International Journal of Applied Science; Vol. 8, No. 4; 2025
ISSN 2576-7240 E-ISSN 2576-7259
https://doi.org/10.30560/ijas.v8n4p98

Eff-TA-Net: A Lightweight Model with Triple Attention for Medical Image Segmentation

Xiaowen Jiang[1] & Xin Wang[1]

[1] School of Electronic Information and Artificial Intelligence, Shaanxi University of Science & Technology, China

Correspondence: Xiaowen Jiang, School of Electronic Information and Artificial Intelligence, Shaanxi University of Science & Technology, Xi'an 710016, China. E-mail: 231611013@sust.edu.cn

Received: November 20, 2025; Accepted: December 4, 2025; Published: December 5, 2025

Abstract

Despite significant progress in deep learning for medical image analysis, the clinical adoption of many segmentation models remains limited due to their large parameter sizes and high computational complexity. Furthermore, these models often struggle to accurately segment regions with low contrast and blurred boundaries, which are common challenges in medical imaging. To address these issues, this paper introduces Eff-TA-Net, a lightweight medical image segmentation model incorporating a triple attention mechanism. Built upon the UNeXt architecture, Eff-TA-Net employs a hybrid encoder combining convolutional and Tok-MLP modules to reduce parameter count, while a triple attention mechanism integrated into skip connections enhances spatial feature perception. Additionally, a composite loss function named DiceBoundary Loss is designed to synergistically optimize the Dice coefficient and a boundary-aware binary cross-entropy loss, thereby improving the model's precision in segmenting lesion boundaries. Experimental results on the ISIC 2018 skin cancer dataset demonstrate that Eff-TA-Net, with only 1.50 million parameters and 2.31 GFLOPs, achieves state-of-the-art segmentation accuracy among comparable models, with an F1-score of 90.03% and an IoU of 85.26%. These findings indicate the strong potential of Eff-TA-Net for practical clinical applications.

Keywords: Medical image segmentation, Lightweight model, Attention mechanism, Loss function, Boundary-aware

1. Introduction

Recent years have witnessed substantial progress in medical image segmentation, driven primarily by advances in deep learning methodologies, particularly convolutional neural networks (CNNs). Models like UNet and its derivatives, with their distinctive encoder-decoder structure and skip connections, have established themselves as foundational benchmarks [1]. This pursuit of high accuracy, however, often comes at a cost: these models frequently employ very deep architectures, leading to large parameter footprints (often 20M to 100M), high computational demands, and slow inference, which limits their practicality in resource-conscious clinical settings like mobile and edge computing[2]. Consequently, developing lightweight architectures that preserve competitive performance has become a prominent research focus. The rise of Vision Transformers (ViTs) offered a new paradigm, though their high computational complexity and need for vast datasets present hurdles [3]. A notable lightweight solution, UNeXt [4], cleverly merges CNNs with MLPs, achieving competitive results with far fewer parameters, thus providing a key design insight. A common trade-off with lightweight models, however, is reduced feature representation power, which can impair performance on challenging aspects of medical images like low contrast and blurry borders. Attention mechanisms are known to sharpen a model's focus on relevant features [5], yet the standard self-attention mechanism is computationally expensive and ill-suited for lightweight frameworks.Consequently, a critical challenge is to integrate efficient attention mechanisms into lightweight networks to enhance capability without overburdening computational resources. Addressing these challenges, the principal contributions of this study are as follows:

(1) We introduce a lightweight segmentation model enhanced with a triple-attention mechanism, boosting spatial feature awareness with negligible computational overhead.

(2) We propose a new composite loss function, DiceBoundary Loss, that refines boundary segmentation by simultaneously optimizing both regional overlap and precise boundary alignment.

(3) Extensive experiments, including comparisons and ablations on public datasets, confirm our model's superior accuracy-efficiency trade-off and the efficacy of its constituent parts

2. Related Work

2.1 Medical Image Segmentation Models

Medical image segmentation, serving as a core component of computer-aided diagnosis (CAD) systems, aims to achieve pixel-level identification of organs, tissues, or lesions within images. A significant milestone was achieved with the Fully Convolutional Network (FCN) [6], which introduced the first end-to-end CNN architecture for semantic segmentation by utilizing only convolutional layers. Building upon this, UNet [7], with its symmetric encoder-decoder architecture and skip connections, became a cornerstone model in the field of medical image segmentation. These skip connections integrate detailed, high-resolution encoder features with the semantically rich decoder features, thereby compensating for the spatial detail loss inherent in pooling operations, making it particularly adept at segmenting intricate biomedical structures. Subsequent research has led to the prolific development of UNet variants along different technical pathways, aiming to further enhance segmentation accuracy and robustness.

