electronics

Article

MWRD (Mamba Wavelet Reverse Diffusion)—An Efficient Fundus Image Enhancement Network Based on an Improved State-Space Model

Tianle Chen [1,†], Yu Li [1,†] and Xuedong Zhang [2,*]

[1] School of Electronic and Electrical Engineering, Wuhan Textile University, Wuhan 430077, China; 2215363002@mail.wtu.edu.cn (T.C.); leewoo@wtu.edu.cn (Y.L.)
[2] School of Information Engineering, Tarim University, Alaer 843300, China
* Correspondence: zxd612@taru.edu.cn
† These authors contributed equally to this work.

Abstract: This study introduces an efficient fundus image enhancement framework based on an improved Mamba model and the Denoising Diffusion Probabilistic Model (DDPM). By integrating wavelet transform for local feature extraction and applying a reverse diffusion process, this approach significantly improves the effectiveness and efficiency of enhancing low-quality fundus images. The model achieves high-precision enhancement of retinal vessel details while substantially reducing the number of parameters. Comprehensive tests on publicly available datasets show that the suggested approach surpasses various advanced low-light image enhancement methods in both quantitative and qualitative assessments.

Keywords: fundus image enhancement; state-space model; diffusion model; wavelet transform

check for updates

Citation: Chen, T.; Li, Y.; Zhang, X. MWRD (Mamba Wavelet Reverse Diffusion)—An Efficient Fundus Image Enhancement Network Based on an Improved State-Space Model. *Electronics* **2024**, *13*, 4025. https://doi.org/10.3390/electronics13204025

Academic Editor: Reza K. Amineh

Received: 22 August 2024
Revised: 30 September 2024
Accepted: 9 October 2024
Published: 13 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Fundus images play a crucial role in assisting ophthalmologists in diagnosing fundus diseases, particularly various retinal disorders. Nevertheless, image quality is affected by multiple factors, including pupil size, refractive media conditions, lighting effects, and camera pixels [1–3]. Low-quality images hinder ophthalmologists from making accurate diagnoses. Figure 1 illustrates this, where the color fundus image in (A) is of good quality, while the color fundus image in (B) is of low quality.These images fail to clearly reveal the retina and vascular structures of the fundus due to factors such as pupil size, refractive media conditions, lighting effects, and camera resolution, resulting in unclear contrast of colors and insufficient brightness. Disease characteristics visible in high-quality images are challenging to detect in low-quality ones:for example, retinal tears, acute zonal occult outer retinopathy (AZOOR), and branch retinal vein occlusion (BRVO). Moreover, the best generating process, reverse diffusion, is applied to maximize the initial output and produce the most lifelike pictures.

Effectively enhancing medical low-light images remains challenging. Early methods focused on optimizing image parameters, whose effectiveness largely depended on the accuracy of manual priors. However, enhancing medical low-light images is more complex, as anatomical and pathological features in medical images follow specific spatial distributions and multiple scales. Although numerous methods for fundus image enhancement have emerged in recent years [4–6], these methods exhibit certain limitations in capturing the structural and pathological features, particularly when dealing with fine-scale details. This is mainly because these features can become blurred or disappear during deep network processing. Therefore, there is room for improvement in existing algorithms in extracting and preserving these critical features.



Figure 1. Illustration of high-quality and low-quality color fundus images. (**A**) High-quality color fundus image, (**B**) low-quality color fundus image.

Currently, learning-based techniques demonstrate improved results by establishing a direct mapping between normal and low-light images for enhancement learning. especially in distortion metrics like Peak Signal-to-Noise Ratio (PSNR),Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) [7,8].

Moreover, diffusion models [9–11] have garnered attention in image restoration due to their generative and generalization capabilities. Nevertheless, numerous challenges remain in the application of diffusion models to improve low-light images. The diversity of Gaussian noise during the inference process leads to chaotic information. Additionally, the large number of parameters in diffusion models results in increased inference time. In order to improve low-light photographs in natural images, Jiang et al. [12] applied wavelet transform to diffusion models, transforming low-light images into the wavelet domain., representing images sparsely, significantly reducing model complexity and computation, and improving training efficiency. However, some detail information might be lost during wavelet transform, especially at larger scale decompositions, potentially affecting the final model performance.

Lately, state-space models (SSMs), particularly the enhanced Mamba model, have demonstrated significant capability in capturing global relationships with linear complexity [13–17]. Unlike Transformers [18], Mamba models do not face quadratic complexity adjustments, greatly reducing parameter counts, achieving great success in the medical imaging field. However, challenges remain in local pixel forgetting and channel redundancy in medical imaging. Local pixel forgetting refers to the phenomenon where certain local details (such as pixel-level features) are overlooked by the network during image processing due to the focus on learning global features. For instance, in fundus images, the tiny vessel details in the retina can easily be neglected in low-light or blurred conditions. Channel redundancy refers to the presence of similar or redundant information across different channels, leading to wasted resources. For example, in processing fundus images, redundant information carried by the color channels may result in less significant enhancement effects. It has been demonstrated that the Mamba model can be combined with various deep learning techniques to achieve stronger network performance [19,20]. By integrating the improved state-space model with the reverse diffusion method of Denoising Diffusion Probabilistic Models (DDPMs), we can retain image details to the maximum extent without significantly increasing computational cost.

Our principal findings are summed up as follows:

1. We combine the improved Mamba module with reverse diffusion and apply it to fundus image enhancement for the first time, significantly reducing the model's parameter count without compromising generation quality.
2. We propose the Mamba Wavelet (MW) module, which uses curved attention modules and wavelet transform to efficiently extract high-frequency detail data. The enhanced Mamba model is then used for successful long-range modeling.
3. Extensive tests using datasets that are accessible to the public show how well our suggested model performs.

2. Related Works

2.1. Medical Low-Light Image Enhancements

To enhance medical images under low-light conditions, various traditional algorithms such as intensity transformation [21], generative adversarial networks (GANs) [22], Retinex theory [23], histogram equalization [24], and diffusion models [11] have been applied. Fundus image enhancement focuses more on emphasizing high-level structural features relevant to disease diagnosis, such as improving the distinction of lesion areas. Li et al. [25] proposed an algorithm using a global encoder-decoder residual module called NuI-Go to remove uneven illumination in color fundus images. Lee et al. [26] collected a dataset containing high-quality and low-quality fundus images from 2017 to 2019 and proposed a CNN-based fundus image enhancement framework. You et al. [27] proposed an algorithm called CycleCBAM for enhancing color fundus images. This algorithm is an improvement of the CycleGAN network [28], introducing the CBAM module [29] into the CycleGAN network and using cycle consistency loss to constrain the transformation between low-quality and high-quality fundus images. Ma et al. [30] introduced an algorithm named StillGAN, which incorporates a brightness constraint loss to ensure overall brightness balance and optimizes brightness variance to improve the dark regions of fundus images. Hu et al. [31] were the pioneers in proposing fundus image enhancement using diffusion models, utilizing its unique regression mechanism for better enhancement effects. Cheng et al. [32] introduced a diffusion model for enhancing fundus images through degradation learning, finding the degradation mapping from high-quality to low-quality photos using a data-driven degradation framework. Bai et al. [33] presented a network architecture for enhancing low-light medical images using CNNs and diffusion models, leveraging the excellent feature extraction capabilities of CNN and the reverse diffusion process of diffusion models to better enhance low-light medical pictures.

