Unlocking Dark Vision Potential for Medical Image Segmentation

Hongpeng Yang[1], **Xiangyu Hu**[1], **Yingxin Chen**[2], **Siyu Chen**[2], **Srihari Nelakuditi**[1], **Yan Tong**[1*], **Shiqiang Ma**[3*] and **Fei Guo**[2*]

[1]Department of Computer Science and Engineering, University of South Carolina
[2]School of Computer Science and Engineering, Central South University
[3]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences

Abstract

Accurate segmentation of lesions is crucial for disease diagnosis and treatment planning. However, blurring and low contrast in the imaging process can affect segmentation results. We have observed that noninvasive medical imaging shares considerable similarities with natural images under low light conditions and that nocturnal animals possess extremely strong night vision capabilities. Inspired by the dark vision of these nocturnal animals, we proposed a novel plug-and-play dark vision network (DVNet) to enhance the model's perception for low-contrast medical images. Specifically, by employing the wavelet transform, we decompose medical images into subbands of varying frequencies, mimicking the sensitivity of photoreceptor cells to different light intensities. To simulate the antagonistic receptive fields of horizontal cells and bipolar cells, we design a Mamba-Enhanced Fusion Module to achieve global information correlation and enhance contrast between lesions and surrounding healthy tissues. Extensive experiments demonstrate that the DVNet achieves SOTA performance in various medical image segmentation tasks.

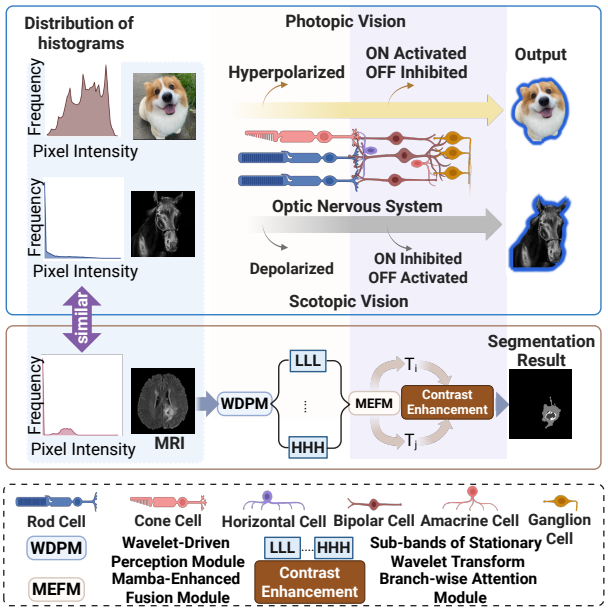

Figure 1: The proposed DVNet based on principles of biological vision aims to enhance lesion segmentation in MRI images by emulating the human vision mechanisms under dim lighting conditions.

1 Introduction

Accurate segmentation of lesion areas in medical images is a key step in diagnosing, planning treatments, and monitoring various diseases [Aglinskas *et al.*, 2022]. The precise excavation of disease markers directly affects subsequent clinical decisions and the effectiveness of interventions [Isensee *et al.*, 2021]. Due to the principles of noninvasive imaging, achieving high-precision segmentation is significantly challenged by factors such as artifacts, low contrast, and inherent tissue characteristics that result in blurry images. Despite the powerful capabilities of deep learning models based on Convolutional Neural Networks (CNNs) [Yang and Yu, 2021; Wang *et al.*, 2024c], Transformers [Dosovitskiy *et al.*, 2021], or Mamba [Gu and Dao, 2024], they struggle to effectively address interference caused by image blurriness, leading to unsatisfactory segmentation results.

Various methods have been attempted to mitigate this issue, including model-based deblurring, image enhancement, and the use of deep learning for feature extraction and image reconstruction [Lian and Wang, 2023]. For instance, some studies utilize CNNs to learn the mapping from blurry images to clear images, while others explore the application of Generative Adversarial Networks (GANs) [Goodfellow *et al.*, 2014] in image deblurring. In addition, efforts have been devoted to improving imaging equipment and acquisition techniques to reduce blurriness.

However, these methods often face common challenges. Firstly, the complexity and diversity of blurry images demand models with high adaptability and generalizability. Current deep learning models often require a large amount of annotated data for training, which is, however, scarce and expensive in the medical field. Secondly, key features may be lost

*Corresponding author tongy@cse.sc.edu, sq.ma@siat.ac.cn, and guofei@csu.edu.cn

due to blurriness, increasing the difficulty for models to learn these features. Lastly, although deep learning models perform well in object detection and classification, they still have limitations in interpreting the physical causes of blurry images and restoring clear images. How to enhance the model's understanding and processing capabilities of blurry images, especially with limited data, remains an open research question.

Through an in-depth understanding of the principles of photoreception and imaging of the eye, we have discovered that noninvasive medical imaging shares similarities with visual imaging in low light conditions, particularly in terms of contrast between the foreground and background [Hofmann and Lamb, 2023], as illustrated in Fig. 1. In addition, medical image processing and analysis exhibit a high degree of consistency with the contrast enhancement mechanisms of scotopic vision. In low light environments, rod cells are much more sensitive to dim light than cells [McKyton *et al.*, 2024]. This characteristic suggests that an effective medical image analysis method needs to capture subtle signals.

Inspired by the scotopic vision perception mechanism, we proposed a dark vision network (DVNet) with a wavelet-driven perception module (WDPM) to better perceive features in blurry image areas, as shown in Fig. 1. The wavelet transform [Tian *et al.*, 2023] can decompose an image into a sum of coarse to fine image features, representing different structures of the image. Using the wavelet transform, the model can more easily extract the structural and detailed information from the original image.

We further proposed a Mamba-Enhanced Fusion Module (MEFM) to simulate the antagonistic receptive fields of horizontal and bipolar cells, which enhance the contrast between the central and peripheral visual fields through increasing the brightness difference and thus achieve a highlight representation of the central area. The Mamba framework [Gu and Dao, 2024] combines state-space models with selective mechanisms, effectively processing long-sequence data, which contributes to obtaining global information and localizing lesions. Our method, under a global view, can enhance the contrast between the lesion area and surrounding healthy tissues, achieving perceptual activation of the lesion area. The attention mechanism endows the model with the ability to interact and associate directly between different positions within the sequence. By adaptively weighting, it enhances the features of key areas in the image and suppresses the visual response of surrounding areas, thereby achieving enhanced contrast. In summary, the key contributions of this paper are as follows.

