Grenze International Journal of Engineering and Technology, June Issue

A Patch-based Transformer Framework for Robust Cross-Dataset Retinal Vessel Segmentation

Pavithra G[1], Swapnil S N[2], M B Girish[3], Riya Keerthana[4], E. Tarun Skanda[5] and Pavan. Nandan[6]
[1-6]Dayananda Sagar College of Engineering, Bengaluru, India.
Email: dr.pavithrag.8984@gmail.com

***Abstract*— Retinal vessel segmentation plays a crucial role in automated diagnosis systems for ophthalmological diseases. However, achieving high accuracy across different datasets remains challenging due to variations in imaging conditions, noise, and vessel complexity.**

This work proposes a transformer-enhanced deep learning framework that combines patch-based learning, Gaussian-weighted reconstruction, and test-time augmentation (TTA) to improve both accuracy and robustness. The model is trained on the DRIVE dataset and evaluated on unseen datasets such as CHASE_DB1, STARE, and HRF without retraining.

The results demonstrate that the proposed approach not only achieves competitive segmentation accuracy but also maintains consistent performance across datasets, highlighting its ability to generalize well in real-world clinical scenarios.

***Index Terms*— Retinal vessel segmentation, Transformer, Deep learning, Patch-based learning, Cross-dataset generalization, Test-time augmentation, Medical image analysis.**

I. Introduction

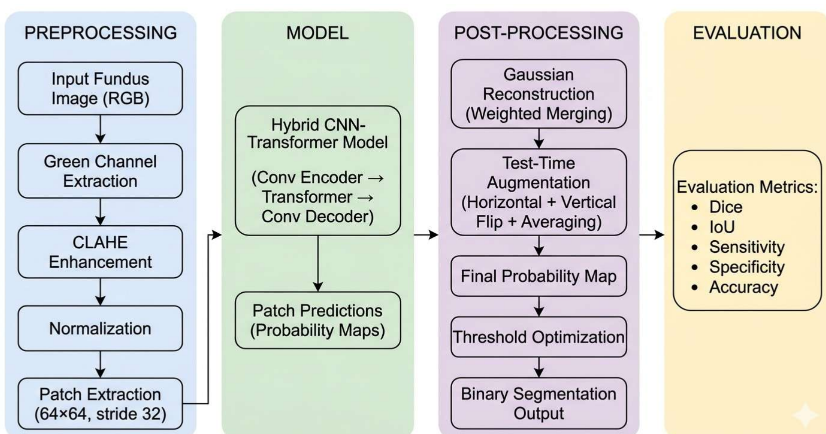

Fig. 1

Grenze ID: 01.GIJET.12.2.436_24

Retinal vessel segmentation is a fundamental step in analyzing fundus images for disease detection. The structure of blood vessels provides important clinical indicators such as vessel width, branching patterns, and abnormalities. Accurate segmentation enables early diagnosis of conditions like diabetic retinopathy and glaucoma. Traditional image processing techniques often fail due to Low contrast between vessels and background, Presence of noise, Thin and discontinuous vessel structures. Deep learning methods, especially CNN-based architectures like U-Net, improved segmentation significantly. However, they are limited by Local receptive fields, Inability to capture long-range dependencies, Poor generalization across datasets. To address these issues, transformer-based models are introduced. Transformers use self-attention mechanisms to capture global relationships in images. This work combines CNNs and transformers to leverage Local feature extraction (CNN), Global context modeling (Transformer). Additionally, practical challenges like cross-dataset variability are addressed through improved training and inference strategies.

II. Scope and Objectives of the Work

Scope:
This project focuses on building a complete segmentation pipeline, including:

- Data Preprocessing: Enhancing raw fundus images to improve vessel visibility.
- Model Architecture Design: Developing a hybrid CNN-transformer model for better feature representation.
- Patch-Based Learning: Dividing images into smaller regions to capture fine vessel structures.
- Post-Processing Techniques: Using Gaussian reconstruction and TTA to refine predictions.
- Cross-Dataset Evaluation: Testing model robustness across multiple datasets without retraining.