2.2. State-Space Models

SSM have recently been introduced as models for state-space transformation in deep learning [15,17,34]. When paired with HiPPO [35] initialization and drawing upon continuous state-space models from control systems, promising results have been obtained while using LSSL [34] to handle long-range dependence issues. One of the first deep state-space models for describing long-range relationships is the Structured State-Space Sequence Model (S4) [15,36]. Mamba [14], a data-dependent SSM that scales linearly with input length and features a selective mechanism and effective hardware architecture, has recently beat Transformers in natural language challenges. Additionally, groundbreaking work has applied Mamba to visual tasks such as image classification [19,37], video analysis [38], and image segmentation [39–41]. Guo et al. [42] combined Mamba with channel attention for image restoration, demonstrating the capabilities of the Mamba model as a powerful visual foundation model. However, its performance in enhancing fundus images, particularly in preserving vascular details, remains to be improved. By creating a Mamba-based fundus image improvement network model, we further investigate Mamba's potential in medical picture enhancement in this study.

2.3. Diffusion Model

Diffusion models [11], have attained remarkable outcomes in image generation by refining the lower bound of the likelihood function through parameterized Markov chains, enabling them to generate more accurate target distributions compared to other generative models (i.e., GANs) [22]. Consequently, many image recovery tasks, including super-resolution [43], inpainting [44], and deblurring [45], have been introduced into diffusion models. For image enhancement tasks, WCDM [12] combined wavelet transform with diffusion models to design an efficient and robust diffusion restoration framework. Hu et al. [31] first applied diffusion models to fundus image denoising. Cheng et al. [32] designed the LED model, using a data-driven approach to learn the mapping relationship between unpaired high- and low-quality images to achieve enhancement and applying conditional diffusion models to learn the reverse enhancement process. Bai et al. [33] proposed a network framework for medical image enhancement based on CNN and diffusion models, leveraging the excellent reverse diffusion process of diffusion models to generate enhanced medical images. Diffusion models by themselves, however, may result in the distortion and loss of features in medical images. To address this issue, we decided to design additional modules to learn local details of the fundus, generating more realistic fundus images.

3. Methods

This section presents the MWRD model.

3.1. Preliminaries

3.1.1. Mamba

Currently, all SSMs, such as S4 [15] and Mamba, map the input data or sequence $x(t) \in R \longrightarrow y(t) \in R$ through an intermediate state $h(t) \in R^N$. The input sequence x represents time-series data, which is arranged in chronological order and is typically used to describe the variation of one or more variables over time. Here, $\in$ indicates that an element belongs to a certain set, where R denotes the set of real numbers, while N represents the set of natural numbers. The transformation process can be described using an equation as follows:

$$\begin{aligned} h'(t) &= \alpha h(t) + \beta x(t) \\ y(t) &= \gamma h(t) \end{aligned} \tag{1}$$

The matrix β influences the input $x(t)$, while the matrix α affects the previous state $h(t)$. The derivative $h'(t)$ represents the change in state, which indicates the representation of the next state. $y(t)$ is the defined output equation, describing how the state is transformed into the output through the matrix γ: this means predicting the system's future state using the input sequence and the prior state.

Since computers are inherently discrete, discrete time and discrete space data must be processed to facilitate simulation and analysis of the model on a computer. Specifically, the discrete parameters of the model are represented by $\bar{\alpha}$ and $\bar{\beta}$, while Δ denotes the discrete time step. The discretization is performed using the Zero-Order Hold (ZOH) method, as described by the following equations:

$$\begin{aligned} \bar{\alpha} &= \exp(\Delta\alpha) \\ \bar{\beta} &= (\Delta\alpha)^{-1}(\exp^*(\Delta\alpha) - I) \ldots \Delta\beta \end{aligned} \tag{2}$$

Equations (3) and (4), which represent linear recurrence and global convolution, respectively, may be used to calculate SSM-based models after discretization.

$$\begin{aligned} h'(t) &= \bar{\alpha} h(t) + \bar{\beta} x(t) \\ y(t) &= \gamma h(t) \end{aligned} \tag{3}$$

$$\begin{aligned} K' &= \left\{ \gamma\bar{\beta}, \gamma\bar{\alpha}\bar{\beta}, \ldots, \gamma\bar{\alpha}^{L-1}\bar{\beta} \right\} \\ y &= x * \bar{K} \end{aligned} \tag{4}$$

Here, $\bar{K} \in R^L$ represent a structured convolution kernel, '. . . ' indicating the computation of higher-order terms in the matrix or sequence, and L represents the length of the input sequence x.

3.1.2. Diffusion Models

This section provides a brief introduction to diffusion models. For further information, see [11]. Among the generative models that may produce samples from Gaussian noise are diffusion models.

Starting with a sample x_0 and its associated distribution $q(x_0)$, the forward process produces a sequence of samples $\{x_T\}_{t=1}^{T}$ derived from $q(x_0)$ with the samples being corrupted by Gaussian noise. This process can be represented as

$$q(x_t \mid x_{t-1}) = N\Big(x_t; \sqrt{1-\beta}x_{t-1}, \beta_t I\Big), \qquad t = 1, 2, 3, \ldots, T, \tag{5}$$

where $[\beta_1, \ldots\ldots, \beta_T]$ represents a predefined or learned variance schedule. Sampling from $q(x_0)$ at time step t is represented as

$$q(x_t \mid x_0) = N(x_t; \sqrt{\alpha_t}x_0, 1-\alpha_t I), \tag{6}$$

where $\alpha_t = \prod_{i=1}^{t}(1-\beta_i)$. The process $p(x_0 \mid x_T)$used to learn the recovery of images is a key step in the diffusion model, and this process $p(x_{t-1} \mid x_t)$ can be represented by the following equation:

$$p_\theta(x_{t-1} \mid x_t) = N\big(x_{t-1}; \mu_\theta(x_t, t), \tilde{\beta}_t I\big), \tag{7}$$

where $\tilde{\beta} = \frac{1-\alpha_{t-1}}{1-\alpha_t}\beta_t$, where θ represents the learnable parameters. Thus, the reverse process aiming to denoise corrupted samples can be expressed as

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{1-\beta_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\alpha_t}} \in \theta(x_t, t)\right) \tag{8}$$

where ζ_θ represents the noise estimation network. We further rephrase Equation (6) as $x_t(x_0, \zeta) \in \sqrt{\alpha_t}x_0 + \sqrt{1-\alpha_t}\zeta$, using the reparameterization trick, The reparameterization trick refers to restructuring the sampling process of a random variable into a differentiable form by introducing additional variables, commonly used in variational inference, to enable optimization through gradient descent where $\zeta \sim N(0, I)$. DDPM employs an optimization strategy that approximates the true data distribution by maximizing its variational lower bound. The formula is presented as follows:

$$L_\theta(x_0, t) = E_{x_0, \zeta, t}\left[\left\|\zeta - \zeta_\theta\Big(\sqrt{\alpha_t}x_0 + \sqrt{1-\alpha_t}\zeta, t\Big)\right\|^2\right] \tag{9}$$

In practice, DDPM minimizes the objective function by randomly sampling from $[0, T]$, where T is a hyperparameter representing the total time steps.

