

中图分类号: TP18; TP391.41 文献标识码: A 文章编号: 1006-8961(2026)07-2660-14

论文引用格式: Wang J C, Zhou S B, Pan F, Jiang X Q, Huang R and Xie Y H. 2026. Two-stage mask-guided network for car surface specular high-light removal by using the shifted window attention mechanism. Journal of Image and Graphics, 31(7):2660-2673(王嘉成, 周树波, 潘峰, 蒋学芹, 黄荣, 谢颖华. 2026. 窗口移位注意力的两阶段掩码引导汽车表面高光去除网络. 中国图象图形学报, 31(7):2660-2673)[DOI: 10.11834/jig.250353]

窗口移位注意力的两阶段掩码引导 汽车表面高光去除网络

王嘉成, 周树波, 潘峰, 蒋学芹, 黄荣, 谢颖华*

东华大学信息与智能科学学院, 上海 201600

摘要: 目的 镜面反射会干扰视觉感知系统,降低目标检测和三维重建等任务的性能。尽管现有高光去除方法已有进展,但大多依赖合成数据训练,难以有效处理真实场景中的反射。针对这一问题,本文构建了多光照条件下的汽车高光数据集,并提出掩码引导的高光去除网络(mask guided Transformer U-Net, MG-TransUNet),显著提升高光去除效果。**方法** 首先,构建了汽车高光数据集,覆盖多种光照条件及不同强度、分布的高光图像对,并提供高光掩码。并且提出协同优化的两阶段网络架构,第1阶段采用轻量级U形网络(U-shaped network, U-Net)实现高光检测,生成高光区域掩码;第2阶段将掩码作为空间先验,指导基于改进型窗口移位注意力(shifted window attention Transformer, SWAT)的U-Net高光去除网络。通过端到端联合训练,实现检测与去除网络的参数协同优化。**结果** 本文在自建数据集和3个公开数据集PSD(paired specular-diffuse image dataset)、SHIQ(specular highlight image quadruples)和SSHR(synthetic specular highlight removal dataset)上进行了系统评估,与8种先进方法进行对比。在PSD数据集上,本文方法以26.37 dB的峰值信噪比(peak signal-to-noise ratio, PSNR)位列第2;在SHIQ数据集上以35.24 dB的PSNR值优于第2名(34.13 dB)1.11 dB;在SSHR数据集上以36.54 dB的PSNR值领先第2名(36.16 dB)0.38 dB;在自建数据集上以25.14 dB的PSNR值优于第2名(25.03 dB)0.11 dB。此外,消融实验表明掩码引导网络和SWAT模块均能提升高光去除性能。**结论** 本文构建了面向真实场景的汽车高光数据集,提出了基于窗口移位注意力机制的掩码引导网络(MG-TransUNet),本文方法在自建数据集上性能达到最优,并在多个公开数据集上取得了具有竞争力的表现。代码地址:https://github.com/chengWULUQI/MG-Trans_Unet。

关键词: 高光检测;高光去除;高光数据集;U-Net模型;窗口移位注意力机制

Two-stage mask-guided network for car surface specular highlight removal by using the shifted window attention mechanism

Wang Jiacheng, Zhou Shubo, Pan Feng, Jiang Xueqin, Huang Rong, Xie Yinghua*

School of Information and Intelligent Science, Donghua University, Shanghai 201600, China

Abstract: **Objective** Specular highlight detection and removal, with the primary objective of accurately identifying high-light regions exhibiting mirror-like reflection characteristics and precisely recovering the underlying pixel information, constitute a fundamental research topic in computer vision and image processing. This technology demonstrates considerable

收稿日期: 2025-08-15; 修回日期: 2026-01-09; 预印本日期: 2026-01-16

* 通信作者: 谢颖华 yh_xie@dhu.edu.cn

基金项目: 国家自然科学基金项目(62001099); 中央高校基本科研业务费专项资金项目(2232023D-30)

Supported by: National Natural Science Foundation of China(62001099); Fundamental Research Funds for the Central Universities(2232023D-30)

practical value across various visual applications, including object detection, medical image analysis, industrial defect identification, and 3D imaging. From an optical perspective, the specular highlight phenomenon originates from the directional reflection behavior of incident light on smooth non-Lambertian surfaces, and its physical mechanism is theoretically describable through Fresnel equations. Traditional highlight removal methods typically rely on strict physical assumptions, which primarily encompass three classical approaches: 1) chromaticity space analysis-based techniques, 2) illumination estimation-based algorithms, and 3) polarization information-utilizing solutions. However, these methods exhibit limited adaptability to complex and variable highlight removal requirements in real-world scenarios because of their constraints regarding specific scenes or object types. By contrast, deep learning-based highlight detection and removal techniques adopt a data-driven approach to automatically learn the complex nonlinear mapping relationships between highlight reflections and scene content, thereby effectively circumventing the dependency on prior assumptions inherent in traditional methods. The shifted window attention mechanism demonstrates unique advantages in highlight removal tasks and effectively addresses issues, such as texture information loss and color distortion, caused by specular highlights. This mechanism integrates local window partitioning with cross-window interaction, thus reducing computational complexity while maintaining global modeling capability. Building upon the exceptional modeling capability demonstrated by the shifted window attention mechanism during the feature extraction stage, this study proposes a two-stage mask-guided highlight removal network architecture. The proposed framework employs the shifted window attention mechanism to extract multiscale cross-channel features and utilizes highlight masks as prior knowledge to guide subsequent pixel recovery in highlight regions, thereby achieving highlight localization and removal. Notably, existing deep learning methods face challenges regarding insufficient training data. Current publicly available highlight datasets predominantly employ synthetic generation or laboratory-controlled lighting acquisition methods, failing to meet the requirements of real-world applications. To address this limitation, this study additionally constructs an in-situ automotive highlight dataset captured under multiple illumination conditions. This dataset has the following characteristics: 1) coverage of diverse natural illumination conditions, 2) inclusion of complex background interference, and 3) precise pixel-level annotations, thereby providing a data foundation for enhancing model generalization performance in real-world scenarios. **Method** In this study, we construct a real highlight dataset with cars as the main object, which is based on common cars in daily life with different highlight regions and intensities. Each pair of highlight images has its corresponding real highlight mask image. For highlight detection and removal, we propose a mask-guided two-stage highlight removal network (MG-TransUNet). This network explicitly divides the highlight detection and removal task into two phases. In the first phase, we train with a lightweight U-Net architecture to extract the highlight regions in the highlight images and generate the highlight mask map. In the second phase, a shifted window attention Transformer network that is based on the U-Net architecture is designed to recover the diffuse images of highlight regions by using the predicted highlight mask map as an a priori guide. Our experimental dataset is selected from three major public datasets (i. e., paired specular-diffuse image dataset (PSD), specular highlight image quadruples (SHIQ) and synthetic specular highlight removal dataset (SSHR)) with our proposed automotive dataset. Among the four datasets, only SHIQ and automobile datasets provide highlight masks. To make our method trainable on the PSD and SSHR datasets, we subtract and threshold the highlight images and the corresponding specular no-highlight images from the two datasets to obtain the highlight mask for each pair of images. The highlight detection network is trained with the highlight images and masks at the first stage, and the obtained highlight masks are used as a priori information to guide the highlight removal task at the second stage. Our mask-guided highlight removal network is implemented by the Pytorch toolbox. The input images and ground truth maps are resized to 256×256 for training, the momentum parameter is set to 0.99, and the initial learning rate is set to 2×10^{-4} , which is reduced using a decay coefficient of 0.8 every 5 cycles until it reaches 1×10^{-6} . The stochastic gradient descent learning process is accelerated using an NVIDIA GeForce GTX3090 GPU device, and 100 iterations requires about half a day. **Result** Experiments are conducted on the proposed automotive dataset and three publicly available datasets to compare eight state-of-the-art methods. The quantitative evaluation metrics include the structural similarity index measure and peak signal-to-noise ratio (PSNR). Experimental results show that the PSNR value ranks second on the PSD dataset. It is improved by 1.11 compared with that in the model with second-best performance on the SHIQ dataset and by 0.38 compared with that in the model with the second-best performance on the SSHR dataset. Comparative

experiments are also performed on automotive datasets to validate the effectiveness of the proposed dataset. **Conclusion** We propose a car highlight dataset for real scenarios and a mask-guided network incorporating a window-shift attention mechanism to perform the highlight removal task. The experimental results show that our highlight removal network can accurately recognize highlight regions and remove highlight images. The code link: https://github.com/chenWULUQI/MG-Trans_Unet.