Objectives:

- Develop a segmenatation model capable of handling complex vessel patterns
- Improve detection of thin and low – contrast vessels
- Incorporate global attention mechanisms using transformers
- Reduce segmentation artifacts using Gaussian reconstruction
- Improve prediction reliability using test-time augmentation
- Ensure generalization across datasets
- Validate system performance through quantitative and ablation studies

III. Proposed Methodology and Block Diagram

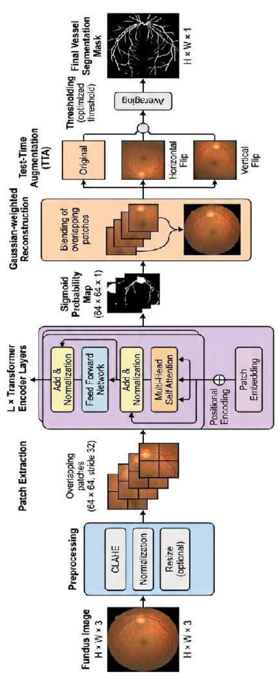

Fig 2

The proposed retinal vessel segmentation framework follows a structured pipeline consisting of preprocessing, patch-based learning, transformer-enhanced segmentation, and post-processing. Initially, the input fundus images are preprocessed by extracting the green channel, which provides better contrast for vessel structures. Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to enhance local contrast, followed by normalization of pixel values to ensure stable model training.

To effectively capture fine vessel details, each image is divided into overlapping patches of size 64 × 64 with a stride of 32. This patch-based approach increases the number of training samples and improves the model's ability to detect thin and low-contrast vessels. Data augmentation techniques such as horizontal and vertical flipping and rotation are also applied to improve generalization and reduce overfitting.

The segmentation model is based on a transformer-enhanced U-Net architecture. The encoder extracts hierarchical local features using convolutional layers, while the decoder reconstructs the segmentation map using upsampling and skip connections. A transformer module is introduced at the bottleneck to capture global contextual dependencies through self-attention, enabling better continuity of vessel structures.

The model is trained using a hybrid loss function combining Dice loss and Focal loss to address class imbalance and improve accuracy. During inference, predictions are generated patch-wise and merged using Gaussian-weighted reconstruction, which assigns higher importance to central pixels and reduces boundary artifacts. Additionally, test-time augmentation (TTA) is applied by averaging predictions from flipped versions of the input image to improve robustness. Finally, a threshold value of 0.3 is applied to obtain the final binary segmentation output, ensuring optimal performance.

IV. RESULT AND DISCUSSIONS

A. Performance on DRIVE Dataset

The proposed framework is evaluated on the DRIVE dataset to assess segmentation performance under standard conditions. The model achieves a Dice coefficient of 0.804, demonstrating its ability to accurately capture retinal vessel structures, including thin and low-contrast vessels. Table I shows the performance of proposed framework on the DRIVE dataset.

Qualitative segmentation results are shown in Fig.3, demonstrating the ability of the proposed framework to accurately capture both major and fine vessel structures.

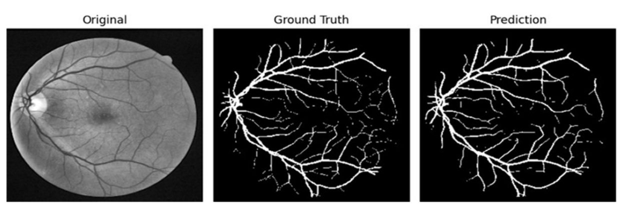

Fig. 3. Segmentation result upon the DRIVE datasets showing input image, ground truth, and predicted vessel map.

TABLE I: PERFORMANCE OF THE PROPOSED FRAMEWORK ON THE DRIVE DATASET

Model	Dice	IoU	Sensitivity	Specificity
Proposed Method	0.804	0.670	0.823	0.980

All results are obtained using the same preprocessing and inference pipeline to ensure fair comparison. The ROC curve shown in Fig.4 indicates strong classification performance with a high area under the curve (AUC) of 0.969.

V. ADVANTAGES AND DISADVANTAGES OF THE PROPOSED SYSTEM

A. Advantages

- Global Context Understanding: Transformer module captures long-range dependencies, improving continuity of vessel structures.
- Improved Detection of Fine Vessels: Patch-based learning helps in identifying thin and low-contrast vessels effectively.
- Reduced Boundary Artifacts: Gaussian-weighted reconstruction ensures smooth merging of overlapping patches.
- High Robustness: Test-Time Augmentation (TTA) reduces prediction variance and improves consistency.

- Better Generalization: Performs well across multiple datasets without retraining.
- High Segmentation Accuracy: Combination of Dice loss and Focal loss improves overall performance.

Disadvantages:

- High Computational Complexity: Transformer modules increase memory and processing requirements.
- Increased Inference Time: Patch-based processing and TTA make prediction slower.
- Resource Intensive: Requires powerful hardware (GPU) for efficient training and testing.
- Performance Drop on Unseen Data: Slight decrease in accuracy due to variations in datasets.
- Parameter Sensitivity: Performance depends on tuning parameters like patch size, stride, and threshold.
- Complex Implementation: Integration of multiple components makes the system harder to design and optimize.

