Enhanced Skin Lesion Segmentation: DeepLabV3 and U-Net with Spatial Attention Mechanisms

Su Myat Thwin
Department of Computer Science and Engineering
Ewha Womans University
Seoul 03760, Republic of Korea
sumyatthwin@ewhain.net

Hyun-Seok Park
Department of Computer Science and Engineering
Ewha Womans University
Seoul 03760, Republic of Korea
neo@ewha.ac.kr

***Abstract*— Accurate segmentation of skin lesions is crucial for early skin cancer detection and treatment. This study introduces a novel approach that enhances DeepLabv3 and U-Net architectures with spatial attention mechanisms to improve skin lesion segmentation. Using the ISIC dataset, the study shows that integrating attention mechanisms into these models significantly enhances their performance. DeepLabV3 captures multi-scale contextual information with its atrous spatial pyramid pooling (ASPP), while U-Net excels at capturing fine-grained details through its encoder-decoder structure. The spatial attention mechanisms enable the models to focus dynamically on the most relevant image regions, improving accuracy and robustness. The models are trained on the diverse ISIC dataset, using data augmentation techniques like rotation, scaling, and color jittering to handle variability in lesion appearance and limited data size. A composite loss function balancing pixel-wise accuracy and boundary precision guides the training process. Experimental results demonstrate that the attention-enhanced DeepLabV3 and U-Net models outperform their baseline versions achieving higher accuracy and Intersection over Union (IoU) scores. This study highlights the potential of attention-enhanced DeepLabV3 and U-Net models for skin lesion segmentation, suggesting they could be valuable tools for dermatologists in early skin cancer detection and treatment. The proposed models not only improve segmentation accuracy but also offer scalable solutions for clinical applications.**

Keywords—DeepLabV3, U-Net, Spatial Attention Mechanism, Skin Lesion Segmentation

I. Introduction

Skin cancer, particularly melanoma, poses a significant health risk due to its high mortality rate and potential for metastasis. Early detection and accurate diagnosis are critical for improving patient outcomes, as they allow for timely and effective treatment. Dermoscopic imaging has become a standard tool for dermatologists to examine skin lesions. However, manual analysis of dermoscopic images is labor-intensive and subject to inter-observer variability. Automated skin lesion segmentation, which delineates lesions from healthy skin, can enhance diagnostic accuracy and efficiency, making it a crucial step in the skin cancer detection pipeline.

Recent advancements in deep learning have revolutionized medical image analysis, particularly in the area of image segmentation. Convolutional neural networks (CNNs) such as DeepLabv3 and U-Net have emerged as powerful tools for segmentation tasks. DeepLabv3, with its atrous spatial pyramid pooling (ASPP), is adept at capturing multi-scale contextual information, making it effective for identifying features at various scales. U-Net, known for its encoder-decoder architecture, excels at capturing fine-grained details by integrating high-resolution features from the encoder with the upsampled features in the decoder.

Despite their strengths, both DeepLabv3 and U-Net face challenges in accurately segmenting skin lesions due to the high variability in lesion appearance, size, shape, and color. To address these challenges, attention mechanisms have been introduced into deep learning models. Attention mechanisms enhance feature representation by allowing the model to dynamically focus on the most relevant regions of the input image, thereby improving segmentation performance.

In this study, we propose a novel approach that enhances both DeepLabv3 and U-Net architectures with attention mechanisms to improve skin lesion segmentation accuracy. By leveraging the complementary strengths of these architectures and incorporating attention mechanisms, our method aims to achieve superior segmentation accuracy and robustness.

The ISIC (International Skin Imaging Collaboration) dataset is utilized for training and evaluating our models. The ISIC dataset is a comprehensive collection of dermoscopic images that represent a wide variety of skin lesions, making it an ideal resource for this study. To address the challenges of limited dataset size and variability in lesion appearance, we employ extensive data augmentation techniques, including rotation, scaling, and color jittering.

Our proposed method is evaluated using standard performance metrics, such as Accuarcy and Intersection over Union (IoU) scores, to demonstrate its effectiveness. Experimental results show that the attention-enhanced DeepLabv3 and U-Net models significantly outperform their baseline versions and other state-of-the-art methods in terms of segmentation accuracy. The attention mechanisms contribute to reducing false positives and negatives, further enhancing segmentation reliability. The main contributions of this paper are as follows:

Integration of Attention Mechanisms: We incorporate attention mechanisms into both DeepLabv3 and U-Net architectures, enabling the models to dynamically focus on relevant regions and improving their feature representation capabilities.

