

基于区域一致性注意力的单幅图像阴影去除算法

黄鑫庆

北京建筑大学理学院, 北京

收稿日期: 2026年3月22日; 录用日期: 2026年4月12日; 发布日期: 2026年4月24日

摘要

单幅图像去阴影是计算机视觉领域的重要研究任务。针对现有深度学习去阴影算法在处理复杂阴影时, 极易产生阴影区与非阴影区特征相互干扰, 从而导致边界色彩不一致和局部结构失真的问题, 本文提出了一种融合区域一致性注意力的生成式图像去阴影算法。首先, 引入具备强大数据分布拟合能力的扩散模型作为基线框架, 将去阴影过程建模为条件引导的逐步去噪生成过程。其次, 为了消除非阴影背景与核心阴影区在特征提取阶段的相互污染, 在网络的编码器与解码器阶段创新性地设计了区域一致性注意力模块(RCAM)。该模块利用阴影掩码(Mask)对自注意力机制的计算范围进行严格的空域约束, 确保特征聚合仅在各自区域内部进行, 有效提升了光照恢复的全局色彩一致性。在ISTD数据集上的大量对比实验表明, 本文方法在峰值信噪比(PSNR)和结构相似性(SSIM)等客观指标上均优于主流对比算法, 能够生成色彩自然、边界过渡平滑的高质量无阴影图像。

关键词

阴影去除, 扩散模型, 区域一致性, 注意力机制, 图像修复

Single Image Shadow Removal Algorithm Based on Region-Consistent Attention

Xinqing Huang

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

Received: March 22, 2026; accepted: April 12, 2026; published: April 24, 2026

Abstract

Removing shadows from a single image is an important research task in the field of computer

vision. In response to the problem that existing deep learning shadow removal algorithms are prone to interference between shadow and non-shadow features when dealing with complex shadows, resulting in inconsistent boundary colors and local structural distortion, this paper proposes a generative image shadow removal algorithm that integrates region-consistent attention. Firstly, a diffusion model with strong data distribution fitting ability is introduced as the baseline framework to model the shadow removal process as a condition-guided gradual denoising generation process. Secondly, in order to eliminate the mutual contamination between non-shadow background and core shadow area in the feature extraction stage, a region consistent attention module (RCAM) was innovatively designed in the encoder and decoder stages of the network. This module uses shadow masks to strictly constrain the computational range of the self-attention mechanism, ensuring that feature aggregation only occurs within their respective regions, effectively improving the global color consistency of lighting restoration. A large number of comparative experiments on the ISTD dataset show that our method outperforms mainstream comparison algorithms in objective indicators such as peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), and can generate high-quality shadowless images with natural colors and smooth boundary transitions.

Keywords

Shadow Removal, Diffusion Models, Region Consistency, Attention Mechanism, Image Restoration

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

阴影是由于光源被物体遮挡而在自然场景中形成的普遍光学现象。虽然阴影在真实世界中极为常见，但它们会显著降低图像的视觉质量，并导致图像特征表示发生畸变。因此，阴影的存在严重制约了诸如文本识别、目标分割以及视觉跟踪等诸多下游高级计算机视觉任务的性能。实现高效、高保真的图像去阴影，不仅是改善人类视觉感知的关键，也是提升复杂视觉系统鲁棒性的必要前提。尽管深度学习技术近年来极大地推动了去阴影领域的发展，但该任务仍然充满挑战。由于光照条件的多样性、遮挡物几何形状的复杂性以及阴影边界的模糊性，现有方法在复原大面积阴影时，往往容易出现色彩不一致、结构失真和过渡不自然等问题。近年来，虽然去噪扩散概率模型(DDPMs)在图像修复与生成任务中取得了突破性进展，但由于阴影区域存在严重的结构遮挡和纹理丢失，直接采用常规的扩散模型进行去阴影极易导致边界过渡生硬和感知不一致。其核心痛点在于：标准的深度神经网络(尤其是基于全局注意力的模型)在进行特征提取时，往往会将阴影像素与非阴影像素进行无差别的特征聚合，这种“跨区域的特征污染”是导致色彩断层和结构伪影的直接原因。为了解决上述问题，本文提出了一种融合区域一致性注意力(RCAM)的生成式去阴影算法。本文的核心贡献如下：

