

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2025)01-0066-17

论文引用格式: Huang Y, Cheng B, Fang S J and Liu X. 2025. Shadow removal with wavelet-based non-uniform diffusion model. Journal of Image and Graphics, 30(01):0066-0082(黄颖, 程彬, 房少杰, 刘歆. 2025. 用于阴影去除的小波非均匀扩散模型. 中国图象图形学报, 30(01):0066-0082)[DOI:10.11834/jig.230904]

用于阴影去除的小波非均匀扩散模型

黄颖^{1,2}, 程彬^{1*}, 房少杰², 刘歆²

1. 重庆邮电大学计算机科学与技术学院, 重庆 400065; 2. 重庆邮电大学软件工程学院, 重庆 400065

摘要: 目的 现有的阴影去除方法通常依赖于像素级重建, 旨在学习阴影图像和无阴影图像之间的确定性映射关系。然而阴影去除关注阴影区域的局部恢复, 容易导致在去除阴影的同时破坏非阴影区域。此外, 现有的大多数扩散模型在恢复图像时存在耗时过长和对分辨率敏感等问题。为此, 提出了一种用于阴影去除的小波非均匀扩散模型。**方法** 首先将图像通过小波分解为低频分量与高频分量, 然后针对低频和高频分量分别设计扩散生成网络来重建无阴影图像的小波域分布, 并分别恢复这些分量中的各种退化信息, 如低频(颜色、亮度)和高频细节等。**结果** 实验在3个阴影数据集上进行训练和测试, 在SRD(shadow removal dataset)数据集中, 与9种代表性方法进行比较, 在非阴影区域和整幅图像上, 峰值信噪比(peak signal-to-noise ratio, PSNR)、结构相似性(structural similarity index, SSIM)和均方根误差(root mean square error, RMSE)均取得最优或次优的结果; 在ISTD+(augmented dataset with image shadow triplets)数据集中, 与6种代表性方法进行比较, 在非阴影区域上, 性能取得了最佳, PSNR和RMSE分别提高了0.47 dB和0.1。除此之外, 在SRD数据集上, ShadowDiffusion方法在生成不同分辨率图像时性能有明显差异, 而本文方法性能基本保持稳定。此外, 本文方法生成速度与其相比提高了约4倍。**结论** 提出的方法能够加快扩散模型的采样速度, 在去除阴影的同时, 恢复出阴影区域缺失的颜色、亮度和丰富的细节等信息。

关键词: 阴影去除; 扩散模型(DM); 小波变换; 双分支网络; 噪声调度表

Shadow removal with wavelet-based non-uniform diffusion model

Huang Ying^{1,2}, Cheng Bin^{1*}, Fang Shaojie², Liu Xin²

1. School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;

2. School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract: **Objective** Shadows are a common occurrence in optical images captured under partial or complete obstruction of light. In such images, shadow regions typically exhibit various forms of degradation, such as low contrast, color distortion, and loss of scene structure. Shadows not only impact the visual perception of humans but also impose constraints on the implementation of numerous sophisticated computer vision algorithms. Shadow removal can assist in many computer vision tasks. It aims to enhance the visibility of shadow regions in images and achieve consistent illumination distribution between shadow and non-shadow regions. Currently, deep learning-based shadow removal methods can be roughly divided into two categories. One typically utilizes deep learning to minimize the pixel-level differences between shadow regions and their corresponding non-shadow regions, aiming to learn deterministic mapping relationships between shadow and non-shadow images. However, the primary focus of shadow removal lies in locally restoring shadow regions, often overlooking

收稿日期: 2024-01-08; 修回日期: 2024-04-29; 预印本日期: 2024-05-06

* 通信作者: 程彬 2606301180@qq.com

基金项目: 国家自然科学基金项目(62471076)

Supported by: National Natural Science Foundation of China (62471076)

the essential constraints required for effectively restoring boundaries between shadow and non-shadow regions. As a result, discrepancies in brightness exist between the restored shadow and non-shadow areas, along with the emergence of artifacts along the boundaries. Another approach involves using image generation models to directly model the complex distribution of shadow-free images, avoiding the direct learning of pixel-level mapping relationships, and treating shadow removal as a conditional generation task. While diffusion models have garnered significant attention due to their powerful generation capabilities, most existing diffusion generation models suffer from issues such as time-consuming image restoration and sensitivity to resolution when recovering images. Inspired by these challenges, a wavelet non-uniform diffusion model (WNDM) is proposed, which combines the advantages of wavelet decomposition and the generation ability of diffusion models to solve the above problems. **Method** First, the image is decomposed into low-frequency and high-frequency components via wavelet decomposition. Then, diffusion generation networks are designed separately for low-frequency and high-frequency components to reconstruct the wavelet domain distribution of shadow-free images and restore various degraded information within these components, such as low-frequency (color, brightness) and high-frequency details. Wavelet transform can decompose the image into high-frequency and low-frequency images without sacrificing information, and the spatial size of the decomposed images is halved. Thus, modeling diffusion in the wavelet domain not only greatly accelerates model inference but also captures information that may be lost in the pixel domain. Furthermore, considering the complexity of the distribution of low-frequency and high-frequency components and their sensitivity to noise, for example, high-frequency components exhibit sparsity, making it easier for neural networks to learn their features. Hence, this study devises two separate adaptive diffusion noise scheduling tables tailored for low-frequency and high-frequency components. The branch for low-frequency diffusion adjustment independently fine-tunes the low-frequency information within shadow images, whereas the branch for high-frequency diffusion adjustment independently refines the high-frequency information within shadow images, resulting in the generation of more precise low-frequency and high-frequency images, respectively. Additionally, the low-frequency and high-frequency diffusion adjustment branches are consolidated to share a denoising network, thus streamlining model complexity and optimizing computational resources. The difference lies in the design of two prediction branches in the final layer of this network. These branches consist of several stacked convolution blocks, each predicting the low-frequency and high-frequency components of the shadow-free image, respectively. Finally, high-quality shadow-free images are reconstructed using inverse wavelet transform. **Result** The experiments were conducted on three shadow removal datasets for training and testing. On the shadow removal dataset (SRD) dataset, comparisons were made with nine state-of-the-art shadow removal algorithms. The proposed model achieved the best or second-best results in terms of peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and root mean square error (RMSE) in both non-shadow regions and the entire image. On the dataset with image shadow triplets dataset (ISTD), the performance was the best in non-shadow regions, with improvements of 0.36 dB in PSNR, 0.004 in SSIM, and 0.04 in RMSE compared with the second-best model. It ranked second in performance across all metrics for the entire image. On the augmented dataset with image shadow triplets (ISTD+), compared with six state-of-the-art shadow removal algorithms, the performance was the best in non-shadow regions, with improvements of 0.47 dB in PSNR and 0.1 in RMSE. Additionally, regarding the advanced shadow removal diffusion model ShadowDiffusion, the RMSE for the entire image was 3.63 on the SRD dataset when generating images of 256×256 pixels resolution. However, a significant performance drop occurred when generating images of the original resolution of 840×640 pixels, with RMSE increasing to 7.19. By contrast, our approach yielded RMSE values of 3.80 and 4.06 for images of dimensions 256×256 pixels and 840×640 pixels, respectively, showcasing consistent performance. Additionally, the time required to generate a single original image of size 840×640 pixels was reduced by roughly fourfold compared with ShadowDiffusion. Furthermore, our method was expanded to address image raindrop removal tasks, delivering competitive results on the RainDrop dataset. **Conclusion** In this paper, the proposed method accelerates the sampling time of the diffusion model. While removing shadows, it restores missing color, brightness, and rich details in shadow regions. It treats shadow removal as an image generation task in the wavelet domain and designs two adaptive diffusion flows for the low-frequency and high-frequency components of the image wavelet domain to address the degradation of low-frequency (color, brightness) and high-frequency detail information caused by shadow images. Benefiting from the frequency decomposition of the wavelet transform, WNDM does not learn from the

entangled pixel space domain but effectively separates and trains them separately, thereby generating more refined low-frequency and high-frequency information for reconstructing the final image. Extensive experiments on multiple datasets demonstrate the effectiveness of WNDM, achieving competitive results compared with state-of-the-art methods.

Key words: shadow removal; diffusion model(DM); wavelet transform; dual-branch network; noise schedule

0 引言

阴影是在部分或完全被光遮挡的情况下拍摄光学图像时普遍存在的现象。这类图像中的阴影区域通常会产生各种退化,如低对比度、色彩失真和细节丢失等。阴影图像不仅对人类的视觉体验产生影响,还限制了许多高级计算机视觉算法的应用,如目标识别(Long和Hua, 2017)、目标检测和跟踪(Mikic等, 2000)等。

阴影去除旨在提高图像阴影区域的可见性,追求阴影和非阴影区域之间的一致照明分布。目前的阴影去除方法大致分为两种:基于物理先验的传统方法和基于深度学习的方法。早期的阴影去除方法(Finlayson等, 2006; Liu和Gleicher, 2008; 唐磊等, 2008)依赖于理想化制作的手工先验信息,无法有效地处理复杂真实场景中的阴影(Khan等, 2016)。近年来,随着大规模阴影去除数据集的公布和深度学习的兴起,这些问题已经得到了部分解决。基于深度学习的方法直接学习退化的阴影图像和相应的高质量无阴影图像之间的映射关系,从而比传统方法更具鲁棒性。然而,这些方法难以逃脱拟合的本质,通常深度学习用于最小化阴影区域与其对应的非阴影区域之间的像素级差异,旨在学习阴影图像和非阴影图像之间的确定性映射关系。但阴影去除关注阴影区域的局部恢复,阴影区域和非阴影区域的边界部分缺乏必要的约束,从而导致恢复的图像阴影区域与非阴影区域亮度不一致、细节丢失以及边界伪影问题。为了解决这个问题,图像生成模型被用来直接对无阴影图像的复杂分布进行建模,从而避免直接学习阴影区域到无阴影区域的映射关系。考虑到阴影去除是在阴影图像引导下恢复无阴影图像的过程,因此可将阴影去除视为具有条件信息的图像生成任务。

目前,深度生成模型如生成对抗网络(generative adversarial network, GAN)(Wang等, 2018)在阴影去除任务中展现出良好的效果,这些方法致力于