Comprehensive Evaluation: Using the ISIC dataset, we rigorously evaluate the performance of our models through extensive experiments and demonstrate significant improvements over baseline models and existing state-of-the-art methods.

Data Augmentation Techniques: We employ a variety of data augmentation techniques to address the challenges posed by the limited size and variability of the ISIC dataset, enhancing the generalizability and robustness of our models.

Clinical Relevance: Our study highlights the potential clinical application of the proposed models, providing a valuable tool for dermatologists in the early detection and treatment of skin cancer.

The rest of this paper is structured as follows: the literature reviews are presented at section 2, the theorical background is described at section 3, proposed segmentation system is illustrated at section 4, experimental results are discussed at section 5, and the last section concludes the proposed system and its future directions.

II. Literature Reviews

In this research [1], we use transfer learning using pre-trained weights from the Microsoft COCO dataset to develop a Mask R-CNN model for lesion segmentation, first performing image processing. We next train a Mask R-CNN model for the job of skin lesion classification using these trained weights, which are preserved and utilized in the process. The International Skin Imaging Collaboration 2018 (ISIC 2018) benchmark datasets are used in our investigations, and the same criteria are used for evaluation as in ISIC 2018. According to our findings, lesion classification obtains a balanced multiclass accuracy of 80%, while lesion border segmentation reaches an accuracy of 96%.

In this study [2], we used the ISIC2018 dataset, which contains 3,533 pictures of skin lesions spanning benign, malignant, nonmelanocytic, and melanocytic tumors, to train convolutional neural networks (CNNs) to detect malignant and benign tumors. First, preprocessing involved enhancing the photos with ESRGAN and then resizing, normalizing, and augmenting them. Based on the combined outcomes of several iterations, the CNN technique was used to categorize skin lesions. We then refined a number of transfer learning models, such as Inception-ResNet, ResNet50, and Inception-V3. The use of ESRGAN in the preprocessing stage is what makes this research unique. The outcomes of our specially created model were on par with the pre-trained models. The efficacy of our methodology was tested on the ISIC2018 dataset, yielding 83.2% accuracy with the CNN model and 83.7% accuracy with ResNet50, 85.8% with InceptionV3, and 84% with Inception-ResNet.

We provide a sophisticated skin lesion segmentation model in this paper that is built on a modified conditional generative adversarial network (cGAN) [3]. In the cGAN encoder, our method incorporates a new factorized channel attention (FCA) block that combines residual 1-D kernel factorized convolution with a channel attention mechanism. This lowers computing load and improves the separation between lesion and non-lesion features. In order to create scale-variant filters for a scale-invariant representation, we additionally use a multi-scale input technique. When tested on skin lesion challenge datasets from ISBI2016, ISBI2017, and ISIC2018, the model outperforms state-of-the-art algorithms in terms of Dice coefficient and intersection over union (IoU) score.

Significant progress has been achieved in automatic skin lesion classification by Deep Convolutional Neural Networks (DCNNs), which is important for improving diagnostic accuracy in communities with limited access to specialists. The combination of AI and web-based dermoscopic imaging offers a viable and approachable technique for the investigation of skin lesions in the future. A plugin with AI capabilities can greatly benefit the dermatological community. In this publication [4], we report our work on automated melanoma region segmentation in dermoscopic pictures using DCNNs, training and evaluating them on the public dataset HAM10000. Our goal is to create a web application that gives lab technicians and general practitioners likely diagnosis for skin lesions. This automation will speed up the process of follow-up diagnosis and treatment by making it easier to identify high-risk individuals quickly.

In this work [5], we provide a deep learning-based method for segmenting melanoma. Utilizing an altered U-net network in conjunction with post-processing methods, our approach shows excellent lesion segmentation efficacy. PH2 and DermIS are two public datasets that we used for our tests. On PH2 we obtained an average Dice coefficient of 0.933, and on DermIS we obtained an average of 0.872. Our suggested strategy is extremely promising, surpassing existing approaches with remarkable outcomes and excelling high-performance procedures documented in the literature.