- 1) 构建了一个基于残差参数化扩散模型的去阴影框架，利用扩散模型的强大生成先验来恢复阴影区域的缺失纹理。
- 2) 创新性地提出了区域一致性注意力模块(RCAM)。该模块通过引入掩码引导的软调制机制，严格限制注意力计算发生在同质区域(阴影或非阴影)内部，从而有效抑制了区域间的特征干扰。
- 3) 在 ISTD 数据集上的实验证明，本文方法在保持结构保真度和区域色彩一致性方面取得了显著优势。

2. 相关工作

2.1. 传统与基于深度学习的去阴影方法

早期的阴影去除方法主要依赖于手工提取的先验特征或物理光照模型[1]-[3]。然而,由于现实世界光照条件的复杂性,这些基于简单假设的传统方法往往难以泛化。随着深度学习的兴起,基于卷积神经网络(CNN)和生成对抗网络(GAN)[4]的方法成为了主流。例如,DSC[5]网络利用方向感知的空间上下文特征进行阴影处理;Mask-ShadowGAN[6]则通过循环一致性设计,摆脱了对成对训练数据的严格依赖。尽管取得了显著进展,但大多数CNN框架在长距离特征建模上存在局限,导致大面积阴影区域容易出现色彩漂移和纹理模糊。

2.2. 扩散模型在图像恢复中的应用

去噪扩散概率模型[7]因其稳定的训练过程和卓越的分布建模能力,在图像修复领域展现出巨大潜力。其中,残差去噪扩散模型(RDDM)[8]通过解耦残差与噪声扩散过程,为图像恢复任务提供了一个强有力的基线。然而,常规的扩散调控策略在面对阴影特有的模糊边界和非均匀光照时,仍显不足。因此,本文致力于在扩散模型的基础上,引入强有力的区域一致性约束。

3. 本文方法

3.1. 整体网络架构

本文提出了一种专为去阴影任务设计的区域一致性扩散框架,其整体网络架构如图1所示。该框架建立在残差扩散公式的基础之上,目标图像被显式地建模为输入阴影图像与一个可学习残差分量的叠加。这种残差建模方式使得网络能够专注于估计受阴影影响的差异信息,同时最大程度保留光照良好区域的特征。

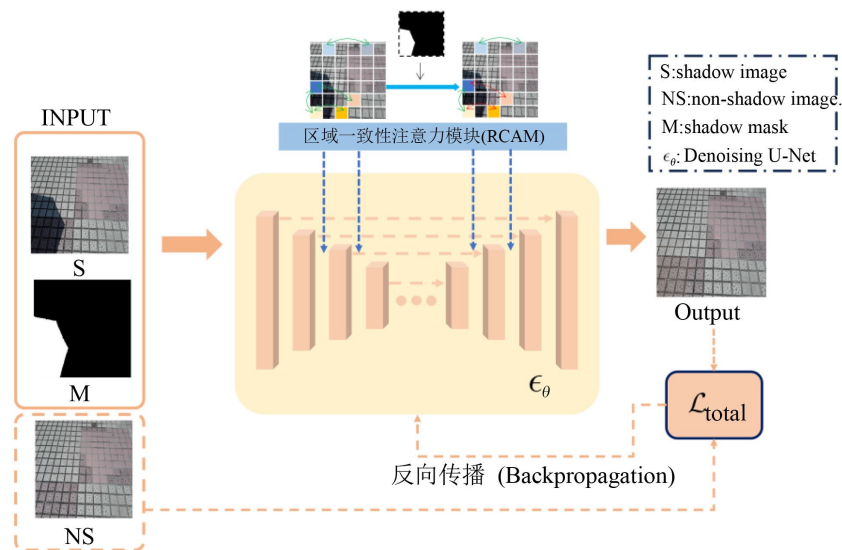


Figure 1. Overall architecture of a diffusion model that integrates region-consistent attention

图 1. 融合区域一致性注意力的扩散模型整体架构

网络的主体架构采用 U-Net 结构(表示为 ϵ_θ)。在特征编码阶段,输入的阴影图像与二值阴影掩码共同送入编码器以提取多尺度特征。为了防止在特征下采样和上采样过程中阴影区域与非阴影区域发生相