To capture more complex feature representations, ResUNet [8] incorporated residual modules into UNet. These residual or shortcut connections facilitate gradient flow in deep networks, mitigating the vanishing gradient problem and thus allowing for the training of substantially deeper models. Inspired by DenseNet, DenseUNet [9] implemented dense connections within the encoder, promoting feature reuse and enhancing gradient flow during backpropagation. UNet++ [10] enhanced multi-scale feature exploitation by designing a nested architecture with dense skip connections, which reduces the semantic disparity between features from the encoder and decoder pathways. UNet3+ [11] further proposed full-scale skip connections and deep supervision, allowing each layer in the decoder to directly receive feature maps from all scales of the encoder, achieving genuine full-scale feature fusion. To enable models to adaptively focus on target regions, Attention UNet [12] embedded attention gate modules within the skip connections, suppressing irrelevant background responses without extra supervision signals. FCD-R2U-net [13] combined recurrent convolutional residual networks with UNet, using a recurrent structure to simulate an attention mechanism and enhance contextual information capture.

In recent years, Transformer models have been introduced to computer vision due to their powerful global context modeling capabilities. As a pioneer, TransUNet [14] utilized a Transformer as the encoder to extract global context and combined it with a CNN decoder, achieving significant performance gains in medical image segmentation. Demonstrating the viability of Transformer-based designs, Swin-UNet [15] implemented a fully translational U-shaped network using Swin Transformer blocks, showcasing strong potential for medical image segmentation.

However, although these models continually achieve state-of-the-art results on specific datasets, their performance gains often rely on deeper networks, more complex connections, or computationally intensive modules. However, these enhancements often result in substantially larger model sizes and greater computational demands, which can hinder their practical deployment in resource-limited clinical environments..

2.2 Lightweight Medical Image Segmentation Models

The practical need for computationally efficient solutions has propelled lightweight model design to the forefront of recent research efforts. Research on model lightweighting primarily follows three directions. The first is compact architecture design. This line of work focuses on designing efficient network modules to build lightweight backbones. For instance, the MobileNet [16] and ShuffleNet [17] series significantly reduce computational cost and parameters through operations like depthwise separable convolution and channel shuffle, often employed as encoders in segmentation networks.

The second direction involves architectures based on Transformers and MLPs. The rise of Vision Transformer (ViT) [18] introduced a new paradigm for image segmentation. Architectures such as TransUNet [19] integrate the strength of CNNs in capturing local features with the global receptive field of Transformers. However, the inherent computational complexity of ViT and its reliance on large-scale datasets limit its utility in the medical image domain. As an emerging alternative, the UNeXt [20] model ingeniously integrates CNNs with Tokenized Multi-Layer Perceptrons (MLPs), achieving competitive performance with a very low parameter count, providing a significant reference for lightweight model design. Finally, methods like knowledge distillation and model compression aim to transfer knowledge from large "teacher models" to small "student models," but these typically involve complex multi-stage training pipelines.

Despite significant progress in existing lightweight research, a common challenge persists: lightweight design often comes at the cost of compromised feature representation capacity. Particularly when handling complex

scenarios common in medical images, such as low contrast and blurred boundaries, existing lightweight models still lack sufficient accuracy for detail and boundary segmentation. Pure lightweight architectures lack focus on critical spatial regions, while traditional attention mechanisms are difficult to integrate due to their excessive computational overhead. Therefore, researching how to integrate efficient attention mechanisms with lightweight backbone networks, while strengthening boundary constraints at the loss function level, becomes key to enhancing model performance. Motivated by these limitations, we introduce a triple attention mechanism and a boundary-aware loss function to enhance the detail segmentation performance of lightweight models, while maintaining computational efficiency.

3. Eff-TA-Net Model

3.1 Model Architecture

As illustrated in Figure 1, the overall architecture of Eff-TA-Net adopts a classic encoder-decoder schema with five hierarchical levels and incorporating skip connections. The primary innovations lie in the lightweight design of the encoder and the attention-enhanced feature fusion within the skip connections.