Key words: highlight detection; highlight removal; highlight dataset; U-Net model; shifted window attention mechanism

0 引言

镜面高光反射是光线在非朗伯物体光滑表面上发生定向反射时产生的光学现象,其物理机制可由菲涅尔方程(Born和Wolf,1999)描述。这种光学效应在计算机视觉领域会引发颜色失真、纹理细节丢失和几何形变等负面效应,从而显著降低目标检测(Fu等,2024)、医疗图像分割(Li等,2024;Zhu等,2024;Zhang等,2023)、缺陷检测(Jiang等,2024)及三维成像(Li等,2023)等视觉任务性能。因此,高光区域检测与去除已成为计算机视觉领域的重要挑战。

现有高光检测与去除方法可分为传统方法和深度学习方法两类。传统检测方法主要基于光学物理模型与低层图像特征的结合,其核心基于3个假设:1)理想白光光源;2)二色反射模型(Shafer,1985)的线性可分离性;3)表面材质的均匀反射特性。然而,这类方法存在较大的局限性。首先,由于缺乏语义理解能力,难以准确区分镜面反射与高亮漫反射材质;其次,在汽车表面检测等实际应用场景中,曲面多角度反射与环境光的复杂相互作用往往导致误检(如将白色标识误判为高光)和漏检问题。典型传统高光去除方法主要包括基于色度分析(Lin等,2002;Shen等,2008)、照明估计(Finlayson等,2001)以及偏振信息(Lei等,2020;Wen等,2021)的方法。实际图像高光抑制结果表明,这些方法难以应用于真实世界下的高光去除任务。

近年来,基于深度学习的高光检测与去除技术取得重要进展。通过端到端学习,深度学习方法能够自动学习高光反射与场景内容之间的复杂非线性映射,减少对物理假设的依赖。在这之中,卷积神经网络(convolutional neural network, CNN)(Fu等,2020;Chen等,2024;Huang等,2022)通过局部感受野捕获高光的空间分布特征;生成对抗网络(generative adversarial network, GAN)(Muhammad等,2020;Hu等,2022;Wu等,2022;Anwer等,2023)通过对抗

训练机制改善了高光去除结果的视觉效果;Guo等人(2024)和Wu等人(2023)通过空间—通道双重注意力机制提升了高光区域定位的能力;而新兴的扩散模型(Pan等,2024)通过渐进式去噪过程进一步提高了对复杂反射现象的建模能力。然而,现有深度学习方法仍面临数据依赖性挑战。现有公开数据集主要存在两类局限:合成数据集(如SHIQ(specular highlight image quadruples)(Fu等,2021),SSHR(synthetic specular highlight removal dataset)(Fu等,2023))容易导致模型在真实场景中泛化性能下降;而真实场景数据集(如PSD(paired specular-diffuse image dataset)(Wu等,2022))又因场景单一而难以支撑大型模型训练需求。因此,构建真实场景的高质量数据集成为深度学习高光去除的一大关键。

基于此,本文构建了一个用于检测和去除汽车表面高光的汽车高光数据集。该数据集包含842对高分辨率图像,图像采集涵盖阳光直射、夜间路灯照明和地下车库灯光条件,涵盖不同类型的车辆(轿车、越野车、卡车)和不同的表面材料(金属漆、亚光饰面、玻璃),每幅图像都配有像素级高光掩码和通过偏振成像获得的无镜面反射图像。同时,本文提出了掩码引导的高光去除网络MG-TransUNet(mask guided Transformer U-Net),该网络第1阶段采用轻量级U-Net实现高光检测,生成高光区域掩码;第2阶段设计基于SWAT(shifted window attention Transformer)的U-Net架构,引入窗口移位注意力模块和掩码引导特征选择机制,以实现高光深层特征提取。本文主要贡献可以总结如下:1)构建首个多种真实光照条件下的汽车高光数据集,用于汽车表面图像的高光检测与去除研究,弥补了现有高光去除领域真实数据稀缺的不足。2)提出一种端到端两阶段网络,通过窗口移位注意力机制融合全局与局部特征,实现了高光检测与漫反射图像恢复。3)在3个主流高光去除数据集(PSD、SHIQ、SSHR)和自建汽车数据集上进行实验验证,定量结果表明本文方法相比现有方法更具优势。

1 相关工作

1.1 传统高光检测方法

传统高光检测研究主要基于阈值分割方法,其核心是通过设定光学特征阈值实现高光区域的初步定位。典型研究包括: Park 和 Kak(2003)提出截断最小二乘法检测彩色图像的镜面高光; Yang 等人(2013)在色调—饱和度—明度(hue-saturation-intensity, HSI)颜色空间提高检测精度,通过区域增长算法聚合具有相似漫反射色度的相邻像素,但未解决高光与白色材质的语义歧义问题; Akashi 和 Okatani(2015)提出基于稀疏非负矩阵分解的方法,通过非负性与稀疏性约束分离反射分量,为非负矩阵分解(non-negative matrix factorization, NMF)在高光相关任务中的应用提供了重要思路; Zhang 等人(2019)通过设定亮度阈值将最亮像素归类为高光,并用 NMF 求解高光检测问题。这类方法计算高效,但存在语义理解的缺陷,难以应用在自动驾驶、工业检测等复杂场景中。

1.2 传统高光去除方法

传统高光去除方法主要基于物理模型、颜色空间及纹理细节等先验信息对高光特征进行建模。Shen 和 Zheng(2013)通过 HSI 颜色空间从每个像素中分离出漫反射与镜面分量; Guo 等人(2018)将问题建模为 NMF,增强算法的鲁棒性; 高如新和李雪颖(2018)提出了一种改进的双边滤波去高光算法,该方法基于双色反射模型,通过阈值分割分类像素,并分别估计其最大漫反射色度,最终将色度相似度融入双边滤波的值域核,在有效抑制镜面反射的同时,较好地保持了图像边缘与颜色保真度; Fu 等人(2019)进一步提出稀疏约束的优化框架,从单幅图像估计漫反射与高光层; Hao 等人(2025)利用暗通道先验建立修正无镜面反射图像,再提取修改后图像的最大色度值指导镜面反射的去除; Shi 等人(2025a)将偏振光源与控制系统相结合,以动态调节照明并消除镜面高光。然而,这类方法存在两大局限: 1) 缺乏语义理解能力,难以区分镜面高光与高亮漫反射材质(如玻璃或金属)。2) 当场景中存在各向异性反射材质(如金属)时,分离误差显著增加。

1.3 基于深度学习的高光检测和去除方法

随着深度学习方法的发展,神经网络在高光检

测和去除领域得到了广泛应用。在高光检测领域, Fu 等人(2020)构建了真实场景高光数据集,并通过多尺度上下文对比的特征提取方法提升了不同尺度高光的检测精度; Anwer 等人(2022)设计了基于 U-Net 架构的高光分割网络(specular segmentation network, SpecSeg); Wu 等人(2023)则提出两阶段处理框架,先通过分割网络生成高光掩码,再将其作为先验信息指导后续去除网络。在高光去除领域, Hu 等人(2024)提出基于改进二色反射模型的高光去除网络,设计自适应高光感知网络,针对性解决高光去除中颜色不一致和内容还原不真实的问题; Chen 等人(2024)使用预训练模型学习金属表面的高光特征; Huang 等人(2022)提出了跨场景一致性框架,可同时处理合成图像、人脸图像等多种数据类型; Shi 等人(2025b)利用编码器—解码器结构实现联合检测与去除。近年来,生成对抗网络(GAN)和基于自注意力机制的 Transformer 架构(周丽娟和毛嘉宁, 2023)凭借其在生成质量、特征提取和长程依赖建模等方面的卓越性能,正推动该领域研究取得新的突破; Muhammad 等人(2020)提出 Spec-Net(specularity network)和 Spec-CGAN(specularity conditional generative adversarial network)网络处理低色度人脸图像; Hu 等人(2022)利用 MG-CycleGAN(mask-guided cycle-consistent generative adversarial network)网络实现未配对数据训练; Anwer 等人(2023)结合多偏振角度信息识别高光像素; 最新研究探索了混合架构: Guo 等人(2024)提出双通道混合注意力网络(dual-hybrid attention network for specular highlight removal, DHAN-SHR),引入新的混合注意机制捕获和处理不同尺度的特征信息; Wu 等人(2023)设计 Unet-Transformer 网络实现高光去除; Pan 等人(2024)首次将扩散模型应用到高光去除领域。这些方法提升了高光去除网络的图像恢复效果,但在不可预测的成像环境(如强烈日光照射)中性能下降。因此,现有方法面临两大挑战: 1) 难以处理大面积高光及高强度高光; 2) 图像恢复后高光与背景区域间颜色不一致。这些问题仍待后续研究解决。