互污染, 本文在编码器和解码器的中间层(具体为第二和第三层)嵌入了区域一致性注意力模块(RCAM)。解码器最终融合多尺度特征, 输出用于扩散逆向过程的噪声预测值以及光照残差图, 从而生成最终的无阴影图像。

特别地, 为了清晰说明系统的运行机制, 图 1 对模型的训练与推理流程进行了显式区分。图中实线路径代表通用的推理阶段(Inference), 网络仅依赖输入的阴影图像与掩码, 即可通过前向传播直接生成去阴影结果; 而虚线路径则代表仅在训练阶段(Training)激活的监督闭环。在训练过程中, 生成的去阴影图像将与额外引入的真实无阴影标签(如图左下角虚线框所示)共同送入损失计算模块, 以计算总损失(\mathcal{L}_{total}), 并最终通过反向传播机制优化网络参数。

3.2. 区域一致性注意力模块(RCAM)

在常规的 Transformer 架构中, 标准的自注意力机制(Self-Attention)允许特征图中的每个空间标记(Token)与所有其他标记进行无差别的全局交互。尽管这种机制极大地扩展了网络的感受野, 但在单幅图像去阴影任务中却存在明显的缺陷: 由于阴影区域与非阴影区域在亮度和色彩分布上存在显著差异, 无差别的全局注意力会促使阴影像素错误地聚合非阴影区域的特征, 进而导致复原图像在边界处出现严重的“色彩渗漏(Color Bleeding)”与结构伪影。

为了从根本上消除这种跨区域的特征干扰, 本文设计了区域一致性注意力模块(RCAM), 其内部网络结构如图 2 所示。该模块的核心思想是引入阴影掩码(Mask)作为空间先验, 通过软调制权重矩阵对注意力分数的计算范围进行严格约束, 强制特征的聚合仅在同质区域(即阴影内部或非阴影内部)闭环完成。

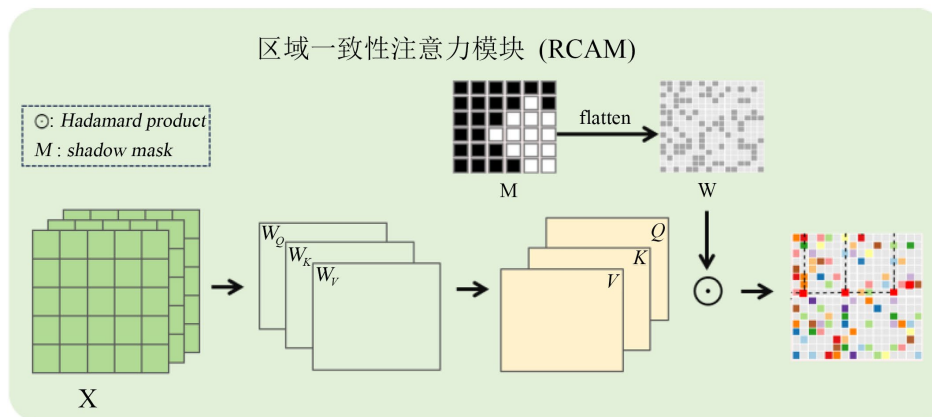


Figure 2. Region-consistent attention module (RCAM) structure diagram

图 2. 区域一致性注意力模块(RCAM)结构图

具体而言, 给定经编码器提取的深层输入特征图 $X \in \mathbb{R}^{B \times C \times H \times W}$, RCAM 首先利用三组独立的 1×1 卷积层对其进行线性映射, 生成查询矩阵(Query)、键矩阵(Key)和值矩阵(Value): $Q = W_q \cdot X$, $K = W_k \cdot X$, $V = W_v \cdot X$ 。其中, W_q, W_k, W_v 分别代表对应的卷积可学习参数。随后, 计算原始的全局注意力相关性分数矩阵 A :

$$A = \frac{Q \cdot K^T}{\sqrt{d_k}} \quad (1)$$

其中, d_k 为通道维度的缩放因子, 用于防止梯度消失。为了阻隔异质区域间的特征流动, 本文引入了与特征图分辨率相匹配的二值阴影掩码 $M \in \{0, 1\}^{B \times 1 \times H \times W}$, 并将其展平为一维序列 $m \in \{0, 1\}^{B \times N}$ (其中

