International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Improving Retinal Blood Vessel Segmentation Accuracy with Hybrid Attention-Based CNNs

Chakradhar Bandla

Submitted:10/03/2024 **Revised**: 25/04/2024 **Accepted**: 02/05/2024

Abstract: Accurate retinal blood vessel segmentation is crucial for diagnosing and monitoring various ocular and systemic diseases. While convolutional neural networks (CNNs) have shown potential in this area, their performance is often hindered by the intricate and subtle structures of retinal vasculature. This paper introduces a hybrid attention-based CNN architecture designed to overcome these challenges and improve segmentation accuracy. The model incorporates both spatial and channel attention mechanisms within a U-Net framework, enabling it to focus on the most relevant features in retinal images. By integrating attention gates and Squeeze-and-Excitation (SE) blocks, the network is better equipped to detect fine and complex blood vessels while reducing interference from irrelevant background information. Experimental evaluations on two public datasets—STARE, and DRIVE—demonstrate that the proposed method outperforms both attention-based and non-attention-based architectures, achieving state-of-the-art results. Specifically, the model attains Accuracy scores of 0.9876 and 0.9797 on the respective datasets . These results highlight the potential of the proposed approach in enhancing the accuracy and robustness of retinal blood vessel segmentation, making it a promising tool for clinical applications.

Keywords: *Attention mechanisms, Convolutional Neural Networks (CNNs), Deep learning, Diabetic retinopathy, Hypertensive retinopathy, Image segmentation, Retinal blood vessel segmentation*

1. Introduction

Blood vessels are essential components of the circulatory system, delivering oxygen and nutrients to tissues while removing waste products. Accurate segmentation of blood vessels from medical images is critical for various diagnostic and therapeutic procedures, particularly in diagnosing and managing ophthalmic conditions like diabetic retinopathy, glaucoma, and hypertensive retinopathy.

Accurate segmentation of blood vessels from medical images is critical for various diagnostic and therapeutic procedures, particularly in diagnosing and managing ophthalmic conditions like diabetic retinopathy, glaucoma, and hypertensive retinopathy. However, traditional methods for retinal vessel segmentation often struggle with challenges such as varying vessel widths, low contrast, and image noise [1]-[4].

Blood vessel segmentation has been a focus of medical image analysis due to its importance in disease diagnosis. Early methods relied on image processing techniques like edge detection, thresholding, and morphological operations, but their effectiveness was limited by image quality and variability. With the advent of machine learning, algorithms like Support Vector Machines (SVMs), Random Forests, Decision Trees, and Artificial Neural Networks (ANNs) were introduced, though these approaches depended heavily on handcrafted features [5]-[8].

Coppell-75019, Texas.
ORCID ID: 0009-0000-9901-9171
**Email:bandlachakradhar3@gmail.com*

Recently, convolutional neural networks (CNNs) have become powerful tools for medical image analysis, including retinal vessel segmentation [9]. Various U-Net variations have been developed to address challenges like capturing fine vessel structures, handling varying vessel sizes, and working with limited training data. Key U-Net variants include:

Attention U-Net: Incorporates attention mechanisms to focus on the most relevant features, enhancing segmentation accuracy, particularly for thin or low-contrast vessels [10].

Residual U-Net: Introduces residual connections to mitigate the vanishing gradient problem, improving the model's robustness and accuracy in capturing intricate vessel networks [11].

Dense U-Net: Combines DenseNet with U-Net to enhance feature propagation and reuse, improving segmentation across different scales and vessel thicknesses [12].

Recurrent U-Net (R2U-Net): Adds recurrent connections to iteratively refine segmentation predictions, effectively capturing elongated and continuous vessel structures [13].

Multi-Scale U-Net: Uses multi-scale processing to capture both fine details and broader contextual information, improving the detection of vessels with significant size variation [14].

3D U-Net: Extends U-Net to process 3D input volumes, capturing spatial context in three dimensions, with potential applications in 3D retinal imaging modalities [15].

Residual Attention U-Net: Combines residual connections and attention mechanisms, improving the segmentation of complex vessel structures with varying thickness [16].

