

探索单图像反射去除的光谱特性

Pengbo Guo^{1,2}, Chengxu Liu¹, Guoshuai Zhao¹, Xingsong Hou¹, Jialie Shen³, Xueming Qian¹

¹Xi'an Jiaotong University

²Shanghai Innovation Institute

³City, University of London

Abstract

消除由入射光与反射介质相互作用引起的反射仍然是图像恢复领域的一个病态问题。主要的挑战在于捕获图像中反射和透射成分的重叠，这使得准确区分和恢复干净背景的任务变得复杂。现有的方法通常仅在图像域处理反射去除，忽略了反射光谱特性的变化，这阻碍了它们有效辨别反射的能力。本文从光谱学习的新视角出发，提出光谱码本以重建反射图像的光学光谱。通过感知光谱中不同光源之间的波长差异，可以有效地区分反射。为了利用重建的光谱，我们设计了两个光谱先验细化模块来在空间维度重新分配像素，并沿波长维度自适应增强光谱差异。此外，我们提出了光谱感知变换器以联合恢复频谱和像素域中的透射内容。在三个不同的反射基准上的实验结果证明了我们的方法优于最先进的模型，并展示了其优越性和泛化能力。

介绍

作为一种日常生活中的普遍物理现象，反射通常由透明或有光泽的材料（如玻璃或金属）引起，从而产生破坏图像质量并干扰后续下游视觉任务的不良伪影 (Sinha et al. 2012; Piccinelli et al. 2024; Wang et al. 2025)。因此，去除反射一直是低级视觉中的一个关键任务。

现有方法主要集中在特定场景上，例如，透明反射 (Hu and Guo 2021, 2023; Zhu et al. 2024)，塑料薄膜反射 (Tang et al. 2024)，或镜面高光 (Fu et al. 2021)，通过精心设计的手工先验或网络来移除反射。然而，这些方法面临着以下挑战：1) 由于训练数据不足导致的严重过拟合使得模型将反射模式识别为简单的色彩映射，错误地将饱和度高的区域识别为反射，而忽视了那些饱和度较低的区域；2) 作为一种自然界中最普遍的物理现象之一，反射在不同场景中并不会孤立出现。各

种反射退化是入射光与不同反射介质 (Born and Wolf 2013) 相互作用的视觉结果。将它们分开处理会降低模型的鲁棒性，并带来额外的成本。因此，探索不同光源的差异表示对于移除反射至关重要。

根据反射背后的物理原理，当入射光穿过反射介质时，反射光的能量分布表现出显著差异 (Born and Wolf 2013; Zheng, Sato, and Sato 2015; Nguyen et al. 2018; Qiu et al. 2023)。受此启发，我们在三种不同的反射场景中进行了实验，利用反射光和透射光之间的波长差异来观察反射区域的变化。如图 1所示，通过采样光谱差异明显的特定频带，可以获得没有反射退化的清晰光谱图像。更多证据可以在补充材料中找到。

基于这一观察，我们专注于反射光的光学特性，发现光谱信息可用于放大反射和透射成分之间的差异可以有效用于解决光源分离问题。因此，我们的目标是通过学习输入反射图像的光谱来指导反射去除，利用波长差异来分离不同的成分。具体来说，我们使用配对的 RGB-光谱图像来训练我们的 Spectral Codebook (光谱码本)，该模型描述了各波段上的光谱分布。在去除反射的过程中，可以使用预训练的 Spectral Codebook 直接重建光谱而无需增加额外成本。

为了去除捕获图像中嵌入的反射，我们提出了一种新颖的单图像去反射网络，该网络利用光谱指导来区分每个组件之间的差异。我们的方法的关键见解在于利用反射的光谱特征，在 RGB 和光谱域联合学习以消除反射退化并重建传输内容。图 2展示了总体概述，其中包括光谱先验重构 (SPR) 模块和光谱感知转换器 (SAFormer)。具体来说，SPR 首先重构给定的反射图像的光谱，并在空间和波长维度上对其进行优化以放大每个组件之间的差异并消除冗余信息。随后，差分离散标记器 (DST) 分别编码具有不同模式的输入特征，互补引导多头自注意力 (CG-MSA) 协同恢复干净的背景内容。引入 CC-FFN 来补偿在差异化学学习过程中丢失的上下文信息。