高光去除领域的数据集发展经历了以下关键阶段: Fu 等人(2020)率先提出了真实世界高光数据集,但仅提供高光掩码而缺少真实无高光图像,限制了其在去除任务中的应用; Fu 等人(2021)通过合成方法扩展了日常闪亮材料的高光—无高光图像对;

Wu 等人(2022)利用偏振成像技术创建首个多角度图像数据集 PSD;Fu 等人(2023)进一步提出基于合成渲染的 SSHR 数据集;Guo 等人(2024)将 PSD、SHIQ 和 SSHR 数据集合并构建了新数据集,并用于训练 DHAN-SHR 网络;Zhang 等人(2024)开发了多光源捕获系统,构建了大规模 NSH(natural specular highlight dataset)数据集。然而,这些数据集存在两类局限:1)基于合成渲染的数据集(如 SHIQ、

SSHR)存在域偏移问题;2)实验室受控环境下采集得到的数据集(如 PSD、NSH)难以模拟真实复杂光照。因此,未来研究需着力构建真实光照条件下的数据集。

1.4 现有公开数据集

现有高光去除领域的公开数据集有 PSD、SHIQ 和 SSHR,而 NSH 数据集尚未公开。表 1 总结了这些数据集的基本信息。

表 1 不同数据集的基本信息

Table 1 Basic information on different datasets

数据集	数量	高光掩码	数据类型	发表年份	SSIM	采集光源
SHIQ	16 K	√	合成	2021	0.925 8	无
PSD	11.7 K	×	真实	2021	0.857 0	灯光
SSHR	130 K	×	合成	2023	0.892 6	无
NSH	30 K	×	真实	2024	—	灯光
本文	842	√	真实	2026	0.734 1	自然光及灯光

注:√和×分别表示包含和不包含,“—”表示无相关信息。

1)PSD数据集。该数据集在实验室可控光源条件下,通过多角度偏振成像采集了13 380组真实场景高光图像对,解决了合成数据的域偏移问题。然而该数据集存在两大局限:(1)尽管样本总量达10 000对,但其在不同偏振角度对同一场景重复拍摄,导致该数据集场景多样性不足;(2)受限于实验室光照环境,光照条件和材质类型单一。

2)SHIQ数据集。该数据集从现有闪亮材料数据集中选取25 000幅高光图像,使用鲁棒主成分分析方法生成对应的无高光图像,最终通过人工评价方式来确保数据质量。虽然这种合成方法能够提供丰富的材质类型和多样场景,但仍存在域偏移问题。

3)SSHR数据集。该数据集基于改进的二值反射模型原理并用开源渲染软件生成无高光图像,包含不同材料表面和光照条件的图像。但其仅包含前景分割图像(背景为纯色),缺少真实场景的复杂背景信息。

4)NSH数据集。该数据集和PSD数据集均采用偏振成像技术获取无镜面高光图像,但NSH数据集使用多光源图像捕获系统,解决了PSD数据集中图像对错位和非高光区域亮度不一致等问题。

综上所述,现有高光数据集采集方式主要有两种:一是基于合成渲染(如SHIQ、SSHR);二是在实

验室受控光源下采集(如PSD、NSH)。这些数据集均缺乏自然光照条件下的真实数据。因此,本研究构建了首个包含自然光照场景的汽车高光数据集。该数据集采集环境包括阳光直射、地下车库LED(light-emitting diode)照明及停车场混合光照。如表1所示,在现有的公共数据集中,只有汽车数据集考虑到了在阳光直射条件下的物体表面高光。本文还比较了现有几个数据集的高光图像与无高光图像之间的结构相似度(structural similarity index, SSIM)指标。结果显示,汽车数据集有着最低的SSIM值,说明汽车数据集的高光区域更为复杂。

2 数据集构建

2.1 理论背景

从光学本质而言,光具有偏振特性,但自然光通常表现为非偏振状态,这是因为其由大量偏振方向随机分布的光波叠加而成(Born和Wolf, 1999)。因此,通过添加线性偏振片,可将非偏振光转换为偏振光。由二值反射模型(Shafer, 1985)可知,相机捕获到的图像I可表示为

$$I = I_{\text{spec}} + I_{\text{diff}} \tag{1}$$

式中, I_{spec} 是镜面反射分量, I_{diff} 是漫反射分量。当在

相机前放置偏振片时,由马吕斯定律可知,通过偏振器的镜面反射分量强度 I'_{spec} 与偏振角 θ 之间的关系可表示为

$$I'_{\text{spec}} = I_{\text{spec}} \times \cos^2 \theta \tag{2}$$

最终,在偏振器后捕捉到的图像 I 可表示为

$$I = I_{\text{spec}} \times \cos^2 \theta + I_{\text{diff}} \tag{3}$$

式中,当 θ 等于 $\frac{\pi}{2}$ 时, $I = I_{\text{diff}}$,此时捕获到的图像滤除了镜面反射分量 I_{spec} 。因此,使用偏振相机调整偏振角度可以获取高光抑制的图像。

2.2 数据采集

采用数码相机结合高消光比圆形偏振片实现偏振去高光图像的采集,通过三脚架固定以减小图像配对误差,整个数据集采集系统由图 1 所示。

与在实验室受控环境采集的 PSD 和 NSH 数据集不同,本文在真实场景中拍摄汽车在阳光直射、路灯照明和地下车库 LED 光源等多种光照条件产生的高光图像。每组图像对的采集都按照先拍摄原始高光图像、再旋转偏振片获取对应的无高光图像的顺序,最终获得 842 对高光图像对。如图 2 所示,汽车数据集涵盖了金属曲面车身的多向异性高光和车窗玻璃的高光—倒影混合现象等复杂情况。与合成数据集相比,本数据集的高光图像对表现出更大的

色度偏移和更高的纹理复杂度,为高光去除算法的评估提供了更接近实际应用的测试基准。



图 1 数据集采集系统
Fig. 1 Data collection system



图 2 汽车数据集中不同光照条件下的图像对
Fig. 2 Images in car datasets under different lighting conditions
((a) natural light; (b) street lamp illumination; (c) underground garage light)

3 方法

3.1 总体架构

如图 3 所示, MG-TransUnet 是一个端到端两阶

段网络, 网络采用双分支协同设计。先将高光图像输入至检测分支, 通过轻量级 U-Net 生成像素级高光掩码; 去除分支在通道维度上将高光图像与分割掩码拼接, 融合原始特征与高光区域特征, 并构建增强型 U 型架构以实现高光区域图像恢复。

