

Phase and structure based input enhancement for retinal vessel segmentation

Ondrej Straka¹, Filip Zubek¹, Michal Kovac¹, Jarmila Pavlovicova¹, Veronika Kurilova^{1,2}

This paper presents a study of retinal vessel segmentation that combines traditional image processing methods with deep learning. Two edge detection techniques, Phase Stretch Transform (PST) and B-COSFIRE filters. Both were applied as preprocessing steps to enhance vessel structures before segmentation using a U-net. PST parameters were optimized via a parallel genetic algorithm, and the resulting vessel maps replaced one channel in the RGB color space of the original fundus images. The method was evaluated on five public datasets (DRIVE, CHASE DB1, HRF, TREND, and FIVES) using standard segmentation metrics. Results show improved performance on DRIVE, CHASE and HRF datasets, and comparable or slightly improved performance on other datasets, indicating that spectral and structural enhancement can be complementary to deep learning methods without increasing computational complexity.

Keywords: segmentation, retinal vessels, eye, U-net, PST, B-COSFIRE

1 Introduction

Effective segmentation of retinal blood vessels enables timely diagnosis of ophthalmological and systemic diseases, including diabetic retinopathy [1]. While numerous deep learning-based approaches have been proposed, U-Net and its variants remain among the most widely adopted architectures for biomedical image segmentation [2]. Initial approaches to segmenting vessels involve traditional methods, such as thresholding the green channel in RGB color space, morphological operations, and edge detection methods. These methods provided the foundation for the retinal vessel segmentation, but do not use contextual information. Edge detection algorithms do not distinguish between vessel edges and edges from other physiological or pathological findings [3, 4].

As a next step, deep learning models such as U-Net enabled the use of contextual information, resulting in more accurate segmentations. This approach requires high-quality expert-labeled ground-truth (GT) annotations, where fine vessels span only a few pixels in width. Such masks are difficult to obtain and thus rare [5].

One possible solution is a combination of these two approaches, built upon our previous work [6], in which we demonstrated that a mask created by modifying the phase component of the Discrete Fourier Transform (DFT) image spectrum and subsequently used to replace the blue channel in the RGB color space can improve the vessel segmentation performance of U-Net models. The goal of the current study is to extend our previous work and examine the impact of incorporating additional

vessel-enhancing features (the Phase Stretch Transform (PST) [7, 8] and B-COSFIRE (Bar-selective Combination of Shifted Filter Responses) filters [9, 10]), into the input representation in form of preprocessing method, and subsequently evaluating its impact on the learning dynamics and performance of standard U-Net architectures for retinal vessel segmentation. Additionally, we validated our approach across multiple publicly available datasets of retinal blood vessel segmentation, enabling a more robust and comprehensive evaluation.

2 Related work

Retinal vessel segmentation is a foundational task in automated retinal analysis, offering critical insight into systemic and ocular diseases. Methods range from classical hand-crafted pipelines to modern deep learning-based models that leverage complex network architectures and enhanced input processing.

U-Net and its many derivatives remain the most widely used backbone in retinal vessel segmentation. These architectures combine encoder-decoder pathways with skip connections, enabling dense, spatially precise predictions. However, performance limitations have prompted various innovations.

Galdran et al. [11] investigated the necessity of high model complexity for achieving state-of-the-art performance in retinal vessel segmentation. The authors demonstrate that a minimal U-Net model with approximately 34,000 parameters can achieve competitive results when trained under a rigorous protocol. They

¹ Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Ilkovicova 3, Bratislava 812 19, Slovakia

² Faculty of Medicine, Slovak Medical University in Bratislava, Limbova 12, 833 03 Bratislava, Slovakia

*correspondence to veronika.kurilova@stuba.sk

further introduce a cascaded architecture, called Little W-Net, consisting of two sequential U-Nets with approximately 68,000 parameters in total. Despite its simplicity, this architecture consistently outperforms existing approaches with significantly fewer parameters. The evaluation focuses on three widely used retinal fundus image datasets: DRIVE, CHASE-DB1, and HRF. On these datasets, the Little W-Net achieves area under the ROC curve (AUC), Dice coefficient, and Matthews Correlation Coefficient (MCC) scores that surpass those of previously published methods, including deep architectures with tens of millions of parameters. Specifically, the model achieves AUCs of 98.10% on DRIVE, 98.47% on CHASE-DB1, and 98.25% on HRF, with corresponding Dice and MCC values consistently exceeding 80%. The findings indicate that carefully designed and well-trained lightweight models can achieve state-of-the-art performance in retinal vessel segmentation, challenging the necessity of increasingly complex architectures and providing a strong baseline for future research in efficient and reproducible medical image analysis.

Kun Sun et al. [12] propose SDAU-Net, an improved U-Net-based architecture for retinal vessel segmentation that effectively addresses the challenges of segmenting small vessels with complex morphology and dealing with low-contrast or lesion-affected images. Their approach introduces a Series Deformable Convolution (SDC) module to adaptively extract features from vessels of varying shapes and thicknesses, improving robustness and mitigating vanishing/exploding gradient issues. Additionally, they incorporate a Lightweight Attention Mechanism (LAM) and Dual Attention Mechanism (DAM) to enhance feature discrimination, particularly by improving global feature interaction and combining spatial and channel dependencies. The model was evaluated on four public datasets: DRIVE, STARE, CHASE DB1, and IOSTAR. SDAU-Net achieved superior segmentation accuracy across all datasets, with the highest sensitivity on STARE (89.73%) and best overall accuracy on IOSTAR (98.48%). Comparative and ablation experiments confirm that both SDC and attention modules contribute significantly to performance gains, especially in segmenting fine vessels in complex or low-contrast regions. Furthermore, cross-dataset and cross modality experiments demonstrate SDAU-Net's strong generalization and extendibility. Despite added complexity, the network remains computationally efficient, making it viable for clinical application.