VI. APPLICATIONS

- Medical Diagnosis
- Ophthalmology Research
- Clinical Decision Support Systems
- Automated Screening Systems
- Biomedical Image Analysis
- AI-based Healthcare Tools
- Telemedicine Applications

VII. CONCLUSIONS

The proposed retinal vessel segmentation framework combines a transformer-enhanced U-Net with patch-based learning and advanced post-processing techniques to achieve accurate and robust segmentation. The model effectively captures both local and global features, improving the detection of thin and complex vessel structures. Techniques such as Gaussian-weighted reconstruction and test-time augmentation further enhance the smoothness and reliability of the results. The model achieves strong performance on the DRIVE dataset and maintains good generalization across unseen datasets. Overall, the framework provides a reliable solution for automated retinal vessel segmentation in medical applications.

ACKNOWLEDGEMENT

The authors would like to thank the following 6th Semester students of Dayananda Sagar College of Engineering for their contribution. Madhura, Prathiksha, Nandan, Girish, Basavraj Patil, Purab jain, E Tarun, Nikhil, Dhanya, Keerthana, Pavan, Sharanpatil, Suryansh, Bharath, Riya and Rakshitha

REFERENCES

[1] G. Saravanan, R. Punitha Gowri, Arvinfh S. V, Dheena R. S, and Krishna Kumar M, “Survey on Remote Monitoring System,” 2024 9th International Conference on Communication and Electronics Systems (ICCES).

[2] Li Tung Tan, Raihanah Syamimi Mohammad Shahrizal, Muhammad Naqib’Aqil Mohd Nadzri, Norashikin Yahya, and Suhaila Badarol Hisham, “Remote Patient Monitoring System,” 2022 IEEE 5th International Symposium in Robotics and Manufacturing Automation (ROMA).

[3] Md Ameenul Hasan and Megha P. Arakeri, “Remote Patient Monitoring System Using IoT and Artificial Intelligence: A Review,” 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC).

[4] K., Jyotish Chandra; Jishnu; Ashirwaad; Sarbesh; G., Pavithra; Ninawe, Swapnil S. Ninawe,” Flower Classification using Pre-trained ResNet Models in Computer Vision” Grenze International Journal of Engineering & Technology (GIJET), 2024, Vol 10, Issue 2,Part 6, p6737. https://www.scopus.com/record/display.uri?eid=2-s2.0-85209065920&origin=resultslist

[5] Shweta Goyal, Sachine Sharma, Pushpendra Singh Kharayat, Sugam Gupta-"E-Waste: Promoting the Need for Green Electronics", 2023 International Conference on Power Energy, Environment & Intelligent Control (PEEIC), IEEE, 2023.

[6] Tejas, R.; L., Venkatesh; D. N., Shivamadu; Ravishankar, K.; Ninawe, Swapnil S. Ninawe; G. H., Samyama Gunjal; G., Pavithra; Manjunath, T. C., “Design & Development of an Automatic Defense System for Warfare Applications” Grenze International Journal of Engineering & Technology (GIJET), 2024, Vol 10, Issue 2,Part 6, p6711. https://www.scopus.com/record/display.uri?eid=2-s2.0-85209068806&origin=resultslist

[7] Shobha A.S. & T.C. Manjunath, “Design & development of transmitted & encrypted datas using SDN and energy self-healing concepts used in RF energy harvesting wireless sensor nets”, 2022 Third IEEE International Conference on

Intelligent Computing Instrumentation and Control Technologies (ICICICT), Kannur, Kerala, India, pp. 686-689, 11-12 August 2023
[8] G Pavithra, Swapnil S Ninawe, KV Sandeep, Iffath Fawad, TC Manjunath, “Development of hybrid algorithms for the detection of Infectious Diseases from the X-Ray’s images of the Chest Structure of Human Beings” 2024 2nd International Conference on Networking, Embedded and Wireless Systems (ICNEWS). https://www.scopus.com/record/display.uri?eid=2-s2.0-85210423361&origin=resultslist
[9] Abolfazl Younesi, Hamed Shahbazi Fard, Alireza Belal Yengikand, Vahid Pezeshki-"Survey on IoT-based waste Management Systems", 2022 8th International Conference on Web Research (ICWR), IEEE, 2022.
[10] Swapnil S. Ninawe, Pavithra G., T.C. Manjunath, Sandeep K.V., Iffath Fawad, Chetan Umadi, Padmavathi M., Ashwini Gowda H.B., “A Method of Designing Authentication Scheme for Aadhaar User” Journal of Information Systems Engineering and Management, vol 10 issue 11, pp 97 – 108, 2025
[11] J. Schlemper et al., "Attention gated networks for medical image analysis," Medical Image Analysis, 2019.
[12] Q. Li et al., "Deep learning for retinal vessel segmentation: A review," IEEE Access, 2020.