In this paper [6], we introduce a convolutional deep neural network, called SC-DeLeNet, which accomplishes two main tasks: (i) using a segmentation sub-network, it segments lesion regions from unaffected skin tissue in dermoscopic images; (ii) using transferred parameters from the segmentation sub-network, it classifies each image according to its medical condition. A 'Classification Feature Extraction' mechanism in the classification sub-network makes use of training segmentation feature maps for lesion prediction, while the segmentation sub-network uses an EfficientNet-B4 backbone as the encoder. The 'Feature Coalescing Module' in the classification architecture integrates features from the encoder and decoder. Meanwhile, the '3D-Layer Residuals' block improves classification by establishing a parallel pathway for low-dimensional, high-variance features. After fine-tuning on a publicly available dataset, the mean accuracy for classification was 0.9103, outperforming traditional classifiers, while the mean Dice score for segmentation was 0.9494, exceeding current techniques.

We presented ResBCU-Net, a CNN-based neural network intended for medical picture segmentation, in this research [7]. With the addition of residual blocks, batch normalization, and bi-directional ConvLSTM, ResBCU-Net expands upon the U-Net design. Furthermore, we suggest a more sophisticated variant, ResBCU-Net (d = 3), which makes use of densely connected layers in its bottleneck area. We used the ISIC 2018 dataset, a publicly accessible collection of 2,594 photos of melanoma skin cancer, to train and assess our neural network. According to our findings, ResBCU-Net outperforms other cutting-edge techniques in image segmentation accuracy.

This paper [8] introduced MFSNet (Multi-Focus Segmentation Network), an AI framework for deep learning-based supervised skin lesion segmentation. MFSNet begins by preprocessing raw RGB photos of skin lesions in order to eliminate noise and artifacts. It uses the Parallel Partial Decoder (PPD) module to construct a global segmentation mask after using the Res2Net backbone to extract deep features. To generate the final segmentation output, the network incorporates multi-scale maps and convolutional features using Reverse Attention (RA) and Boundary Attention (BA) modules. Tests conducted on the ISIC 2017 and HAM10000 datasets demonstrate how robust and effective MFSNet is, outperforming existing state-of-the-art techniques.

For CNN-based melanoma identification, this study [9] presented the use of Self-attention Progressive Growing of

GANs (SPGGANs) to produce high-resolution 256 x 256 skin lesion images. Because lesion features vary and training is inconsistent, traditional GANs have difficulty producing high-resolution images. In order to ensure precise and consistent characteristics throughout the image, SPGGAN uses self-attention to overcome these issues. To increase stability, the Two-Timescale Update Rule (TTUR) is used (SPGGAN-TTUR). Tests conducted on the HAM10000 dataset demonstrate that SPGGAN-TTUR greatly improves the skin lesion classifiers' sensitivity (recall). In particular, sensitivity increases by 2.5% over data supplemented using conventional techniques and by 5.6% over non-augmented data. Sensitivity gains for melanoma detection are 13.8% and 8.6% higher than the top conventional DA methods and non-augmented approaches, respectively.

In this research [10], full resolution convolutional networks (FrCN), a revolutionary segmentation technique, are presented. Without the requirement for pre- or post-processing, the FrCN technique learns high-resolution characteristics for every pixel in the incoming data directly. Two publicly available datasets, PH2 and the ISBI 2017 Challenge, were used to assess the technique When compared to well-known segmentation methods like FCN, U-Net, and SegNet, FrCN fared better. It achieved 84.79% and 95.08% on the PH2 dataset, and an average Jaccard index of 77.11% and an overall accuracy of 94.03% on the ISBI 2017 test dataset. In terms of segmentation accuracy, FrCN outperformed FCN, U-Net, and SegNet by 1.31%, 3.89%, and 2.27%, and in the Jaccard index, by 4.94%, 15.47%, and 7.48%, respectively. In addition, FrCN performed better in the ISBI 2017 dataset than FCN, U-Net, and SegNet, with segmentation accuracy of 95.62% for benign cases, 90.78% for melanoma patients, and 91.29% for instances with seborrheic keratosis.

III. THEORICAL BACKGROUND

This section presents the background theories for this segmentation system.

