Skin Lesion Segmentation Using Deep Learning

Minh Huu-Tuan Nguyen[1,2], Loc Hoang Ngo[1,2],
Thien Tat-Bao Nguyen[1,2]

[1]Faculty of Information Science and Engineering, University of Information Technology, Ho Chi Minh City, Vietnam.
[2]Vietnam National University, Ho Chi Minh City, Vietnam.

Contributing authors: 21520348@gm.uit.edu.vn; 20520618@gm.uit.edu.vn; thienntb@uit.edu.vn;

Abstract

Leveraging image processing techniques and deep neural networks, skin lesion segmentation endeavors to automate and enhance the analysis of dermatological images, thereby facilitating more efficient and reliable clinical decision-making processes. This research paper utilizes and compares various deep learning models used in skin lesion segmentation task, using two datasets from the International Skin Imaging Collaboration (ISIC). The models are compared in terms of model's size, computing complexity and performance on the datasets using five evaluation metrics. Our experimental results show that U-NeXt outperform traditional U-Net and Attention U-Net in terms of both performances, model's size and computing complexity. Our source code is publicly available on Github *

Keywords: Skin Lesion Segmentation, Deep Learning

1 Introduction

Skin cancer poses a significant global health concern and stands as one of the leading cancer types worldwide. Early cancer diagnosis can significantly improve a patients chances of being cured. Skin lesion segmentation is a critical task in medical imaging and computer vision aimed at precisely delineating and characterizing regions of interest within dermatological images. Skin lesions encompass a broad spectrum of

*https://github.com/chains229/Skin-Lesion-Segmentation

abnormalities, including benign moles, malignant melanomas, and various dermatological conditions. Segmentation plays a pivotal role in early detection, diagnosis, and treatment planning by providing clinicians with accurate spatial information about lesions' boundaries and characteristics. However, skin lesion segmentation is challenging due to several dermoscopic image issues, e.g., illumination changes, low contrast of images, distinct texture, position, shapes, color, and boundaries of skin lesions. In addition, visual artifacts in dermoscopic images, such as air bubbles, hair, ruler markers, and blood vessels, make skin cancer segmentation extremely difficult. Some examples of theses issues can be shown in Figure 1.

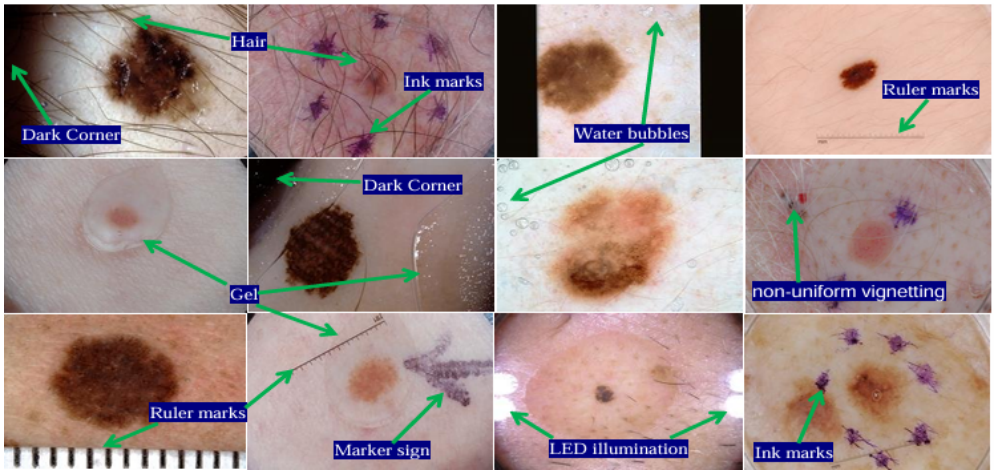


Fig. 1 Some issue examples in ISIC 2017's test dataset.

Deep learning has been proven to be a pioneer in fast and accurate pattern recognition in medical applications. Consequently, it can provide specialists with valuable help during the diagnosis stages. Leveraging advanced image processing techniques and deep neural networks, skin lesion segmentation endeavors to automate and enhance the analysis of dermatological images, thereby facilitating more efficient and reliable clinical decision-making processes. This research paper utilizes and compares various deep learning models used in skin lesion segmentation task, using two datasets from the International Skin Imaging Collaboration (ISIC).