Shin et al. [13] propose a novel deep-learning system for blood vessel segmentation called the Vessel Graph Network (VGN). Unlike standard CNN-based approaches that mainly rely on local pixel appearance, VGN explicitly incorporates vessel connectivity by

introducing a graph neural network (GNN) into the segmentation pipeline. A graph is constructed from an initial CNN prediction by sampling vertices and connecting them using geodesic distances, enabling information propagation along likely vessel paths. The CNN captures local texture cues, while the graph attention network (GAT) models global vessel structure, and both feature types are combined in a U-Net-like inference module. Experiments on four retinal vessel datasets and a coronary X-ray angiography dataset show that VGN achieves state-of-the-art or comparable performance in average precision and AUC, with statistically significant improvements over baselines. The method particularly improves detection of weak vessels and reduces some false positives by enforcing structural consistency.

In this paper, we examine two image processing techniques used as preprocessing steps to enhance vessel information in retinal images. The goal is to extend our previous research [6] and explore alternative methods, validating them on multiple public datasets.

3 Methodology

The proposed methodology is based on the idea that traditional vessel-enhancement techniques can complement deep learning segmentation models without modifying the network architecture or increasing computational complexity. Instead of changing the U-net itself, we modify the input representation by integrating vessel-focused structural information directly into one of the RGB channels, as a preprocessing step. The workflow begins with background normalization. Each fundus image is processed using either a dataset-provided mask or an automatically generated circular mask to suppress non-retinal regions. This step is essential, because background noise or illumination artefacts can significantly disrupt traditional edge detection methods such as PST or B-COSFIRE. Once the background is normalized, the image is suitable for subsequent processing.

After normalization, the image is enhanced by two independent edge detection methods. The Phase Stretch Transform [7] with our genetic algorithm for parameter optimization and B-COSFIRE [9]. Each method outputs an edge map that highlights the retinal vasculature. To maintain consistency with the intensity distribution in the original RGB channels, the resulting edge maps are inverted so that vessel pixels correspond to low-intensity values. This decision ensures that vascular structures appear consistently dark across all three channels, which simplifies the learning dynamics of convolutional filters in the segmentation network. The whole preprocessing pipeline is shown in Fig. 1.

