WaveDE: Wavelet-Routed Dual-Expert Network for Fundus Photograph Enhancement

Haitao Nie[1*], Dongjie Wu[2*], and Bin Xie[1✉]

[1] School of Automation, Central South University, Changsha, Hunan, China
xiebin@csu.edu.cn
[2] School of Computer Science and Engineering, Central South University, Changsha, Hunan, China

Abstract. Fundus photography is essential for retinal disease screening and diagnosis, yet low-quality images are common in clinical practice, undermining both clinical reading and automated analysis. Fundus image enhancement faces two key challenges: (i) real degradations are often compound, causing enhancement models trained under single-degradation settings to generalize poorly; and (ii) restoring low- and high-frequency components involves conflicting objectives, making it difficult to jointly achieve global consistency, fine-detail recovery, and noise suppression. We propose WaveDE, a wavelet-routed dual-expert network for fundus photograph enhancement. Discrete wavelet transform-based band routing decomposes decoder features into low- and high-frequency subbands processed by two specialized experts: the low-frequency expert performs global correction and provides structural priors, while the high-frequency expert leverages them to enhance textures and suppress noise. A high-frequency-injected VAE decoder further recovers fine details lost to latent compression, and an anatomical consistency regularization enforces structural fidelity of vessels and the optic disc. On multiple public datasets, WaveDE achieves the best PSNR (39.02, 95% CI: 38.90–39.16) and SSIM (0.98, 95% CI: 0.97–0.98) among 12 state-of-the-art methods, with consistent gains across 2 classification and 4 segmentation datasets; out-of-distribution evaluations further confirm cross-domain robustness. The code and model are available at https://github.com/CowboyH/WaveDE.

Keywords: Fundus photograph enhancement · Compound degradations · High—low frequency decoupling · Dual-expert network · Wavelet-based frequency routing.

1 Introduction

Color fundus photography is a key modality for screening and diagnosing ophthalmic diseases. However, real-world acquisition frequently introduces degradations such as uneven illumination, blur, and noise, compromising both clinical

* These authors contributed equally to this work.
✉ Corresponding author: xiebin@csu.edu.cn

readability and the performance of AI-based analysis. Developing effective fundus image quality enhancement methods is therefore essential.

Fundus image quality enhancement has attracted increasing attention in recent years [23,4]. Most existing methods focus on a single degradation and develop specialized models, such as illumination correction [14], artifact suppression [15], deblurring [1] and dehazing [28]. More recently, studies have started to address compound degradations that commonly occur in clinical acquisition [5,17], e.g., underexposure accompanied by noise, or glare co-existing with blur. To better recover fine structures, frequency-aware strategies have been introduced to strengthen high-frequency representations [16] and enforce multi-scale constraints [18]. In addition, anatomical localization priors, such as vessel and optic disc cues, are used to constrain the enhancement process and reduce structural distortion [30,4,15].

However, fundus image quality enhancement still faces two key challenges: (1) **Degradation-specific settings.** Most methods are tailored to a single degradation, while low-quality fundus images often exhibit mixed degradations (e.g., underexposure with noise), causing performance to drop under varying combinations and severities. (2) **Balancing global correction and detail recovery.** Enhancement requires both global appearance correction and fine-detail restoration, but these objectives can conflict during training; without explicit disentangled modeling, many methods converge to a suboptimal trade-off and underperform on both aspects.

To address the above challenges, we propose WaveDE, a wavelet-routed dual-expert network for fundus image enhancement. Our core insight is twofold: (i) different degradations predominantly reside in distinct frequency bands (e.g., global corruptions in low-frequency and local artifacts in high-frequency subbands), so discrete wavelet transform-based decomposition naturally disentangles compound degradations into dedicated expert streams; and (ii) explicit frequency decoupling also alleviates the inherent optimization conflict between global correction and fine-detail recovery. Specifically, our contributions are threefold:

1. A **frequency-decoupled dual-expert module** in the diffusion U-Net decoder, where a low-frequency expert performs global correction and provides structural priors, and a high-frequency expert leverages them for texture enhancement and noise suppression.
2. An **enhanced VAE decoder** with multi-scale high-frequency subband injection to compensate for detail loss from latent-space compression.
3. An **Anatomical Consistency Regularization** using frozen anatomical teacher networks to preserve the topological integrity of vessels and the optic disc, reducing structural hallucinations.

