Journal of computer science and software applications
https://www.mfacademia.org/
ISSN:2377-0430
Vol. 4, No. 4, 2024

Advanced Medical Image Segmentation with Multi-Scale Feature Fusion and Attention Mechanisms

John Smith

Arizona State University

John0319@asu.edu

Abstract:

Skin diseases and lung inflammation present significant diagnostic challenges, necessitating precise and efficient methods to aid clinicians. The early detection of pulmonary nodules is particularly crucial, yet traditional diagnostic methods often lead to radiologist fatigue and increased error rates. Deep learning has emerged as a promising tool to enhance clinical diagnostics, but conventional techniques struggle with the accurate segmentation of small and irregularly shaped lesions, leading to reduced accuracy. This paper proposes a novel lesion image segmentation method based on the UNeXt architecture, integrating multi-scale feature fusion and attention mechanisms. The multi-scale feature fusion ensures comprehensive feature extraction, while the attention mechanism enhances the focus on critical regions, improving edge feature sensitivity. Additionally, a joint loss function is employed to optimize the model's performance. The proposed method aims to significantly improve the segmentation accuracy of pulmonary nodules, providing a robust tool for early diagnosis and treatment, thereby demonstrating the transformative potential of deep learning in clinical diagnostics.

Keywords:

UNeXt; Edge Feature Fusion; Multi-layer Perceptron; Medical Image Segmentation.

1. Introduction:

Skin diseases often occur in daily life and cause great trouble to patients. Doctors need to diagnose the diseased areas during their work to help patients recover as soon as possible. At the same time, lung inflammation is a serious disease that may be caused by long-term living habits and genetic mutations. Lung lesions initially manifest as inflammation, but over time, the accumulation of many diseases in the lungs may manifest as inflammation, so the initial diagnosis of pulmonary nodules is very important. When diagnosing pulmonary nodules in the early stage, the large number of patients usually leads to excessive fatigue of doctors. Therefore, doctors are required to repeatedly check CT images to better determine the area of pulmonary nodules and ensure that the diagnosis results are correct. In recent years, with the continuous development of deep learning, this technology has been widely used in clinical diagnosis to assist doctors in quickly detecting CT images and reducing misdiagnosis rates. However, due to the small size and irregular shape of the lesion area, traditional methods have certain limitations when segmenting images. They may not be able to accurately locate the nodule position, resulting in low model segmentation accuracy, and may even increase with the network model layer. Deepening loses some information, so that more features cannot be effectively obtained.

To this end, this paper proposes a lesion image segmentation method based on UNeXt combined with multi- scale feature fusion and attention mechanism. First, we introduce a multi - scale feature fusion mechanism based on the UNeXt An attention mechanism is introduced based on X t; finally, we use a joint loss function to improve the sensitivity of the network model to edge features.

2. Related Work

Recent advancements in deep learning, particularly in the field of medical image analysis, have paved the way for significant improvements in diagnosing and managing various diseases. Liu et al. conducted a comprehensive study on feature extraction utilizing convolutional neural networks (CNNs), emphasizing the model's efficiency in capturing intricate data patterns from raw inputs. Their findings indicated substantial improvements in accuracy and processing speed, establishing a benchmark for subsequent research in this domain [1]. Hu et al. explored the application of deep learning models for early warning systems in cardiovascular diseases. By leveraging extensive medical datasets, their research demonstrated that deep learning algorithms could predict potential cardiovascular events with high precision, underscoring the transformative potential of these models in preventative healthcare [2].

Yang et al. focused on the diagnosis of pulmonary nodules using deep learning models. Their study revealed that CNNs could effectively differentiate between benign and malignant nodules, thereby supporting radiologists in making more accurate diagnostic decisions [3]. This aligns closely with our research objectives, as accurate segmentation and detection of lung lesions are crucial for early intervention and treatment. Similarly, Xiao et al. investigated the classification of cancer cytopathology images, using breast cancer as a case study. Their research showcased the robustness of CNNs in medical image classification tasks, particularly in distinguishing cancerous cells from non-cancerous ones, thereby aiding in early cancer detection [4].

Sun et al. proposed optimization techniques for natural language processing (NLP) models using multimodal deep learning. They highlighted the integration of text and image data to enhance the model's understanding and processing capabilities, which is crucial for applications requiring comprehensive data analysis [5]. Yan et al. discussed the use of neural networks for survival prediction across various cancer types. Their work underscored the versatility of deep learning models in handling diverse datasets and providing accurate survival predictions, which are critical for patient prognosis and treatment planning [6].

