

## 适应或灭亡：自适应稀疏变换器，具有用于图像恢复的细心特征细化

周世豪<sup>1,2</sup> 陈多生<sup>1</sup> 潘金山<sup>3</sup> 施静蕾<sup>1\*</sup> 杨巨峰<sup>1,21</sup> 南开大学计算机学院 VCIP & TMCC & DISec<sup>2</sup> 南开国际高级研究院（深圳·福田）<sup>3</sup> 南京大学计算机科学与工程学院科学技术学院

zhoushihao96@mail.nankai.edu.cn, duoshengchen@mail.nankai.edu.cn, sdluran@gmail.com

jinglei.shi@nankai.edu.cn, yangjufeng@nankai.edu.cn

### Abstract

基于 Transformer 的方法在图像恢复任务中取得了可喜的性能，因为它们能够对远程依赖性进行建模，这对于重新覆盖清晰图像至关重要。尽管各种有效的注意力机制设计已经解决了与使用变压器相关的密集计算，但通过考虑所有可用的标记，它们通常涉及来自不相关区域的冗余信息和噪声交互。在这项工作中，我们提出了一种自适应稀疏变换器（AST）来减轻不相关区域的噪声交互并消除空间域和通道域中的特征冗余。AST 包括两个核心设计，即自适应稀疏自注意力（ASSA）块和特征细化前馈网络

（FRFN）。具体来说，ASSA 是使用双分支范式自适应计算的，其中引入稀疏分支来过滤低查询关键匹配分数对聚合特征的负面影响，而密集分支则确保有足够的信息流通过网络进行学习-歧视性表述。同时，FRFN 采用增强和简化方案来消除通道中的特征冗余，增强清晰潜在图像的恢复。常用基准的实验结果证明了我们的方法在雨条纹去除、真实雾霾去除和雨滴去除等多项任务中的多功能性和竞争性能。代码和预训练模型可在 <https://github.com/joshyZhou/AST> 获取。

### 一、简介

图像恢复的目的是从退化的图像中恢复清晰的图像。现有的基于 CNN 的方法[6,55,103]取得了显着的进步。然而，他们的基本单位，

\* 通讯作者。

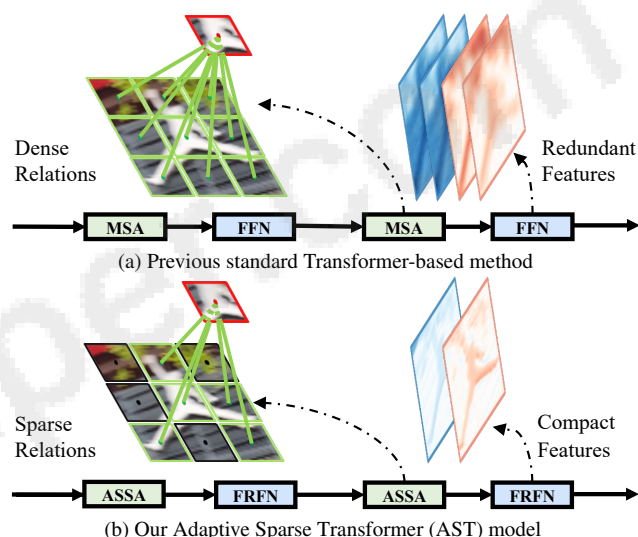


图 1. 工作流程比较。(a) 标准的基于 Transformer 的方法将所有可用的 token 纳入多头自注意力 (MSA) 计算中，并使用前馈网络 (FFN) 来处理冗余特征。(b) 所提出的自适应稀疏变换器 (AST) 包括一个自适应稀疏自注意力 (ASSA) 块，用于过滤掉不相关标记中的噪声交互，以及一个特征细化前馈网络 (FRFN)，以减少冗余隐藏在频道中。

卷积具有受限的感受野，并且在建模远程依赖性时效率较低。虽然最近基于 Transformer 的 [73] 架构通过结合自注意力机制来探索全局相关性来解决这一限制，但它在实际应用中遇到了高计算复杂性。尽管尝试设计有效的注意力机制 [37,82,100] 为了解决计算挑战，障碍仍然存在，原因有两个：1) 标准 Transformers [37, 100] 采用密集的注意力关系来聚合特征，这会无意中在不相关的区域引入噪声交互，如图 1 所示。2) 冗余信息[77, 113] 在密集聚合的特征中