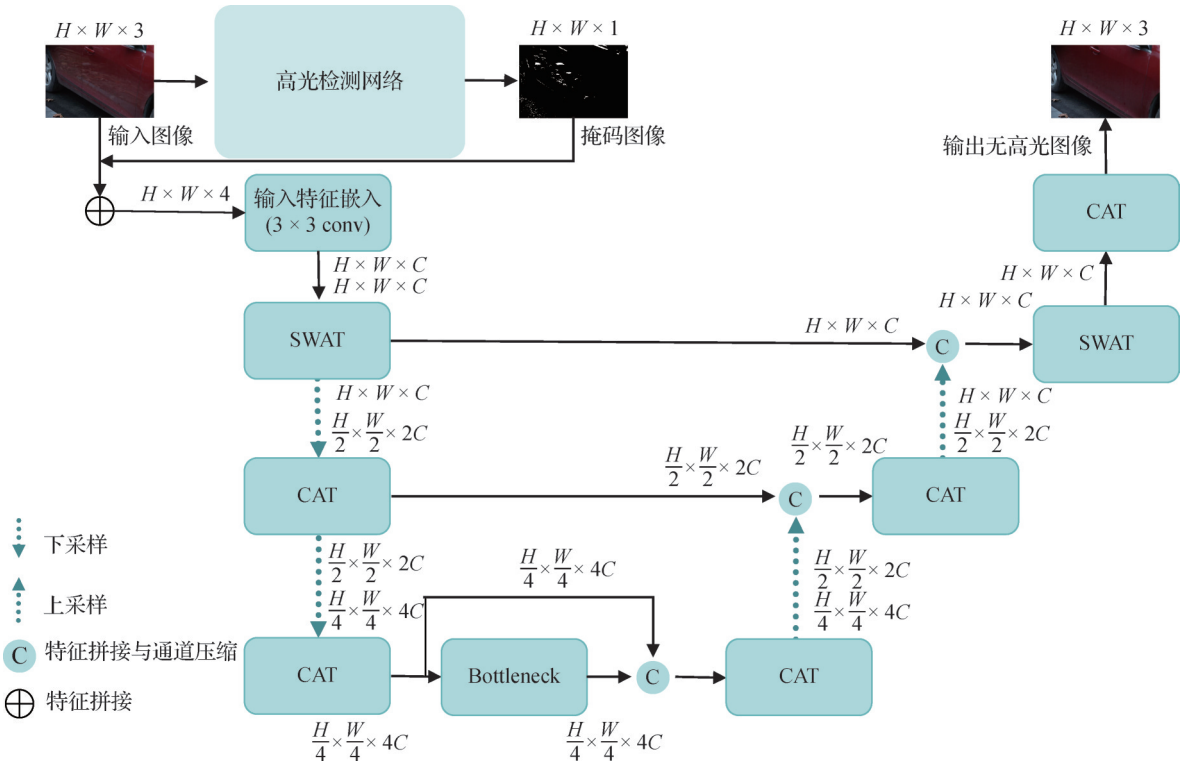


图 3 MG-TransUnet 网络整体架构

Fig. 3 MG-TransUnet network overall architecture

3.2 高光检测网络

高光检测网络如图 4 所示, 该网络以镜面高光图像 I_{spec} 作为输入, 输出对应的高光区域掩码 M , 实现过程可由 $M = N_{\text{detect}}(I_{\text{spec}})$ 给出, 其中, $I_{\text{spec}} \in \mathbb{R}^{H \times W \times 3}$ 是

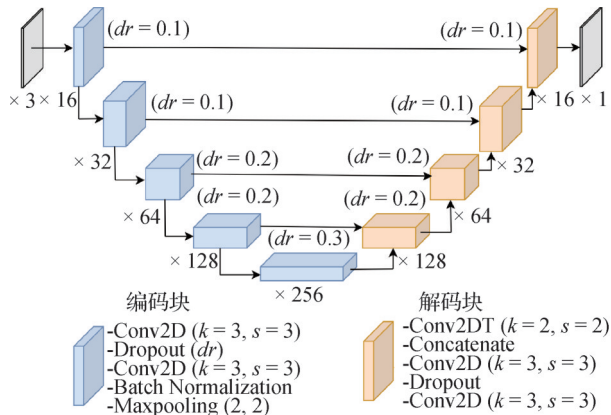


图 4 高光检测网络

Fig. 4 Specular highlight detection network

输入高光图像, $M \in \mathbb{R}^{H \times W \times 1}$ 是输出分割掩码, H 和 W 分别代表图像的高度和宽度, N_{detect} 代表高光检测网络。该网络采用四级对称编码器—解码器结构, 并通过跳跃连接实现多尺度特征的有效融合。编码器部分采用渐进式特征提取策略: 每个编码模块包含两个 3×3 卷积层 (步长为 3)、批归一化层、ReLU (rectified linear unit) 激活函数和 2×2 最大池化层, 其中第 1—5 层分别采用 0.1、0.1、0.2、0.2、0.3 的 dropout 率以增强鲁棒性; 解码器部分通过转置卷积 (卷积核 k 为 2、步长 s 为 2) 逐步上采样, 配合 3×3 卷积层和跳跃连接的特征融合, 最终输出与输入同尺寸的高光掩码。

3.3 高光去除网络

在高光去除任务中, 本文旨在恢复被高光遮挡的底层图像信息, 提出一种基于 Transformer 的高光去除框架, 如图 3 所示, 该网络以镜面高光图像 I_{spec}

和分割掩码 $\mathbf{M}$ 作为输入, 通过端到端的学习输出无镜面反射图像 $\mathbf{I}_{\text{diff}}$, 这一过程可简化表示为 $\mathbf{I}_{\text{diff}} = N_{\text{removal}}(\mathbf{I}_{\text{spec}}, \mathbf{M})$, 其中, $\mathbf{I}_{\text{spec}} \in \mathbf{R}^{H \times W \times 3}$ 表示输入高光图像, $\mathbf{M} \in \mathbf{R}^{H \times W \times 1}$ 是输出分割掩码, $\mathbf{I}_{\text{diff}} \in \mathbf{R}^{H \times W \times 3}$ 为输出无高光图像, N_{removal} 代表高光去除网络。该网络核心在于: 1) 特征提取阶段引入窗口移位注意力模块 (shifted window attention Transformer, SWAT), 该模块通过局部窗口计算与跨窗口信息交互机制, 在保持线性计算复杂度的同时捕获长程空间依赖关系; 2) 编码器路径集成通道注意力模块 (channel attention Transformer, CAT), 通过建立跨通道依赖关系实现特征选择, 从而增强网络对高光区域的鉴别能力。

3.3.1 基于窗口移位的注意力模块(SWAT)

如图3所示, 基于窗口移位的注意力模块 (SWAT) 位于U形架构的最顶层, 其目标是对输入图像进行特征提取, 并增强网络对高光区域的建模能力。整体流程表示为

$$\mathbf{Y} = \mathbf{F} + \text{SWA}(\text{LN}(\mathbf{F}))$$

$$\text{SWAT}(\mathbf{F}) = \mathbf{Y} + \text{FFN}(\text{LN}(\mathbf{Y})) \quad (4)$$

式中, $\mathbf{F}$ 表示输入特征, LN 表示层归一化, SWA 表示窗口注意力块 (shifted window attention, SWA), $\mathbf{Y}$ 表示第1阶段的输出, FFN 表示前馈神经网络 (feed-forward network, FFN), SWAT 表示窗口移位注意力模块。计算过程可分为两个阶段。第1阶段将输入特征 $\mathbf{F}$ 进行层归一化, 并输入至 SWA 模块中以建模局部上下文关系, 随后将该模块输出与原始输入特征 $\mathbf{F}$ 进行残差连接, 得到中间特征表示 $\mathbf{Y}$; 第2阶段将 $\mathbf{Y}$ 进行层归一化, 并输入至由3个卷积层构成的 FFN 模块中进行非线性映射, 最后将输出结果与 $\mathbf{Y}$ 相加, 完成第2次残差连接, 得到最终输出。该设计通过双阶段的注意力增强与前馈建模, 有效提升了模型对复杂高光区域特征的表达能力。

SWA 模块设计了两个级联的窗口注意力块, 用于增强模型对局部细节与空间结构的感知能力。如图5所示, 在第1个窗口注意力块中, 针对输入特征 $\mathbf{F} \in \mathbf{R}^{H \times W \times C}$ (C 表示通道数量), 首先, 将其划分为 $H \cdot W/M^2$ 个不重叠的 $M \times M$ 大小的局部窗口。为确保窗口划分后的数量为整数, 划分前对输入特征图进行零填充 (类似于卷积操作中的填充处理)。完成窗口划分与张量重塑后, 输入特征的维度变为 $(H \cdot W/M^2) \cdot C \cdot M^2$ 。随后, 将特征 $\mathbf{F}$ 投影至 $\mathbf{Q}_s$ (查

询) 和 $\mathbf{K}_s$ (键) 空间上, $\mathbf{Q}_s$ 通过重塑和转置操作变为 $(H \cdot W/M^2) \cdot M^2 \cdot C$ 维度, $\mathbf{K}_s$ 与 $\mathbf{V}$ (值) 通过重塑操作变为 $(H \cdot W/M^2) \cdot C \cdot M^2$, 再计算 $\mathbf{Q}_s$ 与 $\mathbf{K}_s$ 之间的自注意力权重, 以捕捉局部区域内的细粒度结构特征与微小变化。再将得到的注意力权重与 $\mathbf{V}$ 相乘, 以生成加权特征表示并得到维度为 $(H \cdot W/M^2) \cdot C \cdot M^2$ 的特征图, 最终, 将该特征图重塑回原始输入维度得到窗口注意力块的输出特征图。该窗口注意力块具体计算过程为

$$\text{SWA}_{(\mathbf{Q}_s, \mathbf{K}_s, \mathbf{V})} = \mathbf{V} \cdot \text{softmax}\left(\frac{\mathbf{Q}_s^T \times \mathbf{K}_s}{M}\right) \quad (5)$$