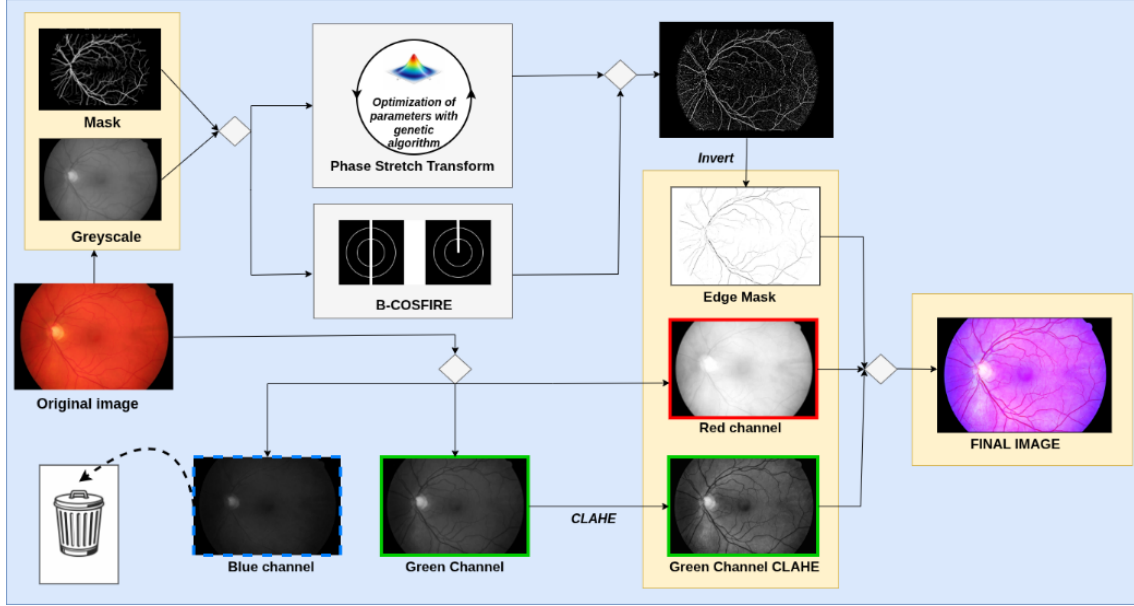


Fig. 1. Visualisation of preprocessing pipeline

The inverted edge map (see Fig. 2) replaces the blue channel of the original image. This substitution preserves the three-channel input format required by the ResNet-34 encoder while injecting vessel-specific structural information into the representation. Because many pre-trained encoders are designed for natural RGB images, retaining the three-channel structure avoids architectural changes and maintains compatibility with established training pipelines. Additionally, we explored a combined representation of the PST and BCOSFIRE filter vessel masks. While each method can produce false positives or miss certain vessel segments under challenging conditions, their combination can potentially reduce these weaknesses by reinforcing consistent detections and suppressing method specific noise. We evaluated two fusion strategies. First, we applied a logical AND (Pixel-wise logical AND) operation between the two masks, retaining only pixels where both PST and BCOSFIRE filter agree, resulting in a conservative mask emphasizing high-confidence vessel structures. Second, we performed an additive fusion by summing (Pixel-wise sum) both masks and rescaling the resulting image back to the $[0, 255]$ range. In this setting, pixels detected by both methods obtain higher intensities, pixels detected by only one method receive intermediate intensities, and low values correspond to regions where neither method detects vessel-like structures. After creating these modified inputs, we train a U-net with a ResNet-34 encoder on all datasets in two modes: using the original unmodified images and using

the PST- or B-COSFIRE- and combined-enhanced versions. This setup enables a direct comparison of segmentation performance and allows us to examine whether the injected structural information leads to measurable improvements.

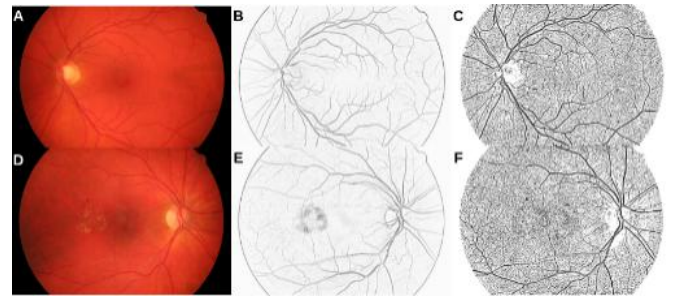


Fig. 2. Comparison of B-COSFIRE (B, E) and PST (C, F) on healthy retina (A) and retina with diabetic retinopathy (D)

3.1 Datasets

For our experiments, we used five different public datasets. We aimed to utilize public datasets that are heterogeneous, vary in size and capturing method (fundus camera vs. smartphone), and contain healthy and diseased eyes. Summarized information about all the datasets is in Tab. 1.

Table 1. Summary of the datasets used in the experiments

Dataset	Number of images	Resolution (px)	Unique features
DRIVE	40	565×584	Most used database
HRF	45	3504×2336	High-resolution images, include healthy, diabetic retinopathy, and glaucoma cases
CHASE_DB1	28	999×960	Captured from children, introduces age-related variability, multi-ethnicity, tortuous vessels
TREND	82	2560×1920	Mobile-acquired images, old and young patients
FIVES	800	2048×2048	Large dataset

The DRIVE dataset [14] consists of 40 retinal fundus images (565×584 pixels), equally divided into training and test sets. The images were acquired from a diabetic retinopathy screening program in the Netherlands and included manual vessel annotations by two independent experts. The dataset has well-defined vascular structures, high contrast between vessels and the background, and reliable ground-truth annotations, making it a benchmark dataset for segmentation tasks.

The HRF [15] dataset is unique for its high-resolution (3504×2336 pixels) images, which provide finer vascular detail than other fundus datasets. It contains 45 images divided into three categories: healthy, diabetic retinopathy, and glaucoma cases. The HRF dataset is valuable for evaluating segmentation models under varying image quality conditions and disease states. The higher resolution also makes it useful for training deep-learning models that require fine-grained vascular segmentation.

The FIVES [16] dataset includes 800 retinal fundus images with manually annotated vascular structures. What sets FIVES apart is its size and the inclusion of images with different color distributions, contrast settings and lightning conditions. This diversity ensures that models trained on FIVES can generalize better across different imaging hardware and clinical environments. The CHASE_DB1 [17] dataset consists of 28 retinal fundus images (999×960 pixels) captured from multi-ethnic children in the CHASE (Child Heart and Health Study in England) project. This dataset introduces age-related variability, altogether with slight differences in children’s vasculature compared to adults.

The CHASE project also includes images with tortuous retinal vessels, as it investigates whether retinal vessel tortuosity is associated with early cardiovascular risk factors. These variabilities are important for the model’s ability to generalize across different vessel structures and demographics.

TREND [18] dataset contains various retinal fundus images focusing on vessel tree segmentation. A unique aspect of TREND is that it consists of mobile-acquired images, introducing additional variability in illumination, motion blur, and focus inconsistencies. This makes TREND an essential dataset for evaluating model robustness in real-world imaging conditions, where lighting and acquisition settings may vary significantly.

For the FIVES and DRIVE datasets, we followed the original train–test split as defined by the dataset authors. For the HRF dataset, we selected the first five images from each of three disease categories. For the CHASE_DB1 and we randomly selected 9 images for testing. When constructing the test subset for the TREND dataset, we ensured that it included eyes from the “young” and “old” groups, while excluding images labeled as bad quality, as they lack segmentation masks and are unsuitable for evaluation. Final TREND test set consists of 18 randomly selected eyes from which 2 belong to old patients. Similarly, for the FIVES dataset, we further filtered out images of poor quality (e.g., blurry scans, shadows, or other artifacts obscuring parts of the retina).