2 Methodology

Given paired low/high-quality fundus images $(\mathbf{x}, \mathbf{y})$, we learn G_θ such that $\hat{\mathbf{y}} = G_\theta(\mathbf{x}) \approx \mathbf{y}$, while minimizing a reconstruction loss between $\hat{\mathbf{y}}$ and $\mathbf{y}$. Unlike iterative diffusion models, we treat image degradations as implicit noise

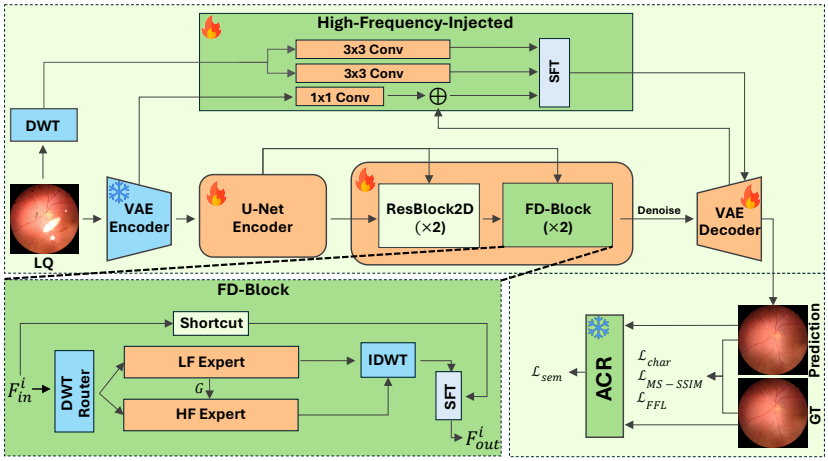

Fig. 1. Overview of WaveDE. Given a low-quality fundus image $\mathbf{x}$, the diffusion U-Net decoder incorporates FD-Blocks for frequency-decoupled restoration, while the High-Frequency-injected VAE decoder recovers fine details lost to latent compression. ACR enforces anatomical consistency of vessels and the optic disc.

and perform a single deterministic denoising step. The core components include a High-Frequency-Injected (HFI) Enhanced VAE Decoder, a Frequency-Decoupled Dual-Expert Block (FD-Block), and Anatomical Consistency Regularization (ACR) (Figure 1).

2.1 High-Frequency-Injected Enhanced VAE Decoder

VAEs tend to lose high-frequency details (*e.g.*, vessels, optic disc) during downsampling. To address this, we inject high-frequency priors extracted from the input $\mathbf{x}$ via multi-level 2D discrete wavelet transform (DWT) into the decoder. At decoder stage i, the DWT subbands $\mathbf{HF}_x^{(l)}$ whose spatial resolution matches the current feature map are injected alongside encoder skip features:

$$\begin{aligned} \tilde{\mathbf{h}}_i &= \mathbf{h}_i + \gamma_{\text{skip}} \cdot \mathcal{H}_i(\mathbf{a}_i), \\ \mathbf{h}'_i &= \tilde{\mathbf{h}}_i \odot \Big(1 + \tanh\big(\mathcal{C}_\gamma(\mathbf{HF}_x^{(l)})\big)\Big) + \tanh\big(\mathcal{C}_\beta(\mathbf{HF}_x^{(l)})\big), \end{aligned} \tag{1}$$

where $\mathbf{a}_i$ denotes the frozen VAE encoder feature aligned via $\mathcal{H}_i(\cdot)$, and $\mathcal{C}_\gamma$, $\mathcal{C}_\beta$ are independent Conv–SiLU–Conv branches producing scale and shift maps via Spatial Feature Transform (SFT) [26]. Zero-initialized final layers ensure near-identity initialization. During training, the VAE encoder is frozen; only the decoder is fine-tuned.

2.2 Frequency-Decoupled Dual-Expert Block

To decouple composite degradations and resolve the global-local optimization conflict, we introduce FD-Block in the Stable Diffusion v1.5 (SD) [21] U-Net de-

coder, replacing the original ResBlock2D at resolutions $\geq 16\times16$ while preserving the pretrained prior at lower resolutions. Given input $\mathbf{F}_{in} \in \mathbb{R}^{B\times C\times H\times W}$, DWT (Biorthogonal 1.3) decomposes it into a low-frequency component $\mathbf{F}_{LL}$ and concatenated high-frequency subbands $\mathbf{F}_{HF} = \text{Concat}(\mathbf{F}_{HF}^{LH}, \mathbf{F}_{HF}^{HL}, \mathbf{F}_{HF}^{HH})$, which are processed by dedicated experts. The low-frequency (LF) expert outputs $\mathbf{Y}_{LF}$ and produces a structural guidance map $\mathbf{G}$ to constrain the high-frequency (HF) expert, enabling deterministic band routing without learned gating. The frequency-band residuals $(\mathbf{Y}_{LF} - \mathbf{F}_{LL}, \mathbf{Y}_{HF} - \mathbf{F}_{HF})$ are reconstructed via IDWT and added to the input via a layer-scale shortcut: $\mathbf{F}_{out} = \mathbf{F}_{in} + \gamma \cdot \text{IDWT}(\mathbf{Y}_{LF} - \mathbf{F}_{LL}, \mathbf{Y}_{HF} - \mathbf{F}_{HF})$.