- By delving into the mechanisms of scotopic vision, we proposed a novel DVNet with a WDPM and a MEFM to enhance the model's ability to perceive blurred areas in non-invasive medical imaging. To the best of our knowledge, DVNet is the first medical image segmentation algorithm inspired by scotopic vision.
- The DVNet is plug-and-play and can be deployed in various architectures based on CNNs, Transformers, or Mamba. Extensive experiments have demonstrated that the DVNet can consistently enhance the performance of its baseline model by enhancing the model's visual activation of lesion areas through contrast enhancement.

2 Related Work

2.1 U-Net and Its Variants

U-Net [Ronneberger *et al.*, 2015] has become a cornerstone in medical image segmentation due to its symmetric encoder-decoder structure and skip connections that enhance detail preservation. However, U-Net and its variants face limitations in modeling explicit long-range dependencies within images, which is critical for precise segmentation. To address the limitations of U-Net, researchers explored the integration of Transformers into U-Net. The integration of U-Net and Transformer, e.g., TransUNet [Chen *et al.*, 2024a], leverages the strengths of both architectures to enhance segmentation performance: restoring local spatial information through U-Net and modeling global context by Transformer.

2.2 Mamba in Medical Image Segmentation

Most recently, Mamba [Gu and Dao, 2024], a state space model, has emerged as a promising alternative to CNNs and Transformers in medical image segmentation. Mamba excels in modeling long-range interactions while maintaining linear computational complexity, which is particularly advantageous for handling large medical images. Mamba-UNet [Wang *et al.*, 2024d] further synergizes U-Net with Mamba's capabilities by adopting a pure Visual Mamba-based encoder-decoder structure, which facilitates comprehensive feature learning and captures both intricate details and broader semantic contexts within medical images. Mamba-UNet [Wang *et al.*, 2024d] has been shown to outperform various types of UNet in medical image segmentation under the same hyperparameter setting.

2.3 Perception of Blurry Information

A variety of methods have been proposed to deal with the issue of blurry medical images. An iterative edge attention network was introduced in [Wang *et al.*, 2022b], which employs an edge attention preservation module to reduce noise and help the edge flow focus on boundary-related information. Ma et al. [Ma *et al.*, 2024c] developed an edge feature fusion module to integrate the spatial correlation of the edges of the injury in adjacent images, achieving more precise edge segmentation. Qiao et al. [Qiao *et al.*, 2022] proposed an edge extraction method for medical images based on improved Local Binary Pattern combined with edge-aware filtering to suppress noise and enhance contrast while preserving edges.However, these methods typically improve edge detection capabilities by eliminating noise or incorporating additional information but do not fundamentally change the visual contamination issues caused by imaging blurriness.

2.4 Segmentation and Detection in the Dark

Low illumination can lead to degraded image quality including image blur and increased noise due to the low signal-to-noise ratio and reduced contrast between targets and backgrounds [Chen *et al.*, 2023]. To overcome these issues, a DNF model [Jin *et al.*, 2023] was proposed to decouple domain-specific subtasks, fully utilize the unique properties of the

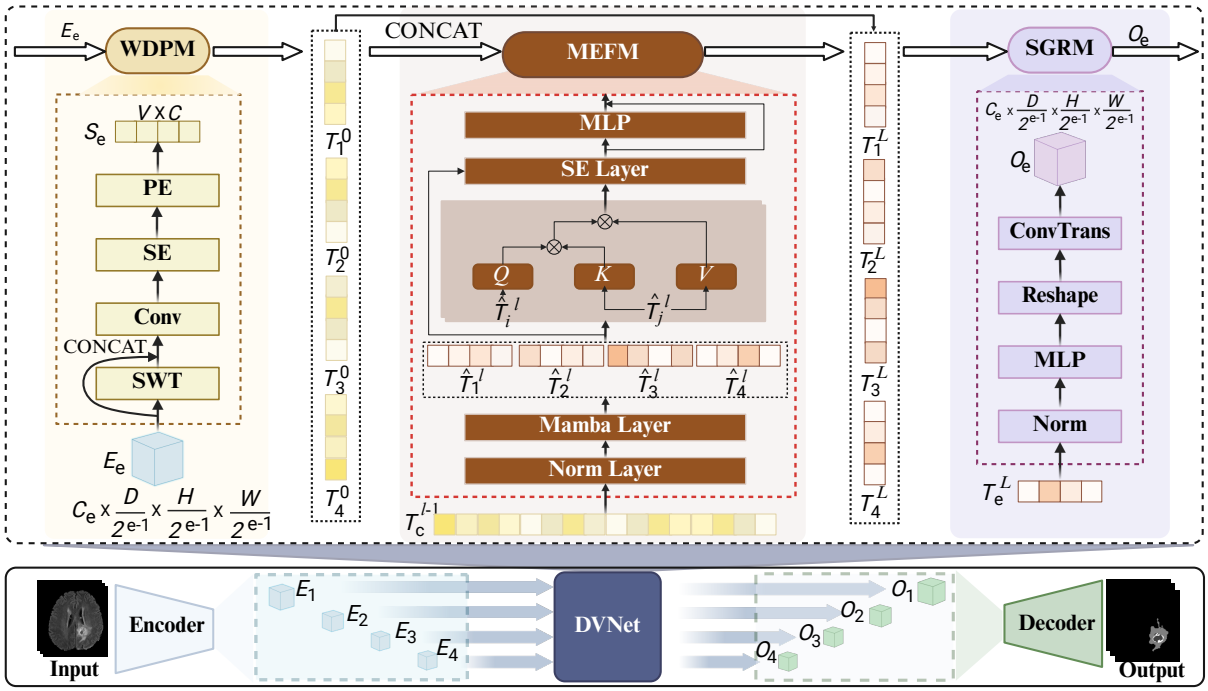

Figure 2: An overview of the proposed DVNet. The DVNet is implanted into the skip connection structure and plug-and-play. Three novel modules, i.e., WDPM, MEFM, and SGRM, are integrated to heighten the contrast between lesions and their surrounding healthy tissues, thereby improving lesion segmentation.

