

EDDS-Unet: An Encoder-Decoder Double Skip Connection Scheme for Skin Lesion Segmentation

*Thi-Thao Tran, Minh-Nhat Trinh, Nhu-Toan Nguyen, Van-Truong Pham**

Hanoi University of Science and Technology, Ha Noi, Vietnam

**Corresponding author email: truong.phamvan@hust.edu.vn*

Abstract

The paper presents an approach for effective skin lesion segmentation from dermoscopic images. Aiming at transferring the weights trained from a network originally designed for image classification task, this study proposes to utilize the first layers of EfficientNet as the encoding layers of a U-Net based architecture. Besides, we introduce an encoder-decoder double skip connection scheme, a new skip connection architecture for extracting useful spatial details of skin lesions from the encoding layers. By the double skip scheme, the approach not only fuses information from the encounter layer in the encoder path to the corresponding layer in the decoder path, but also takes into account information of the proceeding encoding layer. In addition, we propose a new decoder network using the Residual blocks and Convolutional Block Attention Module (CBAM) blocks to handle the gradient vanishing problem as well as penalize the weight of each layer. The proposed Encoder-Decoder Double Skip with the Unet architecture, namely EDDS-Unet, has shown promising performance when evaluated on the official ISIC 2017 challenge and the PH2 databases. The proposed method achieves high evaluation scores with the Dice Similarity Coefficients of 0.907 for the ISIC 2017 and 0.950 for the PH2 databases without pre-or post-processing steps.

Keywords: Skin lesion segmentation, skip connection, u-net, efficientNet.

1. Introduction

Image segmentation is of paramount importance in enormous image processing, and object recognition applications, especially in medical image analysis [1]. Having the segmented objects, it is possible to analyze the changes in sizes and shapes of interested objects. In the case of skin cancer diagnosis, skin lesion segmentation (SLS) from dermoscopic images is a central component of the computer-aided diagnosis system that helps experts in the analysis process [2]. However, obtaining accurate segmentation results for skin lesions is nontrivial since the lesions often have irregular and blurred edges. In addition, the presence of artifacts as well as cutaneous features like hairs, air bubbles, blood vessels, and ink frames also causes difficulties for the segmentation task [3].

In recent years, advances in deep learning, specifically convolutional neural networks (CNNs) contribute significant development in not only classification works but also segmentation works. In deep learning-based segmentation tasks, Ronneberger *et al.* [4] introduced the U-Net architecture which has been a remarkable model for segmenting medical images. U-Net owes its name to its symmetric U-shape that consists of three main parts: encoder, decoder, and the skipped connection encoder-decoder. The U-Net has been a popular network for image segmentation and has also been adapted to other image processing fields such as image translation, image saliency

detection. The CNN approach and its variants have also shown promising success for medical image segmentation issues [5]. In particular, for SLS, a vast number of CNN-based methods have been presented. For instance, inspired from deep residual networks, Yu *et al.* [6] introduced a technique for both SLS and melanoma classification problems. Yuan *et al.* [7] introduced as a new loss function called the Jaccard distance for training the deep FCN model for SLS. Based on the U-Net structure, Tang *et al.* [8] introduced a multi-stage scheme for segmenting skin lesion images. More recently, Ünver and Ayan [9] have proposed a real-time algorithm for skin segmentation by integrating the YOLO with GrabCut approaches. Du *et al.* [10] showed that the contextual information is effectively disseminated to higher resolution layers due to the plenty of feature maps in the decoder part. Tran and Pham [11] proposed a new CNN model in which the fuzzy energy-based shape distance is used as a new loss function for skin lesion segmentation. The crucial benefit of the U-Net architecture makes it suitable for many biomedical segmentation problems. A good review for skin lesion segmentation by CNN approach can be found at [12]. Inspired by the successes of the U-Net architecture [4] and EfficientNet [13] for image classification tasks, in the current study, we proposed an approach network with the new idea of a double skip connection encoder-decoder scheme, namely EDDS, which stands for encoder-decoder double skip connection scheme. To

this end, we leverage the power of EfficientNet in extracting informative features for the encoder path and simultaneously increase the layer in the bottleneck layer, combining with the Convolutional Block Attention Module (CBAM) block [14] to penalize the weight for each layer. Besides, the decoder path is also rebuilt by modifying Residual block after using the proposed way of skip connection layers from the encoder path, which is different from the traditional method [4] that uses a single skip connection extracted from the encoder to the decoder of the same resolution layers. Moreover, in addition to using the traditional skip connection, we also use an additional connection from the preceding layer. In this way, the model could retain crucial spatial information. Since a manipulation detection approach needs to also analyze higher-level image cues along with detecting smooth edges, a combination of spatial resolution cues from different layers would potentially boost detection performance. In this regard, the proposed framework uses a fusion of low-level and preceding level representations for manipulation detection. Instead of a simple concatenation, we added the CBAM block [14] after fusing to weight each skip layer to get better performance.