捕捉输入的多模态分布,以生成在感知上更为真实的输出图像。生成对抗网络通过对抗性机制同时训练生成器和鉴别器,因此GAN存在训练不稳定性,容易导致模型崩溃、不收敛以及梯度消失等问题。此外,由于GAN属于隐式生成模型,基于GAN的阴影去除方法通常会通过训练后直接拟合生成无阴影图像,从而导致可能产生额外的内容和伪影。近期,扩散模型(diffusion model, DM)(Sohl-Dickstein等, 2015; Ho等, 2020)因其在图像生成(Song等, 2021)和恢复任务(Whang等, 2022; Li等, 2022)中优异的性能而备受关注。DM依赖于去噪编码器的分层架构,使其能够迭代地进行反转扩散过程,并实现从随机采样的高斯噪声到目标图像或潜在分布的高质量映射,而不会遇到先前生成模型中存在的稳定性和模型崩溃问题。然而,由于扩散模型生成图像时需要多次迭代网络,这些基于像素空间建模的扩散模型方法通常无法避免采样时间长、模型性能对图像分辨率较为敏感等问题。而小波变换可以将图像分解为低频分量和高频分量,低频和高频分量初始尺寸的减少提高了网络的处理速度。因此,将扩散模型与小波变换相结合可以加快模型的采样速度。除此之外,对小波域扩散建模可以避免从信息纠缠的像素空间域中学习,使模型能够捕捉到像素空间可能丢失或失真的信息,从而提高图像重建质量。

本文为阴影去除任务设计了一种基于小波域的非均匀扩散模型,该模型将阴影去除问题定义为小波分解和条件图像生成。具体来说,首先使用 K 次2D离散小波变换将图像转换到小波域,该变换将每个输入分解为代表全局信息的低频分量和表示稀疏的高频分量。与目前大多数专注于空间域的扩散模型不同,该模型对无阴影图像小波分解后的低频分量和高频分量分别建模。考虑到低频分量和高频分量在扩散过程中对噪声的敏感度不一样,对低频分量和高频分量采用不同的扩散速度进行扩散,分别解决阴影图像低频和高频分量中的各种退化,包括暗照明、颜色偏差、场景内容丢失等问题。

本文主要贡献如下:1)提出一种基于小波的非

均匀扩散模型(wavelet non-uniform diffusion model, WNDM),该模型利用扩散模型的生成能力和小波变换的优势来实现稳健高效的阴影去除;2)发现同量级噪声对不同频率信息的影响有着显著的区别,并对低频和高频分量分别制定了不同的噪声调度,以适应不同频率信息的扩散生成;3)实验结果表明,WNDM更好地生成小波域中的低频和高频信息,并对其重建后的质量有一定程度的提高。大量实验表明了该方法的有效性。

1 相关工作

1.1 阴影去除

阴影去除早期的工作主要通过各种手工制作先验,例如图像梯度、照明信息和区域信息,以增强阴影区域的照明。例如,Liu和Gleicher(2008)通过消除阴影光照变化引起的梯度,并对纹理特征进行采样实现阴影去除。然而这些方法都是在理想的假设下构建的,在转移到真实复杂场景下会导致明显的性能退化。

近年来,利用从庞大数据中学习映射的强大能力,基于深度学习的方法取得了优异的性能。例如,BMNet(bijective mapping network)(Zhu等,2022a)利用可逆生成网络通过阴影生成的反向映射进行阴影去除。ShadowFormer(Transformer for shadow removal)(Guo等,2023a)基于Transformer利用非阴影区域信息帮助阴影区域恢复。尽管这些方法展现出不错的性能,但本质上仍然学习的是阴影图像到无阴影图像之间的确定性映射关系,通常会带来颜色恢复不正确、产生边界伪影以及模型泛化性不强等问题。

随着生成对抗网络、流模型(Lugmayr等,2020)和其他生成模型的发展,图像生成模型可以产生高质量的结果。ST-CGAN(stacked conditional generative adversarial network)(Wang等,2018)通过将两个条件生成对抗网络组合起来同时学习优化阴影检测和去除任务。Hu等人(2019)采用未配对数据对模型进行训练,并使用对抗性学习和循环一致性约束思想学习阴影图像和无阴影图像之间的潜在关系。然而,由于GAN独特的对抗性训练,这些方法极易导致模型训练不稳定等问题。最近,基于扩散模型的图像生成方法(Song等,2021)研究产生了优异的

结果。ShadowDiffusion(Guo等,2023b)基于扩散模型通过阴影掩膜迭代细化并结合生成先验的信息来实现阴影去除。尽管与基于GAN和流模型的生成方法相比,基于扩散模型的方法可以获得更高质量的图像,但由于在生成图像过程中需多次迭代网络,通常会需要更多的计算资源和推理时间。本文提出将扩散模型与小波变换相结合,利用小波分解的优势加速模型的推理。

1.2 扩散模型

扩散模型受到热力学的启发,可以通过逐步添加高斯噪声的方式在高斯分布与目标数据分布之间构建一条马尔可夫链。由于其对数据分布建模的优异能力,广泛应用于众多图像生成和恢复任务。受条件生成模型的启发,Saharia等人(2022)提出一种通用的图像转换框架,以解决具有条件扩散概率模型的图像恢复问题。Yi等人(2023)将Retinex分解模型和扩散模型相结合,通过扩散模型增强照明图和反射率图来实现低照度图像增强。Phung等人(2023)和Jiang等人(2023)通过结合小波变换的优势,以解决在各自任务中扩散模型采样时间长等问题。

在扩散模型的前向过程中需要设定一个噪声调度表来指导每一步加噪声的量级,因此许多工作对此展开了研究。IDDPM(improved denoising diffusion probabilistic model)(Nichol和Dhariwal,2021)探究了在扩散过程中的多种噪声调度设置方式。Yue等人(2023)在图像超分辨率任务中制定了一个详细的噪声调度表,通过多参数的选择,灵活地控制扩散过程中的移动速度和噪声强度。Chen(2023)强调噪声调度表的重要性,发现同量级噪声对不同分辨率图像的影响有着显著的区别,提出在对不同分辨率图像进行扩散建模时,应采用不同的噪声调度表。与之前工作不同,本文发现在小波域中,相同噪声对不同频率信息的影响有着显著的区别,因此本文对低频和高频分量分别制定了不同的噪声调度,以适应不同频率信息的扩散生成。

2 方法

本文提出的WNDM方法框架如图1所示,该方法充分结合了扩散模型的生成能力和小波变换的优势。由于小波变换可以在不牺牲信息的前提下无损

地将图像分解成高频图像和低频图像,分解后的图像空间大小减半,因此对小波频域进行扩散建模不仅可以极大地提升模型效率还可以捕捉到图像像素域中可能丢失的信息。此外,高频分量具有稀疏性,神经网络往往更加容易学习其特征,考虑到低频分量分布和高频分量分布的复杂程度以及对噪声的敏

感度不同,本文为低频分量和高频分量设计了两条自适应的扩散噪声调度表。低频扩散调整分支单独调整阴影图像中的低频信息,高频扩散调整分支单独调整阴影图像中的高频信息,从而各自生成更加精细化的低频和高频图像,最后利用逆小波变换重建出高质量的无阴影图像。

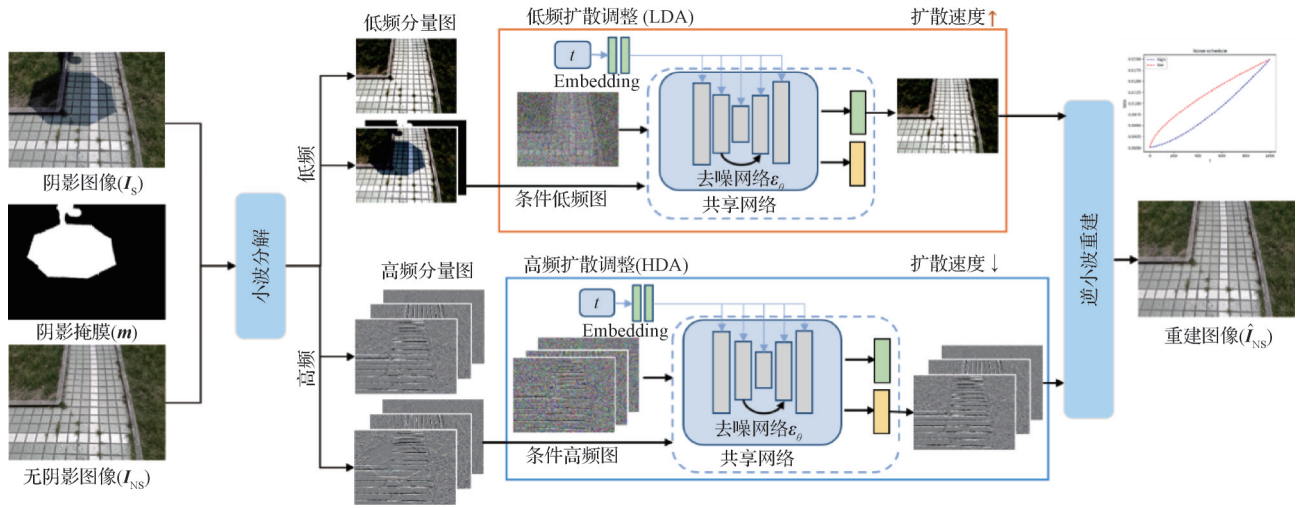


图1 本文提出的WNDM方法结构示意图

Fig. 1 Diagram of the proposed WNDM structure

2.1 离散小波输入模块

本文将阴影去除任务视为图像生成任务,而生成模型生成一幅高质量的无阴影图像不仅需要生成出精确的全局信息(光照、颜色等),还需要生成出精细的局部细节信息(纹理等)。若基于图像原始像素空间直接建模,会导致模型生成的内容顾此失彼,生成图像中的全局信息和细节信息不能很好均衡,最终生成不正确的颜色和亮度信息。因此,本文期望构建两条扩散生成流,分别对全局信息和细节信息进行精细化的建模生成,而离散小波变换可以无损地将图像分解出低频全局信息和高频细节信息,这为实现该方法提供了充分的可行性。