3.2 Phase stretch transform

The Phase Stretch Transform (PST) is an image processing algorithm introduced in papers [7] and [8] for feature detection and enhancement. Unlike traditional edge detection methods, PST is particularly adept at uncovering subtle features in images. The process begins with transforming the image into the frequency domain using the Fourier transform. Next, a nonlinear operation stretches the phase of the transformed image to enhance its details. Finally, an inverse Fourier transform is applied to revert the modified image back to the spatial domain, resulting in sharper edges and more pronounced features.

In mathematical terms, the phase signal transform is defined by Eqn. (1). $A[n, m]$ represents the final phase image, and $B[n, m]$ the original image. Parameters n and m correspond to image coordinates in the spatial domain. FFT2 denotes the two-dimensional Fast Fourier Transform, and IFFT2 represents the two-dimensional Inverse Fast Fourier Transform. The term $\tilde{L}[p, q]$ refers to the localisation kernel in the frequency spectrum domain, often exemplified by the Gaussian filter. $\tilde{K}[p, q]$ denotes the warped phase kernel, described as a non-linear frequency-dependent phase filter. Such a filter can be seen in Fig. 3.

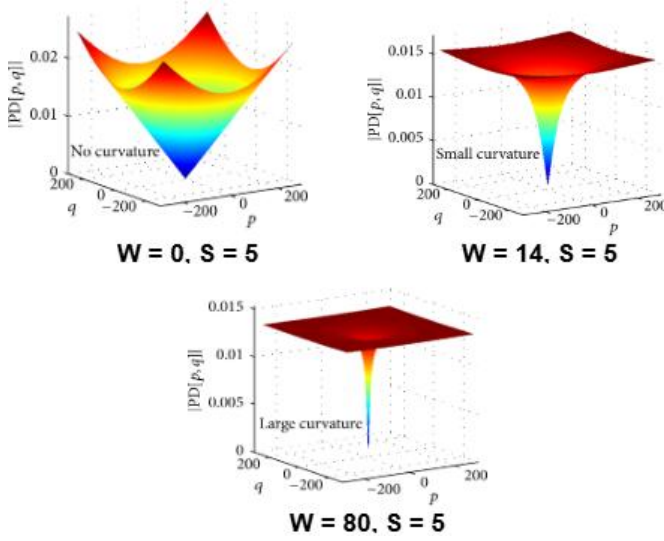


Fig. 3. Visualization of the warping effect on PST kernels from [8]

Implementation of the PST involves transformation of the phase kernel $\varphi[p, q]$ to $\varphi_{\text{polar}}[r]$, as seen in Eqn. (2). Parameters S and W are real-valued numbers corresponding to the strength and warp of the filter.

$$A[n, m] = \mathcal{F}\{\text{IFFT2}\{\tilde{K}[p, q] \cdot \tilde{L}[p, q]\} \cdot \text{FFT2}\{B[n, m]\}\} \quad (1)$$

$$\tilde{K}[p, q] = e^{j \cdot \varphi[p, q]}$$

$$\varphi[p, q] = \varphi_{\text{polar}}[r, \theta] = \varphi_{\text{polar}}[r]$$

$$\varphi_{\text{polar}}(r) = S \cdot \frac{Wr \tan^{-1}(Wr) - \frac{1}{2} \ln(1 + (Wr)^2)}{Wr_{\text{max}} \tan^{-1}(Wr_{\text{max}}) - \frac{1}{2} \ln(1 + (Wr_{\text{max}})^2)} \quad (2)$$

PST has multiple variable parameters that can be adjusted to optimize the filter's performance. We used a genetic algorithm to find specific parameter values for the dataset. This approach requires the GA to determine the optimal parameters beforehand. After determining the optimal global parameters, we generate the PST edge masks to enhance vessel information in the original image.

3.3 Genetic algorithm

Genetic algorithm is a widely used method for simple parameter optimization. It is capable of fast parameter optimization of simpler and medium-sized use cases like ours. Instead of an empirical approach, we decided to use GA to better approximate the PST hyper-parameters on the dataset used, which is key to improving the accuracy of the proposed method.

Based on preliminary experiments, we chose an island-model parallel genetic algorithm (hereafter, GA) for PST hyperparameter optimization. The GA employs a three-island architecture with migration according to the migration scheme shown in Fig. 4. Island A serves as a collector that refines locally optimal parameter sets, whereas Islands B and C focus on global and local search, respectively. To increase the GA's search ability, we employed the Reset of the Island C and the warming mechanism [19] on Island B. The migration, as shown in Fig. 4, resets on island C every 15th generation, and warming on island B every 10th generation.

