

基于跨区域特征交互与掩码引导的单幅图像去阴影算法研究

黄鑫庆

北京建筑大学理学院, 北京

收稿日期: 2026年3月23日; 录用日期: 2026年4月30日; 发布日期: 2026年5月13日

摘要

单幅图像去阴影旨在恢复被光源遮挡区域的底层内容, 是计算机视觉领域重要且极具挑战性的任务。现有多数深度学习去阴影算法在特征修复时, 往往过度依赖阴影区域内部的局部特征映射, 忽视了同一图像中非阴影区域所蕴含的丰富且干净的背景先验, 导致大面积阴影的复原结果极易出现纹理模糊与结构失真。针对此问题, 本文提出了一种基于跨区域特征交互与掩码引导的单幅图像去阴影算法。该算法构建了一个多尺度的编码器-解码器网络, 并创新性地网络的最深层(瓶颈层)引入了跨区域特征感知注意力模块(SACA)。该模块以二值阴影掩码作为空间约束, 将非阴影区域的高质量精细特征作为键(Key)和值(Value), 将阴影区域特征作为查询(Query)。通过交叉注意力机制, 网络能够跨越空间距离, 将背景处完整的纹理与结构信息精准对齐并传递至阴影目标区域, 有效打破了阴影重构的信息瓶颈。在公开的去阴影数据集上进行的广泛定性与定量实验表明, 本文方法在客观评价指标上优于现有主流算法, 生成的无阴影图像具有更高的纹理清晰度与更平滑的边界过渡。

关键词

图像去阴影, 特征交互, 注意力机制, 纹理恢复, 特征对齐, 掩码引导

Single Image Shadow Removal via Cross-Region Feature Interaction and Mask Guidance

Xinqing Huang

School of Science, Beijing University of Civil Engineering and Architecture, Beijing

Received: March 23, 2026; accepted: April 30, 2026; published: May 13, 2026

Abstract

Single image shadow removal aims to recover the underlying content in regions occluded by light sources, which remains a crucial yet challenging task in computer vision. During feature restoration, most existing deep learning-based shadow removal algorithms rely heavily on local feature mapping within the shadow regions. However, they tend to overlook the rich and clean background priors inherent in the non-shadow regions of the same image, which often leads to texture blurring and structural distortion when dealing with large-area shadows. To address this issue, this paper proposes a single image shadow removal algorithm based on cross-region feature interaction and mask guidance. Specifically, we construct a multi-scale encoder-decoder network and innovatively introduce a Structure-Aware Cross Attention (SACA) module at the deepest bottleneck layer. With the binary shadow mask acting as a spatial constraint, this module utilizes the high-quality fine features from the non-shadow regions as the Key and Value, while taking the shadow region features as the Query. Through a cross-attention mechanism, the network bridges spatial distances to accurately align and propagate the intact texture and structural information from the background to the target shadow regions, thereby effectively breaking the information bottleneck in shadow reconstruction. Extensive qualitative and quantitative experiments on public shadow removal datasets demonstrate that the proposed method outperforms mainstream algorithms in objective evaluation metrics, generating shadow-free images with clearer textures and smoother boundary transitions.

Keywords

Image Shadow Removal, Feature Interaction, Attention Mechanism, Texture Restoration, Feature Alignment, Mask Guidance

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

阴影是自然场景中由于光源被遮挡而产生的普遍光学现象。尽管阴影在日常生活中极为常见，但它们会显著改变目标物体的表面亮度和色度，导致图像特征表示发生严重畸变。因此，阴影的存在不仅降低了图像的视觉质量，还极大地限制了诸如目标检测、语义分割和视觉跟踪等下游计算机视觉任务的性能。实现高效、高保真的单幅图像去阴影，是提升复杂视觉系统鲁棒性的关键环节。

近年来，随着深度学习技术的快速发展，图像去阴影任务取得了显著的进展。现有的主流方法大多采用卷积神经网络(CNN)或生成对抗网络(GAN)来学习阴影到无阴影的端到端映射。然而，由于真实场景中光照条件的多样性以及阴影边界的复杂性，现有模型在处理大面积阴影或具有复杂纹理的背景时，依然面临着巨大的挑战。其核心痛点在于：传统的卷积架构受限于局部感受野，难以在空间上建立长距离的特征依赖关系。当阴影区域由于光照严重衰减而丢失底层纹理时，网络仅靠阴影区域内部的局部特征往往难以“凭空”重构出逼真的细节，最终导致复原图像在阴影区域出现纹理模糊、色彩失真或边界过渡生硬等伪影现象。