哈尔小波(Haar, 1911)由于其简单性在实际应用中被广泛采用,它涉及两种类型的操作:离散小波变换(discrete wavelet transform, DWT)和离散小波逆变换(inverse wavelet transform, IWT),本文采用此方法将图像从像素空间域转换到小波频域之中。给定一幅阴影图像 $I_s \in \mathbf{R}^{H \times W \times C}$, 将其转换到小波域中, 具体为

$$L_s^1, \{V_s^1, H_s^1, D_s^1\} = DWT(I_s) \quad (1)$$

式中, $L_s^1, V_s^1, H_s^1, D_s^1 \in \mathbf{R}^{\frac{H}{2} \times \frac{W}{2} \times C}$, 分别表示阴影图像经1次小波分解得到的低频分量以及在垂直、水平和对角线上的高频分量。同理,通过逆小波变换重建出图像的计算为

$$I_s = IWT(L_s^1, V_s^1, H_s^1, D_s^1) \quad (2)$$

由于小波一次分解后的图像空间维度是上一层图像的1/4,若小波分解K次后,则空间维度变为原始空间的 $(1/4)^K$,因此对小波域中的分量进行扩散建模可以显著减少扩散模型的推理时间和计算资源。本文将阴影图像 I_s 、阴影掩膜图像 m 、无阴影图像 I_{ns} 通过式(1)小波分解后的低频分量和高频分量分别送入到低频扩散流和高频扩散流中进行扩散建模。

2.2 小波域中的双分支扩散生成

如图1所示,本文利用扩散模型分别对无阴影图像的低频和高频分量的分布进行建模,而阴影图像的低频和高频分量分别作为扩散模型的条件引导信息。总的来说,该模型可以分为两条流,即低频扩散调整流(low-frequency diffusion adjustment, LDA)和高频扩散调整流(high-frequency diffusion adjust-

ment, HDA)。

本文采用去噪扩散概率模型(denoising diffusion probabilistic model, DDPM)(Ho等, 2020)对低频分量和高频分量的分布进行建模, 它可以被描述为正向扩散和逆向去噪两个过程。其建模原理是通过在前向过程中逐步添加高斯噪声来破坏训练数据, 然后在逆向过程中通过去除添加的噪声过程来学习如何恢复数据, 图2展示了扩散模型的正向扩散和逆向去噪过程。在扩散过程中, 将无阴影图像分解的低频和低频分量表示为 $I_0 \in \mathbf{R}^{H \times W \times C}$ (在 LDA 中 $C=3$, 在 HDA 中 $C=9$)。将阴影图像和阴影掩膜分解的低频和低频图像表示为 I_c , 并分别与低频扩散分支和高频扩散分支中的噪声图像通道连接作为条件信息引导扩散模型的去噪生成。

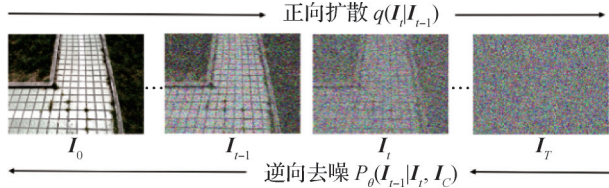


图2 扩散框架

Fig. 2 Framework of diffusion

2.2.1 正向扩散过程

正向扩散过程可以看做是一个马尔可夫链, 它将高斯噪声逐步添加到数据中, 步骤 t 时刻的数据状态仅取决于 $t-1$ 时刻的数据。因此, 任意 $t \in [0, T]$, 可以得到噪声图像的数据分布, 具体为

$$q(I_t|I_0) = N(I_t; \sqrt{\bar{a}_t} I_0, (1 - \bar{a}_t) I) \quad (3)$$

式中, $\bar{a}_t = \prod_{i=0}^t \alpha_i$, $\alpha_i = 1 - \beta_i$, β_i 是预定义的噪声调度表, I_0 和 I_t 分别表示未加噪的初始图像和加 t 步噪声后的图像。当第 t 步的分布 $q(I_t|I_0)$ 接近正态分布 $N(0, I)$ 时, 则认为完成了扩散的正向过程。

2.2.2 逆向去噪过程

逆向去噪过程是从纯噪声的高斯分布中恢复原始分布的过程。与正向扩散过程类似, 逆向去噪过程也是逐步进行的。在第 t 步时, 对数据 I_t 执行去噪操作是以在条件图像 I_c 的指导下得到 I_{t-1} 的概率分布过程。因此, 给定 I_t 可以将条件概率分布 I_{t-1} 公式化为

$$p_\theta(I_{t-1}|I_t, I_c) = N(I_{t-1}; \mu_\theta(I_t, I_c, t), \sigma_t^2 I) \quad (4)$$

式中, $\mu_\theta(I_t, I_c, t)$ 是第 t 步估计去除噪声的均值, σ_t^2 是方差, 本文在 LDA 和 HDA 中遵循 DDPM(Ho等, 2020)的设置将方差设置为固定值, 因此只需要训练一个神经网络预测式(4)中分布的均值 μ_θ , 即可完成扩散模型的逆向去噪过程。又根据 DDPM, 预测 μ_θ 可以优化成预测噪声 ϵ_θ 。

在逆向去噪过程的每一步 ($t \in [0, T]$) 中, 对网络预测的噪声 ϵ_θ 与正向扩散过程中实际添加的噪声 $\epsilon \sim N(0, I)$ 的均方根误差函数进行优化。因此, 逆向去噪过程的损失函数为

$$L_{DF} = \left\| \epsilon - \epsilon_\theta(\sqrt{\bar{a}_t} I_0 + \sqrt{1 - \bar{a}_t} \epsilon, I_c, t) \right\|_F^2 \quad (5)$$

当网络训练完成后, 可从高斯噪声图像 $I_T \sim N(0, I)$ 迭代网络去噪。为了减少模型的采样时间, 本文选择 DDIM(denoising diffusion implicit model)(Song等, 2021)作为扩散模型的采样策略。它与 DDPM 有相同的优化目标函数, 但却可以在几乎不牺牲图像质量的前提下进行跳跃式去噪。

当前逆向去噪过程中的去噪网络通常使用结合注意力模块的 U-Net(Ronneberger等, 2015)结构。在 LDA 和 HDA 中, 本文设计的去噪网络借鉴了 SR3 模型(Saharia等, 2023)的主干网络, 并遵循扩散去噪网络的设计。由于传统的 U-Net 网络中的下采样和上采样方法通常采用插值或池化, 这种方法会导致信息的丢失。小波变换可以无损地将图像下采样, 分解出低频和高频分量, 因此本文利用该特性, 将去噪网络的下采样使用小波分解, 上采样使用小波逆变换, 使网络可以较好地地区分低频和高频信息。

除此之外, 为了减少模型复杂度和计算资源, 本文将 LDA 和 HDA 共享去噪网络, 不同之处是在该网络的最后一层设计了两个预测分支。该分支由几个叠加的卷积块组成, 分别预测低频分量的噪声 ϵ_θ^L 和高频分量的噪声 ϵ_θ^H 。因此, 整个扩散模型网络的训练总目标函数为

$$L_T = L_{DF}^L + \lambda L_{DF}^H \quad (6)$$

式中, L_{DF}^L 是低频分量扩散损失函数, L_{DF}^H 是高频分量扩散损失函数, λ 是平衡低频扩散分支和高频扩散分支的权重因子。

综上所述, 整个扩散生成过程是从阴影图像的小波频域恢复无阴影图像小波频域的过程。

2.3 基于小波域的非均匀扩散

对于扩散模型来说, 扩散过程中的噪声调度表

至关重要,它影响着模型生成能力的好坏,许多工作(Nichol和Dhariwal, 2021; Chen, 2023)都研究了不同的噪声调度表对于扩散模型性能的影响。在扩散的前向过程中,加噪声实际上起到一种掩盖信息的作用,若每一步添加的噪声太多,则会增加网络学习去噪的难度,在逆向过程中,网络就很难估计出每步应该去除的噪声;若每一步添加的噪声太少,则在加噪 T 步后,图像无法转换成高斯噪声,导致无法对数据的分布精确建模,极大地影响模型的性能。

根据图3可以发现,相同噪声对图像的低频和低频信息有着不同的影响, L_0 是小波分解后低频全局信息, V_0, H_0, D_0 是高频细节信息,相比于低频信息,噪声对高频信息的影响更大,即掩盖的信息更多。由于本文的目标是构建小波域的双流扩散模型,该模型能够对低频信息和高频信息进行精细化的生成,因此在模型的低频扩散分支和低频扩散分支上理应有不同的扩散速度。

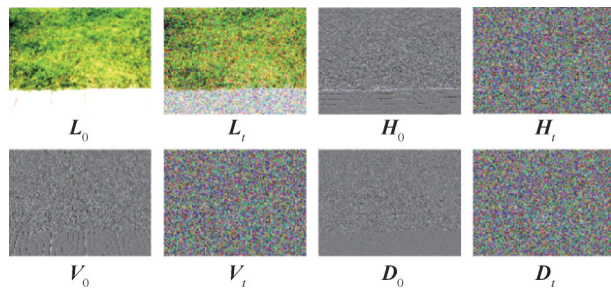


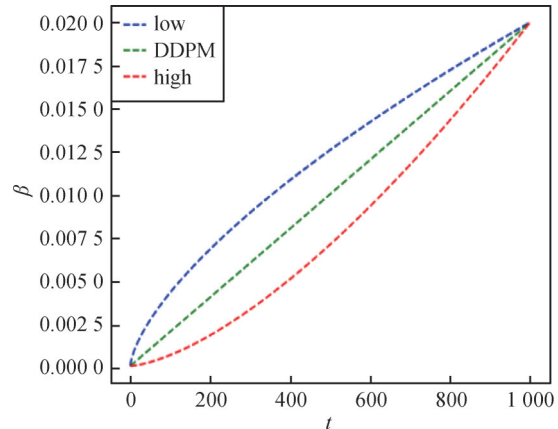
图3 对小波变换分解出的低频和高频分量添加相同的噪声得到的可视化结果

Fig. 3 Visualization results obtained by adding the same noise to the low-frequency and high-frequency components decomposed by wavelet transform