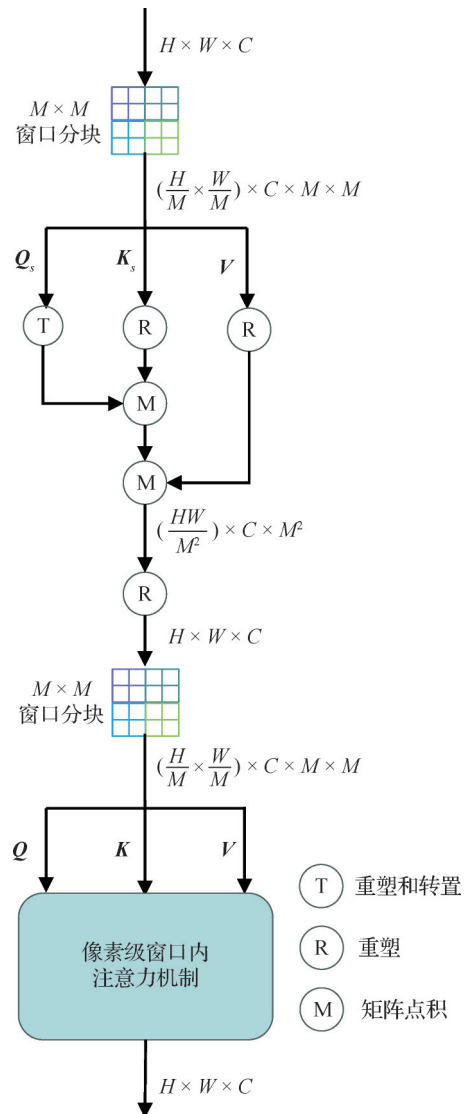


图5 SWAT模块示意图

Fig. 5 Architectural diagram of the SWAT

与第1个注意力块不同, 第2个注意力块采用移位窗口的划分策略。具体而言, 该模块首先对输入特征图执行循环移位操作, 使得注意力机制的感受

野能够覆盖前一阶段窗口中未被充分建模的边界区域。随后,在位移后的特征图上进行窗口划分,确保注意力计算仍能映射回原始像素位置。这一设计不仅保持了窗口计算的高效性,同时增强了模型对边缘细节的建模能力,从而提升了局部特征与全局上下文的融合效果。

3.3.2 通道注意力模块(CAT)

该模块通过编码器路径中的渐进式下采样和解码器路径中的对称上采样操作,实现了多尺度通道级上下文信息的层次化提取。针对跨通道上下文关系建模的挑战,采用注意力机制对高层语义特征进行建模。具体而言,该模块通过建立通道间的长程依赖关系,动态学习各通道特征在当前任务中的权重分布,从而实现对关键通道特征的增强与非必要特征的抑制。CAT根据其与其他通道的上下文关系衡量每个通道的重要性。具体为

$$CA_{(\hat{Q}, \hat{K}, \hat{V})} = \hat{V} \cdot \text{softmax}\left(\frac{\hat{Q} \cdot \hat{K}^T}{\sqrt{\hat{C}/h}}\right) \quad (6)$$

如图6所示, $\hat{Q}$ 、 $\hat{K}$ 和 $\hat{V}$ 表示特征图 $\hat{F} \in \mathbf{R}^{\hat{H} \times \hat{W} \times \hat{C}}$ 的投影,该投影被调整为原始输入图像的分辨率 1/2 或 1/4,以对应于U形结构中的不同比例层级。

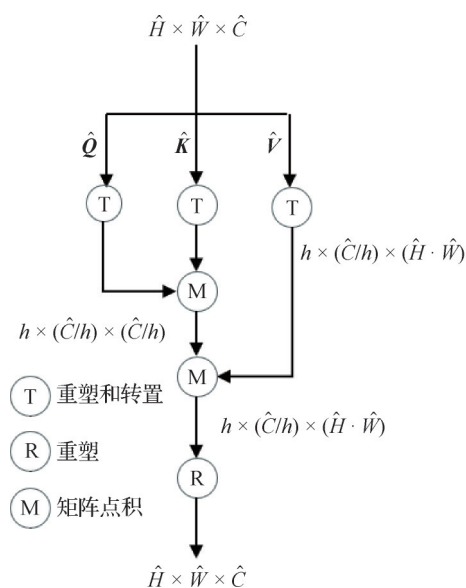


图6 CAT模块示意图

Fig. 6 Architectural diagram of the CAT

然后,这些投影被重塑为 $h \times (\hat{C}/h) \times (\hat{H} \cdot \hat{W})$ 的维度,其中, h 表示注意头的数量, $\hat{C}$ 表示通道数量,再计算 $\hat{Q}$ 与 $\hat{K}$ 之间的自注意力权重,以捕捉不同通道空间内的重要程度。最后用注意力权重与 $\hat{V}$ 相乘

得到维度为 $h \times (\hat{C}/h) \times (\hat{H} \cdot \hat{W})$ 的特征表示,并重塑回 $\hat{H} \times \hat{W} \times \hat{C}$ 大小输出。CAT在通道维度上运行,使网络能够根据其上下文相关性动态调整对不同通道的强调。实验结果表明,该机制有效提升了模型在复杂光照条件下对高光区域的感知能力与细节恢复性能。

3.4 损失函数

高光检测网络本质上是一个像素级二分类问题,其目标是判定每个像素是否属于高光区域,从而推断出整个高光区域的空间位置,因此在损失函数的设计上,本文选用了相似系数损失(Dice loss)与焦点损失(focal loss)的组合对检测网络进行优化指导。其中,相似系数损失用于缓解前景与背景类别不平衡带来的训练偏差,增强模型对高光区域的敏感性,具体的定义为

$$L_{\text{Dice}}(p, \hat{p}) = 1 - 2 \sum p \cdot \hat{p} / (\sum p + \sum \hat{p}) \quad (7)$$

式中, p 是真实掩码图的二值表示, $\hat{p}$ 是预测概率且 $p \in \{0, 1\}$, $0 \leq \hat{p} \leq 1$ 。

但是, Dice loss 通过从区域层面优化预测掩码与真实标签之间的交并比来保证预测高光区域的整体完整性,对于边界的精细化程度关注不足。因此引入了 focal loss 作为辅助损失去驱动模型学习如何更精确地分割高光区域的边缘细节,从而生成边界更准确的掩码。其定义为

$$L_{\text{Focal}}(p, \hat{p}) = -\alpha \cdot (1 - p_i)^\gamma \cdot \log(p_i) \quad (8)$$

$$p_i = \begin{cases} 1 - \hat{p} & p = 0 \\ \hat{p} & p = 1 \end{cases}$$

式中, α 是平衡超参数,用于平衡正负样本的重要性。 γ 是调节超参数,用于调节难易样本的损失占比。

最终,整个高光检测网络的损失可定义为

$$L_{\text{Detect}} = L_{\text{Dice}} + L_{\text{Focal}} \quad (9)$$

高光去除任务的目标是尽可能地使网络输出与真实漫反射图像保持一致,这不仅要求实现高精度重建,同时也需保留图像的亮度、对比度及结构等关键视觉属性。为此,本文引入了一个复合损失函数,综合考虑了像素保真度与结构一致性两个方面。具体而言,该复合损失函数由峰值信噪比(peak signal-to-noise ratio, PSNR)与 SSIM 共同构成。PSNR 作为衡量像素级重建误差的指标,用于确保重建图像在数值上的精确性; SSIM 侧重于图像结构信息的保

持,有助于提升高光去除后的感知质量。两者结合,有效提升了模型在高光去除任务中的整体表现。最终整个网络的损失函数为

$$L_{\text{Removal}} = \omega \cdot L_{\text{ssim}} + L_{\text{psnr}} + L_{\text{Dice}} + L_{\text{Focal}} \quad (10)$$