3.2. Proposed Framework

As illustrated in Figure 2, our approach utilizes the generative capabilities of diffusion models, the advantages of wavelet transform, and the excellent local feature extraction and long-range modeling abilities of the improved Mamba. The following section will detail the proposed MWRD model framework.

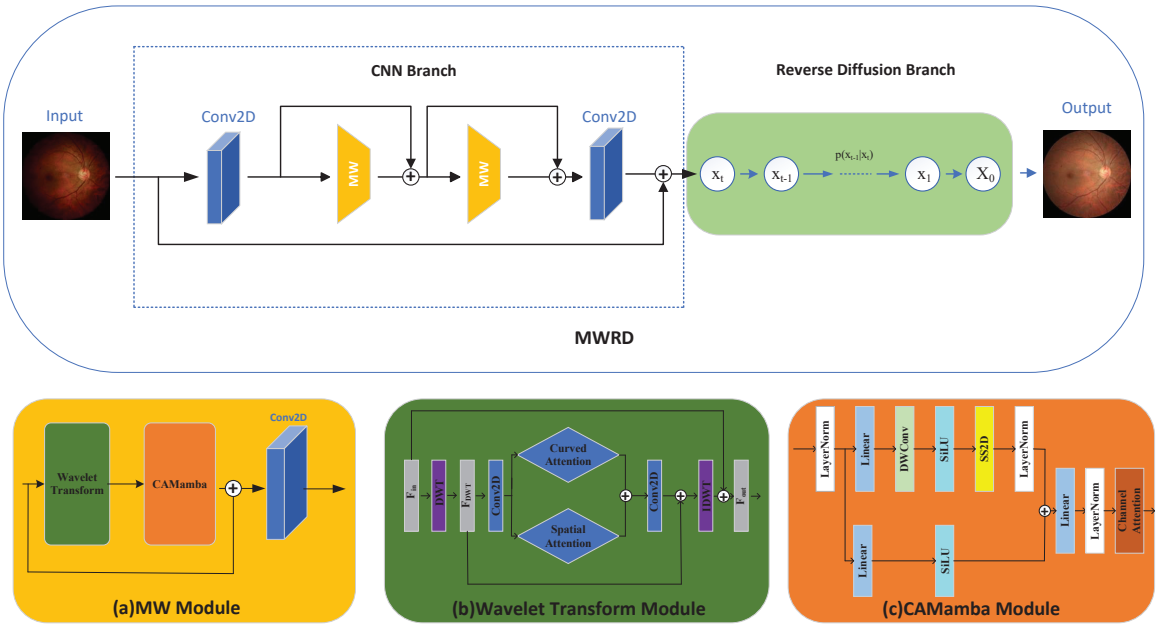

Figure 2. An outline of our MWRD. After the CNN branch extracts the shallow picture output, Markov chain inference is used by the DDPM branch to further optimize the image. (**a**) shows the MW module, which includes the wavelet transform module, CAMamba module, and a convolution layer with residual connections for better learning of image features. The attention learning and feature recovery wavelet transform module is displayed in (**b**), where DWT and IDWT stand for discrete wavelet transform and inverse wavelet transform, respectively. (**c**) shows the CAMamba module for channel attention. DWConv represents depthwise convolution.

3.2.1. CAMamba

Due to the existing challenges of local pixel forgetting and channel redundancy in the Mamba module, inspired by Guo et al. [42], we introduce residual structures and channel attention [46] mechanisms in the Mamba module to focus on learning different channel representations, which are crucial for subsequent channel selection and avoiding channel redundancy. The addition of residual structures and convolutional layers in the module enhances local feature learning, compensating for the issue of local pixel forgetting in the Mamba module.

In Figure 2c, to handle the input, the CAMamba module first applies a layer normalization layer. It is worth noting that after layer normalization, the input feature map is divided into two branches for processing. The input travels via an activation function and a linear layer in the first branch of CAMamba. In the second branch, the input is passed via an activation function; after passing through a depthwise separable convolution and a linear layer, the image input is fed into the two-dimensional selective scanning module (SS2D) [19] for further learning. Figure 3 presents an illustration of the SS2D structure. After applying layer normalization to the features, the outputs of the two branches are merged through multiplication. Finally, the output of the convolution-CAMamba block is obtained by adding residual connections and using a linear layer for feature mixing. Notably, the activation function in the CAMamba module defaults to SiLU [47]. At the end of the network, we use another convolution operation and a channel attention mechanism to enhance the learning of channel information. The features learned before and after the channel attention are merged, and a convolutional layer is then used to merge the feature maps facilitating information exchange between channels to generate the final output of CAMamba. This process can be summarized as follows:

$$F_{l+1}^{D} = CA\left(Conv\left(LN\left(Z^{l}\right)\right)\right) + s' \cdot Z^{l} \tag{10}$$

The conventional Mamba processes incoming data in a causal manner, extracting information from a segment of the data scan. While this approach is effective for sequential NLP tasks, it poses considerable difficulties when applied to non-causal data, like images. To overcome this, we propose the 2D Selective Scan Module (2D-SSM), which, following the method in [19], leverages two-dimensional spatial information more efficiently, as demonstrated in Figure 3.