Further advancements were made by Zhang et al., who introduced a multi-scale image recognition strategy based on CNNs. This approach improved recognition accuracy by analyzing images at multiple scales, thus capturing finer details and enhancing the overall performance of image recognition systems [7]. Mei et al. examined the efficiency optimization of large-scale language models in NLP tasks. Their findings indicated that deep learning techniques could significantly reduce computational costs while maintaining high performance, which is essential for deploying these models in real-world applications [8].

Xiao et al. explored the incorporation of attention mechanisms in deep learning models for mining medical textual data. Their study demonstrated that attention-enhanced models could better capture relevant information from vast amounts of text, improving the accuracy of medical data analysis [9]. Yan et al. investigated customized medical decision algorithms using graph neural networks. They provided insights into how these networks could model complex relationships in medical data, leading to more precise and personalized medical decisions [10].Gao et al. presented an enhanced encoder-decoder network for image semantic segmentation. Their work focused on reducing information loss during the segmentation process, thereby improving the fidelity and usability of segmented images in various applications [11].

Zhan et al. developed innovative techniques for recognizing time-related expressions using LSTM networks. Their study provided new methods for accurately processing temporal data, which is crucial for applications in natural language understanding and processing [12]. Lastly, Yang et al. introduced a novel image recognition method combining deep learning-generated adversarial networks with traditional algorithms. This hybrid approach enhanced the robustness and accuracy of image recognition systems, making them more resilient to adversarial attacks [13].

3. Data and methods

3.1 Network Architecture

In the research, in order to better realize the accurate real- time segmentation task and avoid the problem of losing local low-dimensional features through direct large-area upsampling, which leads to the loss of too many features on the segmentation boundary and the inability to restore complete edge information, only through features The graph superimposes feature information in the channel dimension to retain feature information[14]. This will cause the last few layers of feature maps to be too bloated, causing the model to require a large amount of calculations. Based on these problems, we adopt a multi-branch feature fusion network for medical image segmentation[15]. We first propagate contextual information to higher-resolution layers through progressive upsampling to obtain preliminary low-level semantic features. We avoid superimposing feature information in the channel dimension of UNet's series of related models, and choose the method of feature map multiplication to fuse features ; therefore, most of the feature information is well preserved, and boundary information can be effectively obtained , effectively reducing the number of failures. The designed skip link uses more detailed low-dimensional feature information as a supplement to feature fusion to ensure that the accuracy is slightly better than UNet[16] ,ResUNet++ [17]and other networks run much faster than other models; it also has the advantages of high training efficiency and strong generalization ability.

