深度协同 Mamba 携手记忆专家，实现全天图像反射分离

Siyan Fang
华中科技大学 siyanfang@hust.edu.cn

Long Peng
中国科学技术大学 l
ongp2001@mail.ustc.edu.cn

Yuntao Wang华中科技大学
yuntaowang@hust.edu.cn

Ruonan Wei华中科技大学
ruonan2765@gmail.com

Yuehuan Wang*华中科技大学
yuehwang@hust.edu.cn

arXiv:2601.00322v1 [cs.CV] 1 Jan 2026

摘要

图像反射分离旨在从混合图像中分离出透射层和反射层。现有方法依赖单幅图像中的有限信息，当两层的对比度相近时，容易将它们混淆，而这一挑战在夜间尤为严峻。为解决这一问题，我们提出深度-记忆解耦网络（DMDNet）。该网络采用深度感知扫描（DAScan）引导 Mamba 关注显著结构，促进信息沿语义连贯性流动，从而构建稳定状态。与 DAScan 协同工作时，深度协同状态空间模型（DS-SSM）根据深度调节状态激活的敏感度，抑制干扰层解耦的模糊特征扩散。此外，我们提出记忆专家补偿模块（MECM），利用跨图像历史知识引导专家提供针对特定层的补偿。为解决夜间反射分离数据集缺乏的问题，我们构建了夜间图像反射分离（NightIRS）数据集。大量实验表明，DMDNet 在白天和夜间均优于最先进方法。

项目主页： https://github.com/fashyon/DMDNet

1 引言

通过玻璃等透明介质拍摄图像时，通常会出现反射伪影，这不仅会损害视觉质量，还会降低下游视觉任务的性能 [1–5]。图像反射分离的任务旨在将混合图像 $\boldsymbol{I}$ 分解为透射层 $\boldsymbol{T}$ 和反射层 $\boldsymbol{R}$，其中 $\boldsymbol{T}$表示玻璃后方的场景， $\boldsymbol{R}$ 表示玻璃表面上的反射内容。早期研究主要依赖梯度稀疏性 [6] 和反射模糊性 [7, 8], 等物理先验，利用基于物理假设的手工设计约束进行处理。然而，这些方法仅在受限场景下有效。随着深度学习 [9–39], 方法的发展，例如 Zhang 等人 [40] 和 DSIT [41] ，研究者开始从数据中学习 $\boldsymbol{T}$ 和 $\boldsymbol{R}$ 的隐式先验，以实现分离。然而，由于单幅图像中的信息有限，当 $\boldsymbol{T}$ 和 $\boldsymbol{R}$ 呈现相似对比度时，这些方法往往会遇到瓶颈。这一挑战在夜间场景中尤为严重。白天，充足的自然光照会增强 $\boldsymbol{T}$ 并抑制 $\boldsymbol{R}$，从而形成两个层次之间清晰的对比。夜间的光照来自随机分布的人工光源，导致不均匀光照条件。因此，由于全局光照不足， $\boldsymbol{T}$ 显得更暗，而局部强

*通讯作者。

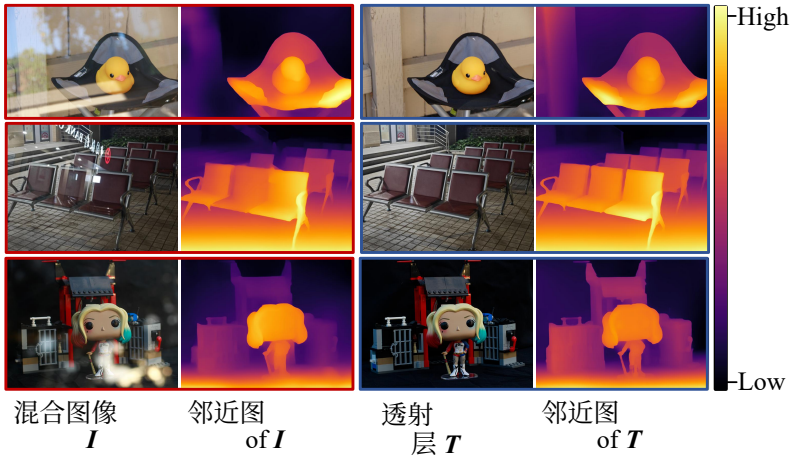


图 1：通过深度估计获取的白天、夜间和室内场景的邻近度图。深度估计能够穿透反射遮挡，捕捉 T 的底层结构。

照射到玻璃表面的光线会产生眩光和散射高光。因此，T和 R 呈现出相似的对比度，使得二者更难分离。

尽管这些困难并未得到直接解决，一些研究尝试通过引入额外的物理线索来进行补偿，例如多视图图像 [42, 43], 偏振滤光片 [44, 45], 红外摄像机 [46, 47], 和闪光照明 [48, 49]。然而，这类方法需要受控环境和额外设备，限制了其在实际应用中的灵活性。为摆脱对外部硬件的依赖，一些研究引入人机交互，例如语言提示 [50, 51]和手动区域标注 [52, 53]。然而，这些方法耗时费力。

深度估计无需额外硬件或人工干预，即可提供物理线索。对混合图像 [54] 进行深度估计后，我们观察到，所得邻近图会突出显示与 T对应的连贯清晰结构，同时自然抑制与 R 相关的模糊透明叠加层，如图 1 所示。这表明，高邻近度值往往包含显著结构。这些结构通常跨越较大的空间范围，例如建筑物的轮廓或一排椅子；要充分利用这些线索，需要具备捕捉长程依赖能力的模型。

Mamba [55] 凭借其状态空间模型（SSM）的高效长程建模能力， [56, 57], 在多个领域取得了令人瞩目的结果。VMamba [58] 通过四向扫描将这一能力引入视觉领域。然而，这种扫描策略在图像反射分离任务中存在两项局限：

（1）结构连续性受损。透射场景通常由连贯的轮廓、形状和纹理构成，例如窗户的边缘或人脸的曲线。固定顺序扫描会将这些内容割裂，导致结构线索失真，同时妨碍模型将这些语义实体作为整体进行感知。

（2）误差传播。在 SSM 中，较早扫描区域的状态会持续影响后续区域。如果首先传播的是模糊特征，其不确定性便会扩散至整幅图像，加剧分离误差。

为解决这些问题，我们提出深度协同解耦 Mamba（DSMamba）。其深度感知扫描策略（DA-Scan）针对 T和 R分别定制扫描策略，使模型在建模早期就能接触到显著结构，从而有助于建立语义连续性。与 DA-Scan 协同工作，我们设计了深度协同状态空间模型（DS-SSM），用于调节状态演化的活动，同时抑制模糊区域中的激活，以防止错误信息扩散。

为克服单幅图像信息有限的问题，我们引入记忆专家补偿模块（MECM），以利用跨图像历史知识。每位专家都配备一个用于存储特征模式的记忆库，MECM 会动态激活最相关的专家，以提供有针对性的补偿。例如，对于 T，可以激活专精于纹理细节和结构轮廓的专家；对于 R，则可以使用处理稀疏高光和模糊重影的专家。

为解决夜间图像反射分离数据集稀缺的问题，我们构建了夜间图像反射分离（NightIRS）数据集。该数据集包含在夜间反射条件下获取的 1,000 个图像三元组。

该数据集捕捉了夜间成像的独特复杂性，包括不均匀照明、强烈的人工光源以及多种反射伪影，而这些因素往往被现有公开数据集所忽略。
总体而言，本工作的贡献如下:

- 我们提出了 DSMamba，其中 DA-Scan 与 DS-SSM 协同工作，引导 Mamba 关注结构显著性并抑制错误传播。
- 我们引入 MECM，利用跨图像历史记忆进行针对性补偿。
- 我们构建了 NightIRS 数据集，用于评估夜间反射分离。
- 实验结果表明，DMDNet 的表现优于最先进方法（SOTAs）。

2 相关工作

图像反射分离。 早期研究 [6, 8] 依赖手工设计先验，但这类先验仅适用于简单情况。深度学习方法 [40, 41] 利用大规模数据学习从受污染图像到清晰图像的映射，但由于单幅图像所包含的信息有限，在复杂场景中往往表现不佳。为引入物理线索，一些方法利用多视图图像、偏振 [45], 闪光灯 [48], 或红外摄像机 [47], , 但这些方法需要额外硬件，因此不适用于互联网图像。为避免这一问题，Zhong 等人 [50] 引入了语言提示，而 FIRM[53] 依赖手动区域标注。然而，这些方法需要人工干预，从而限制了自动化程度。相比之下，深度估计无需外部传感器即可提供物理线索。Elnenaey 等人 [59] 将深度图粗略量化为四个等级，并将其与输入图像拼接以提供引导。DGR^2-Net [60] 对深度图进行全局池化，然后将其与输入图像拼接，用于双目反射去除。然而，这些方法缺乏细粒度深度引导，导致效果不足。更重要的是，它们忽略了深度图中蕴含的、对图像反射分离至关重要的结构显著性。

视觉 Mamba。 由于 Mamba 在长序列建模方面表现出色，近年来已被广泛应用于各种视觉任务 [56, 58, 61–63]。MambaIR [64] 和 VMambaIR [65] 是最早将 Mamba 引入图像复原领域的研究之一。随后，MambaIRv2 [66] 提出语义引导邻域交互机制，以促进信息传递。TAMambaIR [57] 提出多方向感受野扩展方案，以增强建模能力。然而，这些方法缺乏对几何结构敏感的动态状态建模策略，因此在反射分离中难以区分不同层。

混合专家（MoE）。 MoE 通过采用多个专家实现自适应计算，已广泛应用于图像复原任务。MoCE-IR [67] 设计了计算复杂度各异的专家模块，以适应不同的退化类型。FAME [68] 采用频率自适应 MoE 架构，对低频和高频分量采用不同的动态处理策略。然而，这些方法缺乏跨图像记忆，因此难以补偿单幅图像中的受污染信息。
记忆增强方法。 一些研究探索了用于图像复原的记忆机制。例如，Xu 等人 [69] 提出了一种纹理记忆，用于存储图像块样本，以指导纹理合成。ER^{2}Net [70] 利用记忆模块修复眼镜反光区域。然而，较高的计算成本限制了其只能一次性使用，使其不适合部署多个专家。此外，这些方法仅限于全局匹配或局部建模，缺乏能够实现自适应专家行为的统一机制。

3 方法

3.1 深度-内存解耦网络

深度-记忆解耦网络（DMDNet）由编码分支、深度语义调制分支（DSBranch）和解码分支组成，如图2所示。编码分支采用互门控交互模块（MuGI） [71] 来提取 $\boldsymbol{T}$和 $\boldsymbol{R}$的特征，其中 $\boldsymbol{E}_T^i$， $\boldsymbol{E}_R^i \in \mathbb{R}^{C_i \times \frac{H}{2^{i-1}} \times \frac{W}{2^{i-1}}}$，
$i \in \{1, 2, 3, 4, 5\}$。这里， C_i表示第 i层的通道数，而 H和 W分别表示输入图像 $\boldsymbol{I}$的高度和宽度。DSBranch利用深度语义特征$\boldsymbol{D}_S^3 \in \mathbb{R}^{96 \times \frac{H}{4} \times \frac{W}{4}}$， $\boldsymbol{D}_S^4 \in \mathbb{R}^{256 \times \frac{H}{8} \times \frac{W}{8}}$， $\boldsymbol{D}_S^5 \in \mathbb{R}^{512 \times \frac{H}{16} \times \frac{W}{16}}$，对为解码分支编码的特征进行调制。解码分支通过深度-记忆解耦块（DMBlock）和邻近图$\boldsymbol{P}_M^i \in \mathbb{R}^{C \times \frac{H}{2^{i-1}} \times \frac{W}{2^{i-1}}}$，
$i \in \{1, 2, 3, 4, 5\}$，实现 $\boldsymbol{T}$与 $\boldsymbol{R}$ 的分离。如图3(a)所示，DMBlock由DSMamba、MECM和EFFN [65]组成。

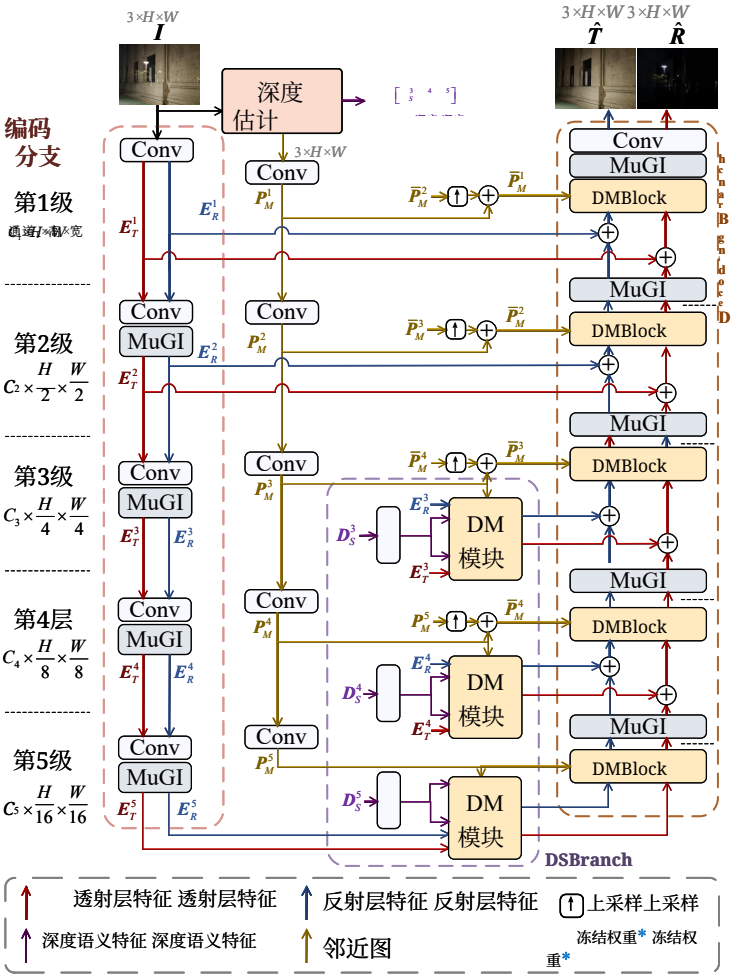


图2：深度-记忆解耦网络（DMDNet）。DMDNet 采用 DMBlock，利用深度和内存线索解耦 $\boldsymbol{T}$ 和 $\boldsymbol{R}$。