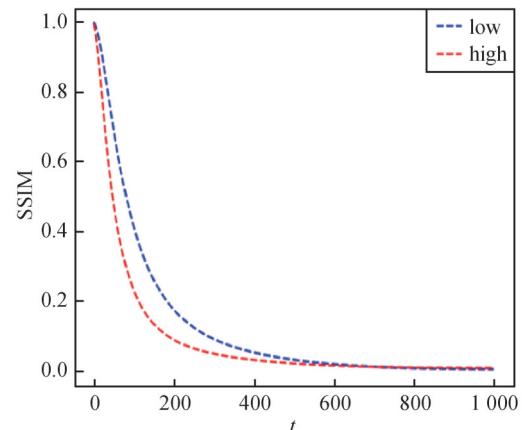
扩散模型控制正向扩散速度通常是根据噪声调度表实现的。当前大多数扩散模型都遵循DDPM中的先验设定,将噪声调度表 $\beta_t \in (0.0001, 0.02)$ 进行1000步的线性采样,如图4(a)所示。

将无阴影图像小波分解后的低频和高频图像都应用该时间表,得到扩散过程中各分量图像与原始干净图像的结构相似性(structural similarity index, SSIM)值,如图4(b)所示。SSIM是衡量两幅图像之间结构相似度的指标,更能体现噪声对图像的整体破坏程度。由此可知,在整个扩散过程中,高频图像在同一时刻 t 下往往会比低频图像掩盖更多的信息。当 t 足够大时,由于低频和高频图像添加的高斯噪声足够

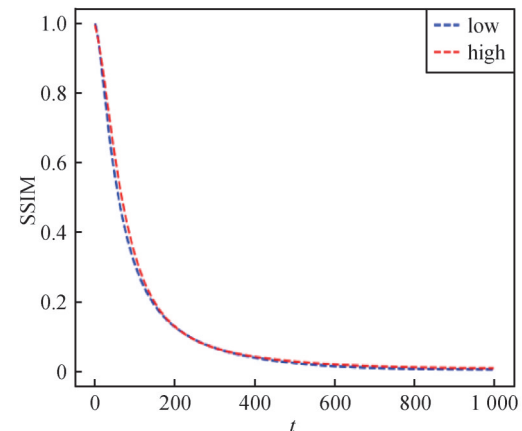
多,已经变成了满足正态分布的高斯噪声图像,所以之后无论添加多少高斯噪声,它们的SSIM值都会趋于接近。



(a) DDPM的扩散系数以及本文改进后的低频和高频扩散系数



(b) 低频和高频分量以相同扩散系数在 T 步加噪过程中各自噪声图像与原始干净图像的SSIM值



(c) 改进后非均匀扩散系数在 T 步加噪过程中各自噪声图像与原始干净图像的SSIM值

图4 噪声调度表的图解

Fig. 4 Illustration of the proposed noise schedule ((a) diffusion coefficient of DDPM and ours; (b) SSIM values of noisy image and clean image for low-frequency and high-frequency components; (c) SSIM values of non-uniform diffusion between noisy image and clean image)

根据2.2节中介绍的基于小波的扩散生成过程中, LDA和HDA共享去噪网络, 因此为了简化网络的学习难度, 期望在每个步骤 t 处, 无论对低频还是高频信息网络应该去除相接近的噪声影响。为此本文重新设计了低频和高频扩散分支的噪声调度表, 它能让在整个扩散过程中的每个步骤 t 处, 噪声对低频和高频信息的影响是接近的, 如图4(c)所示。

通过低频和高频分支的非均匀扩散, 定义了分别适应自身的扩散速度, 导致模型能够在低频分量和高频分量上生成更加精细化的结果。具体方法为

$$\beta_i^L = N(\beta_i)^{\gamma_L}, \quad \beta_i^H = N(\beta_i)^{\gamma_H} \quad (7)$$

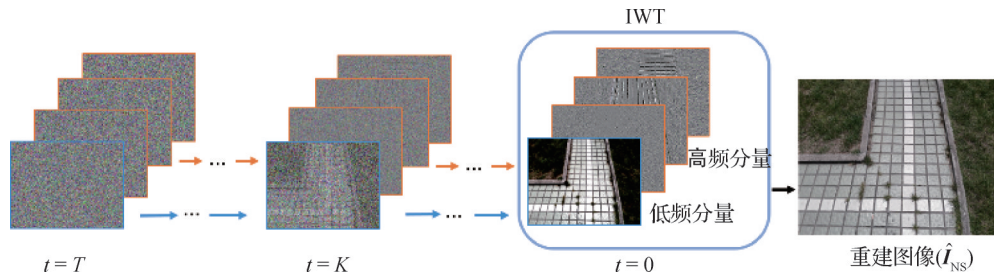


图5 WNDM生成图像过程

Fig. 5 Image generation process of WNDM model

3 实验结果及分析

3.1 实验设置

3.1.1 训练细节

本文所有的实验均采用PyTorch1.8.1框架和单个NVIDIA RTX A5000 GPU实现。在训练阶段主要有两种类型数据增强。1) 首先将图像缩减为 286×286 像素, 再从 286×286 像素的图像随机裁剪出 256×256 像素部分; 2) 从水平和垂直两个方向进行随机翻转。训练期间, 使用Adam优化器优化所提出的网络, 其中, 第1阶和第2阶动量分别设置为0.9和0.999, 网络初始学习率为 3×10^{-5} , 并在整个训练阶段保持恒定, 批处理大小为8, 训练迭代轮次(epoch)设置为1500。扩散模型的扩散步数 T 设为1000, 采样步数为25, 损失权重 λ 为0.7。

3.1.2 数据集

本文在3个具有代表性的数据集上进行了阴影去除实验, ISTD(image shadow triplets dataset)(Wang等, 2017)数据集、ISTD+(augmented dataset with image shadow triplets)(Le和Samaras, 2019)数据集和SRD

式中, β_i^L 和 β_i^H 表示低频和高频分量的噪声调度表, $N(\cdot)$ 表示将输入归一化到0至1之间。 β_i^L 和 β_i^H 具体的值由伽马变换系数 γ_L 、 γ_H 控制, 本文中 γ_L 和 γ_H 的默认设置为0.8和1.25。

图5展示了本文基于小波变换的非均匀扩散模型生成无阴影图像过程。该模型从小波空间而不是像素空间进行去噪, 并对低频和高频分量执行不同的去噪调度。在每个步骤 t , 通过去噪网络生成 $t-1$ 步噪声较小的样本 I_{t-1} , 通过 T 个步骤后, 生成原始干净的低频和高频分量 I_0 , 并通过逆小波变换(IWT)重建最终图像。

(shadow removal dataset)(Qu等, 2017)。ISTD数据集由1870个图像三元组(阴影图像、无阴影图像和阴影掩膜)组成, 该数据集划分为1330个训练三元组和540个测试三元组。ISTD+数据集调整了ISTD数据集中阴影图像和无阴影图像之间的色差, 因此该数据集的训练和测试三元组的数量与ISTD相同。SRD数据集分别包含2680个训练对和408个测试对的阴影和无阴影图像, 由于SRD不提供阴影掩膜, 因此本文遵循之前大多数方法在训练和测试阶段直接使用DHAN(Cun等, 2020)提供的公开SRD阴影掩膜。

3.1.3 评估指标

为了量化和比较模型的性能, 遵循先前大多数方法, 选择 256×256 像素分辨率的阴影去除结果, 并选择3个客观的评价指标: LAB颜色空间中的均方根误差(root mean square error, RMSE)、峰值信噪比(peak signal to noise ratio, PSNR)和结构相似性(SSIM)。给定一个阴影去除结果和相应的真值图, RMSE测量它们的平均像素误差, PSNR测量它们的平均像素相似度, SSIM测量它们的结构相似度。RMSE值越低表示生成结果越好, PSNR和SSIM值越高表示生成结果越好。

3.2 在SRD数据集上实验

表1展示了在SRD数据集上与近期最佳方法的定量比较结果,包括非深度学习的方法(Guo等, 2013)和基于深度学习的方法DSC(direction-aware spatial context)(Hu等, 2020)、DHAN(dual hierarchical aggregation network)(Cun等, 2020)、DC-ShadowNet(domain-classifier shadow removal network)(Jin等, 2021)、Auto-Exposure(Fu等, 2021)、Zhu等人(2022b)、BMNet(Zhu等, 2022a)、ShadowDiffusion

(Guo等, 2023b)、ShadowFormer(Guo等, 2023a)。为了确保比较的公平性,所有的结果均使用论文作者发布的指标或官方结果,结果均采用四舍五入取值法。本文方法在整幅图像区域和无阴影区域的所有指标上均取得了最优或次优的结果,特别是与目前最新方法ShadowDiffusion相比,无阴影区域的RMSE从3.44显著降低到3.11,SSIM从0.985提升到0.990,验证了本文方法在阴影去除任务中取得了较为优异的性能。

表1 SRD数据集上不同方法的定量比较
Table 1 Quantitative comparison of different methods on the SRD dataset

方法	阴影区域(S)			无阴影区域(NS)			整幅图像(ALL)		
	PSNR/dB↑	SSIM↑	RMSE↓	PSNR/dB↑	SSIM↑	RMSE↓	PSNR/dB↑	SSIM↑	RMSE↓
输入图像	18.96	0.871	36.69	31.47	0.975	4.83	18.19	0.830	14.05
Guo 等人(2013)	—	—	29.89	—	—	6.47	—	—	12.60
DSC(Hu等, 2020)	30.65	0.960	8.62	31.94	0.965	4.41	27.76	0.903	5.71
DHAN(Cun等, 2020)	33.67	0.978	8.94	34.79	0.979	4.80	30.51	0.949	5.67
DC-ShadowNet(Jin等, 2021)	34.00	0.975	7.70	35.53	0.981	3.65	31.53	0.955	4.65
Auto-Exposure(Fu等, 2021)	32.26	0.966	9.55	31.87	0.945	5.74	28.40	0.893	6.50
Zhu 等人(2022b)	34.94	0.980	7.44	35.85	0.982	3.74	31.72	0.952	4.79
BMNet(Zhu等, 2022a)	35.05	0.981	6.61	36.02	0.982	3.61	31.69	0.956	4.46
ShadowDiffusion(Guo等, 2023b)	38.72	<u>0.987</u>	4.98	37.78	0.985	<u>3.44</u>	34.73	0.970	3.63
ShadowFormer(Guo等, 2023a)	<u>36.91</u>	0.989	5.90	36.22	<u>0.989</u>	<u>3.44</u>	32.90	<u>0.958</u>	4.04
WNDM(本文)	36.39	0.985	<u>5.33</u>	<u>37.14</u>	0.990	3.11	<u>32.94</u>	0.970	<u>3.80</u>

