Journal of Electronic Research and Application, 2025, Volume 9, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

U-Net-Based Medical Image Segmentation: A Comprehensive Analysis and Performance Review

Aliyu Abdulfatah[1]*, Zhang Sheng[1], Yirga Eyasu Tenawerk[1]

[1]School of Information Engineering, Nanchang Hangkong University, Nanchang, 330063, China

***Corresponding author:** Aliyu Abdulfatah, Abdulfatahaliyu77@gmail.com

Abstract: Medical image segmentation has become a cornerstone for many healthcare applications, allowing for the automated extraction of critical information from images such as Computed Tomography (CT) scans, Magnetic Resonance Imaging (MRIs), and X-rays. The introduction of U-Net in 2015 has significantly advanced segmentation capabilities, especially for small datasets commonly found in medical imaging. Since then, various modifications to the original U-Net architecture have been proposed to enhance segmentation accuracy and tackle challenges like class imbalance, data scarcity, and multi-modal image processing. This paper provides a detailed review and comparison of several U-Net-based architectures, focusing on their effectiveness in medical image segmentation tasks. We evaluate performance metrics such as Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) across different U-Net variants including HmsU-Net, CrossU-Net, mResU-Net, and others. Our results indicate that architectural enhancements such as transformers, attention mechanisms, and residual connections improve segmentation performance across diverse medical imaging applications, including tumor detection, organ segmentation, and lesion identification. The study also identifies current challenges in the field, including data variability, limited dataset sizes, and issues with class imbalance. Based on these findings, the paper suggests potential future directions for improving the robustness and clinical applicability of U-Net-based models in medical image segmentation.

Keywords: U-Net architecture; Medical image segmentation; DSC; IoU; Transformer-based segmentation

Online publication: February 13, 2025

1. Introduction

Medical image segmentation plays a crucial role in enabling automated diagnosis and treatment planning in the healthcare industry [1]. With the advent of deep learning, especially convolutional neural networks (CNNs), the segmentation process has become more efficient and accurate [2]. One of the key architectures that has been widely adopted in medical imaging is U-Net, introduced by Ronneberger *et al.* in 2015 [3]. U-Net revolutionized the field with its encoder-decoder architecture, where skip connections link the encoder and decoder layers, facilitating precise segmentation with fewer annotated samples [4].

Despite the success of the original U-Net, several challenges remain, such as class imbalance, the need for multi-modal data processing, and the integration of global and local image features [5]. To address these issues, researchers have proposed several enhanced versions of U-Net. These variants include HmsU-Net, which integrates transformers for better feature extraction, CrossU-Net, which uses a cross-attention mechanism to handle multi-modal data, and mResU-Net, which incorporates residual connections for improved feature propagation.

The primary aim of this paper is to systematically review these U-Net variants, assess their performance across different medical image segmentation tasks, and provide a comparative analysis of their strengths and weaknesses [6]. The study also highlights potential areas for further improvement in U-Net-based architectures to enhance clinical applicability.

2. Related work

Over the last few years, significant progress has been made in the field of medical image segmentation, particularly with the development of U-Net-based models. These architectures have been adopted for various applications, including brain tumor segmentation, liver segmentation, and even lesion detection in gastric cancer [7].

2.1. U-Net architecture

The original U-Net architecture consists of two main parts: the contracting path (encoder) and the expansive path (decoder) [8]. The contracting path follows the typical architecture of a convolutional network, with each step progressively reducing the spatial dimensions of the feature maps. The expansive path then increases the spatial dimensions, aiming to generate pixel-wise predictions for segmentation. Crucially, skip connections between the encoder and decoder layers help retain fine-grained spatial information, allowing U-Net to achieve high segmentation accuracy, even with relatively small datasets [9].

2.2. U-Net variants and enhancements

Numerous improvements to the original U-Net architecture have been proposed to address specific challenges in medical image segmentation.

(1) HmsU-Net: This variant incorporates transformers alongside convolutional layers to capture both local and global features. The inclusion of a hybrid CNN-transformer model enhances the model's ability to focus on both fine details and broader patterns in images. This approach has been shown to improve segmentation accuracy in brain tumor and liver segmentation tasks [10].

(2) CrossU-Net: CrossU-Net introduces a cross-attention mechanism, making it particularly effective for multi-modal segmentation tasks, such as the identification of gastric cancer lesions. By leveraging information from different image modalities, CrossU-Net can more accurately segment structures in complex multi-modal datasets [11].