3.2 深度协同解耦 Mamba

为解决 Mamba 固定扫描策略的局限性，我们提出了深度协同解耦 Mamba（DSMamba）。如图 3(b)所示，DSMamba 由深度感知扫描（DAScan）和深度协同状态空间模型（DS-SSM）组成。DAScan采用深度感知区域扫描（DA-RScan）处理 $\boldsymbol{T}$，并采用深度感知全局扫描（DA-GScan）处理 $\boldsymbol{R}$。

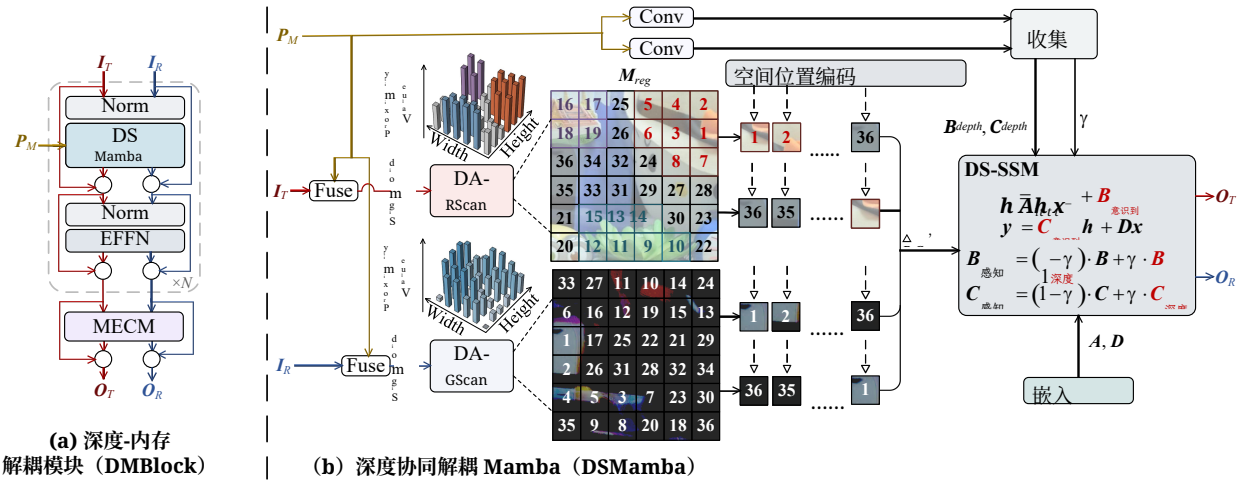
DA-RScan 遵循“先大区域、 + 由近及远”的方案。具体而言，邻近图被划分为区域扫描图 $\boldsymbol{M}_{reg}$。区域按面积从大到小依次扫描，因为较大的区域表示更显著的语义，背景区域则最后扫描，以确保完整性。这种基于区域的方案保留了同一物体内像素的语义连续性。在每个区域内，像素按由近及远的顺序扫描，优先处理结构显著结构。

DA-GScan 遵循“全局由近及远”的方案，从全局最近的像素扫描到最远的像素。该方案强调全局结构显著性，这与 $\boldsymbol{R}$ 的稀疏且不连续分布相契合，能够增强对反射特征的建模。最后，按相反顺序应用逆向 DAScan，以补充结构线索。

原始状态空间 模型（SSM）采用 Mamba 的统一状态更新机制，对所有区域进行处理， d as:

$$\boldsymbol{h}_t = \boldsymbol{A}\boldsymbol{h}_{t-1} + \boldsymbol{B}\boldsymbol{x}_t, \quad \boldsymbol{y}_t = \boldsymbol{C}\boldsymbol{h}_t + \boldsymbol{D}\boldsymbol{x}_t \tag{1}$$

其中 $\boldsymbol{A} \in \mathbb{R}^{N\times N}$、 $\boldsymbol{B} \in \mathbb{R}^{N\times 1}$、 $\boldsymbol{C} \in \mathbb{R}^{1\times N}$ 和 $\boldsymbol{D} \in \mathbb{R}$。 N是状态大小。该机制缺乏结构感知，因此难以分离 $\boldsymbol{T}$ 与 $\boldsymbol{R}$ 错综交织的区域。为克服这一



(a) 深度-内存解耦模块（DMBlock）　　(b) 深度协同解耦 Mamba（DSMamba）

图 3：DMBlock 和 DSMamba。DSMamba 通过 DAScan 优先处理显著结构，并协同调制状态激活。数字表示正向扫描顺序。

限制，同时与 DAScan 协同，我们设计了 DS-SSM，其状态更新定义如下：

$$\begin{aligned} \boldsymbol{h}_t &= \boldsymbol{A}\boldsymbol{h}_{t-1} + \boldsymbol{B}_{aware}\boldsymbol{x}_t, \\ \boldsymbol{y}_t &= \boldsymbol{C}_{aware}\boldsymbol{h}_t + \boldsymbol{D}\boldsymbol{x}_t, \\ \boldsymbol{B}_{aware} &= (1-\boldsymbol{\gamma})\cdot\boldsymbol{B} + \boldsymbol{\gamma}\cdot\boldsymbol{B}_{depth}, \\ \boldsymbol{C}_{aware} &= (1-\boldsymbol{\gamma})\cdot\boldsymbol{C} + \boldsymbol{\gamma}\cdot\boldsymbol{C}_{depth} \end{aligned} \tag{2}$$

其中，$\boldsymbol{\gamma}$ 是一个取值范围为 0–1 的权重图，由邻近图推导而来。$\boldsymbol{B}_{depth}$ 和 $\boldsymbol{C}_{depth}$ 是深度引导状态矩阵，分别控制状态更新的幅度以及状态对输出的贡献。

在结构显著区域，较大的 $\boldsymbol{\gamma}$ 会增强 $\boldsymbol{B}_{depth}$ 和 $\boldsymbol{C}_{depth}$的影响，加快清晰结构的整合，并强化其对输出的引导作用。相反，在结构模糊区域，则会抑制这种干预，以防止模糊特征传播。

空间位置编码。为了在扫描过程中强化位置特异性，DSMamba采用基于二维正弦和余弦函数的空间位置编码（SPE）：

$$\begin{aligned} \boldsymbol{PE}_x &= [\sin(x\cdot f_i),\ \cos(x\cdot f_i)], \\ \boldsymbol{PE}_y &= [\sin(y\cdot f_i),\ \cos(y\cdot f_i)] \end{aligned} \tag{3}$$

其中 x 和 y 表示归一化空间坐标，f_i 表示不同的频带 .

结合水平和垂直编码后，可得到位置嵌入 $\boldsymbol{PE}\in\mathbb{R}^{H\times W\times d_{inner}}$，其中$d_{inner}$ 是状态空间模型的通道维度。该嵌入会根据扫描顺序重新排列，并添加到状态特征中，为状态建模提供位置信息。

3.3 记忆专家补偿模块

为了利用跨图像累积知识进行针对性补偿，我们引入了记忆专家补偿模块（MECM），如图4所示。MECM由专家门控 [68] 和记忆专家组成。专家门控负责从 N_{Exp} 个候选项中选择最相关的 N_{Exp}^{K} 个专家。记忆专家执行特征检索和演化，从而利用历史知识实现自适应补偿。

记忆专家由全局模式交互流（GPStream）和空间上下文细化流（SCStream）组成。GPStream进一步分为全局模式调整和记忆演化。

对于全局模式调整，首先将输入图像 $\boldsymbol{I}\in\mathbb{R}^{B\times C\times H\times W}$池化为全局表示 $\boldsymbol{I}_G\in\mathbb{R}^{B\times C}$，并用它计算与记忆库 $\boldsymbol{Mem}\in\mathbb{R}^{M\times C}$的相似度，得到相似度分数矩阵 $\boldsymbol{S}\in\mathbb{R}^{B\times M}$，其中 B是批量大小，M是记忆项的数量。我们分别沿记忆维度和图像维度应用 Softmax，得到 $\boldsymbol{S}_I$ 和 $\boldsymbol{S}_M$。这里，$\boldsymbol{S}_I$ 表示每张图像在所有记忆项上的匹配分布，而 $\boldsymbol{S}_M$ 表示每个记忆项对图像的贡献。接下来，$\boldsymbol{S}_I$ 用于

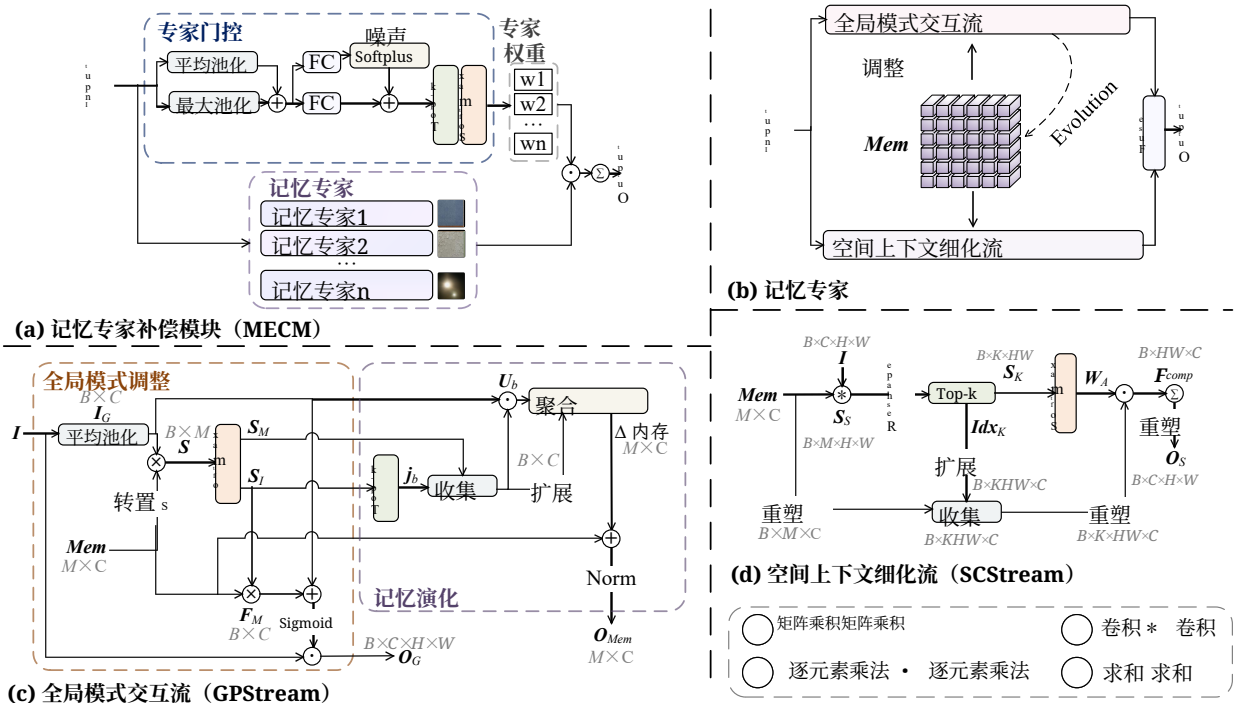


图4：记忆专家补偿模块（MECM）及其组成部分，该模块利用跨图像历史知识指导解耦。每个记忆专家由用于全局调整和记忆演化的 GPStream，以及用于空间级细化的 SCStream 组成。

对 $\boldsymbol{Mem}$进行加权聚合，生成记忆响应特征 $\boldsymbol{F}_M \in \mathbb{R}^{B\times C}$。最后， $\boldsymbol{F}_M$与 $\boldsymbol{I}_G$ 交互，生成一个用于调制输入 $\boldsymbol{I}$的注意力掩码，从而产生全局补偿 $\boldsymbol{O}_G$。

记忆演化旨在提供反馈并更新记忆库。对于每个图像样本 $b \in [1, B]$，从匹配矩阵 $\boldsymbol{S}_I$中选出响应最强的记忆索引 $j_b \in [1, M]$。相应的分数 $\boldsymbol{S}_M[b, j_b]$用作权重，与全局表示 $\boldsymbol{I}_G[b]$相乘，得到更新向量 $\boldsymbol{U}_b \in \mathbb{R}^C$。所有 $\boldsymbol{U}_b$向量沿其对应的索引 j_b聚合，形成记忆增量 $\Delta\boldsymbol{Mem} \in \mathbb{R}^{M\times C}$:

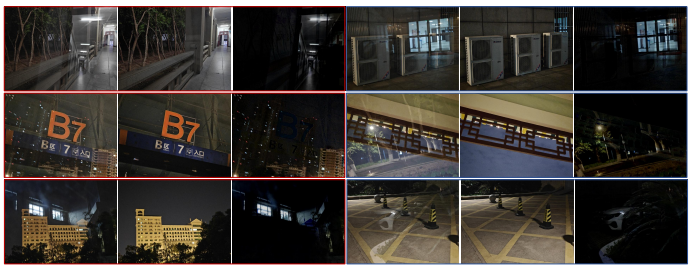
$$\Delta\boldsymbol{Mem}[m] = \sum_{b\in[1,B],\ j_b=m} \boldsymbol{U}_b, \quad m \in [1, M] \tag{4}$$

最后，以残差方式更新记忆库，得到更新后的记忆 $\boldsymbol{O}_{Mem}$.