注:加粗和下划线字体分别表示各列最优和次优结果。↑(↓)表示越高(越低)越好,“—”表示原文未提供数据。

为验证本文方法的视觉效果,图6给出了SRD数据集上的可视化结果。值得一提的是,SRD数据集的图像与ISTD数据集相比具有更复杂的纹理和颜色分布。由于本文将阴影去除视为图像生成任务,对图像的低频和高频独立地非均匀扩散生成,因此生成后的无阴影图像无论在高频细节和低频全局内容上都取得了优异的结果。如图6的最后一行,由于草地上的阴影遮挡住了大部分细节信息,其他方法在去除阴影时通常会产生模糊或者错误的结果,而本文方法可以重建出草地的细节信息和正确的颜色。此外,当图像阴影区域和非阴影区域的亮度差异较小时,此部分的阴影容易被现有方法忽略,如图6的第4个示例右下角处。当图像整体亮度较小、结构信息缺失严重时,现有方法无法恢复出图像的颜色和亮度信息并有效地去除阴影,如图6第2个

示例中,Auto-Exposure在无阴影区域产生了错误的颜色信息,ShadowDiffusion在无阴影区域产生了错误的亮度信息,并且阴影完全未被去除。而得益于本文方法的低频信息和高频信息独立非均匀重建,无论是颜色和照明信息(低频)还是纹理细节信息(高频),本文方法都能够重建出更高质量的无阴影图像。

3.3 在ISTD和ISTD+数据集上实验

表2和表3分别展示了在ISTD、ISTD+数据集上与近期最佳方法的定量比较结果,包括非深度学习的方法(Guo等, 2013)和基于深度学习的方法ST-CGAN(Wang等, 2018)、DHAN(Cun等, 2020)、Auto-Exposure(Fu等, 2021)、BMNet(Zhu等, 2022a)、ShadowDiffusion(Guo等, 2023b)、ShadowFormer(Guo等, 2023a)等。本文方法在ISTD和ISTD+数据集上同

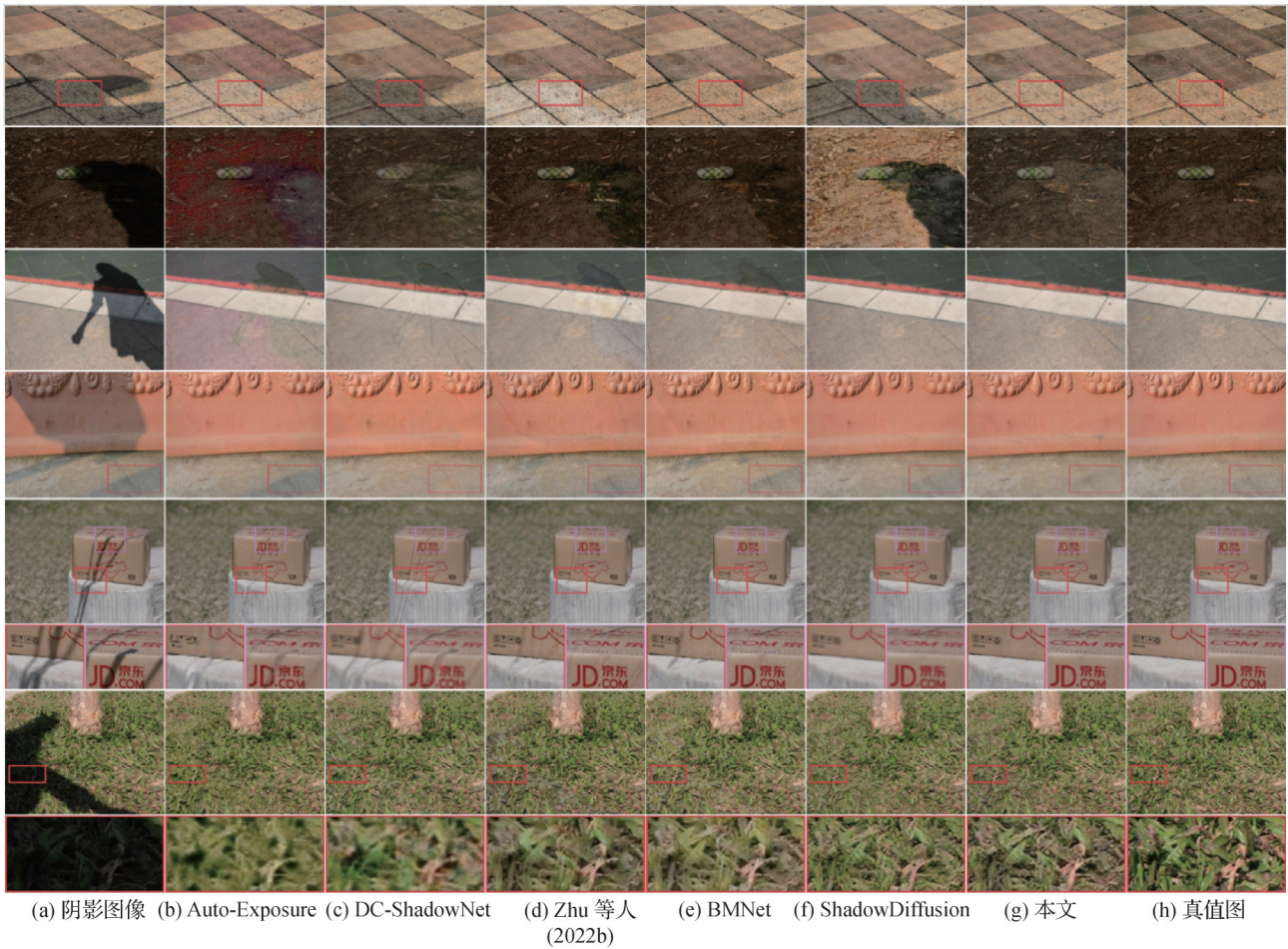


图 6 在SRD数据集测试集上对 state-of-the-art 方法与本文方法的可视化比较

Fig. 6 Visual comparison of the state-of-the-art methods and ours on the test set of the SRD dataset ((a) shadow images; (b) Auto-Exposure; (c) DC-ShadowNet; (d) Zhu et al. (2022b); (e) BMNet; (f) ShadowDiffusion; (g) ours; (h) ground truth)

表 2 ISTD数据集上不同方法的定量比较

Table 2 Quantitative comparison of different methods on the ISTD dataset

方法	阴影区域(S)			无阴影区域(NS)			整幅图像(ALL)		
	PSNR/dB↑	SSIM↑	RMSE↓	PSNR/dB↑	SSIM↑	RMSE↓	PSNR/dB↑	SSIM↑	RMSE↓
输入图像	22.40	0.936	32.10	27.32	0.976	7.09	20.56	0.893	10.88
Guo 等人(2013)	27.76	0.964	18.65	26.44	0.975	7.09	23.08	0.91	9.26
ST-CGAN(Wang等,2018)	33.74	0.981	9.99	29.51	0.958	6.05	27.44	0.929	6.65
DHAN(Cun等,2020)	35.53	0.988	7.49	31.05	0.971	5.30	29.11	0.954	5.66
DC-ShadowNet(Jin等,2021)	31.69	0.976	11.43	28.99	0.958	5.81	26.38	0.922	6.57
Auto-Exposure(Fu等,2021)	34.71	0.975	7.91	28.61	0.880	5.51	27.19	0.945	5.88
Zhu 等人(2022b)	36.95	0.987	8.29	31.54	0.975	4.55	29.85	0.960	5.09
BMNet(Zhu等,2022a)	35.61	0.988	7.60	32.80	0.976	4.59	30.28	0.959	5.02
ShadowDiffusion(Guo等,2023b)	40.15	0.994	4.13	<u>33.70</u>	<u>0.977</u>	<u>4.14</u>	32.33	0.969	4.12
ShadowFormer(Guo等,2023a)	<u>37.03</u>	0.985	6.76	32.20	0.953	4.44	30.47	0.935	4.79
WNDM(本文)	36.67	<u>0.990</u>	<u>6.51</u>	34.06	0.981	4.10	<u>31.37</u>	<u>0.967</u>	<u>4.50</u>

注:加粗和下划线字体分别表示各列最优和次优结果。↑(↓)表示越高(越低)越好。

样取得了优异的性能。如在表3中,本文方法在无阴影区域和整幅图像的RMSE指标上均取得了最佳的结果。如表3中ISTD+数据集上的无阴影区域结果所示,一些阴影去除方法会在去除阴影后破坏非阴影区域。凭借本文提出的低频、高频非均匀扩散生成的优点,可以对图像重建出更精细化的低频和高频信息,从而生成的无阴影图像具有更加正确的颜色、亮度和纹理信息。此外,图7展示了在ISTD和ISTD+数据集上的可视化比较。

表3 ISTD+数据集上不同方法的定量比较
Table 3 Quantitative comparison of different methods on the ISTD+ dataset

方法	RMSE		
	阴影区域	无阴影区域	整幅图像
输入图像	40.2	2.6	8.5
ST-CGAN(Wang等,2018)	13.4	7.7	8.7
DHAN(Cun等,2020)	11.2	7.1	7.8
Auto-Exposure(Fu等,2021)	6.6	3.8	4.2
BMNet(Zhu等,2022a)	5.6	2.5	3.0
ShadowDiffusion(Guo等,2023b)	4.9	2.3	2.7
ShadowFormer(Guo等,2023a)	5.2	2.3	2.8
WNDM(本文)	5.3	2.2	2.7