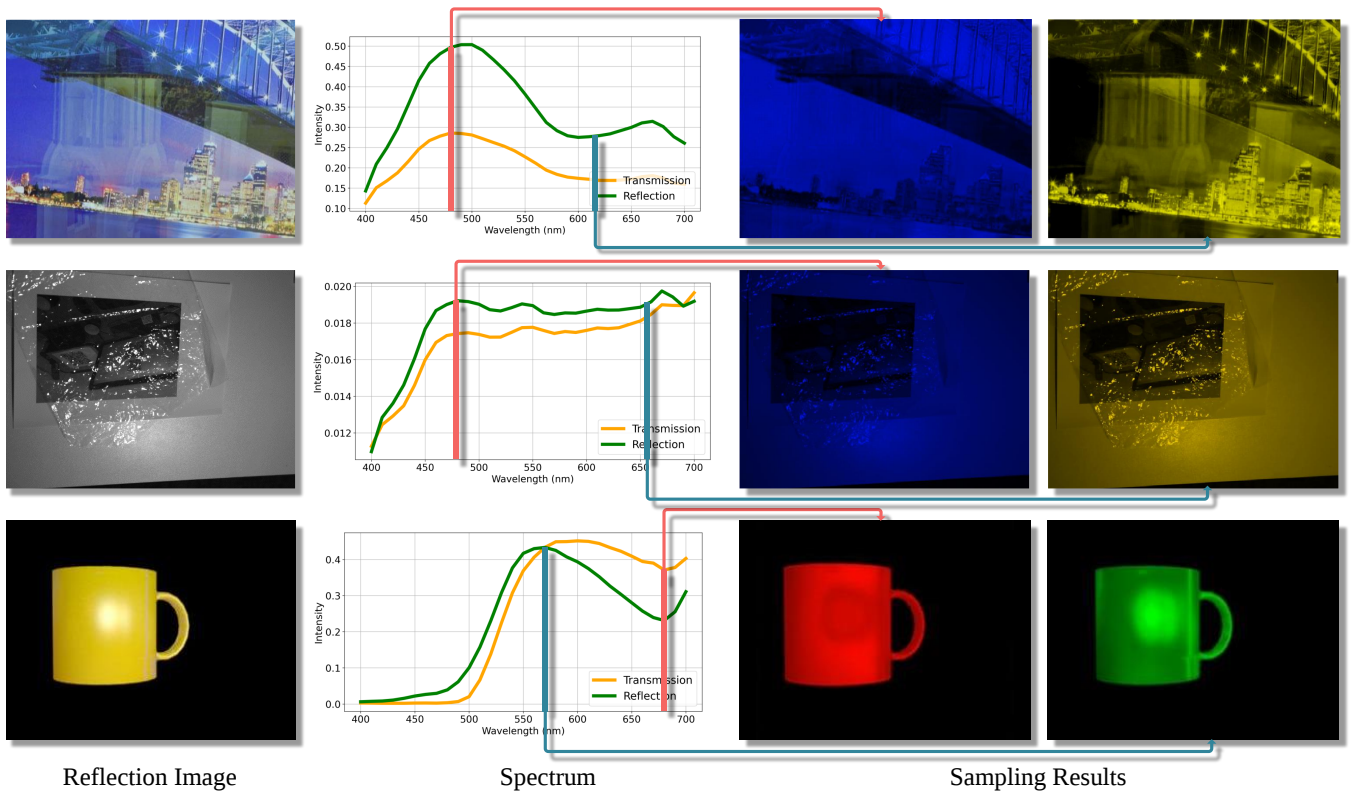


图 1: 我们使用最新的 SOTA 光谱重建方法 MST++ (Cai et al. 2022) 来估计给定反射和透射图像在各种场景下的光谱。通过对强度差异较大的光谱带进行采样, 我们可以获得几乎无反射的结果。这样的观察表明利用光谱引导去反射过程的必要性。**红色** 线表示在最大强度差异处 (无反射) 的采样, 而**蓝色** 线表示在最小差异处 (全反射) 的采样。

我们的贡献可以总结如下:

- 我们探索了光谱波长差异与反射成像之间的内在关系, 强调了使用具有显著能量差异的光谱带有效抑制反射的潜力, 为去除反射提供了新的见解。
- 我们提出了光谱码本, 以在没有额外计算成本的情况下重建输入图像的光谱, 并提出两个光谱先验细化模块来增强其在指导去除反射方面的有效性。
- 我们设计了一种新型的频谱感知 Transformer, 开创性地将频谱学习融入到反射去除任务中。通过在 RGB 和频谱域联合学习, 我们的方法有效地消除了反射并重构了干净的透射内容。

相关工作

单图像反射去除

较少的数据约束使得基于单幅图像的方法更受欢迎, 但也由于信息有限而带来了挑战。为了解决这个问题, 传统方法 (Levin and Weiss 2007; Yang et al. 2019) 依赖于手工设计的先验知识, 假设传输层和反射层具有不同的统计分布。然而, 对这种刚性假设的依赖限制了

它们的鲁棒性, 当遇到违反这些先验的复杂场景时, 往往会导致失败。

为了克服这些限制, 最近的工作很大程度上转向了基于深度学习的解决方案。Hu 和 Guo(2021) 同时恢复反射和透射成分, 使用 ReLU 整流器来增强两个分支之间的交互。此外, Hu 和 Guo (2023) 提出了一种可学习的残余项, 有效地在分解过程中捕获残差信息。Zhong 等人 (2024) 引入了对反射图像中不同成分的语言描述来更高效地指导去除过程。Zhu 等人 (2024) 设计了一个最大滤波器来定位反射区域, 然后使用位置引导来移除反射。然而, 大多数方法专注于复杂模块的设计, 忽视了探索处理反射的基本特征。

光谱重构

高光谱图像 (HSI) 捕获了细微的光谱特征, 与传统的 RGB 图像相比提供了更丰富的纹理和结构细节。然而, 依赖于光谱扫描 (Huang et al. 2022) 的传统 HSI 采集通常由于昂贵的设备成本和较长的拍摄时间而不切实际。

最近, 基于学习的从单个 RGB 图像进行高光谱重

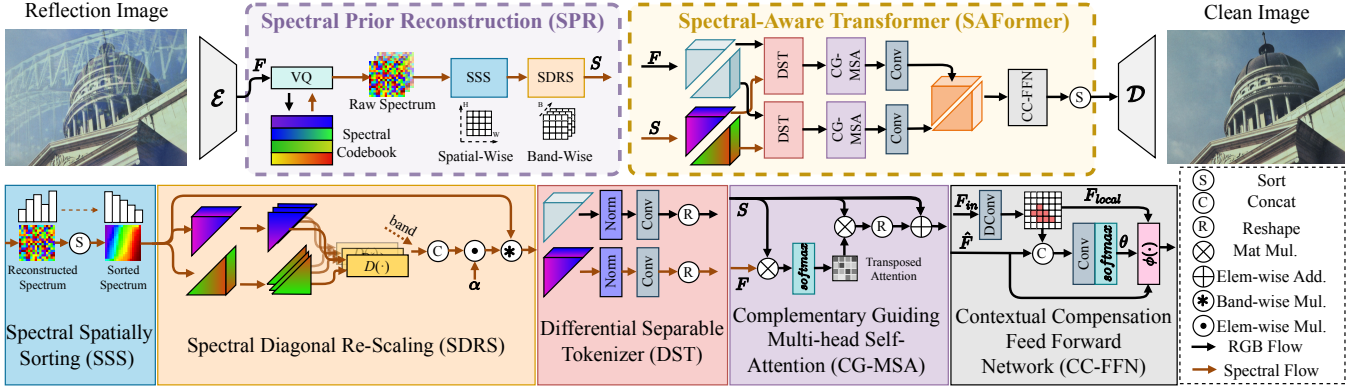


图 2: 我们提出的方法概述。光谱先验重建 (SPR) 模块利用光谱码本重构输入反射图像的光谱, 并自适应地增强其效果。我们的光谱感知变换器 (SAFormer) 差异性学习跨域排序特征。