SCStream 专注于空间上下文补偿。首先，将记忆库 $\boldsymbol{Mem}$重塑为卷积核，并与输入图像 $\boldsymbol{I}$进行卷积，以获得相似度图 $\boldsymbol{S}_S \in \mathbb{R}^{B\times M\times H\times W}$。 $\boldsymbol{S}_S[b, m, h, w]$表示位置（$h$， w）与第 m 个记忆项之间的相似度。接下来，对于每个空间位置，选出 Top-k 个最相关的记忆项。具体而言， $\boldsymbol{Idx}_K$， $\boldsymbol{S}_K \in \mathbb{R}^{B\times K\times HW}$（其中 $HW = H \times W$）分别表示每个像素对应的 Top-k 个记忆项的索引和相似度分数。使用 softmax 对相似度分数 $\boldsymbol{S}_K$ 进行归一化，得到注意力权重 $\boldsymbol{W}_A \in \mathbb{R}^{B\times K\times HW}$，以表示每个像素与 Top-k 个记忆项之间的匹配程度。然后，利用 $\boldsymbol{Idx}_K$从记忆库中检索相应的记忆特征。检索得到的记忆张量记为 $\boldsymbol{Mem}_K \in \mathbb{R}^{B\times K\times HW\times D}$，其中包含与每个像素位置关联的 Top-k 个记忆项的特征。加权和得到补偿特征 $\boldsymbol{F}_{comp} \in \mathbb{R}^{B\times HW\times D}$，最终输出 $\boldsymbol{O}_S$ 通过重塑获得。加权和的计算方式如下:

$$\boldsymbol{F}_{comp}[b, hw, d] = \sum_{k=1}^{K} \boldsymbol{W}_A[b, k, hw] \cdot \boldsymbol{Mem}_K[b, k, hw, d] \tag{5}$$

每个专家采用不同的卷积，将 GPStream 和 SCStream 的特征融合起来，从而捕捉特定的语义关系并实现自适应细化。



I ***T*** ***R*** ***I*** ***T*** ***R***

图 5：NightIRS 数据集中的示例。 $\boldsymbol{I}$、 $\boldsymbol{T}$和 $\boldsymbol{R}$ 分别表示混合图像、透射层和反射层。

方法	自然（20）	真实（20）	野外（55）	明信片（199）	纯色 (200)	平均值
	峰值信噪比↑结构相似性↑学习感知图像块相似度↓	峰值信噪比↑结构相似性↑学习感知图像块相似度↓	峰值信噪比↑结构相似性↑学习感知图像块相似度↓	峰值信噪比↑结构相似性↑学习感知图像块相似度↓	峰值信噪比↑结构相似性↑学习感知图像块相似度↓	峰值信噪比↑结构相似性↑学习感知图像块相似度↓
BDN (ECCV'18)	18.83 0.737 0.242	18.68 0.728 0.284	22.02 0.822 0.181	20.54 0.857 0.177	22.68 0.856 0.125	20.55 0.800 0.202
ERRNet (CVPR'19)	20.43 0.756 0.172	23.03 0.810 0.156	23.87 0.848 0.132	21.81 0.874 0.152	24.72 0.896 0.095	22.77 0.837 0.141
IBCLN (CVPR'20)	23.78 0.784 0.145	21.59 0.764 0.210	24.46 0.885 0.134	22.95 0.875 0.155	24.74 0.893 0.097	23.50 0.840 0.148
LANet (ICCV'21)	23.55 0.811 0.115	22.51 0.815 0.145	26.06 0.900 0.109	24.14 0.907 0.106	24.30 0.898 0.087	24.11 0.866 0.112
YTMT (NIPS'21)	20.77 0.769 0.178	22.86 0.807 0.158	25.07 0.892 0.116	22.40 0.881 0.147	24.70 0.899 0.092	23.16 0.850 0.138
DMGN (TIP'21)	20.63 0.764 0.167	20.28 0.763 0.215	21.34 0.774 0.152	22.65 0.879 0.151	23.27 0.872 0.102	21.63 0.810 0.157
HGNet(TNNLS'23)	25.23 0.824 0.111	23.65 0.818 0.155	26.88 0.897 0.109	23.56 0.900 0.124	25.00 0.900 0.092	24.86 0.868 0.118
DSRNet（ICCV'23）	21.62 0.781 0.149	23.41 0.805 0.147	24.35 0.893 0.117	24.66 0.911 0.111	26.10 0.914 0.071	24.03 0.861 0.119
RDRNet(CVPR'24)	24.44 0.820 0.107	21.29 0.769 0.190	26.48 0.905 0.101	23.65 0.891 0.146	25.93 0.912 0.080	24.36 0.860 0.125
DSIT (NIPS'24)	26.05 0.830 0.128	24.34 0.823 0.136	27.55 **0.920** **0.081**	**26.01** 0.921 0.103	26.62 0.922 0.075	26.11 0.883 0.105
RDNet (CVPR'25)	25.77 0.828 0.108	**25.13** **0.838** **0.117**	27.59 0.915 0.085	25.95 **0.921** **0.088**	26.59 0.922 0.069	26.21 0.885 0.094
DMDNet（我们的方法）	**26.68**					

表1：公开数据集上透射层的定量比较。DMDNet取得了最佳平均性能。**粗体**和下划线分别表示 Top-1 和 Top-2 结果。 ↑ 表示越高越好，而 ↓表示越低越好。

3.4 夜间图像反射分离数据集

夜间图像反射分离（NightIRS）数据集包含 1,000 组三元夜间反射图像。每组三元图像由 $\boldsymbol{I}$、 $\boldsymbol{T}$ 和 $\boldsymbol{R}$ 组成，如图 5 所示。我们使用不同厚度的玻璃和亚克力板引入反射干扰。为确保照明多样性，数据集采集自各种夜间环境，例如路灯、霓虹灯牌、亮灯建筑和低光照自然环境。为捕捉反射的几何变化，我们考虑了不同的相机到玻璃的距离和观看角度。该数据集还提供高分辨率版本（NightIRS-HR），为夜间反射分离提供可扩展的基准。

4 实验

4.1 实现细节

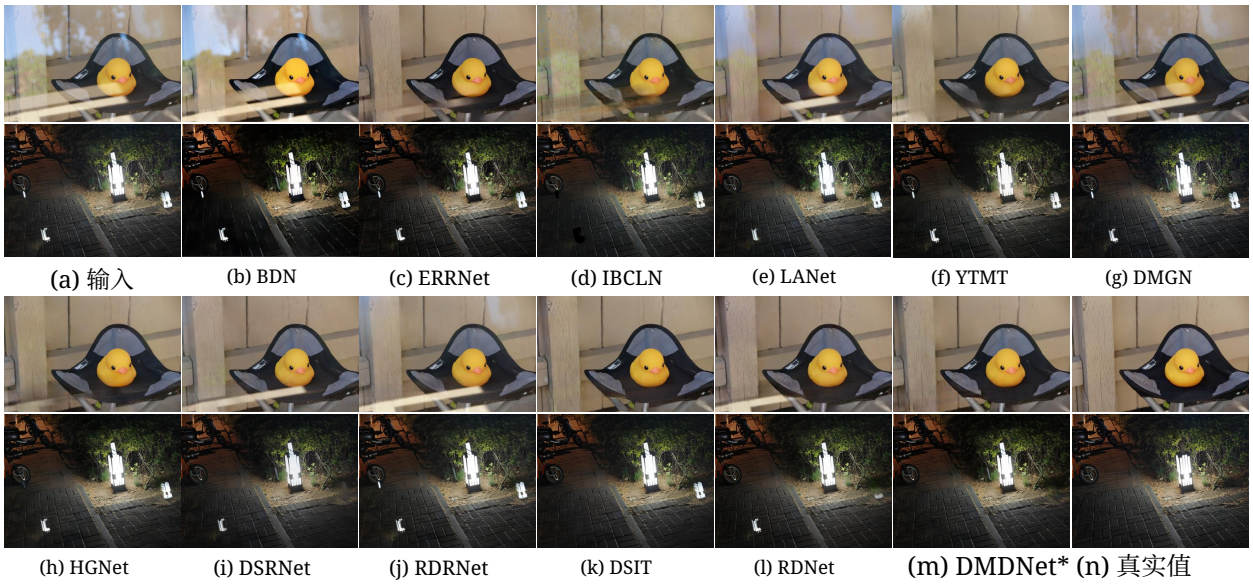
通道维度设为 C_1， C_2， C_3， C_4， $C_5 = [48, 96, 192, 384, 768]$。在 MECM 中， $N_{Exp} = 4$ 和 $N_{Exp}^K = 2$。我们采用批量大小为 1，并将图像裁剪为 352×352 图像块。训练过程中采用随机水平翻转进行数据增强。使用 Adam优化器 [72] 进行模型优化，初始学习率为 10^{-4}。模型训练 60 个训练轮次，并分别在第 30 和第 50 个训练轮次将学习率降至 5×10^{-5} 和 10^{-5}。所有实验均在单块 NVIDIA RTX 4090 GPU 上进行。更多详情请参阅补充材料。

4.2 数据集和评估指标

遵循先前的工作 [41, 71, 73, 74], ，我们使用来自 PASCAL VOC 数据集的 7,643 对图像[75], 200 、来自 Nature 数据集的图像对 [76], 以及来自 Real 数据集的 89 对图像 [40]训练模型。Nature 和 Real 数据集中剩余的图像，以及 SIR^2 数据集[77], 中的 Wild、Postcard 和 Solid 子集，还有 NightIRS 数据集，均用于测试。为避免 GPU 内存溢出，我们将 Real 数据集中的图像缩放至较长边为 420 像素，同时保持原始宽高比。

方法	透射层			反射层			参数 (M)↓	浮点运算次数 (G)↓
	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓		
BDN（ECCV'18）	20.52	0.680	0.293	8.79	0.082	0.843	75.16	12.70
ERRNet (CVPR'19)	22.43	0.767	0.180	N/A	N/A	N/A	18.95	116.72
IBCLN (CVPR'20)	23.16	0.803	0.196	20.54	0.292	0.701	21.61	98.16
LANet (ICCV'21)	23.68	0.817	0.171	21.61	0.280	0.472	10.93	83.81
YTMT (NIPS'21)	23.03	0.799	0.186	24.96	0.500	0.503	76.90	110.98
DMGN (TIP'21)	22.88	0.799	0.174	24.77	0.488	0.508	45.49	116.85
HGNet (TNNLS'23)	23.60	0.817	0.170	N/A	N/A	N/A	14.51	82.08
DSRNet（ICCV'23）	23.39	0.813	0.175	24.80	0.404	0.499	124.6	90.21
RDRNet（CVPR'24）	24.04	0.824	0.185	N/A	N/A	N/A	29.09	5.14
DSIT（NIPS'24）	24.61	0.827	0.168	27.18	0.569	0.372	131.76	74.18
RDNet（CVPR'25）	25.08	0.831	0.149	27.93	**0.636**	0.309	266.43	66.10
DMDNet（我们的方法）	**25.24**	**0.832**	**0.144**	**28.37**	0.633	**0.286**	87.22	39.33

表2：在 NightIRS 数据集上的最先进方法定量比较。处理一张 128×128 RGB图像的浮点运算次数。



(a) 输入 (b) BDN (c) ERRNet (d) IBCLN (e) LANet (f) YTMT (g) DMGN

(h) HGNet (i) DSRNet (j) RDRNet (k) DSIT (l) RDNet (m) DMDNet* (n) 真实值

图6：与最先进方法在透射层上的定性比较。我们的 DMDNet 在日间和夜间场景中都能最有效地去除反射。夜间图像取自 NightIRS 数据集。

为确保公平性，所有输出图像均以无损PNG格式保存，并在RGB色彩空间中计算评估指标，包括PSNR [78], 、结构相似性 [79], 和学习感知图像块相似度 [80],, 分别从像素级、结构和感知视角评估图像质量。

4.3 性能评估

我们将 DMDNet 与 11 种方法进行比较，包括 BDN [81], ERRNet [82], IBCLN [76], LANet [74], YTMT [83], DMGN [84], HGNet [85], DSRNet [71], RDRNet [86], DSIT [41], 和 RDNet[73]。表1展示了在公开数据集上的定量比较，这些数据集主要由日间场景构成，结果表明 DMDNet 的平均性能最佳。表2展示了在 NightIRS 数据集上的定量比较。DMDNet 在透射层和反射层上获得了数量最多的排名第一指标，体现出其对夜间反射的适应能力，同时参数量和浮点运算次数（FLOPs）也保持在合理范围内。

图6展示了透射层的定性比较。我们的DMDNet实现了最有效的恢复，在保留结构细节的同时抑制了日间场景中的残余反射。即使在反射与场景内容十分相似的夜间条件下，DMDNet也能有效去除反射。

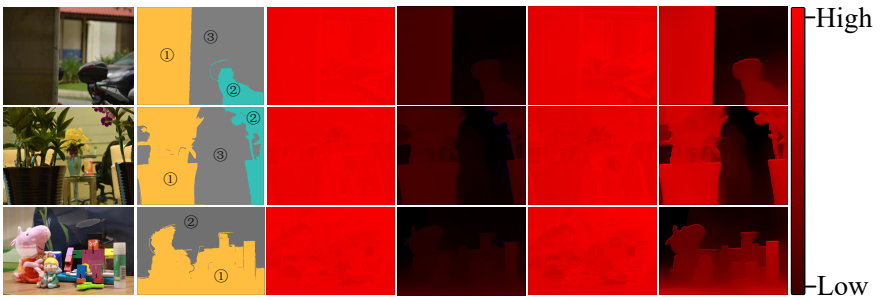


图7：DSMamba 可视化。 M_{reg} 展示了区域级扫描顺序。(c)–(f)展示了状态空间矩阵。B_{depth} 和 C_{depth} 更侧重于显著结构。

方法	透射层			反射层			参数 (M)↓	浮点运算次数 (G)↓
	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓		
MambaIR	25.56	0.880	0.106	22.09	0.500	0.420	103.61	42.43
VMambaIR	25.89	0.884	0.100	22.06	0.490	0.414	83.76	38.05
MambaIRv2	24.84	0.868	0.118	21.66	0.490	0.445	88.38	40.38
DSMamba（我们的方法）	**26.27**	**0.889**	**0.093**	**22.31**	**0.522**	**0.403**	87.22	39.33

表 3：比较 在公开数据集上与 Mamba 变体进行比较。浮点运算次数以一张 128×128 RGB 图像为单位计算 e.