注:加粗字体表示各列最优结果。

3.4 与ShadowDiffusion的进一步比较

ShadowDiffusion(Guo等,2023b)是当前基于扩

散模型效果最优的阴影去除方法,其结合阴影物理先验信息并逐步迭代细化阴影掩膜,为扩散生成提供了更加精确的指导信息,在去除较低分辨率图像阴影时取得了优异的结果。然而,由于其对图像像素空间进行扩散建模,本文方法在去除更高分辨率的阴影图像时,效果上更具优势。表4展示了其与本文方法在去除不同分辨率图像阴影时的定量比较,图8展示了视觉对比。ShadowDiffusion和本文方法训练时均将原图缩放至256×256像素在SRD数据集进行训练。可以观察到,本文方法无论生成256×256像素分辨率图像还是生成840×640像素(原图)都取得了几乎一致性的效果。而ShadowDiffusion虽然在生成256×256像素分辨率图像时,取得了不错的效果,但生成原图分辨率大小的图像时,性能迅速恶化,特别是阴影区域,阴影几乎不能去除。由于ShadowDiffusion对256×256像素分辨率图像的像素空间进行扩散建模,因此在生成同样分辨率大小图像时会取得很好的效果,但若生成高分辨率的原图大小图像时,由于模型对像素空间的分辨率较为敏感(模型并不具备超分效果),因此其性能发生恶化。而本文方法由于对小波空间的低频和高频分量进行扩散建模,因此对图像分辨率并不敏感,对不同分辨率的阴影去除效果更具鲁棒性。

除此之外,扩散模型由于需要迭代生成图像,因此采样时间长是其一大劣势,而本文在小波空间进行扩散建模可以极大地缩减采样时间。根据表4,本文方法与ShadowDiffusion相比,在生成一幅840×

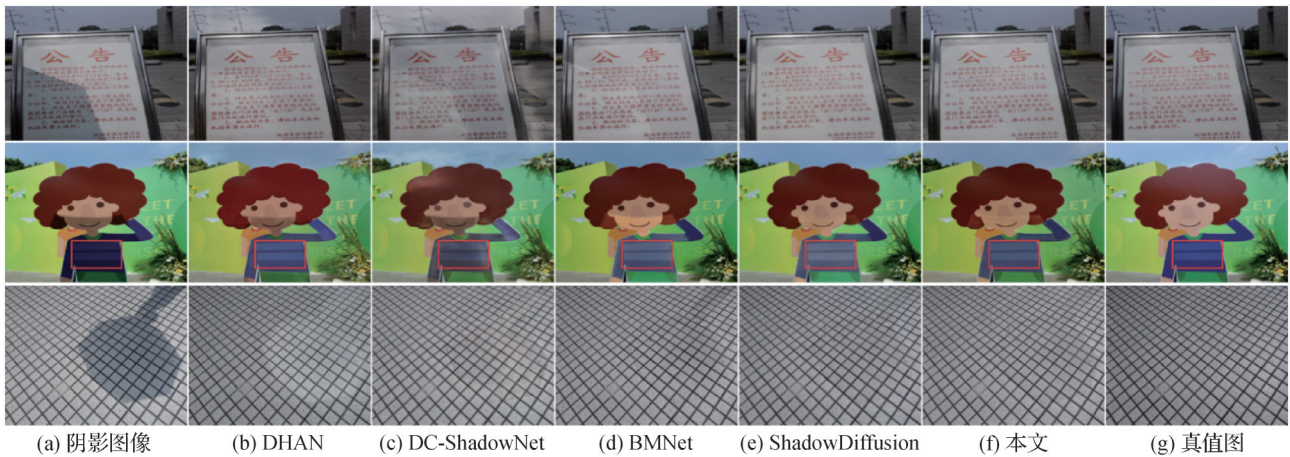


图7 在ISTD和ISTD+数据集测试集上对state-of-the-art方法与本文方法的可视化比较

Fig. 7 Visual comparison of the state-of-the-art methods and ours on the test sets of the ISTD and ISTD+ datasets
((a) shadow images; (b) DHAN; (c) DC-ShadowNet; (d) BMNet; (e) ShadowDiffusion; (f) ours; (g) ground truth)

表 4 ShadowDiffusion 与本文方法的定量比较
Table 4 Quantitative comparison of ShadowDiffusion and ours

方法	分辨率/像素	RMSE ↓			生成时间/s ↓
		阴影区域(S)	无阴影区域(NS)	整幅图像(ALL)	
ShadowDiffusion	256 × 256	4.98	3.44	3.63	0.92
	840 × 640	9.50	6.19	7.19	6.97
WNDM(本文)	256 × 256	5.33	3.11	3.80	0.39
	840 × 640	5.75	3.37	4.06	1.65

注: ↓ 表示越低越好。

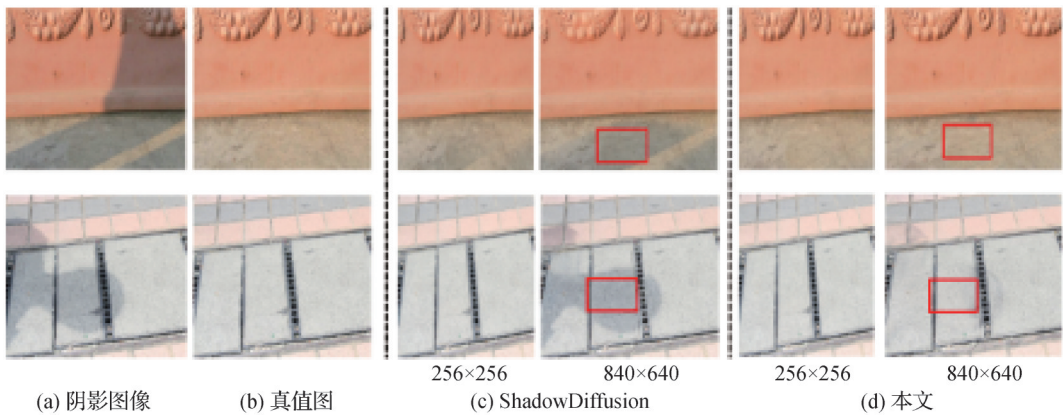


图 8 ShadowDiffusion 与本文方法在生成不同分辨率无阴影图像的质量对比

Fig. 8 Comparison of ShadowDiffusion and ours in generating shadowless images of different resolutions
((a) shadow images; (b) ground truth; (c) ShadowDiffusion; (d) ours)

640 像素分辨率的图像时,提升了约 4 倍的速度。为与更多方法比较生成图像速度,表 5 展示了基于不同类别的阴影去除方法处理 256 × 256 像素分辨率图像的时间。

表 5 不同方法生成 256 × 256 像素分辨率图像的时间比较
Table 5 Time comparison of different methods for generating 256 × 256 resolution images

方法	主要类型	生成时间/s
Auto-Exposure	CNN	0.08
BMNet	INN	0.07
ShadowFormer	Transformer	0.15
ShadowDiffusion	Diffusion	0.92
WNDM(本文)	Diffusion	0.39

3.5 消融实验

3.5.1 小波域扩散生成效果分析

为了验证对小波域进行扩散建模的优势,本文在 ISTD+数据集上分别对小波空间和像素空间扩散

建模进行实验。除了扩散目标不同,该两组实验的其余网络组件和参数均相同。表 6 提供了该两组实验的结果,可以发现对图像小波空间扩散建模的 RMSE、PSNR 指标在阴影区域和无阴影区域均有提升,且在不同分辨率下的阴影去除结果更具稳定性。

3.5.2 小波变换尺度分析

为了研究不同的小波分解尺度对模型性能的影响,表 7 报告了在 SRD 数据集上整幅图像的 PSNR、RMSE、SSIM 结果,实验中除了小波分解尺度不同,其余变量均相同。结果表明,随着小波分解尺度 K 的增加,模型的推理时间逐级减少,但性能也整体下降。推理时间减少是由于随着小波尺度的增加,对小波域扩散的空间维度减少了 4^k 倍。由于本文方法在低频和高频扩散流上共享去噪网络, K 的增加会加剧低频和高频扩散建模分支的通道失衡,如 $K = 1$ 时,低频扩散建模的通道数为 3,高频为 9;而当 $K = 3$ 时,低频和高频的建模通道数分别为 3 和 189。因此,这会增加网络的学习难度,导致最终性能下降。

表7表明,当 $K = 1$ 时,生成一幅 256×256 像素分辨率的图像也仅需 0.39 s 。因此,综合模型的性能和效率,选择 $K = 1$ 作为本文实验的默认设置。

表6 对小波域扩散建模影响的消融研究
Table 6 Ablation studies on the influence of wavelet domain diffusion mode

方法	分辨率 /像素	阴影区域		无阴影区域		整幅图像	
		P/dB $\uparrow$	R $\downarrow$	P/dB $\uparrow$	R $\downarrow$	P/dB $\uparrow$	R $\downarrow$
像素空间建模	256×256	37.14	6.4	38.35	2.4	33.91	3.0
	640×480	35.20	7.7	35.54	3.7	31.08	4.3
小波空间建模	256×256	38.03	5.3	39.37	2.2	34.85	2.7
	640×480	37.40	5.8	38.98	2.3	34.48	2.9

注: $\uparrow(\downarrow)$ 表示越高(越低)越好, P 为 PNSR, R 为 RMSE。

表7 不同的小波分解尺度对模型性能的影响
Table 7 The influence of different wavelet decomposition scales on model performance

小波 尺度	PSNR /dB $\uparrow$	SSIM $\uparrow$	RMSE $\downarrow$	时间 $\downarrow$ /s	
				256×256 /像素	840×640 /像素
$K = 1$	32.94	0.970	3.80	0.39	1.65
$K = 2$	32.64	0.966	3.97	0.18	0.56
$K = 3$	31.58	0.943	4.81	0.11	0.35

注: $\uparrow(\downarrow)$ 表示越高(越低)越好, P 为 PNSR, R 为 RMSE。

3.5.3 扩散控制系数 γ_L 和 γ_H 对模型的影响分析

为了验证对小波域低频和高频分量进行非均匀扩散建模的优势以及探究扩散速度控制系数 γ_L 和 γ_H 对模型的影响,图9展示了不同控制系数对噪声调度表的影响,表8和图10分别展示了在ISTD数据集上对性能影响的客观比较与可视化比较。表8中,(a)代表原始的均匀扩散,(b)—(e)分别代表低频和高频不同程度的非均匀扩散。从图9可知, γ_L 和 γ_H 的值越小代表扩散过程中每一步加噪的程度更大,图像更快接近高斯噪声,即加速扩散,反之则减速扩散。经过实验发现,当 $\gamma_L = 0.8$ 、 $\gamma_H = 1.25$,即低频略微加速扩散,高频略微减速扩散,本文方法取得较好结果,因此将该组参数作为本文方法的默认设置。

