恶魔藏于统一性中: 探索用于图像复原的Transformer中的多样化学习者

周世豪[1] 李大宇[1] 潘金山[3] 周俊成[1] 时晶磊[1,4] 杨巨峰[1,2]
[1] 南开大学，计算机学院，VCIP & TMCC & DISSec
[2] 鹏程实验室
[3] 南京理工大学，计算机科学与工程学院
[4] 大连理工大学，SCCI重点实验室

Abstract

基于*Transformer* 的方法在图像复原任务中受到了极大关注，其中的核心组件多头注意力（*MHA*），在捕获多样特征和恢复高质量图像方面发挥着至关重要的作用。在多头自注意力机制中，各注意力头基于均匀划分的子空间独立执行注意力计算，该过程引发的表征冗余问题限制了模型获取理想输出的能力。本文提出通过构建多样化特征学习单元以及设计跨头的多维交互机制改进传统多头自注意力机制，进而构建了一种针对图像复原任务的分层多头注意力驱动的*Transformer* 模型（*HINT*）。*HINT* 模型包含两大核心模块，分别是分层多头注意力（*HMHA*）模块与查询-键缓存更新（*QKCU*）模块，旨在针对性地解决原始多头自注意力架构固有的表征冗余问题。具体而言，*HMHA*通过利用各头从不同大小、包含不同信息的子空间中学习，提取多样化上下文特征。此外，*QKCU*包含层内和层间机制，通过促进层内及跨层注意力头间的增强交互，进一步缓解冗余问题。我们在***5*** 类图像复原任务的***12*** 个基准数据集（包括低光增强、去雾、去雪、去噪、去雨）上展开了广泛的实验，以验证*HINT*的优越性。源代码已随文公开于*https://github.com/joshyZhou/HINT*。

1. 引言

在图像复原领域，基于*Transformer* 框架*[19, 59]*的解决方案表现出卓越性能。这类范式的成功源自于自注意力机制（*SA*）能够对像素的非局部关系进行建模，从而恢复图像的全局结构。现有众多的研究致力于设计高效的自注意力（*SA*）机制变体以实现高质量复原结果 *[37, 57, 82]*。而值得注意的是，多头注意力（*MHA*）作为核心基础组件，通过在均匀划分的子空间中并行执行自注意力计算，既实现了高计算效率，又增强了特征捕获的多样性。

传统多头注意力（*MHA*）机制存在表征冗余的固有缺陷。自然语言处理（*NLP*）领域的研究 *[48, 49, 64]* 表明，少数注意力头对最终决策的贡献占主导地位，其余头则可被裁剪而不影响性能。在本文中，我们

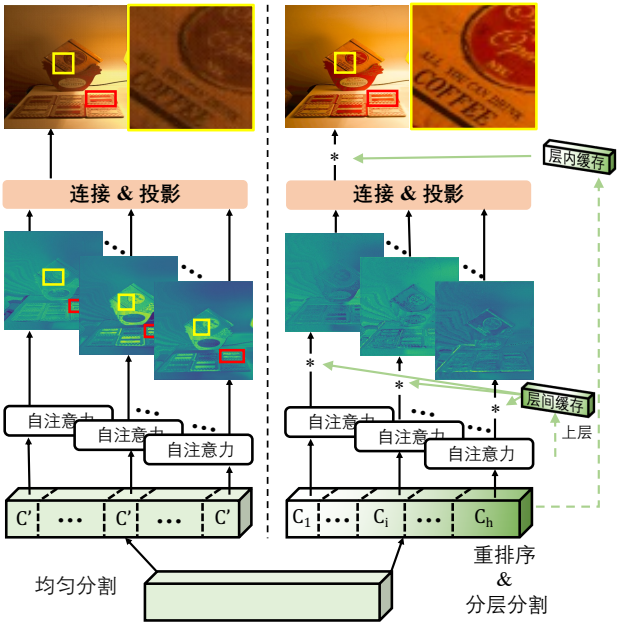


Figure 1. 原始多头自注意力（vanilla MHA [2, 37, 82]，左图）与本文提出的配备QKCU 模块的分层多头注意力（HMHA，右图）在低光增强任务中的对比。标准MHA 将相同尺寸大小(C')的子空间均分给h个注意力头，各头独立计算注意力。这种机制导致不同的头倾向于聚焦相同区域（红色框），而忽视部分退化区域的复原（黄色框），最终输出结果丢失细节并引入模糊效果。与之对应的是，HMHA 在分层子空间划分前执行重排序操作，促使模型学习多样化的表征。QKCU 模块通过层内-层间交互机制增强头间信息流动，对HMHA中的特征进行调制，最终促成更好的输出。

确认该问题在图像复原任务中同样存在，并进一步发现：表征冗余问题的根源在于不同注意力头对相同区域的聚焦。如图 *1* 所示，标准多头注意力（*MHA*）机制为h个注意力头分配相同维度(C') 的子空间，各头以并行且完全独立的方式执行注意力计算。不同注意力头的可视化特征揭示了原始*MHA* 的核心问题，不同的头倾向于聚焦相同区域，造成表征冗余（红色框），同时忽视对部分退化区域的复原，导致细节丢失与模糊（黄色框）。

针对这一问题，我们从两个方面对多头注意力（*MHA*）进行改进。首先，考虑到不同的注意力头从统一规模大小的子空间中学习，而这些子空间包含相似信息，最终导致多头聚焦相同区域的冗余。为缓解这一问题，我们提出在进行层次化子空间划分前引入基于通道相似性的重排序策略。通过这种方式，每个子空间承载与其他子空间相互独立的信息，且各子空间的尺寸规模差异化。该设计将注意力头建模为差异化学习者，构建了分层多头注意力模块（即*HMHA*），以替代原始多头注意力模块。与传统*MHA* 相比，该模块可有效提取差异化表征。其次，考虑到注意力头间缺乏协作进一步加剧了表征冗余问题。我们提出通过查询- 键缓存更新（*QKCU*）机制，从层内与层间角度增强注意力头间的交互。具体而言，层内缓存充当门控模块，通过筛选各注意力头捕获的聚合特征，针对性增强其中的有效信息。另一方面，层间缓存通过历史注意力分数，调制当前层各注意力头的注意力计算得分。层内与层间调制均基于输入的特性，赋予了*HMHA* 更强的学习多样化上下文表征能力。基于分层多头注意力（*HMHA*）与查询- 键缓存更新（*QKCU*）上述两大核心组件，我们提出了一种由分层多头注意力驱动专为图像复原任务设计的*Transformer* 模型（*HINT*）。

总体而言，本文的主要贡献总结如下：

- 我们提出了一种分层多头注意力驱动的*Transformer* 模型*HINT*，用于去除图像中的退化元素。*HINT* 验证了在多头注意力（*MHA*）机制中探索多样性学习者，以及通过层间/ 层内方式增强头间交互对图像复原问题的有效性。
- *HINT* 引入分层多头注意力（*HMHA*），通过使模型从不同子空间学习差异化上下文特征，缓解标准*MHA* 中的冗余问题。此外，*HINT* 融入查询- 键缓存更新（*QKCU*）机制，结合层内/ 层间方案增强头间交互。
- *HINT* 在***5*** 个图像复原任务（低光增强、去雾、去雪、去噪、去雨）的***12*** 个基准数据集上开展了广泛定性与定量评估，结果表明其在图像恢复质量和模型复杂度方面均优于现有算法。