4.4 消融研究

4.4.1 DSMamba可视化分析

图7（b）可视化了由DA-RScan生成的扫描区域图 M_{reg}。划分区域与结构布局高度契合。图7（c）-（d）显示，原始状态空间矩阵 B 和 C 的激活呈均匀分布，缺乏判别性关注。相比之下， B_{depth} 和 C_{depth} 增强了显著结构区域中的激活，同时抑制了模糊区域中的响应，从而提升了状态演化的结构感知能力。值得注意的是， B_{depth} 比 C_{depth}更暗，因为它更严格地调节输入对状态的影响，因此总体激活值更低。

4.4.2 与 Mamba 变体的比较

为进行公平比较并使这些方法适用于反射分离，我们将 DSMamba 替换为 MambaIR[64], 、VMambaIR [65], 和 MambaIRv2 [66] ， 同时保持训练策略不变。如表 3 所示，MambaIR 和 VMambaIR 受固定扫描顺序限制，难以分离重叠层。MambaIRv2 的注意力状态空间设计容易受到与场景内容语义相似的反射干扰。相比之下，我们的 DSMamba 在 T和 R 恢复任务上均优于这些变体。

4.4.3 DSMamba 消融研究

如表4所示，当对 T使用 DA-RScan、对 R使用 DA-GScan 时，可获得最佳性能，优于 VMamba 中的原始四向扫描策略。结果还证明了 DS-SSM 优于原始 SSM，并验证了 SPE 的有效性。

4.4.4 MECM 的消融研究

如表5所示，GPStream 和 SCStream 均有助于提升性能。此外，增加记忆专家的总数 N_{Exp} 可以提供更多样化的特征先验，而选择适当数量的 top-k专家 N_{Exp}^{K}则能够实现有效的专家路由并降低计算成本。设置 $N_{Exp} = 4$、 $N_{Exp}^{K} = 2$可实现令人满意的平衡。

5 结论

我们提出 DMDNet，以应对透射层和反射层对比度相近时的分离难题，尤其是在夜间场景中。我们提出 DSMamba，利用 DAScan 优先处理结构显著

扫描策略		状态空间模型	SPE	平均值			参数 (M)↓	浮点运算次数 (G)↓
T	R			峰值信噪比 ↑	结构相似性 ↑	学习感知图像块相似度 ↓		
DA-RScan	DA-GScan	DS-SSM	✓	**26.27**	**0.889**	**0.093**	87.22	39.33
DA-RScan	DA-RScan	DS-SSM	✓	25.99	0.886	0.098	87.22	39.33
DA-GScan	DA-GScan	DS-SSM	✓	25.87	0.886	0.100	87.22	39.33
DA-GScan	DA-RScan	DS-SSM	✓	<u>26.09</u>	<u>0.887</u>	<u>0.096</u>	87.22	39.33
DA-RScan	DA-GScan	DS-SSM	×	25.66	0.882	0.105	87.22	39.33
DA-RScan	DA-GScan	原始	✓	25.78	0.884	0.098	83.29	38.55
原始	原始	DS-SSM	✓	25.69	0.884	<u>0.096</u>	89.36	39.21

表4：DSMamba消融研究。在公开数据集的透射层上报告结果。

GP-流	SC-流	N_{Exp}^{K}	N_{Exp}	平均值			参数 (M)↓	浮点运算次数 (G)↓
				峰值信噪比 ↑	结构相似性 ↑	学习感知图像块相似度 ↓		
✓	✓	2	4	**26.27**	**0.889**	**0.093**	87.22	39.33
×	×	0	0	24.93	0.882	0.100	55.92	35.85
×	✓	2	4	25.91	0.884	0.100	80.98	37.66
✓	×	2	4	<u>26.07</u>	<u>0.887</u>	<u>0.096</u>	80.98	37.67
✓	✓	1	3	25.93	0.884	0.098	79.39	37.60

表5：MECM 消融研究。结果报告于公开数据集的透射层上。

区域，并利用 DS-SSM 增强这些区域对状态演化的影响，同时抑制干扰的扩散。我们引入 MECM，使专家能够自适应地利用跨图像知识来弥补层恢复的不足。此外，我们构建了 NightIRS 数据集，用于评估夜间反射分离。实验结果表明，DMDNet 在全天场景中均优于最先进方法。其局限之一是依赖有监督训练数据，未来工作将探索无监督方法。

参考文献

[1] Alexander Kirillov、Eric Mintun、Nikhila Ravi、Hanzi Mao、Chloe Rolland、Laura Gustafson、Tete Xiao、Spencer Whitehead、Alexander C Berg、Wan-Yen Lo 等。分割一切。载于*IEEE/CVF*国际计算机视觉会议论文集，第4015–4026页，2023年。[2] Yuntao Wang、Jinpu Zhang、Ruonan Wei、Wenbo Gao 和 Yuehuan Wang。Mfrgn：用于地面到空中地理定位的多尺度特征表示泛化网络。载于第*32*届*ACM*国际多媒体会议论文集，第2574–2583页，2024年。[3] Long Peng、Yang Cao、Yuejin Sun 和 Yang Wang。面向压缩图像分类的轻量级自适应特征去漂移。*IEEE* 多媒体汇刊，26：6424–6436，2024年。[4] Junhao Xiao、Yi Chen、Xiao Feng、Ruoyu Wang 和 Zhiyu Wu。Recnet：基于视图校正的零售场景密集目标检测优化。载于*ICASSP 2025 - 2025*年*IEEE*国际声学、语音与信号处理会议（*ICASSP*），第1–5页，2025年。[5] Jiancheng Pan、Yanxing Liu、Xiao He、Long Peng、Jiahao Li、Yuze Sun 和 Xiaomeng Huang。先增强后搜索：利用基础模型进行跨域少样本目标检测的增强搜索策略。载于计算机视觉与模式识别会议论文集，第1548–1556页，2025年。[6] Anat Levin 和 Yair Weiss。利用稀疏先验，在单幅图像中通过用户辅助分离反射。*IEEE* 模式分析与机器智能汇刊，29(9)：1647–1654，2007年。

[7] Qingnan Fan、Jiaolong Yang、Gang Hua、Baoquan Chen 和 David Wipf。用于单幅图像反射去除和图像平滑的通用深度架构。发表于*IEEE*国际计算机视觉会议论文集，第 3238–3247 页，2017年。

[8] Yang Yang、Wenye Ma、Yin Zheng、Jian-Feng Cai 和 Weiyu Xu。通过凸优化快速抑制单幅图像反射。发表于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第 8141–8149 页，2019年。

[9] Long Peng、Aiwen Jiang、Qiaosi Yi 和 Mingwen Wang。用于单幅图像去雨的累积雨密度感知网络。*IEEE*信号处理快报，27:406–410，2020年。

[10] Long Peng、Aiwen Jiang、Haoran Wei、Bo Liu 和 Mingwen Wang。通过渐进式结构增强约束实现的集成式单幅图像去雨网络。信号处理：图像通信，99:116460，2021年。

[11] Long Peng、Yang Wang、Xin Di、Xueyang Fu、Yang Cao、Zheng-Jun Zha 等。通过中心-周边协同卷积提升图像去雨效果。载于*AAAI* 人工智能会议论文集，第39卷，第6470–6478页，2025年。[12] Yuhong He、Long Peng、Lu Wang 和 Jun Cheng。用于单幅图像去雨的潜在退化表征约束。载于*ICASSP 2024-2024 IEEE* 声学、语音与信号处理国际会议（*ICASSP*），第3155–3159页。IEEE，2024年。[13] Yuhong He、Aiwen Jiang、Lingfang Jiang、Long Peng、Zhifeng Wang 和 Lu Wang。通过空频交互实现的双路径耦合图像去雨网络。载于*2024*年 *IEEE*图像处理国际会议（*ICIP*），第1452–1458页。IEEE，2024年。[14] Haodian Wang、Long Peng、Yuejin Sun、Zengyu Wan、Yang Wang 和 Yang Cao。面向递归式低光照图像增强的亮度感知。*IEEE*人工智能汇刊，5(6):3034–3045，2023年。[15] Yang Wang、Long Peng、Liang Li、Yang Cao 和 Zheng-Jun Zha。面向图像曝光校正的解耦与聚合。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第18115–18124页，2023年。[16] Alexander Yakovenko、George Chakvetadze、Ilya Khrapov、Maksim Zhelezov、Dmitry Vatolin、Radu Timofte、Youngjin Oh、Junhyeong Kwon、Junyoung Park、Nam Ik Cho 等。AIM 2025低光照 RAW 视频去噪挑战赛：数据集、方法与结果。*arXiv*预印本 *arXiv:2508.16830*，2025年。[17] Xin Jin、Chunle Guo、Xiaoming Li、Zongsheng Yue、Chongyi Li、Shangchen Zhou、Ruicheng Feng、Yuekun Dai、Peiqing Yang、Chen Change Loy 等。MIPI 2024少样本 RAW 图像去噪挑战赛：方法与结果。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第1153–1161页，2024年。[18] Yawei Li、Yulun Zhang、Radu Timofte、Luc Van Gool、Lei Yu、Youwei Li、Xinpeng Li、Ting Jiang、Qi Wu、Mingyan Han 等。NTIRE 2023高效超分辨率挑战赛：方法与结果。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第1922–1960页，2023年。[19] Bin Ren、Yawei Li、Nancy Mehta、Radu Timofte、Hongyuan Yu、Cheng Wan、Yuxin Hong、Bingnan Han、Zhuoyuan Wu、Yajun Zou 等。第九届 NTIRE 2024高效超分辨率挑战赛报告。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第6595–6631页，2024年。[20] Yingqian Wang、Zhengyu Liang、Fengyuan Zhang、Lvli Tian、Longguang Wang、Juncheng Li、Jungang Yang、Radu Timofte、Yulan Guo、Kai Jin 等。NTIRE 2025光场图像超分辨率挑战赛：方法与结果。载于计算机视觉与模式识别会议论文集，第1227–1246页，2025年。[21] Long Peng、Yang Cao、Renjing Pei、Wenbo Li、Jiaming Guo、Xueyang Fu、Yang Wang 和 Zheng-Jun Zha。通过自适应方向梯度卷积实现高效真实场景图像超分辨率。*arXiv*预印本 *arXiv:2405.07023*，2024年。[22] Long Peng、Wenbo Li、Renjing Pei、Jingjing Ren、Jiaqi Xu、Yang Wang、Yang Cao 和 Zheng-Jun Zha。迈向真实场景超分辨率的真实感数据生成。*arXiv*预印本 *arXiv:2406.07255*，2024年。[23] Long Peng、Wenbo Li、Jiaming Guo、Xin Di、Haoze Sun、Yong Li、Renjing Pei、Yang Wang、Yang Cao 和 Zheng-Jun Zha。揭示隐藏细节：面向真实场景超分辨率的 RAW 数据增强范式。*arXiv*预印本 *arXiv:2411.10798*，2024年。[24] Long Peng、Anran Wu、Wenbo Li、Peizhe Xia、Xueyuan Dai、Xinjie Zhang、Xin Di、Haoze Sun、Renjing Pei、Yang Wang 等。像素到高斯：采用二维高斯建模的超快速连续超分辨率。*arXiv*预印本 *arXiv:2503.06617*，2025年。[25] Chen Wu、Ling Wang、Long Peng、Dianjie Lu 和 Zhuoran Zheng。摒弃高倍率下采样：一种超高清图像复原的新型设计范式。载于*2025*年 *IEEE/CVF*冬季计算机视觉应用会议（*WACV*），第2390–2399页。IEEE，2025年。[26] Zhibo Du、Long Peng、Yang Wang、Yang Cao 和 Zheng-Jun Zha。FC3DNet：用于高效去摩尔纹的全连接编码器-解码器。载于*2024*年 *IEEE*图像处理国际会议（*ICIP*），第1642–1648页。IEEE，2024年。[27] Andrey Ignatov、Georgy Perevozchikov、Radu Timofte、Wu Pan、Song Wang、Dong Zhang、Zhao Ran、Xiaochen Li、Shichang Ju、Diankai Zhang 等。移动 GPU 上的 RGB 照片增强，Mobile AI 2025挑战赛：报告。载于计算机视觉与模式识别会议论文集，第1922–1933页，2025年。[28] Marcos V Conde、Zhijun Lei、Wen Li、Ioannis Katsavounidis、Radu Timofte、Min Yan、Xin Liu、Qian Wang、Xiaoqian Ye、Zhan Du 等。压缩 AVIF 图像的实时 4K 超分辨率。AIS 2024挑战赛综述。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第5838–5856页，2024年。

