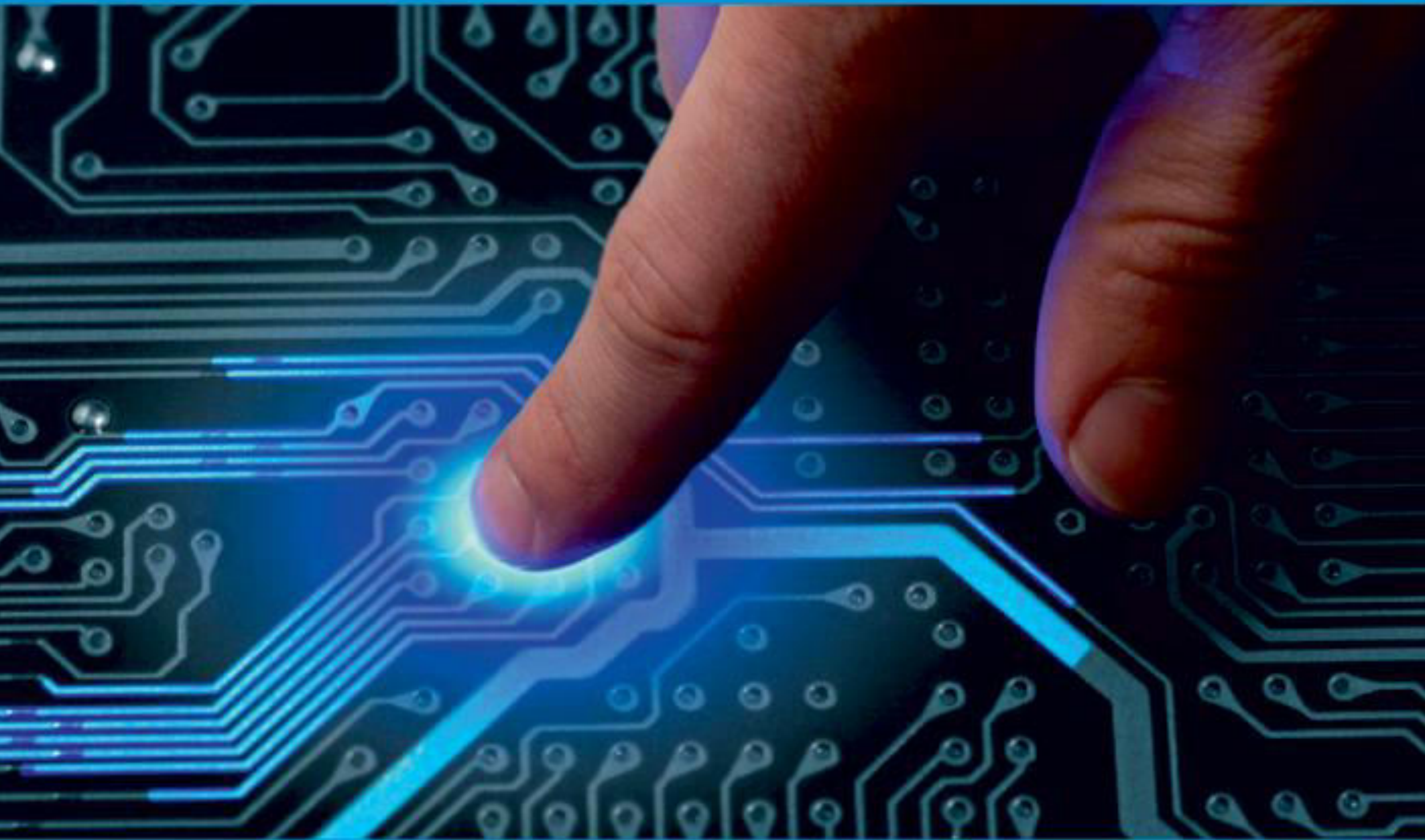
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204161 |

A Van-Based Multi-Scale Cross-Attention Mechanism for Skin Lesion Segmentation Network