Despite their success, CNN-based methods still face challenges in accurately segmenting intricate retinal vessels, particularly in low-contrast or overlapping areas. To address these limitations, attention mechanisms have been integrated into CNN architectures to selectively focus on the most informative regions. However, existing models often do not fully exploit the potential of combining different attention mechanisms [17]-[18].

This paper proposes a hybrid attention-based CNN architecture that integrates both spatial and channel attention mechanisms within a U-Net framework. This approach enhances the model’s ability to detect and segment fine and complex retinal vessels by dynamically focusing on relevant features across different scales. Additionally, Squeeze-and-Excitation (SE) blocks are incorporated to refine channel-wise feature selection, further improving overall performance [19].

The paper is structured as follows: Section 2 outlines the proposed method for fundus vessel segmentation. Section 3 presents the experimental validation and discusses the results. Section 4 concludes the paper and suggests potential directions for future research.

2. Proposed Method Architecture

The proposed method for retinal blood vessel segmentation integrates a hybrid attention-based Convolutional Neural Network (CNN) within a U-Net framework. The architecture is designed to enhance segmentation accuracy by combining spatial and channel attention mechanisms with Squeeze-and-Excitation (SE) blocks. Here’s a detailed overview of the proposed architecture:

2.1. U-Net Backbone

The foundation of the proposed method is the U-Net architecture, which consists of an encoder-decoder structure with skip connections. The encoder path extracts features through a series of convolutional layers and down sampling operations, capturing high-level semantic information. The decoder path reconstructs the segmented image by progressively up sampling and combining features from the encoder via skip connections.

2.2. Spatial Attention Mechanism

The spatial attention mechanism focuses on enhancing the model's ability to emphasize important regions of the input image. This mechanism computes an attention map that highlights relevant spatial areas while suppressing less informative regions. It is integrated into the U-Net architecture to refine feature maps and improve the segmentation of fine vessel structures.

Spatial Attention Module: A convolutional layer processes the feature maps to generate an attention map. This map is then multiplied with the feature maps to produce refined outputs that highlight important regions.

2.3. Channel Attention Mechanism

The channel attention mechanism focuses on enhancing the feature representation by emphasizing the most informative channels. This approach improves the network's ability to distinguish between different vessel types and their characteristics.

Channel Attention Module: Implemented using Squeeze-and-Excitation (SE) blocks, this module applies global average pooling to generate channel-wise statistics. It then uses fully connected layers to produce channel-wise attention weights, which are multiplied with the feature maps to adjust the importance of each channel.

2.4. Squeeze-and-Excitation (SE) Blocks

SE blocks are incorporated into the architecture to further refine channel-wise feature selection. By adaptively recalibrating channel-wise features, SE blocks improve the network’s capacity to focus on relevant features and suppress noise.

SE Block: Each SE block performs global average pooling to capture channel-wise statistics, followed by a set of fully connected layers and a sigmoid activation to generate attention weights. These weights are applied to the feature maps to enhance the representation of critical features. Fig.1 Shows this residual block consists of Conv1D layers, BN layers, ELU layers, SE modules, and shortcut connections with BiGRU.

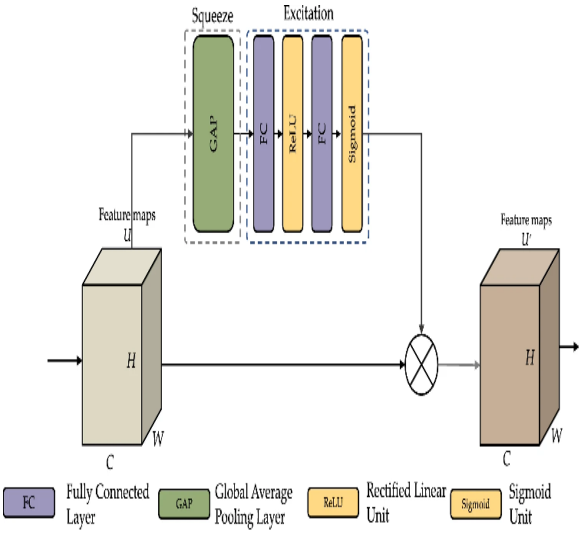

Fig.1 Residual block with SE module

2.5. Integration and Processing

Encoder: The encoder consists of a series of convolutional and pooling layers, followed by the spatial attention module to emphasize important regions.