2. 相关工作

2.1. 图像复原

在非理想环境中捕获的图像常导致低质量输出，进而对下游任务造成负面影响 *[5, 20, 28]*。图像复原通过从退化图像（如雾 *[25, 32, 56]*、雨痕 *[22, 51, 66]*、低光照 *[2, 70, 75]*）中复原清晰图像，提供了一种可行的解决方案。过去几十年来，图像复原领域见证了从传统手工方法 *[9, 21]*到基于学习的*CNN* 模型 *[38, 45, 79]*的范式转变。为实现复原性能的提升，研究者们提出了多样的高效模块与先进的架构设计。其中，带跳跃连接的残差特征学习 *[23, 40, 86]*、用于分层表征的编解码架构 *[10, 30, 80]*，以及聚焦重要信号的注意力机制 *[17, 55, 83]*，已成为主流组件。

近年来，基于*Transformer* 的模型 *[59]* 已被适配于底层视觉任务，在多种图像复原任务 *[2, 57, 74]*）中取得显著成效。*IPT [3]* 是首个将视觉*Transformer*架构 *[19]* 引入底层视觉任务的开创性工作，并取得了令人惊艳的结果。原始自注意力的平方级计算复杂度阻碍其处理高分辨率输入，促使研究者探索降低计算负载的解决方案。针对这一问题，*Restormer [82]* 提出沿通道维度计算注意力分数。另一种方法是基于窗口的注意力 *[43]*，该设计被*Uformer [69]*、*SwinIR [37]* 等模型应用。尽管这些注意力机制成功减轻了计算负担，但仍依赖原始多头注意力（*MHA*）设计 *[59]*，导致冗余问题 *[47, 49, 73]* 并进一步限制模型表征能力。

2.2. 多头注意力

作为*Transformer* 的基础组件，多头注意力（*MHA*）在捕捉多样关系、并在实际应用中取得优异性能方面发挥关键作用。遗憾的是，学界逐渐认识到并非所有注意力头都会对目标任务的均等贡献 *[60]*。现有方法尝试通过引入注意力头间的交互或协作（如 *[1, 35, 85]*）解决这一问题，但由于各头仍独立运算，其表达能力提升仍受限 *[73]*。另一种潜在补救方法是对各注意力头的查询、键、值投影进行调制 *[12, 41]*。尽管这种方法调整了底层信息流，但因其静态本质，仍缺乏对输入的适应性 *[73]*。近期，一些工作探索了动态调制注意力分数的方法 *[48, 54, 64]*，在不显著增加计算负载的前提下实现了模型表达能力的提升。

针对图像复原问题，本研究在动态调制注意力分数的基础思想（如 *[48, 73]*）之上，进一步尝试缓解原始*MHA* 的表达能力受限问题。具体来说，标准*MHA* 为每个注意力头分配相同维度大小的子空间 *[3, 37, 82]*），这种做法限制了各头学习独特特征的能力并最终导致特征冗余。与传统做法不同的是，我们提出通过分层多头注意力（*HMHA*）学习层次化表示，利用不同注意力头在不同规模大小的子空间中提取多样的上下文信息。此外，我们引入*QKCU* 机制，融合层内与层间调制的双重方案，增强注意力头间的交互。与现有复原方法采用静态投影方式调制注意力分数 *[8, 88]*不同，*HINT* 通过动态调制实现了更优的模型表达能力。

3. 方法

整体架构。如图 *2*所示，所提出的*HINT* 采用编解码器（*encoder-decoder*）架构。给定退化输入图像$\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$，首先通过卷积层提取浅层特征$\mathbf{F}_s \in \mathbb{R}^{H\times W\times C}$，其中$H$、$W$和$C$分别表示高度、宽度和通道数。浅层特征随后通过$N_1$层级的复原模型处理，生成深层特征$\mathbf{F}_d \in \mathbb{R}^{H\times W\times C}$。在编码器和解码器的每一层级，有$N_2$个基本块，并配备用于下采样或上采样的卷积层。遵循现有工作 *[29, 88, 89]*，我们采用非对称设计：省略编码器部分的自注意力机制，以提升复原性能。因此，编码器中的基本块仅包含前馈网络（*FFN [82]*），而解码器的块则同时包含本文提出的分层多头注意力（*HMHA*）和*FFN*。我们采用1×1卷

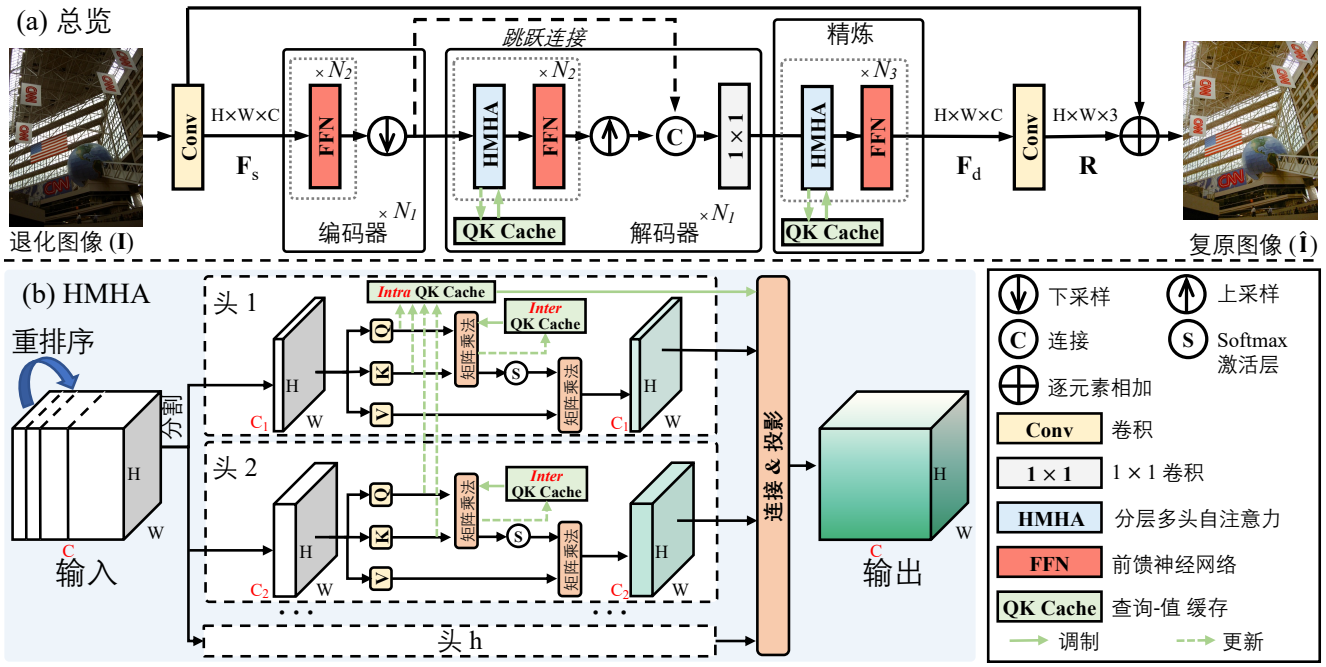


Figure 2. 分层多头注意力驱动的Transformer 模型（HINT）示意图：（a）HINT的整体架构（b）分层多头注意力（HMHA）。

积的跳跃连接操作，以在解码器层中利用编码器的特征。此后，我们设计了一个由N_3个基本块组成的精炼阶段，以进一步增强所学习到的特征。最后，一个3×3卷积层对深层特征$\mathbf{F}_d$进行处理，以生成残差图像$\mathbf{R} \in \mathbb{R}^{H \times W \times 3}$。复原图像通过残差图像与退化图像相加得到，即$\hat{\mathbf{I}} = \mathbf{I} + \mathbf{R}$。