地图可能会进一步阻碍模型关注信息特征。最近,人们努力过滤掉不相关区域中的噪声交互,并删除特征表示中的冗余信息[8,114]。这些方法要么采用 Top-K 选择操作来选择最有用的标记[8],要么项目在执行自注意力计算之前将特征映射到超像素空间[114]。由于参数 K 可能对特定的恢复任务敏感,并且在超像素空间中进行的自注意力机制考虑了所有标记之间的关系,因此它们仍然可能遇到与特征图冗余相关的挑战。在实践中,设计一种有效的机制来识别特征图信息流中最有价值的特征,同时对特定恢复任务表现出较低的敏感性。标准 Transformers [82,100] 通常考虑所有查询键对注意关系来聚合特征。不幸的是,由于并非所有查询标记都与键中的相应标记密切相关,因此利用所有相似性对于清晰的图像重建来说是无效的。直观地说,开发一个稀疏 Transformer 来选择令牌之间最有用的交互可以增强特征聚合。为了实现注意力的稀疏性,基于 squaredReLU 的激活[67]似乎是一个可行的解决方案。它消除了具有负相关性的相似性,而不考虑像[8]这样的特定参数设置。然而,一些特定的设计[23,85]经常需要放松稀疏性以减轻信息损失[66],这与使用稀疏自注意力而不是标准密集自注意力的动机相矛盾。因此,我们探索另一种范式,以确保减少噪声表示特征,并尽可能保留信息丰富的特征。鉴于此,我们提出了一种高效的基于 Transformer 的模型,称为自适应稀疏变换器 (AST),用于图像恢复。AST 引入了两个关键模块:自适应稀疏自注意力模块 (ASSA) 和特征细化前馈网络 (FRFN)。简而言之,ASSA 由两个分支组成:稀疏自注意力分支 (SSA) 和密集自注意力对应分支 (DSA)。具体来说,SSA 用于过滤掉 token 之间不相关的交互,而 DSA 则用于增强-确保必要的信息流经整个网络。我们以自适应方式为每个分支分配权重,使模型能够适应两个分支的影响。与标准的自注意力方法相比,这种设计可以实现更有效的特征聚合,但计算负担有限。另一方面,我们开发了一种简单而有效的常规前馈网络 [11] 替代方案,即 FRFN,以增强特征表示为了更好地恢复潜在图像。简而言之,FRFN 通过增强和简化方案执行特征转换。增强特征图的信息部分,然后

使用门机制减少冗余。同时,FRFN 在抑制沿通道维度的冗余信息方面补充了 ASSA,而 ASSA 则减少了空间域的冗余。由于两个互补组件的合作,AST 捕获了最具代表性的特征,同时在一定程度上抑制了信息量较小的特征。总的来说,这项工作的关键贡献有三个方面:

- 我们提出了 AST,一种高效的 Transformer-基于模型,促进最有用的信息向前流动,提取更有建设性的特征来恢复清晰的图像。
- AST 结合了一个 ASSA 块,其中包括一个密集的自我注意力分支和一个稀疏的分支,以自适应地捕获令牌之间的信息交互,同时保留基本信息。此外,我们开发了一种基于特征转换方案的新特征细化前馈网络 (FRFN),即增强有价值的特征,同时抑制信息较少的特征。
- 进行了综合实验以消除几种类型的退化:雨纹、雾霾和雨滴,展示了我们 AST 设计的优越性。此外,我们还提供广泛的消融研究来突出设计贡献。

## 2.相关工作

图像恢复。高质量图像对于下游应用程序获得令人满意的性能至关重要,例如识别 [28,76,101]、分割 [97,108,110]、表示学习 [42,84,112] 和重建 [117,118] 以图像 [45,83,115] 和视频 [107,109,111] 的形式。在过去的几十年里,研究界见证了从传统的基于先验的模型 [20,92,103] 到基于学习的方法 [40,50,95] 的巨大范式转变,因为它们在消除各种退化方面具有令人印象深刻的表现,例如雨条纹 [14,39,63],阴霾 [18,60,116],雨滴 [54,71,93] 等。性能的提升可归因于不同的架构结构 [64] 和先进的组件 [21,25,27] 受到高级视觉任务的启发。例如,U形网络设计和跳跃连接被广泛应用于获得分层多尺度表示 [9,29,98] 并学习残差信号 [17,44,106]。尽管基于 CNN 的网络取得了令人印象深刻的结果,但它们仍然面临卷积运算的感受野有限的问题。为了解决这个限制,最近的工作 [10,53,68] 探索了注意力机制以获得更好的恢复性能。例如,SPANet [78] 扩展了 IRNN 模型以显式生成雨纹的注意力图。RCAN [105] 设计了一种通道注意力机制来强调更多信息特征。NTIRE 挑战报告 [49,80] 和最近的评论 [31,41,104] 总结了更多网络架构设计。