事实上，在一幅自然图像中，非阴影区域通常保留着完整、清晰的结构和材质信息，这些干净的特征能够为阴影区域的修复提供极其可靠的“向导”。然而，现有的大多数去阴影网络未能显式地建模阴

影与非阴影区域之间的交互关系，未能充分释放非阴影区域的特征潜力。

为了打破这一信息瓶颈，本文提出了一种基于跨区域特征交互的图像去阴影算法。本文的核心贡献如下：

(1) 构建了一个基于编码器-解码器架构的深度去阴影网络，旨在通过长距离特征依赖来解决大面积阴影的纹理恢复难题。

(2) 创新性地网络的深层瓶颈处提出了跨区域特征感知注意力模块(SACA)。该模块以阴影掩码为空间先验，显式地将非阴影区域的高质量特征“搬运”并对齐到阴影区域，实现了精准的纹理传播。

(3) 在公开的去阴影数据集上进行了广泛的实验。结果表明，本文方法在峰值信噪比(PSNR)和结构相似性(SSIM)上均取得了优异的表现，能够生成纹理清晰、边界自然的无阴影图像。

2. 相关工作

2.1. 传统图像去阴影方法

早期的图像去阴影方法主要依赖于物理光照模型或手工提取的先验特征。例如，部分研究通过计算图像的梯度域或熵最小化来区分阴影边界与反射率边界，进而实现阴影的擦除。然而，这些方法通常建立在较为理想的物理假设之上(如均匀光源或纯色背景)，当面对真实世界中复杂的光照变化和多样化的物体材质时，往往容易失效，泛化能力较弱。

2.2. 基于深度学习的图像去阴影方法

随着大规模数据集的构建，基于数据驱动的深度学习方法逐渐主导了去阴影领域。早期的深度去阴影模型如 *DeshadowNet* [1]通过提取多尺度上下文特征来预测阴影遮罩并恢复光照。随后，生成对抗网络(GAN)被引入该任务，如 *Mask-ShadowGAN* [2]利用循环一致性损失实现了无监督或弱监督的去阴影训练；*DSC* [3]网络则探索了方向感知的空间上下文特征以提升阴影区域的结构连贯性。

尽管上述方法取得了显著进步，但大多依赖于局部卷积操作的堆叠。近期，部分学者开始引入 *Transformer* [4]架构以捕获长距离依赖。然而，现有模型依然缺乏对“阴影”与“非阴影”两个异质区域之间信息流动的显式调控。本文与上述方法的核心区别在于，我们将去阴影任务建模为一个跨区域的特征检索与映射过程，通过 *SACA* 模块显式地利用干净背景引导阴影重构，从而大幅提升了纹理恢复的上限。

3. 本文方法

3.1. 整体网络架构

为了充分利用非阴影区域的特征来引导阴影区域的修复，本文提出了一种基于跨区域特征交互的深度编码器-解码器网络，整体采用 *U-net* 架构[5]，其整体框架如图 1 所示。

如图 1 所示，网络的输入包含两部分：退化的 RGB 阴影图像 $I_m \in \mathbb{R}^{3 \times H \times W}$ 以及对应的二值阴影掩码 $M \in \mathbb{R}^{1 \times H \times W}$ (其中像素值为 1 表示阴影区域，0 表示非阴影区域)。将 I_m 与 M 在通道维度上进行拼接后，送入特征编码器。本文的编码器(Encoder)由 4 个连续的卷积模块(Encoder Block 1~4)组成，通过逐步的下采样操作，提取从浅层边缘到深层语义的多尺度特征表示。随着网络深度的增加，特征图的空间分辨率逐渐降低。在到达网络的瓶颈层(即 Encoder Block 4 之后)时，特征图尺寸降至 $H/8 \times W/8$ ，此时感受野达到最大，但阴影区域的底层纹理信息也极度匮乏。为了打破这种大面积阴影带来的信息孤岛，本文在瓶颈层嵌入了跨区域特征感知注意力模块(*Structure-Aware Cross Attention, SACA*)。该模块利用深层特征的全局感受野，跨越空间距离，将非阴影区(NS)的精细纹理对齐并传播至阴影区(S)。随后，经过 *SACA* 模