K.GANDHIMATHI, GOKUL V, HARIHARASUDHAN D, KARTHIKEYAN G,

Assistant Professor, Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

ABSTRACT: Skin cancer is one of the most common types of cancer. Excessive exposure to ultraviolet radiation can cause genetic mutations in superficial skin cells, leading to uncontrolled growth and the development of skin cancer. Melanoma, which arises from melanocyte cells, grows rapidly and poses a higher risk than non-melanoma skin cancer. Early treatment of patients can greatly reduce the mortality rate of melanoma. Subsequently, a lightweight and efficient channel attention (ECA) module is introduced during the encoder's feature extraction stage. The attention module assigns suitable weights to channels through attention learning and effectively captures inter-channel interaction information. Finally, the densely connected atrous spatial pyramid pooling module (DenseASPP) module is inserted between the encoder and decoder paths. This module integrates dense connections and ASPP, as well as multi-scale information fusion, to recognize lesions of varying sizes. The experimental studies in this paper were constructed on two public skin lesion datasets, namely, ISIC-2018 and ISIC-2017. The experimental results show that our model is more accurate in segmenting lesions of different shapes and achieves state-of-the-art performance in segmentation. In comparison to UNet3+, the proposed REDAUNet model shows improvements of 2.01%, 4.33%, and 2.68% in Dice, Spec, and mIoU metrics, respectively. These results suggest that REDAUNet is well-suited for skin lesion segmentation and can be effectively employed in computer-aided systems.

KEYWORDS: skin lesions; image segmentation; deep learning; atrous convolution; UNet3+

I.INTRODUCTION

Melanoma is a relatively aggressive form of skin malignancy that accounts for only about 1% of skin cancers but causes most deaths. There are currently more than 132,000 cases of melanoma skin cancer worldwide each year. The accuracy of diagnosis by patients and dermatologists by using visual inspection is only about 60%. In addition, the shortage of dermatologists per capita prompted the need for computer-aided methods in detecting skin cancer. The American Cancer Society's 2022 homegrown statistics estimate that there will be approximately 99,780 new melanoma cases (about 57,180 cases in men and 42,600 cases in women) and an estimated 7650 deaths from melanoma (about 5080 men and 2570 women). In addition to this, there are other types of cancer. For example, colon cancer, lung cancer, stomach cancer, and so on are still the leading causes of human suffering and death.

With the development of computer vision technology and artificial intelligence, image analyses have been widely used in various scene-parsing tasks. Medical image analyses play vital roles in computer-aided diagnosis and detection. The amount of medical image data acquired is growing faster than the available human expert interpretation. Therefore, automated segmentation techniques are desired in helping physicians achieve accurate and timely imaging-based diagnoses. However, due to insufficient original training samples of medical images or the lack of a clear demarcation line between some subtle lesion areas and normal tissues and organs. They are making the task of skin lesion segmentation more difficult.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204161 |