[29] Jingjing Ren、Wenbo Li、Haoyu Chen、Renjing Pei、Bin Shao、Yong Guo、Long Peng、Fenglong Song 和 Lei Zhu。Ultrapixel：推动超高分辨率图像合成迈向新高度。神经信息处理系统进展，37：111131–111171，2024年。[30] Anran Wu、Long Peng、Xin Di、Xueyuan Dai、Chen Wu、Yang Wang、Xueyang Fu、Yang Cao 和 Zheng-Jun Zha。RobustGS：在低质量条件下统一提升前馈式三维高斯泼溅。*arXiv* 预印本 *arXiv:2508.03077*，2025年。[31] Qiuhai Yan、Aiwen Jiang、Kang Chen、Long Peng、Qiaosi Yi 和 Chunjie Zhang。文本提示引导的图像复原。人工智能工程应用，155：110981，2025年。[32] Haoze Sun、Wenbo Li、Jiayue Liu、Kaiwen Zhou、Yongqiang Chen、Yong Guo、Yanwei Li、Renjing Pei、Long Peng 和 Yujiu Yang。超越像素：文本增强真实世界图像复原的泛化能力。*arXiv* 预印本 *arXiv:2412.00878*，2024年。[33] Haoze Sun、Wenbo Li、Jiayue Liu、Kaiwen Zhou、Yongqiang Chen、Yong Guo、Yanwei Li、Renjing Pei、Long Peng 和 Yujiu Yang。文本增强泛化能力：用于真实世界图像复原的即插即用图像描述生成器。[34] ZhanFeng Feng、Long Peng、Xin Di、Yong Guo、Wenbo Li、Yulun Zhang、Renjing Pei、Yang Wang、Yang Cao 和 Zheng-Jun Zha。PMQ-VE：用于视频增强的渐进式多帧量化。*arXiv* 预印本 *arXiv:2505.12266*，2025年。[35] Shulian Zhang、Yong Guo、Long Peng、Ziyang Wang、Ye Chen、Wenbo Li、Xiao Zhang、Yulun Zhang 和 Jian Chen。VividFace：用于视频人脸增强的高质量高效一步扩散。*arXiv* 预印本 *arXiv:2509.23584*，2025年。[36] Hao Xu、Long Peng、Shezheng Song、Xiaodong Liu、Ma Jun、Shasha Li、Jie Yu 和 Xiaoguang Mao。Camel：资源受限设备上的能耗感知大语言模型推理。*arXiv* 预印本 *arXiv:2508.09173*，2025年。[37] Xiaohua Qi、Renda Li、Long Peng、Qiang Ling、Jun Yu、Ziyi Chen、Peng Chang、Mei Han 和 Jing Xiao。利用扩散模型进行无数据知识蒸馏。*arXiv* 预印本 *arXiv:2504.00870*，2025年。[38] Aiwen Jiang、Zhi Wei、Long Peng、Feiqiang Liu、Wenbo Li 和 Mingwen Wang。DALPSR：利用与退化对齐的语言提示进行真实场景超分辨率。*arXiv* 预印本 *arXiv:2406.16477*，2024年。[39] Long Peng、Wenbo Li、Jiaming Guo、Xin Di、Haoze Sun、Yong Li、Renjing Pei、Yang Wang、Yang Cao 和 Zheng-Jun Zha。利用原始数据提升真实场景超分辨率：一种新视角、数据集和基线。[40] Xuaner Zhang、Ren Ng 和 Qifeng Chen。利用感知损失进行单幅图像反射分离。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第4786–4794页，2018年。[41] Qiming Hu、Hainuo Wang 和 Xiaojie Guo。通过双流交互式 Transformer 实现单幅图像反射分离。神经信息处理系统进展，37：55228–55248，2024年。[42] Tianfan Xue、Michael Rubinstein、Ce Liu 和 William T Freeman。一种无障碍摄影的计算方法。*ACM* 图形学汇刊（*TOG*），34(4)：1–11，2015年。[43] Simon Niklaus、Xuaner Cecilia Zhang、Jonathan T Barron、Neal Wadhwa、Rahul Garg、Feng Liu 和 Tianfan Xue。学习式双视角反射去除。载于*IEEE/CVF*冬季计算机视觉应用会议论文集，第3713–3722页，2021年。[44] Rui Li、Simeng Qiu、Guangming Zang 和 Wolfgang Heidrich。通过多次反射偏振状态追踪实现反射分离。载于欧洲计算机视觉会议，第781–796页。Springer，2020年。[45] Chenyang Lei、Xuhua Huang、Mengdi Zhang、Qiong Yan、Wenxiu Sun 和 Qifeng Chen。在自然场景中利用完美对齐实现偏振反射去除。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第1750–1758页，2020年。[46] Jun Sun、Yakun Chang、Cheolkon Jung 和 Jiawei Feng。利用卷积神经网络进行多模态反射去除。*IEEE*信号处理快报，26(7)：1011–1015，2019年。[47] Yuchen Hong、Youwei Lyu、Si Li、Gang Cao 和 Boxin Shi。融合近红外与RGB图像特征的反射去除。*IEEE* 多媒体汇刊，25：7101–7112，2022年。[48] Chenyang Lei 和 Qifeng Chen。利用无反射纯闪光线索实现鲁棒反射去除。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第14811–14820页，2021年。[49] Tianfu Wang、Mingyang Xie、Haoming Cai、Sachin Shah 和 Christopher A Metzler。Flash-Split：利用闪光线索和潜在扩散分离进行二维反射去除。载于计算机视觉与模式识别会议论文集，第5688–5698页，2025年。

[50] Haofeng Zhong、Yuchen Hong、Shuchen Weng、Jinxiu Liang 和 Boxin Shi。语言引导的图像反射分离。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第24913–24922页，2024年。[51] Yuchen Hong、Haofeng Zhong、Shuchen Weng、Jinxiu Liang 和 Boxin Shi。L-differ：利用基于语言的扩散模型进行单幅图像反射去除。载于欧洲计算机视觉会议，第58–76页。Springer，2024年。[52] Huaidong Zhang、Xuemiao Xu、Hai He、Shengfeng He、Guoqiang Han、Jing Qin 和 Dapeng Wu。通过边缘感知级联网络实现快速的用户引导式单幅图像反射去除。*IEEE* 多媒体汇刊，22(8)：2012–2023，2019年。[53] Xiao Chen、Xudong Jiang、Yunkang Tao、Zhen Lei、Qing Li、Chenyang Lei 和 Zhaoxiang Zhang。FIRM：灵活交互式反射去除。载于 *AAAI* 人工智能会议论文集，第 39 卷，第 2230–2238页，2025年。[54] René Ranftl、Katrin Lasinger、David Hafner、Konrad Schindler 和 Vladlen Koltun。迈向鲁棒的单目深度估计：混合数据集以实现零样本跨数据集迁移。*IEEE* 模式分析与机器智能汇刊，44(3)：1623–1637，2022年。[55] Albert Gu 和 Tri Dao。Mamba：利用选择性状态空间进行线性时间序列建模。*arXiv* 预印本 *arXiv:2312.00752*，2023年。[56] Lianghui Zhu、Bencheng Liao、Qian Zhang、Xinlong Wang、Wenyu Liu 和 Xinggang Wang。视觉 Mamba：利用双向状态空间模型进行高效视觉表征学习。*arXiv* 预印本 *arXiv:2401.09417*，2024年。[57] Long Peng、Xin Di、Zhanfeng Feng、Wenbo Li、Renjing Pei、Yang Wang、Xueyang Fu、Yang Cao 和 Zheng-Jun Zha。引导 Mamba 处理复杂纹理：一种用于图像复原的高效纹理感知状态空间模型。*arXiv* 预印本 *arXiv:2501.16583*，2025年。[58] Yue Liu、Yunjie Tian、Yuzhong Zhao、Hongtian Yu、Lingxi Xie、Yaowei Wang、Qixiang Ye、Jianbin Jiao 和 Yunfan Liu。VMamba：视觉状态空间模型。神经信息处理系统进展，37：103031–103063，2024年。[59] Abdelrahman Elnenaey 和 Marwan Torki。利用多步损失进行单幅图像反射去除。载于*2024* 年 *IEEE/ACS* 第 *21* 届计算机系统与应用国际会议（*AICCSA*），第1–6页。IEEE，2024年。[60] Lingzhi He、Yakun Chang、Runmin Cong、Hongyu Liu、Shujuan Huang、Renshuai Tao 和 Yao Zhao。重新思考深度引导的反射去除。*IEEE* 多媒体汇刊，2025年。[61] Peizhe Xia、Long Peng、Xin Di、Renjing Pei、Yang Wang、Yang Cao 和 Zheng-Jun Zha。S3Mamba：通过可扩展状态空间模型实现任意尺度超分辨率。*arXiv* 预印本 *arXiv:2411.11906*，6，2024年。[62] Xin Di、Long Peng、Peizhe Xia、Wenbo Li、Renjing Pei、Yang Cao、Yang Wang 和 Zheng-Jun Zha。QmambaBSR：利用查询状态空间模型实现突发图像超分辨率。载于计算机视觉与模式识别会议论文集，第23080–23090页，2025年。[63] Yuhong He、Long Peng、Qiaosi Yi、Chen Wu 和 Lu Wang。利用状态空间模型进行图像复原的多尺度表征学习。*arXiv* 预印本 *arXiv:2408.10145*，2024年。[64] Hang Guo、Jinmin Li、Tao Dai、Zhihao Ouyang、Xudong Ren 和 Shu-Tao Xia。MambaIR：基于状态空间模型进行图像复原的简易基线。载于欧洲计算机视觉会议，第222–241页。Springer，2024年。[65] Yuan Shi、Bin Xia、Xiaoyu Jin、Xing Wang、Tianyu Zhao、Xin Xia、Xuefeng Xiao 和 Wenming Yang。VMambaIR：用于图像复原的视觉状态空间模型。*IEEE* 视频技术电路与系统汇刊，2025年。[66] Hang Guo、Yong Guo、Yaohua Zha、Yulun Zhang、Wenbo Li、Tao Dai、Shu-Tao Xia 和 Yawei Li。MambaIRv2：注意力状态空间复原。载于计算机视觉与模式识别会议论文集，第28124–28133页，2025年。[67] Eduard Zamfir、Zongwei Wu、Nancy Mehta、Yuedong Tan、Danda Pani Paudel、Yulun Zhang 和 Radu Timofte。复杂度专家是适用于任意图像复原任务的任务判别学习器。载于计算机视觉与模式识别会议论文集，第12753–12763页，2025年。[68] Xuanhua He、Keyu Yan、Rui Li、Chengjun Xie、Jie Zhang 和 Man Zhou。利用混合专家进行频率自适应全色锐化。载于*AAAI* 人工智能会议论文集，第38卷，第2121–2129页，2024年。

[69] Rui Xu、Minghao Guo、Jiaqi Wang、Xiaoxiao Li、Bolei Zhou 和 Chen Change Loy。《纹理记忆增强的深度分块图像修复》。IEEE图像处理汇刊，30：9112–9124，2021年。[70] Wentao Zou、Xiao Lu、Zhilv Yi、Ling Zhang、Gang Fu、Ping Li 和 Chunxia Xiao。《通过联合学习反射消除与内容修复去除眼镜反射》。IEEE 视频技术电路与系统汇刊，34(10)：10266–10280，2024年。[71] Qiming Hu 和 Xiaojie Guo。《通过成分协同实现单幅图像反射分离》。载于《IEEE/CVF国际计算机视觉会议论文集》，第13138–13147页，2023年。[72] Diederik P Kingma 和 Jimmy Ba。《Adam：一种随机优化方法》。arXiv预印本 arXiv:1412.6980，2014年。[73] Hao Zhao、Mingjia Li、Qiming Hu 和 Xiaojie Guo。《用于单幅图像反射去除的可逆解耦网络》。载于《计算机视觉与模式识别会议论文集》，第26430–26439页，2025年。[74] Zheng Dong、Ke Xu、Yin Yang、Hujun Bao、Weiwei Xu 和 Rynson WH Lau。《位置感知的单幅图像反射去除》。载于《IEEE/CVF国际计算机视觉会议论文集》，第5017–5026页，2021年。[75] Mark Everingham、Luc Van Gool、Christopher KI Williams、John Winn 和 Andrew Zisserman。《PASCAL视觉物体类别（VOC）挑战赛》。国际计算机视觉杂志，88(2)：303–338，2010年。[76] Chao Li、Yixiao Yang、Kun He、Stephen Lin 和 John E Hopcroft。《通过级联细化实现单幅图像反射去除》。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第3565–3574页，2020年。[77] Renjie Wan、Boxin Shi、Ling-Yu Duan、Ah-Hwee Tan 和 Alex C Kot。《单幅图像反射去除算法基准测试》。载于《IEEE国际计算机视觉会议论文集》，第3922–3930页，2017年。[78] Quan Huynh-Thu 和 Mohammed Ghanbari。《PSNR在图像/视频质量评估中的有效范围》。电子快报，44(13)：800–801，2008年。[79] Zhou Wang、Alan C Bovik、Hamid R Sheikh 和 Eero P Simoncelli。《图像质量评估：从误差可见性到结构相似性》。IEEE图像处理汇刊，13(4)：600–612，2004年。[80] Richard Zhang、Phillip Isola、Alexei A Efros、Eli Shechtman 和 Oliver Wang。《深度特征作为感知度量的惊人效能》。载于《IEEE计算机视觉与模式识别会议论文集》，第586–595页，2018年。[81] Jie Yang、Dong Gong、Lingqiao Liu 和 Qinfeng Shi。《深度且双向地观察：一种用于单幅图像反射去除的深度学习方法》。载于《欧洲计算机视觉会议论文集（ECCV）》，第654–669页，2018年。[82] Kaixuan Wei、Jiaolong Yang、Ying Fu、David Wipf 和 Hua Huang。《利用错位训练数据和网络增强实现单幅图像反射去除》。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第8178–8187页，2019年。[83] Qiming Hu 和 Xiaojie Guo。《垃圾还是宝藏？一种用于单幅图像反射分离的交互式双流策略》。神经信息处理系统进展，34：24683–24694，2021年。[84] Xin Feng、Wenjie Pei、Zihui Jia、Fanglin Chen、David Zhang 和 Guangming Lu。《深度掩膜生成网络：用于从叠加图像中恢复背景的统一框架》。IEEE图像处理汇刊，30：4867–4882，2021年。[85] Yurui Zhu、Xueyang Fu、Zheyu Zhang、Aiping Liu、Zhiwei Xiong 和 Zheng-Jun Zha。《用于单幅图像反射去除的色调引导网络》。IEEE神经网络与学习系统汇刊，2023年。[86] Yurui Zhu、Xueyang Fu、Peng-Tao Jiang、Hao Zhang、Qibin Sun、Jinwei Chen、Zheng-Jun Zha 和 Bo Li。《重新审视自然场景中的单幅图像反射去除》。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第25468–25478页，2024年。[87] Karen Simonyan 和 Andrew Zisserman。《用于大规模图像识别的超深卷积网络》。arXiv预印本 arXiv:1409.1556，2014年。[88] Nikolaos Arvanitopoulos、Radhakrishna Achanta 和 Sabine Susstrunk。《单幅图像反射抑制》。载于《IEEE计算机视觉与模式识别会议论文集》，第4498–4506页，2017年。[89] Frank Wilcoxon。《基于排序方法的个体比较》。载于《统计学突破：方法与分布》，第196–202页。Springer，1992年。