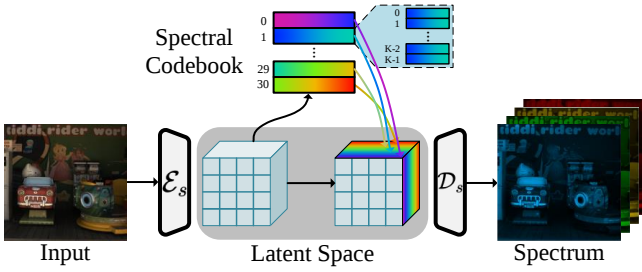


图 3: 训练我们的光谱代码本的配对 RGB-光谱图像示例。

建已成为一个重要的研究方向。熊等人 (2017) 提出了 HSCNN 框架, 该框架利用输入的 RGB 图像和压缩测量来进行光谱重建。蔡等人 (2022) 将 Transformer 架构应用于光谱重建任务。在这项工作中, 我们建模了频谱域中反射率的分布, 利用光谱信息来引导去除反射的过程。

我们的方法

如图 2 所示, 给定通过编码器的反射图像, 光谱先验重建 (SPR) 模块首先利用光谱代码本重构其光谱, 并分别沿空间和波长维度自适应地细化光谱。输入的 RGB 特征和细化后的光谱随后被送入光谱感知变换器 (SAFormer) 以共同去除反射并恢复透射内容。

光谱先验重建

图 1 中的波长变化激励我们利用光谱信息消除由反射引起的退化。不幸的是, 直接重建光谱在正确引导恢复过程方面存在显著挑战, 并且需要一个具有大量参数的额外重建网络。为了解决这个问题, 我们提出光谱码本作为更有效的方法来重建反射光谱。

为了更好地利用我们重建的原始光谱的指导作用, 分别提出了 SSS 和 SDRS 来优化其在空间维度和波长维度上的性能。前者根据光谱强度组织像素, 使得具有相似模式的像素在空间上也相邻。后者自适应地重新缩放各波段之间的差异, 帮助网络更好地建模反射与透射之间的差分表示。

光谱码本。 我们在配对的 RGB 光谱图像上进行 Spectral Codebook 的预训练, 以实现光谱重建。具体的实验数据集信息可以在实验部分获得。为了与离散光谱数据对齐, 我们将整个代码本分为 31 个相等的部分, 每一部分对应一个可见光谱带 (例如, 400-410nm, 410-420nm 等等) (Kaya, Can, and Timofte 2019)。如图 3 所示, 给定一个 RGB 图像 $\mathbf{x} \in \mathbb{R}^{3 \times H \times W}$, 首先将其输入编码器以提取特征 $\hat{\mathbf{z}} = \mathcal{E}_s(\mathbf{x}) \in \mathbb{R}^{n_z \times H \times W}$, 其中 H , W 和 n_z 分别表示特征的高度、宽度和潜在向量的维度。接下来, 对于第 i 个频带, 找到 $\hat{\mathbf{z}}$ 中每个像素的最近邻 $\mathbf{z}_k^i \in \mathbb{R}^{n_z}$ 在相应的光谱码本部分 $\mathcal{Z}^i \in \mathbb{R}^{K \times n_z}$ 中, K 是第 i 个码本部分的长度, 得到频带感知矢量量化 (Band-Wise VQ) 表示 $\tilde{\mathbf{z}}$:

$$\tilde{\mathbf{z}}_{hw}^i = \mathbf{q}(\hat{\mathbf{z}}_{hw}) = \arg \min_{\mathbf{z}_k^i \in \mathcal{Z}^i} (\|\hat{\mathbf{z}}_{hw} - \mathbf{z}_k^i\|_2), \quad (1)$$

$$\tilde{\mathbf{z}} = \text{Concat}(\tilde{\mathbf{z}}^i), i \in \{0, 1, \dots, 30\},$$

其中 $\hat{\mathbf{z}}_{hw}$ 表示空间域中的每个像素, $\mathbf{q}(\cdot)$ 表示逐元素量化, 而 $\text{Concat}(\cdot)$ 是连接操作。解码器 $\mathcal{D}_s$ 将量化表示 $\tilde{\mathbf{z}}$ 映射回 RGB 空间。整体的重建范式可以表述为:

$$\mathbf{S} = \mathcal{D}_s(\mathbf{z}') = \mathcal{D}_s(\mathbf{q}(\mathcal{E}_s(\mathbf{x}))), \quad (2)$$

其中 $\mathbf{S} \in \mathbb{R}^{31 \times H \times W}$ 表示重构光谱, 31 是光谱带的数量。在恢复过程中, 仅使用光谱码本根据提取的 RGB 特征来重构光谱。