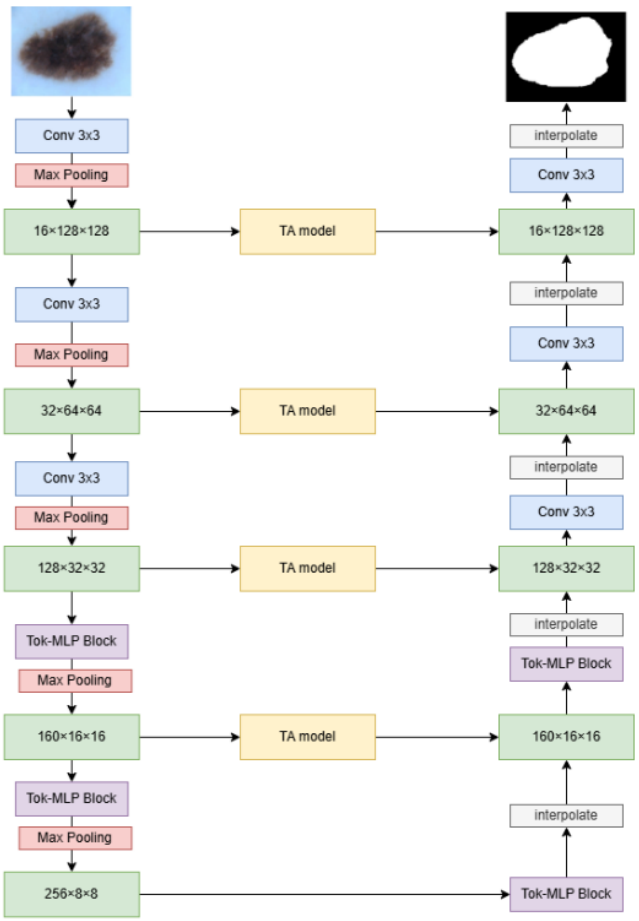

Figure 1. Overall framework of the Eff-TA-Net model

A hybrid encoder forms the core of the model, where convolutional blocks and token-mixing MLP (Tok-MLP) blocks are arranged in an alternating pattern across the network stages. The number of channels progressively increases through the layers according to the sequence: 32, 64, 128, 160, and 256. This configuration ensures that strong local features are captured initially by the convolutional layers, while the subsequent MLP-based modules efficiently model long-range interactions. The synergy between these components enables a powerful feature extraction process without a significant parameter overhead.

A symmetric decoder path is designed to reconstruct the segmentation map. Upsampling is performed using bilinear interpolation, a choice made to maintain parameter efficiency, foregoing the use of transposed convolution.

3.2 Core Modules

3.2.1 Tok-MLP Module

Central to the lightweight design is the Tok-MLP block. This module orchestrates a sequence of operations—including feature shifting along both spatial axes, tokenization, MLP projection, and depthwise separable convolution (DWConv)—to efficiently capture global contextual relationships. The integration of DWConv is critical, as it implicitly encodes positional information into the MLP-processed features. This approach effectively mitigates the fixed-input-size limitation associated with standard ViT positional embeddings and is inherently parameter-efficient. Its structure is illustrated in Figure 2.

The computational workflow of the module is summarized as follows. Given an input feature map $X \in R^{B\times C\times H\times W}$, where B is the batch size, C is the number of channels, and $H \times W$ are the spatial dimensions. Features are shifted along the x-axis and tokenized to obtain T_W.

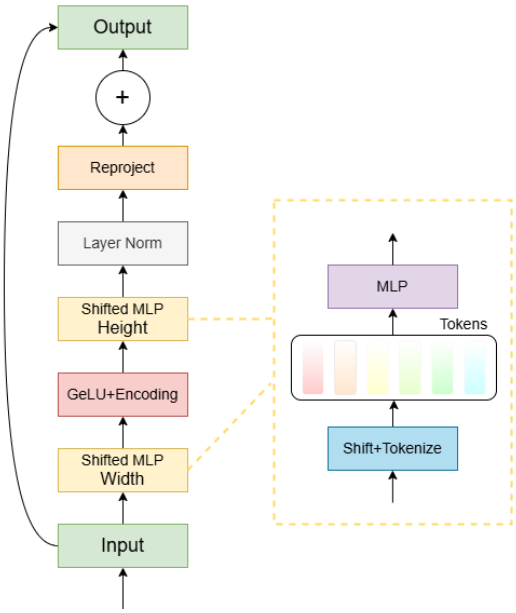

Figure 2. Structure of the Tok-MLP module

$$X_{shift} = Shift(X) \tag{1}$$

$$T_w = Tokenize(X_{shift}) \tag{2}$$

The resulting T_W is subsequently passed through an MLP, followed by processing with a depthwise separable convolutional (DWConv) layer.