块强化的特征被送入解码器(Decoder)。解码器同样包含 4 个层级(Decoder Block 4~1)，通过上采样操作逐步恢复特征的空间分辨率为 $H/2 \times W/2$ 直至 $H \times W$ 。在此过程中，网络通过虚线所示的跳跃连接(Skip Connections)，将前 3 个编码器层级的多尺度高分辨率细节与解码器特征进行融合，最终输出具有高保真纹理的无阴影图像(Output)。

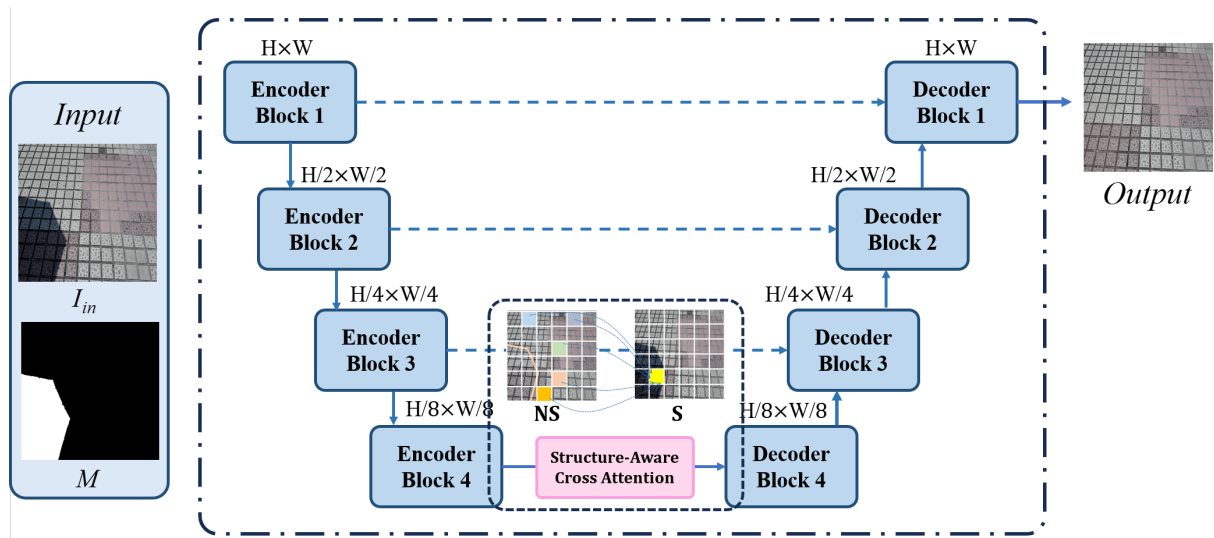


Figure 1. Overall architecture of the shadow removal network based on cross-region feature interaction
图 1. 基于跨区域特征交互的去阴影网络整体架构

3.2. 跨区域特征感知注意力模块(SACA)

在传统的编解码架构中，阴影区域的特征表征往往会因为多次下采样而变得极度匮乏。考虑到自然图像中阴影与非阴影区域在物理材质上通常具有高度的连续性，充分利用非阴影区的干净特征来引导阴影区的修复是突破信息瓶颈的关键。基于此，本文在网络的瓶颈层设计了跨区域特征感知注意力模块(SACA)，其详细的内部网络构造如图 2 所示。