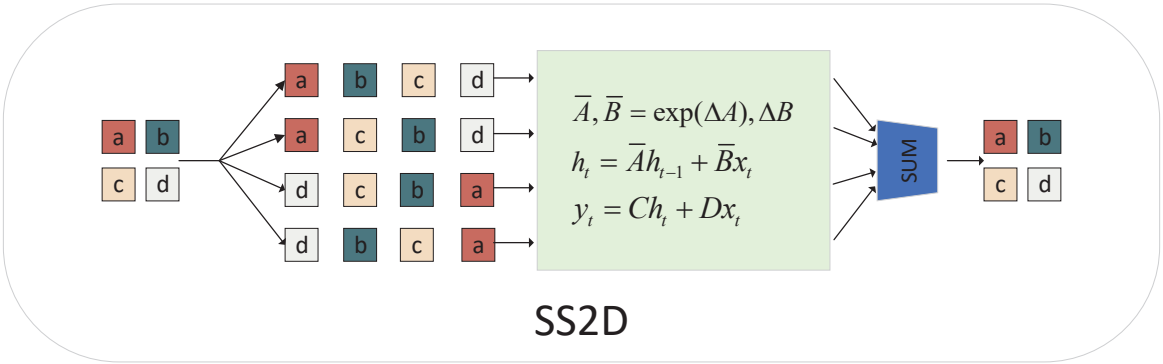

Figure 3. An illustration of the SS2D function. Following the first patch segmentation of the input picture, four directions of scanning are applied to the two-dimensional image features: horizontal, vertical, main diagonal, and secondary diagonal. One-dimensional sequences are created by flattening the two-dimensional visual characteristics. Subsequently, the outputs are merged to construct a two-dimensional feature map as the final output.

3.2.2. MW (Mamba Wavelet Block)

The Mamba Wavelet (MW) block serves as the central module of our CNN network. As shown in Figure 2b, it first consists of a wavelet transform module that transforms the input image Fin into the wavelet space FDWT, using the discrete wavelet transform to extract high-frequency detail information (DWT). This is accomplished by DWT by gradually breaking down the signal into various scales of high-frequency and low-frequency components. The advantage of DWT is that it efficiently provides multi-scale resolution, allowing the network model to learn more details under different illuminations in image enhancement. Motivated by Bai et al. [33], we incorporate bending attention mechanisms and spatial attention mechanisms following the wavelet transform to better capture local illumination and insufficient lighting. With c representing the feature location coordinates, $Curve_{n-1}$ representing the pixel-wise curve parameters, and $ILn(c)$ as the curve function, we can derive the curve estimation formula as

$$\frac{ILn(c)}{ILn-1(c)} = Curve_{n-1}(1 - ILn - 1(c)) \tag{11}$$

We use bending attention blocks and parallel spatial attention blocks to build our bending dual attention module. The convolutional features' spatial dependencies are used in spatial attention (SA) [48]. This attention mechanism places greater emphasis on spatial features. Then, deep feature learning is propagated through feature selectors and dual attention modules. The learned feature maps are then inverse wavelet transformed (IDWT) to obtain new feature images. Finally, the output image is concatenated with the input image Fin, then reduced in dimensionality to obtain Fout, and input into the CAMamba module for enhanced feature learning.

3.2.3. Reverse Diffusion Process

Diffusion models may be used with other DL-based techniques in some publications [49,50] to lower training costs and for applications later on. In our study, we maximize the shallow output of the CNN branch by combining the DDPM reverse diffusion process in a smart way. Among the aforementioned publications [33,49,50], the

combination of diffusion models with other deep learning techniques has shown remarkable effectiveness in improving image quality and supporting therapeutic applications. The learning goal Y^* of the CNN is assumed to be i_0 in our approach. For the LLIE job, we thus simply need to utilize the opposite technique. The Markov chain is used to simulate the opposite process:

$$p_\theta(i_0 : T) = p(i_T)\prod_{t=1}^{T} p_\theta(i_{t-1} \mid i_t) \tag{12}$$

$$p_\theta(i_{t-1} \mid i_t) = N(i_{t-1}; \mu_\theta(i_t, t)), \sum_\theta(i_t, t) \tag{13}$$

where $p_\theta(i_t - 1 \mid i_t)$ is a parameterized Gaussian distribution, and its mean $\mu_\theta(i_t, t)$ and variance $\sum_\theta(i_t, t)$ are given by the trained network. Meanwhile, we directly utilize the reverse diffusion process as the output structure of the overall network. In order to achieve the expected picture Y, the shallow output is refined through the reverse diffusion branch. By employing just pixel-level loss to the final output picture, we further streamline the optimization function and increase the effectiveness of training and convergence.

3.2.4. Overall Network Architecture

Figure 2 illustrates the overall architecture of our MWRD network, which primarily consists of a CNN module for feature extraction and a reverse diffusion branch for generating enhanced images. The CNN module is composed of a shallow feature extractor (SFE), a feature extraction block MW, and an output module (POUT). The SFE is a Conv2D convolutional layer, while the stacked MW blocks enhance the extraction and learning of image features. POUT is another convolutional layer responsible for reconstructing the extracted features into image pixels, utilizing residual connections to optimize the network. Therefore, given a low-light image $x \in R^{3\times W\times H}$, where W and H represent the width and height, respectively, the CNN branch can be represented as

$$\begin{aligned} F_{SFE} &= H_{SFE}(X) \\ F_{MW} &= H_{MW}(F_{SFE}) \\ F_{POUT} &= H_{POUT}(F_{MW}) + x \end{aligned} \tag{14}$$

3.3. Loss Function

To fairly compare with previous works [21,33,51,52], we optimize the MWRD for image SR with L1 loss, which can be represented as

$$L = \left\| I_{HQ} - I_{LQ} \right\|_1 \tag{15}$$

where $\|\cdot\|_1$ represents the L1 norm. For image denoising, we use the Charbonnier [53], where $\varepsilon \in 10^{-3}$:

$$L = \sqrt{\left\| I_{HQ} - I_{LQ} \right\|^2 + \varepsilon^2} \tag{16}$$

4. Experiments

4.1. Dataset

We trained the MWRD model on the EyeQ dataset [54]. It has 28,792 fundus pictures classified as "good", "usable", and "reject" according to three quality levels. A total of 2400 photos were chosen at random; 2000 were utilized for training and 400 for testing. To simulate low-light conditions, we employed random gamma correction and reduced illumination. [55,56].

4.2. Evaluation Metrics

PSNR [57] is one of the common metrics used to measure image quality, particularly in the field of image compression. The main comparison is made between the original and processed/compressed image quality differences. PSNR represents the degree of picture distortion on a logarithmic scale and is computed using the mean squared error (MSE)

between the original and processed images. Higher PSNR values suggest improved image quality and reduced distortion. The PSNR calculation formula is shown in Equation (17):

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX^2}{MSE}\right) \tag{17}$$

where MSE stands for the mean squared error and MAX is used to calculate the maximum energy of the signal. A measure for comparing the similarity of two pictures that takes into account contrast, structural, and brightness variations is SSIM [58]. SSIM is calculated based on a weighted combination of luminance, contrast, and structural similarity, providing a better simulation of the human visual system's characteristics. The SSIM formula is provided in Equation (18):

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{18}$$

where x and y represent the two images, μ_x and μ_y are their mean values, σ_x and σ_y represent their standard deviations, σ_{xy} is their covariance, and C_1 and C_2 are constants to avoid division by zero.