A. Transfer Learning

A machine learning technique called transfer learning involves applying a model that has been trained on one job to a related task in a different domain [11]. It makes use of information from a source task to enhance performance and lessen the requirement for in-depth training on the new work, which is particularly helpful when there is a shortage of labeled data. For the target job, pre-trained models like as ImageNet or BERT extract features, and the new classifier's parameters are the only things that need to be modified. Further model adaptation can be achieved by fine-tuning some layers with a lower learning rate. This method improves performance and generalization, especially when there is a lack or high cost of labeled data for the intended task.

B. DeepLabV3

DeepLabV3 [12] is a Deep Neural Network (DNN) architecture that categorizes individual pixels into predetermined groups in order to perform semantic image segmentation. Semantic segmentation offers pixel-level knowledge, in contrast to object detection. To capture hierarchical characteristics, DeepLabV3 uses a backbone architecture similar to MobileNetV2 or Xception. It then expands the receptive field without raising parameters by using atrous (dilated) convolutions. In order to capture multi-scale information, it also incorporates the Atrous Spatial Pyramid Pooling (ASPP) module, which combines numerous parallel atrous convolutions with varying dilation rates. This improves segmentation precision by enabling DeepLabV3 to make precise, context-aware predictions. Its design is presented at Figure 1.

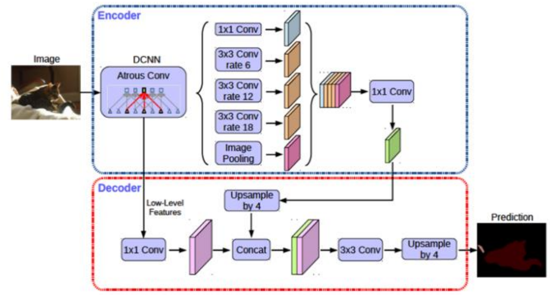

Fig 1. DeepLabV3

C. U-Net

A deep neural network architecture called U-Net [13] was unveiled in 2015 with the purpose of semantic segmentation—particularly in the context of medical picture analysis. The two approaches are the contracting path (encoder), which uses convolutional layers and ReLU activation to extract features and capture context, and the expanding path (decoder), which upsamples feature maps to restore spatial resolution. To preserve fine features and avoid disappearing gradients, skip connections are used to connect the encoder and decoder, while a bottleneck layer joins these pathways. The last layer generates a probability map for binary segmentation tasks using a sigmoid activation function. Its architecture is illustrated at Figure 2.

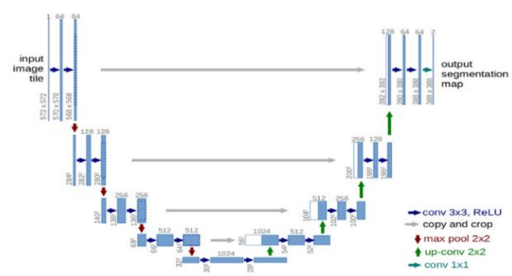

Fig 2. U-Net

D. Spatial Attention Mechanism

By creating attention maps based on learned weights or filters, a spatial attention module in deep learning improves a model's capacity to concentrate on significant spatial regions within input data [14]. With the help of these maps, the network is able to dynamically modify its emphasis, highlighting important elements and minimizing unimportant ones. These modules provide hard (discrete) or soft (continuous) attention, compute attention weights for spatial locations, and integrate into a variety of neural network topologies, including transformers and CNNs. They are essential for tasks requiring spatial localization, like object detection and semantic segmentation, as they efficiently guide the network's attention, increasing task efficiency and accuracy. All things considered, spatial attention modules greatly improve the interpretability and performance of deep learning models.