3.1. 分层多头注意力（HMHA）

标准*MHA* 为每个注意力头分配相同大小且包含相似信息的子空间，这阻碍了模型学习独特特征的能力并导致信息冗余。为解决这一问题，*HINT* 构建分层多头注意力（*HMHA*）以探索不同维度的子空间，促使各注意力头学习多样化的上下文信息。

首先回顾主流方法 *[37, 82, 88]*所采用的多头注意力（*MHA*）中的缩放点积注意力机制 *[59]*。给定归一化的输入张量$\mathbf{X} \in \mathbb{R}^{H \times W \times C}$，其注意力分数计算如下：

$$\begin{aligned} \mathit{Attention(Q,K,V)} &= \mathit{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \\ \mathbf{Q} = \mathbf{X}\mathbf{W}_Q, \mathbf{K} &= \mathbf{X}\mathbf{W}_K, \mathbf{V} = \mathbf{X}\mathbf{W}_V, \end{aligned} \tag{1}$$

其中，$\mathbf{W}_Q \in \mathbb{R}^{C \times d_k}$, $\mathbf{W}_K \in \mathbb{R}^{C \times d_k}$, 和 $\mathbf{W}_V \in \mathbb{R}^{C \times d_v}$分别为查询（$Q$）、键（$K$）和值（$V$）的线性投影矩阵。值得注意的是，在自注意力中，键与值的维度相等，即$d_k = d_v$

传统*MHA [19, 82]* 利用多个注意力头并行执行缩放点积注意力。通过在不同子空间上执行注意力函数，模型得以增强表征能力。具体而言，标准*MHA*采用h个注意力头在不同子空间中学习（*Q,K,V*）的表示，即每个头的子空间维度为d_k/h。*MHA*的计算公式化为：

$$\begin{aligned} \mathit{MultiHead(X)} &= \mathit{Concat}(\mathrm{H}_1, \mathrm{H}_2, ..., \mathrm{H}_\mathrm{h})\mathbf{W}_p, \\ \mathrm{H}_\mathrm{i} &= \mathit{Attention}(\mathbf{X}\mathbf{W}_Q^i, \mathbf{X}\mathbf{W}_K^i, \mathbf{X}\mathbf{W}_V^i), \end{aligned} \tag{2}$$

其中，$\mathbf{W}_Q^i \in \mathbb{R}^{C \times d_k/h}$, $\mathbf{W}_K^i \in \mathbb{R}^{C \times d_k/h}$, 和 $\mathbf{W}_V^i \in \mathbb{R}^{C \times d_v/h}$ 为第i个头的投影矩阵。输出投影矩阵$\mathbf{W}_p \in \mathbb{R}^{d_v \times d_{out}}$ 负责聚合所有头捕获的特征。

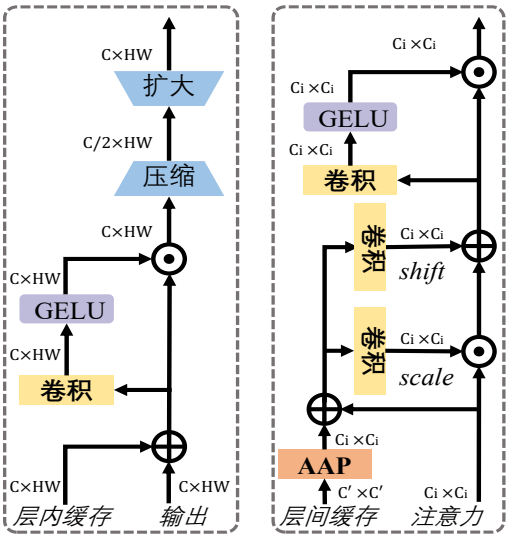
为增强*MHA* 的表达能力，我们引入分层表征学习过程。具体而言，所提出的*HMHA*通过将通道空间划分为$C = [C_1, C_2, \ldots, C_h]$（满足$C_1 \le C_2 \le ... \le C_h$），为每个注意力头分配差异化的子空间。在执行通道划分前，首先基于通道相似性进行重排序，确保每个头聚焦于独特的语义特征。随后，各头在专属子空间内执行点积注意力，通过不同维度的子空间操作捕获多样化的上下文信息。

3.2. 查询-键缓存更新机制

HMHA 机制促使多注意力头学习包含多样化上下文信息的特征，而这些头仍如传统方法 *[37, 82]*一样保持独立工作。这种方式下，存在层内及跨整个模型的协作缺失的问题，最终，模型难以实现最佳图像恢复结果。为缓解这一问题，我们提出*QKCU* 机制，增强层内及层间的注意力头间交互，如图 *3* 所示。

层内查询-键缓存调制与更新。给定两个输入：层内缓存$\mathbf{F}_{\mathbf{intra}} \in \mathbb{R}^{C \times HW}$和*HMHA*输出$\mathbf{F}_{\mathbf{out}} \in \mathbb{R}^{C \times HW}$，首先计算两者特征和：$\mathbf{F}^{\mathbf{s}}_{\mathbf{intra}} = \mathbf{F}_{\mathbf{intra}} + \mathbf{F}_{\mathbf{out}}$。为在信息流中选择性保留最具关键信息的元素，我们引入一种门控机制，表示为：

$$\mathbf{F}_{\mathbf{gated}} = \mathit{GELU}(\mathit{Conv}(\mathbf{F}^{\mathbf{s}}_{\mathbf{intra}})) \odot \mathbf{F}^{\mathbf{s}}_{\mathbf{intra}}, \tag{3}$$



(a) Intra Modulation (b) Inter Modulation

Figure 3. 查询- 键缓存更新机制

其中，$\mathbf{F_{gated}} \in \mathbb{R}^{C\times HW}$为门控机制输出结果，⊙表示逐元素乘法，Conv(·) 和GELU(·) 分别表示为卷积操作与*GELU* 激活函数 *[26]*。随后，我们通过特征压缩与重建对该变换后的特征进行适配:

$$\mathbf{F^o_{Intra}} = \mathrm{Conv_{up}}(\mathrm{Conv_{down}}(\mathbf{F_{gated}})), \tag{4}$$

其中，$\mathbf{F^o_{Intra}} \in \mathbb{R}^{C\times HW}$为调制后的输出特征，$\mathrm{Conv_{up}}(\cdot)$ 和$\mathrm{Conv_{down}}(\cdot)$分别为将通道维度投影至高维和低维的卷积操作，该设计促使模型聚焦数据的关键特征。