RAW and sRGB domains, and employ a feedback mechanism for feature propagation across stages, avoiding information loss caused by image-level data flow. Alternatively, an improved SFNet algorithm, i.e., SFNet-N [Wang *et al.*, 2022a], was proposed specifically for semantic segmentation of low-light autonomous driving road scenes, which includes a light enhancement network and a segmentation network with strong feature extraction capabilities to identify objects in dark environments. Wang et al. [Wang *et al.*, 2024b] proposed a method for multi-object tracking under low-light conditions by utilizing an adaptive low-pass downsampling module and a degradation suppression learning strategy to learn illumination-invariant features from low-light videos. Although these methods achieved success in different tasks under low-light natural scenes, they have not yet been applied to medical image analysis, which has similar conditions.

3 Methodology

DVNet represents a novel approach to mimicking the principles of dark vision by improving the sensitivity of the model to subtle image features and addressing the inherent blurriness in medical images. In this section, we elaborate on the three major components of the proposed DVNet: (1) WDPM, (2) MEFM, and (3) Spatial-Guided Refinement Module (SGRM). These components work synergistically to emulate the remarkable ability of nocturnal animals to perceive and process visual information in low light environments. Fig. 2 illustrates an overview of the proposed model.

3.1 Wavelet-Driven Perception Module

In dark vision, the ability of rod cells to perceive low contrast regions relies mainly on enhancing signals of different frequencies, which is critical to capture structural details in dim environments. However, traditional skip connections, commonly used in U-shaped networks, simply propagate low-level semantic information, limiting their ability to effectively process different frequency signals in blurry images. To address this limitation, the WDPM is developed by integrating a 3D stationary wavelet transform (SWT) module [Chen *et al.*, 2012] into each branch of our U-shaped network. SWT preserves spatial resolution while decomposing feature maps into distinct frequency subbands, enabling the network to distinguish between structural details and noise. As shown in Fig. 3, this enhances feature representation at a deeper semantic level and aligns with the principles of dark vision.

As illustrated in Fig. 2, given a 3D input volume $I \in \mathbb{R}^{C_0 \times D \times H \times W}$ with C_0 input channels, D slices (depth), spatial dimensions H (height) and W (width), the U-shaped architecture processes the input through its encoder structure, generating embedding $E_e \in \mathbb{R}^{C_e \times \frac{D}{2^{e-1}} \times \frac{H}{2^{e-1}} \times \frac{W}{2^{e-1}}}$ at different layers $(e \in \{1, ..., 4\})$ corresponding to varying levels and resolutions of features.

Then, in each branch, the input feature maps are first processed through a 3D SWT, which decomposes the volumetric data into multiple subbands, capturing both high-frequency edge details and low-frequency structural information. The decomposed subbands are concatenated with the original input feature maps. This operation combines the complemen-

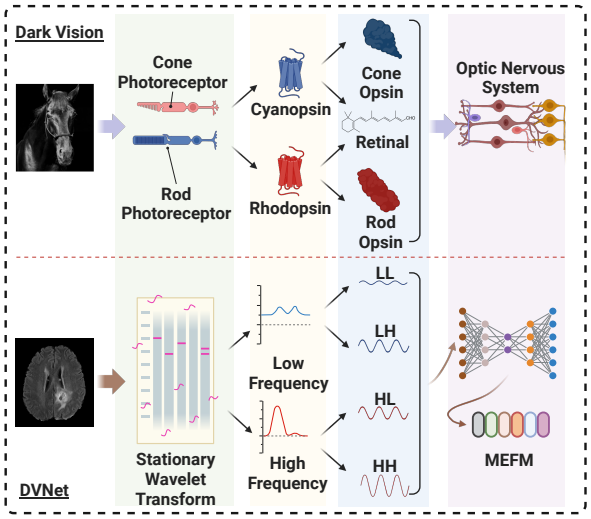

Figure 3: An illustration of the WDPM. By employing wavelet transform, the WDPM decomposes medical images into subbands of varying frequencies, mimicking the sensitivity of photoreceptor cells to different light intensities.

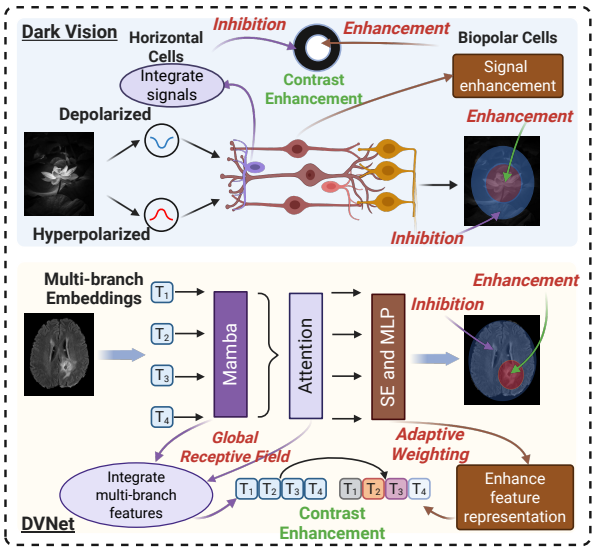

Figure 4: An illustration of the MEFM. Analogous to the working mechanism of the visual neural system, MEFM integrates attention mechanisms to achieve global information correlation and enhance the contrast between lesion areas and surrounding healthy tissues.

tary information from the frequency domain (SWT subbands) with the spatial features of the raw input, allowing the network to leverage both perspectives for improved feature representation. The concatenated features are passed through a 3D convolution layer to integrate spatial and frequency-domain information, extracting more discriminative features relevant for lesion segmentation. To enhance channel-wise feature dependencies, the convolution output is refined using a 3D Squeeze and Excitation (SE) block [Hu *et al.*, 2018]. This recalibrates the feature maps by dynamically weighting the channels based on their importance, ensuring that critical features are emphasized. Finally, to enable patches from multiple stages to focus on the corresponding regions in the original embeddings, the embeddings are tokenized by reshaping features from different scales into token sequences with patch sizes $\frac{P}{2^{e-1}}$ [Wang *et al.*, 2024a]. The channel dimensions are all changed to $C = 128$. Then, a spatial embedding module is used to capture positional information, improve spatial awareness of the model, and output the token sequences of embeddings $S_e \in \mathbb{R}^{V \times C}, V = \frac{DHW}{P^3}$.