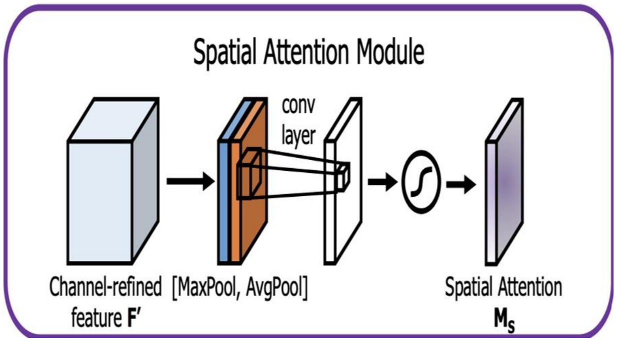

Fig 3. Spatial Attention Module

IV. System Design

Using a spatial attention module and deep U-Net and DeepLabv3 models, this system creates a skin lesion segmentation system based on a skin lesion dataset. Figure 4 describes the flow diagram of the system. To collect images from the International Skin Imaging Collaboration (ISIC) [15] dataset on Kaggle, which consists of 438 melanoma images, 376 basal cell carcinoma images, and 181 squamous cell carcinoma images, a systematic approach is followed. Initially, polyline annotation is conducted using LabelMe, a tool designed for precise annotation of medical images. Each image is annotated to outline the boundaries of skin lesions, capturing detailed spatial information crucial for subsequent segmentation tasks. Following annotation, JSON files containing annotation data are processed to generate mask images, where pixels inside lesion boundaries are labeled as foreground, while outside pixels remain background.

The dataset is then split into training and validation sets using an 80%-20% ratio to ensure both sets are representative of the overall distribution of melanoma, basal cell carcinoma, and squamous cell carcinoma images. Images are resized to a standardized 256x256 resolution to facilitate uniformity in input dimensions across the dataset, ensuring compatibility with most deep learning models. Subsequently, pixel intensities in both RGB images and corresponding mask images are normalized. RGB images are typically normalized to have values between 0 and 1, while mask images are binarized to maintain consistency in pixel values for effective model training. Once preprocessing is complete, the dataset is loaded into the training and validation pipelines, ensuring proper batching and shuffling to enhance model generalization. The data is converted into tensor format suitable for training, leveraging frameworks TensorFlow to expedite computations on GPUs. With the data prepared, a segmentation model, such as U-Net or DeepLabv3, is selected and trained on the training set. During training, the model iteratively learns to predict accurate segmentation masks that delineate skin lesions from background.

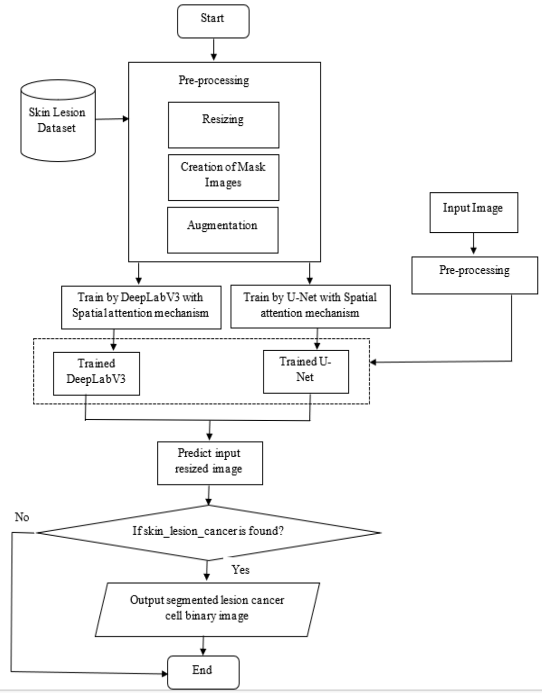

Fig 4. System Flow Diagram

Combining segmentation with spatial attention mechanisms for skin lesion segmentation using U-Net and DeepLabv3 involves integrating attention modules to enhance the models' ability to focus on relevant features within the input images. In the context of U-Net, which is renowned for its encoder-decoder architecture, spatial attention can be incorporated at various stages. This typically includes inserting attention blocks after each convolutional layer to selectively emphasize informative spatial regions while suppressing less relevant areas. The spatial attention module, often implemented as a combination of average and max pooling followed by convolutional layers and sigmoid activation, allows the network to dynamically adjust its focus during both encoding and decoding phases.

Conversely, DeepLabV3, built upon a deep convolutional neural network (CNN) backbone, ResNet leverages dilated convolutions for capturing multi-scale contextual information. Integrating spatial attention into DeepLabv3 involves embedding attention mechanisms directly within the network structure, augmenting its inherent capability to exploit both local and global context for precise segmentation. Here, attention modules can be strategically placed to refine feature maps before classification, enabling the model to allocate more attention to crucial spatial features crucial for accurate lesion segmentation.