A measure called LPIPS [59] is used to evaluate how two pictures vary perceptually, measuring image similarity through feature representations learned by deep learning models. Unlike traditional pixel-level metrics, LPIPS better captures perceptual differences as recognized by the human visual system, making it widely used in tasks such as image generation and style transfer. The LPIPS calculation is not a simple mathematical formula but rather the distance between features computed by the model. The LPIPS value is typically obtained by calculating the feature distance between pictures using a pre-trained model.

4.3. Implementation Details

We compared the performance of MWRD with the following state-of-the-art (SOTA) LLIE methods: LIME [21], LLFlow [51], LED [32], EnlightenGAN [60], MIRNet [52], StillGAN [30], DDPM [11], and LLCaps [33]. We trained our model for 200 epochs using the Adam optimizer with a batch size of 4 and a learning rate of 1×10^{-4}.

Three widely used metrics for assessing picture quality were utilized in our evaluation: LPIPS [59], SSIM [58], and PSNR [57]. Furthermore, we conducted external validation on 100 genuine pictures taken from the eyeQ dataset [54] in order to assess performance on real data. These were low-brightness photos that we excluded from our original trials.

All of our experiments were conducted using the PyTorch framework on an NVIDIA RTX 3090 GPU. The results are averaged over a three-fold cross-validation.

4.4. Results

4.4.1. Comparison with Other Enhancement Models

We compared the performance of MWRD against existing methods, as illustrated in Table 1 and Figure 4. Our approach outperformed all others across every metric. Specifically, it achieved a PSNR improvement of 2.69 dB and a 93.16% SSIM on the EyeQ dataset compared to LLCaps [33]. PSNR effectively reflects the quality of the enhanced fundus images in comparison to the original ones, while SSIM measures differences in brightness and structure. Compared to the original DDPM model [11], our method achieves a 7.95 dB increase in PSNR and a 31.41% improvement in SSIM. Our method also surpasses the high-performing LED [32] method with an 8.44% and 0.56% improvement in PSNR and SSIM, respectively, and it shows superior performance in LPIPS, trailing the best LLCaps [33] model by only 0.50%. All metrics indicate optimal performance. Figure 4 presents the qualitative results on the EyeQ dataset [54], comparing our method with other approaches. Additionally, our model significantly leads in parameter efficiency. The redesigned network modules achieve a lightweight fundus image enhancement framework, delivering high

performance with reduced training time. Our model achieves much higher image quality than LIME [21] without significantly increasing parameter count and shows substantial parameter efficiency compared to other high-quality image enhancement models. Furthermore, during external validation, our approach achieved the top results on an actual low-light dataset. In the table, the arrows indicate the direction of performance improvement: an upward arrow indicates that an increase in the metric value is favorable, while a downward arrow indicates that a decrease in the value is favorable.

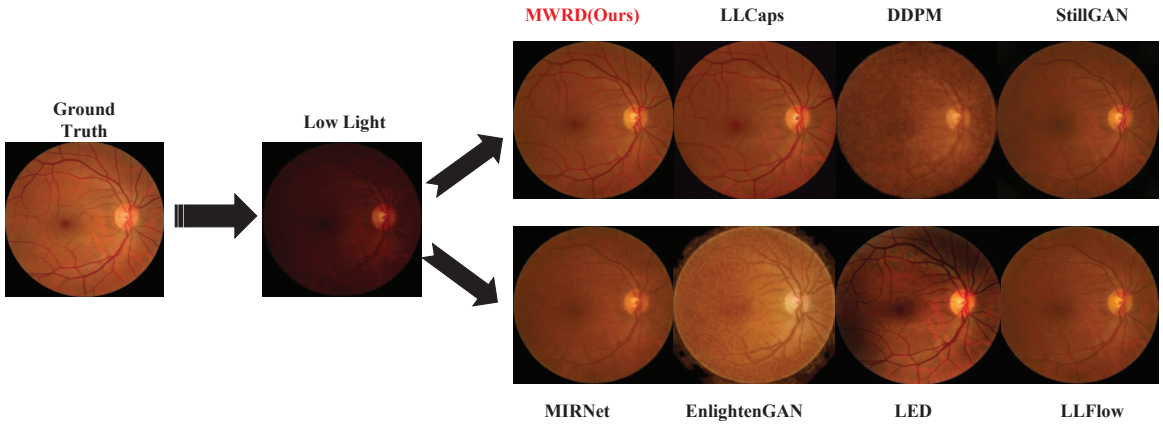

Figure 4. Quantitative results of MWRD compared to SOTA methods on the EyeQ dataset [54].

Additionally, to verify the robustness and generalization ability of our MWRD model, we performed experiments on another publicly available endoscopy dataset, Kvasir-Capsule [61], following the data processing methods outlined in [33]. As illustrated in Table 2, we can observe our experimental performance on the Kvasir-Capsule dataset, demonstrating that our model performs strongly on the PSNR and SSIM metrics. This further demonstrates that our model has strong generalization capabilities and significant potential for application in other fields.

Table 1. A comparison between the proposed MWRD and other state-of-the-art methods.

Models	eyeQ			Parameters
	PSNR ↑	SSIM ↑	LPIPS ↓	
LIME [21]	10.46 dB	0.2719	0.5863	1.78 M
LLFlow [51]	27.44 dB	0.8964	0.0918	28.16 M
LED [32]	29.49 dB	0.9264	0.0758	56.47 M
EnlightenGAN [60]	26.54 dB	0.8432	0.1913	11.15 M
MIRNet [52]	28.73 dB	0.8920	0.0656	31.79 M
StillGAN [30]	23.44 dB	0.8693	0.1108	10.24 M
DDPM [11]	24.03 dB	0.7089	0.4551	64.22 M
LLCaps [33]	29.29 dB	0.9200	0.0598	119.83 M
MWRD (Ours)	31.98 dB	0.9316	0.0601	3.79 M

Table 2. Comparison of MWRD with other SOTA methods on the Kvasir–Capsule dataset [61].