式中,ω设置为0.4。

4 实 验

4.1 实验环境

本网络使用 NVIDIA GeForce GTX3090 显卡并在 PyTorch 中实现,使用 Adam 优化器在训练集上训练网络 100 个 epoch,初始学习率设置为 2×10^{-4} ,每 5 个周期使用 0.8 的衰减系数减小至 1×10^{-6} 。

4.2 实验结果

4.2.1 高光检测网络

本网络第 1 阶段需要高光掩码作为先验指导,而现有公开数据集中仅 SHIQ 数据集提供了真实的高光掩码,因此,针对 SSHR 和 PSD 数据集,通过对高光图像与对应的无高光图像进行像素级差分并阈值化生成对应的高光掩码。

为验证高光检测网络的性能,本文在 SHIQ 数据集上与现有多种算法进行了对比实验,并采用准确率(accuracy, Acc)作为评价指标。表 2 汇总了各方法在 SHIQ 数据集上的性能表现。从表 2 可以看出,本网络在准确率指标上优于其他先进方法(state-of-the-art, SOTA),能够生成准确的高光掩码并满足对

表 2 高光检测网络与其他方法在 SHIQ 数据集的对比

Table 2 Comparison of results between highlight detection network and other methods on SHIQ dataset

方法	Acc
Orhlr(Shi 等,2025b)	0.90
Shdnet(Fu 等,2020)	0.87
Unet-Transformer	<u>0.97</u>
本文	0.98

注:加粗、下划线字体表示各列最优、次优结果。

高光区域精确定位的需求。

4.2.2 高光去除网络

本文将高光检测网络与高光去除网络进行联合训练,并在三大公开数据集上进行了测试。表 3 展示了联合高光检测和高光去除网络与现有高光去除方法在三大公开数据集上训练和测试的实验结果。从表 3 可以看出,本文方法在 SHIQ 和 SSHR 数据集上的两大评价指标均优于所有对比方法,而在 PSD 数据集上仅次于 HighlightRNet,说明本文方法能够有效地提升高光去除任务的性能。

4.2.3 汽车高光数据集

汽车高光数据集有高光强和大面积高光场景,为了验证现有高光去除方法能否有效地去除这类复杂场景中的高光,本文选取了 3 个公开网络在汽车数据集上进行了实验,具体结果如表 4 所示。可以看出,本文方法在所构建的汽车数据集上同样展现出良好的性能,能够应对更为复杂的真实场景下的

表 3 本文算法和现有 SOTA 算法在三大公开数据集的对比结果

Table 3 Comparison of results between our algorithm and SOTA algorithms on three major public datasets

方法	PSD		SHIQ		SSHR	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
SHIQ(Fu 等,2021)	21.52	0.88	<u>34.13</u>	0.86	26.98	0.90
PSD(Wu 等,2022)	21.80	0.88	23.42	0.92	25.73	0.89
MG-GAN(Hu 等,2022)	22.12	0.82	30.94	0.97	28.40	0.87
Unet-Transformer(Wu 等,2023)	23.93	0.86	31.57	<u>0.96</u>	33.45	0.95
TSHR(Fu 等,2023)	22.76	<u>0.90</u>	25.58	0.93	28.63	0.94
DHAN(Guo 等,2024)	25.17	0.87	33.64	0.97	<u>36.16</u>	<u>0.96</u>
HighlightRNet(Zhang 等,2024)	29.79	0.92	30.23	0.93	30.07	<u>0.96</u>
L ² HRNet(Pan 等,2025)	—	—	32.71	<u>0.96</u>	—	—
本文	<u>26.37</u>	0.88	35.24	0.97	36.54	0.97

注:加粗、下划线字体表示各列最优、次优结果。“—”表示未公开在该数据集的结果。

高光去除任务。然而,从最终恢复图像的 PSNR 指标来看,现有方法在该数据集上的表现与其在三大公开数据集上的结果存在差距。这一现象说明,在面对光照强度更大、高光区域分布更广泛的场景时,当前高光去除方法仍面临挑战,需进一步提升其在复杂自然环境下的泛化能力与鲁棒性。

为提供更直观的验证,图 7 展示了 MG-TransUnet 的可视化效果。可以清晰地观察到,对于大面积高光与复杂纹理高光,本模型均能有效地抑制高光,同时精准地恢复其下的车身结构与表面纹理,其输出结果与真实无光图像在视觉上高度一致。

表 4 本文算法和其他算法在汽车数据集上的对比结果
Table 4 Comparison of results between our algorithm and other algorithms on the car dataset

方法	PSNR/dB	SSIM
PSD	22.66	0.74
MG-GAN	<u>25.03</u>	0.80
DHAN	23.78	<u>0.82</u>
本文	25.14	0.83

注:加粗、下划线字体表示各列最优、次优结果。

这一方面佐证了前述定量指标的可靠性,另一方面也直观体现了本文方法在真实场景的实用价值。

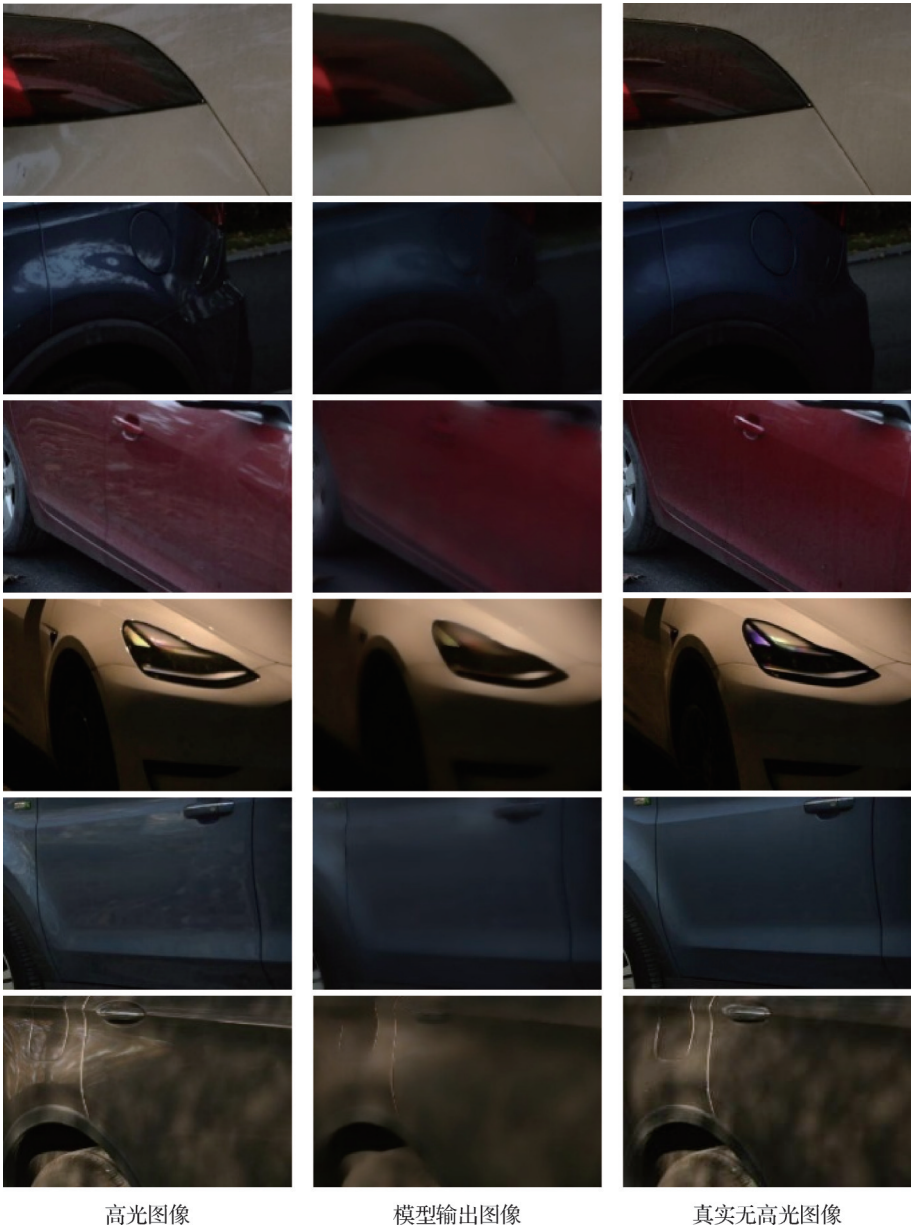


图 7 MG-TransUnet 算法在汽车数据集上的实验结果

Fig. 7 Experimental results of MG-TransUnet on the car dataset

4.3 消融实验

本研究提出的网络基于 U-Net 结构实现。为了准确地评估各模块对高光去除性能的影响,本文用标准的纯 U-Net 模型重新训练,并基于此结构引入不同的模块进行对比实验,从而验证各组件的有效性。表 5 展示了掩码引导模块与窗口移位注意力模块在标准 U-Net 结构下的实验结果。从表 5 可以看出,掩码引导网络与窗口移位注意力模块均在不同程度上提升了整体网络的性能。在标准 U-Net 模型的基础上,引入掩码引导模块使 PSNR 指标平均提升了 0.993 3 dB,SSIM 指标平均提升了 0.026 7;引入窗口移位注意力模块后,PSNR 指标平均提升了 1.49 dB,