[90] Kangning Yang、Ling Ouyang、Huiming Sun、Jie Cai、Lan Fu、Jiaming Ding、Chiu Man Ho 和 Zibo Meng。《OpenRR-1k：用于真实场景反射去除的可扩展数据集》。*arXiv*预印本 *arXiv:2506.08299*，2025年。[91] Guang-Yong Chen、Chao-Wei Zheng、Guo-Dong Fan、Jian-Nan Su、Min Gan 和 CL Philip Chen。《真实场景图像反射去除：超高清数据集与高效基线》。*IEEE* 视频技术电路与系统汇刊，2024年。

补充材料：深度协同 Mamba 携手记忆专家实现全天图像反射分离

本补充材料进一步介绍所提出的 DMDNet，包括损失函数的设计、实现设置、与最先进方法的扩展比较、补充消融研究、统计显著性分析，以及介绍 NightIRS 数据集的数据附录。

A 损失函数

损失函数由负载损失 $\mathcal{L}_{load}$、记忆匹配损失 $\mathcal{L}_{mem}$ 和外观损失 $\mathcal{L}_{app}$ 组成，分别用于平衡专家的使用、增强记忆匹配效果并确保输出保真度。
负载损失采用变异系数的平方来平衡专家的负载，并防止过度依赖某些专家。其定义如下：

$$\mathcal{L}_{load} = \sum_{X \in \{T,R\}} \lambda_X^{load} \cdot \mathbb{E}\left[\left(\frac{\sigma(W_X)}{\mu(W_X)+\epsilon}\right)^2\right] \tag{6}$$

其中， W_X 表示每个样本在第 X层的所有专家上的选择权重； $\sigma(\cdot)$和 $\mu(\cdot)$分别表示标准差和均值； ϵ 是一个用于避免除以零的小常数； $\mathbb{E}[\cdot]$ 表示期望算子； λ_X^{load}表示 $\boldsymbol{T}$和 $\boldsymbol{R}$的负载损失权重。

记忆匹配损失促使图像特征接近与其最相关的记忆项，同时与相关性较低的记忆项保持明显间隔。为此，$\mathcal{L}_{mem}$ 由三元组损失和均方误差（MSE）损失组成，定义如下：

$$\mathcal{L}_{mem} = \sum_{X \in \{T,R\}} \left[\lambda_X^{triplet} \cdot \max\left(\|I_i - m_i^+\|_2^2 - \|I_i - m_i^-\|_2^2,\ 0\right) + \lambda_X^{align} \left\|I_i - m_i^+\right\|_2^2\right] \tag{7}$$

其中， I_i 表示图像的查询特征。 m_i^+ 和 m_i^- 分别表示与 I_i最相似和第二相似的记忆项。 $\lambda_X^{triplet}$ 和 λ_X^{align} 是权重系数，分别用于平衡三元组项和对齐项的贡献。

外观损失约束复原图像与目标图像在像素空间和感知空间中的相似度。它由两个部分组成：逐像素 L1 损失，以及基于 VGG [87]特征的感知损失：

$$\mathcal{L}_{app} = \lambda_T^{L1} \left\|\hat{T} - T\right\|_1 + \lambda_R^{L1} \left\|\hat{R} - \mathcal{R}\right\|_1 + \lambda_T^{VGG} \left\|VGG(\hat{T}) - VGG(T)\right\|_1 \tag{8}$$

ˆ ˆ其中， T和 R 分别表示恢复后的透射层和反射层，而 T和 R 表示相应的真实值。 λ_T^{L1}、 λ_R^{L1}和 λ_T^{VGG}是权重系数，用于平衡 L1 项和感知项的贡献。

总体损失函数定义如下：

$$\mathcal{L}_{total} = \mathcal{L}_{load} + \mathcal{L}_{mem} + \mathcal{L}_{app} \tag{9}$$

B 更多比较

我们进一步将 DMDNet 与 11 种最先进方法（SOTAs）进行比较，包括 BDN [81], ERRNet [82],IBCLN [76], LANet [74],YTMT [83],DMGN [84],HGNet [85],DSRNet [71],RDRNet [86],DSIT [41],和 RDNet[73]。我们在公开数据集上评估反射层恢复，包括 Nature 数据集 [76], 、Real20 数据集 [40], 以及 SIR^2 数据集 [77], 中的 Wild、Postcard 和 Solid 子集。如表 6 所示，DMDNet 获得的 Top-1 和 Top-2 结果数量最多，在反射恢复方面取得了最佳平均性能。

我们还将 DMDNet 与传统方法 L0-RS [88] 和 Fast-RS [8]进行了比较。如表7所示，我们的 DMDNet 优于这两种方法。此外，图9–13展示了透射层（$\boldsymbol{T}$）和反射层（$\boldsymbol{R}$）的更多定性对比，进一步验证了 DMDNet 的有效性。

方法	自然（20）			真实（20）			野外（55）			明信片（199）			纯色（200）			平均值		
	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓
BDN（ECCV'18）	6.72	0.108	0.850	8.33	0.144	0.739	9.02	0.262	0.768	9.16	0.461	0.687	8.26	0.277	0.745	8.30	0.251	0.758
IBCLN（CVPR'20）	16.96	0.298	0.757	18.60	0.325	0.666	19.54	0.482	0.690	19.54	0.616	0.636	21.83	0.562	0.681	19.29	0.456	0.686
LANet (ICCV'21)	18.79	0.335	0.604	19.57	0.383	0.481	22.31	**0.677**	0.425	**19.61**	**0.699**	**0.513**	**23.98**	**0.754**	0.468	20.85	0.570	0.498
YTMT (NIPS'21)	21.23	0.339	0.647	22.51	0.453	0.531	20.15	0.187	0.429	11.92	0.153	0.811	18.77	0.061	0.553	18.92	0.238	0.594
DMGN (TIP'21)	21.48	0.365	0.630	20.01	0.314	0.623	21.22	0.513	0.458	17.15	0.574	0.607	20.57	0.399	0.522	20.08	0.433	0.568
DSRNet(ICCV'23)	19.80	0.350	0.706	23.43	0.491	0.505	21.71	0.643	0.455	18.47	0.671	0.627	23.16	0.739	0.486	21.31	**0.579**	0.556
DSIT (NIPS'24)	27.53	0.641	0.427	24.41	0.554	0.448	**22.98**	0.556	**0.343**	13.44	0.390	0.634	21.72	0.542	0.422	22.02	0.537	0.455
RDNet (CVPR'25)	28.37	0.657	0.326	**25.67**	0.601	**0.309**	21.44	0.326	0.379	14.56	0.418	0.661	20.22	0.268	0.454	22.05	0.454	0.426
DMDNet（我们的方法）	**28.95**		**0.316**	25.53		0.320	22.19			13.51		0.609	21.38			**22.31**		

表 6：在公开数据集上对反射层进行定量比较。DMDNet 在恢复反射层方面取得了最佳平均性能。**粗体**和下划线分别表示 Top-1 和 Top-2 结果。 ↑表示数值越高越好，而 ↓表示数值越低越好。

方法	公开数据集			NightIRS		
	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓
L0-RS（CVPR 2017）	21.16	0.794	0.206	23.25	0.789	0.231
Fast-RS（CVPR'19）	20.55	0.792	0.205	22.70	0.777	0.244
DMDNet（我们的方法）	**26.27**	**0.889**	**0.093**	**25.24**	**0.832**	**0.144**

表8：与传统方法在透射层上的比较。

为评估人类感知偏好，我们招募了30名具有正常视觉功能的志愿者，对各方法的输出进行主观排名。具体而言，每位参与者评估7组图像，其中包括来自图6的2组和来自图9至13的5组，并从每组中选出排名前三的方法。我们的 DMDNet 获得119票，超过 RDNet 的91票和 DSIT 的87票，这表明 DMDNet 更符合人类视觉感知。

C 补充消融实验

C.1 深度信息消融研究

为评估深度估计精度的影响，我们将深度模型 MiDaS v3.1 Next-ViT-L [54]替换为两个精度较低的变体：v3.0 DPT-H 和 v2.1 DPT-Small。如表8所示，高精度深度估计能够提升性能，而较低精度的深度估计则会导致性能下降。即使使用精度较低的深度估计，模型仍优于主流基线 RDRNet，这表明我们的 DSMamba 能够从粗略几何结构中有效提取有用线索。进一步从网络中移除深度先验会导致性能明显下降，凸显了深度信息在所提出方法中的重要作用。

为探究深度信息对其他方法的影响，我们将深度作为额外先验引入 RDNet [73] 和 DSIT [41],，将其与输入拼接，并采用卷积融合。我们加载原始模型作为预训练权重，并按照其官方设置进行训练。如表8所示，与原始版本相比，这些方法的性能有所下降，表明引入深度会扰乱其对特征的理解。相比之下，DSMamba 能够有效利用深度信息，充分发挥深度先验的优势。

C.2 状态空间建模策略的消融研究。

我们通过改变 $\boldsymbol{B}$ 和 $\boldsymbol{C}$的公式，研究了不同的状态空间建模策略，结果总结于表9。使用不包含深度信息的原始矩阵 $\boldsymbol{B}$ 和 $\boldsymbol{C}$ 时，模型的性能有限。引入由深度推导出的矩阵 $\boldsymbol{B}_{depth}$ 和 $\boldsymbol{C}_{depth}$ 可增强表征能力，但将它们直接加到原始矩阵（$\boldsymbol{B}=\boldsymbol{B}+\boldsymbol{B}_{depth}, \boldsymbol{C}=\boldsymbol{C}+\boldsymbol{C}_{depth}$）中无法获得最优结果。相比之下，采用深度协同状态空间模型（DS-SSM），即 $\boldsymbol{B}_{aware}=(1-\gamma)\boldsymbol{B}+\gamma\boldsymbol{B}_{depth}$ 和$\boldsymbol{C}_{aware}=(1-\gamma)\boldsymbol{C}+\gamma\boldsymbol{C}_{depth}$，可获得最高的整体性能。这些结果表明，DS-SSM 能够在保持计算效率的同时提升层分离质量。

C.3 通道数量的消融研究。

我们针对 DMDNet 主架构中的不同通道配置进行了消融研究，如表10所示。减少通道数量能显著缩小模型规模并降低计算成本，但会导致性能明显下降。增加通道数量通常能够提升性能，但通道规模过大会使参数量和浮点运算次数（FLOPs）急剧增加，却无法带来进一步收益，这主要是由于

方法	深度	公开数据集						NightIRS					
		透射层			反射层			透射层			反射层		
		峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓
DMDNet	v3.1	**26.27**	**0.889**	**0.093**	**22.31**	0.522	**0.403**	**25.24**	**0.832**	**0.144**	**28.37**	0.633	**0.286**
DMDNet	v3.0	25.53	0.880	0.104	22.19	**0.534**	0.414	24.57	0.827	0.151	27.31	0.616	0.319
DMDNet	v2.1	24.98	0.875	0.110	21.87	0.513	0.423	24.40	0.821	0.163	28.32	**0.681**	0.318
DMDNet	-	23.89	0.853	0.131	21.24	0.505	0.466	23.97	0.815	0.174	26.39	0.564	0.339
DSIT	v3.1	22.75	0.849	0.179	21.81	0.547	0.510	23.81	0.811	0.209	26.71	0.518	0.413
RDNet	v3.1	21.74	0.821	0.193	19.69	0.306	0.495	21.72	0.701	0.226	26.27	0.548	0.371

表8：不同方法在深度估计质量和深度集成方面的消融研究。

状态空间模型建模策略	**平均值**			**参数** (M)	浮点运算次数 (G)
	峰值信噪比↑	结构相似性↑	学习感知图像块相似度↓		
$\boldsymbol{B}_{aware} = (1-\gamma)\boldsymbol{B} + \gamma\boldsymbol{B}_{depth}$, $\boldsymbol{C}_{aware} = (1-\gamma)\boldsymbol{C} + \gamma\boldsymbol{C}_{depth}$	**26.27**	**0.889**	**0.093**	87.22	39.33
$\boldsymbol{B}_{aware} = \boldsymbol{B}$, $\boldsymbol{C}_{aware} = \boldsymbol{C}$	25.78	0.884	0.098	83.29	38.55
$\boldsymbol{B}_{aware} = \boldsymbol{B} + \boldsymbol{B}_{depth}$, $\boldsymbol{C}_{aware} = \boldsymbol{C} + \boldsymbol{C}_{depth}$	25.83	0.885	0.097	87.11	39.28
$\boldsymbol{B}_{aware} = (1-\gamma)\,\boldsymbol{B} + \gamma\,\boldsymbol{B}_{depth}$, $\boldsymbol{C}_{aware} = \boldsymbol{C}$	26.24	0.886	0.095	87.01	39.33
$\boldsymbol{B}_{aware} = \boldsymbol{B}$, $\boldsymbol{C}_{aware} = (1-\gamma)\,\boldsymbol{C} + \gamma\,\boldsymbol{C}_{depth}$	26.04	0.887	0.098	87.01	39.33

表9：状态空间建模策略的消融研究。结果在公开数据集的透射层上报告。所提出的 DS-SSM 取得了最佳整体性能。

过参数化引入了冗余。(48，96，192，384，768)这一设置在复原质量与效率之间实现了最佳权衡，证明其适用于 DMDNet 架构。

C.4 损失函数的消融研究。

我们进一步对 DMDNet 的损失函数进行了消融研究，结果汇总于表11。如表格第一行所示，我们的设置按照特定权重比例结合了负载损失、记忆匹配损失和外观损失，取得了最佳整体性能。移除某些损失项或调整其权重都会导致性能下降，这证明了我们所设计的损失函数的有效性。