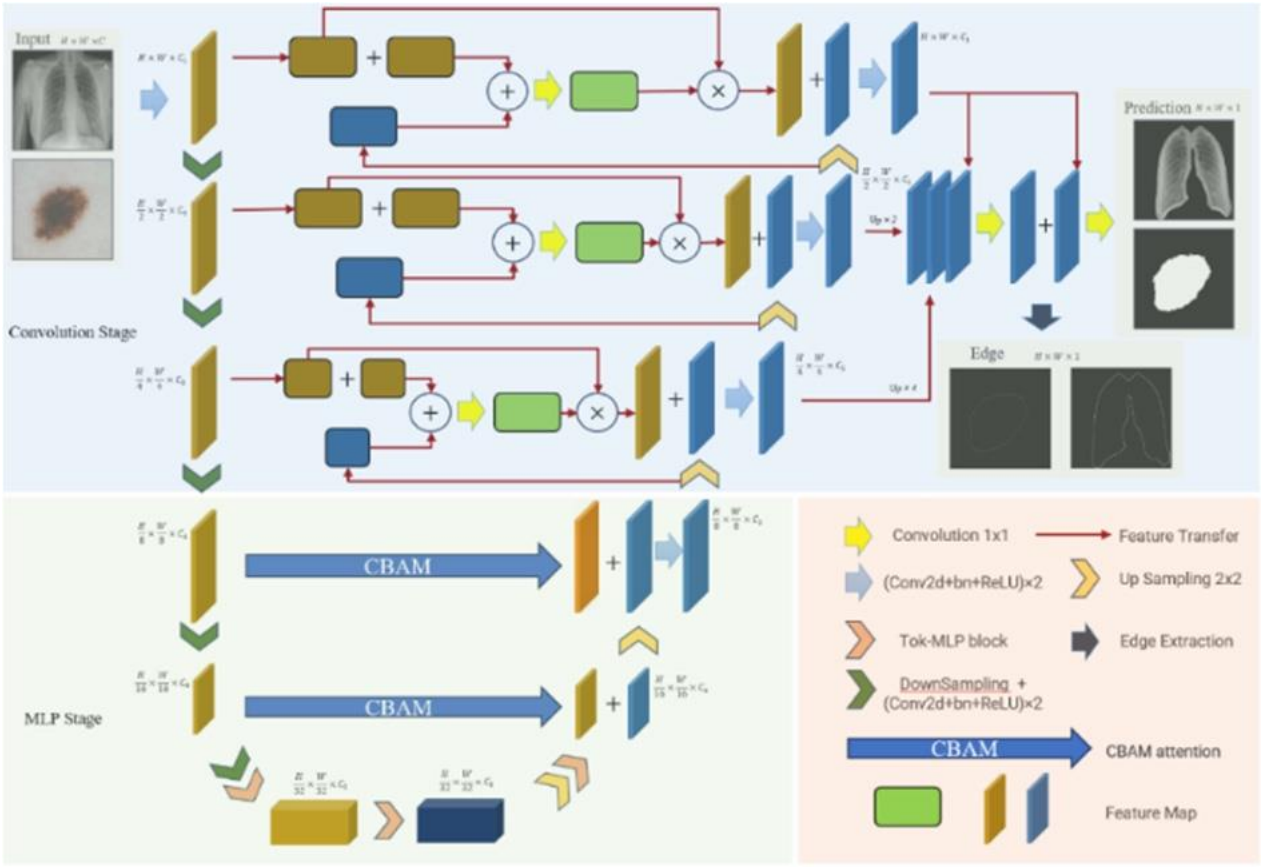

Figure 1. SC-UneXt structure

Overall framework Our network consists of three parts: encoder, multi-scale cross-skip connection and decoder. In order to better integrate semantic and scale- inconsistent features and further improve the segmentation effect, we propose a cross-joint attention-guided Multi-scale fusion scheme, which solves the problems that arise when fusing features of different scales.

Table 2. Number of data collection in each health category

3.2 Convolutional Attention Modules

Convolutional Block Attention Module (CBAM) consists of channel attention and spatial attention. Among them, channel attention strengthens the connection between features in different channels by using maximum pooling and average pooling to pay attention to the information between different channels; spatial attention improves the spatial connection of the network by paying attention to space. Combining the channel and spatial attention mechanisms, adaptive feature extraction can be achieved. The CBAM model structure is shown in Figure 2.

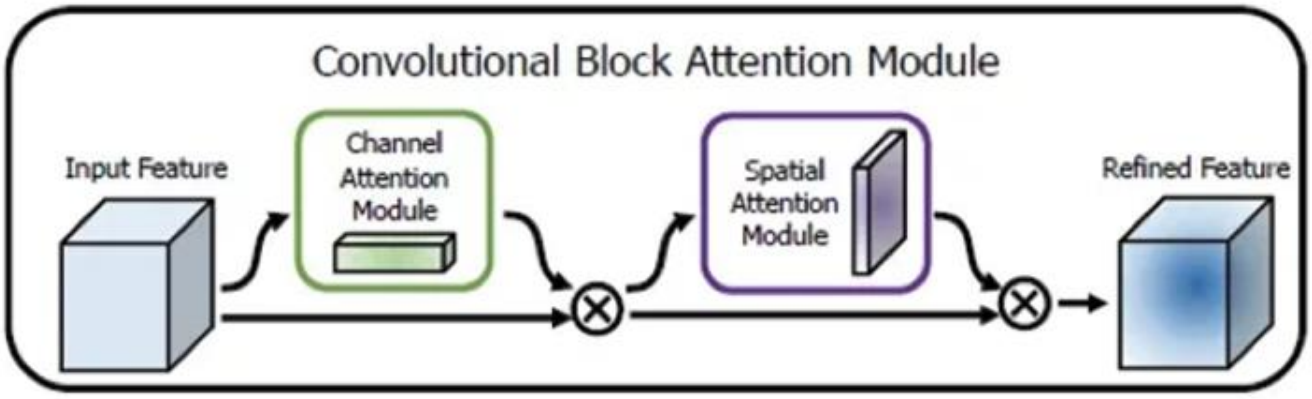

Figure 2. CBAM module

CBAM can be expressed by the following formula:

$$F' = M_C(F) \otimes F$$
$$F'' = M_s(F') \otimes F' \quad (1)$$

Among them, F is the feature map, Mcand Msrepresents channel-based and spatial-based attention respectively, ⊗ represents element-wise multiplication, F' and F"represents the output feature map after channel attention and spatial attention respectively. Since the input and output sizes of the CBAM module are the same, it can be inserted anywhere in an existing model.