$$Y = f\left(DWConv(MLP(T_w))\right) \tag{3}$$

The result Y is shifted along the y-axis and tokenized to produce T_H :

$$Y_{shift} = Shift(Y) \tag{4}$$

$$T_H = Tokenize(Y_{shift}) \tag{5}$$

Finally, T_H is activated, processed by another MLP, and combined with a residual connection. The resultant features are standardized using Layer Normalization.

$$Y = f\left(LN\left(T + MLP(GELU(T_H))\right)\right) \tag{6}$$

3.2.2 TA Module

The structure of the proposed Triple Attention (TA) module is depicted in Figure 3. Its core innovation integrates a triple attention mechanism with channel recalibration.

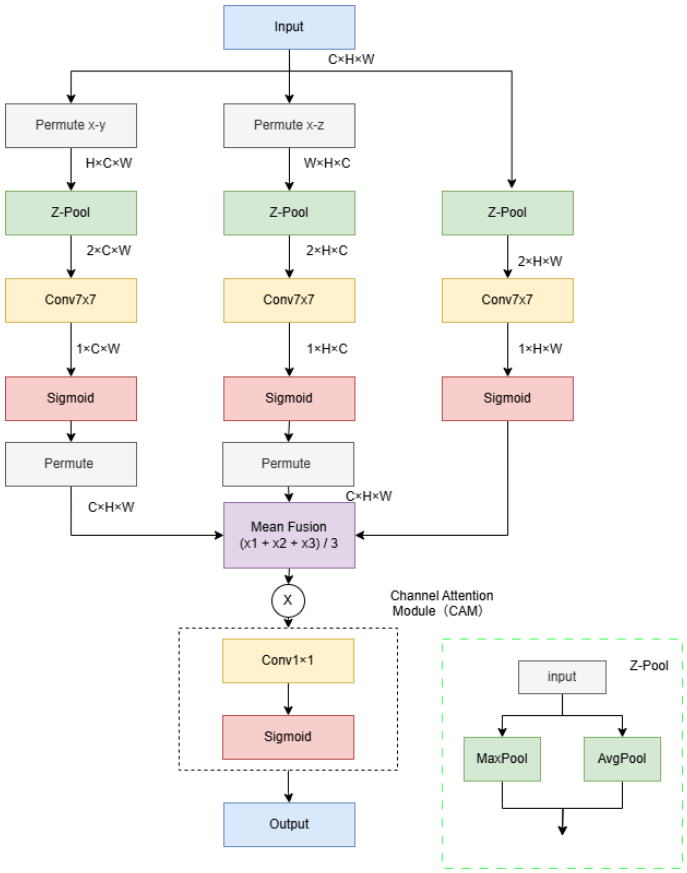

Figure 3. Structure of the TA module

The Eff-TA-Net model incorporates the proposed Triple Attention (TA) module into the skip connections of its first four layers. The TA module is built upon a Triplet Attention mechanism, which captures spatial dependencies based on directional modeling rather than complex self-attention. It operates by computing attention maps across three orthogonal spatial planes to enhance the model's perception of critical regions.

Given an input feature map $T \in R^{C\times H\times W}$, the Triplet Attention module constructs spatial attention maps along the following three planes: the $x-y$ plane (Height-Width direction), the $x-z$ plane (Channel-Width direction), the $y-z$ plane (Channel-Height direction).