Each island has a population size of 50 individuals, where each individual represents optimized parameters according to Eqn. (3).

$$\text{individual} = [S, W, \sigma_{\text{LPF}}, \text{thresh}_{\text{min}}, \text{thresh}_{\text{max}}] \quad (3)$$

To evaluate the population, a fitness function is used that employs the RGB image and the corresponding ground-truth masks from the training dataset. For each individual representing a parameter set, the PST is calculated for a batch of the training dataset. Calculated masks are then evaluated against ground truth masks, yielding TP = True Positives, FP = False Positives, FN = False Negatives, and TN = True Negatives, and precision. Final fitness is then calculated based on the chosen loss function. In case of experimental loss, the fitness is calculated according to (4), where $w_1 = -2$, $w_2 = 1.25$, $w_3 = 1.5$, $w_4 = 0$. As the fitness function being minimized, TP is rewarded through a negative weight, whereas FP and TN are penalized via positive weights. FN was intentionally omitted, since noisy PST responses make FN overwhelmingly dominant. FN values are consistently high, and including them would bias the

GA toward trivial solutions with minimum vessels. Assigning any meaningful weight would prevent the GA from effectively exploring the parameter space. The best solution is considered the individual with the best fitness after the 50th generation.

$$fitness = w_1 TP + w_2 TN + w_3 FP + w_4 FN \quad (4)$$

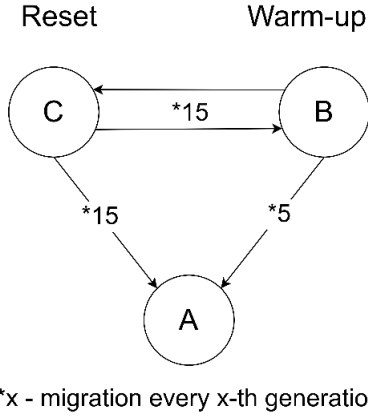


Fig. 4. Parallel genetic algorithm structure and migration

3.4 Bar-selective combination of shifted filter responses

B-COSFIRE [9, 10] is a trainable filter designed to detect bar-like patterns by combining the responses of Difference-of Gaussians (DoG) filters placed collinearly as shown in Fig. 5. The configuration is driven by a prototype pattern. In this case, a straight bar, from which the spatial arrangement of strong DoG responses is extracted. These positions, defined in polar coordinates relative to the pattern center, form the filter's support.

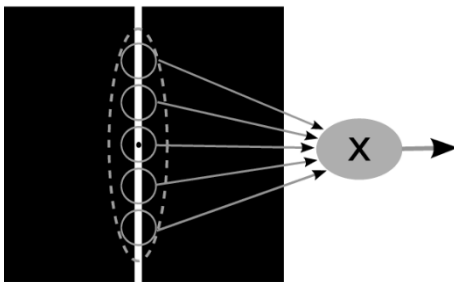


Fig. 5. Visualisation of B-COSFIRE response collection mechanism [9]

At runtime, the image is convolved with the same DoG filter. The B-COSFIRE filter samples the responses at the configured positions, applies spatial blurring to allow for tolerance, and shifts them toward the center. Next, responses are combined using a geometric mean. This results in a high response only when the image locally resembles the prototype.

The filter's selectivity is determined by the scale of the DoG kernel, the radial and angular tolerance parameters, and the number and spacing of support points. Rotation invariance is achieved by replicating the configuration at multiple orientations. The method is efficient, biologically inspired, and suitable for detecting elongated structures with variable curvature and intensity profiles.

3.5 Evaluation metrics

We employ a set of standard metrics commonly used in most papers to ensure the best comparability. The provided metrics are shown in Eqns. (5). BAcc stands for balanced accuracy and MCC stands for Matthews correlation coefficient.

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN}$$

$$\text{BAcc} = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right)$$

$$\begin{aligned} \text{MCC} &= \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \end{aligned} \quad (5)$$

4 Results

The results (see Tabs. 2 and 3) indicate that the model achieved better results for DRIVE, CHASE DB1 and HRF datasets and similar performance for TREND and FIVES on the original and the preprocessed datasets. One possible explanation for not helping on the TREND is the limited dataset size. Additionally, the overall quality of the images and annotations may play a significant role. Inconsistent labeling (e.g., variable boundary definitions or missing structures) and low contrast or noisy scans can introduce substantial uncertainty into the training data. This noise may limit the effectiveness

of preprocessing techniques and reduce the model's ability to learn robust, generalizable features, ultimately constraining achievable performance. In the case of FIVES, we see great results for both original and preprocessed datasets, which indicates that size of

FIVES dataset helped so much that additional preprocessing was not needed. These factors together likely contribute to the comparable performance and highlight the need for larger or higher-quality datasets to achieve more substantial performance differences.

Table 2. Comparison of results on different commonly used retinal vessel segmentation datasets

Experiment	DRIVE			CHASE_DB1			HRF		
	F1	AUC	MCC	F1	AUC	MCC	F1	AUC	MCC
Original	0.7724	0.9573	0.7535	0.8208	0.9376	0.8059	0.8155	0.9625	0.7926
PST	0.7834	0.9593	0.7644	0.8178	0.9291	0.8033	0.8362	0.9860	0.8017
B-COSFIRE	0.6320	0.9648	0.6282	0.6959	0.9457	0.8055	0.8097	0.9597	0.7863
Pixel-wise sum	0.7822	0.9665	0.7632	0.8210	0.9339	0.8061	0.8305	0.9719	0.7962
Pixel-wise logical AND	0.7855	0.9633	0.7667	0.8204	0.9483	0.8055	0.8290	0.9619	0.7977

Table 3. Comparison of results for different preprocessing methods on the TREND and FIVES datasets

Experiment	TREND			FIVES		
	F1	AUC	MCC	F1	AUC	MCC
Original	0.6687	0.8960	0.6544	0.9396	0.9936	0.9342
PST	0.6611	0.8792	0.6465	0.9279	0.9857	0.9213
B-COSFIRE	0.6686	0.8915	0.6543	0.9284	0.9870	0.9219
Pixel-wise sum	0.6663	0.8402	0.6519	0.9332	0.9940	0.9271
Pixel-wise logical AND	0.6668	0.8553	0.6524	0.9327	0.9938	0.9266