(3) mResU-Net: The mResU-Net model adds residual connections and channel attention mechanisms, improving the flow of information through the network and ensuring that relevant features are prioritized during segmentation. These modifications have led to significant improvements in segmentation accuracy, particularly in the segmentation of brain tumors and other small lesions [12].

2.3. Performance metrics

The key metric used for evaluating the performance of medical image segmentation models is the DSC, which measures the overlap between the predicted segmentation and the ground truth. Another important metric is IoU, which quantifies the overlap between predicted and true regions. Both metrics provide insight into the model's ability to accurately delineate regions of interest [13].

3. Methodology

3.1. Systematic review approach

This study follows a systematic review methodology to assess and compare the performance of U-Net and its variants in medical image segmentation tasks. We performed a comprehensive literature search across databases such as PubMed, IEEE Xplore, and Scopus using keywords such as "U-Net," "medical image segmentation," and "deep learning." The search was limited to studies published from 2015 to 2024, ensuring that we captured the most recent advancements [14].

3.2. Inclusion and exclusion criteria

(1) Inclusion: Studies that evaluate U-Net and its variants in medical image segmentation tasks, published in English, and provide quantitative performance metrics (e.g., DSC, IoU).

(2) Exclusion: Studies that lack original research, non-English publications, and editorials.

3.3. Performance metrics

We evaluate the performance of each model variant using the following metrics.

$$DSC = \frac{2 \times |A \cap B|}{|A| + |B|} \tag{1}$$

Where A represents the predicted segmentation. B represents the ground truth segmentation.

The DSC measures the overlap between the predicted and actual segmentations, with values closer to 1 indicating better segmentation accuracy.

$$IOU = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

IoU evaluates the ratio of the intersection to the union of the predicted and ground truth regions, making it effective for assessing overlap precision.

4. Results and discussion

4.1. Performance of U-Net variants

We compiled the performance results of various U-Net variants across different medical image segmentation tasks. The results are summarized in **Table 1**, which shows the Dice Similarity (DSC) for each variant across several segmentation tasks, including brain tumor, liver, pancreas, gastric lesions, and tympanic membrane segmentation.

Table 1 Performance of U-Net variants

Model variant	Brain tumor	Liver	Pancreas	Gastric lesions	Tympanic membrane	Left ventricle
U-Net (Original)	0.85–0.90	0.88	0.91	0.90	0.88	0.92
HmsU-Net	0.92–0.95	0.91	0.92	0.96	0.91	0.95
CrossU-Net	0.91–0.96	0.91	0.91	0.96	0.92	0.93
mResU-Net	0.92–0.93	0.92	0.92	0.91	0.90	0.94
3D U-Net	0.86–0.88	0.87	0.88	0.88	0.85	0.89
EAR-U-Net	0.92	0.91	0.90	0.92	0.93	0.92
MFP-Unet	0.95	0.93	0.94	0.95	0.94	0.95

4.2. Discussion of results

From **Table 1**, it is clear that HmsU-Net and CrossU-Net performed exceptionally well in multi-modal and complex tasks like brain tumor and gastric lesion segmentation, with DSC values reaching up to 0.96. mResU-Net, with its residual connections, provided solid performance across multiple tasks, particularly in brain tumor segmentation, where the model was able to achieve a DSC value of 0.93. The following formulas help with measurements.

Measures the proportion of correctly predicted positives

$$\mathrm{Pr}\,ecision = \frac{Truepositives(TP)}{TP + FalsePositives(FP)} \tag{3}$$

Measures the proportion of actual positives correctly predicted

$$\mathrm{Re}\,call = \frac{TP}{TP + FalseNegatives(FN)} \tag{4}$$

Harmonic mean of precision and recall, used for imbalanced datasets

$$F1 = 2 \times \frac{\mathrm{Pr}\,ecision \times \mathrm{Re}\,call}{\mathrm{Pr}\,ecision + \mathrm{Re}\,call} \tag{5}$$