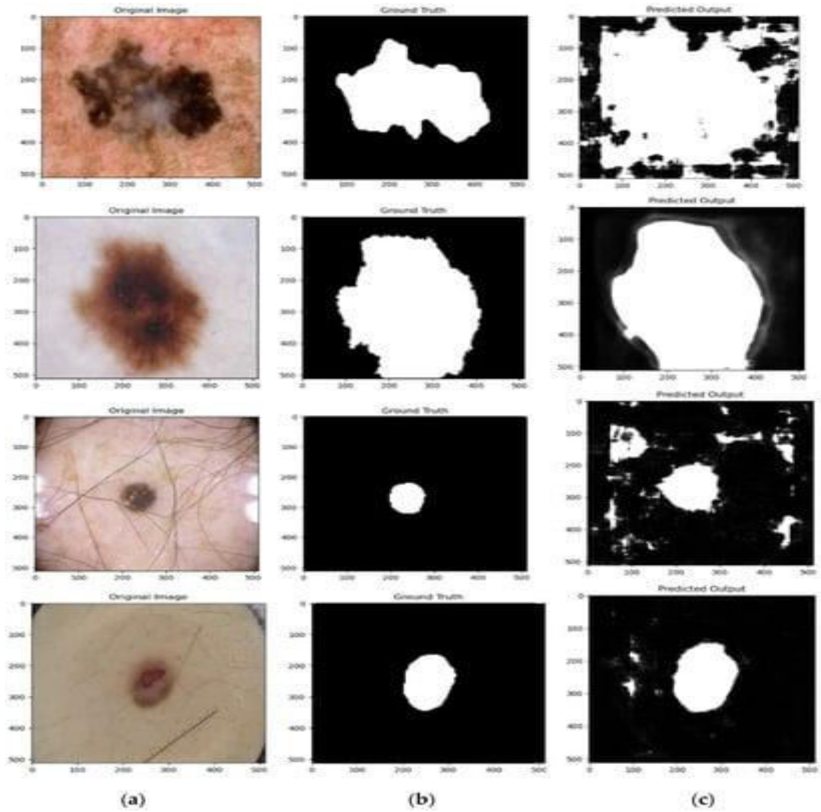


Fig 1: skin lesion segmentation

In recent years, with in-depth research on deep learning theory, convolutional neural network-based deep learning methods for image recognition and classification have shown excellent performance [10,11,12], including the recently popular BP neural network algorithm for image processing [13]. Moreover, with respect to multi-level dilated residual network [14] for processing skin lesions and MRIs, Long et al. [7] proposed an FCN architecture based on CNNs to solve the semantic level image segmentation problem by performing end-to-end pixel-level classification of the input raw images. Most medical images are large, so the feature vector obtained by training using raw images is large. It also has high requirements for computer performances, leading to substantial computational costs. Fischer et al. [15] proposed U-Net, which consists of mutually symmetric systolic and dilated paths. Among them, the systolic path is used to obtain context information and the dilated path is used for precise localizations. In the dilation path, feature vectors are fused with corresponding low-level features to add multi-scale information. Finally, the overlap tile strategy alleviates the computational resource issue. High IoU values of 0.9203 and 0.7756 were obtained on the PhC-U373 and DIC-HELa datasets, respectively. Later, Zhou et al. [16] proposed a new architecture U-Net++ that enables flexible feature fusion by redesigning multiple dense skip connections, reducing the semantic gap between feature representations and encoder sub-networks. Moreover, the multi-scale feature aggregation of U-Net++ can synthesize the segmentation results step by step, thus improving the accuracy and accelerating the convergence speed of the network.

II.RELATED WORK

As the complexity of computer vision tasks and task demands increase, deeper [17] convolutional neural networks are required for feature extraction. As a result, the gradient disappearance problem sometimes occurs during feature propagation. Huang et al. [18] proposed Dense Net, which not only alleviates the gradient vanishing problem but also enhances the feature's propagation and can dramatically reduce the number of parameters. Subsequently, Zhang et al. [19] proposed Res-Net, which utilizes skip connections (Identity mapping) to alleviate the vanishing gradient problem while increasing the

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |
|| Volume 12, Issue 4, April 2024 ||
| DOI: 10.15680/IJIRCCE.2024.1204161 |

network's depth. The authors of [20] proposed PSP-Net using a pyramid pooling module to aggregate global contextual information from different regions to increase the target receptive field. Later, Ibtehaz et al. [21] proposed MultiResUNet to introduce contextual multi-scale information into the U-Net architecture via different residual modules, adding local detail information.

However, FCNs and CNN models face the same issue: a lack of long-term global correlation modeling capabilities. The main reason is that CNN extracts local information simply and cannot measure global relevance efficiently. A transformer [22] is an essential model in natural language processing, and was used initially to improve NMT (neural machine translation) models using attention mechanisms. The transformer network has a cleaner structure and is quicker in training and inferencing. The transformer focuses on extracting global information but weakens local information, so it also has some disadvantages in medical image segmentation tasks. How to properly highlight foreground information, weaken background information, and how to better jointly model local information and global correlation dependence become focuses of the study. The authors of [23] combined the transformer structure with the U-Net model, using the transformer's powerful encoding ability and U-Net's local localization ability to complete the segmentation of multiple abdominal organs and the heart. Extensive experiments demonstrate that TransU-Net outperforms the original U-Net architecture in various image classification tasks.

Based on existing approaches, in this paper, we propose a novel CNN for medical image segmentation. The training results on three different datasets outperformed the current state-of-the-art models in three main areas of work:

- Standard convolution is replaced by dilated convolution; original image information of varying resolution sizes is introduced into the encoder at all levels;
- Feature fusion at each level uses hybrid attention for detail enhancement of feature vectors in both channel and spatial dimensions;
- Slicing experiments are conducted to verify the contribution of HRA, dilation convolution, and cross-validation pieces relative to the MHAU-Net model.

Image pre-processing Since the lesion areas in the original dermoscopic images vary in shape, size, and pixel intensity, some lesion areas are hidden under human hair or shadows, which will inevitably affect segmentation results, thereby reducing the generalization ability of the model. Therefore, to minimize the impact of these factors on the model segmentation performance, we introduce an image preprocessing method.