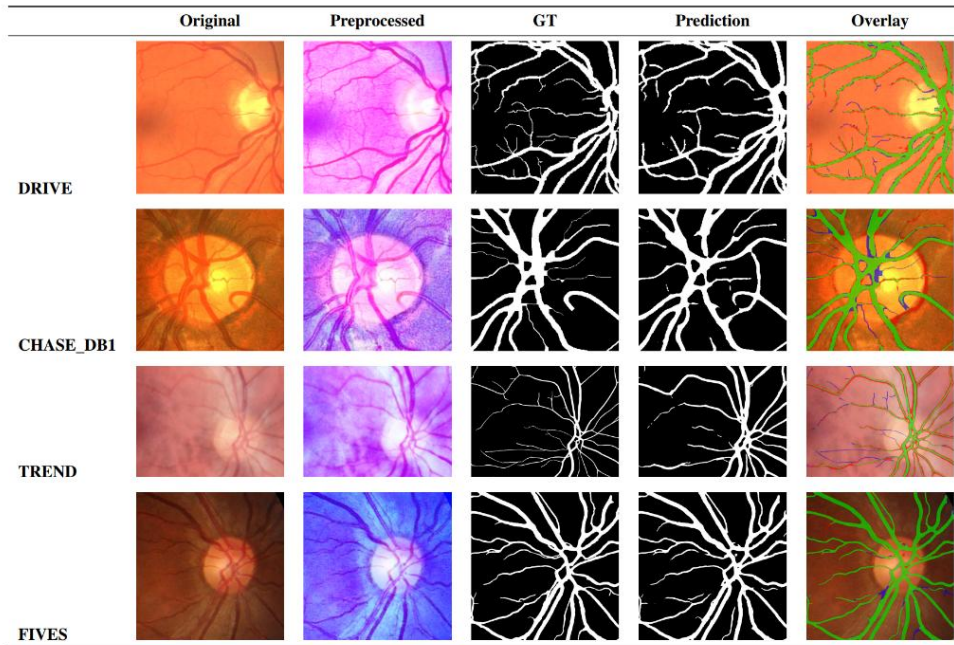


Fig. 6. Examples of original images and predicted vessels from used datasets. Second column (Preprocessed) are the results of our preprocessing method. The last column (Overlay) represents visualization between vessel predictions and GT masks (TP(green), FN(Blue), FP(Red)).

A qualitative evaluation was performed on several samples from the TREND dataset, comparing results obtained using the original and the modified data. The model produces nearly identical segmentation outputs in both cases, with similar distributions of false positives (red) and false negatives (blue). These errors are most pronounced at vessel bifurcations and in peripheral regions with thinner vessels. A closer examination suggests that part of these discrepancies originates in the image quality rather than in the model's performance, see Fig. 6. The ground truth vessel masks in the TREND dataset are often noticeably thinner than the actual vessel structures visible in the fundus images. As a result, predictions that visually correspond well to the vessel boundaries are sometimes penalized as false positives. This systematic mismatch between annotation thickness and the appearance of real vessels likely contributes to the comparable quantitative results observed across datasets. These findings highlight the importance of high-quality and geometrically consistent annotations, particularly when evaluating fine vascular structures in small datasets. The results on the FIVES dataset show only minimal differences between the original and modified versions. This suggests that the applied preprocessing or augmentation methods did not lead to a substantial improvement in segmentation performance. A plausible explanation lies in the considerably larger size of the FIVES dataset. It contains around 800 images, which is an order of magnitude more than in the other datasets used in this study. Such a volume of training data likely enabled the model to generalize effectively even from the original images, reducing the marginal benefit of additional preprocessing. In other words, once the dataset provides sufficient variability and representation of vessel patterns, the contribution of preprocessing becomes less significant. The results, therefore, indicate that the model may have reached a performance plateau on this dataset, where further improvements are limited more by data and annotation quality than by preprocessing strategies. See Fig. 6 for results example.

While most FIVES images exhibit accurate alignment between ground truth masks and corresponding fundus images, a few samples exhibit possible annotation inconsistencies (see Fig. 7). In particular, certain vessel structures appear slightly shifted relative to the underlying image content. This suggests that at least some ground-truth masks may not be perfectly registered with the image data. Such spatial inaccuracies can negatively affect evaluation metrics and lead to apparent false positives and false negatives, even when the model output visually aligns with the vessel pattern. This observation highlights the importance of verifying annotation accuracy and spatial alignment within large datasets, as even small positional errors can distort the assessment of segmentation performance.