$N = H \times W$)。基于展平后的掩码标签, RCAM 为特征空间中的任意两个像素点对 (i, j) 动态构建了一个软调制权重矩阵 $W \in \mathbb{R}^{N \times N}$, 其计算规则如下:

$$W_{i,j} = \begin{cases} 1.0, & \text{if } |m_i - m_j| < \tau \\ \gamma, & \text{otherwise} \end{cases} \quad (2)$$

其中, τ 为区域判别阈值(本文经验性地设置为 0.1), 用于判断两点是否属于同一物理区域; γ 为跨区域抑制惩罚系数(本文设为 0.2)。与直接将异质区域权重重置为零的硬截断策略相比, 这种软调制机制不仅有效削弱了错误特征的引入, 还保留了微弱的跨界信息交互, 有助于维持阴影边界的自然平滑过渡。在获取调制矩阵 W 后, 将其与 Softmax 激活后的注意力分数进行哈达玛乘积(Hadamard Product)操作, 实现对注意力分布的逐元素重新加权:

$$Attn^{mod} = \text{Softmax}(A) \odot W \quad (3)$$

最后, 利用调制后的注意力权重矩阵对值矩阵 V 进行加权聚合, 并通过引入残差连接(Residual Connection)输出最终的特征图 Y :

$$Y = Attn^{mod} \cdot V + X \quad (4)$$

通过上述设计, RCAM 模块在特征传递的底层物理逻辑上实现了阴影与背景的解耦, 为后续扩散模型的精准噪声预测与高质量纹理生成奠定了坚实的基础。

3.3. 损失函数

为了提升模型在复原过程中的语义保真度与结构一致性, 本文在基础扩散模型损失之外, 引入了针对去阴影任务的联合损失函数, 其完整定义如下:

$$\mathcal{L}_{total} = \mathcal{L}_{base} + \lambda_1 \mathcal{L}_{region} + \lambda_2 \mathcal{L}_{vgg} + \lambda_3 \mathcal{L}_{grad} \quad (5)$$

其中, \mathcal{L}_{base} 为扩散模型的残差与噪声预测损失。 \mathcal{L}_{region} 为区域一致性损失, 用于强制约束预测阴影区与真实阴影区在均值 μ 和标准差 σ 上的统计学一致性, 其公式为:

$$\mathcal{L}_{region} = \alpha \left\| \mu(\hat{I}; S) - \mu(I; S) \right\| + \beta \left\| \sigma(\hat{I}; S) - \sigma(I; S) \right\| \quad (6)$$

其中, \mathcal{L}_{grad} 为梯度一致性损失, 用于锐化边界并对齐结构细节。 \mathcal{L}_{vgg} 为感知损失, 利用预训练 VGG 网络提取深层特征, 以提升人类视觉感知的真实感。

4. 实验结果与分析

4.1. 实验环境与数据集

本文在公开的 ISTD 基准数据集上进行了广泛的实验。ISTD 包含 1330 张训练图像和 540 张测试图像(均包含对应的阴影图、掩码及真值图)。所有输入图像在训练和测试阶段均统一缩放至 256×256 像素。网络训练采用 AdamW 优化器, 总迭代步数为 80,000 步, 初始学习率设置为 8×10^{-5} 。客观评价指标采用平均绝对误差(MAE, 越小越好)、峰值信噪比(PSNR, 越大越好)和结构相似性(SSIM, 越大越好)。

4.2. 定量结果对比

为了客观评估本文方法的性能, 将其与近年来主流的去阴影模型进行了定量对比, 包括 Mask-ShadowGAN、DSC、DC-ShadowNet [9] 以及基于扩散模型的强基线 RDDM。实验结果如表 1 所示。

从定量数据可以看出, 本文方法在整体区域(All)取得了 31.17 dB 的最高 PSNR 值。与同样基于扩散架构的 RDDM 模型相比, 本文模型在阴影核心区域(S.)的 PSNR 从 36.63 dB 提升至 37.12 dB。这一显著提升充分证明, 通过显式引入 RCAM 模块进行区域隔离, 能够极大地减少复原过程中的色彩漂移误差, 实现高精度的像素级恢复。

