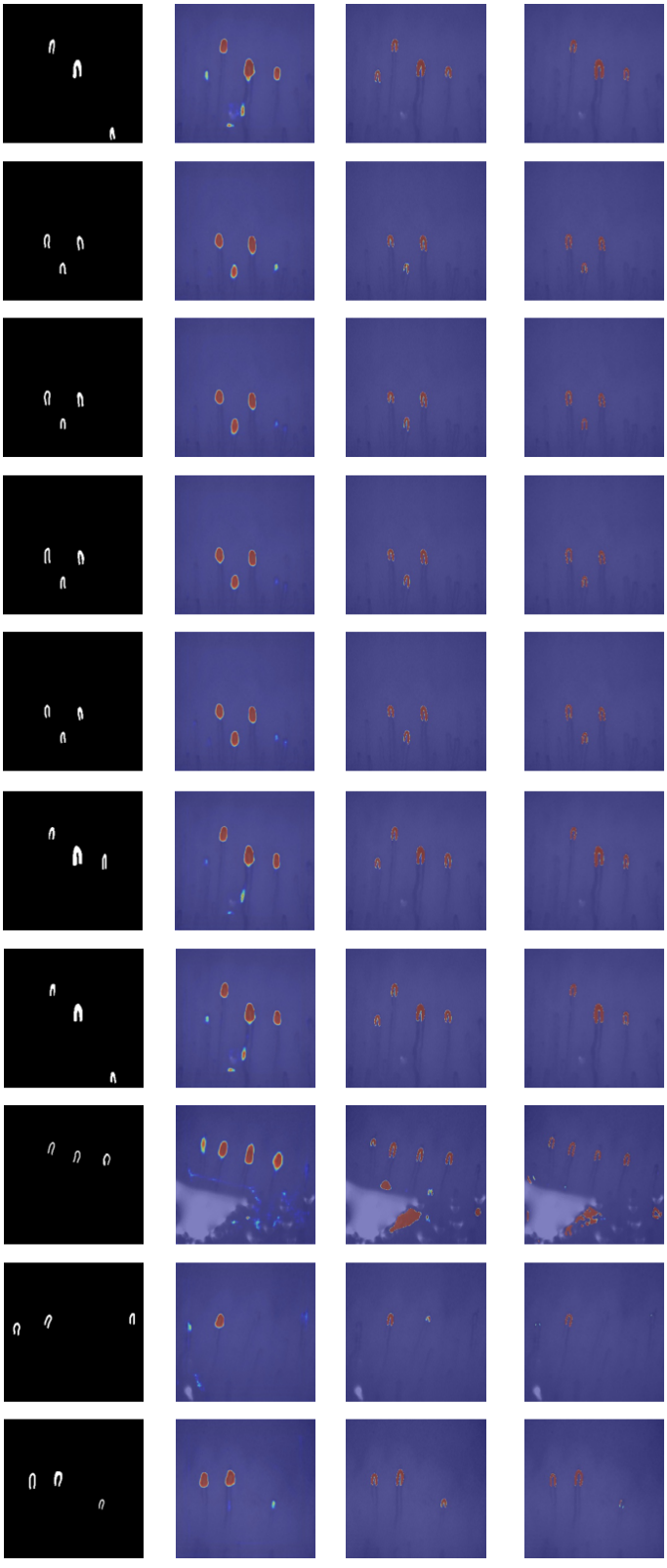


FIGURE 8. Grad-CAM visualization. We compare Grad-CAM [35] visualization among all methods on all test samples. Grad-CAM visualizes the importance of the spatial location information in the convolution layers using the gradients. (a) Ground truth. (b) SegNet. (c) U-Net (d) DA-CapNet.

V. RESULTS AND DISCUSSION

A. RESULTS

Our proposed method was evaluated using the collected dataset in terms of the metrics mentioned in Section 3C. The three baselines used in this study are the adaptive Gaussian threshold, SegNet, and the original U-Net. First, all results are presented in terms of the IoU or the Jaccard index, as shown in Fig. 4. From the whisker diagram, the DA-CapNet outperforms the other methods. DA-CapNet resulted in an increase of 0.43 in IoU scores over the adaptive Gaussian threshold, 0.09 over the original U-Net, and SegNet. The overall performance in terms of IoU is 0.218, 0.556, 0.559, and 0.6431 for adaptive Gaussian threshold, SegNet, original U-Net, and DA-CapNet, respectively.