利用光谱码本的优势有三点：首先，光谱表示在空间上是稀疏的而光谱相似 (Cai et al. 2022)，我们的按带矢量量化可以更好地捕捉空间差异同时减轻对冗余波长维度的过度强调。其次，光谱码本分别重建每个光谱带以与训练数据的离散分布相一致；同时，每一部分都可以恢复其对应波段的详细光谱表示。第三，我们的光谱码本可以无缝集成到反射去除网络中，实现对编码特征进行光谱重建而不引入额外的网络组件。与其他直接重构方法的比较见表 5。

光谱空间排序。 不同于噪声或雾霾等常见的退化现象，反射包含了丰富的结构和纹理细节。由于 CNN 固有的平移不变性，在从包含反射和透射的局部区域学习时，网络经常遇到不稳定的梯度，导致恢复失败 (Sun et al. 2024; Cui et al. 2024)。因此，我们提出了光谱空间排序 (SSS)，在不修改卷积范式的情况下缓解这一问题，同时也有助于后续模块区分不同的成分。

形式上，给定一个输入频谱特征 $S \in \mathbb{R}^{C \times H \times W}$ ，其中 C 表示其通道，我们在高度 hi 和宽度 wi 维度上进行排序如下：

$$S', id = \text{Sort}_{wi}(\text{Sort}_{hi}(S)), \quad (3)$$

其中 $\text{Sort}(\cdot)$ 表示排序操作， id 代表排序顺序，而 S' 指示排序后的输出频谱。排序后的光谱根据光谱差异被聚类到左上角和右下角两个区域，从而将空间分布与光谱中的图案差异关联起来。

值得注意的是，为了使光谱和 RGB 特征对齐以进行互补学习，我们也根据顺序 id 对输入的 RGB 特征进行了排序。在将输出送入解码器之前，我们将输出特征恢复到它们原来的顺序。这样的设计平滑了局部反射区域中的梯度，防止了由于恢复过程中梯度爆炸可能产生的模糊或伪影 (Dong et al. 2021)。

谱对角重新缩放。 为了增强反射和传输组件在波长维度上的差异，即，特征的通道维度，我们引入了光谱对角线重新缩放 (SDRS) 模块来自适应地重新缩放光谱。

具体来说，我们将排序后的光谱特征 S' 沿反对角线在空间上分割为两个三角区域：左上方部分和右下方部分，分别表示为 S^{tl} 和 S^{br} 。然后我们计算这两个区域在每个频带上的光谱差 D_i ：

$$S^{tl}, S^{br} = \text{Split}(S'_i),$$

$$D_i(S^{tl}, S^{br}) = \left| \sum_{h=1}^{H_{tl}} \sum_{w=1}^{W_{tl}} S_{hw}^{tl} - \sum_{h=1}^{H_{br}} \sum_{w=1}^{W_{br}} S_{hw}^{br} \right|, \quad (4)$$

$$D = \text{Concat}(D_i), i \in \{0, 1, \dots, B-1\},$$

其中 S_i 表示第 i 个频带， B 代表总频带数，而 $\text{Split}(\cdot)$ 是特征分割操作。为了确保自适应重新缩放，我们引

入了一个可学习的缩放因子 α ，它被应用于逐元素乘法 $\odot$ 与计算出的频谱差值 $D \in \mathbb{R}^{B \times 1}$ ，然后通过一个 $\text{softmax}(\cdot)$ 函数获得每个频带的重新缩放权重 $W \in \mathbb{R}^{B \times 1}$ ：

$$W = \text{softmax}(D \odot \alpha). \quad (5)$$

最后，我们将重新缩放权重 W 与输入光谱 S 相乘以获得精细的频谱特征 S_r ：

$$S_r = W \otimes S, \quad (6)$$

其中 $\otimes$ 表示通道逐元素乘法。重新缩放的光谱在波长维度上表现出更大的差异，减少了冗余，并帮助模型精确地定位反射区域。

频谱感知变压器

由于强大的长距离模型能力，基于 Transformer 的方法可以建模完整的反射结构以实现更彻底的去除。然而，自注意力机制难以适应性地关联不同模式下的远距离特征 (Sun et al. 2024)，限制了其处理空间变化反射的能力。此外，不当结合光谱可能导致从原始像素域中丢失颜色细节。

为了解决这些挑战，我们采用了一种双域学习变压器，使优化后的频谱先验能够有效引导像素域的恢复。具体而言，我们的 SAFormer 包括差分离散标记器 (DST)，它分别编码来自不同组件的特征，并且使用互补引导多头自注意力 (CG-MSA) 在频谱和像素域中共同去除反射。为了补偿排序过程中丢失的上下文信息，我们提出了上下文补偿前馈网络 (CC-FFN) 来弥补这一点。

可分离微分标记器。 传统的 Transformer 在标记化过程中仅考虑位置编码，导致反射和传输成分被投影到同一个标记中，进一步加剧了全局注意力中的混淆。为了解决这个问题，我们提出了 DST 来独立提取单一模式特征，而不受其他组件的干扰。

在公式方面，给定排序后的输入 RGB 特征 $F_r \in \mathbb{R}^{C \times H \times W}$ 和优化后的光谱 $S_r \in \mathbb{R}^{C \times H \times W}$ ，DST 应用 $\text{Split}(\cdot)$ 操作沿反对角线将特征分为左上部分 F^{tl} 和右下部分 F^{br} ，促进对具有单一模式的特征进行分词化，

$$F^{tl}, F^{br} = \text{Split}(F_r),$$

$$S^{tl}, S^{br} = \text{Split}(S_r). \quad (7)$$

对于每一部分，DST 使用 1×1 卷积核在通道维度上聚合特征。随后，我们应用一个 3×3 深度可分离卷积来编码排序特征的空间模式相关上下文。最后，我们将投影重塑，得到最终输出的标记：

$$Q^{tl} = W_Q^{tl} S^{tl}, K^{tl} = W_K^{tl} F^{tl}, V^{tl} = W_V^{tl} F^{tl},$$

$$Q^{br} = W_Q^{br} S^{br}, K^{br} = W_K^{br} F^{br}, V^{br} = W_V^{br} F^{br}, \quad (8)$$