For segmentation performance validation, we apply the approach to the official ISIC 2017 Challenge and PH2 datasets. The numerical results are assessed using standard Dice (DSC) and Jaccard similarity coefficients and statistical parameters, including accuracy, sensitivity, and specificity. Comparative experiments illustrate that our proposed technique attains remarkably high accuracy and outperforms other state-of-the-art (SOTA) methods. In particular, for the official ISIC 2017 challenge data, the proposed approach achieves the DSC of 0.907 without using any pre-/post-processing stages, the highest score reported so far, to the best of our survey.

Contributions of this study can be pointed out as: First, we introduce a double encoder-decoder skip connection scheme to improve accuracy. Second, we propose rebuilding the decoder part of the U-Net architecture by modifying Residual block after using the proposed skip connection layers from the encoder path. Third, we utilize the EfficientNet layers as encoding layers of the proposed U-Net architecture.

2. Material and Methods

2.1. Network Architecture

This section illustrates the network architecture presented in Fig. 1, which includes two paths: encoders and decoders. The role of the encoder path is to compress meaningful features to enhance the model's performance. Since training a complicated model from scratch with plenty of parameters demand a large dataset, in this work, we utilize the idea of transfer learning by exploiting pre-trained EfficientNetB3 layers [13] as the encoder path.

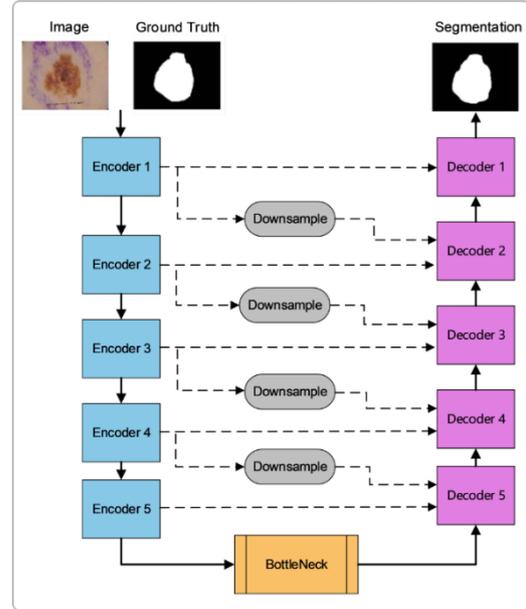


Fig. 1. The architecture of the encoder-decoder double skip connection scheme (EDDS-Unet).

However, collecting an immense amount of labelled data is challenging, especially in the case of medical images. Therefore, the idea of transfer learning could help in inheriting the capacity of pre-trained models for new problems with the limited amount of data and annotations work.

From that point, the first five blocks of EfficientB3 are employed as pre-trained models to construct the encoder path. More specifically, the first layer consists of convolutional 3×3 , a batch normalization (BatchNorm) layer, and Swish activation. Then five improved structures of MobileNetV2 convolutional blocks are added with Swish activation, rather than ReLU activation as in the original block. In addition, the squeeze-and-excited phase [13] is utilized inside to increase the network representational ability via dynamic channel-wise feature re-calibration. Besides, the head part of EfficientNet is modified by substituting the last Global Average Pooling layer and all Dense layers by two convolutional 1×1 filters, a BatchNorm, and then a Leaky ReLU, which is constructed according to the Residual block of ResNet [15]. The detailed structures of Encoder and sub-blocks of the Encoder block, as well as the BottleNeck of the proposed scheme, are presented in Fig. 2.

Regarding the decoder path, the output of the previous encoding layer is firstly upsampled, which is then concatenated with the feature maps copied from the corresponding encoder path with the same size via skip connection structure. Unlike the conventional skipped connection encoder-decoder, in our proposed skip connection scheme, along with feature maps from the corresponding layer in the encoder, the feature maps from the preceding encoder layer are also fused in skip connection layer after being downsampled to a

similar size, as in the gray blocks in Fig. 1. In other words, our proposed model processes two types of feature maps.