SSIM 指标平均提升了 0.033 3。由此可见,窗口移位注意力模块在性能提升效果上相较掩码引导模块更为显著。最终,将两种模块联合构建的完整网络架构在各项指标上均取得了最佳实验结果,验证了各子模块在高光去除任务中的有效性与协同增强作用。

此外,为了更直观地验证本文提出的窗口移位注意力模块在高光区域特征提取方面的有效性,本研究对该模块的输出特征图进行可视化,并与原始高光图像进行对比分析,从而评估其高光特征提取能力。实验结果如图 8 所示,SWAT 模块提取出的特征热力图能够准确地显示高光区域,达到对高光区域特征建模的要求。

表 5 掩码引导网络与窗口移位注意力模块的消融实验结果

Table 5 The ablation experiment of mask guided network and shifted window attention Transformer module

方法	PSD		SHIQ		SSHR	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
U-Net	24.34	0.83	32.81	0.91	35.18	0.91
U-Net + 掩码引导网络	25.78	<u>0.86</u>	33.64	0.94	35.89	0.93
U-Net + 窗口移位注意力模块	<u>26.16</u>	<u>0.86</u>	<u>34.37</u>	<u>0.95</u>	<u>36.27</u>	<u>0.94</u>
本文方法	26.37	0.88	35.24	0.97	36.54	0.97

注:加粗、下划线字体表示各列最优、次优结果。

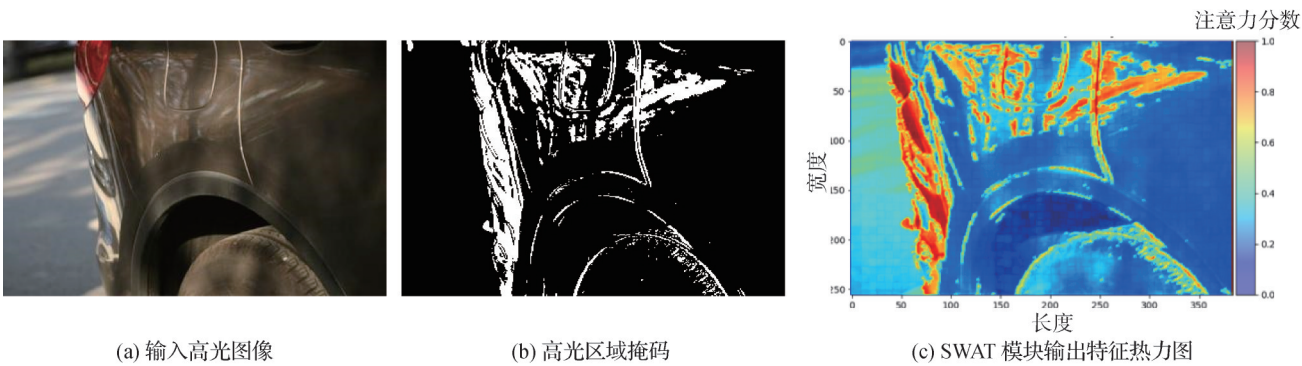


图 8 SWAT 模块输出特征热力图与输入高光图像比较结果

Fig. 8 Comparison results between SWAT module output feature heatmap and input highlight image
((a) input highlight image; (b) highlight area mask; (c) SWAT module outputs feature heatmap)

5 结 论

本文构建了首个面向室外自然光照条件下汽车表面高光检测与去除的小型高光图像数据集。该数据集采集于自然环境中的白天直射光、夜间路灯照明以及地下车库等多种真实场景,涵盖了太阳光与

灯光等光源,弥补了现有高光去除研究中缺乏真实场景数据的不足。此外,本数据集还包含大量具有曲面金属材质和玻璃反射属性的图像,为高光检测与去除任务带来了更高挑战。

基于该数据集,本文进一步提出了两阶段高光去除网络。在三大主流公开数据集上的实验结果表明,该网络能够精准地定位高光区域,并有效恢复被

高光覆盖的像素信息。同时本文在所构建的汽车高光数据集上对当前主流的高光去除方法进行了系统评估,实验发现,在处理光照强度较高、区域分布较广的高光场景时,现有方法的性能仍存在明显提升空间。

参考文献(References)

- Akashi Y and Okatani T. 2015. Separation of reflection components by sparse non-negative matrix factorization//Proceedings of the 12th Asian Conference on Computer Vision. Singapore, Singapore: Springer: 611-625 [DOI: 10.1007/978-3-319-16814-2_40]
- Anwer A, Ainouz S, Saad M N M, Ali S S A and Meriaudeau F. 2022. SpecSeg network for specular highlight detection and segmentation in real-world images. *Sensors*, 22 (17): #6552 [DOI: 10.3390/s22176552]
- Anwer A, Ainouz S, Saad N M, Ali S S A and Meriaudeau F. 2023. Joint network for specular highlight detection and adversarial generation of specular-free images trained with polarimetric data. *Neurocomputing*, 559: #126769 [DOI: 10.1016/j.neucom.2023.126769]
- Born M and Wolf E. 1999. Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light. 7th ed. Cambridge, USA: Cambridge University Press
- Chen J, Song Y, Li D P, Lin X X, Zhou S H and Xu W Q. 2024. Specular removal of industrial metal objects without changing lighting configuration. *IEEE Transactions on Industrial Informatics*, 20 (3): 3144-3153 [DOI: 10.1109/tii.2023.3297613]
- Finlayson G D, Hordley S D and Hubel P M. 2001. Color by correlation: a simple, unifying framework for color constancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23 (11): 1209-1221 [DOI: 10.1109/34.969113]
- Fu G, Zhang Q, Lin Q F, Zhu L and Xiao C X. 2020. Learning to detect specular highlights from real-world images//Proceedings of the 28th ACM International Conference on Multimedia. Seattle, USA: Association for Computing Machinery: 1873-1881 [DOI: 10.1145/3394171.3413586]
- Fu G, Zhang Q, Song C F, Lin Q F and Xiao C X. 2019. Specular highlight removal for real-world images. *Computer Graphics Forum*, 38 (7): 253-263 [DOI: 10.1111/cgf.13834]
- Fu G, Zhang Q, Zhu L, Li P and Xiao C X. 2021. A multi-task network for joint specular highlight detection and removal//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 7752-7761 [DOI: 10.1109/cvpr46437.2021.00766]
- Fu G, Zhang Q, Zhu L, Lin Q F, Wang Y H, Fan S Y, et al. 2024. Towards high-resolution specular highlight detection. *International Journal of Computer Vision*, 132 (1): 95-117 [DOI: 10.1007/s11263-023-01845-3]
- Fu G, Zhang Q, Zhu L, Xiao C X and Li P. 2023. Towards high-quality specular highlight removal by leveraging large-scale synthetic data//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 12857-12865 [DOI: 10.1109/ICCV51070.2023.01181]
- Gao R X and Li X Y. 2018. Specular highlight removal using improved bilateral filtering. *Journal of Image and Graphics*, 23 (1): 9-17 (高如新, 李雪颖. 2018. 双边滤波的改进高光去除. 中国图象图形学报, 23 (1): 9-17) [DOI: 10.11834/jig.170272]
- Guo J, Zhou Z J and Wang L M. 2018. Single image highlight removal with a sparse and low-rank reflection model//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 282-298 [DOI: 10.1007/978-3-030-01225-0_17]
- Guo X J, Chen X H, Luo S H, Wang S Q and Pun C M. 2024. Dual-hybrid attention network for specular highlight removal//Proceedings of the 32nd ACM International Conference on Multimedia. Melbourne, Australia: Association for Computing Machinery: 10173-10181 [DOI: 10.1145/3664647.3680745]
- Hao C P, He Y, Li Y F, Niu X B and Wang Y. 2025. A fast specular removal method for a single real image. *Displays*, 87: #102930 [DOI: 10.1016/j.displa.2024.102930]
- Hu G W, Zheng Y F, Yan H R, Hua G and Yan Y C. 2022. Mask-guided cycle-GAN for specular highlight removal. *Pattern Recognition Letters*, 161: 108-114 [DOI: 10.1016/j.patrec.2022.06.014]
- Hu K, Huang Z Y F and Wang X J. 2024. Highlight removal network based on an improved dichromatic reflection model//Proceedings of 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Seoul, Korea (South): IEEE: 2645-2649 [DOI: 10.1109/ICASSP48485.2024.10447916]
- Huang Z Y F, Hu K and Wang X J. 2022. M2-Net: multi-stages specular highlight detection and removal in multi-scenes [EB/OL]. [2025-05-26]. <https://arxiv.org/pdf/2207.09965.pdf>
- Jiang S R, Cheng L, Yuan H W and Li X. 2024. Highlight removal emphasizing detail restoration. *Applied Sciences*, 14 (6): #2469 [DOI: 10.3390/app14062469]
- Lei C Y, Huang X H, Zhang M D, Yan Q, Sun W X and Chen Q F. 2020. Polarized reflection removal with perfect alignment in the wild//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 1750-1758 [DOI: 10.1109/CVPR42600.2020.00182]
- Li W, Jia F C and Liu W J. 2024. EndoSRR: a comprehensive multi-stage approach for endoscopic specular reflection removal. *International Journal of Computer Assisted Radiology and Surgery*, 19 (6): 1203-1211 [DOI: 10.1007/s11548-024-03137-8]
- Li X, Liu Z Q, Cai Y D, Pan C Y, Song J W, Wang J S, et al. 2023. Polarization 3D imaging technology: a review. *Frontiers in Physics*, 11: #1198457 [DOI: 10.3389/fphy.2023.1198457]
- Lin S, Li Y Z, Kang S B, Tong X and Shum H Y. 2002. Diffuse-specular separation and depth recovery from image sequences//