Table 1. Comparison table of ISTD dataset
表 1. ISTD 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)
MaskShadowGan	12.67	6.68	7.41	-	-	-	-	-	-
DSC	9.48	6.14	6.67	0.967	-	-	33.45	-	-
DC-ShadowNet	11.43	5.81	6.57	0.976	0.958	0.922	31.69	28.99	26.38
ST-CGGAN [10]	9.99	6.05	6.65	0.981	0.951	0.929	33.74	29.51	27.44
FusionNet [11]	7.77	5.56	5.92	0.975	0.880	0.945	34.71	28.61	27.19
BMNet [12]	7.60	4.59	5.02	0.988	0.976	0.959	35.61	32.80	30.28
DMTN [13]	7.00	<u>4.28</u>	<u>4.72</u>	0.990	<u>0.979</u>	0.965	35.83	<u>33.01</u>	30.42
RDDM	6.92	4.36	4.77	0.987	0.978	0.961	36.63	32.95	<u>30.78</u>
Ours	6.56	4.11	4.49	<u>0.989</u>	0.980	<u>0.964</u>	37.12	33.35	31.17

4.3. 定性视觉分析

图 3 展示了本文方法与其他模型在复杂场景下的视觉对比。观察视觉对比结果可以发现, 传统的方法(如 DC-ShadowNet)由于缺乏有效的区域解耦机制, 在处理如瓷砖路面等存在高对比度边界的区域时, 容易残留明显的阴影轮廓(红框所示)。相比之下, 本文方法(蓝框所示)得益于区域一致性注意力(RCAM)的强力约束, 能够生成极其干净的去阴影结果。复原后的区域不仅边界过渡自然, 而且底层纹理与周边非阴影环境高度融合, 展现出了极强的鲁棒性。

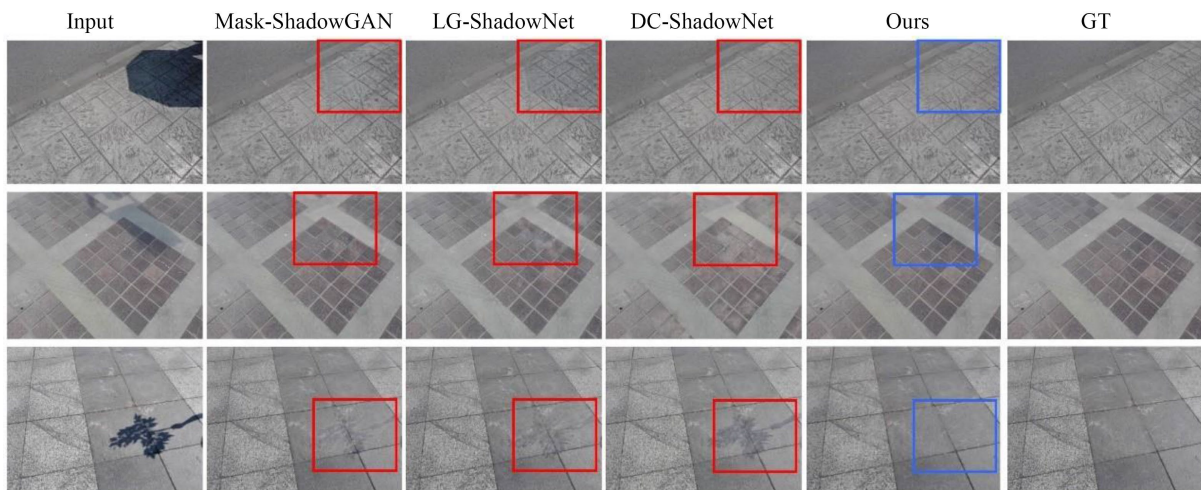


Figure 3. Visual comparison results with other methods
图 3. 与其他方法的视觉对比结果

4.4. 消融实验

为了验证 RCAM 模块在网络架构中的核心作用，本文在相同实验设置下开展了模块消融实验。我们对比了纯扩散模型基线(Base Model)与添加了 RCAM 模块(+RCAM)的模型性能(见表 2)。