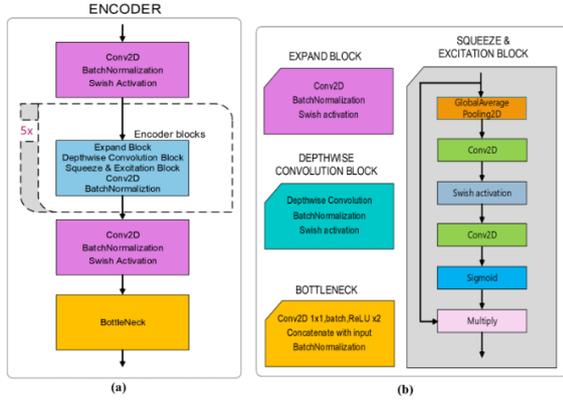


Fig. 2. Structures of (a) the Encoder and (b) sub-blocks of the Encoder block and the Bottleneck of the proposed scheme.

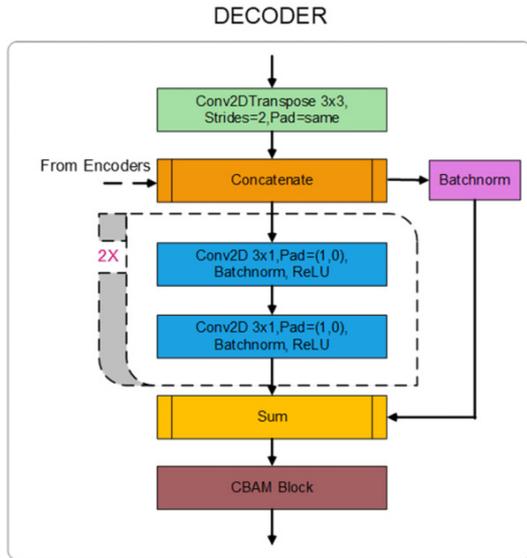


Fig. 3. The structure of the Decoder in the proposed scheme.

Let $X_e \in R^{F_b \times W_b \times H_b}$ denote the set of output feature maps copied from the b^{th} Mobile Block in encoder, and $X_d \in R^{F_b \times W_b \times H_b}$. It is worth to mention that $W_b = W_d = 1/2 \times W_{b-1}$ and $H_b = H_d = 1/2 \times H_{b-1}$. Firstly, X_d is passed to an up-convolutional layer where an up-sampling and a 2×2 convolution operation are performed. In this way, the size of each feature map is increased, denoted X_b^{up} , while the number of feature channels is reduced, i.e., X_{e-1} is downsampled to $X_{e-1}^{down} \in R^{F_b \times W_b \times H_b}$. Following that, X_b^{up} is concatenated with X_{e-1}^{down} and X_e to yield X_c^d . Next, X_c^d is fed to the modified version of the Residual

block, which has two sub-blocks containing convolutional 3×1 and 1×3 filters, followed by Leaky ReLU and then a BatchNorm. After that, the output is added with input to integrate the memory or information from the first model's layers to the final one, producing \bar{X}_c^d . Finally, \bar{X}_c^d goes through CBAM block [14] with a ratio of 8, which increases a small number of parameters for each channel so that the network can refine the importance of each feature map with minor cost computing. Consequently, it helps discard redundant information while enhancing necessary features in up-sampling operations. More details about the decoder structure are presented in Fig. 3.

2.2. Training

The proposed neural network is trained with the TPUv2 on Google Colab and programmed with the TensorFlow framework. As mentioned in the previous subsection, we employ the idea of transfer learning in the encoders. To show the effectiveness of using a pre-trained weight, we also compare the approach when trained from scratch with the mini-batch size of 64. In more detail, in the case of using pre-trained weights, we used weights from ImageNet. For this case, we run the first 100 epochs with the encoder weights frozen and use the Nadam optimization algorithm with default configuration (learning rate (LR) begin with 0.001). Additionally, the LR would be reduced by 30% if the validation performance was not enhanced for five epochs. An early stopping callback would be applied if the average loss of the training process did not decrease for 16 consecutive epochs. Then, we unfreeze the encoder path and continue training the whole model with the learning rate set to 0.0005 in order to prevent wreaking havoc on the representations learned by the model so far. The number of epochs for training is 50, but the process would be stopped early if the validation performance was not enhanced for 21 epochs. Hence, the total iteration maximum number is 150. In terms of training from scratch, we train the whole model in 250 epochs with the Nadam optimizer (LR set to 0.001 initially). Eventually, the training process lasted approximately 5 hours, and each sample took nearly one second to segment in the testing phase.