3.2 Mamba-Enhanced Fusion Module

As shown in Fig. 4, MEFM is inspired by the working mechanisms of horizontal cells and bipolar cells in the visual neural system, where the Center-Surround Antagonism [Ankri *et al.*, 2020] enables edge detection and contrast enhancement within the visual cortex. By integrating attention mechanisms through Mamba, MEFM achieves global information correlation and enhances the contrast between lesions and surrounding healthy tissues.

MEFM integrates multi-branch embeddings effectively while preserving their unique characteristics. As illustrated in Fig. 2, MEFM consists of L stacked layers, each performing a sequence of operations to fuse and refine features from all branches iteratively. For each layer $l \in [1, 2, ..., L]$, the embeddings from all branches are concatenated along the channel axis to form a unified feature representation.

$$\begin{aligned} T_e^0 &= S_e, \\ T_c^l &= \text{concat}([T_e^l]), \end{aligned} \tag{1}$$

where $T_c^l \in \mathbb{R}^{V \times 4C}$ and e is the index of branches. This operation aggregates complementary information across branches, preparing the features for subsequent fusion.

The concatenated embeddings are passed through a Mamba layer, which plays a crucial role in dynamically learning and balancing frequency-specific weights to achieve complementary integration of features, enhancing the model's adaptability to diverse input characteristics. After fusion, the combined features are split and grouped to ensure that each branch $\hat{T}_i^l$ focuses on its specific representation while maintaining inter-branch relationships.

$$\begin{aligned} \hat{T}_c^l &= \text{Mamba}(T_c^{l-1}), \\ [\hat{T}_1^l, \hat{T}_2^l, \hat{T}_3^l, \hat{T}_4^l] &= \text{Split}(\hat{T}_c^l) \end{aligned} \tag{2}$$

For each branch, its embedding $\hat{T}_i^l \in \mathbb{R}^{l \times C}$ is treated as the query (Q), while the embeddings from the other branches $\hat{T}_j^l = concat(\hat{T}_k^l | k \neq i) \in \mathbb{R}^{l \times 3C}$ serve as the key (K) and value (V). This attention mechanism allows each branch to focus selectively on relevant features from the other branches, enriching its representation through inter-branch interactions. The process is repeated for all branches, ensuring comprehensive attention-based feature refinement.

$$Q_i^l = \hat{T}_i^l W_{Q,i}^l, \quad K_i^l = \hat{T}_j^l W_{K,i}^l, \quad V_i^l = \hat{T}_j^l W_{V,i}^l \tag{3}$$

$$A_i^l = \text{softmax}\left(\frac{Q_i^l {K_i^l}^\top}{\sqrt{d_i^l}}\right) V_i^l, \tag{4}$$

Models	BraTS 2020			BraTS 2023		
	WT	TC	ET	WT	TC	ET
	DSC↑ / HD95↓	DSC↑ / HD95↓	DSC↑ / HD95↓	DSC↑ / HD95↓	DSC↑ / HD95↓	DSC↑ / HD95↓
SegResNet [Myronenko, 2019]	90.78 / 5.99	86.91 / 5.29	**80.64** / 4.53	92.02 / 4.07	89.10 / 4.08	83.66 / 3.88
SegResNet + DVNet	**91.02 / 4.87**	**87.80 / 4.92**	79.51 / **4.34**	**93.82 / 3.64**	**92.19 / 2.81**	**88.32 / 2.69**
UNETR [Hatamizadeh *et al.*, 2022]	89.44 / 7.07	83.91 / 8.40	78.45 / 6.70	92.19 / 6.17	86.39 / 5.29	84.48 / 5.03
UNETR + DVNet	**90.05 / 5.74**	**84.98 / 5.90**	**79.60 / 5.25**	**93.20 / 4.05**	**91.16 / 3.44**	**87.50 / 3.16**
SwinUNETR [Hatamizadeh *et al.*, 2021]	90.43 / 5.65	**86.40** / 7.01	**79.94** / 5.92	92.71 / 5.22	87.79 / 4.42	84.21 / 4.48
SwinUNETR + DVNet	**90.58 / 5.28**	85.72 / **6.99**	79.48 / **5.07**	**93.71 / 4.03**	**92.22 / 2.72**	**88.40 / 2.68**
SegMamba [Xing *et al.*, 2024]	90.90 / 4.87	**87.90** / 5.37	77.61 / 6.12	93.61 / 3.37	92.65 / 3.85	87.71 / 3.48
SegMamba + DVNet	**91.22 / 4.78**	87.48 / **4.52**	**79.56 / 4.94**	**93.99 / 3.22**	**92.68 / 3.54**	**88.87 / 2.85**

Table 1: Performance comparison of DVNet-enhanced U-shaped segmentation models on BraTS 2020 and BraTS 2023. Metrics include DSC (higher is better) and HD95 (lower is better) for WT, TC, and ET.

where $W^l_{Q,i}$, $W^l_{K,i}$, and $W^l_{V,i}$ are learnable weight matrices for each l^{th} layer and i^{th} branch, and d^l_i is the feature dimension for scaling.

Each branch's output from the attention mechanism is further refined using a SE block to recalibrate channel-wise dependencies dynamically and a Multi-Layer Perceptron (MLP) to enhance non-linear feature representation and improve discriminative ability.

$$\tilde{T}^l_i = \text{SE}(\hat{T}^l_i + A^l_i) \tag{5}$$

$$T^l_i = \text{MLP}(\tilde{T}^l_i) + \tilde{T}^l_i \tag{6}$$

This process is repeated for L layers, with the output of each layer serving as input to the next. The iterative refinement allows MEFM to progressively improve feature quality, ensuring robust fusion and representation.