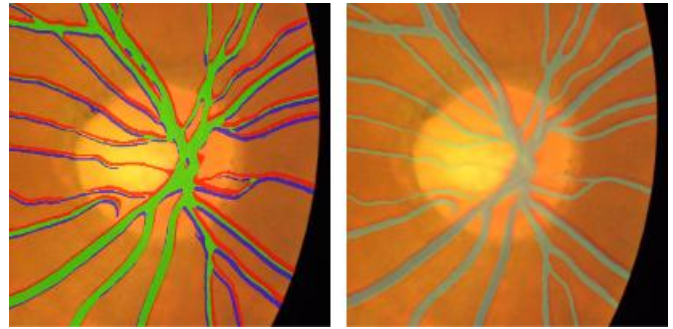


Fig. 7. Annotation error in image 023 in test set of FIVES dataset, visualization between vessel predictions and GT masks (left), original image with GT overlay (right) (TP(Green), FN(Blue), FP(Red))

The quantitative results on the DRIVE dataset show that the model achieves stable but only moderate improvements when using edge-based preprocessing methods such as PST or B-COSFIRE. This outcome can likely be attributed to the relatively low spatial resolution of the DRIVE images (approximately 565×584 pixels), which is significantly lower than in the other datasets used in this study. Due to the limited resolution, fine vessels, represented by only a few pixels in width (see Fig. 6), make it challenging for edge-enhancement techniques to accurately capture and emphasize them. Consequently, the potential benefits of preprocessing aimed at highlighting vessel boundaries become less pronounced, and the model performance remains close to that obtained with the original images.

For the CHASE DB1 dataset, the model achieved quantitative results comparable to those of the DRIVE dataset, but visual inspection revealed a distinct, recurring artifact not observed in any other dataset. Specifically, several false positive responses appear around the optic disc region, likely caused by strong local intensity gradients between the disc and the surrounding darker retinal area. In CHASE DB1, we observed an error in segmenting the optic disc border during prediction in both scenarios, with or without our preprocessing. False prediction is caused by a noticeable darker pigmentation around the disc, and this difference is even larger in our augmented images, (see Fig. 6). Moreover, in some ground truth masks, parts of the optic disc itself are annotated as vessels, which further contributes to over-segmentation in this region (see Fig. 8).

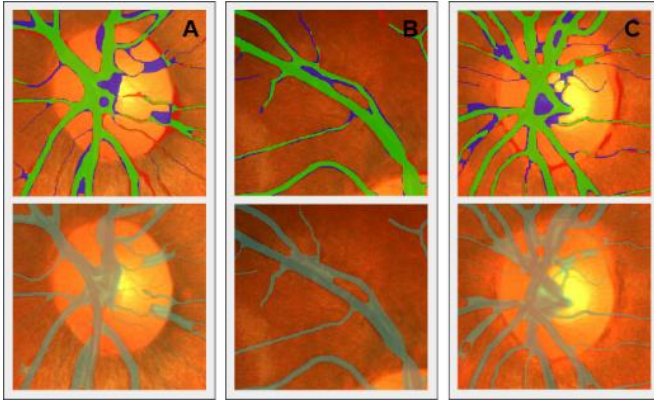


Fig. 8. Possible errors in GT masks in CHASE_DB1, visualization between vessel predictions and GT masks (first row), (TP(Green), FN(Blue), FP(Red)) and overlaid with GT mask (second row).

Column A: larger regions in optic disk annotated as vessels; column B: gap between two vessels annotated as vessel; column C: central part of the disk fully annotated as a singular large vessel and rectangular artifact in one vessel in top right.

Another recurring observation is that in regions where multiple vessels run in parallel or closely overlap, the segmentation masks tend to merge them into a single thick structure, suggesting possible limitations in the ground truth annotations or in the model's ability to distinguish adjacent thin vessels. Overall, these findings highlight how dataset-specific imaging conditions and annotation styles can significantly influence segmentation behavior and the perceived benefit of prepro-

cessing. While the proposed method did not achieve the best performance in every experimental setting, it yielded promising results on selected data sets.

Specifically, our approach reached performance levels comparable to, and in some cases exceeding, those of state-of-the-art models, suggesting its potential effectiveness under certain dataset conditions. This can be seen in Tab. 4.

Table 4. Results comparison with other authors

Experiment	DRIVE			CHASE_DB1			HRF		
	F1	AUC	MCC	F1	AUC	MCC	F1	AUC	MCC
Graph neural network [13]	0.8263	0.9801	-	0.8034	0.9830	-	0.8151	0.9838	-
lw-net [11]	0.8279	0.9810	0.8024	0.8169	0.9847	0.7974	0.8103	0.9825	0.7909
SDAU-Net [12]	-	0.9834	-	-	0.9858	-	-	-	-
PST	0.7834	0.9593	0.7644	0.8178	0.9291	0.8033	0.8362	0.9860	0.8017
Pixel-wise sum	0.7822	0.9665	0.7632	0.8210	0.9339	0.8061	0.8305	0.9719	0.7962

5 Conclusion

The proposed method explores combinations of traditional vessel-enhancement techniques with deep learning to improve retinal vessel segmentation performance. Our results show that while PST and B-COSFIRE preprocessing steps produced consistent outputs across all evaluated datasets, their overall contribution to quantitative performance was dataset dependent rather than universally beneficial.

The most notable improvement was observed in the HRF dataset, where high-resolution, consistent imaging conditions enabled the filters to effectively emphasize fine-scale vascular structures. In contrast, for datasets such as TREND or FIVES, the preprocessing method did

not yield noticeable gains, likely because our method is not well-suited to scenarios with greater variability in imaging and annotations. These findings suggest that dataset quality and annotation fidelity remain critical bottlenecks when evaluating modifications in input representation. The proposed enhancements demonstrated clear improvements at higher resolutions, confirming their potential to enhance segmentation performance under suitable conditions. Although these benefits were not consistently observed across all experimental settings, the findings highlight an important takeaway: traditional image processing techniques can still play a valuable role in modern deep learning pipelines, particularly when working with limited datasets or structures that are only faintly visible.