2.3. Loss Function

In this study, for training the network, we employ the variant of the Tversky loss function for handling extremely imbalanced database as well as small regions of interest in skin lesions. The Tversky loss is inspired by the Tversky similarity index (TI) [16], a generalization of the DSC expressed as the following:

$$TI_c = \frac{\sum_{i=1}^N p_{ic} g_{ic} + \varepsilon}{\sum_{i=1}^N (p_{ic} + \alpha p_{ic} g_{ic} + \beta p_{ic} g_{ic}) + \varepsilon} \quad (1)$$

where p_{ic} is the probability that pixel i in the predicted label belongs to the lesion class c , and $p_{i\bar{c}}$ refers to the probability the pixel i belongs to non-lesion class \bar{c} . The same applied to the pixel of the ground truth g_{ic} and $g_{i\bar{c}}$. In the current work, we adopt the Focal Tversky Loss (FTL) in [16] as:

$$FTL_c = \sum_c (1 - TI_c)^\gamma \quad (2)$$

where hyperparameter $\alpha = 0.7$, $\beta = 0.3$, and $\gamma = 4/3$ as suggested by [16].

3. Experimental Results

3.1. Dataset

In this work, we first evaluate the method on part I of the ISIC 2017 database [17] which contains 2000 dermoscopic training images and the corresponding lesion masks. The validation set contains 150 skin mirror-image pairs, including dermoscopic images and corresponding segmentation masks. The test data consists of 600 image pairs, with the test images used for the testing phase and the ground truth masks used for segmentation evaluation. The images and masks have the different sizes, varying from 540×722 to 4499×6748 pixels. In addition, we also evaluate the proposed approach with the PH2 dataset [18] which includes 200 image pairs of the size of 768×560 . Each image pair includes a skin lesion dermoscopic image with a given mask from expert annotation. For efficient computation for training and reducing memory usage, we scaled all images and masks to 192×256 then divided them by 255 to normalize the data. To increase the number of training data, we horizontally and vertically flip the training images and their corresponding ground truth masks.

3.2. Performance Evaluation

The Jaccard Coefficient (JAC), Dice Similarity Coefficient (DSC), the Accuracy (Acc), the Sensitivity (Sen), and the Specificity (Spe) are taken into implementation to evaluate the effectiveness of the proposed approach. The formulations of these evaluation metrics are expressed as follows:

$$DSC = \frac{2 \times TP}{2 \times TP + FN + FP} \quad (3)$$

$$JAC = \frac{TP}{TP + FN + FP} \quad (4)$$

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \quad (5)$$

$$Sen = \frac{TP}{TP + FN} \quad (6)$$

$$Spe = \frac{TN}{TN + FP} \quad (7)$$

where TP , FN , TN , and FP are the number of true positives, false negatives, true negatives and false positives, respectively.

3.3. Results

3.3.1. Results on ISIC 2017 challenge dataset

Some representative obtained results on the official ISIC 2017 testing dataset are presented in Fig. 4. In this figure, input images are indicated in the first column, the obtained results and ground truths are provided in the succeeding columns of the figure. The overlays between predictive segmentations and the ground truth masks are given in the last column. The segmentations as well as the overlays show good agreements between the predictive labels and ground truths.

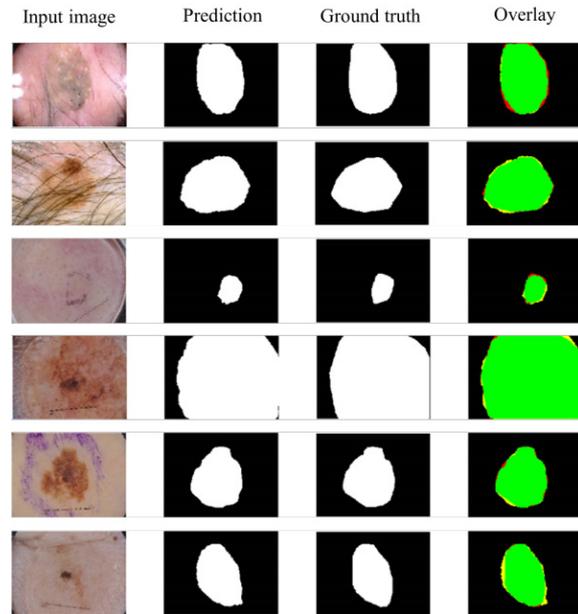


Fig. 4. Representative segmentation results on the ISIC2017 database. From left to right: Input images, predicted masks, ground truths, and the overlays between the predictions and ground truths.