接下来，为更新每层的层内查询-键缓存，我们按如下方式计算每个头的查询*(**Q**)* 投影和键*(**K**)* 投影之和:

$$\begin{aligned} \mathbf{F}^i &= \mathbf{Q}^i + \mathbf{K}^i, \\ \mathbf{F_{intra}} &= \mathrm{Concat}(\mathbf{F}^1, ..., \mathbf{F}^h) \end{aligned} \tag{5}$$

层间查询-键缓存调制与更新。给定两个输入张量：层间缓存$\mathbf{F_{inter}} \in \mathbb{R}^{C'\times C'}$和查询-键点积注意力$\mathbf{F_{att}} \in \mathbb{R}^{C_i\times C_i}$，我们首先调整$\mathbf{F_{inter}}$的尺寸，得到$\hat{\mathbf{F}}_{\mathbf{inter}} \in \mathbb{R}^{C_i\times C_i}$。然后，我们计算用于逐像素调制的“缩放”和“偏移”分量:

$$\begin{aligned} \mathbf{F_{shift}} = \mathbf{F^s_{inter}}\mathbf{W_{shift}}, \mathbf{F_{scale}} = \mathbf{F^s_{inter}}\mathbf{W_{scale}}, \\ \mathbf{F_m} = \mathbf{F_{scale}} \odot \mathbf{F_{att}} + \mathbf{F_{shift}}, \end{aligned} \tag{6}$$

其中，$\mathbf{F_m} \in \mathbb{R}^{C_i\times C_i}$为调制后特征图，$\mathbf{W_{scale}}$和$\mathbf{W_{shift}}$分别为经学习得到的$\mathbf{F_{scale}} \in \mathbb{R}^{C_i\times C_i}$（缩放分量）与$\mathbf{F_{shift}} \in \mathbb{R}^{C_i\times C_i}$（偏移分量）的投影矩阵。类似地，该特征通过门控机制进一步变换，其可定义为:

$$\mathbf{F^o_{inter}} = \mathit{GELU}(\mathit{Conv}(\mathbf{F_m})) \odot \mathbf{F_m}, \tag{7}$$

其中，$\mathbf{F^o_{inter}} \in \mathbb{R}^{C_i\times C_i}$为层间输出特征。

Table 1. LOL-v2 [77]数据集上的低光增强定量比较，* 表示无监督方法。

Method	LOL-v2-real		LOL-v2-syn		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
*EnGAN [27] (TIP'21)	18.23	0.617	16.57	0.734	17.4	0.676
*RUAS [39] (CVPR'21)	18.37	0.723	16.55	0.652	17.46	0.688
*QuadPrior [67] (CVPR'24)	20.48	0.811	16.11	0.758	18.30	0.785
KinD [87] (MM'19)	14.74	0.641	13.29	0.578	14.02	0.610
Uformer [69] (CVPR'22)	18.82	0.771	19.66	0.871	19.24	0.821
Restormer [82] (CVPR'22)	19.94	0.827	21.41	0.830	20.68	0.829
MIRNet [81] (ECCV'20)	20.02	0.820	21.94	0.876	20.98	0.848
Sparse [77] (TIP'21)	20.06	0.816	22.05	0.905	21.06	0.861
MambaIR [24] (ECCV'24)	21.25	0.831	25.55	0.929	23.40	0.880
SNR-Net [75] (CVPR'22)	21.48	0.849	24.14	0.928	22.81	0.889
IGDFormer [71] (PR'25)	22.73	0.833	25.33	0.937	24.03	0.885
Retinexformer [2] (ICCV'23)	22.80	0.840	25.67	0.930	24.24	0.885
MambaLLIE [72] (NeurIPS'24)	22.95	0.847	25.87	0.940	24.41	0.894
HINT (Ours)	**23.11**	**0.884**	**27.17**	**0.950**	**25.14**	**0.917**

接下来，为更新跨层的层间查询- 键缓存，我们执行逐层缓存计算过程，并对历史缓存进行优化。当前层的缓存情况计算如下:

$$\mathbf{F}^l_{\mathbf{inter}} = \sum_{i=1}^{h} \mathrm{R}(\mathbf{Q}^i{\mathbf{K}^i}^T)C_i/C, \tag{8}$$

其中，*R(·)*是一个将每个头的特征图调整为统一形状的函数，C_i/C是每个特征的分配权重，其中C为通道总数。基于当前层缓存，我们按如下方式渐进式更新层间缓存:

$$\mathbf{F_{inter}} = \alpha\mathbf{F_{inter}} + (1-\alpha)\mathbf{F}^l_{\mathbf{inter}}, \tag{9}$$

其中，超参数α控制层间缓存内信息流的程度。

4. 实验

在本节中，我们针对***5*** 类图像恢复任务（涵盖低光增强、去雾、去雪、去噪、去雨）***12*** 个基准数据集上对*HINT* 进行了评估由于篇幅限制，额外结果（图像去雨与去噪）及详细实验设置见补充材料。

4.1. 实验设置

实验细节。 *HINT*由一个N_1 = *4*层的编码器-解码器组成，其中编码器和解码器采用相同的模块结构：N_2 = *[4, 6, 6, 6]*。在第四层级，编码器与解码器模块沿用 *[69]*的设计，统一为一个瓶颈层。精炼阶段包含N_3= *4*个模块，嵌入特征维度C 设置为*48*. 注意力头数量为*4*，维度比例设置为*[1, 2, 2, 3]*。*HINT* 采用基于皮尔逊相关系数计算相似性的重排序策略。公式 (9) 中的超参数α 依照实验经验取值为*0.9*。我们采用*AdamW* 优化器训练*HINT*。采用广泛使用的损失函数 *[63, 89]* 约束模型训练。

评测指标。 我们采用峰值信噪比（*PSNR*） *[68]*、结构相似性（*SSIM*）等常用指标，以评估有参考图像的基准上的恢复结果。此外，采用无参考指

(a) Input (b) Reference (c) KinD [87] (d) RUAS [39] (e) MIRNet [81] (f) Sparse [77] (g) Restormer [82] (h) HINT (Ours)

Figure 4. LOL-v2 [77] 数据集上的低光增强定性结果。上例源自合成子集，下例源自真实子集。与其他方法相比，HINT 恢复的图像生动逼真，且未引入明显颜色失真。

(a) Input (b) Reference (c) HDCW [7] (d) JSTASR [6] (e) Uformer [69] (f) AST [88] (g) HINT (Ours)

Figure 5. Snow100K [42] 数据集上的去雪定性结果HINT 实现了清晰的去雪效果，而其他方法生成的图像仍残留明显雪痕。

Table 2. Snow100K 数据集 [42] 上的去雪定量结果。

Method	JSTASR ECCV'20 [6]	All in One CVPR'20 [36]	Uformer CVPR'22 [69]	DesnowNet TIP'18 [42]	HDCW ICCV'21 [7]	TransWeather CVPR'22 [58]
PSNR	23.12	26.07	29.80	30.50	31.54	31.82
SSIM	0.86	0.88	0.93	0.94	0.95	0.93
Method	NAFNet ECCV'22 [4]	AST CVPR'24 [88]	FocalNet ICCV'23 [13]	SFNet ICLR'23 [14]	ConvIR-S TPAMI'24 [62]	HINT - (Ours)
PSNR	32.41	32.50	33.53	33.79	33.79	**34.14**
SSIM	0.95	**0.96**	0.95	0.95	0.95	0.94