Models	PSNR ↑	SSIM ↑	LPIPS ↓
LIME [21]	12.07 dB	0.2966	0.4401
LLFlow [51]	29.69 dB	0.9257	0.0774
LED [32]	27.87 dB	0.8973	0.0971
EnlightenGAN [60]	27.15 dB	0.8503	0.1769
MIRNet [52]	31.23 dB	0.9577	0.0436
StillGAN [30]	28.28 dB	0.9130	0.1302
DDPM [11]	25.17 dB	0.7316	0.4089
LLCaps [33]	35.24 dB	0.9634	0.0374
MWRD (Ours)	36.44 dB	0.9782	0.0150

4.4.2. Qualitative Analysis

As illustrated in Figure 5, we used the IterNet [62] model to segment blood vessels in degraded test images from the EyeQ dataset. IterNet is a deep learning model for image segmentation that accurately identifies vascular regions in fundus images. By segmenting blood vessels in the degraded test images from the EyeQ dataset, this allows us to assess how well image enhancement algorithms preserve vascular structures. Firstly, it is evident that directly training DDPM [11] for LLIE does not achieve satisfactory image restoration with the structure of DDPM images being significantly disrupted. EnlightenGAN [60] demonstrates subpar performance in structural restoration. While LED [32] performs well, our method only differs slightly in restoring the peripheral vessels of the fundus thanks to the improved Mamba module's detailed learning of fundus images. We evaluated the segmentation experiments on the enhanced results of the EyeQ dataset in Figure 5, and we used commonly employed evaluation metrics in segmentation experiments, namely mean Intersection over Union (mIoU) and Dice similarity coefficient (Dice), to assess the segmentation performance of our enhanced images. Our experimental results are presented in Table 3. As observed, our model demonstrates a notable advantage in both mIoU and Dice metrics. Furthermore, our method outperforms LLFlow [51] and LLCaps [33] in terms of illumination restoration. The fundus vessel segmentation pictures also demonstrate how well our approach performs in restoring structure and lighting from low-light images.

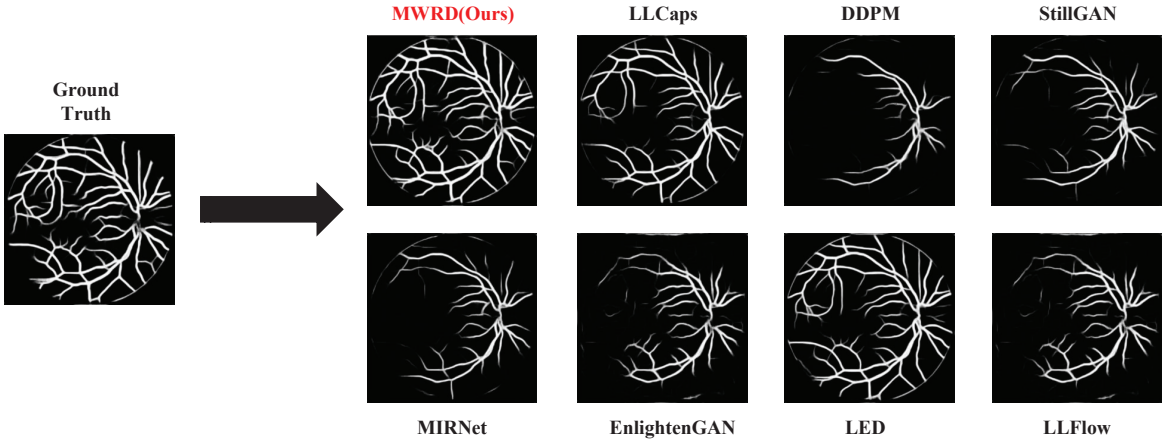

Figure 5. Comparison of vascular segmentation results before and after enhancement of fundus images.

Table 3. Comparison of the proposed MWRD with other SOTA methods on segmentation metrics.

Model	mIoU(%)	Dice(%)
LLFlow [51]	52.81	69.12
LED [32]	96.50	98.22
EnlightenGAN [60]	79.14	88.36
MIRNet [52]	36.49	53.47
StillGAN [30]	34.06	50.81
DDPM [11]	21.55	35.46
LLCaps [33]	89.68	94.56
MWRD (Ours)	97.70	98.84

4.5. Ablation Studies

Furthermore, as Table 4 illustrates, we performed ablation tests on the eyeQ dataset to illustrate the effectiveness of our network structure and its various modules. We (i) took the dual-branch attention and wavelet transform out of the MW module in order to track any changes in performance. The results from the ablation experiments indicate that the wavelet transformation module has a significant impact on multiple metrics. The wavelet transform creates multiple mixed frequency domain spaces during the learning of fundus

image features. By integrating these mixed frequency domain spaces, it effectively captures high-frequency details of the fundus image, improving the quality of the generated fundus images. From the data, we can see that after removing the wavelet transform module, the PSNR decreased by 3.71 dB and the SSIM decreased by 0.033. (ii) We removed the CAMamba module from the MW module and experimented with stacking the CAMamba module. By combining the standard Mamba network structure, we found that stacking the CAMamba module twice achieved optimal metrics. We also experimented with stacking it three times, but this significantly increased the number of parameters, leading to gradient explosion. The Mamba module with the added channel attention mechanism better learns the global features of fundus images, optimizing the generation quality of these images. (iii) We removed the reverse diffusion branch. The removal of the reverse diffusion branch showed the most noticeable improvement in the quality of the generated fundus images. Instead of using the forward inference process of the diffusion model, which has a huge parameter count, we directly used the reverse generation module. By combining features learned from the preceding network, it generates higher-quality fundus images, showing significant improvement compared to using CNN alone without reverse diffusion. Additionally, to verify the advantages of our CAMamba module over the standard Mamba module, we conducted comparative experiments under the same conditions, and the results are illustrated in Table 5. Our CAMamba module showed clear advantages across all metrics. The experimental results show that a considerable performance reduction occurs when any component is absent. The significant increases in quantitative measurements provide additional evidence of each component's design's efficacy.

Table 4. Our MWRD ablation experiments on the eyeQ dataset [54].

Wavelet Transform	CAMamba	Reverse Diffusion	PSNR ↑	SSIM ↑	LPIPS ↓
×	×2	✓	28.07 dB	0.8986	0.1016
✓	×2	×	30.43 dB	0.9157	0.0667
×	×2	×	27.48 dB	0.7751	0.2015
✓	×	✓	29.16 dB	0.8872	0.1642
×	×	✓	25.48 dB	0.7049	0.2046
✓	×	×	26.94 dB	0.8318	0.1589
×	×1	✓	30.10 dB	0.9238	0.0770
✓	×2	✓	31.78 dB	0.9316	0.0601

Table 5. Comparative experiments between our MWRD and the standard Mamba on the eyeQ dataset [54].