Furthermore, we compared the performances of DA-CapNet against adaptive Gaussian threshold, SegNet, and original U-Net in terms of precision. Fig. 5 shows that our DA-CapNet outperforms the other methods significantly. Finally, the recall scores of DA-CapNet, adaptive Gaussian threshold, SegNet, and original U-Net are shown in Fig. 6, where the best result performance is shown by DA-CapNet. Subsequently, we compared the performances between the original U-Net and DA-CapNet. Ground truth images and their segmentation results using original U-Net and DA-CapNet are shown in Table 1. All these experiments with three evaluation metrics demonstrated that DA-CapNet achieved the best performance compared with the other methods. By adding the dual attention, it increases the network parameters from original U-Net with 485,813 parameters to 489,047 parameters in DA-CapNet. Despite this slightly increases parameter, we believe that our dual attention module helps to improve the performance results.

In addition to the quantitative results, we provide the qualitative results on the segmented images using our method compared with the adaptive Gaussian threshold, SegNet, and original U-Net, as shown in Fig. 7. The adaptive Gaussian threshold produces the worst performance, particularly in the eight-row image of Fig. 7b was owing to ambient noise. The SegNet and original U-Net could segment the nailfold capillary better than the adaptive Gaussian threshold. However, the segmented images by SegNet and original U-Net are not as accurate as the images produced by DA-CapNet. Following these results, we apply the Grad-CAM [35] to visualize SegNet, original U-Net, and DA-CapNet using images from the test set. Therefore, we can visualize the importance of the spatial location information in the convolution layers using the gradients. The Grad-CAM visualization can be found in Fig. 8.

B. DISCUSSION

In this subsection, we further discuss the dual attention model used in our proposed method. An experiment was conducted to demonstrate the effectiveness of the proposed dual attention module for the improvement of feature representation. The performance of DA-CapNet was compared with that of the original U-Net integrated with only the SE or CBAM module. It is noteworthy that these three models have the same underlying architecture except the attention module and were trained using the same setting for a fair comparison. Experimental results with different attention modules are shown in Table 2. The U-Net + CBAM module performed better compared with the U-Net + SE module by 0.06 points in terms of mean IoU. However, DA-CapNet performed even better by 0.08 points and 0.02 points compared with the U-Net + SE and U-Net + CBAM modules, respectively. The proposed DA-CapNet is the most effective segmentation approach, defeating U-NET with a single SE module or CBAM module without additional complicated modules.

TABLE 1. Comparison results between the original U-net and DA-CapNet.

Name	Method	IoU	Precision	Recall
Image 1	U-Net	0.52	0.65	0.72
	DA-CapNet	0.56	0.72	0.72
Image 2	U-Net	0.57	0.75	0.83
	DA-CapNet	0.63	0.82	0.87
Image 3	U-Net	0.63	0.84	0.91
	DA-CapNet	0.67	0.87	0.88
Image 4	U-Net	0.64	0.79	0.84
	DA-CapNet	0.67	0.82	0.79
Image 5	U-Net	0.63	0.69	0.83
	DA-CapNet	0.66	0.72	0.85
Image 6	U-Net	0.67	0.90	0.88
	DA-CapNet	0.7	0.92	0.88
Image 7	U-Net	0.64	0.66	0.73
	DA-CapNet	0.67	0.68	0.73
Image 8	U-Net	0.43	0.08	0.77
	DA-CapNet	0.65	0.56	0.66
Image 9	U-Net	0.42	0.55	0.51
	DA-CapNet	0.63	0.77	0.47
Image 10	U-Net	0.44	0.80	0.72
	DA-CapNet	0.62	0.82	0.58

TABLE 2. Comparison result of different attention module integrated on U-Net. All three methods have the same architecture except the attention module.

Method	IoU	Precision	Recall
U-Net + CBAM	0.62	0.65	0.81
U-Net + SE	0.56	0.60	0.71
Dual Attention U-Net	0.64	0.77	0.75

VI. CONCLUSION

We herein proposed a new deep-learning-based method for a nailfold capillary segmentation, named DA-CapNet. The DA-CapNet is an improvement of the well-known U-Net model through the integration of a dual attention module into several layers of U-Net; it yielded better feature representations from an input image with large variability caused by insufficient focusing and low visibility of capillaries compared with conventional methods. The proposed method was evaluated on a new collected dataset for the nailfold capillary segmentation. Extensive experimental results demonstrated its superiority compared with the state-of-the-art methods. For the future work, we plan to combine the DA-CapNet with an object detection network for detecting and counting white blood cell to implement a human immune monitoring condition.