We used a morphological manipulation approach to remove artifacts from the original dermoscopic images. First, the input RGB image is converted into a grayscale image. The morphological operation with black hat transform is used [24], followed by artifact removal using a thresholding operation (as shown in **Figure 2** see legend information for details). We continuously adjusted the experimental parameters and selected a cross-shaped two-dimensional array of size 25×25 as the structural element, which has the middle row and column consisting of 1 and the remaining elements composed of 0. All images are resized to a shape of 256×256 using bilinear interpolation to achieve faster convolution operations and to solve the excessive memory consumption problem.

III.METHODS

In general, the method for increasing the receptive field and reducing the amount of computation in deep neural networks is down-sampling. However, down-sampling sacrifices part of the spatial resolution and loses some information, which limits the effect of semantic segmentation. In contrast, atrous convolutions [30] enable effectively increasing target receptive field without increasing model parameters and without changing the size of the feature map. In addition, we introduce residual connectivity [10]. Residual connection not only reduces the complexity of model training to minimize overfitting but also prevents the gradient from vanishing. The RA convolutional network is proposed by combining the above two methods. In the RA module, we replace the standard convolution in the original CNN with dilated convolutions. On the one hand, the receptive field increases, and significant targets can be detected and segmented. On the other, the increased resolution compared with down-sampling can accurately locate the target. Combining residual connections can improve the mobility of information and prevent serious information loss. Significantly, RA can be integrated into other convolutional neural networks, which is a crucial reference for improving the propagation of feature vectors.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204161 |

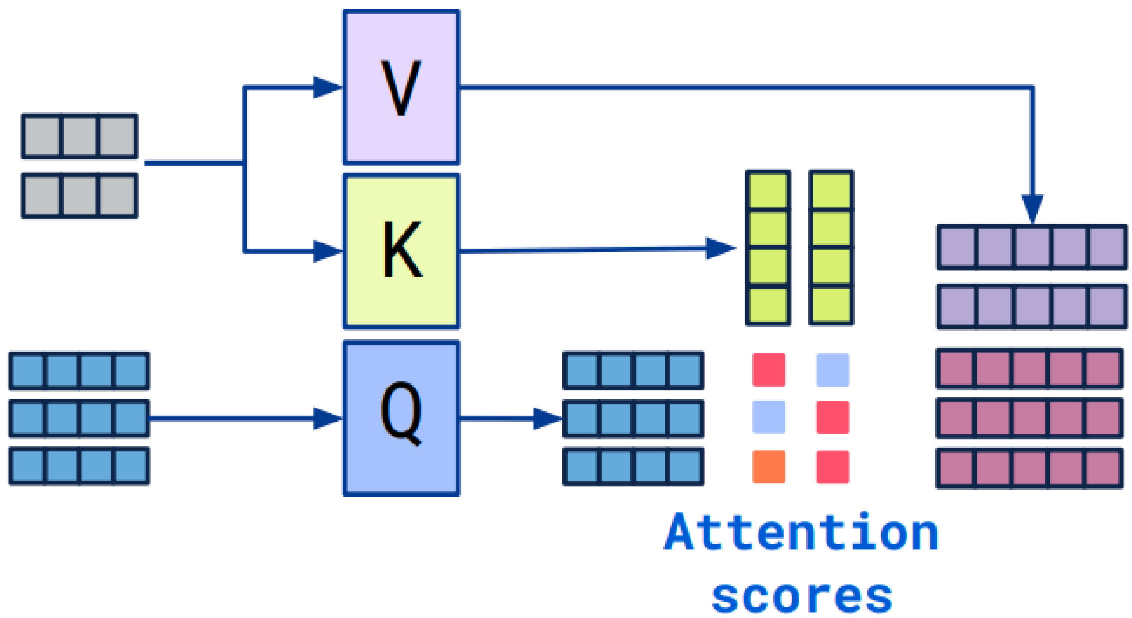


Fig 2: Self-attention

The encoding stage uses RA blocks with different dilation rates for feature information extraction (as shown Then, the feature vectors are input into HRA, and the dependencies between channels are established. Using the dependency relationship between feature channels, the feature representation of specific semantics can be improved to generate channel attention maps. The spatial attention module encodes a vast range of contextual information into local features, thus enhancing their expressive power. The spatial relations among the elements generate a spatial attention graph. HRA has powerful feature representation capabilities that can be integrated into other CNN architectures. However, frequently using channels and spatial attention mechanisms increases spatial and time complexities. The high resolution low-level features and the smaller field of perception of individual pixels enable the use of more fine-grained feature information to capture more small targets. The validation shows that increasing too many attention mechanisms does not bring about significant improvements but instead increases the training burden. Therefore, we choose to use attention mapping after three more low-level features of RA, R_3A, and R_4A.