Mamba	CAMamba	PSNR ↑	SSIM ↑	LPIPS ↓
✓	×	29.75 dB	0.9072	0.0815
×	✓	93.16 dB	0.9316	0.0601

5. Discussion

This study proposes a new fundus image enhancement method utilizing the improved Mamba model and the DDPM reverse diffusion process, significantly enhancing the details of low-quality fundus images, particularly in the recognition and enhancement of retinal vessels. Experimental results demonstrate that this method maintains high performance with fewer parameters, indicating that the technical innovations in this study effectively address challenges related to global quality imbalance and local detail blurring. Compared to existing low-light image enhancement methods, such as LLCaps and LLFlow, our approach shows clear advantages in quantitative metrics like PSNR and SSIM. Additionally, the improved Mamba model better preserves local details than traditional diffusion models, which is particularly important in medical image processing. The primary contribution of this study lies in the introduction of a new image enhancement framework, which not only has broad potential applications in low-light image processing but also provides clearer image support for clinical ophthalmic diagnosis. By effectively improving the quality

of fundus images, our method could enhance the early diagnosis rate of eye diseases. However, while the method performs excellently in enhancing low-quality images, certain limitations remain. For instance, the enhancement effect may diminish in cases of boundary blurring. Furthermore, the model's generalization capability requires improvement along with the need for further development in improving other types of medical image data. Future research may focus on optimizing the reverse diffusion process or integrating more multi-scale feature extraction techniques to tackle more complex image enhancement tasks. We also plan to apply this method to other types of medical images, such as CT and MRI, to explore its applicability in broader medical fields.

6. Conclusions

We introduced MWRD, an end-to-end fundus image enhancement framework composed of a CNN branch and a reverse diffusion process. The CNN branch consists of stacked MW modules, where the core wavelet transform and CAMamba module effectively extract high-frequency detail information and adequately supplement the learning of local feature details. The shallow output is further used for additional optimization through the reverse diffusion process, producing pictures that are most like the genuine thing. Studies comparing and ablation show that our approach and design result in a lightweight model while also bringing about notable performance gains in picture quality and a large reduction in model parameters. Additional research on medical picture segmentation has shown that this approach is feasible for use in clinical settings. Future research might involve applying our approach to a range of medical contexts, including the use of clinical deep learning models, surgical augmented reality, and robotic surgery.

Author Contributions: Conceptualization, T.C. and Y.L.; Methodology, T.C.; Validation, Y.L.; Writing—original draft, T.C.; Writing—review and editing, Y.L. and X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Bingtuan Science and Technology Program (2022DB005).

Data Availability Statement: Publicly available datasets were used in this study. The eyeQ dataset can be found here: https://github.com/hzfu/EyeQ?tab=readme-ov-file (accessed on 20 December 2023). The Kvasir-Capsule dataset can be found here: https://osf.io/dv2ag/ (accessed on 18 January 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Raj, A.; Tiwari, A.K.; Martini, M.G. Fundus image quality assessment: Survey, challenges, and future scope. *IET Image Process.* **2019**, *13*, 1211–1224. [CrossRef]
2. Wang, S.; Jin, K.; Lu, H.; Cheng, C.; Ye, J.; Qian, D. Human visual system-based fundus image quality assessment of portable fundus camera photographs. *IEEE Trans. Med. Imaging* **2015**, *35*, 1046–1055. [CrossRef] [PubMed]
3. Shome, S.K.; Vadali, S.R.K. Enhancement of diabetic retinopathy imagery using contrast limited adaptive histogram equalization. *Int. J. Comput. Sci. Inf. Technol.* **2011**, *2*, 2694–2699.
4. Cheng, J.; Li, Z.; Gu, Z.; Fu, H.; Wong, D.W.K.; Liu, J. Structure-preserving guided retinal image filtering and its application for optic disk analysis. *IEEE Trans. Med. Imaging* **2018**, *37*, 2536–2546. [CrossRef]
5. Dai, P.; Sheng, H.; Zhang, J.; Li, L.; Wu, J.; Fan, M. Retinal fundus image enhancement using the normalized convolution and noise removing. *Int. J. Biomed. Imaging* **2016**, *2016*, 5075612. [CrossRef] [PubMed]
6. Sahu, S.; Singh, A.K.; Ghrera, S.; Elhoseny, M. An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE. *Opt. Laser Technol.* **2019**, *110*, 87–98.
7. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep retinex decomposition for low-light enhancement. *arXiv* **2018**, arXiv:1808.04560.
8. Al Sobbahi, R.; Tekli, J. Comparing deep learning models for low-light natural scene image enhancement and their impact on object detection and classification: Overview, empirical evaluation, and challenges. *Signal Process. Image Commun.* **2022**, *109*, 116848. [CrossRef]
9. Song, Y.; Ermon, S. Generative modeling by estimating gradients of the data distribution. *arXiv* **2019**, arXiv:1907.05600.
10. Song, Y.; Sohl-Dickstein, J.; Kingma, D.P.; Kumar, A.; Ermon, S.; Poole, B. Score-based generative modeling through stochastic differential equations. *arXiv* **2020**, arXiv:2011.13456.
11. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840–6851.