Methods	Venue	<i>Nature</i> (20)		<i>Real</i> (20)		<i>SIR</i> ² (454)		<i>Average</i> (494)	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Input Image	-	20.44	0.785	18.96	0.733	22.76	0.885	22.51	0.884
BDN	ECCV 2018	18.83	0.738	18.64	0.726	21.61	0.854	21.50	0.844
FRS	CVPR 2019	20.01	0.756	18.63	0.719	22.23	0.867	21.99	0.867
Zhang et al.	CVPR 2018	22.31	0.804	20.16	0.767	23.07	0.869	22.92	0.862
ERRNet	CVPR 2019	22.57	0.807	20.67	0.781	22.97	0.885	22.85	0.877
RMNet	CVPR 2019	21.08	0.730	19.93	0.718	21.66	0.843	21.57	0.834
Kim et al.	CVPR 2020	20.10	0.759	20.22	0.735	23.57	0.877	23.30	0.886
IBCLN	CVPR 2020	23.90	0.787	21.42	0.769	24.05	0.888	23.94	0.878
YTMT	NeurIPS 2021	20.69	0.777	22.94	0.815	23.57	0.889	23.43	0.882
LANet	ICCV 2021	23.51	0.810	23.40	<u>0.826</u>	23.04	0.898	23.07	0.891
PNACR	ACM MM 2023	23.93	0.807	22.57	0.806	24.14	0.894	24.06	0.888
DSRNet	ICCV 2023	21.24	0.789	<u>24.23</u>	0.820	25.45	0.901	24.65	0.893
Zhu et al.*	CVPR 2024	<u>25.96</u>	<u>0.843</u>	23.82	0.817	25.45	0.910	25.40	0.904
Zhong et al.†	CVPR 2024	23.87	0.812	24.05	0.824	25.86	0.919	25.72	0.914
L-DiffER†	ECCV 2024	23.95	0.831	23.77	0.821	25.18	0.911	25.08	0.905
Ours	-	25.99	0.850	24.36	0.834	25.86	<u>0.912</u>	25.81	<u>0.907</u>

表 1: 三个反射数据集的定量比较。**粗体**和 下划线 分别表示最佳和第二好的性能。* 表示使用额外训练数据的方法，而 † 表示使用额外语言标签。

其中 W_Q, W_K, W_V 分别表示查询、键和值的投影卷积。通过 DST，我们将光谱特征作为查询 $Q \in \mathbb{R}^{m \times HW}$ ，将 RGB 特征作为键 $K \in \mathbb{R}^{m \times HW}$ 和值 $V \in \mathbb{R}^{m \times HW}$ ，其中 m 表示标记的维度。

互补引导多头自注意力。 为了确保频谱先验能够正确引导恢复，并和谐地融合来自两个域的特征，我们提出了 CG-MSA 以互补恢复退化的反射区域。给定 DST 输出标记 Q, K 和 V ，我们的 CG-MSA 依次对它们进行矩阵乘法运算，得到最终输出：

$$\text{CG-MSA}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{m}}\right)V. \quad (9)$$

为了减轻乘法后重叠反射区域相似性的快速增加，我们在 $\text{softmax}(\cdot)$ 函数之前引入了 $\sqrt{m}$ 将输入缩放到合理范围。我们的 CG-MSA 的整体计算可以表述如下，

$$\begin{aligned} F_o^{tl} &= \text{CG-MSA}(\text{LN}(Q^{tl}, K^{tl}, V^{tl})) \oplus F^{tl}, \\ F_o^{br} &= \text{CG-MSA}(\text{LN}(Q^{br}, K^{br}, V^{br})) \oplus F^{br}, \\ \hat{F} &= F_o^{tl} \oplus F_o^{br}, \end{aligned} \quad (10)$$

其中 $F_o^{tl} \in \mathbb{R}^{C \times H \times W}$ 和 $F_o^{br} \in \mathbb{R}^{C \times H \times W}$ 表示分离的 RGB 特征， $\hat{F} \in \mathbb{R}^{C \times H \times W}$ 代表最终输出， $\oplus$ 表示逐元素相加。由于两个域中的特征表现出相似的分布，这样

的设计有利于具有更清晰光谱指导的互补学习。此外，不同光谱带对纹理的不同敏感性可以进一步丰富局部细节。

$$\begin{aligned} F_{local} &= \text{DConv}(F_{in}), \\ \theta &= \text{softmax}(\text{Conv}_{1 \times 1}(\text{Concat}(F_{local}, \hat{F}))), \\ F_{out} &= \theta \hat{F} \oplus (1 - \theta) F_{local}. \end{aligned} \quad (11)$$

通过这种方式，CC-FFN 有效地以极小的代价弥补了排序学习过程中丢失的上下文。

上下文补偿前馈网络。 虽然我们的方法解决了病态的重叠成分学习问题，但它可能会导致上下文信息的丢失。例如，如果一个物体的一部分受到反射影响，DST 可能将其分别投影到不同的标记中，从而降低结构完整性。因此，我们提出了上下文补偿前馈网络 (CC-FFN)，它在排序前后交互融合特征以弥补局部上下文。

具体而言，如图 2 所示，CC-FFN 采用了一个可变形卷积层 (Zhu et al. 2019) $\text{DConv}(\cdot)$ 来强调输入特征 $F_{in} \in \mathbb{R}^{C \times H \times W}$ 中的局部上下文特征 $F_{local} \in \mathbb{R}^{C \times H \times W}$ 。然后，CC-FFN 将局部特征 F_{local} 与来自 CG-MSA 的输出特征 $\hat{F}$ 连接起来，并使用一个 1×1