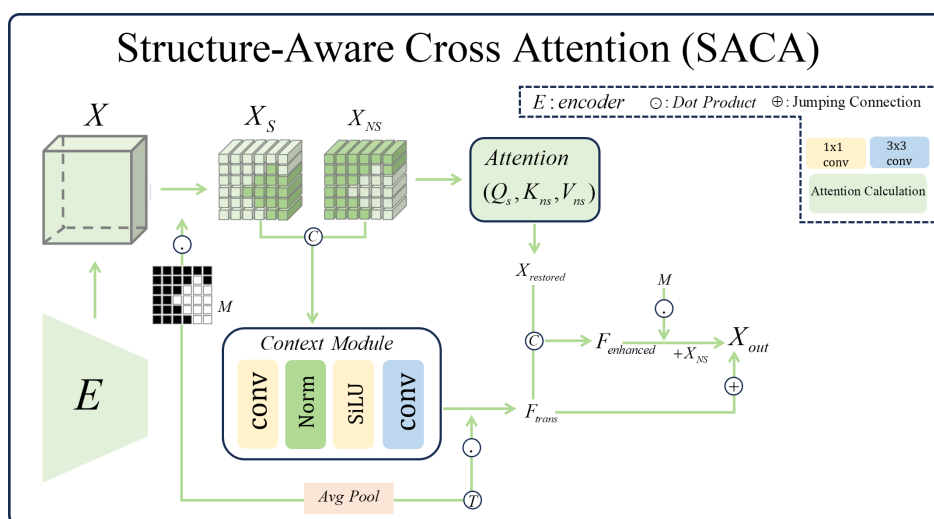


Figure 2. Structure diagram of the Structure-Aware Cross Attention (SACA) module
图 2. 跨区域特征感知注意力模块(SACA)结构图

如图 2 所示, SACA 模块的核心思想是通过阴影掩码将特征空间显式解耦, 并利用交叉注意力机制 (Cross-Attention) 完成纹理的跨界传播。设输入到瓶颈层的深层特征图为 $X \in \mathbb{R}^{C \times \frac{H}{8} \times \frac{W}{8}}$, 对应的下采样二值掩码为 M' (1 代表阴影, 0 代表非阴影)。首先, 利用掩码将全局特征 X 在空间维度上分离为阴影特征 X_s 与非阴影特征 X_{ns} :

$$X_s = X \odot M' \quad (1)$$

$$X_{ns} = X \odot (1 - M') \quad (2)$$

其中 \odot 表示空间维度的逐元素相乘。

在跨区域注意力计算阶段, 特征被划分为“待修复区”与“引导字典”两个独立分支。具体而言, 阴影特征 X_s 作为待修复目标, 通过 1×1 卷积映射为查询矩阵 Q_s ; 而非阴影特征 X_{ns} 作为包含干净纹理的引导源, 被映射为键矩阵 K_{ns} 和值矩阵 V_{ns} :

$$Q_s = W_q \cdot X_s \quad (3)$$

$$K_{ns} = W_k \cdot X_{ns}, V_{ns} = W_v \cdot X_{ns} \quad (4)$$

其中 W_q, W_k, W_v 为对应的可学习线性投影权重。

随后, 计算阴影区与非阴影区之间的跨区域相关性矩阵 A_{cross} 。该矩阵精准度量了阴影区每一个像素位置与背景区所有像素位置在结构语义上的相似度:

$$A_{cross} = \text{Softmax} \left(\frac{Q_s \cdot K_{ns}^\top}{\sqrt{d}} \right) \quad (5)$$

其中 \sqrt{d} 为缩放因子, 用于维持梯度的稳定性。在获取相似度矩阵后, 利用该矩阵对背景值特征 V_{ns} 进行加权聚合。这一步在物理意义上等效于根据结构相似性, 将背景区域的有效纹理特征“复制”并填补到阴影区域对应的空间位置:

$$Y_{fill} = A_{cross} \cdot V_{ns} \quad (6)$$

最后, 为了保证网络在传播过程中的特征完整性, 将重构后的阴影特征 Y_{fill} 与原始输入特征 X 进行通道拼接 (Concatenation), 并通过残差连接输出 SACA 模块的最终特征表示 Y :

$$Y = \text{Conv}_{out} (Y_{fill} \oplus X) + X \quad (7)$$

其中 \oplus 表示特征拼接, Conv_{out} 为用于特征平滑降维的卷积层。通过上述精细的特征交互机制, SACA 模块在最底层的语义空间中有效打破了阴影重构的信息壁垒。

3.3. 损失函数

为了端到端地优化所提出的去阴影网络, 本文设计了联合损失函数来约束网络的训练过程, 确保复原图像在像素级与感知级均具备极高的逼真度。总损失函数 \mathcal{L}_{total} 定义如下:

$$\mathcal{L}_{total} = \mathcal{L}_{char} + \lambda \mathcal{L}_{per} \quad (8)$$