3.3 Spatial-Guided Refinement Module

The SGRM is designed to restore the dimensions of the unified embeddings of each branch to the original input size, allowing seamless integration with the final reconstruction pipeline. As illustrated in Fig. 2, SGRM employs an MLP to project the feature representation of each branch, T^L_e. Then, using a 3D transposed convolution operation, the refined embeddings are reshaped and expanded to their original spatial dimensions. This operation reconstructs the volumetric structure of the input image by reversing the patch embedding and tokenization process.

$$O'_e = \text{MLP}(T^L_e), \tag{7}$$

$$O_e = \text{ConvTranspose3D}(\text{Reshape}(O'_e)) + E_e. \tag{8}$$

where O_e represents the final output of the e^{th} branch.

This module ensures that the spatial integrity of the original input image is preserved while maintaining the enhanced feature representation learned during the earlier stages. By directly coupling embedding refinement with spatial reconstruction, SGRM effectively bridges the gap between the unified branch embeddings and the final output.

4 Experiments

4.1 Datasets

This study conducted an in-depth analysis using multiple datasets across diverse domains to evaluate DVNet's robustness and adaptability. These datasets include both 2D and 3D imaging modalities, ensuring a comprehensive assessment of our model's performance.

The BraTS 2020 and BraTS 2023 datasets [Menze *et al.*, 2014; Kazerooni *et al.*, 2024; Henry *et al.*, 2021] are widely recognized benchmarks for brain tumor segmentation tasks. These datasets provide 3D brain MRI volumes with precise ground truth annotations, including four modalities for each volume: T1-weighted, T1-contrast enhanced, T2-weighted, and FLAIR. Ground truth annotations include three segmentation targets: Whole Tumor (WT), Enhancing Tumor (ET), and Tumor Core (TC). The BraTS 2020 & 2023 datasets provide foundational and diversity benchmarks, challenging the model's adaptability to varying imaging conditions and glioma types. For a fair comparison, we follow the evaluation protocol outlined in SegMamba [Xing *et al.*, 2024] and utilize public implementations of the baseline methods in comparison to retrain their networks under the same settings for BraTS 2020 & 2023 datasets, respectively. The data is split randomly: 70% of the 3D volumes for training, 10% for validation, and the remaining 20% for testing.

In addition to 3D segmentation, DVNet was also evaluated on the Abdomen MRI dataset [Ji *et al.*, 2022] and Microscopy Cell Dataset [Ma *et al.*, 2024b] for 2D segmentation tasks. The Abdomen MRI dataset was used for the MICCAI 2022 AMOS Challenge [Ji *et al.*, 2022], designed for segmentation of abdominal organs. Microscopy Cell dataset was used for the NeurIPS 2022 Cell Segmentation Challenge [Ma *et al.*, 2024b]. In accordance with the settings of xLSTM-Unet [Chen *et al.*, 2024b], we employed 60 labeled MRI scans for training and 50 scans for testing in the Abdomen MRI dataset and 1,000 images for training and 101 images for evaluation in the Microscopy Cell dataset.

4.2 Implementation Details

DVNet is implemented based on SegMamba [Xing *et al.*, 2024] for 3D tasks and xLSTM [Chen *et al.*, 2024b] for 2D tasks. For 3D tasks, the input crop size is set to $(64 \times 64 \times 64)$ with a batch size of 2. The optimization process uses cross-entropy loss with a stochastic gradient descent optimizer and a polynomial learning rate scheduler (initial learning rate of 0.01 and decay of 1×10^{-5}). Training runs for 1,000 epochs with data augmentations including brightness, gamma, rotation, scaling, mirror, and elastic deformation. For 2D tasks, the loss function is the sum of Dice loss and cross-entropy loss, optimized with the AdamW optimizer (weight decay:

Models	WT		TC		ET		Avg	
	DSC ↑	HD95 ↓	DSC ↑	HD95 ↓	DSC ↑	HD95 ↓	DSC ↑	HD95 ↓
SegResNet [Myronenko, 2019]	92.02	4.07	89.10	4.08	83.66	3.88	88.26	4.01
UX-Net [Lee *et al.*, 2022]	93.13	4.56	90.03	5.68	85.91	4.19	89.69	4.81
MedNeXt [Roy *et al.*, 2023]	92.41	4.98	87.75	4.67	83.96	4.51	88.04	4.72
UNETR [Hatamizadeh *et al.*, 2022]	92.19	6.17	86.39	5.29	84.48	5.03	87.68	5.49
SwinUNETR [Hatamizadeh *et al.*, 2021]	92.71	5.22	87.79	4.42	84.21	4.48	88.23	4.70
SwinUNETR-V2 [He *et al.*, 2023]	93.35	5.01	89.65	4.41	85.17	4.41	89.39	4.51
SegMamba [Xing *et al.*, 2024]	93.61	3.37	92.65	3.85	87.71	3.48	91.32	3.56
DVNet + SegMamba	**93.99**	**3.22**	**92.68**	**3.54**	**88.87**	**2.85**	**91.85**	**3.20**

Table 2: Comparison of 3D segmentation performance across different SOTA models on BraTS 2023. Metrics include DSC (higher is better) and HD95 (lower is better) for WT, TC, ET, and the average.

0.05). Learning rates are set empirically for each dataset: Abdomen MRI (0.005) and Microscopy (0.0015). Batch size is set to 30 for Abdomen MRI and 12 for Microscopy. All models are trained for 1,000 epochs for 2D tasks.

4.3 Evaluation Metrics

The Dice Similarity Coefficient (DSC) and 95% Hausdorff Distance (HD95) are adopted for quantitative comparison on the BraTS 2020 & 2023 datasets. The DSC and Normalized Surface Distance (NSD) are used as evaluation metrics for Abdomen MRI dataset. The F1 score is used for cell segmentation on the Microscopy dataset.

4.4 Experimental Results

To evaluate the effectiveness of DVNet, we conducted comparisons in different scenarios with appropriate baselines.