For future work, we plan to investigate patch-based training strategies and multi-task learning frameworks, potentially integrating vessel enhancement as an auxiliary learning signal rather than a fixed preprocessing step. This direction may improve sensitivity to fine vasculature while reducing dependence on handcrafted image operations. Expanding evaluation to larger and more diverse datasets would further validate the scalability and generalization potential of this approach.

Acknowledgement

This work was supported by the Slovak Research and Development Agency under the project APVV-22-606 AI4EYE.

References

- [1] C. Du, J. Zhuang, and X. Huang, "Deep learning technology in vascular image segmentation and disease diagnosis," *Journal of Intelligent Medicine*, vol. 1, no. 1, pp. 6–41, 2024. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jim4.15>
- [2] R. Azad, E. K. Aghdam, A. Rauland, Y. Jia, A. H. Avval, A. Bozorgpour, S. Karimijafarbigloo, J. P. Cohen, E. Adeli, and D. Merhof, "Medical image segmentation review: The success of u-net," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 10 076– 10 095, 2024.
- [3] T. Mapayi, S. Viriri, and J. R. Tapamo, "Adaptive thresholding technique for retinal vessel segmentation based on glcm-energy information," *Computational and Mathematical Methods in Medicine*, vol. 2015, p. 597475, 2015, epub 2015 Feb 24. [Online]. Available: <https://doi.org/10.1155/2015/597475>
- [4] J. Dash and N. Bhoi, "A thresholding based technique to extract retinal blood vessels from fundus images," *Future Computing and Informatics Journal*, vol. 2, no. 2, pp. 103–109, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2314728817300272>
- [5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.
- [6] O. Straka, F. Zubek, M. Kovac, J. Pavlovicova, and V. Kurilova, "Retinal vessel segmentation by u-net using modification of the phase component of the dft image spectrum," in *2025 International Symposium ELMAR*, 2025, pp. 311–314.
- [7] M. Suthar, H. Asghari, and B. Jalali, "Feature enhancement in visually impaired images," *IEEE Access*, vol. 6, pp. 1407–1415, 2018, epub 2017 Dec 5. [Online]. Available: <https://doi.org/10.1109/ACCESS.2017.2779107>
- [8] M. H. Asghari and B. Jalali, "Edge detection in digital images using dispersive phase stretch transform," *International Journal of Biomedical Imaging*, vol. 2015, p. 687819, 2015, epub 2015 Mar 23. [Online]. Available: <https://doi.org/10.1155/2015/687819>
- [9] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable cosfire filters for vessel delineation with application to retinal images," *Medical Image Analysis*, vol. 19, no. 1, pp. 46–57, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1361841514001364>
- [10] G. Azzopardi and N. Petkov, "Automatic detection of vascular bifurcations in segmented retinal images using trainable cosfire filters," *Pattern Recognition Letters*, vol. 34, no. 8, pp. 922–933, 2013, *computer Analysis of Images and Patterns*. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167865512003625>
- [11] A. Galdran, A. Anjos, J. Dolz, H. Chakor, H. Lombaert, and I. B. Ayed, "State-of-the-art retinal vessel segmentation with minimalistic models," *Scientific Reports*, vol. 12, no. 1, p. 6174, 2022, published April 13, 2022. [Online]. Available: <https://doi.org/10.1038/s41598-022-09675-y>
- [12] K. Sun, Y. Chen, Y. Chao, J. Geng, and Y. Chen, "A retinal vessel segmentation method based improved u-net model," *Biomedical Signal Processing and Control*, vol. 82, p. 104574, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809423000071>
- [13] S. Y. Shin, S. Lee, I. D. Yun, and K. M. Lee, "Deep vessel segmentation by learning graphical connectivity," *Medical Image Analysis*, vol. 58, p. 101556, Dec 2019.
- [14] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501– 509, 2004.
- [15] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images," *International Journal of Biomedical Imaging*, vol. 2013, pp. 1–11, 2013.
- [16] K. Jin, X. Huang, J. Zhou, Y. Li, Y. Yan, Y. Sun, Q. Zhang, Y. Wang, and J. Ye, "FIVES: A Fundus Image Dataset for Artificial Intelligence based Vessel Segmentation," *Scientific Data*, vol. 9, no. 1, p. 475, 2022. [Online]. Available: <https://doi.org/10.1038/s41597-022-01564-3>
- [17] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "Chase db1: Retinal vessel reference dataset," 2012. [Online]. Available: <https://researchdata.kingston.ac.uk/96/>
- [18] N. Popovic, S. Vujosevic, M. Radunovic, M. Radunović, and T. Popovic, "TREND database: Retinal images of healthy young subjects visualized by a portable digital non-mydiatic fundus camera," *PLOS ONE*, vol. 16, no. 7, p. e0254918, July 2021.
- [19] F. Zubek, M. Ernek, I. Kenický, and I. Sekaj, "Neuro-evolution of frequency controllers in the power-electricity system island," *IFAC-PapersOnLine*, vol. 58, no. 9, pp. 49– 54, 2024, 18th IFAC Conference on Programmable Devices and Embedded Systems PDES 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2405896324004592>

Received 30 January 2026