Table 3. SOTS [33] 上的去雾定量对比。

Method	EPDN CVPR'19 [52]	FDGAN AAAI'20 [18]	AirNet CVPR'22 [34]	InstructIR ECCV'24 [11]	Restormer CVPR'22 [82]	NAFNet ECCV'22 [4]
PSNR	22.57	23.15	23.18	30.22	30.87	30.98
SSIM	0.863	0.921	0.900	0.959	0.969	0.970
Method	FSNet TPAMI'24 [15]	PromptIR NeurIPS'23 [50]	DehazeFormer TIP'23 [56]	AdaIR ICLR'25 [16]	NDR-Restore TIP'24 [78]	HINT - (Ours)
PSNR	31.11	31.31	31.78	31.80	31.96	**32.24**
SSIM	0.971	0.973	0.977	**0.981**	0.980	**0.981**

标*MANIQA* *[76]* 衡量模型对真实场景输入的恢复性能。在表格中，我们分别加粗 和带下划线 标注最佳与次佳结果。

4.2. 主要结果

低光增强。 我们在表 *1* 中，将提出的*HINT* 与现有最优低光增强方法在*LOL-v2* 数据集上进行对比。在真实与合成子集上，*HINT* 在*PSNR* 和*SSIM* 指标上均显著优于其他方法。特别地，在两个子集上取平均后，*HINT* 的*PSNR* 较*Retinexformer* *[2]* 实现了*0.9 dB* 的显著提升，而后者是首个专为低光增强设计的*Transformer* 框架。与基于多种架构的通用图像恢复方案（如*CNN* *[81]*、*Transformer* *[69, 82]*、*Mamba* *[24]*）相比，*HINT* 取得了至少*1.74 dB* 的*PSNR* 增益。值得注意的是，较最新算法*IGDFormer* *[71]*，*HINT* 在*PSNR* 上实现了*1.11 dB* 的显著性能提升。定性对比如图 *4* 所示。在上行子图中，现有方法未能恢复出令人满意的细节，普遍存在模糊与伪影残留。比较方法要么面临过曝/ 欠曝问题 *[39, 77]*（如图 *4 d/f*），要么难以还原真实色彩 *[81, 82, 87]*（如图 *4 c/e/g*）。对于底部真实案例，*HINT* 恢复的图像与参考图像（*GT*）更接近，而其他算法或引入明显模糊痕迹 *[77, 81, 82, 87]*（如图 *4 c/e/f/g*），或导致欠曝问题 *[39]*（如图 *4 d*）。

去雪。 针对图像去雪任务，我们在*Snow100K* 数据集 *[42]* 上开展实验。*HINT* 在所有方法中取得了最佳*PSNR* 成绩，且*SSIM* 结果同样具有竞争力。具体而言，较最新通用恢复框架*AST* *[88]*，*HINT* 在*PSNR* 上实现了*1.64 dB* 的显著提升。此外，较专为去雪设计的方法 *[6, 7, 42]*，*HINT* 展现了更优的性能表现。此外，较旨在解决复杂天气退化的方法 *[36, 58]*，*HINT* 在去雪任务中展现出优势。对于包括基于*CNN* *[4, 13, 14, 62]* 和*Transformer* *[69, 88]* 在内的通用恢复框架，*HINT* 在*PSNR* 上实现了至少*0.35 dB* 的性能增益。视觉对比如图 *5* 所示，其中*HINT* 恢复出无雪痕残留的清晰结果（图 *5g*）去雪方法 *[6, 7]* 的输出则不令人满意（图 *5 c*和 *5d*）。基于*Transformer* 的方法 *[69, 88]* 的结果表现出明显的雪纹伪影（图 *5 e* 和*f*）。

去雾。 我们在*SOTS* *[33]* 数据集上进行图像去雾实验，在表 *3* 中对比了*11* 种代表性算法。*HINT* 在*PSNR* 和*SSIM* 指标上均优于所有对比方法。值得注意的是，*HINT* 在*PSNR* 上超过所有专为去雾设计的方法 *[18, 52, 56]* 至少*0.46 dB*。与多合一*(all-in-one)*图像恢复方法 *[16, 50, 78]* 和通用复原方法 *[4, 11, 15, 82]*相比，*HINT* 仍然表现出优势。图 *6* 展示了定性结果比较。相比其他方法存在颜色失真问题 *[34]*（图 *6 c*）或残留雾气 *[50]*（图*d*），*HINT* 恢复出了生动结果。

(a) Input (b) Reference (c) AirNet [34] (d) PromptIR [50] (e) HINT

Figure 6. SOTS [33] 基准上的去雾定性结果，与其他方法相比，HINT 生成的图像更接近参考图像。

Table 4. 不同自注意力机制的消融实验

Model	W-MSA [69]	MDTA [82]	HMHA Ours
PSNR/SSIM	24.19/0.941	26.42/0.948	27.17/0.950

Table 5. HMHA 的消融实验

	Ranking Strategy	Params	PSNR	SSIM
(a)	No-Ranking [82]	24.76	26.42	0.948
(b)	Random Shuffle [84]	24.87	26.54	0.949
(c)	HMHA (Ours)	24.87	27.17	0.950

Table 6. QKCU的消融实验

	IntraCache	InterCache	Params	PSNR	SSIM
(a)			21.34	26.47	0.949
(b)	√		23.82	26.67	0.949
(c)		√	22.39	26.72	0.949
(d)	√	√	24.87	27.17	0.950

4.3. 分析与讨论

我们已验证了在基于*Transformer* 的模型中配备*QKCU* 和*HMHA* 机制能在多个复原任务上带来性能提升。接下来，我们进一步进行实验以研究分析所提出组件的效果。我们在*LOL-v2-syn* 子集 *[77]* 上训练了搭载不同组件的低光增强模型*HINT*的变体进行消融研究，。为确保公平对比，所有模型均在相同实验设置下完成训练，且*FLOPs* / 运行时间基于*256*×*256* 输入图像计算。

HMHA模块的有效性分析。 为探究本文提出的*HMHA* 是否增强了多样化上下文信息的建模能力，我们将其与两种代表性变体进行对比：（*1*）*W-MSA [41]*，（*2*）*MDTA [82]*。表 *4*展示了定量对比结果。搭载*HMHA* 模块的模型在*PSNR* 和*SSIM* 双指标上均取得最优成绩，显著优于其他变体。具体而言，将*HMHA* 模块替换为*W-MSA* 或*MDTA* 后，性能分别显著下降*2.98 dB* 和*0.75 dB*。为探究*HMHA* 模块的作用，我们进一步可视化了*MDTA* 与*HMHA* 中各注意力头学习到的特征图。如图 *7* 所示，*MDTA* 的注意力头倾向于聚焦相同区域（红色框标注），而*HMHA* 的注意力头在学习不同子空间的差异化表征方面展现显著优势。最终，搭载*HMHA* 的模型生成的结果与参考图像（如图中黄色框标注所示）更为接近。如表 *5* 所示，我们进一步验证了*HMHA* 中重排序策略对学习判别性表征的作用。如表 *5 c* 所示，当我们将重排序策略替换为随机洗牌操作 *[84]*（表 *5 b*）时，*PSNR* 指标有*0.63 dB* 的下降。与之形成对比的是，当模型移除所有排序机制（退化为标准*MHA [82]*，表 *5 a*），*HMHA* 相比该基准模型实现了*0.75 dB* 的性能提升。上述结果证明重排序策略增强模型表达能力的关键作用。