Low-Frequency Expert Vision Mamba (LF-VMamba) LF-VMamba captures long-range context for stable global correction via a 3×3 depthwise convolution , stacked vision state space (VSS) blocks [19], and a Global Illumination Calibration (GIC) module. Each VSS block applies SS2D followed by a gated local FFN with residual connections at both stages. GIC regresses affine parameters $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ via global average pooling for brightness and contrast correction, and derives a structural guidance map from the SS2D gating signal $\mathbf{g}$:

$$\mathbf{X}_{out} = \mathbf{X}_{in} \cdot (1 + \boldsymbol{\gamma}) + \boldsymbol{\beta}, \quad \mathbf{G} = \sigma\big(\text{RMSNorm}(\text{Conv}_{1\times1}(\mathbf{g}))\big). \tag{2}$$

The LF expert outputs a GroupNorm-stabilized residual update: $\mathbf{Y}_{LF} = \mathbf{F}_{LL} + \text{GroupNorm}(\Delta_{LF})$.

Orientation-Selective High-Frequency Expert (OS-HFNet) OS-HFNet recovers fine details while suppressing noise through three successive steps. First, an SE block [7] recalibrates the contribution of LH, HL, and HH subbands:

$$\mathbf{F}_{SE} = \mathbf{F}_{HF} \odot \sigma(\text{MLP}(\text{GAP}(\mathbf{F}_{HF}))) \,. \tag{3}$$

Next, SFT modulation with guidance map $\mathbf{G}$ focuses enhancement on anatomical regions:

$$\mathbf{F}_{mod} = \mathbf{F}_{SE} \odot \Big(1 + \tanh(\mathcal{C}_\gamma(\mathbf{G}))\Big) + \tanh(\mathcal{C}_\beta(\mathbf{G})), \tag{4}$$

where $\mathcal{C}_\gamma$, $\mathcal{C}_\beta$ are independent Conv–SiLU–Conv branches. Then, asymmetric convolutions (1×3, 3×1) extract horizontal and vertical responses $\mathbf{F}_{hor}$, $\mathbf{F}_{ver}$, fused via a selective kernel:

$$\mathbf{F}_{sel} = w_1\mathbf{F}_{hor} + w_2\mathbf{F}_{ver}, \quad (w_1, w_2) = \text{Softmax}\big(\text{MLP}(\text{GAP}(\mathbf{F}_{hor}+\mathbf{F}_{ver}))\big). \tag{5}$$

Finally, gated depthwise convolution performs soft-threshold denoising:

$$\mathbf{Y}_{HF} = \mathbf{F}_{HF} + \text{GroupNorm}\Big(\text{Conv}_{1\times1}(\mathbf{F}_{sel}) \odot \sigma(\text{DWConv}_{3\times3}(\mathbf{F}_{sel}))\Big). \tag{6}$$

2.3 Anatomical Consistency Regularization and Training Loss

To suppress hallucinations and preserve anatomical structures, we introduce a semantic consistency loss $\mathcal{L}_{\text{sem}}$ using two frozen teacher networks: a vessel segmentation teacher T_v (U-Net++ [29]) and an optic-disc detection teacher T_{od} (PraNet [3]), both pretrained on annotated data and removed at inference. We align the enhanced output $\hat{\mathbf{y}}$ with the reference $\mathbf{y}$ in logit space under spatially adaptive weighting to compensate for class-area imbalance:

$$\mathbf{W}_v = 1 + 10 \cdot \sigma(T_v(\mathbf{y})), \quad \mathbf{W}_{od} = 1 + 5 \cdot \sigma(T_{od}(\mathbf{y})), \tag{7}$$

$$\mathcal{L}_{\text{sem}} = \sum_{k \in \{v, od\}} \lambda_k \cdot \mathbb{E}\left[\frac{\mathbf{W}_k}{\mathbb{E}[\mathbf{W}_k] + \epsilon} \odot |T_k(\hat{\mathbf{y}}) - T_k(\mathbf{y})| \right]. \tag{8}$$

The overall training objective combines pixel-level, perceptual, and structural terms:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{char}} + 0.25\, \mathcal{L}_{MS\text{-}SSIM} + 0.1\, \mathcal{L}_{\text{FFL}} + \mathcal{L}_{\text{sem}}, \tag{9}$$

where $\mathcal{L}_{\text{char}}$, $\mathcal{L}_{MS\text{-}SSIM}$, and $\mathcal{L}_{\text{FFL}}$ denote the Charbonnier, multi-scale SSIM, and Focal Frequency losses [10], respectively.