This paper is organized as follows: in section 2, the related works of skin lesion segmentation are described. Section 3 explains the datasets we used in this study. We display the models implemented in Section 4. In section 5, we show the experimental results of the discussed methods. Finally, the contribution of this paper and future works are presented in section 6.

2 Related Work

There are various approaches for Skin Lesion Segmentation (SLS) task. They are categorized into four groups: edge-based, region-based, threshold-based and deep learning-based. In edge-based SLS, an edge filter is applied to the image, pixels are classified as edge or non-edge based on the filter output, and pixels not separated by an edge are assigned to the same class. Some popular edge detection algorithms are watershed algorithm [1], active contours [2, 3] and canny edge detector [4]. In Region-based SLS systems, images are divided into regions or groups of comparable pixels based on their attributes, assuming neighboring pixels should have the same value. K-means and fuzzy C-means clustering [5–9] are the most common region-based SLS methods. Threshold-based SLS can be classified as point-based or pixel-based segmentation, depending on the threshold estimation approaches, and commonly suffers from difficulty in estimating effective thresholds due to dermoscopic artifacts. OTSU [10–12] thresholding technique is the most common threshold-based SLS strategy.

In recent years, deep learning-based systems, especially CNN-based ones, have been widely and successfully used not only in Skin Lesion Segmentation task but also various areas in medical imaging: From breast cancer detection [13, 14], brain disease classification [15, 16] to pneumonia detection from CXR images [17, 18] and lung segmentation [19, 20]. Moreover, researchers in [21, 22] conducted comprehensive surveys on CNN in medical image segmentation task.

In SLS task, CNN-related models have been widely used and produced state-of-the-art results. Many researches utilized traditional CNN-based encoder-decoder architectures to segment the skin lesion images [23–27]. Since then, many studies have combined CNN with other structures like attention module and transformer architecture [28–32], to achieve better performances in SLS task. In this research, we implemented three versions of U-Net, a popular CNN-based encoder-decoder architecture, to evaluate and compare the results in this task.

3 Datasets

In this research, we use two datasets from the International Skin Imaging Collaboration (ISIC 2017 [33] and ISIC 2018 [34]) to train and evaluate our models. They were initially used for the ISIC Challenge, which aimed to help participants develop image analysis tools to enable the automated diagnosis of melanoma from dermoscopic images, in the corresponding years. Since then, these datasets have been widely used to evaluate models constructed for the Skin Lesion Segmentation task.

These datasets contain a collection of dermoscopic images of skin lesions and were collected from major international clinical centres and from various devices within each centre. ISIC2017 and ISIC2018 have 2150 and 2694 dermoscopy images with segmentation mask labels, respectively. Masks were created by an expert clinician, using either a semi-automated process (using a user-provided seed point, a user-tuned flood-fill algorithm, and morphological filtering) or a manual process (from a series of user-provided polyline points). We randomly divide datasets in a ratio of $7:3$ as our experiments' training and testing sets. For ISIC2017, there are 1500 images in training sets and 650 images in testing sets. For ISIC2018, there are 1886 images in training

sets and 808 images in testing sets. A sample image and its corresponding mask is shown in table 1

(a) Image

(b) Mask

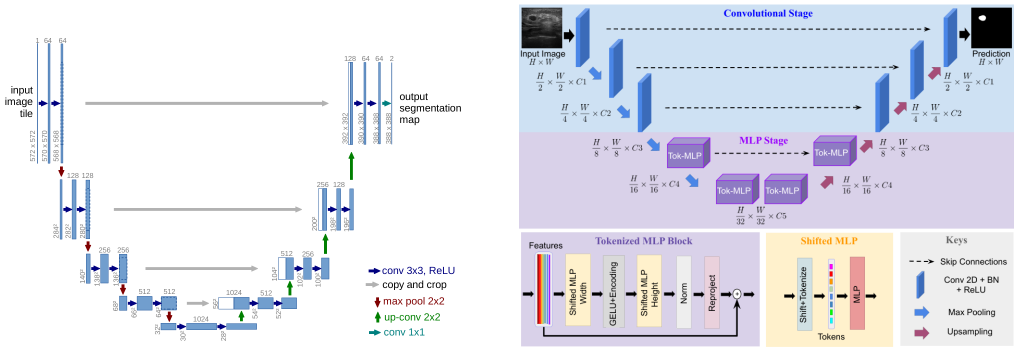
Table 1 A Sample In ISIC 2017 Dataset.