根据表8中的(a)和(e)可验证非均匀扩散的有

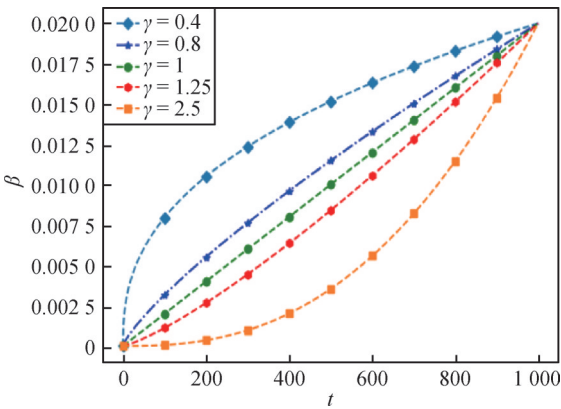


图9 不同扩散控制系数对 β_t (噪声调度表)的影响

Fig. 9 The influence of different diffusion control coefficients on β_t (noise schedule)

表8 低频分支和高频分支的扩散控制系数对模型性能的影响
Table 8 The influence of diffusion control coefficients of low frequency and high frequency branches on model performance

	γ_L	γ_H	阴影区域		无阴影区域		整幅图像	
			P/dB $\uparrow$	R $\downarrow$	P/dB $\uparrow$	R $\downarrow$	P/dB $\uparrow$	R $\downarrow$
(a)	1.0	1.0	36.15	6.87	32.96	4.54	30.56	4.88
(b)	0.4	1.25	35.74	7.14	32.33	4.73	30.01	5.08
(c)	0.8	2.5	36.42	6.67	33.62	4.38	31.03	4.76
(d)	0.4	2.5	35.48	7.35	31.80	4.85	29.59	5.21
(e)	0.8	1.25	36.67	6.51	34.06	4.10	31.37	4.50

注: $\uparrow(\downarrow)$ 表示越高(越低)越好, P 为 PNSR, R 为 RMSE。

效性,图10(a)(e)表明,对低频和高频信息进行非均匀扩散可以产生更加精细化的低频信息和高频细节。根据表8中的(b)和(e),当 γ_L 从0.8降到0.4,即低频分量的扩散速度更快时,模型性能发生显著下降。经分析发现,若低频扩散速度过快,扩散正向过程每步添加的噪声过多时,会增加网络在逆向去噪过程中的难度,导致生成不正确的低频信息,如图10(b)中的阴影区域会产生不正确的颜色和亮度信息。根据表8中的(c)和(e),当 γ_H 从1.25升到2.5时,高频分量进行过缓的减速扩散,模型同样会发生性能下降。经分析,若扩散正向过程中每步添加的噪声过少,可能 T 步后高频分量仍未接近高斯噪声,导致不正确的扩散建模,因此会生成不精确的高频信息,如图10(c)中的阴影边缘处产生严重的伪影。

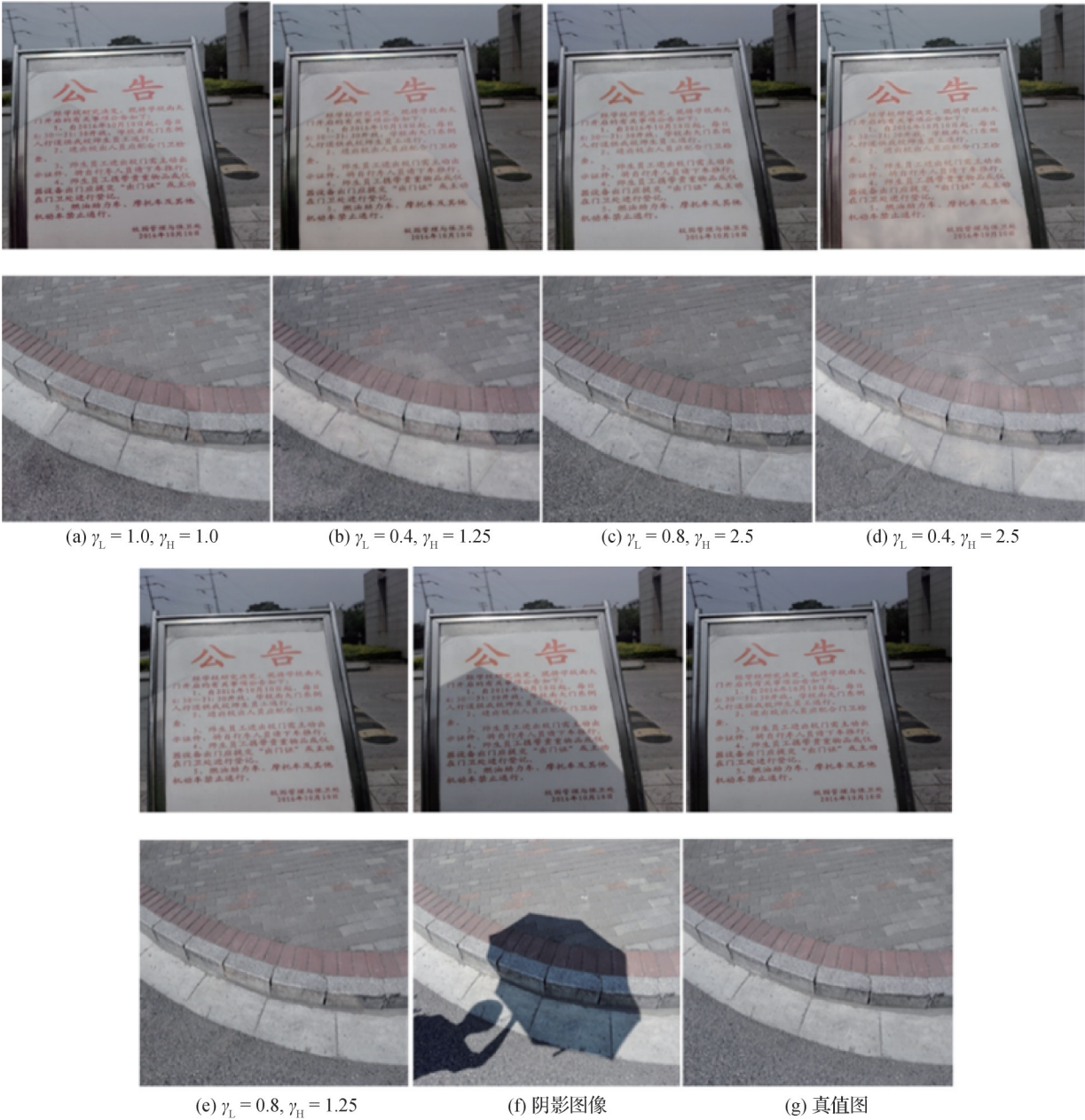


图 10 γ_L 和 γ_H 对模型效果的可视化对比

Fig. 10 Visual comparison of model performance with different γ_L and γ_H ((a) $\gamma_L = 1.0, \gamma_H = 1.0$; (b) $\gamma_L = 0.4, \gamma_H = 1.25$; (c) $\gamma_L = 0.8, \gamma_H = 2.5$; (d) $\gamma_L = 0.4, \gamma_H = 2.5$; (e) $\gamma_L = 0.8, \gamma_H = 1.25$; (f) shadow images; (g) ground truth)

3.6 扩展到其他图像恢复任务

本文的 WNDM 方法可以很容易地应用于其他图像恢复任务。为了验证本文方法的通用性,在图像去雨任务(胡明娣等, 2022)中进行了实验验证。WNDM 应用于图像雨滴去除任务中时只需要移除小波输入模块中的掩膜条件信息即可。表 9 展示了本文方法在 RainDrop 数据集上与其他方法的定量比较,可以发现 WNDM 在图像去雨任务中同样取得了良好的结果。图 11 提供了雨滴去除的视觉效果。

表 9 在 RainDrop 数据集上实验的定量比较

Table 9 Quantitative comparison of experiments on the RainDrop dataset

方法	PSNR/dB	SSIM
AttnGAN(Qian等, 2018)	31.59	0.917
CCN(Quan等, 2021)	31.44	0.947
IDT(Xiao等, 2023)	31.87	0.931
RainDropDiff(Özdenizci和Legenstein, 2023)	32.29	0.942
WNDM(本文)	31.77	0.928

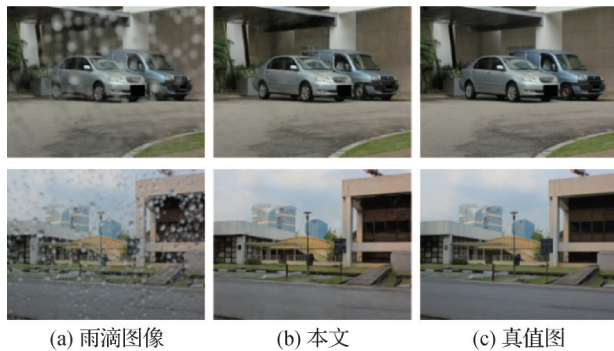


图 11 RainDrop 数据集上的可视化结果

Fig. 11 Visualization results on the RainDrop dataset
((a) raindrop images; (b) ours; (c) ground truth)

4 结 论

本文提出了一种基于小波变换的非均匀扩散模型,即 WNDM。其将阴影去除任务视为小波域的图像生成任务,对图像小波域的低频和低频分量设计了两条自适应的扩散流,分别解决阴影图像带来的低频(颜色、亮度)和高频细节退化信息。得益于小波变换的频率分解,WNDM 不是从纠缠的像素空间域学习,而是将其分离,分别进行有效的训练,从而为重建最终图像生成更加精细化的低频和高频信息。本文在多个数据集上进行了广泛的实验,证明了 WNDM 的有效性,与最先进方法相比,取得了有竞争力的结果。此外,本文方法还具有一定的通用性,除了阴影去除领域,还可以应用于其他图像恢复任务,如图像雨滴去除。在图像去雨数据集上的实验表明,本文方法在该任务中也表现出了优异的性能。

本文尚存在若干不足之处。首先,提出的 WNDM 模型依赖于配对数据集的监督式训练;其次,由于并未结合传统阴影物理模型,导致模型在阴影去除任务可解释性不够强。因此,未来的研究工作将着重结合阴影物理先验模型和条件扩散模型,为基于未配对数据集的扩散生成提供更加丰富的信息,以恢复出更高质量的无阴影图像。