(1) Charbonnier 损失 (\mathcal{L}_{char}): 为了提高模型对异常值的鲁棒性, 本文采用 Charbonnier 损失代替常规的 L1 或 L2 损失, 用于约束网络输出 I_{out} 与真实无阴影标签 I_{gt} 在像素级别的差异:

$$\mathcal{L}_{char} = \sqrt{\|I_{out} - I_{gt}\|^2 + \epsilon^2} \quad (9)$$

其中 ϵ 为防止梯度爆炸的极小常数 (本文设为 10^{-3})。

(2) 感知损失(\mathcal{L}_{per}): 为了消除由于像素级损失导致的图像过度平滑问题, 提升人类视觉感知的真实感, 本文引入了基于预训练 VGG-16 网络的感知损失:

$$\mathcal{L}_{per} = \sum_j \frac{1}{C_j H_j W_j} \left\| \phi_j(I_{out}) - \phi_j(I_{gt}) \right\|_1 \quad (10)$$

其中 $\phi_j(\cdot)$ 表示 VGG-16 网络第 j 层提取的特征图, C_j, H_j, W_j 分别为其通道数、高度与宽度。 λ 为平衡各项损失权重的超参数(本文实验中经验性地设为 0.1)。

4. 实验结果与分析

4.1. 实验环境与数据集

本文在公开的单幅图像去阴影基准数据集 AISTD 上对所提算法进行验证。原有的 ISTD 数据集虽然应用广泛, 但其提供的无阴影标签图像在阴影与非阴影区域之间存在一定的色彩不一致问题。AISTD 数据集通过专业的色彩匹配算法对此进行了修正, 从而能够更精准地评估网络在色彩与纹理恢复上的真实性能。该数据集共包含 1870 组高质量的图像三元组(阴影图像、阴影掩码、无阴影图像), 其中 1330 组用于网络训练, 剩余的 540 组用于测试评估。

4.2. 定量结果对比

为了验证本文算法的优越性, 将其与近年来主流的深度学习去阴影算法(包括 DSC [3]、DHAN [6]、DC-ShadowNet [7]、SG-ShadowNet [8]、FusionNet [9]以及 DeS3 [10])在 AISTD 测试集上进行了定量对比。所有客观评价数据均整理自原论文或其官方开源代码的复现结果, 具体如表 1 所示。

由表 1 可知, 本文方法在 AISTD 数据集上取得了最佳的综合性能表现。在 MAE 指标上, 本文方法在阴影、非阴影和全局区域分别达到了 6.13、2.49 和 3.15, 显著降低了图像的整体像素误差。在反映结构保真度的 SSIM 指标上, 本文方法在全局区域达到了 0.967 的最优水平。尤其值得注意的是, 本文方法在非阴影区域(NS)的 PSNR 达到了 38.52 dB, 远超其他所有对比方法。这表明我们的方法不仅能够高效擦除阴影, 还能在极大程度上保护并增强未受阴影遮挡的高亮度细节区域, 有效避免了传统算法中常见的背景特征退化问题。

Table 1. Comparison table of AISTD Dataset

表 1. AISTD 数据集对比表

Method	MAE↓ (S.)	MAE↓ (NS.)	MAE↓ (All)	SSIM↑ (S.)	SSIM↑ (NS.)	SSIM↑ (All)	PSNR↑ (S.)	PSNR↑ (NS.)	PSNR↑ (All)
DHAN [6]	9.49	7.39	7.74	0.988	0.972	0.958	33.08	27.28	25.75
DCShadowNet [7]	10.83	3.44	4.70	0.977	0.973	0.939	32.20	34.45	29.17
DSC [3]	7.72	5.04	5.59	0.984	0.969	0.944	34.64	31.26	29.00
FusionNet [9]	6.60	3.80	4.20	-	-	-	36.04	31.16	29.45
SG-ShadowNet [8]	5.9	2.9	3.4	-	-	-	-	-	-
Auto-Exposure [9]	6.57	3.83	4.27	0.976	0.874	0.838	36.30	31.10	29.44
PaletteDiff [11]	15.4	7.82	6.41	-	-	-	-	-	-
RepaintDiff [12]	12.9	10.66	24.9	-	-	-	-	-	-
DeS3 [10]	6.5	3.4	3.94	0.989	0.972	0.958	36.49	34.7	31.8
Ours	6.13	2.49	3.15	0.989	0.983	0.967	37.78	38.52	34.32