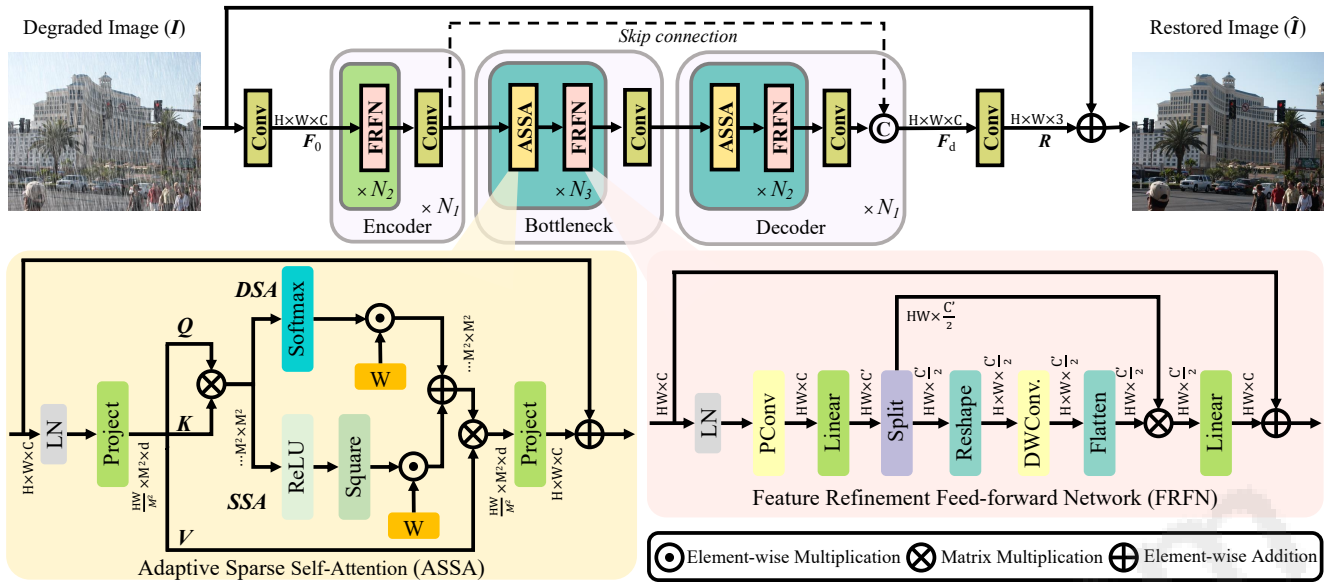


图 2. 我们的自适应稀疏变压器 (AST) 概述。它主要由自适应稀疏自注意力 (ASSA) 和特征细化前馈网络 (FRFN) 组成。LN 表示层归一化，Conv 表示卷积运算。

视觉变压器。由于 Transformer [73] 在自然语言处理领域表现出了卓越的性能，基于 Transformer 的架构被引入计算机视觉社区 [74, 79, 90]。IPT [4] 是基于 Transformer 的图像恢复开创性工作，它通过将输入图像划分为小块并按顺序处理它们来解决计算挑战。尽管如此，普通自注意力的二次复杂性仍然阻碍了 Transformer 应用于高分辨率图像。为了缓解这个问题，restormer [100] 中开发了通道注意力，它沿着通道维度进行注意力计算，降低了计算成本。另一个潜在的补救措施是基于窗口的注意力 [46]，例如 Uformer [82] 采用的方法，它设计了一个局部增强的基于窗口的 Transformer，将局部性引入到 Transformer 架构中。SwinIR [37] 还利用基于窗口的注意力，并引入了一种转移机制以实现更多的跨窗口交互。此外，GRL [35] 结合了窗口注意力和通道注意力形成了一个强大的模型。

尽管这些有效的注意力品种有效地解决了密集计算的问题，并且在消除各种退化方面表现良好，但特征图中不相关的表示或冗余仍然严重阻碍了更好的性能 [8, 114]。为此，DRS-former [8] 在注意力机制中设计了一个 top-k 通道选择算子，以选择信息最丰富的 token 进行计算。类似地，CODE [114] 将特征投影到超像素空间中以减少空间和通道域中的冗余。然而，参数 “k” 的具体选择可能对不同的图像恢复任务敏感。

此外，在超像素空间中执行注意力机制仍然涉及所有可用的标记，可能会在不相关的区域引入不需要的交互。总的来说，我们的 AST 和现有方法之间的主要区别是双重的。一方面，我们引入了一种自适应稀疏自注意力机制，通过选择信息最丰富的交互来减少冗余。采用 square ReLU 激活代替 softmax 层的思想来实现稀疏自注意力。我们没有像之前的作品 [24, 36, 102] 那样设计复杂的组件来放松稀疏性，而是探索一种简单而有效的两分支架构来解决信息丢失问题。通过这种方式，我们的模型充分利用了 SSA 的备用分数，而无需因基于 ReLU 的 SSA 过于稀疏的性质而努力从有限的信息中学习令人满意的表示。另一方面，我们开发了 AST 的另一个关键组件，即特征细化前馈网络。为了缓解特征图中隐藏的冗余信息，它采用了增强和缓解方案，即沿通道维度增强最有用的特征并消除信息量较少的部分。