12. Jiang, H.; Luo, A.; Fan, H.; Han, S.; Liu, S. Low-light image enhancement with wavelet-based diffusion models. *Acm Trans. Graph. (TOG)* **2023**, *42*, 1–14. [CrossRef]
13. Fu, D.Y.; Dao, T.; Saab, K.K.; Thomas, A.W.; Rudra, A.; Ré, C. Hungry hungry hippos: Towards language modeling with state space models. *arXiv* **2022**, arXiv:2212.14052.
14. Gu, A.; Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv* **2023**, arXiv:2312.00752.
15. Gu, A.; Goel, K.; Ré, C. Efficiently modeling long sequences with structured state spaces. *arXiv* **2021**, arXiv:2111.00396.
16. Mehta, H.; Gupta, A.; Cutkosky, A.; Neyshabur, B. Long range language modeling via gated state spaces. *arXiv* **2022**, arXiv:2206.13947.
17. Smith, J.T.; Warrington, A.; Linderman, S.W. Simplified state space layers for sequence modeling. *arXiv* **2022**, arXiv:2208.04933.
18. Vaswani, A. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762.
19. Zhu, L.; Liao, B.; Zhang, Q.; Wang, X.; Liu, W.; Wang, X. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv* **2024**, arXiv:2401.09417.
20. Ruan, J.; Xiang, S. Vm-unet: Vision mamba unet for medical image segmentation. *arXiv* **2024**, arXiv:2402.02491.
21. Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **2016**, *26*, 982–993. [CrossRef] [PubMed]
22. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *arXiv* **2014**, arXiv:1611.02163.
23. Li, M.; Liu, J.; Yang, W.; Sun, X.; Guo, Z. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Trans. Image Process.* **2018**, *27*, 2828–2841. [CrossRef] [PubMed]
24. Liu, Y.F.; Guo, J.M.; Yu, J.C. Contrast enhancement using stratified parametric-oriented histogram equalization. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *27*, 1171–1181. [CrossRef]
25. 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, Seattle, WA, USA, 12–16 October 2020; pp. 1478–1487.
26. Lee, K.G.; Song, S.J.; Lee, S.; Yu, H.G.; Kim, D.I.; Lee, K.M. A deep learning-based framework for retinal fundus image enhancement. *PLoS ONE* **2023**, *18*, e0282416. [CrossRef]
27. You, Q.; Wan, C.; Sun, J.; Shen, J.; Ye, H.; Yu, Q. Fundus image enhancement method based on CycleGAN. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 4500–4503.
28. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
29. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
30. Ma, Y.; Liu, J.; Liu, Y.; Fu, H.; Hu, Y.; Cheng, J.; Qi, H.; Wu, Y.; Zhang, J.; Zhao, Y. Structure and illumination constrained GAN for medical image enhancement. *IEEE Trans. Med. Imaging* **2021**, *40*, 3955–3967. [CrossRef]
31. Hu, D.; Tao, Y.K.; Oguz, I. Unsupervised denoising of retinal OCT with diffusion probabilistic model. In Proceedings of the Medical Imaging 2022: Image Processing, San Diego, CA, USA, 20–24 February 2022; Volume 12032, pp. 25–34.
32. Cheng, P.; Lin, L.; Huang, Y.; He, H.; Luo, W.; Tang, X. Learning enhancement from degradation: A diffusion model for fundus image enhancement. *arXiv* **2023**, arXiv:2303.04603.
33. Bai, L.; Chen, T.; Wu, Y.; Wang, A.; Islam, M.; Ren, H. Llcaps: Learning to illuminate low-light capsule endoscopy with curved wavelet attention and reverse diffusion. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Vancouver, BC, Canada, 8–12 October 2023; pp. 34–44.
34. Gu, A.; Johnson, I.; Goel, K.; Saab, K.; Dao, T.; Rudra, A.; Ré, C. Combining recurrent, convolutional, and continuous-time models with linear state space layers. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 572–585.
35. Gu, A.; Dao, T.; Ermon, S.; Rudra, A.; Ré, C. Hippo: Recurrent memory with optimal polynomial projections. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1474–1487.
36. Hasani, R.; Lechner, M.; Wang, T.H.; Chahine, M.; Amini, A.; Rus, D. Liquid structural state-space models. *arXiv* **2022**, arXiv:2209.12951.
37. Huang, T.; Pei, X.; You, S.; Wang, F.; Qian, C.; Xu, C. Localmamba: Visual state space model with windowed selective scan. *arXiv* **2024**, arXiv:2403.09338.
38. Wang, J.; Zhu, W.; Wang, P.; Yu, X.; Liu, L.; Omar, M.; Hamid, R. Selective structured state-spaces for long-form video understanding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 6387–6397.
39. Islam, M.M.; Hasan, M.; Athrey, K.S.; Braskich, T.; Bertasius, G. Efficient movie scene detection using state-space transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 18749–18758.
40. Ma, J.; Li, F.; Wang, B. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv* **2024**, arXiv:2401.04722.

41. Nguyen, E.; Goel, K.; Gu, A.; Downs, G.; Shah, P.; Dao, T.; Baccus, S.; Ré, C. S4nd: Modeling images and videos as multidimensional signals with state spaces. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 2846–2861.
42. Guo, H.; Li, J.; Dai, T.; Ouyang, Z.; Ren, X.; Xia, S.T. Mambair: A simple baseline for image restoration with state-space model. *arXiv* **2024**, arXiv:2402.15648.
43. Luo, Z.; Gustafsson, F.K.; Zhao, Z.; Sjölund, J.; Schön, T.B. Image restoration with mean-reverting stochastic differential equations. *arXiv* **2023**, arXiv:2301.11699.
44. Lugmayr, A.; Danelljan, M.; Romero, A.; Yu, F.; Timofte, R.; Van Gool, L. Repaint: Inpainting using denoising diffusion probabilistic models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11461–11471.
45. Ren, M.; Delbracio, M.; Talebi, H.; Gerig, G.; Milanfar, P. Multiscale structure guided diffusion for image deblurring. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–3 October 2023; pp. 10721–10733.
46. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
47. Elfwing, S.; Uchibe, E.; Doya, K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Netw.* **2018**, *107*, 3–11. [CrossRef]
48. Zhu, X.; Cheng, D.; Zhang, Z.; Lin, S.; Dai, J. An empirical study of spatial attention mechanisms in deep networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6688–6697.
49. Pandey, K.; Mukherjee, A.; Rai, P.; Kumar, A. Diffusevae: Efficient, controllable and high-fidelity generation from low-dimensional latents. *arXiv* **2022**, arXiv:2201.00308.
50. Wu, J.; Ji, W.; Fu, H.; Xu, M.; Jin, Y.; Xu, Y. Medsegdiff-v2: Diffusion-based medical image segmentation with transformer. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 6030–6038.
51. Wang, Y.; Wan, R.; Yang, W.; Li, H.; Chau, L.P.; Kot, A. Low-light image enhancement with normalizing flow. In Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 27 February—2 March 2022; Volume 36, pp. 2604–2612.
52. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Learning enriched features for real image restoration and enhancement. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part XXV 16; Springer: Berlin/Heidelberg, Germany, 2020; pp. 492–511.
53. Charbonnier, P.; Blanc-Feraud, L.; Aubert, G.; Barlaud, M. Two deterministic half-quadratic regularization algorithms for computed imaging. In Proceedings of the 1st International Conference on Image Processing, Austin, TX, USA, 13–16 November 1994; Volume 2, pp. 168–172.
54. Fu, H.; Wang, B.; Shen, J.; Cui, S.; Xu, Y.; Liu, J.; Shao, L. Evaluation of retinal image quality assessment networks in different color-spaces. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, 13–17 October 2019; Proceedings, Part I 22; Springer: Berlin/Heidelberg, Germany, 2019; pp. 48–56.
55. Li, C.; Guo, C.; Han, L.; Jiang, J.; Cheng, M.M.; Gu, J.; Loy, C.C. Low-light image and video enhancement using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 9396–9416. [CrossRef]
56. Lore, K.G.; Akintayo, A.; Sarkar, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **2017**, *61*, 650–662. [CrossRef]
57. Huynh-Thu, Q.; Ghanbari, M. Scope of validity of PSNR in image/video quality assessment. *Electron. Lett.* **2008**, *44*, 800–801. [CrossRef]
58. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
59. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 586–595.
60. Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. Enlightengan: Deep light enhancement without paired supervision. *IEEE Trans. Image Process.* **2021**, *30*, 2340–2349. [CrossRef] [PubMed]
61. Smedsrud, P.H.; Thambawita, V.; Hicks, S.A.; Gjestang, H.; Nedrejord, O.O.; Næss, E.; Borgli, H.; Jha, D.; Berstad, T.J.D.; Eskeland, S.L.; et al. Kvasir-Capsule, a video capsule endoscopy dataset. *Sci. Data* **2021**, *8*, 142. [CrossRef] [PubMed]
62. Li, L.; Verma, M.; Nakashima, Y.; Nagahara, H.; Kawasaki, R. Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Village, CO, USA, 1–5 March 2020; pp. 3656–3665.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.