4.3. 定性视觉结果与局限性分析

为了直观评估所提算法的去阴影效果，本文在 AISTD 和 SRD 数据集上进行了视觉对比(如图 3、图 4 所示)。

在 AISTD 数据集中(图 3)，本文方法能够鲁棒地去除硬阴影和柔和的投射阴影。即便在背景存在强烈色彩变化的复杂场景下，得益于 SACA 模块精准的特征搬运，本文方法依然能在不引入伪影的情况下消除阴影，复原图像边界清晰、纹理与色彩高度一致。

在 SRD 数据集中(图 4)，真实的阴影掩码常伴有噪声或边界对齐偏差。实验表明，即便在这些不完美掩码的干扰下，本文算法依然展现出了极强的鲁棒性，如图 4 前两行所示。跨区域特征对齐机制有效抑制了常规方法中容易出现的严重边界伪影，证明了本算法在复杂真实场景下的可靠性。

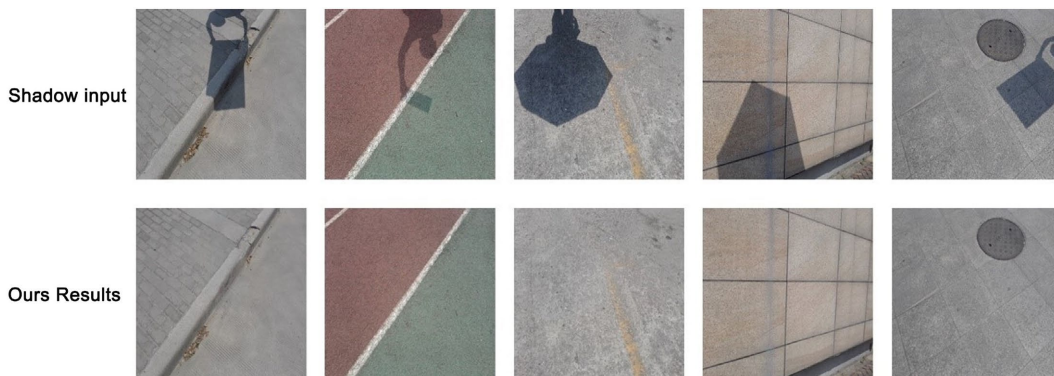


Figure 3. Visual qualitative results of the AISTD dataset
图 3. AISTD 数据集视觉定性结果

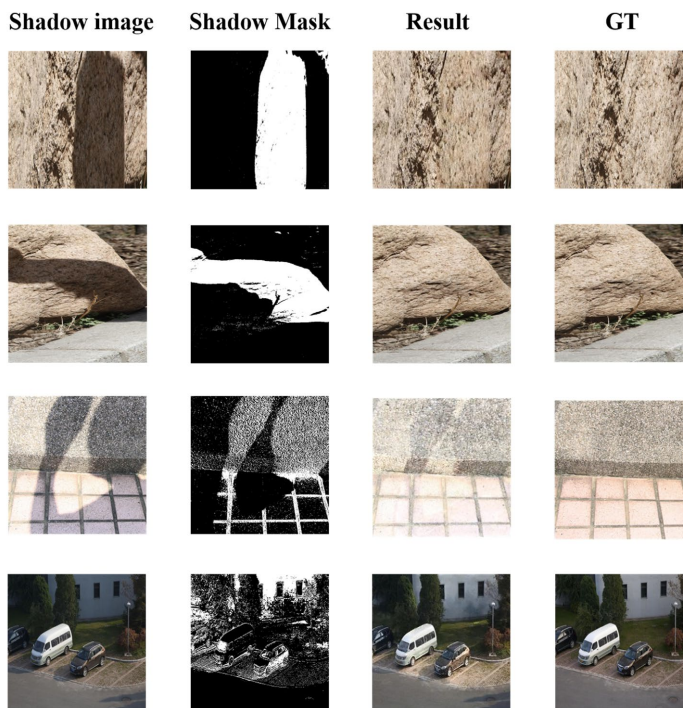


Figure 4. Visual qualitative results of the SRD dataset
图 4. SRD 数据集视觉定性结果

然而, 本文算法也存在一定的局限性, 即特征重构过程高度依赖输入掩码的质量。如图 4 后两行所示, 当面临极端情况(即输入掩码质量极差, 存在大面积雪花状破碎噪声或严重错位)时, 去阴影效果会受到较大影响。其本质原因在于, 极端劣质的掩码会严重误导 SACA 模块在深层空间中的特征解耦, 导致网络将错误的背景纹理提取并搬运至目标区域, 从而使最终复原结果中残留明显的阴影轮廓或色彩伪影。