QKCU 模块的有效性分析。 为验证本文提出的*QKCU* 模块有效性，在表 *6* 中对不同变体开展消融实验。禁用*QKCU* 中的层内缓存（*IntraCache*）或跨层缓存（*InterCache*），*PSNR* 指标分别下降*0.45 dB* 和*0.5 dB*，这印证了两大模块在图像恢复任务中重建高质量细节的不可或缺性。综上，*QKCU* 机制在仅增加*16.5*% 参数的少量代价下，实现*0.7 dB* 的*PSNR* 性能净增益。

模型效率分析。 如表 *7* 所示，我们从计算复杂度（浮点运算量*FLOPs* / 参数数量）、推理延迟（运行时间）、性能表现（*PSNR*）三个维度，对模型效率展开全面对比。*HINT* 在实现最高*PSNR* 指标的同时，模型复杂度（参数/ *FLOPs*）显著低于*CNN* 架构的*MIRNet [81]*、*Transformer* 架构的*IPT [3]* 与*Restormer [82]*。

真实场景泛化性评估。 为验证*HINT* 模型在无真实标注的真实场景下的适用性，我们基于*DICM [31]*、*MEF [46]*、*NPE [65]*和*VV [61]*四大无无真实标注基准数据集开展测试。定量对比结果汇总于表 *8*，其中*HINT* 优于所有对比方法。此外，如图 *8* 所示，*HINT* 恢复了视觉上令人愉悦的效果，而对比技术分别遭遇欠曝/ 过曝问题 *[75, 82]*、触发色彩失真 *[77]*，或残留显著伪影 *[69]*。

高层视觉任务应用。 我们在*ExDark* 数据集 *[44]* 上开展了低光目标检测实验。*HINT* 在*LOL-v2-syn* 子集上进行预训练后，直接用于增强低光图像，并以*YOLO-v3* 模型 *[53]* 作为检测器。增强后的图像在定性（图 *9*）和定量（表 *9*）层面均显示为下游任务带来增益。

Table 7. LOL-v2-syn 数据集 [66] 上的模型效率分析

Method	IPT [3]	MIRNet [81]	Uformer [69]	Restormer [82]	HINT
FLOPs/G	6887	785	**12.00**	144.25	126.92
Parameters/M	115.3	31.76	**5.29**	26.13	24.87
Run-times/s	2.23	0.19	**0.13**	0.29	0.28
PSNR/dB	18.30	21.94	19.66	21.41	**27.17**

Table 8. 真实世界数据集上的定量对比（MANIQA 指标 [76]）。上述结果均基于模型在LOL-v2-syn 数据集 [66] 上预训练的最佳权重测试得出。

Dataset	DICM [31]	MEF [46]	NPE [65]	VV [61]	Mean↑
SNR-Net [75]	0.465	0.527	0.480	0.239	0.428
Uformer [69]	0.526	0.634	0.515	0.356	0.508
Restormer [82]	0.535	0.641	0.491	0.356	0.507
Sparse [77]	0.458	0.532	0.324	0.341	0.413
HINT	**0.583**	**0.642**	**0.547**	**0.448**	**0.555**

5. 总结

我们提出了*HINT*，一种基于*Transformer* 架构的图像

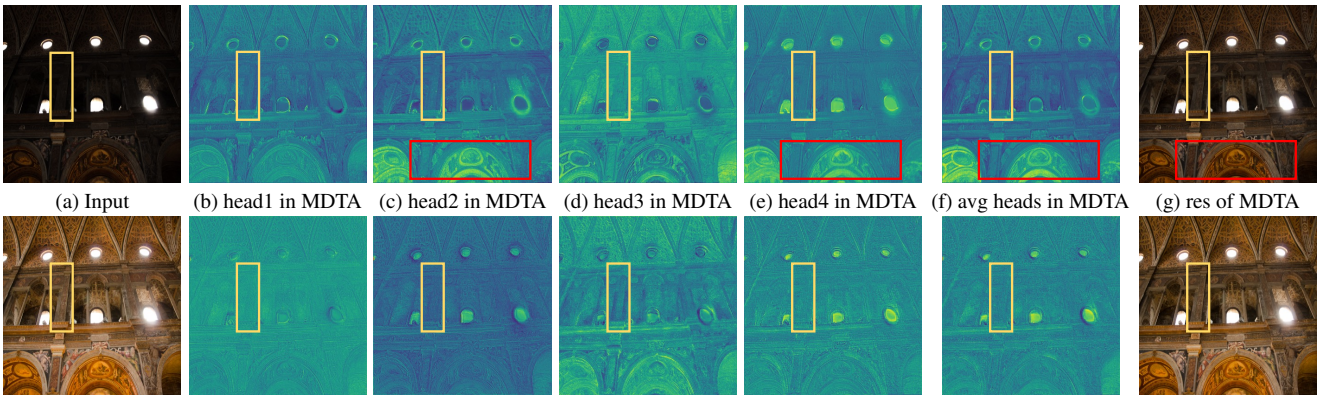


Figure 7. 特征可视化顶行展示了MDTA 中每个头学习到的特征图，而底部的特征图则是来自HMHA 的结果。MDTA 中的头专注于相同区域（红色 框），而HMHA 中的对应头在学习不同子空间的表示上展现优势。因此，配备本文提出的HMHA 的模型恢复出更接近参考图像的结果（黄色 框）。

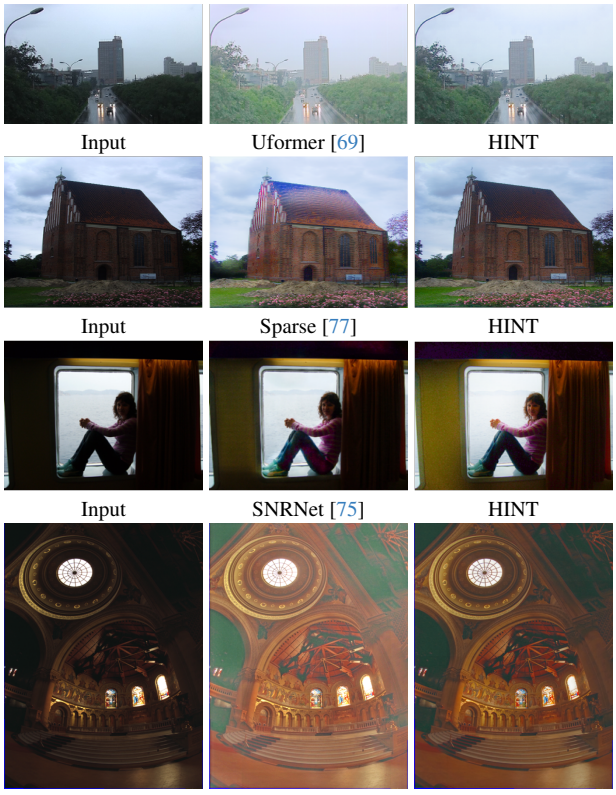


Figure 8. 真实世界基准数据集的可视化结果, 从上到下依次为:NPE [65], DICM [31], VV [61], and MEF [46] 。HINT 模型恢复了更自然的结果，反观对比技术：Restormer [82] 与SNRNet [75] 遭遇欠曝/ 过曝失衡，Sparse [77] 触发色彩失真，Uformer [69] 则残留显著伪影。