IV.RESULT ANALYSIS

o evaluate MHAU-Net, we conducted experiments on three public medical image datasets. In this paper, data augmentation techniques, including vertical flip and transpose were used in advance for all datasets participating in the experiments. However, we do not establish the validation dataset. Since we use a cross-validation strategy, some data are randomly divided as the validation set in each training round. Cross-validation enables an increase in the randomness of the validation dataset and the training parameters are adjusted in time, thus effectively improving the generalization performance of the model.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204161 |

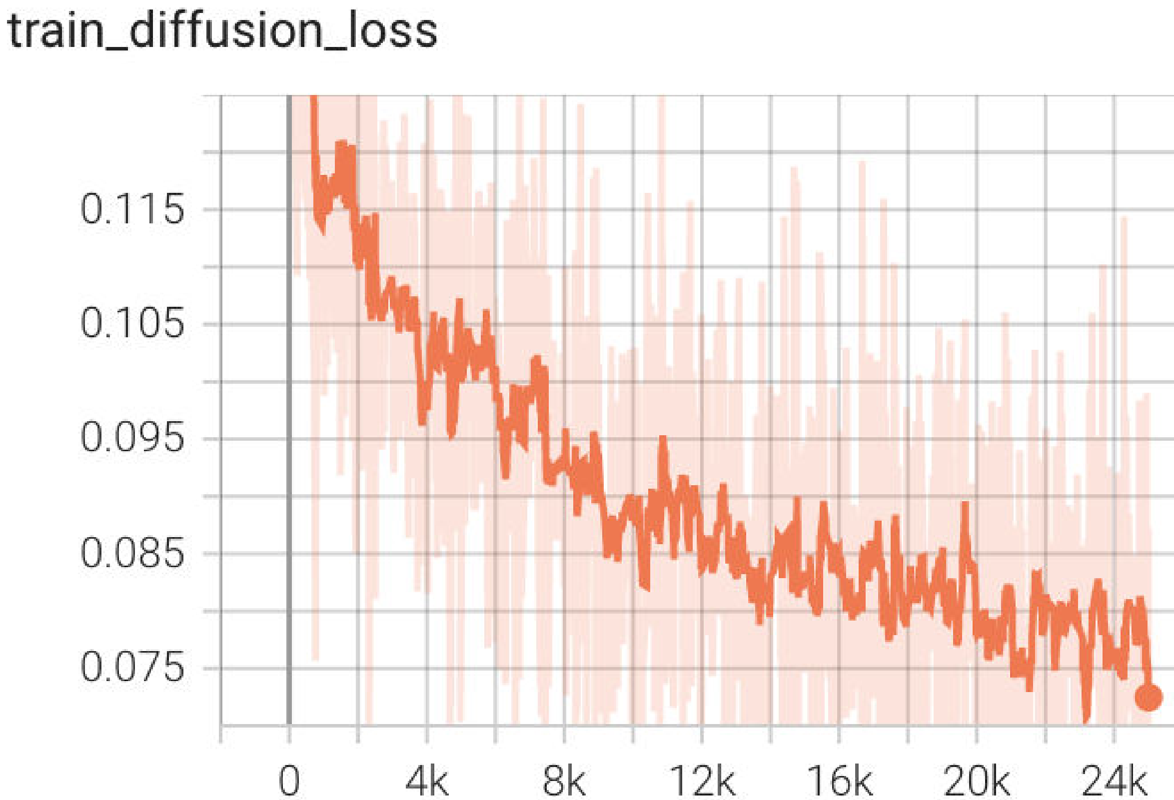


Fig 3: Result analysis

MHAU-Net architecture proposed in this paper achieved satisfactory results, it could be concluded that the segmentation maps generated by the MHAU-Net outperformed the other architectures in capturing the boundary information, demonstrating that the segmentation masks generated in the MHAU-Net showed more precise information in the target area than the existing models. The full convolutional network has more room for improvements in capturing skin lesion locations and edge details.