ACKNOWLEDGMENT

This work was supported in part by the Basic Science Research Programs through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT under Grant NRF-2017R1A5A1015596, in part by the Ministry of Education under Grant NRF-2017R1D1A1B03031485, and in part by the Research Grant of Kwangwoon University in 2019.

REFERENCES

[1] A. Bourquard, I. Butterworth, A. Sánchez-Ferro, L. Giancardo, L. Soenksen, C. Cerrato, R. Flores, and C. Castro-Gonzalez, "Analysis of white blood cell dynamics in nailfold capillaries," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 7470–7473.

[2] A. Bourquard, A. Pablo-Trinidad, I. Butterworth, A. Sánchez-Ferro, C. Cerrato, K. Humala, M. F. Urdiola, C. Del Rio, B. Valles, J. M. Tucker-Schwartz, E. S. Lee, B. J. Vakoc, T. P. Padera, M. J. Ledesma-Carbayo, Y. B. Chen, E. P. Hochberg, M. L. Gray, and C. Castro-González, "Non-invasive detection of severe neutropenia in chemotherapy patients by optical imaging of nailfold microcirculation," *Sci. Rep.*, vol. 8, no. 1, pp. 1–12, 2018.

[3] M. Etehad Tavakol, A. Fatemi, A. Karbalaie, Z. Emrani, and B.-E. Erlandsson, "Nailfold capillaroscopy in rheumatic diseases: Which parameters should be evaluated?" *BioMed. Res. Int.*, vol. 2015, pp. 1–17, 2015.

[4] M. Cutolo, A. Sulli, M. E. Secchi, S. Paolino, and C. Pizzorni, "Nailfold capillaroscopy is useful for the diagnosis and follow-up of autoimmune rheumatic diseases. A future tool for the analysis of microvascular heart involvement?" *Rheumatology*, vol. 45, no. 4, pp. iv43–iv46, Oct. 2006.

[5] O. Wilhelmsson, "Evaluation of video stabilisation algorithms in dynamic capillaroscopy," M.S. thesis, School Elect. Eng. Comput. Sci., KTH Roy. Inst. Technol., Stockholm, Sweden, 2018.

[6] A. Karbalaie, M. Etehadtavakol, F. Abtahi, A. Fatemi, Z. Emrani, and B.-E. Erlandsson, "Image enhancement effect on inter and intra-observer reliability of nailfold capillary assessment," *Microvascular Res.*, vol. 120, pp. 100–110, Nov. 2018.

[7] A. Karbalaie, Z. Emrani, A. Fatemi, M. Etehadtavakol, and B.-E. Erlandsson, "Practical issues in assessing nailfold capillaroscopic images: A summary," *Clin. Rheumatol.*, vol. 38, no. 9, pp. 2343–2354, Sep. 2019.

[8] F. Isgro, F. Pane, G. Porzio, R. Pennarola, and E. Pennarola, "Segmentation of nailfold capillaries from microscopy video sequences," in *Proc. 26th IEEE Int. Symp. Comput.-Based Med. Syst.*, Jun. 2013, pp. 227–232.

[9] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2015, pp. 234–241.

[11] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2014, pp. 818–833.

[12] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[13] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[14] W. Li, F. Jia, and Q. Hu, "Automatic segmentation of liver tumor in CT images with deep convolutional neural networks," *J. Comput. Commun.*, vol. 3, no. 11, pp. 146–151, 2015.

[15] H. R. Roth, A. Farag, L. Lu, E. B. Turkbey, and R. M. Summers, "Deep convolutional networks for pancreas segmentation in CT imaging," *Proc. SPIE*, vol. 9413, Mar. 2015, Art. no. 94131G.

[16] R. M and C. M. Sujatha, "Hybrid LSM-based image segmentation and analysis of morphological variations of the brainstem in alzheimer MR images," *IEIE Trans. Smart Process. Comput.*, vol. 7, no. 2, pp. 124–131, Apr. 2018.

[17] B. Khagi and G.-R. Kwon, "CNN model performance analysis on MRI images of an OASIS dataset for distinction between healthy and Alzheimer's patients," *IEIE Trans. Smart Process. Comput.*, vol. 8, no. 4, pp. 272–278, Aug. 2019.