To quantify the performance of our technique, we compared segmentation scores, including JAC and DSC, to those reported in previous work. For the ISIC 2017 challenge database, we compared the results with those reported in Tschandl *et al.* [19], Tu *et al.* [2], Song *et al.* [20], Zafar *et al.* [21], Xie *et al.* [22], Jahanifar *et al.* [23]. We presented the average Dice and Jaccard similarity scores of comparative methods for the data set in Table 1. In [21], the Res-UNet, which is a combination of the ResNet and the U-Net, achieves the DSC of 0.858 and the JAC of 0.772.

Table 1. Comparative segmentation with other state-of-the-arts on the ISIC2017 database

Method	DSC	JAC
Zafar <i>et al.</i> [21]	0.858	0.772
Tu <i>et al.</i> [2]	0.862	0.768
Song <i>et al.</i> [20]	0.856	0.765
Xie <i>et al.</i> [22]	0.879	0.800
Jahanifar <i>et al.</i> [23]	0.879	0.806
Tschandl <i>et al.</i> [19]	0.886	0.813
Proposed EDDS-Unet	0.907	0.841

However, Tschandl *et al.* [19] proposed an encoder-decoder architecture employing ResNet34 with transfer learning approach as encoding layers, which outperforms previous works with the DSC of 0.886. On the other hand, by comparing quantitatively, the proposed approach achieves the most remarkable segmentation results in terms of both DSC and JAC scores (0.907 and 0.841, respectively). This illustrates the effectiveness of our EDDS-Unet for SLS.

3.3.2. Results on PH2 dataset

Like the case of the ISIC 2017 database, the experiments on the PH2 data [18] are also evaluated the segmentation performance by our approach further. Some representative results on the test images of PH2 data using the proposed EDDS-Unet architecture are shown in Fig. 5. As can be observed from this figure, the obtained prediction masks are well agreed with the ground truth masks, with high overlaps between the two masks of each image in the last column of this figure.

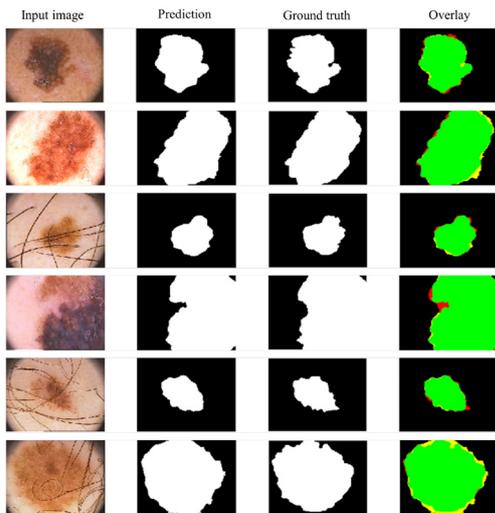


Fig. 5. Representative segmentation results on the PH2 database. From left to right: Input images, predicted masks, ground truths, and the overlays between the predictions and ground truths.

Table 2. Comparative segmentation with other state-of-the-arts on the PH2 database

Method	DSC	JAC
Yuan <i>et al.</i> [7]	0.915	0.815
Bi <i>et al.</i> [24]	0.921	0.859
Xue <i>et al.</i> [25]	0.924	0.851
Tu <i>et al.</i> [2]	0.932	0.863
Proposed EDDS-Unet	0.950	0.906

For the PH2 database, we compared results with DSC and JAC scores from the following works: Yuan *et al.* [7], Bi *et al.* [24], Xue *et al.* [25], and Tu *et al.* [2]. In this case, we used the same strategy as Tu *et al.* [2] for training to utilize 50 images for training, 50 images for validation, and the remaining 100 images for the testing stage. The quantitative results are represented in Table 2. As the observation from Table 2, all existing methods are evidently superior on segmenting skin lesions of the PH2 dataset with the DSC all over 0.910. However, it is worth emphasizing that our result has the dominant ability in handling SLS (0.950 DSC and 0.906 JAC), which again demonstrates the efficacy of the proposed approach.

4. Ablation Studies

To further show the performance of our double skip connection (EDDS-Unet), we compare the results when removing the second connection (the connection with Downsample in Fig. 1). The model with second connection ablation is called EDSS (stand for encoder-decoder with single skip connection). The two schemes, EDDS and EDSS, are conducted using the proposed network architecture with the same training protocol and data. In addition, the process for the two schemes is also evaluated when training from scratch and reusing the pre-trained weight from ImageNet. Some representative results of the two approaches are represented in Fig. 6. As shown in the figure, prediction results by the EDDS scheme most agree with the ground truths even with challenging images, especially when using the pre-trained weights, as obviously shown in the third, fifth, and last rows.