For each plane, the input tensor is first permuted to align the relevant dimensions (e.g., for the x-z plane, the shape becomes $C \times W \times H$). Average pooling is then applied to preserve statistical features along the target direction. The resulting feature vector is subsequently compressed and transformed by a 7×7 convolutional layer (Conv 7×7). Directional attention weights A_{xy}, A_{xz}, and A_{yz} are generated by applying a Sigmoid activation function.

$$A_{xy} = \text{Permute}\left(\sigma\left(\text{Conv}_{7\times7}\left(\text{Z-Pool}\left(\text{Permute}_{\text{xy}}(T)\right)\right)\right)\right) \tag{7}$$

$$A_{xz} = \text{Permute}\left(\sigma\left(\text{Conv}_{7\times7}\left(\text{Z-Pool}\left(\text{Permute}_{\text{xz}}(T)\right)\right)\right)\right) \tag{8}$$

$$A_{yz} = \sigma\left(\text{Conv}_{7\times7}\left(\text{Z-Pool}(T)\right)\right) \tag{9}$$

The three directional attention maps are first integrated via a weighted average.

$$A = \frac{A_{xy} + A_{xz} + A_{yz}}{3} \tag{10}$$

The fused attention map A is then applied to the original input tensor through element-wise multiplication, thereby enhancing the features at critical spatial locations:

$$T' = T \odot A \tag{11}$$

To further augment the model's feature representation by enabling channel-wise selection, a lightweight channel recalibration module is introduced after the spatial attention. This module, comprised of a 1×1 convolutional layer followed by a Sigmoid activation function, generates a weighting factor for each channel.

$$M = \sigma\big(Conv_{1\times1}(T')\big) \tag{12}$$

$$Output = T' \times M \tag{13}$$

3.3 Loss Function

To simultaneously optimize both regional overlap and boundary alignment accuracy in the segmentation results, this paper introduces a composite loss function, termed DiceBoundary Loss, for training the Eff-TA-Net model. Similar hybrid loss strategies combining region and boundary terms have demonstrated effectiveness in various medical image segmentation tasks[21][22].

The Dice Loss function measures the degree of overlap between the prediction and ground truth masks, formulated as:

$$L_{\text{Dice}} = 1 - \frac{2 \cdot \sum y_{true} \cdot y_{pred} + \epsilon}{\sum y_{true}^2 + \sum y_{pred}^2 + \epsilon} \tag{14}$$

where y_{true}denotes the binary ground truth label (0 or 1), y_{pred} represents the predicted probability (ranging from 0 to 1), and ϵ is a small constant included to ensure numerical stability by preventing division by zero.

Accurate delineation of object boundaries is critical in medical image analysis, as misalignment can significantly impact clinical interpretation even if the overall regional overlap is high. Boundary loss enhances the model's focus on boundary structures and is formulated using a binary cross-entropy (BCE) scheme over boundary pixels:

$$L_{\text{Boundary}} = -\sum_{i=1}^{N}\big[B_G^{(i)} \cdot \log B_P^{(i)} + \big(1 - B_G^{(i)}\big) \cdot \log\big(1 - B_P^{(i)}\big)\big] = BCE(B_P, B_G) \tag{15}$$

Here, $B_G^{(i)}$ denotes the predicted boundary probability at the i-th pixel, and $B_P^{(i)}$is the corresponding ground truth boundary label.

The final composite loss function integrates the advantages of both regional and boundary-aware optimization. The composite loss function combines the Dice Loss and Boundary Loss through a weighted summation:

$$L_{DiceBoundary} = L_{Dice} + \lambda L_{Boundary} \tag{16}$$

Here, λ serves as a balancing factor that adjusts the influence of the boundary loss component against the regional overlap component.

4. Experiments and Results Analysis

4.1 Experimental Setup

We assessed Eff-TA-Net using the ISIC 2018 benchmark dataset containing 2,594 dermoscopic images. The dataset was partitioned randomly into training and validation sets with an 8:2 ratio, and all images were rescaled to 256×256 pixels. To improve generalization and prevent overfitting, we applied data augmentation techniques including random 90-degree rotations, horizontal and vertical flips, and rescaling, followed by pixel-wise normalization.

The model was implemented using the PyTorch framework and trained for 100 epochs. Optimization was performed using the Adam optimizer with a learning rate of 0.0001 and a momentum of 0.9. All experiments were conducted under consistent software and hardware conditions to ensure a fair comparison.