In both architectures, the integration of spatial attention enhances segmentation performance by effectively guiding the models to focus on critical details while reducing noise and irrelevant information. This approach not only improves the models' ability to delineate skin lesions accurately but also enhances their interpretability by highlighting regions of interest within medical images, crucial for clinical diagnosis and treatment planning. Fine-tuning the parameters of these attention mechanisms and adapting them to the specific characteristics of skin lesion datasets further optimizes segmentation outcomes, ensuring robust and reliable performance in medical image analysis tasks.

In order to measure the segmentation performance of the trained model, measures such as accuracy and Intersection over Union (IoU) are computed using the validation set. The capacity of the model to precisely recognize and segment various kinds of skin lesions is further validated by the visualization of segmentation results and qualitative analysis. This comprehensive workflow ensures robust and reliable segmentation of skin lesions from medical images, supporting applications in dermatology and clinical decision-making.

V. Experimental Results

This method uses the ISIC dataset (a skin lesion dataset from Kaggle) to segment photos of melanoma, squamous cell carcinoma, and basal cell carcinoma for classification. Important measures including accuracy and the Intersection over Union (IoU) coefficient are used to assess the system's performance. A unique loss function called unet3p_hybrid_loss is used by U-Net and DeepLabV3. It combines Jaccard loss, SSIM loss (Structural Similarity Index), and Focal loss. This hybrid loss function optimizes spatial overlap, structural similarity, and concentrated attention on difficult-to-classify data, hence improving the model's capacity to segment skin lesions properly. Figure 5, and Figure 6 show the segmentation with DeepLabV3, and U-Net by utilizing spatial attention mechanism.

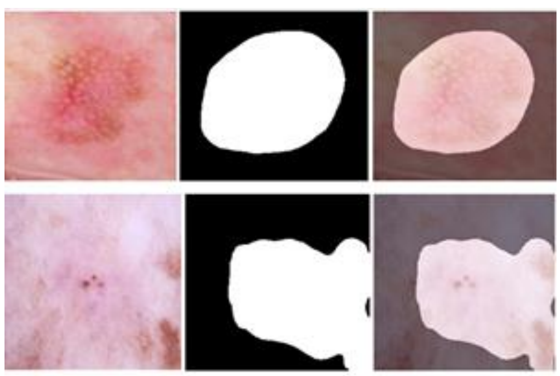

Fig 5. Segmentation with DeepLabV3 and Spatial Attention Mechanism (a) Input Image (b) Predicted masks (c) Overlay masks

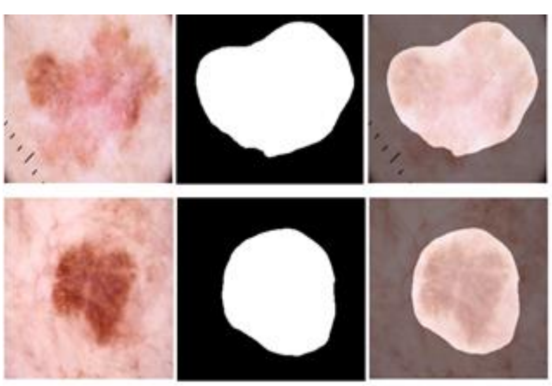

Fig 6. Segmentation with U-Net and Spatial Attention Mechanism (a) Input Image (b) Predicted masks (c) Overlay masks

Table 1 displays the comparison performance findings between U-Net and DeepLabV3. The accuracy outcomes of U-Net and DeepLabV3 are shown in Figure 7. Figure 8 presents U-Net and DeepLabV3 IoU results.

TABLE I. Performance Results Without Attention Mechanism

	IoU	**Accuracy**
DeepLabV3	0.76	0.80
U-Net	0.59	0.78

TABLE II. Performance Results With Attention Mechanism

	IoU	**Accuracy**
DeepLabV3	0.79	0.83
U-Net	0.62	0.80

Based on the results, the system that uses DeepLabV3 with a spatial attention mechanism outperforms U-Net with a spatial attention mechanism in terms of accuracy and IoU outcomes. As a result, employing an attention mechanism improves system performance compared to not utilizing one.