3 Experiments

3.1 Dataset and Implementation Details

We collected 29,704 high-quality (HQ) fundus images from 16 public datasets. Eight datasets are summarized in [13]: DeepDRiD (898), RFMiD (3,091), Messidor (1,712), ODIR (5,945), APTOS (3,376), SUSTech-SYSU (1,219), IDRiD (450), and PAPILA (488). Four datasets are summarized in [24]: FIVES (400), G1020 (1,003), Chákṣu (1,008) and ARIA(63). The remaining datasets include ROP (5,953) [25], SMDG (2,720) [11] , DRD (1,270) [2] and DRIMDB (108) [22]. PAPILA and ARIA are used for out-of-distribution (OOD) evaluation. Here, the numbers in parentheses indicate the sample size.

Inspired by [23], we synthesize 15 degradations (5 single + 10 pairwise compound) to simulate real-world LQ fundus imaging: illumination and color shift (brightness, contrast, and saturation), halo (colorized Gaussian-smoothed ring), dark hole (blurred dark patch), spot (irregular blobs), and blur (Gaussian filtering), applied globally or locally. Generating two LQ variants per in-distribution HQ image yields 58,306 paired samples.

The 58,306 paired HQ–LQ images are resized to 256×256 and split into train/val/test sets (46640/2898/8768). All splits are strictly disjoint: no image overlaps between any training stage (enhancement, ACR teachers, and downstream evaluators) and any test split. The optic-disc teacher is trained on the Chákṣu/SMDG enhancement-training split; the vessel teacher on DRIVE, LES-AV, TREND and DRHAGIS, not used in any other stage. We implement our method on an NVIDIA RTX PRO 6000 (96 GB) and train with AdamW (30,000 steps, batch size 32, cosine annealing with 1,000 warmup steps). The learning rate

is 5×10^{-5} for non-FD-Block modules and 2×10^{-4} for FD-Blocks. At inference, WaveDE runs a single forward pass without the teachers (10 GB, 100 ms/image on a 24 GB RTX 4090). Baselines follow official settings with early stopping. Metrics include PSNR and SSIM for restoration, accuracy and F1-Score for classification, and Dice for segmentation. Baselines span diffusion- [27,31,21,6], GAN- [17,5,30,9], Transformer- [20,8,12], and CNN-based [16] methods; MDA, OTE and GFE are SOTA fundus-enhancement methods. Degradation subsets for image quality assessment are defined as follows: Blur (blur, blur+color, blur+halo, blur+hole, blur+spot), Color (color, blur+color, color+halo, color+hole, color+spot), and Local artifact (halo, hole, spot, blur+halo, blur+hole, blur+spot, color+halo, color+hole, color+spot, halo+hole, halo+spot, hole+spot). Downstream models (ConvNeXt-Tiny for classification, U-Net++ for vessel and PraNet for OD segmentation) are trained once on HQ images, then frozen and applied to each method's enhanced outputs without retraining.

Table 1. Image quality assessment and ablation (PSNR (dB)/SSIM (%)). Top-3 in **bold**, blue, underline; italics: 95% CI. Samples: Total (8,768), Blur (2,792), Color (2,910), Local artifact (7,048). Ablation baseline: Stable Diffusion (SD); ‡: p<0.001