Plug-and-play: To demonstrate the effectiveness of DVNet as a plug-and-play module, DVNet was plugged into four state-of-the-art (SOTA) models (SegResNet [Myronenko, 2019], UNETR [Hatamizadeh *et al.*, 2022], SwinUNETR [Hatamizadeh *et al.*, 2021] and SegMamba [Xing *et al.*, 2024]). Their performance with and without DVNet was compared on the BraTS 2020 and BraTS 2023 datasets. As shown in Table 1, DVNet consistently improves segmentation accuracy (DSC) and boundary precision (HD95) over baseline models by improving feature extraction and addressing challenges such as blurry and low-contrast regions. For example, on BraTS 2023 dataset, DVNet improves the DSC score of SegResNet [Myronenko, 2019] from 92.02% to 93.82% for WT and reduces its HD95 from 4.07 to 3.64, demonstrating its ability to refine predictions and improve boundary delineation. Similarly, when integrated with transformer-based models like SwinUNETR [Hatamizadeh *et al.*, 2021], DVNet achieves notable gains, with DSC of TC increasing from 87.79% to 92.22% and HD95 decreasing from 4.42 to 2.72. These findings validate the robustness and adaptability of DVNet, proving its capability to deliver SOTA performance across diverse datasets and clinical scenarios.

Comparison with SOTA methods for the 3D segmentation task: We also evaluated the performance of DVNet against other SOTA models reported in SegMamba [Xing *et al.*, 2024] on the BraTS 2023 dataset. The results in Table 2 demonstrate the superiority of DVNet over all baseline models for brain tumor segmentation as the DVNet achieves the highest DSC scores and the lowest HD95 values among all methods compared, especially excelling in challenging subregions like ET and TC. Compared to the strongest baseline, SegMamba [Xing *et al.*, 2024], DVNet further enhances segmentation accuracy and improves boundary delineation, particularly for the challenging ET subregion.

Models	Abdomen MRI		Microscopy
	DSC ↑	NSD ↑	F1 ↑
nnU-Net [Isensee *et al.*, 2021]	74.50±11.17	81.53±11.45	53.83±26.57
SegResNet [Myronenko, 2019]	73.17±13.79	80.34±13.86	54.11±26.33
UNETR [Hatamizadeh *et al.*, 2022]	57.47±16.72	63.09±18.58	43.57±25.72
SwinUNETR [Hatamizadeh *et al.*, 2021]	70.28±13.48	76.69±14.42	39.67±26.21
U-Mamba_Bot [Ma *et al.*, 2024a]	75.88±10.51	82.85±10.74	53.89±28.17
U-Mamba_Enc [Ma *et al.*, 2024a]	76.25±10.82	83.27±10.87	56.07±27.84
xLSTM_bot [Chen *et al.*, 2024b]	76.36±10.06	83.22±10.34	58.18±23.86
xLSTM_enc [Chen *et al.*, 2024b]	77.47±9.50	83.74±9.51	60.36±24.35
DVNet + xLSTM_enc	**77.72±15.05**	**84.23±9.22**	**65.54±26.15**

Table 3: Comparison of segmentation performance on 2D tasks.

Models	WT	TC	ET
	DSC↑ / HD95↓	DSC↑ / HD95↓	DSC↑ / HD95↓
DVNet (w/o WDPM)	93.12 / 3.88	91.24 / 3.03	84.75 / 3.38
DVNet (w/o MEFM)	93.13 / 3.87	91.07 / 3.00	84.68 / 3.18
DVNet	**93.82 / 3.64**	**92.19 / 2.81**	**88.32 / 2.69**

Table 4: An ablation study to evaluate impact of WDPM and MEFM on segmentation performance in DVNet on BraTS 2023.

Comparison with SOTA methods for 2D segmentation tasks: Table 3 compares the segmentation performance of DVNet with SOTA models for 2D medical images. On the Abdomen MRI dataset, the DVNet integrated with the xLSTM_enc architecture [Chen *et al.*, 2024b], slightly exceeds xLSTM_enc and achieves the highest performance, with a DSC of 77.58% and an NSD of 83.84%, demonstrating the effectiveness of our plugin-based approach in enhancing segmentation accuracy and boundary precision. For the Microscopy dataset, DVNet again outperforms other SOTA models, achieving an F1 score of 65.54%, a more impressive improvement over xLSTM_enc (60.36%). This demonstrates the robustness of DVNet in handling complex instance segmentation tasks.

In summary, experimental results have demonstrated the effectiveness of DVNet as a plug-and-play module, consistently enhancing segmentation accuracy and boundary preci-

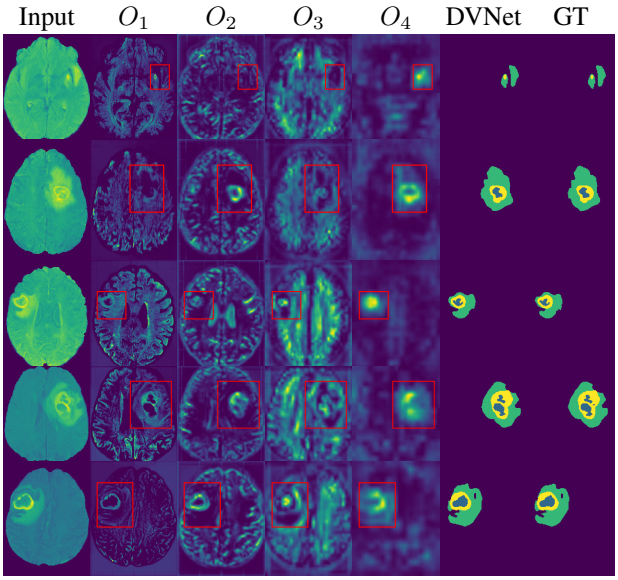

Figure 5: Visualization of the segmentation results of the proposed DVNet. Each row corresponds to a single case, and the columns represent (from left to right): the input image, the outputs of the four branches (O_1 to O_4), the final prediction by DVNet, and the ground truth segmentation mask. The activation maps of O_1 to O_4 demonstrate a trend that as DVNet is more embedded into the baseline model, the segmentation model more focuses on the lesion area and weakens or suppresses the surrounding normal tissue area.

Models	WT	TC	ET
	DSC↑ / HD95↓	DSC↑ / HD95↓	DSC↑ / HD95↓
DVNet (w/ LLL)	93.52 / 3.84	91.87 / 3.72	87.84 / 3.32
DVNet (w/ HHH)	93.31 / 4.56	91.77 / 4.64	87.90 / 3.97
DVNet	**93.82 / 3.64**	**92.19 / 2.81**	**88.32 / 2.69**

Table 5: An ablation study to evaluate effect of different frequency components in WDPM on segmentation performance.

sion for both 3D and 2D tasks.