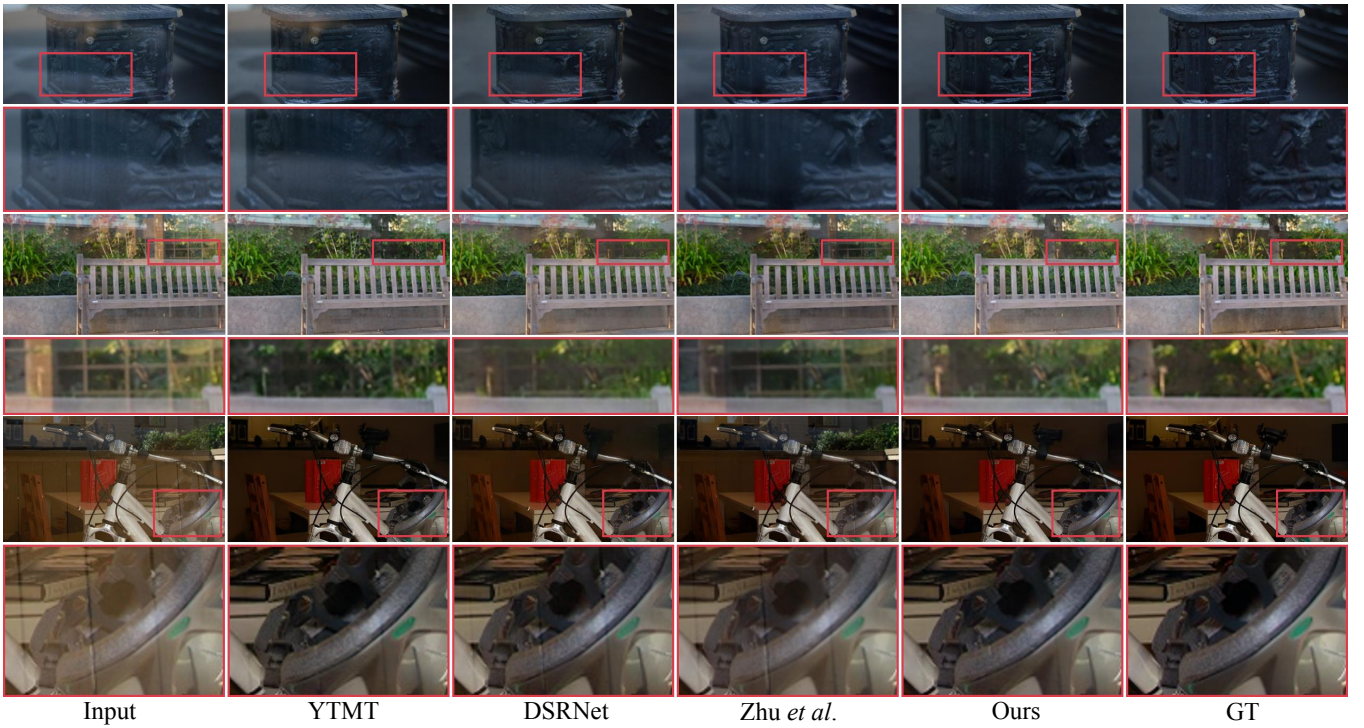


图 4: 从真实世界反射数据集恢复的传输图像的可视化结果 (Zhang, Ng, and Chen 2018)。

卷积层 $\text{Conv}_{1 \times 1}(\cdot)$ 来估计特征融合比 θ ，以加权融合 F_{local} 和 $\hat{F}$ 。

实验

数据集

我们遵循现有设置以保证公平性 (Hu and Guo 2023; Zhu et al. 2024)，使用来自 Real 数据集 (Zhang, Ng, and Chen 2018) 的 90 对真实反射图像，来自 Nature 数据集 (Li et al. 2020) 的 200 对真实图像，以及来自 PASCAL VOC 数据集 (Everingham et al. 2010) 的 7,643 对合成图像。在测试中，我们在 Nature、Real 和 SIR^2 (Wan et al. 2017) 数据集上评估所有方法。

我们进一步在其他两种反射场景中评估我们的方法。包含使用偏振相机从不同角度捕获的塑料薄膜反射图像的数据集 (Tang et al. 2024)。SSHR (Fu et al. 2023) 是一个高光去除数据集，包含 117,000 张训练图像和 18,000 张测试图像。

实现细节

为了预训练我们的光谱代码本，我们使用了 NTIRE 2022 光谱重建数据集 (Arad et al. 2022)，其中包含 1,000 对 RGB-光谱配对。在训练移除网络时，我们使用 Adam 优化器 (Kingma 2014)，并设 $\beta_1 = 0.9$

Methods	PSNR	SSIM	Methods	PSNR	SSIM
SHIQ	21.10	0.7465	Yang	23.243	0.894
Polar-HR	22.18	0.7102	Fu	23.270	0.881
Uformer	31.85	0.9457	JSHDR	26.979	0.895
Restormer	34.51	0.9759	Wu	25.731	0.894
Tang et al.	<u>36.94</u>	0.9850	TSHR	<u>28.633</u>	0.940
Ours	37.15	<u>0.9826</u>	Ours	30.257	<u>0.936</u>

表 2: 薄膜反射数据集的定量比较。表 3: SSHR 数据集上的定量比较。

和 $\beta_2 = 0.999$ 。学习率固定为 10^{-4} 。

与最新方法的比较

为了公平比较，我们遵循 DSRNet 的设置，并与 14 种最先进的方法进行比较。对于反射膜和镜面高光去除，我们在 (Tang et al. 2024) 和 (Fu et al. 2023) 中使用相同的设置。我们从最新的最先进工作及其原始论文中获得性能。

定量比较。 我们的方法在表 1 中的三个反射去除基准上实现了 SOTA 性能。这种优越性源于我们新颖的光谱-像素交互学习，这有效地解耦了复杂的反射伪影。此外，我们的模型甚至超越了使用额外数据 (Zhu et al.