V.CONCLUSIONS

Inspired by U-Net and the attention mechanism, we propose MHAU-Net to address the need for the automated detection of lesion areas in dermoscopy and its related medical fields. MHAU-Net consists of four parts: multi-scale resolution input, hybrid residual attention (HRA), dilated convolution, and atrous spatial pyramid pooling. HRA fully utilizes the benefits of the attention mechanism to achieve feature enhancements and conducts slicing experiments to verify the contribution of network components to the overall network architecture. Use a 5-fold cross-validation strategy during training to improve the generalization performance of the model. We validated MHAU-Net on three datasets with better performance than the BA-Transformer and U-Net. To achieve the goal of model generalizability, the proposed architecture in this paper should be further investigated for improvements to obtain better segmentation results.

REFERENCES

1. Zhou, S.K.; Greenspan, H.; Davatzikos, C.; Duncan, J.S.; Van Ginneken, B.; Madabhushi, A.; Prince, J.L.; Rueckert, D.; Summers, R.M. A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises. *Proc. IEEE* **2021**, *109*, 820–838. [Google Scholar] [CrossRef]

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204161 |

2. Bai, W.; Suzuki, H.; Huang, J.; Francis, C.; Wang, S.; Tarroni, G.; Guitton, F.; Aung, N.; Fung, K.; Petersen, S.E.; et al. A population-based phenome-wide association study of cardiac and aortic structure and function. *Nat. Med.* **2020**, *26*, 1654. [**Google Scholar**] [**CrossRef**]
3. Mei, X.; Lee, H.-C.; Diao, K.-Y.; Huang, M.; Lin, B.; Liu, C.; Xie, Z.; Ma, Y.; Robson, P.M.; Chung, M.; et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nat. Med.* **2020**, *26*, 1224. [**Google Scholar**] [**CrossRef**] [**PubMed**]
4. Khened, M.; Kollerathu, V.A.; Krishnamurthi, G. Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *Med. Image Anal.* **2019**, *51*, 21–45. [**Google Scholar**] [**CrossRef**] [**PubMed**] [**Green Version**]
5. Tang, H.; Chen, X.; Liu, Y.; Lu, Z.; You, J.; Yang, M.; Yao, S.; Zhao, G.; Xu, Y.; Chen, T.; et al. Clinically applicable deep learning framework for organs at risk delineation in CT images. *Nat. Mach. Intell.* **2019**, *1*, 480–491. [**Google Scholar**] [**CrossRef**]
6. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [**Google Scholar**] [**CrossRef**] [**Green Version**]
7. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [**Google Scholar**] [**CrossRef**]
8. Khan, M.A.; Akram, T.; Zhang, Y.-D.; Sharif, M. Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework. *Pattern Recognit. Lett.* **2021**, *143*, 58–66. [**Google Scholar**] [**CrossRef**]
9. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929. [**Google Scholar**]
10. Hesamian, M.H.; Jia, W.; He, X.; Kennedy, P. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *J. Digit. Imaging* **2019**, *32*, 582–596. [**Google Scholar**] [**CrossRef**] [**Green Version**]
11. Li, X.; Chen, H.; Qi, X.; Dou, Q.; Fu, C.W.; Heng, P.A. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes. *IEEE Trans. Med. Imaging* **2018**, *37*, 2663–2674. [**Google Scholar**] [**CrossRef**] [**PubMed**]
12. Dolz, J.; Gopinath, K.; Yuan, J.; Lombaert, H.; Desrosiers, C.; Ben Ayed, I. HyperDense-Net: A Hyper-Densely Connected CNN for Multi-Modal Image Segmentation. *IEEE Trans. Med. Imaging* **2019**, *38*, 1116–1126. [**Google Scholar**] [**CrossRef**] [**PubMed**] [**Green Version**]
13. Huo, Y.; Ma, X.X. Image noise recognition algorithm based on BP neural network. In Proceedings of the 32nd Chinese Control and Decision Conference (CCDC), Hefei, China, 22–24 August 2020. [**Google Scholar**]
14. Gudhe, N.R.; Behravan, H.; Sudah, M.; Okuma, H.; Vanninen, R.; Kosma, V.M.; Mannermaa, A. Multi-level dilated residual network for biomedical image segmentation. *Sci. Rep.* **2021**, *11*, 14105. [**Google Scholar**] [**CrossRef**] [**PubMed**]
15. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015. [**Google Scholar**]
16. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet plus plus: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Trans. Med. Imaging* **2020**, *39*, 1856–1867. [**Google Scholar**] [**CrossRef**] [**Green Version**]