VI. Conclusion

This study investigated DeepLabV3 and U-Net, equipped with spatial attention mechanisms, for skin lesion segmentation using the ISIC dataset. DeepLabV3 utilized atrous convolutions and the ASPP module to capture contextual information effectively, enhanced further by a spatial attention mechanism for precise lesion segmentation. Similarly, U-Net, with spatial attention mechanisms and skip connections, demonstrated proficiency in delineating lesion boundaries and preserving fine details. Evaluation on the ISIC dataset highlighted competitive performance metrics like accuracy and IoU scores, emphasizing the role of attention mechanisms in enhancing segmentation accuracy by emphasizing relevant features and reducing noise. Future research may optimize these architectures for real-time applications and expand their applicability across diverse lesion types, aiming to improve dermatology diagnostics.

As a future work, we will explore techniques such as model pruning and quantization to reduce computational complexity and improve inference speed, which are crucial for real-time deployment. Additionally, I will discuss the implementation of lightweight spatial attention mechanisms that can balance accuracy with processing speed, making the models more efficient for clinical use. I will also consider the use of hardware accelerators like GPUs or AI chips to further optimize real-time performance. Lastly, I intend to include experiments on real-time data streams to validate the practicality and responsiveness of these models in real-world environments. These additions will aim to make the research more comprehensive and applicable to real-time clinical scenarios.

References

[1] L. Lin, and S. Wonseok, “Deep Learning and Transfer Learning for Skin Cancer Segmentation and Classification,” 2021 IEEE 21st International Conference on Bioinformatics and Bioengineering

(BIBE), IEEE, Kragujevac, Serbia, 25-27 October 2021, pp. 83398 – 83414.

[2] W. Gouda, N. U. Sama, and G. A. Waakid, "Detection of Skin Cancer Based on Skin Lesion Images Using Deep Learning," *Volume 10*, MDPI, 24 June 2022.

[3] V. K. Singh, M. Abdel-Nasser, H. A. Rashwan, F. Akram, N. Pandey, A. Lalande, B. Presles, S. Romani, and D. Puig, "FCA-Net: Adversarial learning for skin lesion segmentation based on multi-scale features and factorized channel attention", IEEE Access 2019, 7, 130552–130565.

[4] A. K. Waweru, K. Ahmed, Y. Miao, and P. Kawan, "Deep Learning in Skin Lesion Analysis Towards Cancer Detection", 2020 24th International Conference Information Visualisation (IV), IEEE, Melbourne, Australia, 07-11 September 2020, pp. 740-745.

[5] R. L. Araujo, R. A. Rabelo, J. P. C. Rodrigues, and R. V. Silva, "Automatic Segmentation of Melanoma Skin Cancer Using Deep Learning", 2021 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM), IEEE, Shenzhen, China, 01-02 March 2021, pp. 1-6.

[6] Md.Jahin Alam, Mir Sayeed Mohammad, Md Adnan Faisal Hossain, Ishtiaque Ahmed Showmik, Munshi Sanowar Raihan, Shahed Ahmed, and Talha Ibn Mahmud, "S^2C-DeLeNet: A parameter transfer based segmentation-classification integration for detecting skin cancer lesions from dermoscopic images", Computers in Biology and Medicine, Volume 150, November 2022, 106148.

[7] Noor Badshah, and Asif Ahmad, "ResBCU-Net: Deep learning approach for segmentation of skin images", Biomedical Signal Processing and Control, Volume 71, Part A, January 2022, 103137.

[8] Hritam Basak, Rohit Kundu, and Ram Sarkar, "MFSNet: A multi focus segmentation network for skin lesion segmentation", Pattern Recognition, Volume 128, August 2022, 108673.

[9] Ibrahim Saad Aly Abdelhalim, Mamdouh Farouk Mohamed, and Yousef Bassyouni Mahdy, "Data augmentation for skin lesion using self-attention based progressive generative adversarial network", Expert Systems with Applications, Volume 165, 1 March 2021, 113922.

[10] Mohammed A. Al-masni, Mugahed A. Al-antari, Mun-Taek Choi, Seung-Moo Han, and Tae-Seong Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks", Computer Methods and Programs in Biomedicine, Volume 162, August 2018, Pages 221-231.

[11] J. Brownlee, "A Gentle Introduction to Transfer Learning for Deep Learning", September 16, 2019.

[12] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "DeepLabv3: Rethinking Atrous Convolution for Semantic Image Segmentation", 2018.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", 2015.

[14] https://paperswithcode.com/method/spatial-attention-module

[15] https://www.kaggle.com/datasets/nodoubttome/skin-cancer9-classesisic