Table 9. 低光目标检测ExDark 基准 [44] 的定量结果。HINT 对下游任务有积极影响，平均精度（AP）分数提升7.6%。

Metric	Input	HINT	Δ
Mean (average precision, AP)	45.1	52.7	+7.6

Figure 9. 低光目标检测对比。与低光输入（左）相比，检测器能在经HINT 恢复图像（右）上预测出定位准确边界框。

Figure 10. 失效案例. HINT 在合成数据集上预训练后，不能够应对极微光条件下的输入恢复挑战。

复原方法。*HINT* 在***5*** 类复原任务共计***12*** 个基准数据集上验证了两大设计的有效性：*1*）分层多头注意力（*HMHA*）；*2*）查询- 键缓存更新（*QKCU*）。通过利用*HMHA* 减轻原始多头注意力（*MHA*）中的冗余，并引入*QKCU* 通过层内/ 层间调制增强头间交互，*HINT* 在模型复杂度和准确性上均优于以往先进算法。*HINT* 是在广泛使用的*Transformer* 架构内通过探索高效多头注意力（*MHA*）机制复原高质量图像的工作。本工作为实现卓越复原性能提供了一个有希望的方向，我们期待社区能从中激发兴趣并获益。

局限性. 尽管*HINT* 为图像复原提供了可行解决方案，但解决其引发的失败案例是实现更优结果的一个值得探索的方向如图 *10* 展示了*HINT* 在极微光条件下难以恢复的案例。造成这一现象的可能原因是用于预训练的合成数据集与真实数据之间的差距。收集大规模真实数据集用于进一步训练，是未来工作的一个可能方向。

References

[1] *Karim Ahmed, Nitish Shirish Keskar, and Richard Socher. Weighted transformer network for machine translation.* arXiv preprint arXiv:1711.02132, *2017.* 2

[2] *Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In* ICCV, *2023.* 1, 2, 4, 5

[3] *Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In* CVPR, *2021.* 2, 6

[4] *Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In* ECCV, *2022.* 5

[5] *Tao Chen, Xiruo Jiang, Gensheng Pei, Zeren Sun, Yucheng Wang, and Yazhou Yao. Knowledge transfer with simulated inter-image erasing for weakly supervised semantic segmentation. In* ECCV, *2024.* 2

[6] *Wei-Ting Chen, Hao-Yu Fang, Jian-Jiun Ding, Cheng-Che Tsai, and Sy-Yen Kuo. Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In* ECCV, *2020.* 5

[7] *Wei-Ting Chen, Hao-Yu Fang, Cheng-Lin Hsieh, Cheng-Che Tsai, I Chen, Jian-Jiun Ding, Sy-Yen Kuo, et al. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In* ICCV, *2021.* 5

[8] *Xiang Chen, Hao Li, Mingqiang Li, and Jinshan Pan. Learning a sparse transformer network for effective image deraining. In* CVPR, *2023.* 2

[9] *Sunghyun Cho and Seungyong Lee. Fast motion deblurring.* ACM TOG, *28(5):1–8, 2009.* 2

[10] *Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In* ICCV, *2021.* 2

[11] *Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In* ECCV, *2024.* 5

[12] *Jean-Baptiste Cordonnier, Andreas Loukas, and Martin Jaggi. Multi-head attention: Collaborate instead of concatenate.* arXiv preprint arXiv:2006.16362, *2020.* 2

[13] *Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Focal network for image restoration. In* ICCV, *2023.* 5

[14] *Yuning Cui, Yi Tao, Zhenshan Bing, Wenqi Ren, Xinwei Gao, Xiaochun Cao, Kai Huang, and Alois Knoll. Selective frequency network for image restoration. In* ICLR, *2023.* 5

[15] *Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Image restoration via frequency selection.* TPAMI, *46(2): 1093–1108, 2024.* 5

[16] *Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Shahbaz Khan. AdaIR: Adaptive all-in-one image restoration via frequency mining and modulation. In* ICLR, *2025.* 5

[17] *Xin Deng and Pier Luigi Dragotti. Deep convolutional neural network for multi-modal image restoration and fusion.* TPAMI, *43(10):3333–3348, 2021.* 2

[18] *Yu Dong, Yihao Liu, He Zhang, Shifeng Chen, and Yu Qiao. Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing. In* AAAI, *2020.* 5

[19] *Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In* ICLR, *2021.* 1, 2, 3

[20] *Kaiwen Duan, Song Bai, Lingxi Xie, Honggang Qi, Qingming Huang, and Qi Tian. Centernet++ for object detection.* TPAMI, *46(5):3509–3521, 2024.* 2

[21] *Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, and William T. Freeman. Removing camera shake from a single photograph.* ACM TOG, *25(3):787–794, 2006.* 2

[22] *Xueyang Fu, Jie Xiao, Yurui Zhu, Aiping Liu, Feng Wu, and Zheng-Jun Zha. Continual image deraining with hypergraph convolutional networks.* TPAMI, *45(8):9534–9551, 2023.* 2

[23] *Shuhang Gu, Yawei Li, Luc Van Gool, and Radu Timofte. Self-guided network for fast image denoising. In* ICCV, *2019.* 2

[24] *Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In* ECCV, *2024.* 4, 5

[25] *Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior.* TPAMI, *33(12):2341–2353, 2010.* 2

[26] *Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus).* arXiv preprint arXiv:1606.08415, *2016.* 4

[27] *Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision.* TIP, *30:2340–2349, 2021.* 4

[28] *Chanyoung Kim, Woojung Han, Dayun Ju, and Seong Jae Hwang. EAGLE: eigen aggregation learning for object-centric unsupervised semantic segmentation. In* CVPR, *2024.* 2

[29] *Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In* CVPR, *2023.* 2

[30] *Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In* ICCV, *2019.* 2

[31] *Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms.* TIP, *22(12):5372–5384, 2013.* 6, 7

[32] *Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In* ICCV, *2017.* 2

[33] *Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond.* TIP, *28(1):492–505, 2018.* 5, 6

[34] *Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In* CVPR, *2022.* 5, 6

[35] Jian Li, Baosong Yang, Zi-Yi Dou, Xing Wang, Michael R Lyu, and Zhaopeng Tu. Information aggregation for multi-head attention with routing-by-agreement. In NAACL, 2019. 2

[36] Ruoteng Li, Robby T. Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In CVPR, 2020. 5

[37] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In ICCV Workshops, 2021. 1, 2, 3