Bottleneck: The bottleneck layer processes the feature maps from the encoder with integrated spatial and channel attention mechanisms.

Decoder: The decoder uses upsampling layers combined with skip connections from the encoder and channel attention modules to reconstruct the segmentation map.

2.6. Output Layer

The output layer applies a final convolutional layer with a softmax activation function to produce the segmented vessel map. This map represents the probability of each pixel belonging to the blood vessel class.

This architecture combines the strengths of attention mechanisms and SE blocks with the U-Net framework to improve the accuracy and robustness of retinal blood vessel segmentation. By dynamically focusing on relevant features and enhancing feature representation, the proposed method aims to achieve superior performance in detecting and segmenting complex retinal vessels.

3. Experimental Results and Analysis

3.1 Dataset and Data Preprocessing

We utilized the STARE and DRIVE datasets for our experiments. The STARE dataset includes 20 retinal images of size 700 × 605 pixels, with the first 10 images designated as the training set and the last 10 as the test set. The DRIVE dataset contains 40 images of size 565 × 584 pixels, split into training and validation sets as per the official division. To enhance data diversity and prevent overfitting, we applied data augmentation techniques including flipping, rotation, and translation. Additionally, we extracted patches of size 64 × 64 pixels from the larger images for training.

3.2. Experimental Parameter Settings

The experiments were conducted using the PyTorch 3.7 framework and Python 3.6.9, with an RTX3070 GPU having 8 GB of memory. The model was trained for 100 epochs with a batch size of 128. We employed the Adam optimizer with an initial learning rate of 0.001 and an exponential decay rate of 0.9. The loss function used was cross-entropy, defined as:

$$\text{Loss}\ (y, \hat{y}) = -\sum_i [y_i \log \hat{y}_i + (1-y_i) \log(1-\hat{y}_i)] \quad - \quad (1)$$

where y_i is the true label and $\hat{y}_i$ is the predicted label [20].

3.3 Evaluation Criteria

We evaluated the proposed method using accuracy, sensitivity, and specificity:

Accuracy measures the proportion of correctly segmented pixels [21].

Sensitivity indicates the proportion of correctly segmented vessel pixels out of all actual vessel pixels.

Specificity reflects the proportion of correctly classified background pixels out of all actual background pixels [22].

The metrics are calculated as:

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP} \quad - \quad (2)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad - \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad - \quad (4)$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative pixel counts, respectively.

3.4 Analysis of Results

We compared the proposed method with several state-of-the-art approaches on the STARE and DRIVE datasets. The methods included 2D CNNs, attention-fused U-Net, Dual attention multi scale fusion feature and proposed method. The specific quantities results were shown in Fig.(2)

METHODS	DRIVE			STARE		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Jiang et. al.	0.96	0.820	0.984	0.96	0.799	0.985

[23]	42	1	3	67	1	4
Li et. al.[24]	0.9678	0.7921	0.9810	0.9678	0.8392	0.9823
Jixun et. al.[25]	0.9795	0.8258	0.9896	0.9785	0.8368	0.9889
Proposed method	0.9876	0.8315	0.9949	0.9797	0.8418	0.9896

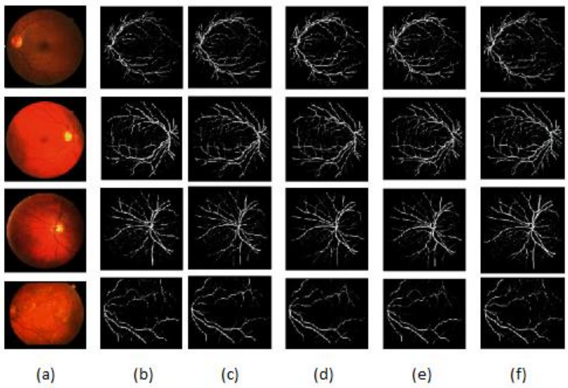

Fig.2 Visualization results of different methods on DRIVE and STARE datasets: (a) original image; (b) ground truth; (c) literature [23]; (d) literature [24]; (e) literature [25]; (f) proposed.