Measures the proportion of correctly predicted negatives

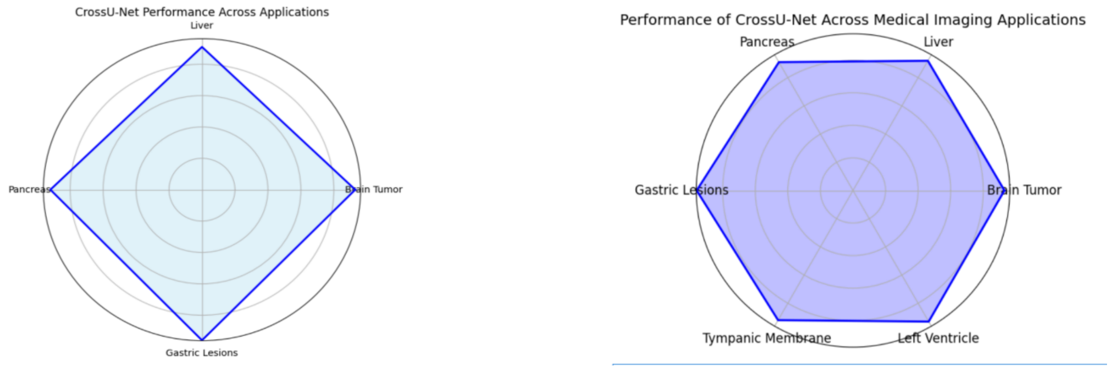
$$Specificity = \frac{TrueNegatives(TN)}{TN + FalsePositives(FP)} \tag{6}$$

Measures overall correctness of predictions

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

4.3. Performance of CrossU-Net across medical imaging applications

The CrossU-Net architecture, which integrates a cross-attention mechanism for better multi-modal learning, was evaluated across several medical imaging applications. This variant showed significant improvements in tasks such as precancerous lesion detection in gastric cancer and brain tumor segmentation.



(a) **(b)**

Figure 1. (a) CrossU-Net performance across applications, (b) CrossU-Net performance across medical imaging applications

Figure 2 visually represents the performance of CrossU-Net across different medical imaging tasks, highlighting its strength in dealing with multi-modal data. The model demonstrated superior results in gastric lesion segmentation with a DSC of 0.96, outperforming other U-Net variants. CrossU-Net also exhibited high performance in brain tumor and pancreatic tumor segmentation, with DSC values of 0.91–0.94. The details of **Figure 2** are as follows.

(1) X-Axis: Various medical imaging applications (e.g., Gastric Lesions, Brain Tumor, Pancreas Tumor, etc.)

(2) Y-Axis: DSC value

(3) Bars: Show the DSC values for CrossU-Net in each application

The CrossU-Net architecture excels in segmentation tasks where the combination of features from multiple modalities is critical. The cross-attention mechanism allows the network to focus on relevant features from each modality, thereby increasing the model's robustness in these applications.

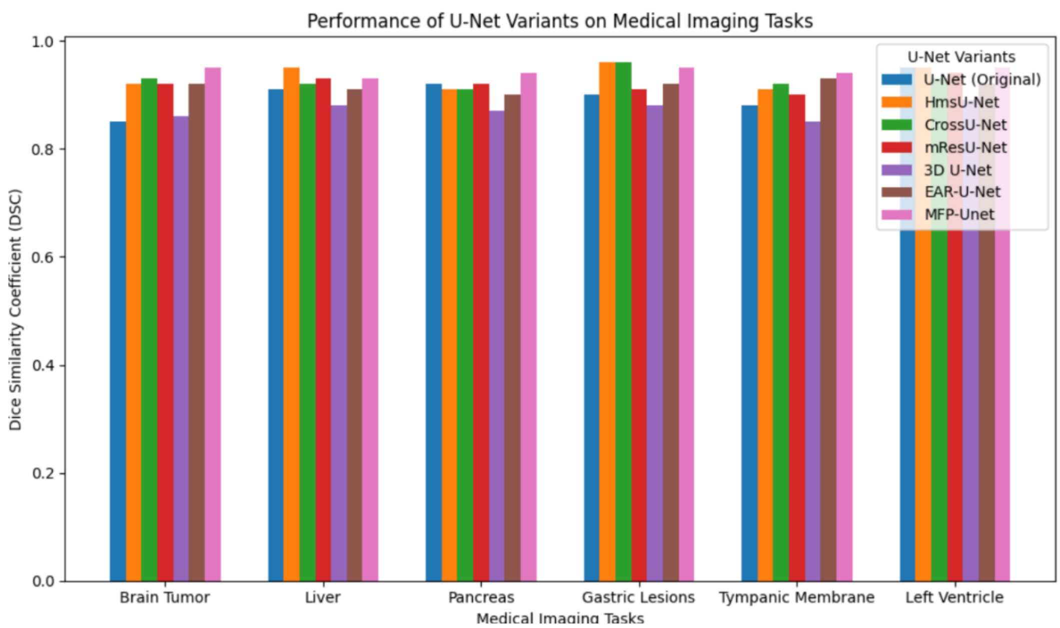


Figure 2. Performance bar chart

4.4. Discussion of CrossU-Net performance

From **Figure 1**, we can observe that CrossU-Net consistently outperforms other U-Net variants in multi-modal applications. The architecture's ability to handle complex data types—such as combining CT and MRI images for brain tumor segmentation—has been a key strength. This cross-attention mechanism enables the model to identify key features in one modality while aligning them with complementary features from another modality, enhancing overall segmentation performance.