4.2 Comparative Experiments

The performance of Eff-TA-Net was rigorously compared with multiple classical and leading-edge segmentation models, with the quantitative results summarized in Table 1. The results demonstrate that Eff-TA-Net achieved the highest F1-score (90.03%) and IoU (85.26%) while maintaining a highly efficient architecture. With only 1.50 million parameters and 2.31 GFLOPs, the model attained an inference speed of 25 ms, matching that of the lightweight baseline UNeXt model. In contrast to TransUNet, which contains 105.32 million parameters, Eff-TA-Net not only reduced the parameter count to 1/70 but also achieved a 4.75% higher IoU, highlighting its superior balance between accuracy and efficiency.

Table 1. Performance and efficiency comparison of different models on the ISIC 2018 dataset.

Model	Params[M]	Speed[ms]	GFLOPs	F1	Iou
UNet	31.13	223	55.84	84.03	74.55
UNet++	9.16	173	34.65	84.96	75.12
ResUNet	62.74	333	94.56	85.60	75.62
MedT	1.60	751	21.24	87.35	79.54
TransUNet	105.32	246	38.52	88.91	80.51
UNeXt	1.48	25	2.29	89.70	80.70
TA- UNeXt	1.50	25	2.31	90.03	85.26

4.3 Ablation Study

Ablation studies were carried out to systematically evaluate the contribution of each component in the TA module. The results are summarized in Table 2.

In the configuration denoted as Eff-TA-Net-TA, the entire triple attention mechanism was removed, reducing the model to a baseline architecture similar to the original UNeXt. This ablation resulted in a decrease in IoU to 81.83%, indicating that the TA module alone contributes a performance improvement of 3.43%.

When only the channel recalibration module was ablated, denoted as Eff-TA-Net-CAM, the IoU dropped to 82.77%. This outcome underscores the importance of this module, responsible for a 2.49% improvement in IoU.

These experimental findings demonstrate that both the triple attention mechanism and the channel recalibration module are effective designs. Furthermore, their synergistic integration is essential for achieving the model's optimal performance.

Table 2. Ablation study on the key components of Eff-TA-Net.

Model	Params[M]	Speed[ms]	GFLOPs	F1	Iou
Eff-TA-Net	1.50	25	2.31	90.03	85.26
Eff-TA-Net-CAM	1.49	25	2.31	89.91	82.77
Eff-TA-Net-TA	1.48	25	2.29	89.72	81.83

5. Conclusion

This paper presents Eff-TA-Net, a lightweight model for medical image segmentation. By employing a hybrid encoder design, the model effectively controls the parameter count while maintaining competitive performance. The introduction of a triple attention mechanism and a channel recalibration module significantly enhances the model's ability to capture critical features in medical images. Experimental results on the ISIC 2018 dataset validate the effectiveness of the proposed approach. Future research will focus on two main directions: first, investigating alternative improvements to the Tok-MLP module to further enhance model efficiency; another direction involves testing the model's adaptability across various medical image types (e.g., MRI, CT) to validate its generalization performance.

References

[1] Yang, J., Wu, Z., Chen, M., Yang, S., Lin, X., Deng, Y., & Liu, X. (2026). D-net: Dynamic large kernel with dynamic feature fusion for volumetric medical image segmentation. *Biomedical Signal Processing and Control*, *113*, Article 108837. https://doi.org/10.1016/j.bspc.2025.108837

[2] Zhong, J., Wu, Z., Chen, C., Sun, Y., He, C., & Zhang, H. (2025). PMFSNet: Polarized multi-scale feature self-attention network for lightweight medical image segmentation. *Computer Methods and Programs in Biomedicine*, *261*, Article 108611. https://doi.org/10.1016/j.cmpb.2025.108611

[3] Khan, A., Alim, A., Hameed, Z., & Aljohani, A. A. (2025). A recent survey of vision transformers for medical image segmentation. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3618215

[4] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

[5] Zhang, J., Xu, Y., & Chen, S. (2025). Advances in attention mechanisms for medical image segmentation. *Computer Science Review*, *56*, Article 100721. https://doi.org/10.1016/j.cosrev.2024.100721

[6] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2015.7298965

[7] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, H. H. K. O. E. J. P. E. G. O. P. R. H. R. J. D. L. V. L. H. S. S. H. K. O. V. W. L. M. E. C. (Eds.), *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

[8] Diakogiannis, F. I., Krestenitis, M., Dimitrakopoulou, D., Thomas, K., Acikgoz, O., Barmpas, K., & Redon, E. (2020). ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *162*, 94–114. https://doi.org/10.1016/j.isprsjprs.2020.01.013