	Total		**Blur**		**Color**		**Local artifact**	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Res [27]	33.08‡ *32.99-33.16*	90‡ *89-91*	33.38‡ *33.22-33.51*	89‡ *89-90*	30.14‡ *29.97-30.30*	90‡ *89-90*	33.43‡ *33.33-33.52*	90‡ *90-91*
Power [31]	20.60‡ *20.45-20.74*	64‡ *63-65*	20.27‡ *20.01-20.51*	60‡ *58-61*	19.46‡ *19.23-19.66*	66‡ *65-68*	20.77‡ *20.61-20.93*	64‡ *63-65*
SD [21]	30.24‡ *30.15-30.32*	88‡ *88-89*	30.38‡ *30.23-30.51*	87‡ *87-88*	28.53‡ *28.38-28.68*	88‡ *87-89*	30.46‡ *30.37-30.55*	88‡ *88-89*
DDPM [6]	20.98‡ *20.85-21.12*	81‡ *81-82*	21.25‡ *21.03-21.49*	80‡ *79-81*	19.20‡ *19.02-19.40*	80‡ *79-81*	21.20‡ *21.04-21.34*	81‡ *81-82*
RSCP [17]	19.03‡ *18.96-19.09*	28‡ *28-29*	19.61‡ *19.50-19.72*	29‡ *28-29*	18.05‡ *17.93-18.18*	27‡ *26-28*	19.00‡ *18.93-19.07*	29‡ *28-29*
MDA [5]	34.36‡ *34.26-34.45*	93‡ *93-94*	34.92‡ *34.75-35.06*	93‡ *92-94*	31.13‡ *30.96-31.30*	92‡ *92-93*	34.62‡ *34.51-34.73*	94‡ *93-94*
OTE [30]	29.78‡ *29.66-29.90*	91‡ *90-92*	31.04‡ *30.86-31.22*	90‡ *90-91*	25.71‡ *25.52-25.89*	89‡ *89-90*	30.01‡ *29.89-30.13*	91‡ *91-92*
P2P [9]	31.23‡ *31.11-31.35*	90‡ *89-90*	31.45‡ *31.28-31.62*	88‡ *87-88*	28.40‡ *28.21-28.58*	89‡ *89-90*	31.55‡ *31.43-31.67*	90‡ *89-91*
PFT [20]	29.88‡ *29.73-30.04*	93‡ *93-94*	31.58‡ *31.32-31.84*	92‡ *91-93*	24.55‡ *24.34-24.76*	90‡ *89-91*	30.08‡ *29.91-30.25*	94‡ *93-94*
RDSTN [8]	28.36‡ *28.26-28.47*	86‡ *86-87*	29.34‡ *29.18-29.50*	87‡ *86-87*	24.45‡ *24.29-24.61*	84‡ *84-85*	28.71‡ *28.60-28.82*	86‡ *86-87*
CMT [12]	24.62‡ *24.52-24.71*	87‡ *86-88*	25.87‡ *25.68-26.04*	86‡ *85-87*	22.14‡ *21.96-22.31*	86‡ *85-87*	24.47‡ *24.37-24.58*	87‡ *87-88*
GFE [16]	30.81‡ *30.72-30.90*	94‡ *94-95*	30.79‡ *30.64-30.92*	92‡ *92-93*	29.84‡ *29.69-30.00*	94‡ *94-95*	30.90‡ *30.81-31.00*	94‡ *94-95*
Ours	**39.02** *38.90-39.16*	**98** *97-98*	**39.27** *39.05-39.47*	**97** *96-98*	**33.73** *33.52-33.93*	**97** *96-98*	**39.64** *39.51-39.78*	**98** *97-98*
w/o HFI	35.04‡ *34.89-35.39*	93‡ *92-93*	35.43‡ *35.19-35.69*	93‡ *92-93*	31.95‡ *31.71-32.18*	93‡ *92-93*	35.18‡ *35.03-35.34*	94‡ *93-94*
w/o FD	36.90‡ *36.70-37.08*	97‡ *96-97*	37.17‡ *36.86-37.51*	95‡ *95-96*	32.25‡ *31.96-32.51*	96‡ *95-97*	37.14‡ *36.93-37.34*	97‡ *96-97*
w/o ACR	38.36‡ *38.16-38.57*	97‡ *97-97*	38.55‡ *38.22-38.91*	96‡ *96-97*	32.59‡ *32.33-32.86*	97‡ *96-97*	38.66‡ *38.44-38.88*	97‡ *96-97*

3.2 Image Quality Assessment

Our method achieves the best PSNR and SSIM across all subsets (Table 1). Performance is consistent across the total, blur, and local-artifact subsets, indicating stable enhancement under non-color degradations. MDA attains the second-best PSNR yet remains 2–5 dB below ours. In the color subset, all methods show lower PSNR with only minor SSIM degradation, suggesting errors are dominated by global color and intensity offsets rather than structural distortions.