D 更多实现细节

损失权重设为 $\lambda_T^{load} = \lambda_R^{load} = 0.008$、 $\lambda_T^{triplet} = \lambda_T^{align} = 0.1$、 $\lambda_R^{triplet} = \lambda_R^{align} = 0.05$、 $\lambda_T^{L1} = \lambda_R^{L1} = 1$ 和 $\lambda_T^{VGG} = 0.02$。对于数据合成，我们采用一种广泛使用的物理模型 [71], ，其公式如下：

$$\boldsymbol{I} = \alpha\boldsymbol{T} + \beta\boldsymbol{R} - \boldsymbol{T} \circ \boldsymbol{R}, \tag{10}$$

其中， $\boldsymbol{I}$ 表示混合图像， $\boldsymbol{T}$ 表示透射层， $\boldsymbol{R}$ 表示反射层， α 和 β 分别表示它们对应的混合系数， $\circ$ 表示 Hadamard 积。

模型在配备 Intel Xeon Platinum 8352V @ 2.10GHz 的设备上训练，运行 Ubuntu 22.04.5LTS，使用 Python 3.10.13、PyTorch 2.1.1 和 CUDA 11.8，并采用单张 NVIDIA RTX 4090 GPU。每项实验均进行两次，并报告最佳性能。在 NVIDIA RTX 6000 Ada GPU 上测试时，DMDNet 处理一张 512×512 RGB图像耗时 0.4 s，占用 4.2 GB GPU 内存，表明其计算成本合理。

E 统计检验

我们采用威尔科克森符号秩检验 [89] 来评估性能差异的显著性。如表12所总结，结果表明，DMDNet在透射层和反射层上的表现通常在统计上显著优于现有方法。

C_1	C_2	C_3	C_4	C_5	平均值 峰值信噪比 ↑	结构相似性 ↑	学习感知图像块相似度 ↓	参数 (M)	浮点运算次数 (G)
48	96	192	384	768	**26.27**	**0.889**	**0.093**	87.22	39.33
32	64	128	256	512	25.86	0.885	<u>0.096</u>	39.36	22.98
64	128	256	512	1024	<u>26.06</u>	<u>0.886</u>	0.097	153.85	62.04

表10：DMDNet 通道数量的消融研究。结果报告基于公开数据集的透射层。

负载损失		记忆匹配损失				外观损失			平均值		
λ_T^{load}	λ_R^{load}	$\lambda_T^{triplet}$	$\lambda_R^{triplet}$	λ_T^{align}	λ_R^{align}	λ_T^{L1}	λ_R^{L1}	λ_T^{VGG}	峰值信噪比 ↑	结构相似性 ↑	学习感知图像块相似度 ↓
0.008	0.008	0.10	0.05	0.10	0.05	1	1	0.02	**26.27**	**0.889**	**0.093**
0	0	0.10	0.05	0.10	0.05	1	1	0.02	25.92	<u>0.885</u>	<u>0.096</u>
0.008	0.008	0	0	0	0	1	1	0.02	25.58	0.879	0.104
0	0	0	0	0	0	1	1	0.02	25.29	0.883	0.099
0	0	0	0	0	0	1	1	0	25.18	0.876	0.125
0.008	0.008	0.10	0.05	0.10	0.05	1	1	0	25.91	0.880	0.121
0.008	0.008	0.16	0.16	0.10	0.10	1	1	0.02	25.25	0.877	0.107
0.008	0.008	0.10	0.10	0.10	0.10	1	1	0.02	<u>26.03</u>	<u>0.885</u>	0.098

表11：针对损失函数的消融研究。结果报告的是公开数据集透射层上的结果。

F 数据附录

我们引入夜间图像反射分离（NightIRS）数据集，以解决夜间场景反射分离基准数据缺失的问题。现有数据集（例如 Nature [76], Real [40], SIR^2 [77], RRW [86]）主要包含全局光照充足的日间场景，无法体现夜间条件下的挑战；在夜间，由于全局光照不足且分散的人造光源的存在，$\boldsymbol{T}$和 $\boldsymbol{R}$ 往往呈现出相似的对比度和重叠结构。

NightIRS 数据集由 1,000 个图像三元组组成，每个三元组包含一张混合图像 $\boldsymbol{I}$、一个透射层 $\boldsymbol{T}$ 和一个反射层 $\boldsymbol{R}$。这些图像使用 Sony LYTIA-T808 拍摄，该设备在低光照条件下具有高灵敏度和高动态范围（HDR）能力，能够忠实记录夜间的细微细节。采集数据时使用三脚架保持稳定，并使用无线遥控快门避免拍摄时产生振动。为引入反射干扰，采用了不同厚度（1 毫米、3 毫米、5 毫米和 8 毫米）、尺寸为 700 毫米 $\times$ 500 毫米的亚克力板和玻璃板。该数据集涵盖多种夜间条件，包括人工照明的城市街道、室内外反射场景以及低光照自然环境，其规模超过 OpenRR-1k [90]（83 个样本）和 RR4K [91]（54 个样本）等反射去除数据集，可为推进夜间反射分离研究提供基准。

NightI 示例 RS 展示于图14中。这些示例涵盖了数据集中多样的夜间条件。

DMDNet 对比	透射层						反射层					
	PSNR		SSIM		学习感知图像块相似度		PSNR		SSIM		学习感知图像块相似度	
	统计量	P值	统计量	P值	统计量	P值	统计量	P值	统计量	P值	统计量	P值
BDN (ECCV'18)	1.4×10^4	6.0×10^{-234}	4.4×10^3	6.0×10^{-242}	1.1×10^3	8.0×10^{-245}	6.4×10^1	1.1×10^{-245}	4.3×10^4	2.9×10^{-209}	2.1×10^3	7.0×10^{-244}
ERRNet (CVPR'19)	3.7×10^4	1.0×10^{-214}	1.6×10^4	7.0×10^{-232}	5.5×10^4	2.7×10^{-200}	N/A	N/A	N/A	N/A	N/A	N/A
IBCLN (CVPR'20)	6.8×10^4	8.0×10^{-190}	7.8×10^4	1.0×10^{-182}	2.1×10^4	2.0×10^{-227}	1.8×10^5	2.2×10^{-112}	2.1×10^5	2.8×10^{-99}	8.4×10^3	2.0×10^{-238}
LANet（ICCV'21）	1.5×10^5	2.2×10^{-135}	1.8×10^5	2.2×10^{-116}	1.3×10^5	1.5×10^{-146}	2.5×10^5	2.2×10^{-76}	3.2×10^5	4.8×10^{-46}	1.0×10^5	7.0×10^{-164}
YTMT（NIPS'21）	6.4×10^4	3.3×10^{-193}	5.0×10^4	6.0×10^{-204}	4.8×10^4	5.5×10^{-206}	5.3×10^4	4.9×10^{-202}	1.3×10^5	1.6×10^{-145}	5.7×10^3	8.0×10^{-241}
DMGN (TIP'21)	5.8×10^4	1.2×10^{-197}	4.5×10^4	2.2×10^{-208}	8.0×10^4	7.5×10^{-181}	2.3×10^5	1.2×10^{-87}	3.2×10^5	6.1×10^{-47}	2.7×10^4	8.6×10^{-223}
HGNet（TNNLS'23）	1.2×10^5	3.6×10^{-152}	1.7×10^5	1.5×10^{-121}	7.6×10^4	8.7×10^{-184}	N/A	N/A	N/A	N/A	N/A	N/A
DSRNet（ICCV'23）	1.6×10^5	1.0×10^{-124}	1.9×10^5	1.3×10^{-106}	1.6×10^5	5.1×10^{-128}	3.4×10^5	5.0×10^{-40}	4.1×10^5	9.5×10^{-20}	3.9×10^4	6.0×10^{-213}
RDRNet（CVPR'24）	2.1×10^5	1.7×10^{-96}	2.7×10^5	3.3×10^{-66}	7.3×10^4	4.1×10^{-186}	N/A	N/A	N/A	N/A	N/A	N/A
DSIT (NIPS'24)	4.1×10^5	5.4×10^{-18}	4.3×10^5	5.5×10^{-14}	2.1×10^5	4.7×10^{-99}	3.8×10^5	4.0×10^{-28}	4.0×10^5	6.4×10^{-21}	1.6×10^5	3.5×10^{-128}
RDNet（CVPR'25）	5.3×10^5	1.1×10^{-1}	5.4×10^5	2.4×10^{-1}	5.4×10^5	2.8×10^{-1}	3.5×10^5	2.2×10^{-37}	3.6×10^5	2.2×10^{-33}	2.6×10^5	4.9×10^{-71}

表12：对两个公开数据集和 NightIRS 数据集的结果进行汇总后，DMDNet 与最先进方法的 Wilcoxon 符号秩检验表明，DMDNet 在透射层和反射层上均较大多数方法取得了统计显著的改进。

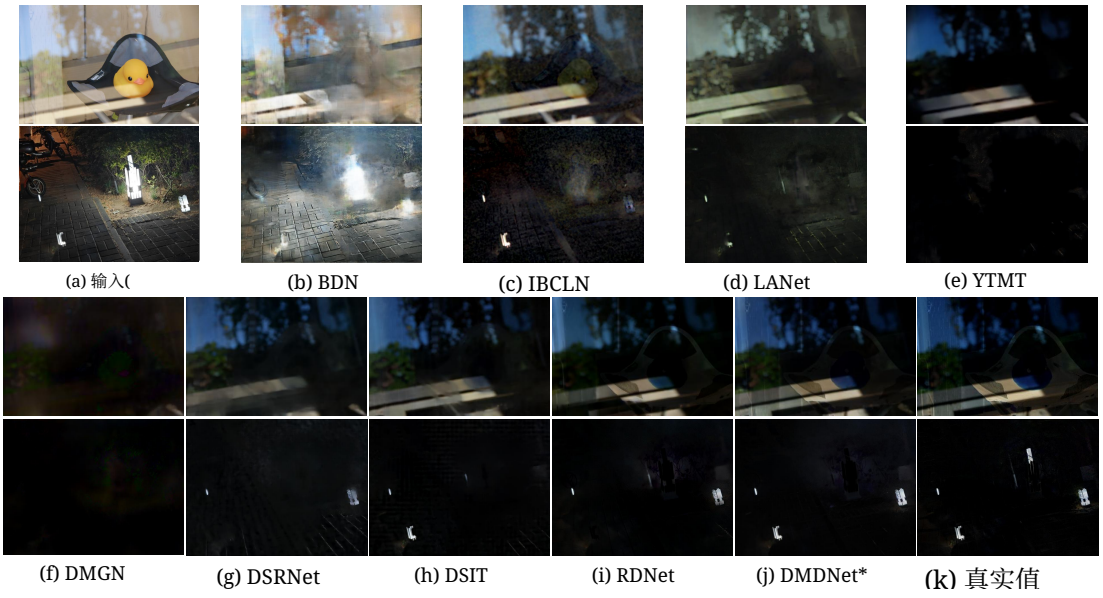


图8：反射层的定性比 较，对应于正文中的图6。与最先进方法相比，我们的DMDNet能

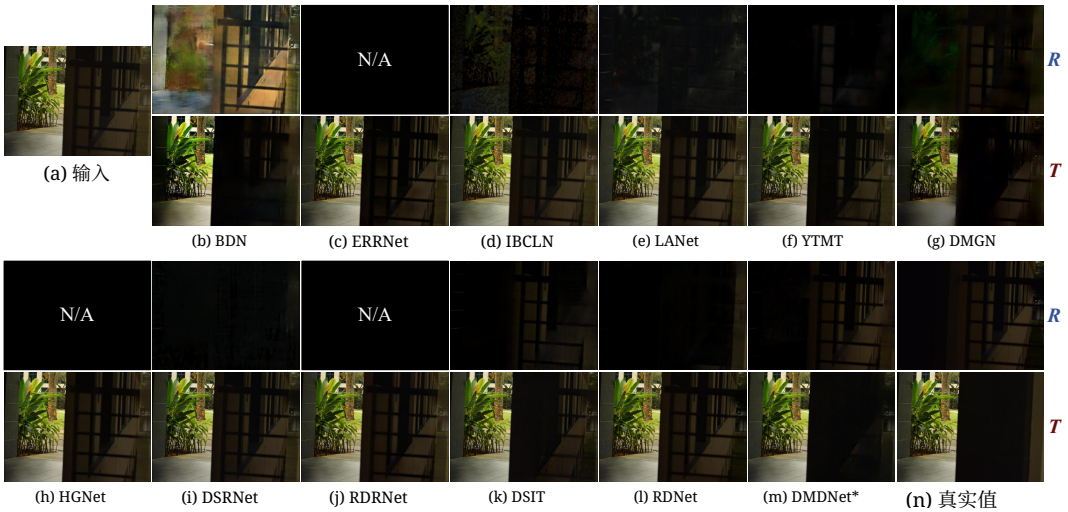


图 9：与日间场景下最先进方法的定性比较。评估中展示了透射层（$\boldsymbol{T}$）和反射层（$\boldsymbol{R}$）。DMDNet 改进了 $\boldsymbol{T}$ 恢复，减少了反射伪影，同时更忠实地恢复了 $\boldsymbol{R}$。“ 不适用 ”表示未提供反射层输出。