3.3 SC Jump Link Module

Segmentation network contain more fine-grained information, which facilitates the segmentation of small lesions. Deep segmentation networks can extract more high-level semantic information, thereby improving segmentation accuracy. In addition, rich multi-scale information integrates the characteristics of different receptive fields, which is beneficial to the segmentation of multi-lesion areas. The skip connection is redesigned to aggregate features of different semantic scales on the decoder subnetwork to form a highly flexible feature fusion scheme. The SC skip connection is an operation that connects simple and effective deep and shallow information fusion. In SC - UNeXt; Jump link takes the first layer as an example:

$$z_1 = f([(2x, +z_2) \otimes x_1] + z_2) \quad (2)$$

Z3 feature information fed back by the jump link is fused with the Z4 advanced semantic features obtained by the convolution operation and the upsampling operation, and then the upsampling operation is performed.

3.4 PFM Pyramid Feature Fusion Module

In order to make the network capable of multi-scale detection, the article uses deconvolution to expand feature layers at different levels to the same size, and then adds them at the element level. The fused feature layer has richer multi- scale features.

4. Experimental Settings

4.1. Dataset

In this experiment, the NIH Chest X-ray Dataset comprises 112,140 X-ray images, each labeled with specific diseases, from 30,105 unique patients. The dataset includes 15 categories (14 diseases and one "none found"). Images are categorized as "no findings" or one or more disease categories, representing 15 prevalent chest pathologies. The original NIH chest X-ray dataset does not include lung field labels. For this study, we randomly selected 2,385 samples, which were then annotated by medical professionals to mark the lung fields, creating a new dataset we refer to as Haut.

The TeHaut dataset features chest X-rays that are severely blurred, occluded, or distorted. Within the Haut dataset, there are 1,637 images of normal individuals and 1,238 images with annotated lung fields exhibiting various

conditions. These include 192 cases of infiltration, 44 cases of atelectasis, 72 cases of effusion, 63 cases of nodules, 53 cases of masses, 23 cases of pneumothorax, and 33 cases of cardiomegaly. Additional cases include 34 instances of pleural thickening, 33 instances of pleural thickening with fibrosis, 23 cases of consolidation, 20 cases of emphysema, 12 cases of edema, 11 cases of pneumonia, 2 cases of hernia, and 317 cases with multiple medical conditions (comprising any combination of the aforementioned diseases).

For the application of EfficientNet-b4, a preprocessing step involves resizing the images to 256×256 pixels.

To bring our experiments as close as possible to point-ofcare imaging, we chose the International Skin Imaging Collaboration (ISIC 2018) to benchmark our results. The ISIC dataset contains camera-acquired skin images and corresponding skin lesion region segmentation maps. The ISIC 2018 dataset consists of 2594 images. We resized all images to a resolution of 512 × 512.

4.2 Experimental Setting

We use the Adam optimizer with a learning rate of 0.0001 and a momentum of 0.9. We also used a cosine annealing learning rate scheduler with a minimum learning rate up to 0.00001. The batch size is set to 8. We trained UNeXt for a total of 400 epochs. We perform an 80-20 random split three times on the dataset and report the mean and variance. Evaluation indicators, we use IoU, Dice Segmentation index to quantify the segmentation ability of SC-UneXt, Dice Similarity Coefficient (DSC), Dice coefficient is a set similarity measure, I O U is used to evaluate the degree of similarity between predictions and true values. Semantic segmentation can be viewed as pixel-level classification. True Positive (TP): The model prediction is a positive example, that is, a positive example. False positive (FP): The model predicts a positive example, but it is a negative example. False Negative (FN): The model prediction is a negative example, but it is a positive example. True Negative (TN): The model prediction is a counterexample, it is a counter example.

$$Dice = \frac{2TP}{2TP+FP+FN}$$
$$IOU = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Dice is usually used to calculate the similarity of two samples. The value range is 0 - 1. The best segmentation result is 1 and the worst value is 0. I OU is calculated as the ratio of the intersection and union of the two sets of real values and predicted values. The larger the ratio, the higher the similarity between the real value and the predicted value, the better the segmentation effect.