[18] P. Hu, F. Wu, J. Peng, Y. Bao, F. Chen, and D. Kong, "Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 3, pp. 399–411, Mar. 2017.

[19] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[20] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[21] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–13, 2018.

[22] C. Park, C. C. Took, and J.-K. Seong, "Machine learning in biomedical engineering," *Biomed. Eng. Lett.*, vol. 8, pp. 1–3, Feb. 2018.

[23] R. F. Mansour, "Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy," *Biomed. Eng. Lett.*, vol. 8, no. 1, pp. 41–57, Feb. 2018.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[25] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 2843–2851.

[26] M. Paradowski, H. Kwasnicka, and K. Borysewicz, "Avascular area detection in nailfold capillary images," in *Proc. Int. Multiconf. Comput. Sci. Inf. Technol.*, Oct. 2009, pp. 419–424.

[27] B. Jones, M. Oral, C. Morris, and E. Ring, "A proposed taxonomy for nailfold capillaries based on their morphology," *IEEE Trans. Med. Imag.*, vol. 20, no. 4, pp. 333–341, Apr. 2001.

[28] R. Nivedha, M. Brinda, K. V. Suma, and B. Rao, "Classification of nailfold capillary images in patients with hypertension using non-linear SVM," in *Proc. Int. Conf. Circuits, Controls, Commun. Comput. (I4C)*, Oct. 2016, pp. 1–5.

[29] K. V. Suma, K. Indira, and B. Rao, "Fuzzy logic based classification of nailfold capillary images in healthy, hypertensive and diabetic subjects," in *Proc. Int. Conf. Comput. Commun. Inform. (ICCCI)*, Jan. 2017, pp. 1–5.

[30] A. Tama, T. R. Mengko, and H. Zakaria, "Nailfold capillaroscopy image processing for morphological parameters measurement," in *Proc. 4th Int. Conf. Instrum., Commun., Inf. Technol., Biomed. Eng. (ICICI-BME)*, Nov. 2015, pp. 175–179.

[31] K. V. Suma, V. Sasi, and B. Rao, "A novel approach to classify nailfold capillary images in indian population using USB digital microscope," *Int. J. Biomed. Clin. Eng.*, vol. 7, no. 1, pp. 25–39, Jan. 2018.

[32] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3156–3164.

[33] J. Park, S. Woo, J.-Y. Lee, and I. S. Kweon, "Bam: Bottleneck attention module," 2018, *arXiv:1807.06514*. [Online]. Available: https://arxiv.org/abs/1807.06514

[34] P. F. Jaeger, S. A. Kohl, S. Bickelhaupt, F. Isensee, T. A. Kuder, H.-P. Schlemmer, and K. H. Maier-Hein, "Retina U-net: Embarrassingly simple exploitation of segmentation supervision for medical object detection," 2018, *arXiv:1811.08661*. [Online]. Available: https://arxiv.org/abs/1811.08661

[35] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 618–626.

YULI SUN HARIYANI received the B.S. degree in telecommunication engineering and the M.S. degree in electrical engineering from Telkom University, Bandung, Indonesia, in 2010 and 2013, respectively. She is currently pursuing the Ph.D. degree with Computer Engineering Department, Kwangwoon University, Seoul, South Korea. Since 2014, she has been a Lecturer with Telkom University, Indonesia. Her research interests include pattern recognition and medical image processing.

HEESANG EOM received the B.S. degree in software engineering from Korea Polytechnic University, Gyeonggi, South Korea. He is currently pursuing the M.S. degree in computer engineering with Kwangwoon University, Seoul, South Korea. His research interests include computer vision, text mining, and machine-learning algorithms.

CHEOLSOO PARK received the B.Eng. degree in electrical engineering from Sogang University, Seoul, South Korea, the M.Sc. degree from the Biomedical Engineering Department, Seoul National University, Seoul, and the Ph.D. degree in adaptive nonlinear signal processing from Imperial College London, London, U.K., in 2012. From 2012 to 2013, he was a Postdoctoral Researcher with the University of California at San Diego. He is currently an Associate Professor with the Computer Engineering Department, Kwangwoon University, Seoul. His research interests include machine learning and adaptive and statistical signal processing, with applications in healthcare, computational neuroscience, and wearable technology.