5. 结论

本文针对单幅图像去阴影任务中大面积纹理丢失与色彩失真的难题, 提出了一种基于跨区域特征交互的去阴影网络。该算法的核心在于引入了跨区域特征感知注意力模块(SACA)。SACA 模块以阴影掩码为空间先验, 显式地将非阴影区的高质量纹理与色彩特征精准迁移至阴影区, 有效克服了局部卷积产生的信息瓶颈。在 AISTD 和 SRD 公开数据集上的大量实验表明, 本文方法在定量指标与定性视觉上均优于现有主流算法, 生成的无阴影图像边界自然、色彩一致, 且对常规的掩码边界偏差展现出了良好的鲁棒性。针对极端劣质掩码下特征重构受限的局限性, 未来我们将重点探索无掩码(Mask-free)的特征交互机制, 以进一步提升算法在复杂且不可控真实场景下的泛化与修复能力。

参考文献

- [1] Qu, L., Tian, J., He, S., Tang, Y. and Lau, R.W.H. (2017) DeshadowNet: A Multi-Context Embedding Deep Network for Shadow Removal. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 3016-3024. <https://doi.org/10.1109/cvpr.2017.248>
- [2] Hu, X., Jiang, Y., Fu, C. and Heng, P. (2019) Mask-ShadowGAN: Learning to Remove Shadows from Unpaired Data. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October 2019-2 November 2019, 2472-2481. <https://doi.org/10.1109/iccv.2019.00256>
- [3] Hu, X., Zhu, L., Fu, C., Qin, J. and Heng, P. (2018) Direction-Aware Spatial Context Features for Shadow Detection. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7454-7462. <https://doi.org/10.1109/cvpr.2018.00778>
- [4] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. <https://arxiv.org/abs/1706.03762>
- [5] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A. Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [6] Cun, X., Pun, C. and Shi, C. (2020) Towards Ghost-Free Shadow Removal via Dual Hierarchical Aggregation Network and Shadow Matting Gan. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 10680-10687. <https://doi.org/10.1609/aaai.v34i07.6695>
- [7] Jin, Y., Sharma, A. and Tan, R.T. (2021) Dc-ShadowNet: Single-Image Hard and Soft Shadow Removal Using Unsupervised Domain-Classifer Guided Network. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 5027-5036. <https://doi.org/10.1109/iccv48922.2021.00498>
- [8] Wan, J., Yin, H., Wu, Z., Wu, X., Liu, Y. and Wang, S. (2022) Style-Guided Shadow Removal. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M. and Hassner, T., Eds., *Lecture Notes in Computer Science*, Springer, 361-378. https://doi.org/10.1007/978-3-031-19800-7_21
- [9] Fu, L., Zhou, C., Guo, Q., Juefei-Xu, F., Yu, H., Feng, W., et al. (2021) Auto-Exposure Fusion for Single-Image Shadow Removal. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 10571-10580. <https://doi.org/10.1109/cvpr46437.2021.01043>
- [10] Jin, Y., Ye, W., Yang, W., Yuan, Y. and Tan, R.T. (2024) Des3: Adaptive Attention-Driven Self and Soft Shadow Removal Using Vit Similarity. *Proceedings of the AAAI Conference on Artificial Intelligence*, **38**, 2634-2642. <https://doi.org/10.1609/aaai.v38i3.28041>
- [11] Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., et al. (2022) Palette: Image-to-Image Diffusion Models. *Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings*, Vancouver, 7-11 August 2022, 1-10. <https://doi.org/10.1145/3528233.3530757>
- [12] Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R. and Van Gool, L. (2022) RePaint: Inpainting Using Denoising Diffusion Probabilistic Models. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 11461-11471. <https://doi.org/10.1109/cvpr52688.2022.01117>