3.3 Ablation Study

Ours achieves the best PSNR (39.02) and SSIM (0.98) (Table 1). HFI (HFI-VAE) contributes the most: removing it yields the lowest PSNR (35.04) and SSIM (0.93). Removing FD (FD-Block) causes clear metric drops across all subsets. Removing ACR (Anatomical Consistency Regularization) yields a smaller but consistent drop, confirming the role of anatomical topology constraints.

Table 2. Disease classification and segmentation results (%). OOD results: gray background. Top-3 in **bold**, blue, underline; italics: 95% CI. **RFMiD classes**: Healthy, Diabetic Retinopathy, Media Haze, Drusen, Optic Disc Cupping. **ROP classes**: Normal, Hemorrhage, Mild Retinopathy of Prematurity, Severe Retinopathy of Prematurity. *: $p<0.05$, †: $p<0.01$, ‡: $p<0.001$. OD: optic disc; HQ: high quality; Acc: accuracy.

	Disease Classification				Segmentation (Dice↑)			
	RFMiD		ROP		OD		Vessel	
	Acc ↑	F1 ↑	Acc ↑	F1 ↑	Messidor	PAPILA	FIVES	ARIA
RAW HQ	81 *75–85*	77 *71–82*	92 *90–94*	91 *89–93*	98 *98–98*	94 *93–94*	74 *74–75*	67 *66–68*
Ours	**79** *75–83*	**75** *71–79*	**86** *85–88*	**84** *82–86*	**97** *97–97*	**92** *92–93*	**69** *66–70*	**65** *63–67*
Res	76 *72–80*	72 *68–77*	74‡ *71–76*	73‡ *71–75*	95‡ *95–96*	90‡ *89–91*	66‡ *64–67*	61‡ *59–62*
Power	63‡ *58–68*	58‡ *53–63*	45‡ *43–48*	47‡ *45–50*	74‡ *69–78*	62‡ *56–67*	53‡ *51–55*	50‡ *47–53*
SD	73† *69–77*	70* *66–74*	61‡ *59–63*	62‡ *60–65*	93‡ *93–94*	89‡ *88–91*	59‡ *57–61*	58‡ *56–60*
DDPM	74† *70–78*	72* *67–76*	65‡ *62–67*	65‡ *63–67*	95‡ *95–96*	87‡ *85–89*	64‡ *62–67*	60† *58–62*
RSCP	44‡ *40–49*	35‡ *31–39*	58‡ *55–60*	51‡ *49–54*	14‡ *10–17*	39‡ *33–44*	24‡ *22–27*	17‡ *15–18*
MDA	75* *71–78*	70† *65–74*	65‡ *63–68*	64‡ *62–67*	94‡ *93–95*	90 *88–92*	66‡ *64–69*	62 *60–64*
OTE	70‡ *66–75*	68‡ *63–72*	57‡ *54–59*	58‡ *56–61*	95‡ *94–95*	86‡ *84–89*	63‡ *60–66*	58‡ *55–60*
P2P	61‡ *57–66*	59‡ *54–63*	59‡ *57–61*	60‡ *58–63*	94‡ *92–95*	81‡ *77–84*	59‡ *55–63*	55‡ *50–58*
PFT	67‡ *63–71*	65‡ *60–69*	73‡ *71–75*	72‡ *69–74*	94‡ *94–95*	87‡ *85–89*	62‡ *58–66*	60‡ *56–62*
RDSTN	26‡ *23–31*	22‡ *18–27*	35‡ *33–37*	36‡ *34–37*	93‡ *93–94*	83‡ *80–86*	39‡ *36–41*	31‡ *29–33*
CMT	59‡ *54–63*	57‡ *52–61*	47‡ *44–49*	47‡ *45–49*	80‡ *75–84*	75‡ *70–79*	49‡ *45–53*	43‡ *38–47*
GFE	75* *71–79*	71* *66–75*	78‡ *76–80*	77‡ *75–79*	95‡ *95–96*	91* *89–92*	64‡ *61–67*	61‡ *59–63*

3.4 Downstream Classification and Segmentation

Our method achieves the best classification and segmentation performance across all evaluated datasets (Table 2). On ROP, it outperforms the second-ranked GFE by a clear margin (accuracy: 0.78→0.86, F1: 0.77→0.84). For segmentation, vessel Dice improves by 0.3 over the second-best, and OOD performance most closely approaches that of segmenting original HQ images.

3.5 Qualitative Visualization Analysis

Our method achieves the smallest spatial, low-frequency, and high-frequency residuals across all subsets (Figure 2). In the color+halo subset, it outperforms competing methods in low-frequency residuals, reflecting superior correction of large-scale illumination and color shift. In the hole+spot subset, our method recovers vessel continuity more accurately under heavy artifact occlusion, whereas MDA exhibits varying degrees of vessel structure loss.