- ceedings of the 7th European Conference on Computer Vision. Copenhagen, Denmark: Springer: 210-224 [DOI: 10.1007/3-540-47977-5_14]
- Muhammad S, Dailey M N, Farooq M, Majeed M F and Ekpanyapong M. 2020. Spec-Net and Spec-CGAN: deep learning models for specular removal from faces. *Image and Vision Computing*, 93: #103823 [DOI: 10.1016/j.imavis.2019.11.001]
- Pan L, Huang Y H, Xie H X, Zhang C, Zhao H W, Shuai H H, et al. 2025. Single document image highlight removal via a large-scale real-world dataset and a location-aware network [EB/OL]. [2025-05-26]. <https://arxiv.org/pdf/2504.14238.pdf>
- Pan L, Zhao H W and Wu R Z. 2024. Towards specular highlight removal through diffusion model//Proceedings of the 7th Chinese Conference on Pattern Recognition and Computer Vision. Urumqi, China: Springer: 497-509 [DOI: 10.1007/978-981-97-8685-5_35]
- Park J B and Kak A C. 2003. A truncated least squares approach to the detection of specular highlights in color images//Proceedings of 2003 IEEE International Conference on Robotics and Automation. Taipei, China: IEEE: 1397-1403 [DOI: 10.1109/ROBOT.2003.1241787]
- Shafer S A. 1985. Using color to separate reflection components. *Color Research and Application*, 10 (4): 210-218 [DOI: 10.1002/col.5080100409]
- Shen H L, Zhang H G, Shao S J and Xin J H. 2008. Chromaticity-based separation of reflection components in a single image. *Pattern Recognition*, 41 (8): 2461-2469 [DOI: 10.1016/j.patcog.2008.01.026]
- Shen H L and Zheng Z H. 2013. Real-time highlight removal using intensity ratio. *Applied Optics*, 52 (19): 4483-4493 [DOI: 10.1364/AO.52.004483]
- Shi W Q, Quan H D and Kong L B. 2025a. Adaptive specular reflection removal in light field microscopy using multi-polarization hybrid illumination and deep learning. *Optics and Lasers in Engineering*, 186: #108839 [DOI: 10.1016/j.optlaseng.2025.108839]
- Shi W Z, Hu Z Q, Chen H, Zhang H J, Yang J L and Li L. 2025b. Orhrl-net: one-stage residual learning network for joint single-image specular highlight detection and removal. *The Visual Computer*, 41 (5): 3359-3370 [DOI: 10.1007/s00371-024-03607-9]
- Wen S J, Zheng Y Q and Lu F. 2021. Polarization guided specular reflection separation. *IEEE Transactions on Image Processing*, 30: 7280-7291 [DOI: 10.1109/tip.2021.3104188]
- Wu Z Q, Guo J W, Zhuang C Q, Xiao J, Yan D M and Zhang X P. 2023. Joint specular highlight detection and removal in single images via Unet-Transformer. *Computational Visual Media*, 9 (1): 141-154 [DOI: 10.1007/s41095-022-0273-9]
- Wu Z Q, Zhuang C Q, Shi J, Guo J W, Xiao J, Zhang X P, et al. 2022. Single-image specular highlight removal via real-world dataset construction. *IEEE Transactions on Multimedia*, 24: 3782-3793 [DOI: 10.1109/tmm.2021.3107688]
- Yang J W, Liu L X and Li S Z. 2013. Separating specular and diffuse reflection components in the HSI color space//Proceedings of 2013 IEEE International Conference on Computer Vision Workshops. Sydney, Australia: IEEE: 891-898 [DOI: 10.1109/ICCVW.2013.122]
- Zhang C, Liu Y L, Wang K and Tian J. 2023. Specular highlight removal for endoscopic images using partial attention network. *Physics in Medicine and Biology*, 68 (22): #225009 [DOI: 10.1088/1361-6560/ad02d9]
- Zhang L, Ma Y D, Jiang Z, He W L, Bao Z Y, Fu G, et al. 2024. High-lightRemover: spatially valid pixel learning for image specular highlight removal//Proceedings of the 32nd ACM International Conference on Multimedia. Melbourne, Australia: Association for Computing Machinery: 10046-10054 [DOI: 10.1145/3664647.3681434]
- Zhang W M, Zhao X, Morvan J M and Chen L M. 2019. Improving shadow suppression for illumination robust face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41 (3): 611-624 [DOI: 10.1109/TPAMI.2018.2803179]
- Zhou L J and Mao J N. 2023. Vision Transformer-based recognition tasks: a critical review. *Journal of Image and Graphics*, 28 (10): 2969-3003 (周丽娟, 毛嘉宁. 2023. 视觉Transformer识别任务研究综述. *中国图象图形学报*, 28 (10): 2969-3003) [DOI: 10.11834/jig.220895]
- Zhu D Z, Huang G H, Yuan X C, Chen X H, Zhong G, Pun C M, et al. 2024. FAQNet: frequency-aware quaternion network for endoscopic highlight removal//Proceedings of 2024 IEEE International Conference on Bioinformatics and Biomedicine. Lisbon, Portugal: IEEE: 1408-1413 [DOI: 10.1109/BIBM62325.2024.10822449]

作者简介

王嘉成,男,硕士研究生,主要研究方向为深度学习和图像处理。E-mail: 2232180@mail.dhu.edu.cn

谢颖华,通信作者,女,副教授,主要研究方向为大数据和数据分析。E-mail: yh_xie@dhu.edu.cn

周树波,男,讲师,主要研究方向为计算机视觉与图像重建。E-mail: zhoushubo@dhu.edu.cn

潘峰,男,讲师,主要研究方向为机器人控制与图像处理。E-mail: fpan@mail.dhu.edu.cn

蒋学芹,男,教授,主要研究方向为机器学习和量子密钥分发。E-mail: xqjiang@dhu.edu.cn

黄荣,男,副教授,主要研究方向为深度学习和机器视觉。E-mail: rong.huang@dhu.edu.cn