[9] Cai, S., Ma, J., Wang, Z., Xie, C., Tang, W., & Zhang, J. (2020). Dense-UNet: A novel multiphoton *in vivo* cellular image segmentation model based on a convolutional neural network. *Quantitative Imaging in Medicine and Surgery*, *10*(6), 1275. https://doi.org/10.21037/qims-19-1090

[10] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2019). UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, *39*(6), 1856–1867. https://doi.org/10.1109/TMI.2019.2959609

[11] Huang, H., Lin, L., Tong, R., Heng, P. A., & Zhang, R. (2020). UNet 3+: A full-scale connected UNet for medical image segmentation. *ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. https://doi.org/10.1109/ICASSP40776.2020.9053405

[12] Das, N., & Das, S. (2024). Attention-UNet architectures with pretrained backbones for multi-class cardiac MR image segmentation. *Current Problems in Cardiology*, *49*(1), Article 102129. https://doi.org/10.1016/j.cpcardiol.2023.102129

[13] Honnahalli, S. S., Tiwari, H., & Chitragar, D. V. (2023). Future Fusion+ UNet (R2U-Net) deep learning architecture for breast mass segmentation. *Engineering Proceedings*, *59*(1), Article 44. https://doi.org/10.3390/engproc2023059044

[14] Chen, J., Chen, X., Cao, H., Xiong, G., Sun, G., & Liu, J. (2024). TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. *Medical Image Analysis*, *97*, Article 103280. https://doi.org/10.1016/j.media.2024.103280

[15] Cao, H., Chen, B., Wang, Y., Lu, X., Wen, B., & Zhang, W. (2022). Swin-UNet: UNet-like pure transformer for medical image segmentation. In M. R. K. W. K. S. J. T. K. L. M. M. S. W. M. T. L. S. W. M. M. J. P. L. T. J. L. R. A. D. A. D. B. P. N. D. L. V. G. (Eds.), *European Conference on Computer Vision* (pp. 574–586). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-25066-8_9

[16] Nan, Y., Niu, J., Ma, C., Liu, C., & Zhang, Y. (2022). A-MobileNet: An approach of facial expression recognition. *Alexandria Engineering Journal*, *61*(6), 4435–4444. https://doi.org/10.1016/j.aej.2021.09.066

[17] Zhang, X., Zhou, X., Lin, L., & Sun, J. (2018). ShuffleNet: An extremely efficient convolutional neural network for mobile devices. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2018.00716

[18] Haruna, Y., Muhammad, K., & Rahman, M. (2025). Exploring the synergies of hybrid convolutional neural network and Vision Transformer architectures for computer vision: A survey. *Engineering Applications of Artificial Intelligence*, *144*, Article 110057. https://doi.org/10.1016/j.engappai.2025.110057

[19] Zee, J. M. V. D., Vercoulen, C., Homan, R. J. R., Klein, M. J. H., Stramigioli, S., Misra, S., & Bult, P. (2025). Evaluation of UNeXt for Automatic Bone Surface Segmentation on Ultrasound Imaging in Image-Guided Pediatric Surgery. *Bioengineering*, *12*(10), Article 1008. https://doi.org/10.3390/bioengineering12101008

[20] Valanarasu, J. M. J., & Patel, V. M. (2022). UNext: MLP-based rapid medical image segmentation network. In R. F. W. J. R. T. A. V. P. G. D. C. T. A. T. H. T. G. H. W. E. G. O. P. P. F. P. A. B. H. M. P. V. V. N. B. K. C. S. C. (Eds.), *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 37–47). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-16443-9_3

[21] Cao, H., Han, T., & Yang, Y. (2024). HDNeXt: Hybrid Dynamic MedNeXt with Level Set Regularization for Medical Image Segmentation. *Proceedings of the Asian Conference on Computer Vision* (pp. 433–449). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-0963-5_24

[22] Li, W., Li, J., Chen, T., Xu, X., He, J., & Feng, Z. (2025). Adaptive window adjustment with boundary DoU loss for cascade segmentation of anatomy and lesions in prostate cancer using bpMRI. *Neural Networks*, *181*, Article 106831. https://doi.org/10.1016/j.neunet.2024.106831