参考文献(References)

Chen T. 2023. On the importance of noise scheduling for diffusion models [EB/OL]. [2024-01-08]. <https://arxiv.org/pdf/2301.10972.pdf>

- Cun X, Pun C M and Shi C. 2020. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting GAN//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI Press: 10680-10687 [DOI: 10.1609/aaai.v34i07.6695]
- Finlayson G D, Hordley S D, Lu C and Drew M S. 2006. On the removal of shadows from images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(1): 59-68 [DOI: 10.1109/TPAMI.2006.18]
- Fu L, Zhou C Q, Guo Q, Juefei-Xu F, Yu H K, Feng W, Liu Y and Wang S. 2021. Auto-exposure fusion for single-image shadow removal//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 10566-10575 [DOI: 10.1109/cvpr46437.2021.01043]
- Guo L Q, Huang S Y, Liu D, Cheng H and Wen B H. 2023a. ShadowFormer: global context helps image shadow removal [EB/OL]. [2024-01-08]. <https://arxiv.org/pdf/2302.01650.pdf>
- Guo L Q, Wang C, Yang W H, Huang S Y, Wang Y F, Pfister H and Wen B H. 2023b. ShadowDiffusion: when degradation prior meets diffusion model for shadow removal//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 14049-14058 [DOI: 10.1109/cvpr52729.2023.01350]
- Guo R Q, Dai Q Y and Hoiem D. 2013. Paired regions for shadow detection and removal. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(12): 2956-2967 [DOI: 10.1109/TPAMI.2012.214]
- Haar A. 1911. Zur theorie der orthogonalen funktionensysteme. Mathematische Annalen, 71(1): 38-53 [DOI: 10.1007/bf01456927]
- Ho J, Jain A and Abbeel P. 2020. Denoising diffusion probabilistic models//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 6840-6851
- Hu M D, Wu Y, Song Y, Yang J B, Zhang R F, Wang H and Meng D Y. 2022. The integrated evaluation and review of single image rain removal based datasets and deep learning methods. Journal of Image and Graphics, 27(5): 1359-1391 (胡明娣, 吴怡, 宋尧, 杨静冰, 张瑞芳, 王红, 孟德宇. 2022. 单幅图像去雨数据集和深度学习算法的联合评估与展望. 中国图象图形学报, 27(5): 1359-1391) [DOI: 10.11834/jig.211153]
- Hu X W, Fu C W, Zhu L, Qin J and Heng P A. 2020. Direction-aware spatial context features for shadow detection and removal. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(11): 2795-2808 [DOI: 10.1109/TPAMI.2019.2919616]
- Hu X W, Jiang Y T, Fu C W and Heng P A. 2019. Mask-shadowGAN: learning to remove shadows from unpaired data//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 2472-2481 [DOI: 10.1109/iccv.2019.00256]

- Jiang H, Luo A, Fan H Q, Han S C and Liu S C. 2023. Low-light image enhancement with wavelet-based diffusion models. *ACM Transactions on Graphics (TOG)*, 42(6): #238 [DOI: 10.1145/3618373]
- Jin Y Y, Sharma A and Tan R T. 2021. De-ShadowNet: single-image hard and soft shadow removal using unsupervised domain-classifier guided network//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*. Montreal, Canada: IEEE: 5007-5016 [DOI: 10.1109/iccv48922.2021.00498]
- Khan S H, Bennamoun M, Sohel F and Togneri R. 2016. Automatic shadow detection and removal from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(3): 431-446 [DOI: 10.1109/TPAMI.2015.2462355]
- Le H and Samaras D. 2019. Shadow removal via shadow image decomposition//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE: 8577-8586 [DOI: 10.1109/iccv.2019.00867]
- Li H Y, Yang Y F, Chang M, Chen S Q, Feng H J, Xu Z H, Li Q and Chen Y T. 2022. SRDiff: single image super-resolution with diffusion probabilistic models. *Neurocomputing*, 479: 47-59 [DOI: 10.1016/j.neucom.2022.01.029]
- Liu F and Gleicher M. 2008. Texture-consistent shadow removal//*Proceedings of the 10th European Conference on Computer Vision*. Marseille, France: Springer: 437-450 [DOI: 10.1007/978-3-540-88693-8_32]
- Long C J and Hua G. 2017. Correlational Gaussian processes for cross-domain visual recognition//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA: IEEE: 4932-4940 [DOI: 10.1109/CVPR.2017.524]
- Lugmayr A, Danelljan M, Van Gool L and Timofte R. 2020. SRFlow: learning the super-resolution space with normalizing flow//*Proceedings of the 16th European Conference on Computer Vision — ECCV 2020*. Glasgow, UK: Springer: 715-732 [DOI: 10.1007/978-3-030-58558-7_42]
- Mikic I, Cosman P C, Kogut G T and Trivedi M M. 2000. Moving shadow and object detection in traffic scenes//*Proceedings of the 15th International Conference on Pattern Recognition*. Barcelona, Spain: IEEE: 321-324 [DOI: 10.1109/ICPR.2000.905341]
- Nichol A Q and Dhariwal P. 2021. Improved denoising diffusion probabilistic models [EB/OL]. [2024-01-08]. <https://arxiv.org/pdf/2102.09672.pdf>
- Özdenizci O and Legenstein R. 2023. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8): 10346-10357 [DOI: 10.1109/TPAMI.2023.3238179]
- Phung H, Dao Q and Tran A. 2023. Wavelet diffusion models are fast and scalable image generators//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE: 10199-10208 [DOI: 10.1109/cvpr52729.2023.00983]
- Qian R, Tan R T, Yang W H, Su J J and Liu J Y. 2018. Attentive generative adversarial network for raindrop removal from a single image//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 2482-2491 [DOI: 10.1109/cvpr.2018.00263]
- Qu L Q, Tian J D, He S F, Tang Y D and Lau R W H. 2017. DeshadownNet: a multi-context embedding deep network for shadow removal//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA: IEEE: 2308-2316 [DOI: 10.1109/CVPR.2017.248]
- Quan R J, Yu X, Liang Y Z and Yang Y. 2021. Removing raindrops and rain streaks in one go//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE: 9143-9152 [DOI: 10.1109/evpr46437.2021.00903]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Saharia C, Chan W, Chang H W, Lee C, Ho J, Salimans T, Fleet D and Norouzi M. 2022. Palette: image-to-image diffusion models//*Proceedings of 2022 ACM SIGGRAPH Conference*. Vancouver, Canada: ACM: #15 [DOI: 10.1145/3528233.3530757]
- Saharia C, Ho J, Chan W, Salimans T, Fleet D J and Norouzi M. 2023. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4713-4726 [DOI: 10.1109/TPAMI.2022.3204461]
- Sohl-Dickstein J, Weiss E A, Maheswaranathan N and Ganguli S. 2015. Deep unsupervised learning using nonequilibrium thermodynamics//*Proceedings of the 32nd International Conference on International Conference on Machine Learning*. Lille, France: JMLR.org: 2256-2265
- Song J M, Meng C L and Ermon S. 2021. Denoising diffusion implicit models [EB/OL]. [2024-01-08]. <https://arxiv.org/pdf/2010.02502v2.pdf>
- Tang L, Zhao C X, Wang H N and Shao W Z. 2008. Shadow removal for road surface images based on anisotropic diffusion Retinex. *Journal of Image and Graphics*, 13(2): 264-268 (唐磊, 赵春霞, 王鸿南, 邵文泽. 2008. 基于各向异性 Retinex 的路面图像阴影消除. *中国图象图形学报*, 13(2): 264-268) [DOI: 10.11834/jig.20080215]
- Wang J F, Li X and Yang J. 2018. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 1788-1797 [DOI: 10.1109/CVPR.2018.00192]
- Wang K F, Gou C, Duan Y J, Lin Y L, Zheng X H and Wang F Y. 2017. Generative adversarial networks: introduction and outlook. *IEEE/CAA Journal of Automatica Sinica*, 4(4): 588-598 [DOI:

10.1109/JAS.2017.7510583]

Wang J, Delbracio M, Talebi H, Saharia C, Dimakis A G and Milanfar P. 2022. Deblurring via stochastic refinement//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 16272-16282 [DOI: 10.1109/CVPR52688.2022.01581]

Xiao J, Fu X Y, Liu A P, Wu F and Zha Z J. 2023. Image De-raining Transformer. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45 (11): 12978-12995 [DOI: 10.1109/tpami.2022.3183612]

Yi X P, Xu H, Zhang H, Tang L F and Ma J Y. 2023. Diff-retinex: rethinking low-light image enhancement with a generative diffusion model [EB/OL]. [2024-01-08].
<https://arxiv.org/pdf/2308.13164.pdf>

Yue Z S, Wang J Y and Loy C C. 2023. ResShift: efficient diffusion model for image super-resolution by residual shifting [EB/OL]. [2024-01-08]. <https://arxiv.org/pdf/2307.12348.pdf>

Zhu Y R, Huang J, Fu X Y, Zhao F, Sun Q B and Zha Z J. 2022a. Bijective mapping network for shadow removal//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Rec-

ognition. New Orleans, USA: IEEE: 5617-5626 [DOI: 10.1109/cvpr52688.2022.00554]

Zhu Y R, Xiao Z Y, Fang Y C, Fu X Y, Xiong Z W and Zha Z J. 2022b. Efficient model-driven network for shadow removal//Proceedings of the 36th AAAI Conference on Artificial Intelligence. Virtually: AAAI Press: 3635-3643 [DOI: 10.1609/aaai.v36i3.20276]

作者简介

黄颖,男,副教授,主要研究方向为图像处理、图像融合、图像质量评估、计算成像、智能信息处理与模式识别。
E-mail:huangying@cqupt.edu.cn

程彬,通信作者,男,硕士研究生,主要研究方向为图像增强和深度学习。E-mail:2606301180@qq.com

房少杰,男,硕士研究生,主要研究方向为图像处理、深度学习和阴影去除。E-mail:s211231012@stu.cqupt.edu.cn

刘歆,女,副教授,硕士生导师,主要研究方向为图像处理和机器学习。E-mail:liuxin@cqupt.edu.cn