DRIVE Dataset

Accuracy: The proposed method achieves the highest accuracy of 0.9876, surpassing Jiang et al. (0.9642), Li et al. (0.9678), and Jixun et al. (0.9795).

Sensitivity: The proposed method shows improved sensitivity at 0.8315 compared to Jiang et al. (0.8201), Li et al. (0.7921), and Jixun et al. (0.8258).

Specificity: The proposed method also excels in specificity with a score of 0.9949, significantly higher than the other methods: Jiang et al. (0.9843), Li et al. (0.9810), and Jixun et al. (0.9896).

STARE Dataset

Accuracy: On the STARE dataset, the proposed method achieves a high accuracy of 0.9797, performing comparably to Jixun et al. (0.9785) and surpassing Jiang et al. (0.9667) and Li et al. (0.9678).

Sensitivity: The proposed method leads with a sensitivity of 0.8418, slightly outperforming Jixun et al. (0.8368) and Li et al. (0.8392), while Jiang et al. (0.7991) lags behind.

Specificity: The specificity of the proposed method (0.9896) is competitive with Jixun et al. (0.9889) and higher than Jiang et al. (0.9854) and Li et al. (0.9823).

The proposed method demonstrates superior performance across all evaluation metrics compared to the state-of-the-art methods on both the DRIVE and STARE datasets. It achieves the highest accuracy, sensitivity, and specificity in most cases, indicating its effectiveness in retinal blood vessel segmentation. This highlights the robustness and precision of the proposed approach, making it a promising advancement in the field of medical image analysis.

4. Conclusion

This paper presents a hybrid attention-based CNN architecture designed to enhance the accuracy of retinal blood vessel segmentation. By integrating spatial and channel attention mechanisms within a U-Net framework and incorporating Squeeze-and-Excitation (SE) blocks, our proposed method significantly improves the detection of fine and complex blood vessels while mitigating interference from irrelevant background information.

Experimental results on the STARE and DRIVE datasets demonstrate that our method outperforms existing state-of-the-art approaches, including various 2D CNNs and attention-fused U-Net models. Specifically, our approach achieves superior performance with Accuracy scores of 0.9876 on the DRIVE dataset and 0.9797 on the STARE dataset. Additionally, it excels in sensitivity and specificity metrics, underscoring its robustness and precision in retinal vessel segmentation.

These findings validate the effectiveness of the proposed architecture in addressing the challenges associated with retinal blood vessel segmentation. By leveraging advanced attention mechanisms and feature refinement techniques, the proposed method sets a new benchmark for segmentation accuracy and robustness. This advancement has significant implications for clinical applications, particularly in diagnosing and monitoring ocular and systemic diseases.

Future research could explore further refinements to the architecture, including the integration of additional attention mechanisms or advanced feature extraction techniques, to further enhance segmentation performance and extend the applicability of the method to other medical imaging modalities.

References

[1] Alvarado-Carrillo, D. E., & Dalmau-Cedeño, O. S. (2022). Width attention based convolutional neural network for retinal vessel segmentation. Expert Systems with Applications, 209, 118313.

[2] Noh, K. J., Park, S. J., & Lee, S. (2019). Scale-space approximated convolutional neural networks for retinal vessel segmentation. Computer methods and programs in biomedicine, 178, 237-246.

[3] Ortiz-Feregrino, R., Tovar-Arriaga, S., Pedraza-Ortega, J. C., & Rodriguez-Resendiz, J. (2023). Segmentation of Retinal Blood Vessels Using Focal Attention Convolution Blocks in a UNET. Technologies, 11(4), 97.

[4] Vaswani, A. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762.