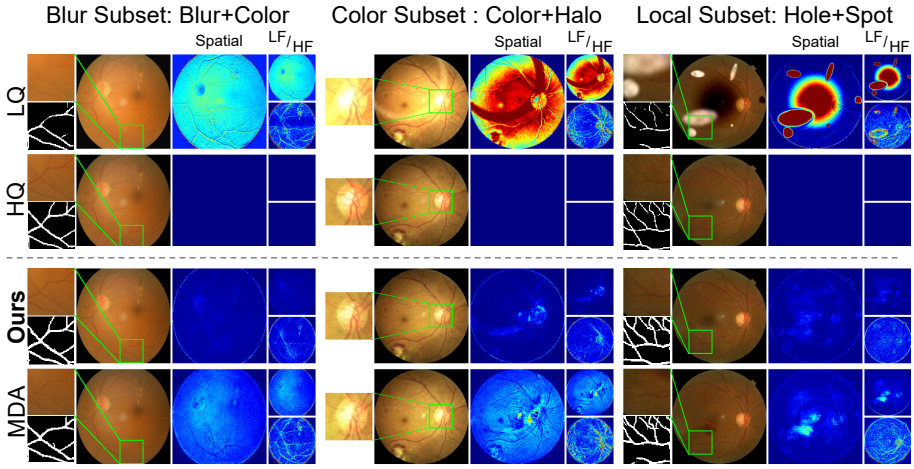

Fig. 2. Visualization of the top-2 models. Blue heatmaps are generated by subtracting the enhanced image from the HQ reference. **Darker (more blue) residuals indicate closer similarity to HQ.** LF: Low-Low band; HF: high-frequency bands.

4 Conclusion

In this work, we propose WaveDE, a wavelet-routed dual-expert network for fundus image enhancement, targeting compound degradations and the global–local optimization conflict. We use the discrete wavelet transform (DWT) for deterministic band routing, decomposing features into frequency-specific components processed by dedicated low- and high-frequency experts. To mitigate detail loss

from latent-space compression, we design a high-frequency-injected VAE decoder. We further introduce anatomical consistency regularization, guided by key anatomical priors, to preserve vessels and the optic disc and reduce structural hallucinations. Experiments show consistent improvements in image quality as well as downstream fundus classification and segmentation performance. A limitation is the reliance on synthetic paired training data, which leaves a synthetic-to-real distribution gap that may affect generalization. The main remaining failure mode is the color subset, where PSNR (33.73 dB) trails the other subsets (39+ dB) while SSIM remains on par with them, suggesting that the remaining errors stem from global color and intensity offsets rather than structural distortion. Overall, WaveDE provides a practical approach to fundus image quality enhancement.

Acknowledgments. This work was supported by the Hunan Provincial Key Research and Development Program (2025JK2115) and the National High Level Hospital Clinical Research Funding (BJ-2025-133).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Dong, J., Pan, J., Yang, Z., Tang, J.: Multi-scale residual low-pass filter network for image deblurring. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12345–12354 (2023)
2. Dugas, E., Jared, Jorge, Cukierski, W.: Diabetic retinopathy detection. https://kaggle.com/competitions/diabetic-retinopathy-detection (2015), kaggle
3. Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pranet: Parallel reverse attention network for polyp segmentation. In: Medical Image Computing and Computer Assisted Intervention. vol. 12266, pp. 263–273 (2020)
4. Guo, E., Fu, H., Zhou, L., Xu, D.: Bridging synthetic and real images: A transferable and multiple consistency aided fundus image enhancement framework. IEEE Transactions on Medical Imaging **42**(8), 2189–2199 (2023)
5. Guo, R., Xu, Y., Tompkins, A., Pagnucco, M., Song, Y.: Multi-degradation-adaptation network for fundus image enhancement with degradation representation learning. Medical Image Analysis **97**, 103273 (2024)
6. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems. vol. 33, pp. 6840–6851 (2020)
7. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)
8. Hu, J., Che, H., Li, Z., Yang, W.: Residual dense swin transformer for continuous depth-independent ultrasound imaging. In: IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 2280–2284 (2024)
9. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)
10. Jiang, L., Dai, B., Wu, W., Loy, C.C.: Focal frequency loss for image reconstruction and synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13919–13929 (2021)