4 Methodology

In this study, we implemented U-Net, UNeXT and Attention U-Net to evaluate on the datasets and compare the results

4.1 UNet [24]

UNet was introduced and evaluated in a medical image segmentation task in 2015 and proved highly promising. The network architecture is illustrated in Figure 4.1. It consists of a contracting path (left side) and an expansive path (right side). The contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3x3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling. At each downsampling step we double the number of feature channels. Every step in the expansive path consists of an upsampling of the feature map followed by a 2x2 convolution (up-convolution) that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution. At the final layer a 1x1 convolution is used to map each 64 component feature vector to the desired number of classes. In total the network has 23 convolutional layers. The number of channels in each stage we are using in this study is 32, 64, 128, 256, 512.



(a) UNet (b) UNeXT

Table 2 Overview of U-Net and U-NeXt architectures.

4.2 UNeXt [35]

UNeXt (published in 2022) is an UNet-based encoder-decoder architecture with two stages: 1) Convolutional stage, and a 2) Tokenized MLP stage. The authors propose a tokenized MLP block where they eficiently tokenize and project the convolutional features and use MLPs to model the representation. To further boost the performance, they propose shifting the channels of the inputs while feeding in to MLPs so as to focus on learning local dependencies. Using tokenized MLPs in latent space reduces the number of parameters and computational complexity while being able to result in a better representation to help segmentation. The overview architecture of this model can be seen in Figure 4.1

The input image is passed through the encoder where the first 3 blocks are convolutional and the next 2 are Tokenized MLP blocks. The decoder has 2 Tokenized MLP blocks followed by 3 convolutional blocks. Each encoder block reduces the feature resolution by 2 and each decoder block increases the feature resolution by 2. We employ the light-weight version of this model, which is called UNeXt-S, and its number of channels is 8, 16, 32, 64, 128 in our experiments, which is the default setting used in its official open source codes.

4.3 Attention U-Net [36]

Attention Gates (AG) is created to automatically learn to focus on target structures of varying shapes and sizes in the input image. They operate by using input features from a specific layer and a gating signal from coarser scales, both of which undergo linear transformations followed by ReLU activation. The combined signal is then processed through a sigmoid activation to generate attention coefficients, which are used to scale the input features via element-wise multiplication. This mechanism allows the model to highlight important regions contextually, improving segmentation accuracy without significant computational overhead. AGs are integrated into the skip connections of the U-Net architecture, filtering features between the encoder and decoder to

enhance the model's performance. They enhance the relevant features while suppressing irrelevant regions, improving the model's focus on critical areas without needing additional supervision.

Attention gates can help U-Net architectures to learn better by focusing on more important features. In this architecture (published in 2018), attention gates scales the input features with attention coefficients, which are computed through additive attention. In other words, areas of higher significance are weighted more than the less significant ones. Consequently, the areas with higher weights get more attention during training. The number of channels in each stage we are using in this study is 64, 128, 256, 512, 1024.

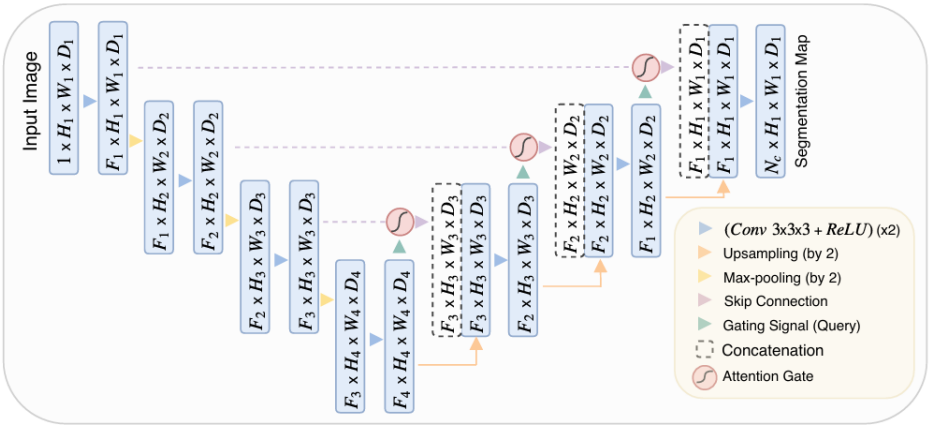


Fig. 2 Overview of Attention U-net architecture.