5. Experiment

5.1. Experimental Results

The network frameworks used in our comparative experiments include the most advanced CNN-based networks, such as U-Net, U-NeXt, and SC-UneXt. Below we will quantitatively and qualitatively compare the test results. analyze. Furthermore, the number of parameters in each network is kept to two decimal places. Evaluation of dermatology datasets. The following table SC-UneXt is the segmentation evaluation indicators Dice, Iou, parameter amount and computing power

consumption:

Table 1. Segmentation evaluation index

Method	FLOPS	Parameters	Dice	IOU
U-Net	25.06 GFLOPs	9.04 MB	84.84%	76.37%
ResUNet++	19.31GFLOPs	7.67 MB	85.73%	77.12%
UNext	0.10 GFLOPs	247.62 KB	86.64%	78.71%
SC-UNeXt	**0.13 GFLOPs**	**243.63 KB**	**87.12%**	**79.13%**

We noticed that the IoU and Dice of SC -UNeXt are 2.74% and 2.76% higher than U-Net respectively, 0.48% and 0.42% higher than U- NeXt respectively, and 1.39% and 2.01% higher than ResUNet++. We noticed, compared with U-Net's 9.04 M parameters, SC -UNeXt's 243.63 KB parameters are also relatively low, roughly the same as UNext. While retaining the lightweight, simple calculation, and fast characteristics of unext, the segmentation effect has been significantly improved from the dice coefficient and IOU coefficient. Regarding the computational power consumption of the model, that is, in terms of flops, it consumes less than Unet, which ensures the lightweight of the model and saves the consumption of computing power.

5.2 Ablation Experiment

In actual situations, if a model is to be put online, the model needs to be repeatedly debugged to prevent the model from performing better only on known data sets and performing poorly on unknown data sets. That is to ensure the generalization ability of the model, which refers to the adaptability of machine learning to fresh samples. Only by ensuring the generalization ability of the model can the construction of the model be meaningful. Therefore, crossvalidation is particularly important throughout the modeling process. Use training set/test set splitting and cross-validation methods to avoid this situation. As shown in the figure below, split the data set into training set/test set, and perform crossvalidation on the training set to obtain the best model parameters., thereby obtaining the score of the model on the test set.

This experiment adopts five-fold cross-validation and conducts cross-validation experiments on SC-UneXt in three directions: jump link, feature fusion, and joint loss function. The experimental results are as follows:

Table 2. Ablation experiment

Model	DSC evaluation index		ISIC 207 data set Val :24		Data volume train: 9 6		Val :24
5 fold cross validation	Fold-1	F old-2	F old -3	F old -4	F old -5	mean	variance
UNeXt	82.61%	84.66%	91.47%	87.61%	83.08%	85.89%	0.14%
SC_UNext(edge_loss)	85.10%	85.53%	88.51%	86.69%	86.84%	86.14%	0.03%
SC_UNext (skip)	85.73%	85.90%	90.68%	82.46%	83.75%	86.53%	0.13%
SC_UNext (fuse)	87.17%	85.07%	90.58%	85.59%	84.92%	86.17%	0.10%
SC_UNext(edge loss+skip)	85.90%	87.98%	90.22%	85.79%	89.14%	86.90%	0.05%
SC_UNext (edge_loss+fuse)	87.98%	80.14%	90.50%	86.48%	88.03%	86.98%	0.20%
SC_UNext (skip+fuse)	82.90%	88.62%	88.60%	84.90%	87.30%	85.90%	0.07%
SC_UNext(edge_loss+skip+fuse)	85.89%	88.11%	89.02%	86.11 %	87.22%	87.29%	0.08%

Through the comparison of ablation experiments, it was found that jump links, feature fusion, and joint loss functions all have an impact on segmentation accuracy. Only under the combined effect of S C-UNeXt at the same time, the DSC evaluation index is the highest and the segmentation effect is

good.

6. Conclusion

In this article, our purpose is to make the network better learn effective features and obtain more accurate lesion segmentation results. We propose an improvement that uses feature fusion and increases the proportion of edge features in the total sample features. network architecture. This network is not only better than other methods in terms of evaluation indicators, but is also cost-effective enough to better help doctors better diagnose the details of these histological images. Future research topics in medical image segmentation will be deep learning to automatically select features from different resolutions, or consider using adversarial training including test images to exploit features in test images without annotations. It achieves high-precision segmentation while also achieving instant feedback and lightweight operations, which plays a role in assisting diagnosis and treatment in the medical process.