基础	弹簧率		SAFormer		峰值信噪比	结构相似性指标	Methods	PSNR	SSIM
	SSS	SDRS	DST	CG-MSA	CC-FFN				
							RGB Stream	24.65	0.893
							Spectrum Stream	23.81	0.886
✓						24.65	无 Spectrum	24.65	0.893
✓	✓					24.99	Weighted Attn	25.06	0.896
✓	✓	✓				25.11	Memory Bank	25.16	0.902
✓			✓	✓	✓	25.47	Sparse Coding	25.19	0.898
✓	✓	✓	✓			25.43	Vanilla VQ	25.32	0.900
✓	✓	✓	✓	✓		25.75			
✓	✓	✓	✓	✓	✓	25.81	Ours	0.907	25.81 0.907

表 4: 我们的方法的消融研究结果。SPR: 光谱先验重建。SAFormer: 频谱感知变压器。

表 5: 网络架构和光谱重建策略的消融研究结果。

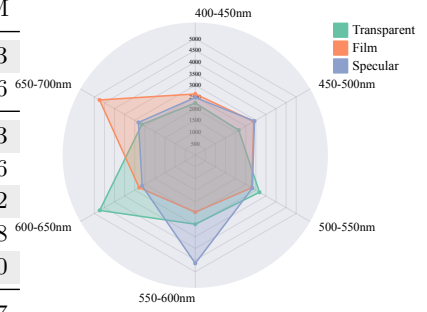


图 5: 不同反射场景中代码激活频率的可视化。

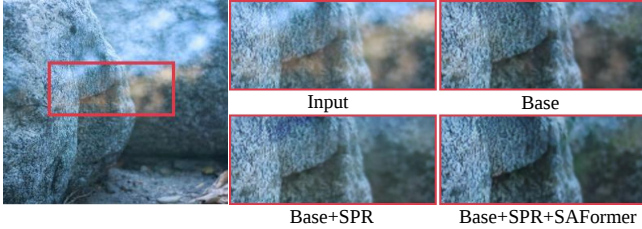


图 6: 我们的方法的消融研究可视化。

2024) 或语言先验 (Zhong et al. 2024; Hong et al. 2024) 的近期方法。我们进一步在表 2 和 3 中的电影反射 (Tang et al. 2024) 和镜面高光移除 (SSHR) (Fu et al. 2023) 数据集上展示了方法的鲁棒性和泛化能力，其中它始终优于现有方法。

定性比较。 为了进一步比较不同方法的视觉质量，我们在图 4 中展示了我们的方法和其他 SOTA 方法在不同反射数据集上恢复的视觉结果。我们的方法提供了更好的视觉质量，表现为反射残留较少，正确地恢复了反射区域的颜色，并且恢复了更多的细节纹理。更多视觉结果可以参考补充材料。

消融研究

在本节中，我们对所提出的方法进行了消融实验，并分析了不同网络架构的影响。此外，我们在不同的场景下进一步评估了我们的光谱码本的可靠性和鲁棒性。更多消融实验结果可以在补充材料中获得。

个体组件。 我们分析了 Tab. 4 和 figure 6 中我们方法的每个组件的影响。我们将带有 ResBlocks 的方法直接移除作为 “Base” 模型，并逐步添加每个组件进行消融实验。SPR 的引入显著提高了 PSNR，从 24.65dB 到 25.11dB，验证了其通过光谱学习消除反射的能力。在 SAFormer 中单独添加每个组件都带来了有利的性能提

升，证明了我们方法中每个组件的有效性。对于 figure 6 中的视觉结果，通过添加 SPR 可以很好地抑制反射退化，证明光谱先验可以有效地帮助模型区分反射分量。SAFormer 的引入消除了残余反射，进一步恢复了丢失的纹理和颜色细节。

网络架构。 为了评估我们 SAFormer 的双域学习范式的可靠性，我们将它的性能与光谱和像素域中的两个单流网络分别进行了比较。表 5 中的结果证实了双流设计的重要性，显著优于两个基线。这种性能提升来自于两个域之间的协同效应：像素流保留了稳健的空间结构，而光谱流提供了恢复所需的精细纹理。

谱代码本。 为了证明我们频谱码本在估计频谱方面的优越性，我们在表 5 中将我们的分带矢量量化与其它类似方法进行了比较。它相对于其他四种方法表现出色。这是因为它们要么无法提供空间上精细的重构，要么无法单独建模不同频段的频谱表示。因此，我们使用带有我们频谱码本的分带矢量量化来最大化重构效果。

图 5 显示了我们在光谱码本中代码的激活频率。不同场景之间显著的分布差异可以被观察到，这证明了我们的光谱码本在学习多样的真实世界反射方面的鲁棒性和敏感性。

结论

在这篇论文中，我们探讨了反射成像的光谱特征，提出了一种新的视角来处理反射问题，通过学习反射和透射成分之间的光谱差异。特别是，我们提出了光谱码本以直接重建输入反射图像的光谱，并设计了两个光谱优化模块以增强其效果。此外，我们提出了 SAformer 在像素域和光谱域中共同学习。这样的设计有效地消除了各种现实场景中的反射，并恢复了被复杂反射退化所遮蔽的详细纹理。广泛的实验结果表明了它的优势以及