[38] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang. Non-local recurrent network for image restoration. In NeurIPS, 2018. 2

[39] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In CVPR, 2021. 4, 5

[40] Xing Liu, Masanori Suganuma, Zhun Sun, and Takayuki Okatani. Dual residual networks leveraging the potential of paired operations for image restoration. In CVPR, 2019. 2

[41] Xiaoyu Liu, Jiahao Su, and Furong Huang. Tuformer: Data-driven design of transformers for improved generalization or efficiency. In ICLR, 2022. 2, 6

[42] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. TIP, 27(6):3064–3073, 2018. 5

[43] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In ICCV, 2021. 2

[44] Yuen Peng Loh and Chee Seng Chan. Getting to know low-light images with the exclusively dark dataset. CVIU, 178: 30–42, 2019. 6, 7

[45] Fangzhou Luo, Xiaolin Wu, and Yanhui Guo. Functional neural networks for parametric image restoration problems. In NeurIPS, 2021. 2

[46] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. TIP, 24(11): 3345–3356, 2015. 6, 7

[47] Paul Michel, Omer Levy, and Graham Neubig. Are sixteen heads really better than one? In NeurIPS, 2019. 2

[48] Tan Nguyen, Tam Nguyen, Hai Do, Khai Nguyen, Vishwanath Saragadam, Minh Pham, Khuong Duy Nguyen, Nhat Ho, and Stanley Osher. Improving transformer with an admixture of attention heads. In NeurIPS, 2022. 1, 2

[49] Tam Minh Nguyen, Tan Minh Nguyen, Dung D. D. Le, Duy Khuong Nguyen, Viet-Anh Tran, Richard Baraniuk, Nhat Ho, and Stanley Osher. Improving transformers with probabilistic attention keys. In ICML, 2022. 1, 2

[50] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. In NeurIPS, 2023. 5, 6

[51] Kuldeep Purohit, Maitreya Suin, AN Rajagopalan, and Vishnu Naresh Boddeti. Spatially-adaptive image restoration using distortion-guided networks. In ICCV, 2021. 2

[52] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In CVPR, 2019. 5

[53] Joseph Redmon. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018. 6

[54] Noam Shazeer, Zhenzhong Lan, Youlong Cheng, Nan Ding, and Le Hou. Talking-heads attention. arXiv preprint arXiv:2003.02436, 2020. 2

[55] Xibin Song, Dingfu Zhou, Wei Li, Yuchao Dai, Zhelun Shen, Liangjun Zhang, and Hongdong Li. Tusr-net: Triple unfolding single image dehazing with self-regularization and dual feature to pixel attention. TIP, 32:1231–1244, 2023. 2

[56] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. TIP, 32:1927–1941, 2023. 2, 5

[57] Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. In ECCV, 2022. 1, 2

[58] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M. Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In CVPR, 2022. 5

[59] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In NeurIPS, 2017. 1, 2, 3

[60] Elena Voita, David Talbot, Fedor Moiseev, Rico Sennrich, and Ivan Titov. Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned. In ACL, 2019. 2

[61] Vassilios Vonikakis, Rigas Kouskouridas, and Antonios Gasteratos. On the evaluation of illumination compensation algorithms. MTA, 77:9211–9231, 2018. 6, 7

[62] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Revitalizing convolutional network for image restoration. TPAMI, 46(12):9423–9438, 2024. 5

[63] Cong Wang, Jinshan Pan, Wei Wang, Jiangxin Dong, Mengzhu Wang, Yakun Ju, and Junyang Chen. Promptrestorer: A prompting image restoration method with degradation perception. In NeurIPS, 2023. 4

[64] Huadong Wang, Xin Shen, Mei Tu, Yimeng Zhuang, and Zhiyuan Liu. Improved transformer with multi-head dense collaboration. TASLP, 30:2754–2767, 2022. 1, 2

[65] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. TIP, 22(9):3538–3548, 2013. 6, 7

[66] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson WH Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In CVPR, 2019. 2, 6

[67] Wenjing Wang, Huan Yang, Jianlong Fu, and Jiaying Liu. Zero-reference low-light enhancement via physical quadruple priors. In CVPR, 2024. 4

[68] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. TIP, 13(4):600–612, 2004. 4

[69] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A gen-

eral u-shaped transformer for image restoration. In CVPR, *2022.* 2, 4, 5, 6, 7

[70] *Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In* BMVC, *2018.* 2

[71] *Yanjie Wen, Ping Xu, Zhihong Li, and Wangtu Xu(ATO). An illumination-guided dual attention vision transformer for low-light image enhancement.* PR, *158:111033, 2025.* 4, 5

[72] *Jiangwei Weng, Zhiqiang Yan, Ying Tai, Jianjun Qian, Jian Yang, and Jun Li. Mamballie: Implicit retinex-aware low light enhancement with global-then-local state space. In* NeurIPS, *2024.* 4

[73] *Da Xiao, Qingye Meng, Shengping Li, and Xingyuan Yuan. Improving transformers with dynamically composable multi-head attention. In* ICML, *2024.* 2

[74] *Jie Xiao, Xueyang Fu, Aiping Liu, Feng Wu, and Zheng-Jun Zha. Image de-raining transformer.* TPAMI, *45(11):12978–12995, 2023.* 2

[75] *Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In* CVPR, *2022.* 2, 4, 6, 7

[76] *Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In* CVPR, *2022.* 5, 6

[77] *Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement.* TIP, *30: 2072–2086, 2021.* 4, 5, 6, 7

[78] *Mingde Yao, Ruikang Xu, Yuanshen Guan, Jie Huang, and Zhiwei Xiong. Neural degradation representation learning for all-in-one image restoration.* TIP, *33:5408–5423, 2024.* 5

[79] *Ke Yu, Xintao Wang, Chao Dong, Xiaoou Tang, and Chen Change Loy. Path-restore: Learning network path selection for image restoration.* TPAMI, *44(10):7078–7092, 2022.* 2

[80] *Zongsheng Yue, Qian Zhao, Lei Zhang, and Deyu Meng. Dual adversarial network: Toward real-world noise removal and noise generation. In* ECCV, *2020.* 2

[81] *Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In* ECCV, *2020.* 4, 5, 6

[82] *Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In* CVPR, *2022.* 1, 2, 3, 4, 5, 6, 7

[83] *Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for fast image restoration and enhancement.* TPAMI, *45(2):1934–1948, 2023.* 2

[84] *Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In* CVPR, *2018.* 6

[85] *Xiaofeng Zhang, Yikang Shen, Zeyu Huang, Jie Zhou, Wenge Rong, and Zhang Xiong. Mixture of attention heads: Selecting attention heads per token. In* EMNLP, *2022.* 2

[86] *Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. In* ICLR, *2019.* 2

[87] *Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In* ACMMM, *2019.* 4, 5

[88] *Shihao Zhou, Duosheng Chen, Jinshan Pan, Jinglei Shi, and Jufeng Yang. Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration. In* CVPR, *2024.* 2, 3, 5

[89] *Shihao Zhou, Jinshan Pan, Jinglei Shi, Duosheng Chen, Lishen Qu, and Jufeng Yang. Seeing the unseen: A frequency prompt guided transformer for image restoration. In* ECCV, *2024.* 2, 4