References

[1] Liu, H., Li, I., Liang, Y., Sun, D., Yang, Y., & Yang, H. (2024). Research on Deep Learning Model of Feature Extraction Based on Convolutional Neural Network. arXiv e-prints, arXiv-2406.

[2] Hu, Y., Hu, J., Xu, T., Zhang, B., Yuan, J., & Deng, H. (2024). Research on Early Warning Model of Cardiovascular Disease Based on Computer Deep Learning. arXiv preprint arXiv:2406.08864.

[3] Yang, Y., Qiu, H., Gong, Y., Liu, X., Lin, Y., & Li, M. (2024). Application of computer deep learning model in diagnosis of pulmonary nodules. arXiv preprint arXiv:2406.13205

[4] Xiao, M., Li, Y., Yan, X., Gao, M., & Wang, W. (2024, March). Convolutional neural network classification of cancer cytopathology images: taking breast cancer as an example. In Proceedings of the 2024 7th International Conference on Machine Vision and Applications (pp. 145-149).

[5] Sun, D., Liang, Y., Yang, Y., Ma, Y., Zhan, Q., & Gao, E. (2024). Research on Optimization of Natural Language Processing Model Based on Multimodal Deep Learning. arXiv preprint arXiv:2406.08838.

[6] Yan, X., Wang, W., Xiao, M., Li, Y., & Gao, M. (2024, March). Survival prediction across diverse cancer types using neural networks. In Proceedings of the 2024 7th International Conference on Machine Vision and Applications (pp. 134-138).

[7] Zhang, H., Diao, S., Yang, Y., Zhong, J., & Yan, Y. (2024). Multi-scale image recognition strategy based on convolutional neural network. Journal of Computing and Electronic Information Management, 12(3), 107-113.

[8] Mei, T., Zi, Y., Cheng, X., Gao, Z., Wang, Q., & Yang, H. (2024). Efficiency optimization of large-scale language models based on deep learning in natural language processing tasks. arXiv preprint arXiv:2405.11704.

[9] Xiao, L., Li, M., Feng, Y., Wang, M., Zhu, Z., & Chen, Z. (2024). Exploration of Attention Mechanism-Enhanced Deep Learning Models in the Mining of Medical Textual Data. arXiv preprint arXiv:2406.00016.

[10] Yan, Y., He, S., Yu, Z., Yuan, J., Liu, Z., & Chen, Y. (2024). Investigation of Customized Medical Decision Algorithms Utilizing Graph Neural Networks. arXiv preprint arXiv:2405.17460.

[11] Gao, Z., Wang, Q., Mei, T., Cheng, X., Zi, Y., & Yang, H. (2024). An Enhanced Encoder-Decoder Network Architecture for Reducing Information Loss in Image Semantic Segmentation. arXiv preprint arXiv:2406.01605.

[12] Zhan, Q., Ma, Y., Gao, E., Sun, D., & Yang, H. (2024). Innovations in Time Related Expression Recognition Using LSTM Networks. International Journal of Innovative Research in Computer Science & Technology, 12(3), 120-125.

[13] Yang, Y., Chen, Z., Yan, Y., Li, M., & Gegen, T. (2024). A new method of image recognition based on deep learning generates adversarial networks and integrates traditional algorithms. Journal of Computing and Electronic Information Management, 13(1), 57-61.

[14] Yao, J., Li, C., Sun, K., Cai, Y., Li, H., Ouyang, W., & Li, H. (2023, October). Ndc-scene: Boost monocular 3d semantic scene completion in normalized device coordinates space. In 2023 IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 9421-9431). IEEE Computer Society.

[15] Xu, R., Zi, Y., Dai, L., Yu, H., & Zhu, M. (2024). Advancing Medical Diagnostics with Deep Learning and Data Preprocessing. International Journal of Innovative Research in Computer Science & Technology, 12(3), 143-147.

[16] Ahmad, P., Jin, H., Alroobaea, R., Qamar, S., Zheng, R., Alnajjar, F., & Aboudi, F. (2021). MH UNet: A multi-scale hierarchical based architecture for medical image segmentation. IEEE Access, 9, 148384-148408.

[17] Diakogiannis, F. I., Waldner, F., Caccetta, P., & Wu, C. (2020). ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS Journal of Photogrammetry and Remote Sensing, 162, 94-114.