5 Experimental Results

5.1 Data Pre-processing and Transformation

In this study, we normalize and resize the images to 256 x 256. We also apply data augmentation strategies which include horizontal flip, vertical flip and random rotation.

5.2 Implementation Details

We trained the models on Kaggle, using P100 GPU. The optimizer which we utilize is AdamW [37] with an initial learning rate of 0.001. A cosine annealing learning rate scheduler is used with a maximum number of iterations of 50 and a minimum learning rate of 0.00001. The training epoch is set to 300, and the batch size is 8.

The loss function we use in this study is the BceDice loss, which can be expressed with this formula:

$$L_{\text{Bce}} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)]$$

$$L_{\text{Dice}} = 1 - \frac{2 \times |A \cap B|}{|A| + |B|}$$

$$L_{\text{BceDice}} = L_{\text{Bce}} + L_{\text{Dice}}$$

Where N is the total number of samples, y_i is the real label, p_i is the prediction. $|X|$ and $|Y|$ represent ground truth and prediction, respectively.

5.3 Evaluation Metrics

In this study, we evaluate the models on five metrics including Mean Intersection over Union (mIoU), Dice similarity score (DSC), Accuracy (Acc), Sensitivity (Sen) and Specificity (Spe). These metrics are calculated using these formulas:

$$\text{mIoU} = \frac{TP}{TP + FP + FN}$$

$$\text{DSC} = \frac{2TP}{2TP + FP + FN}$$

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Sen} = \frac{TP}{TP + FN}$$

$$\text{Spe} = \frac{TN}{TN + FP}$$

Here, TP, FP, FN, TN stand for true positive, false positive, false negative, and true negative.

Moreover, we display and compare the models' size (in million parameters) and the computational complexity is calculated regarding the number of floating point operators (GFLOPs) for each model.

5.4 Results

This section compares the performances of models we employed in section 4 on ISIC 2017 and ISIC 2018 datasets. The experimental results are shown in Table 3 and Table 4.

Table 3 Comparison of model performance on ISIC 2017 dataset

Model	Params	GFLOPs	mIoU	DSC	Acc	Spe	Sen
UNet	7.77	13.78	76.98	86.99	95.65	97.43	86.82
Attention UNet	8.73	16.74	77.56	87.11	95.80	97.49	**87.15**
UNeXt-S	**0.30**	**0.10**	**78.26**	**87.80**	**95.95**	**97.74**	87.04

Table 4 Comparison of model performance on ISIC 2018 dataset

Model	Params	GFLOPs	mIoU	DSC	Acc	Spe	Sen
UNet	7.77	13.78	77.86	87.55	94.05	96.69	85.86
Attention UNet	8.73	16.74	78.43	87.91	94.13	96.23	**87.60**
UNeXt-S	**0.30**	**0.10**	**79.09**	**88.03**	**94.39**	**96.72**	87.15

According to the result, comparing with UNet and Attention UNet, UNeXt has shown the improvement in all-round way. The model is lighter, faster to train and achieves better performance in both datasets.

6 Conclusion

In this study, we utilized CNN-based models to evaluate and compare the performances in the Skin Lesion Segmentation task. Our experimental results show that U-NeXt outperform traditional U-Net and Attention U-Net in terms of both performances, model's size and computing complexity. For future works, we will investigate different models using various backbones, and utilize more modules, especially attention, to enhance the performances in this task. Furthermore, we will try various data pre-processing and augmentation methods and demonstrate their effect on different models' performances.