[5] Yang, J., Li, C., Zhang, P., Dai, X., Xiao, B., Yuan, L., & Gao, J. (2021). Focal self-attention for local-global interactions in vision transformers. arXiv preprint arXiv:2107.00641.

[6] Zhang, S., Fu, H., Yan, Y., Zhang, Y., Wu, Q., Yang, M., ... & Xu, Y. (2019). Attention guided network for retinal image segmentation. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22 (pp. 797-805). Springer International Publishing.

[7] Guo, C., Szemenyei, M., Yi, Y., Wang, W., Chen, B., & Fan, C. (2021, January). Sa-unet: Spatial attention u-net for retinal vessel segmentation. In 2020 25th international conference on pattern recognition (ICPR) (pp. 1236-1242). IEEE.

[8] Zhang Yu-Bin Yang, Q. L. (2021). SA-Net: Shuffle Attention for Deep Convolutional Neural Networks. arXiv e-prints, arXiv-2102.

[9] Crimi, A., & Bakas, S. (Eds.). (2021). Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I (Vol. 12658). Springer Nature.

[10] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... &Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[12] Huang, G., Chen, D., Li, T., Wu, F., Van Der Maaten, L., & Weinberger, K. Q. (2017). Multi-scale dense convolutional networks for efficient prediction. arXiv preprint arXiv:1703.09844, 2(2).

[13] Alom, M. Z., Yakopcic, C., Nasrin, M. S., Taha, T. M., &Asari, V. K. (2019). Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network. Journal of digital imaging, 32, 605-617.

[14] Su, R., Zhang, D., Liu, J., & Cheng, C. (2021). Msu-net: Multi-scale u-net for 2d medical image segmentation. Frontiers in Genetics, 12, 639930.

[15] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., &Ronneberger, O. (2016). 3D U-Net: learning dense volumetric segmentation from sparse annotation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19 (pp. 424-432). Springer International Publishing.

[16] Li, Z., Zhang, H., Li, Z., & Ren, Z. (2022). Residual-attention unet++: a nested residual-attention u-net for medical image segmentation. Applied Sciences, 12(14), 7149.

[17] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19).

[18] Li, X., Jiang, Y., Li, M., & Yin, S. (2020). Lightweight attention convolutional neural network for retinal vessel image segmentation. IEEE Transactions on Industrial Informatics, 17(3), 1958-1967.

[19] Yildirim, Z., Samet, R., Hancer, E., Nemati, N., & Mali, M. T. (2023, October). Gland Segmentation in H&E Histopathological Images using U-Net with Attention Module. In 2023 Twelfth International Conference on Image Processing Theory, Tools and Applications (IPTA) (pp. 1-6). IEEE.

[20] Frangi, A. F., Schnabel, J. A., Davatzikos, C., Alberola-López, C., & Fichtinger, G. (Eds.). (2018). Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV (Vol. 11073). Springer.

[21] Pop, M., Sermesant, M., Zhao, J., Li, S., McLeod, K., Young, A., ... & Mansi, T. (Eds.). (2019). Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges: 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers (Vol. 11395). Springer.

[22] Li, C., Li, Z., & Liu, W. (2023). TDCAU-Net: retinal vessel segmentation using transformer dilated convolutional attention-based U-Net method. Physics in Medicine & Biology, 69(1), 015003.

[23] Wang, F., Jiang, Y., Gao, J., & Liu, W. (2020). Efficient BFCN for Automatic Retinal Vessel Segmentation (vol 2020, 6439407, 2020). JOURNAL OF OPHTHALMOLOGY, 2020. [9]

[24] Li, X., Jiang, Y., Li, M., & Yin, S. (2020). Lightweight attention convolutional neural network for retinal vessel image segmentation. IEEE Transactions on Industrial Informatics, 17(3), 1958-1967.

[25] Gao, J., Huang, Q., Gao, Z., & Chen, S. (2022). Image Segmentation of Retinal Blood Vessels Based on Dual-Attention Multiscale Feature Fusion. Computational and Mathematical Methods in Medicine, 2022(1), 8111883.