For quantitative comparison, we provided evaluation scores for two schemes in the cases of training from scratch and using pre-trained weights in Table 3. The two standard evaluation metrics, including DSC and JAC, are used for comparison. In addition, we also compute the Accuracy, Sensitivity, and Specificity. As we can see from Table 3, the double skip connection scheme (EDDS-Unet) gives the best scores for all metrics and parameters, especially for the Dice and Jaccard scores.

Table 3. Performance metrics on the ISIC2017 database of the proposed encoder-decoder double skip connection scheme (EDDS-Unet) in comparison with those by the traditional single skip connection scheme (EDSS-Unet) for all test data

Metric	Single skip, from scratch	Single skip, pretrained	Double skip, from scratch	Double skip, pretrained
DSC	0.841	0.862	0.866	0.907
JAC	0.752	0.776	0.784	0.841
Acc	0.929	0.942	0.943	0.967
Sen	0.802	0.827	0.847	0.927
Spe	0.968	0.978	0.972	0.980

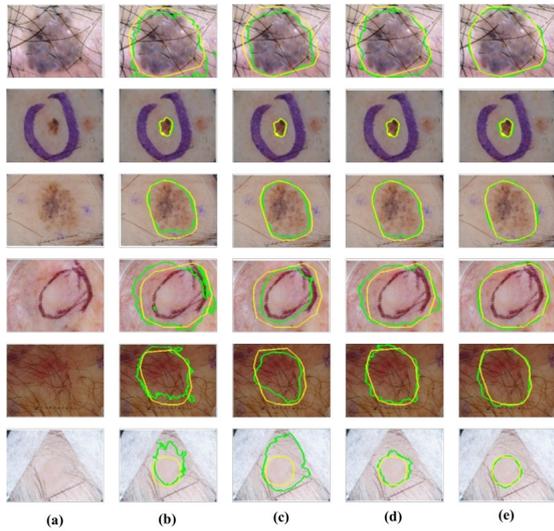


Fig. 6. Representative results on the ISIC2017 database (a) Input images; (b) by EDSS-Unet when trained from scratch; (c) by EDSS-Unet using pre-trained; (d) by EDDS-Unet when trained from scratch; (e) by EDDS-Unet using pre-trained.

For further comparison, we also plot the Receiver Operating Characteristic (ROC) curves along with Area Under Curve (AUC) values using the above schemes in Fig. 7. We can see from the figure that the proposed approach with the double skip connection scheme and using the pre-trained weights obtained the best ROC curve and the highest AUC value, 0.971, compared to 0.939 when using the traditional kip connection scheme.

In PH2 database experiment, the comparative segmentation results when using the double and single skip connections, EDDS and EDSS-Unet, when trained from scratch and using the pre-trained weights, are also compared. Some representative results are given in this experiment are provided in Fig. 8.

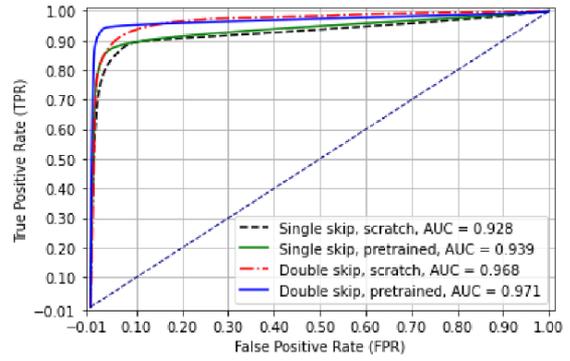


Fig. 7. The ROC curves and AUC values for ISIC2017 database using the studied model in the cases of using double skip (EDDS-Unet) and using a single skip (EDSS-Unet) for all test data.

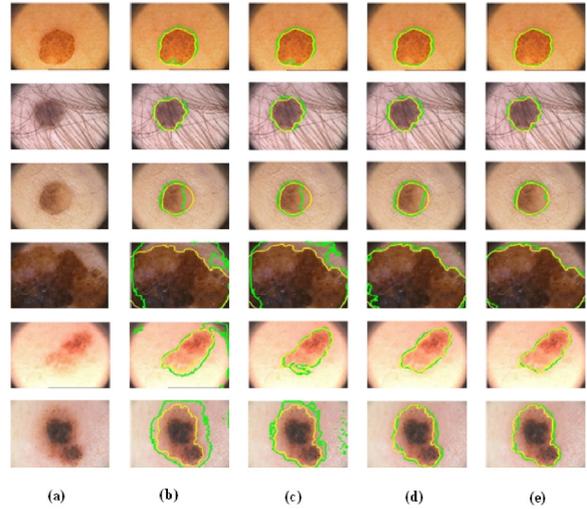


Fig. 8. Results on the PH2 database. (a) Input images; (b) by EDSS-Unet trained from scratch; (c) by EDSS-Unet using pre-trained; (d) by EDDS-Unet trained from scratch; (e) by EDDS-Unet using pre-trained

As can be observed from the figure, though all approaches obtain satisfactory segmentation results, the segmentation contours obtained by the double skip scheme most closely match the ground truth contours created by experts, as obviously shown in the first and third rows of Fig. 8.