References

[1] Chakkaravarthy, A.P., Chandrasekar, A.: An automatic segmentation of skin lesion from dermoscopy images using watershed segmentation. In: 2018 International Conference on Recent Trends in Electrical, Control and Communication (RTECC), pp. 15–18 (2018). https://doi.org/10.1109/RTECC.2018.8625662

[2] Riaz, F., Naeem, S., Nawaz, R., Coimbra, M.: Active contours based segmentation and lesion periphery analysis for characterization of skin lesions in dermoscopy images. IEEE Journal of Biomedical and Health Informatics **23**(2), 489–500 (2019) https://doi.org/10.1109/JBHI.2018.2832455

[3] Ivanovici, M., Stoica, D.: Color diffusion model for active contours - an application to skin lesion segmentation. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5347–5350 (2012). https://doi.org/10.1109/EMBC.2012.6347202

[4] J. H. Jaseema Yasmin, M.M.S.: An improved iterative segmentation algorithm using canny edge detector with iterative median filter for skin lesion border detection. International Journal of Computer Applications **50**(6), 37–42 (2012) https://doi.org/10.5120/7779-0865

[5] Garg, S., Jindal, B.: Skin lesion segmentation using k-mean and optimized fire fly algorithm **80**(5), 7397–7410 (2021) https://doi.org/10.1007/s11042-020-10064-8

[6] Jaisakthi, S.M.: Automated skin lesion segmentation of dermoscopic images using grabcut and k-means algorithms. IET Computer Vision **12**, 1088–10957 (2018)

[7] Alvarez, D., Iglesias, M.: k-Means Clustering and Ensemble of Regressions: An Algorithm for the ISIC 2017 Skin Lesion Segmentation Challenge (2017)

[8] Agarwal, A., Issac, A., Dutta, M.K., Riha, K., Uher, V.: Automated skin lesion segmentation using k-means clustering from digital dermoscopic images. In: 2017 40th International Conference on Telecommunications and Signal Processing (TSP), pp. 743–748 (2017). https://doi.org/10.1109/TSP.2017.8076087

[9] George, Y.M., Aldeen, M., Garnavi, R.: Automatic psoriasis lesion segmentation in two-dimensional skin images using multiscale superpixel clustering. Journal of Medical Imaging **4**(4), 044004 (2017) https://doi.org/10.1117/1.JMI.4.4.044004

[10] Thanh, D.N.H., Hien, N.N., Prasath, V.B.S., Erkan, U., Khamparia, A.: Adaptive thresholding skin lesion segmentation with gabor filters and principal component analysis. (2020). https://api.semanticscholar.org/CorpusID:216342352

[11] Sivaraj, S., Malmathanraj, R., Palanisamy, P.: Detecting anomalous growth of skin lesion using threshold-based segmentation algorithm and fuzzy k-nearest neighbor classifier. Journal of Cancer Research and Therapeutics **16**(1), 40–52 (2020)

[12] Rawas, S., El-Zaart, A.: Hcet-g2: Dermoscopic skin lesion segmentation via hybrid cross entropy thresholding using gaussian and gamma distributions. In: 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS), pp. 1–7 (2019). https://doi.org/10.1109/ICDS47004.2019.8942339

[13] Celik, Y., Talo, M., Yildirim, O., Karabatak, M., Acharya, U.R.: Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. Pattern Recognition Letters **133**, 232–239 (2020) https://doi.org/10.1016/j.patrec.2020.03.011

[14] Hasan, M.K., Aleef, T.A., Roy, S.: Automatic mass classification in breast using transfer learning of deep convolutional neural network and support vector machine. In: 2020 IEEE Region 10 Symposium (TENSYMP), pp. 110–113 (2020). https://doi.org/10.1109/TENSYMP50017.2020.9230708

[15] Talo, M., Yildirim, O., Baloglu, U., Aydin, G., Acharya, U.: Convolutional neural networks for multi-class brain disease detection using mri images. Computerized Medical Imaging and Graphics **78** (2019) https://doi.org/10.1016/j.compmedimag.2019.101673

[16] Gore, D.V., Sinha, A.K., Deshpande, V.: Robust brain diseases classification using cnn and soft computing techniques. In: Kumar, A., Ghinea, G., Merugu, S., Hashimoto, T. (eds.) Proceedings of the International Conference on Cognitive

and Intelligent Computing, pp. 249–261. Springer, Singapore (2022)

[17] Talo, M., Yildirim, O., Baloglu, U.B., Aydin, G., Acharya, U.R.: Convolutional neural networks for multi-class brain disease detection using mri images. Computerized Medical Imaging and Graphics **78**, 101673 (2019) https://doi.org/10.1016/j.compmedimag.2019.101673