光谱学习在图像修复领域的潜力。

参考文献

- Arad, B.; Timofte, R.; Yahel, R.; Morag, N.; Bernat, A.; Cai, Y.; Lin, J.; Lin, Z.; Wang, H.; Zhang, Y.; et al. 2022. Ntire 2022 spectral recovery challenge and data set. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 863–881.
- Born, M.; and Wolf, E. 2013. Principles of optics: electromagnetic theory of propagation, interference and diffraction of light. Elsevier.
- Cai, Y.; Lin, J.; Lin, Z.; Wang, H.; Zhang, Y.; Pfister, H.; Timofte, R.; and Van Gool, L. 2022. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 745–755.
- Cui, Y.; Ren, W.; Cao, X.; and Knoll, A. 2024. Revitalizing convolutional network for image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence.
- Dong, Z.; Xu, K.; Yang, Y.; Bao, H.; Xu, W.; and Lau, R. W. 2021. Location-aware single image reflection removal. In Proceedings of the IEEE/CVF international conference on computer vision, 5017–5026.
- Everingham, M.; Van Gool, L.; Williams, C. K.; Winn, J.; and Zisserman, A. 2010. The pascal visual object classes (voc) challenge. IJCV, 88: 303–338.
- Fu, G.; Zhang, Q.; Zhu, L.; Li, P.; and Xiao, C. 2021. A multi-task network for joint specular highlight detection and removal. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 7752–7761.
- Fu, G.; Zhang, Q.; Zhu, L.; Xiao, C.; and Li, P. 2023. Towards high-quality specular highlight removal by leveraging large-scale synthetic data. In Proceedings of the IEEE/CVF international conference on computer vision, 12857–12865.
- Hong, Y.; Zhong, H.; Weng, S.; Liang, J.; and Shi, B. 2024. L-Differ: Single Image Reflection Removal with Language-based Diffusion Model. In European Conference on Computer Vision.
- Hu, Q.; and Guo, X. 2021. Trash or treasure? an inter-active dual-stream strategy for single image reflection separation. NeurIPS, 34: 24683–24694.
- Hu, Q.; and Guo, X. 2023. Single image reflection separation via component synergy. In Proceedings of the IEEE/CVF international conference on computer vision, 13138–13147.
- Huang, L.; Luo, R.; Liu, X.; and Hao, X. 2022. Spectral imaging with deep learning. Light: Science & Applications.
- Kaya, B.; Can, Y. B.; and Timofte, R. 2019. Towards spectral estimation from a single RGB image in the wild. In Proceedings of the IEEE/CVF international conference on computer vision, 3546–3555. IEEE.
- Kingma, D. P. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Levin, A.; and Weiss, Y. 2007. User assisted separation of reflections from a single image using a sparsity prior. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(9): 1647–1654.
- Li, C.; Yang, Y.; He, K.; Lin, S.; and Hopcroft, J. E. 2020. Single image reflection removal through cascaded refinement. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 3565–3574.
- Nguyen, V. P.; Li, Y.; Zhang, W.; Wang, X.; and Paulus, Y. M. 2018. Multi-wavelength, en-face photoacoustic microscopy and optical coherence tomography imaging for early and selective detection of laser induced retinal vein occlusion. Biomedical optics express, 9(12): 5915–5938.
- Piccinelli, L.; Yang, Y.-H.; Sakaridis, C.; Segu, M.; Li, S.; Van Gool, L.; and Yu, F. 2024. UniDepth: Universal monocular metric depth estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10106–10116.
- Qiu, K.; Shen, J.; Chen, W.; and Zhang, J. 2023. Effectiveness of specular removal from hyperspectral images in the colour spectral measurement of wool fibres. IET Image Processing, 17(11): 3143–3152.
- Sinha, S. N.; Kopf, J.; Goesele, M.; Scharstein, D.; and Szeliski, R. 2012. Image-based rendering for scenes with reflections. ACM Transactions on Graphics (TOG), 31(4): 1–10.

- Sun, S.; Ren, W.; Gao, X.; Wang, R.; and Cao, X. 2024. Restoring images in adverse weather conditions via histogram transformer. In *European Conference on Computer Vision*, 111–129. Springer.
- Tang, J.; Wu, R.; Xu, X.; Hu, S.; and Chen, Y.-C. 2024. Learning to Remove Wrinkled Transparent Film with Polarized Prior. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 24987–24996.
- Wan, R.; Shi, B.; Duan, L.-Y.; Tan, A.-H.; and Kot, A. C. 2017. Benchmarking single-image reflection removal algorithms. In *Proceedings of the IEEE/CVF international conference on computer vision*, 3922–3930.
- Wang, J.; Chen, M.; Karaev, N.; Vedaldi, A.; Rupprecht, C.; and Novotny, D. 2025. Vggt: Visual geometry grounded transformer. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 5294–5306.
- Xiong, Z.; Shi, Z.; Li, H.; Wang, L.; Liu, D.; and Wu, F. 2017. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In *Proceedings of the IEEE/CVF international conference on computer vision*, 518–525.
- Yang, Y.; Ma, W.; Zheng, Y.; Cai, J.-F.; and Xu, W. 2019. Fast single image reflection suppression via convex optimization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8141–8149.
- Zhang, X.; Ng, R.; and Chen, Q. 2018. Single image reflection separation with perceptual losses. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4786–4794.
- Zheng, Y.; Sato, I.; and Sato, Y. 2015. Illumination and reflectance spectra separation of a hyperspectral image meets low-rank matrix factorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1779–1787.
- Zhong, H.; Hong, Y.; Weng, S.; Liang, J.; and Shi, B. 2024. Language-guided Image Reflection Separation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 24913–24922.
- Zhu, X.; Hu, H.; Lin, S.; and Dai, J. 2019. Deformable convnets v2: More deformable, better results. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9308–9316.
- Zhu, Y.; Fu, X.; Jiang, P.-T.; Zhang, H.; Sun, Q.; Chen, J.; Zha, Z.-J.; and Li, B. 2024. Revisiting Single Image Reflection Removal In the Wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 25468–25478.