Moreover, in addition to the qualitative comparison, the quantitative evaluations are also assessed for the two schemes with the two training cases, from scratch and using pre-trained weights. The segmentation scores and parameters are given in Table 4. This table shows that the double skip scheme achieves higher DSC and JAC scores and specificity parameters than the traditional single skip connection scheme.

Table 4. Performance metrics on the PH2 database of the proposed EDDS-Unet compared with those of EDSS-Unet for all test data.

Metric	Single skip, from scratch	Single skip, pretrained	Double skip, from scratch	Double skip, pretrained
DSC	0.924	0.930	0.932	0.950
JAC	0.863	0.873	0.876	0.906
Acc	0.95	0.956	0.955	0.972
Sen	0.967	0.962	0.937	0.970
Spe	0.943	0.953	0.963	0.974

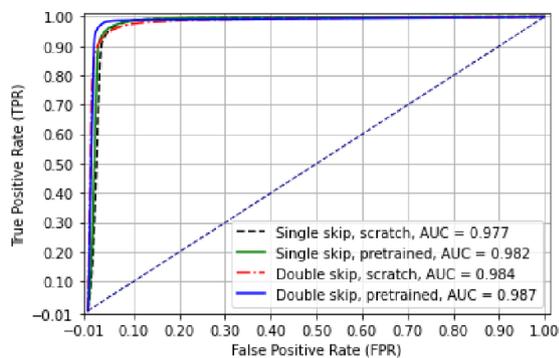


Fig. 9. The ROC curves and AUC values for PH2 database using the studied model in the cases of using double skip (EDDS-Unet) and using a single skip (EDSS-Unet) for all test data.

Furthermore, the ROC curves along with AUC values for all comparative approaches are also plotted in Fig. 9. The figure shows that the AUC values obtained by the proposed Unet-based architecture are all very high, more than 0.977. Nevertheless, the values obtained by the double skip connection are also higher than those obtained by using the single skip connection scheme.

5. Conclusion

We have presented a new approach for automated segmenting skin lesions. The proposed architecture reuses the first layers from the EfficientNet, originally built for the image classification task, for the encoding layers of the modified U-Net architecture. We also introduce a new skipped connection encoder-decoder to the network by fusing feature maps from the interrelated encoder layer and the preceding layer. In addition, we insert modified ResNet blocks and CBAM blocks into the decoding path. Experimental results on ISIC 2017 and the PH2 database show effective skin lesion segmentation with higher Dice and Jaccard similarity indexes than those reported by previous works. Finally, our proposed model, in addition to the skin lesion segmentation problem, can be extended to other segmentation applications.

Acknowledgements

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.05-2021.34.

References

- [1] M. Nasir, M. Attique Khan, M. Sharif, I. U. Lali, T. Saba, and T. Iqbal, An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach, *Microscopy Research and Technique*, vol. 81, pp. 528-543, 2018. <https://doi.org/10.1002/jemt.23009>
- [2] W. Tu, X. Liu, W. Hu, and Z. Pan, Dense-residual network with adversarial learning for skin lesion segmentation, *IEEE Access*, vol. 7, pp. 77037-77051, 2019. <https://doi.org/10.1109/ACCESS.2019.2921815>
- [3] S. Pathan, K. G. Prabhu, and P. C. Siddalingaswamy, Techniques and algorithms for computer aided diagnosis of pigmented skin lesions-A review, *Biomedical Signal Processing Control*, vol. 39, pp. 237-262, 2018. <https://doi.org/10.1016/j.bspc.2017.07.010>
- [4] O. Ronneberger, P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [5] Thi-Thao Tran, Dinh-Thien Vu, Thi Hue Nguyen, and Van-Truong Pham, A CNN-Transformer-based Approach for Medical Image Segmentation, in *International Conference on System Science and Engineering (ICSSE)*, Ho Chi Minh, Vietnam, 2023.
- [6] L. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, Automated melanoma recognition in dermoscopy images via very deep residual networks, *IEEE Transactions on Medical Imaging*, vol. 36, pp. 994-1004, 2016. <https://doi.org/10.1109/TMI.2016.2642839>
- [7] Y. Yuan, M. Chao, and Y. C. Lo, Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance, *IEEE Transactions on Medical Imaging*, vol. 36, pp. 1876-1886, 2017. <https://doi.org/10.1109/TMI.2017.2695227>
- [8] Y. Tang, F. Yang, and S. Yuan, A multi-stage framework with context information fusion structure for skin lesion segmentation, in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 1407-1410. <https://doi.org/10.1109/ISBI.2019.8759535>
- [9] H. M. Ünver and E. Ayan, Skin lesion segmentation in dermoscopic images with combination of YOLO and grabcut algorithm, *Diagnostics*, vol. 9, p. 72, 2019. <https://doi.org/10.3390/diagnostics9030072>
- [10] G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, Medical image segmentation based on u-net: A review,