Table 2. RCAM module ablation experiment
表 2. RCAM 模块消融实验

Method	MAE↓ (S.)	MAE↓ (NS.)	MAE↓ (All)	SSIM↑ (S.)	SSIM↑ (NS.)	SSIM↑ (All)	PSNR↑ (S.)	PSNR↑ (NS.)	PSNR↑ (All)
Base	6.92	4.36	4.77	0.987	0.978	0.961	36.63	32.95	30.78
+RCAM	6.80	4.25	4.58	0.988	0.982	0.964	37.02	33.25	31.07

实验数据显示，在缺乏 RCAM 的情况下，基线模型在全局重构上的表现较弱(PSNR 仅为 30.78 dB)。而引入 RCAM 模块后，模型的结构一致性得到了显著改善，全局 PSNR 提升至 31.07 dB。这一结果有力地佐证了：在深度特征传递过程中，强制约束同质区域内部的注意力交互，是降低跨区域干扰、提升图像恢复质量的关键设计。

5. 总结

综上所述，本文所提出的区域一致性扩散框架巧妙地结合了残差学习机制与扩散模型强大的生成能力。通过在关键特征层级引入 RCAM 模块，网络有效克服了传统去阴影方法中常见的跨区域特征污染与色彩漂移问题。同时，这种严格解耦训练与推理双流程的系统设计，既保证了模型在优化阶段能够充分利用真实物理标签(Ground Truth)进行精准的误差计算与参数迭代，又确保了其在实际推理阶段的高效性与独立性。整体而言，该架构为复杂场景下的高保真阴影去除提供了一个严密且鲁棒的生成范式，并为后续的高质量图像重建奠定了坚实基础。

参考文献

- [1] Finlayson, G.D., Hordley, S.D., Cheng, L. and Drew, M.S. (2006) On the Removal of Shadows from Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **28**, 59-68. <https://doi.org/10.1109/tpami.2006.18>
- [2] Finlayson, G.D., Hordley, S.D. and Drew, M.S. (2002) Removing Shadows from Images. In: Heyden, A., et al., Eds., *European Conference on Computer Vision*, Springer, 823-836. https://doi.org/10.1007/3-540-47979-1_55
- [3] Gryka, M., Terry, M. and Brostow, G.J. (2015) Learning to Remove Soft Shadows. *ACM Transactions on Graphics*, **34**, 1-15. <https://doi.org/10.1145/2732407>
- [4] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2020) Generative Adversarial Networks. *Communications of the ACM*, **63**, 139-144. <https://doi.org/10.1145/3422622>
- [5] 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, 18-22 June 2018, 7454-7462. <https://doi.org/10.1109/cvpr.2018.00778>
- [6] 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, 27-28 October 2019, 2472-2481. <https://doi.org/10.1109/iccv.2019.00256>
- [7] Ho, J., Jain, A. and Abbeel, P. (2020) Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*, 6-12 December 2020, 6840-6851.
- [8] Liu, J., Wang, Q., Fan, H., Wang, Y., Tang, Y. and Qu, L. (2024) Residual Denoising Diffusion Models. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 2773-2783. <https://doi.org/10.1109/cvpr52733.2024.00268>
- [9] 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>
- [10] Wang, J., Li, X. and Yang, J. (2018) Stacked Conditional Generative Adversarial Networks for Jointly Learning Shadow Detection and Shadow Removal. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-22 June 2018, 1788-1797. <https://doi.org/10.1109/cvpr.2018.00192>
- [11] 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>
- [12] Zhu, Y., Huang, J., Fu, X., Zhao, F., Sun, Q. and Zha, Z. (2022) Bijective Mapping Network for Shadow Removal. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 5627-5636. <https://doi.org/10.1109/cvpr52688.2022.00554>
- [13] Liu, J., Wang, Q., Fan, H., Li, W., Qu, L. and Tang, Y. (2023) A Decoupled Multi-Task Network for Shadow Removal. *IEEE Transactions on Multimedia*, **25**, 9449-9463. <https://doi.org/10.1109/tmm.2023.3252271>