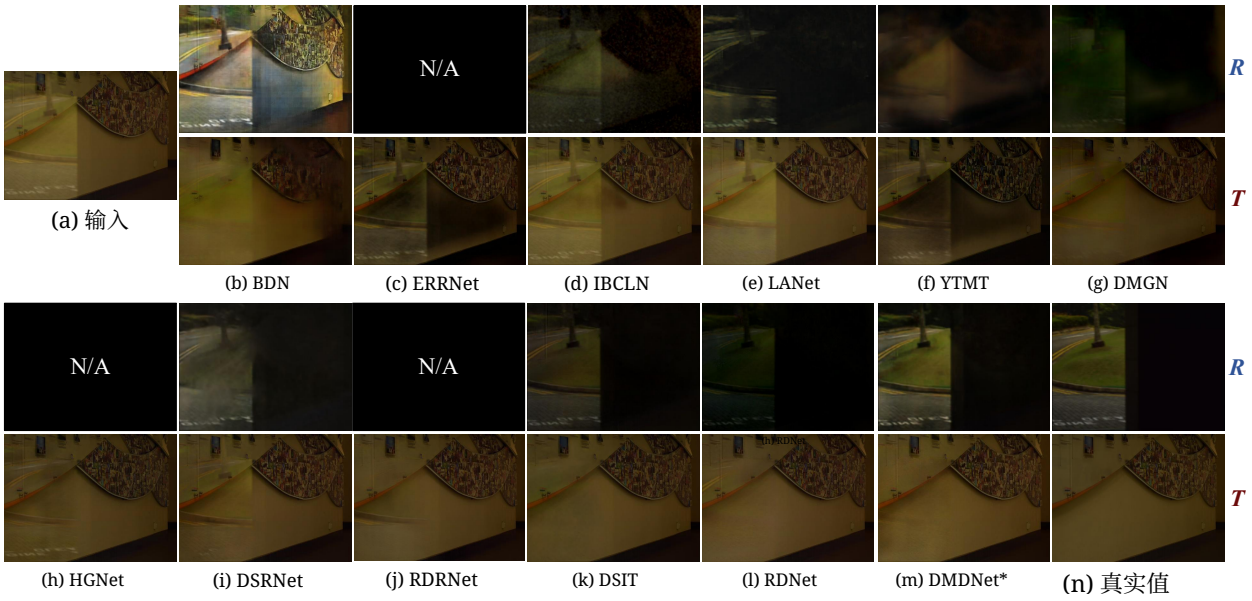


图10：定性与室内场景下最先进方法的比较。DMDNet 能更有效地抑制反射，实现更清晰的 $\boldsymbol{T}$ 恢复与其他方法相比，能够提供更忠实的 $\boldsymbol{R}$ 复原结果。

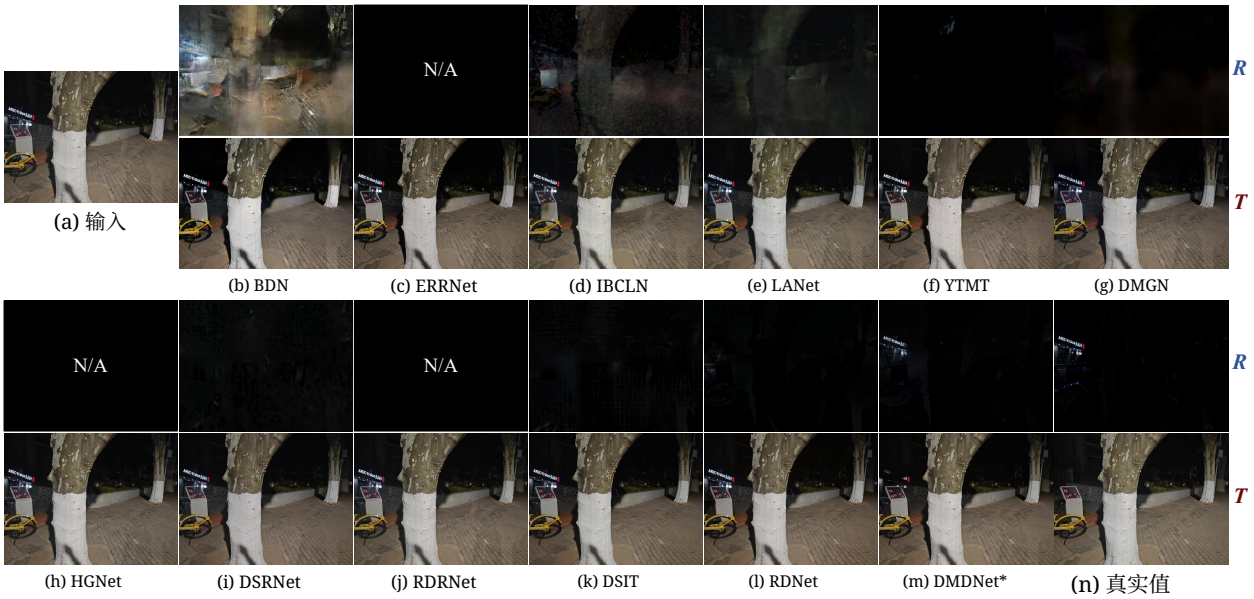


图11：与夜间道路场景下最先进方法的定性比较。DMDNet 能更有效地去除反射，在低光照条件下恢复更清晰的 $\boldsymbol{T}$ 细节，并获得更忠实的 $\boldsymbol{R}$ 复原结果。

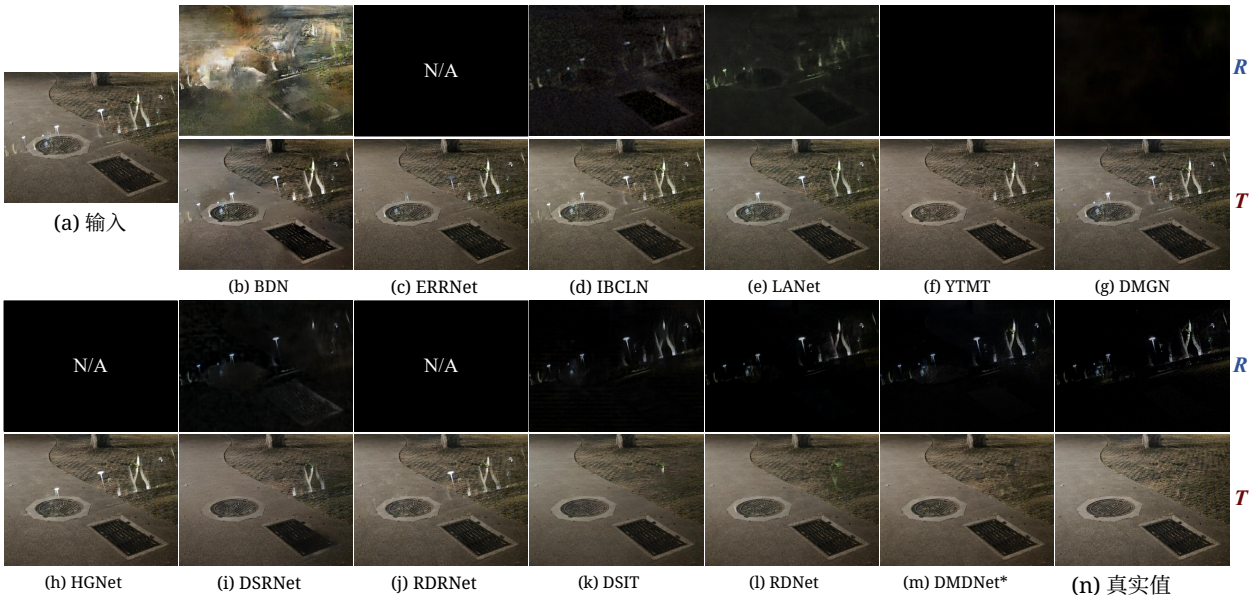


图 12：与夜间地面场景下最先进方法的定性比较。DMDNet 能更有效地抑制反射，恢复更清晰的 $\boldsymbol{T}$ 细节（如路面纹理和井盖），并比其他方法更忠实地 $\boldsymbol{R}$ 恢复图像。

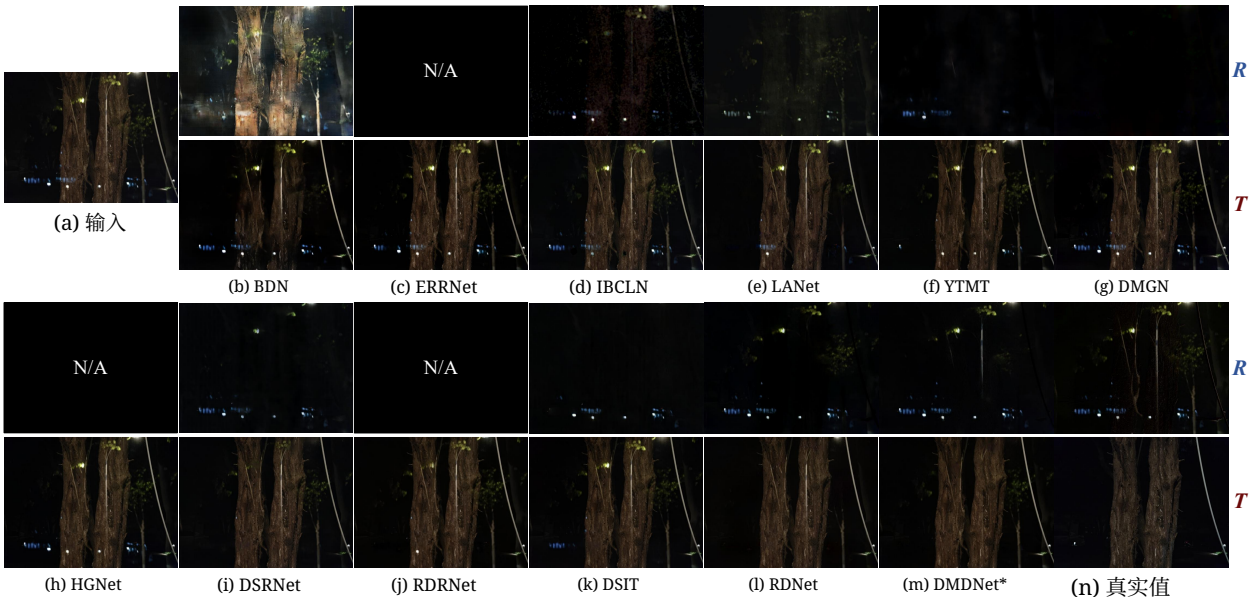


图 13：与夜间自然场景下最先进方法的定性比较。DMDNet 能更有效地抑制散射光产生的反射，恢复更清晰的 $\boldsymbol{T}$ 树木结构，并更忠实地恢复 $\boldsymbol{R}$ 细节。

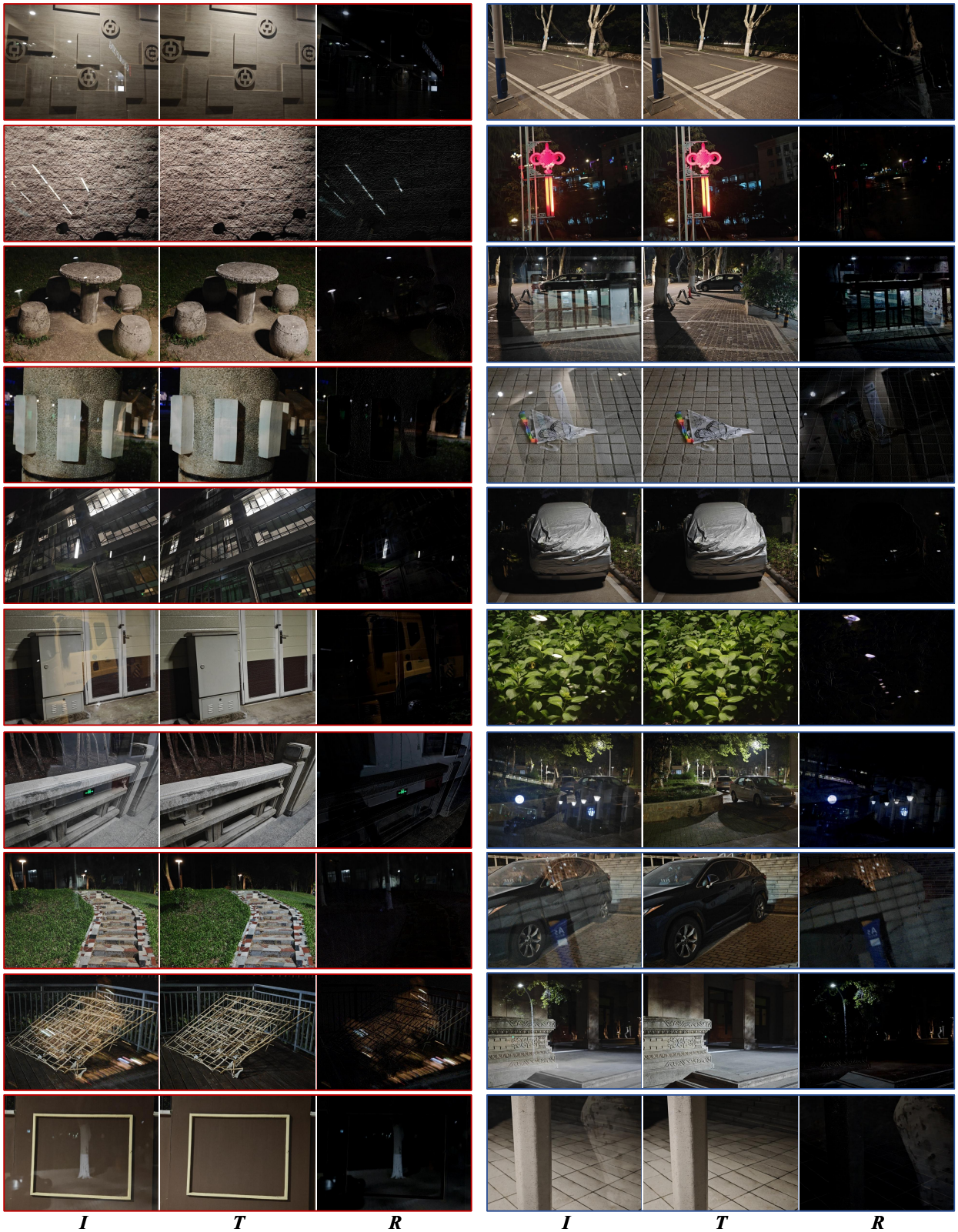


图14：NightIRS 数据集中的示例。 $\boldsymbol{I}$、 $\boldsymbol{T}$ 和 $\boldsymbol{R}$ 分别表示混合图像、透射层和反射层。