11. Kiefer, R., Abid, M., Steen, J., Ardali, M.R., Amjadian, E.: A catalog of public glaucoma datasets for machine learning applications: A detailed description and analysis of public glaucoma datasets available to machine learning engineers tackling glaucoma-related problems using retinal fundus images and oct images. In: Proceedings of the 2023 7th International Conference on Information System and Data Mining. pp. 24–31 (2023)
12. Ko, K., Kim, C.S.: Continuously masked transformer for image inpainting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13169–13178 (2023)
13. Krzywicki, T., Brona, P., Zbrzezny, A.M., Grzybowski, A.E.: A global review of publicly available datasets containing fundus images: Characteristics, barriers to access, usability, and generalizability. Journal of Clinical Medicine **12**(10) (2023)
14. Li, C., Fu, H., Cong, R., Li, Z., Xu, Q.: Nui-go: Recursive non-local encoder-decoder network for retinal image non-uniform illumination removal. In: Proceedings of the 28th ACM International Conference on Multimedia. p. 1478–1487 (2020)
15. Li, H., Liu, H., Fu, H., Shu, H., Zhao, Y., Luo, X., Hu, Y., Liu, J.: Structure-consistent restoration network for cataract fundus image enhancement. In: Medical Image Computing and Computer Assisted Intervention. pp. 487–496 (2022)
16. Li, H., Liu, H., Fu, H., Xu, Y., Shu, H., Niu, K., Hu, Y., Liu, J.: A generic fundus image enhancement network boosted by frequency self-supervised representation learning. Medical Image Analysis **90**, 102945 (2023)
17. Lin, X., Zhou, Y., Yue, J., Ren, C., Chan, K.C., Qi, L., Yang, M.H.: Re-boosting self-collaboration parallel prompt gan for unsupervised image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(11), 9827–9844 (2025)
18. Liu, H., Li, H., Fu, H., Xiao, R., Gao, Y., Hu, Y., Liu, J.: Degradation-invariant enhancement of fundus images via pyramid constraint network. In: Medical Image Computing and Computer Assisted Intervention. pp. 507–516 (2022)
19. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J., Liu, Y.: Vmamba: Visual state space model. In: Advances in Neural Information Processing Systems. vol. 37, pp. 103031–103063 (2024)
20. Long, W., Zhou, X., Zhang, L., Gu, S.: Progressive focused transformer for single image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2279–2288 (2025)
21. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10684–10695 (2022)
22. Şevik, U., Köse, C., Berber, T., Erdöl, H.: Identification of suitable fundus images using automated quality assessment methods. Journal of biomedical optics **19**(4), 046006–046006 (2014)
23. Shen, Z., Fu, H., Shen, J., Shao, L.: Modeling and enhancing low-quality retinal fundus images. IEEE Transactions on Medical Imaging **40**(3), 996–1006 (2021)
24. Silva-Rodríguez, J., Chakor, H., Kobbi, R., Dolz, J., Ben Ayed, I.: A foundation language-image model of the retina (flair): encoding expert knowledge in text supervision. Medical Image Analysis **99**, 103357 (2025)
25. Timkovič, J., Nowaková, J., Kubíček, J., Hasal, M., Varyšová, A., Kolarčík, L., Maršolková, K., Augustýnek, M., Snášel, V.: Retinal image dataset of infants and retinopathy of prematurity. Scientific Data **11**(1), 814 (2024)
26. Wang, X., Yu, K., Dong, C., Loy, C.C.: Recovering realistic texture in image super-resolution by deep spatial feature transform. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)

27. Yue, Z., Wang, J., Loy, C.C.: Resshift: Efficient diffusion model for image super-resolution by residual shifting. In: Advances in Neural Information Processing Systems. vol. 36, pp. 13294–13307 (2023)
28. Zhang, S., Mohan, A., Webers, C.A., Berendschot, T.T.: Mute: A multilevel-stimulated denoising strategy for single cataractous retinal image dehazing. Medical Image Analysis **88**, 102848 (2023)
29. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Transactions on Medical Imaging **39**(6), 1856–1867 (2020)
30. Zhu, W., Qiu, P., Farazi, M., Nandakumar, K., Dumitrascu, O.M., Wang, Y.: Optimal transport guided unsupervised learning for enhancing low-quality retinal images. In: IEEE 20th International Symposium on Biomedical Imaging. pp. 1–5 (2023)
31. Zhuang, J., Zeng, Y., Liu, W., Yuan, C., Chen, K.: A task is worth one word: Learning with task prompts for high-quality versatile image inpainting. In: Proceedings of the European Conference on Computer Vision. pp. 195–211 (2025)