- Journal of Imaging Science and Technology, vol. 64, pp. 20508-1-20508-12, 2020.
<https://doi.org/10.2352/J.ImagingSci.Technol.2020.64.2.020508>
- [11] Thi-Thao Tran and Van-Truong Pham, Fully convolutional neural network with attention gate and fuzzy active contour model for skin lesion segmentation, *Multimedia Tools and Applications*, vol. 81, pp. 13979-13999, 2022.
<https://doi.org/10.1007/s11042-022-12413-1>
- [12] Zahra M., Kumar A., Alceu B., Catarina B., Sandra A., V. Eduardo, et al., A survey on deep learning for skin lesion segmentation, *Medical Image Analysis*, vol. 88, pp. 1-40, 2023.
<https://doi.org/10.1016/j.media.2023.102863>
- [13] M. Tan and Q. V. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, arXiv:1905.11946, 2019.
- [14] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, Cbam: Convolutional block attention module, in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3-19.
https://doi.org/10.1007/978-3-030-01234-2_1
- [15] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
<https://doi.org/10.1109/CVPR.2016.90>
- [16] N. Abraham and N. Khan, A novel focal Tversky loss function with improved attention u-net for lesion segmentation, in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 683-687.
<https://doi.org/10.1109/ISBI.2019.8759329>
- [17] N. Codella, D. Gutman, M. Celebi, B. Helba, M. Marchetti, S. Dusza, et al., Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic), in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 168-172.
<https://doi.org/10.1109/ISBI.2018.8363547>
- [18] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, PH 2-A dermoscopic image database for research and benchmarking, in *2013 35th annual international conference of the IEEE Engineering In Medicine And Biology Society (EMBC)*, 2013, pp. 5437-5440.
<https://doi.org/10.1109/EMBC.2013.6610779>
- [19] P. Tschandl, C. Sinz, and H. Kittler, Domain-specific classification-pretrained fully convolutional network encoders for skin lesion segmentation, *Computers in Biology And Medicine*, vol. 104, pp. 111-116, 2019.
<https://doi.org/10.1016/j.combiomed.2018.11.010>
- [20] L. Song, J. Lin, Z. J. Wang, and H. Wang, Dense-residual attention network for skin lesion segmentation, in *International Workshop on Machine Learning in Medical Imaging*, 2019, pp. 319-327.
https://doi.org/10.1007/978-3-030-32692-0_37
- [21] K. Zafar, S. O. Gilani, A. Waris, A. Ahmed, M. Jamil, M. N. Khan, et al., Skin lesion segmentation from dermoscopic images using convolutional neural network, *Sensors*, vol. 20, p. 1601, 2020.
<https://doi.org/10.3390/s20061601>
- [22] Y. Xie, J. Zhang, Y. Xia, and C. Shen, A mutual bootstrapping model for automated skin lesion segmentation and classification, *IEEE Transactions on Medical Imaging*, vol. 39, pp. 2482-2493, 2020.
<https://doi.org/10.1109/TMI.2020.2972964>
- [23] M. Jahanifar, N. Z. Tajeddin, N. A. Koohbanani, A. Gooya, and N. Rajpoot, Segmentation of skin lesions and their attributes using multi-scale convolutional neural networks and domain specific augmentations, 2018, arXiv preprint arXiv:1809.10243
- [24] L. Bi, J. Kim, E. Ahn, and D. Feng, Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks, 2017, arXiv preprint arXiv:1703.04197.
- [25] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, SegAN: adversarial network with multi-scale L1 loss for medical image segmentation, *Neuroinformatics*, vol. 16, pp. 383-392, 2018.
<https://doi.org/10.1007/s12021-018-9377-x>