[18] Nguyen, H.Q., Lam, K., Le, L.T., Pham, H.H., Tran, D.Q., Nguyen, D.B., Le, D.D., Pham, C.M., Tong, H.T.T., Dinh, D.H., Do, C.D., Doan, L.T., Nguyen, C.N., Nguyen, B.T., Nguyen, Q.V., Hoang, A.D., Phan, H.N., Nguyen, A.T., Ho, P.H., Ngo, D.T., Nguyen, N.T., Nguyen, N.T., Dao, M., Vu, V.: VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations (2022)

[19] Gaál, G., Maga, B., Lukács, A.: Attention U-Net Based Adversarial Architectures for Chest X-ray Lung Segmentation (2020)

[20] Shaziya, H., Shyamala, K.: Pulmonary ct images segmentation using cnn and unet models of deep learning. In: 2020 IEEE Pune Section International Conference (PuneCon), pp. 195–201 (2020). https://doi.org/10.1109/PuneCon50868.2020.9362463

[21] Niyas, S., Pawan, S.J., Kumar, M.A., Rajan, J.: Medical Image Segmentation with 3D Convolutional Neural Networks: A Survey (2022)

[22] Yao, X., Wang, X., Wang, S.-H., Zhang, Y.: A comprehensive survey on convolutional neural network in medical image analysis. Multimedia Tools and Applications **81** (2020) https://doi.org/10.1007/s11042-020-09634-7

[23] Jafari, M., Karimi, N., Nasr-Esfahani, E., Samavi, S., Soroushmehr, S.M.R., Ward, K., Najarian, K.: Skin lesion segmentation in clinical images using deep learning. 2016 23rd International Conference on Pattern Recognition (ICPR), 337–342 (2016)

[24] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation (2015)

[25] Arora, P., Sharma, N., Bhatt, P., Saxena, A.: In: Srivastava, S., Khari, M., Gonzalez Crespo, R., Chaudhary, G., Arora, P. (eds.) Skin Lesion Segmentation Using Deep Convolutional Networks, pp. 111–122. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-76167-7_7 . https://doi.org/10.1007/978-3-030-76167-7$_7$

[26] Hafhouf, B., Zitouni, A., Megherbi, A.C., Sbaa, S.: An improved and robust encoder–decoder for skin lesion segmentation. Arabian Journal for Science and Engineering **47**, 9861–9875 (2022)

[27] Dai, D., Dong, C., Xu, S., Yan, Q., Li, Z., Zhang, C., Luo, N.: Ms red: A novel multi-scale residual encoding and decoding network for skin lesion segmentation.

Medical Image Analysis **75**, 102293 (2022) https://doi.org/10.1016/j.media.2021.102293

[28] Dong, Y., Wang, L., Li, Y.: Tc-net: Dual coding network of transformer and cnn for skin lesion segmentation. PLOS ONE **17**(11), 1–18 (2022) https://doi.org/10.1371/journal.pone.0277578

[29] Wu, H., Chen, S., Chen, G., Wang, W., Lei, B., Wen, Z.: Fat-net: Feature adaptive transformers for automated skin lesion segmentation. Medical Image Analysis **76**, 102327 (2022) https://doi.org/10.1016/j.media.2021.102327

[30] Tran, T.-T., Pham, V.-T.: Fully convolutional neural network with attention gate and fuzzy active contour model for skin lesion segmentation. Multimedia Tools and Applications **81**, 13979–13999 (2022)

[31] Alahmadi, M.D., Alghamdi, W.: Semi-supervised skin lesion segmentation with coupling cnn and transformer features. IEEE Access **10**, 122560–122569 (2022) https://doi.org/10.1109/ACCESS.2022.3224005

[32] Alahmadi, M.D.: Multiscale attention u-net for skin lesion segmentation. IEEE Access **10**, 59145–59154 (2022) https://doi.org/10.1109/ACCESS.2022.3179390

[33] Berseth, M.: ISIC 2017 - Skin Lesion Analysis Towards Melanoma Detection (2017)

[34] Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., Kittler, H., Halpern, A.: Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC) (2019)

[35] Valanarasu, J.M.J., Patel, V.M.: UNeXt: MLP-based Rapid Medical Image Segmentation Network (2022)

[36] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention U-Net: Learning Where to Look for the Pancreas (2018)

[37] Loshchilov, I., Hutter, F.: Decoupled Weight Decay Regularization (2019)