4.5 Ablation Study

Ablation studies have been conducted to demonstrate the effectiveness of different modules in DVNet. Table 4 compares the contributions of the WDPM and MEFM in the proposed DVNet (plug-in SegResNet [Myronenko, 2019]) on the BraTS 2023 dataset.

Removal of WDPM or MEFM results in notable performance degradation with a decrease in DSC score and an increase in HD95, particularly for the more challenging TC and ET subregions. The results indicate reduced segmentation precision and boundary accuracy, underscoring the critical roles of both WDPM and MEFM in boundary refinement and multi-region interaction. In contrast, the full DVNet achieves the best overall performance.

To further explore the contributions of different frequency signals in medical image segmentation, an ablation study was conducted to study the effect of different SWT components. The results shown in Table 5 demonstrate that neither low-frequency nor high-frequency components alone are sufficient for optimal segmentation performance. Low-frequency components (LLL) excel in capturing global structural information but lack the details required for precise boundary delineation. In contrast, high-frequency components (HHH) focus on local details but fail to provide the broader context necessary to segment complex structures and are also susceptible to noise.

Overall, the ablation study highlights the importance of both WDPM and MEFM in enhancing the segmentation performance of DVNet. WDPM strengthens feature representation and target structure recognition; while MEFM improves boundary precision and inter-region consistency. In addition, the full SWT-based DVNet effectively integrates both low-frequency and high-frequency information, leveraging their complementary strengths to achieve superior performance. This highlights the importance of multi-frequency analysis in medical image segmentation tasks, demonstrating DVNet's efficacy in tackling complex medical image segmentation tasks.

4.6 Visualization

To explore the effectiveness of DVNet, we visualize the outputs of DVNet. We randomly selected some images and performed visualization in the inference mode on the BraTS 2023 test dataset. Fig. 5 illustrates the intermediate and final segmentation results of the proposed DVNet for some representative cases. Each row corresponds to a single case, and the columns represent (from left to right): the input image, the outputs of the four branches (O_1 to O_4), the final prediction by DVNet, and the ground truth (GT) segmentation mask. The branch outputs highlight the model's capability to capture features at multiple levels, providing complementary information for the final prediction. The final DVNet predictions demonstrate strong alignment with the ground truth, effectively capturing the boundaries and internal details of the target regions. However, subtle discrepancies in tiny structures indicate potential areas for further improvement. Overall, the visualization results validate the robustness and precision of the proposed multi-branch segmentation approach in addressing complex segmentation tasks.

5 Conclusion

In this work, we are motivated to incorporate the principles of scotopic vision into noninvasive medical image processing and analysis tasks, with the aim of inspiring future research endeavors. Specifically, we introduced DVNet, a novel plug-and-play dark vision enhancement method that teaches models to fully capture and learn key visual features in blurry medical images. By utilizing the multi-scale frequency information obtained from wavelet transform, DVNet can identify lesions in low-contrast areas, using them as attention indices for dense prediction tasks. Our approach refines these indices, constructing global dependencies of visual features in the manner of potential conduction in the optic nerve, reactivating blurred areas and thereby achieving precise segmentation of lesion regions. Extensive experiments demonstrate that DVNet achieves SOTA performance in various medical image segmentation tasks, highlighting its effectiveness and generalizability.

Acknowledgments

This work is supported by grants from the National Natural Science Foundation of China (NSFC 62322215, 62402488), National Science Foundation (MRI-2018966) and South Carolina Research Authority (2331-241-2024811). This study was supported in part by the high-performance computing center of Central South University.

References

[Aglinskas *et al.*, 2022] Aidas Aglinskas, Joshua K. Hartshorne, and Stefano Anzellotti. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. *Science*, 376(6597):1070–1074, 2022.

[Ankri *et al.*, 2020] Lea Ankri, Elishai Ezra-Tsur, Shir R Maimon, Nathali Kaushansky, and Michal Rivlin-Etzion. Antagonistic center-surround mechanisms for direction selectivity in the retina. *Cell reports*, 31(5), 2020.

[Chen *et al.*, 2012] Xuehua Chen, Wei Yang, Zhenhua He, and Wenli Zhong. Adaptive acquisition footprint suppression based on a 3d stationary wavelet transform: A case study from china. *Journal of Applied Geophysics*, 77:1–6, 2012.

[Chen *et al.*, 2023] Linwei Chen, Ying Fu, Kaixuan Wei, Dezhi Zheng, and Felix Heide. Instance segmentation in the dark. *International Journal of Computer Vision*, 131(8):2198–2218, 2023.

[Chen *et al.*, 2024a] Jieneng Chen, Jieru Mei, Xianhang Li, Yongyi Lu, Qihang Yu, Qingyue Wei, Xiangde Luo, Yutong Xie, Ehsan Adeli, Yan Wang, Matthew P. Lungren, Shaoting Zhang, Lei Xing, Le Lu, Alan Yuille, and Yuyin Zhou. Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. *Medical Image Analysis*, 97:103280, 2024.

[Chen *et al.*, 2024b] Tianrun Chen, Chaotao Ding, Lanyun Zhu, Tao Xu, Deyi Ji, Yan Wang, Ying Zang, and Zejian Li. xlstm-unet can be an effective 2d & 3d medical image segmentation backbone with vision-lstm (vil) better than its mamba counterpart. *arXiv preprint arXiv:2407.01530*, 2024.

[Dosovitskiy *et al.*, 2021] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

[Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[Gu and Dao, 2024] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces, 2024.

[Hatamizadeh *et al.*, 2021] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R Roth, and Daguang Xu. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *International MICCAI brainlesion workshop*, pages 272–284. Springer, 2021.

[Hatamizadeh *et al.*, 2022] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R Roth, and Daguang Xu. Unetr: Transformers for 3d medical image segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 574–584, 2022.

[He *et al.*, 2023] Yufan He, Vishwesh Nath, Dong Yang, Yucheng Tang, Andriy Myronenko, and Daguang Xu. Swinunetr-v2: Stronger swin transformers with stagewise convolutions for 3d medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 416–426. Springer, 2023.