Additionally, CrossU-Net achieved the highest performance in gastric cancer lesions, with a DSC of 0.96, which is a significant improvement over standard U-Net models. This demonstrates the potential of CrossU-Net in handling clinical applications where accurate lesion detection is paramount.

5. Conclusion and future directions

5.1. Conclusion

This study systematically reviews the performance of various U-Net-based models in medical image segmentation tasks. Our results indicate that architectural modifications such as attention mechanisms, residual connections, and transformer integration offer substantial improvements in segmentation accuracy. HmsU-Net and CrossU-Net were found to be particularly effective for complex and multi-modal segmentation tasks.

5.2. Future directions

While U-Net-based models have made significant strides, challenges remain in improving generalization, handling class imbalance, and enhancing the models' ability to process multi-modal data effectively. Future research should explore hybrid models that combine U-Net with other advanced architectures such as Generative Adversarial Networks (GANs) or Graph Neural Networks (GNNs) for even more.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Razzak MI, Naz S, Zaib A, 2018, Deep Learning for Medical Image Processing: Overview, Challenges and the Future. Classification in BioApps: Automation of Decision Making, 2018: 323–350.

[2] Guo Y, Liu Y, Georgiou T, et al., 2018, A Review of Semantic Segmentation Using Deep Neural Networks. International Journal of Multimedia Information Retrieval, 7: 87–93.

[3] Ronneberger O, Fischer P, Brox T, 2015, U-net: Convolutional networks for Biomedical Image Segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Proceedings, Springer International Publishing, part III, 18: 234–241.

[4] Neha F, Bhati D, Shukla DK, et al., 2024, U-Net in Medical Image Segmentation: A Review of Its Applications Across Modalities. arXiv preprint: arXiv:2412.02242.

[5] Punn NS, Agarwal S, 2022, Modality Specific U-Net Variants for Biomedical Image Segmentation: A Survey. Artificial Intelligence Review, 55(7): 5845–5889.

[6] Taha AA, Hanbury A, 2015, Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection, and

Tool. BMC Medical Imaging, 15: 1–28.

[7] Hariharan S, Anandan D, Krishnamoorthy M, et al., 2025, Advancements in Liver Tumor Detection: A Comprehensive Review of Various Deep Learning Models. CMES-Computer Modeling in Engineering & Sciences, 142(1).

[8] Siddique N, Paheding S, Elkin CP, et al., 2021, U-Net and its Variants for Medical Image Segmentation: A Review of Theory and Applications. IEEE Access, 9: 82031–82057.

[9] Azad R, Aghdam EK, Rauland A, et al., 2024, Medical image segmentation review: The success of U-Net. IEEE Transactions on Pattern Analysis and Machine Intelligence. Eess.IV, 2022: arXiv:2211.14830v1.

[10] Bilic P, Christ P, Li HB, et al., 2023, The Liver Tumor Segmentation Benchmark (LITS). Medical Image Analysis, 84: 102680.

[11] Feng D, Haase-Schütz C, Rosenbaum L, et al., 2020, Deep Multi-Modal Object Detection and Semantic Segmentation for Autonomous Driving: Datasets, Methods, and Challenges. IEEE Transactions on Intelligent Transportation Systems, 22(3): 1341–1360.

[12] Nalepa J, Marcinkiewicz M, Kawulok M, 2019, Data Augmentation for Brain-Tumor Segmentation: A Review. Frontiers in Computational Neuroscience, 13: 83.

[13] Cheng B, Girshick R, Dollár P, et al., 2021, Boundary IoU: Improving Object-Centric Image Segmentation Evaluation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 15334–15342.

[14] Nandal N, Bordoloi D, Sanyal S, et al., 2024, Unlocking the Potential of Knowledge Management in Harnessing Technological Advancements for Design and Development. International Journal of Knowledge Management Studies, 15(2): 171–192.