[Henry *et al.*, 2021] Theophraste Henry, Alexandre Carré, Marvin Lerousseau, Théo Estienne, Charlotte Robert, Nikos Paragios, and Eric Deutsch. Brain tumor segmentation with self-ensembled, deeply-supervised 3d u-net neural networks: a brats 2020 challenge solution. In *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 6*, pages 327–339. Springer, 2021.

[Hofmann and Lamb, 2023] Klaus Peter Hofmann and Trevor D. Lamb. Rhodopsin, light-sensor of vision. *Progress in Retinal and Eye Research*, 93:101116, 2023.

[Hu *et al.*, 2018] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[Isensee *et al.*, 2021] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

[Ji *et al.*, 2022] Yuanfeng Ji, Haotian Bai, Chongjian Ge, Jie Yang, Ye Zhu, Ruimao Zhang, Zhen Li, Lingyan Zhanng, Wanling Ma, Xiang Wan, et al. Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. *Advances in neural information processing systems*, 35:36722–36732, 2022.

[Jin *et al.*, 2023] Xin Jin, Ling-Hao Han, Zhen Li, Chun-Le Guo, Zhi Chai, and Chongyi Li. Dnf: Decouple and feedback network for seeing in the dark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18135–18144, June 2023.

[Kazerooni *et al.*, 2024] Anahita Fathi Kazerooni, Nastaran Khalili, Xinyang Liu, Debanjan Haldar, Zhifan Jiang, Syed Muhammed Anwar, Jake Albrecht, Maruf Adewole,

Udunna Anazodo, Hannah Anderson, et al. The brain tumor segmentation (brats) challenge 2023: focus on pediatrics (cbtn-connect-dipgr-asnr-miccai brats-peds). *ArXiv*, pages arXiv–2305, 2024.

[Lee *et al.*, 2022] Ho Hin Lee, Shunxing Bao, Yuankai Huo, and Bennett A Landman. 3d ux-net: A large kernel volumetric convnet modernizing hierarchical transformer for medical image segmentation. *arXiv preprint arXiv:2209.15076*, 2022.

[Lian and Wang, 2023] Zuozheng Lian and Haizhen Wang. An image deblurring method using improved u-net model based on multilayer fusion and attention mechanism. *Scientific Reports*, 13(1):21402, 2023.

[Ma *et al.*, 2024a] Jun Ma, Feifei Li, and Bo Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024.

[Ma *et al.*, 2024b] Jun Ma, Ronald Xie, Shamini Ayyadhury, Cheng Ge, Anubha Gupta, Ritu Gupta, Song Gu, Yao Zhang, Gihun Lee, Joonkee Kim, et al. The multimodality cell segmentation challenge: toward universal solutions. *Nature methods*, pages 1–11, 2024.

[Ma *et al.*, 2024c] Shiqiang Ma, Xuejian Li, Jijun Tang, and Fei Guo. Aggregate-aware model with bidirectional edge generation for medical image segmentation. *Applied Soft Computing*, 163:111918, 2024.

[McKyton *et al.*, 2024] Ayelet McKyton, Deena Elul, and Netta Levin. Seeing in the dark: High-order visual functions under scotopic conditions. *Iscience*, 27(2), 2024.

[Menze *et al.*, 2014] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.

[Myronenko, 2019] Andriy Myronenko. 3d mri brain tumor segmentation using autoencoder regularization. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4*, pages 311–320. Springer, 2019.

[Qiao *et al.*, 2022] Shuang Qiao, Qinghan Yu, Zhengwei Zhao, Liying Song, Hui Tao, Tian Zhang, and Chenyi Zhao. Edge extraction method for medical images based on improved local binary pattern combined with edge-aware filtering. *Biomedical Signal Processing and Control*, 74:103490, 2022.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[Roy *et al.*, 2023] Saikat Roy, Gregor Koehler, Constantin Ulrich, Michael Baumgartner, Jens Petersen, Fabian Isensee, Paul F Jaeger, and Klaus H Maier-Hein. Mednext: transformer-driven scaling of convnets for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 405–415. Springer, 2023.

[Tian *et al.*, 2023] Chunwei Tian, Menghua Zheng, Wangmeng Zuo, Bob Zhang, Yanning Zhang, and David Zhang. Multi-stage image denoising with the wavelet transform. *Pattern Recognition*, 134:109050, 2023.

[Wang *et al.*, 2022a] Hai Wang, Yanyan Chen, Yingfeng Cai, Long Chen, Yicheng Li, Miguel Angel Sotelo, and Zhixiong Li. Sfnet-n: An improved sfnet algorithm for semantic segmentation of low-light autonomous driving road scenes. *IEEE Transactions on Intelligent Transportation Systems*, 23(11):21405–21417, 2022.

[Wang *et al.*, 2022b] Kun Wang, Xiaohong Zhang, Xiangbo Zhang, Yuting Lu, Sheng Huang, and Dan Yang. Eanet: Iterative edge attention network for medical image segmentation. *Pattern Recognition*, 127:108636, 2022.

[Wang *et al.*, 2024a] Haonan Wang, Peng Cao, Jinzhu Yang, and Osmar Zaiane. Narrowing the semantic gaps in u-net with learnable skip connections: The case of medical image segmentation. *Neural Networks*, 178:106546, 2024.

[Wang *et al.*, 2024b] Xinzhe Wang, Kang Ma, Qiankun Liu, Yunhao Zou, and Ying Fu. Multi-object tracking in the dark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 382–392, June 2024.

[Wang *et al.*, 2024c] Yifan Wang, Jie Gui, Yuan Yan Tang, and James T Kwok. Cfvnet: An end-to-end cancelable finger vein network for recognition. *IEEE Transactions on Information Forensics and Security*, 2024.

[Wang *et al.*, 2024d] Ziyang Wang, Jian-Qing Zheng, Yichi Zhang, Ge Cui, and Lei Li. Mamba-unet: Unet-like pure visual mamba for medical image segmentation, 2024.

[Xing *et al.*, 2024] Zhaohu Xing, Tian Ye, Yijun Yang, Guang Liu, and Lei Zhu. Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 578–588. Springer, 2024.

[Yang and Yu, 2021] Ruixin Yang and Yingyan Yu. Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis. *Frontiers in oncology*, 11:638182, 2021.