### 3. 提出的方法

#### 3.1. 整体管线

我们的 AST 流程的概述如图 2 所示，给定图像  $I \in \mathbb{R}^{H \times W \times 3}$ ，AST 首先采用卷积层来生成低级特征表示  $F_0 \in \mathbb{R}^{H \times W \times C}$ ，其中  $H \times W, C$  分别是图像分辨率和通道数。接下来，低级表示  $F_0$  通过  $N_1$  级对称编码器-解码器网络并嵌入到深层

特征  $F_d \in \mathbb{R}^{H \times W \times C}$ 。具体来说，编码器内的每个阶段都由  $N/2$  个基本块和一个用于下采样的卷积层组成。编码器中的基本块包括 FRFN。编码器部分中的特征通过恒等连接与解码器中的特征融合。在这里，我们省略了编码器中标准变换器块内的注意机制，因为其低通滤波器特性[56]可能会阻碍学习所需的局部模式，特别是在早期阶段[89]。在解码器侧，每个阶段由  $N/2$  个基本块和用于上采样的单个卷积层组成。解码器中的基本块包括 ASSA 和 FRFN。此外，受[82]的启发，在解码器之前引入了瓶颈阶段，与解码器共享相同的 Transformer 块以捕获更长的依赖关系。最后，AST 使用卷积层从  $F_d$  生成残差图像  $R \in \mathbb{R}^{H \times W \times 3}$ 。恢复图像由退化图像与残差图像之和得到，即  $\hat{I} = I + R$ 。采用 Charbonnier 损失[3]来训练 AST：

$$\ell(I', \hat{I}) = \sqrt{\|I' - \hat{I}\|^2 + \epsilon^2}, \quad (1)$$

其中  $I'$  指的是真实图像，我们通过实验将  $\epsilon$  设置为  $10^{-3}$ 。

### 3.2. AST 块设计

自适应稀疏自注意力。由于普通 Trans-formers [11,73,82] 考虑特征图中的所有标记，因此计算中可能会涉及许多不相关的区域。这样，它不仅计算无信息区域，而且引入了冗余和不相关的特征，从而降低了模型性能。为了解决这个问题，我们引入了基于平方ReLU的自注意力来过滤掉具有低查询键匹配分数负面影响的特征，这也保证了注意力机制[102] (SSA) 的稀疏特性。同时，考虑到基于ReLU的自注意力[66]的过稀疏性，我们引入了另一个密集自注意力分支

(DSA)，它采用softmax层，以帮助保留关键信息。使用这种两分支方案的关键挑战是如何减少噪声特征和冗余信息，同时尽可能正确地保留信息丰富的特征。为此，ASSA 以自适应方式融合两个分支，即自适应地从分支获取特征并通过网络传播它们。给定一个归一化特征图  $X \in \mathbb{R}^{H \times W \times C}$ ，首先将其划分为大小为  $M$  的非重叠窗口  $\times M$ ，产生来自第  $i$  个窗口的展平表示  $X_i \in \mathbb{R}^{M/2 \times C}$ 。然后我们从  $X$  生成查询  $Q$ 、键  $K$  和值  $V$  的矩阵：

$$Q = XW_Q, K = XW_K, V = XW_V, \quad (2)$$

其中查询  $W_Q$ 、键  $W_K$  和值  $W_V \in \mathbb{R}^{C \times d}$  的线性投影矩阵在所有窗口之间共享。注意力计算可以定义为：

$$A = f(QK^T/\sqrt{d} + B)V, \quad (3)$$

其中  $A$  表示估计注意力； $B$  指的是可学习的相对位置偏差， $f(\cdot)$  是评分函数。值得注意的是，继 [46, 82] 之后，我们并行地对不同的“头”进行权重计算，将其连接起来，然后通过线性投影进行融合。然后我们重新审视标准的密集自注意力机制 (DSA)，在大多数现有作品中采用。它采用 softmax 层，考虑所有查询键对来获得注意力分数：

$$DSA = SoftMax(QK^T/\sqrt{d} + B). \quad (4)$$

由于并非所有查询标记都与键中的相应标记密切相关，因此利用所有相似性对于清晰的图像重建来说是无效的。直观地说，开发稀疏自注意力 (SSA) 机制来选择令牌之间有用的交互可以增强特征聚合。为了实现注意力的稀疏性，基于 squaredReLU 的层似乎是一个可行的解决方案。它消除了负分的相似性，并向前传播最有用的信息流：

$$SSA = ReLU^2(QK^T/\sqrt{d} + B). \quad (5)$$

请注意，基于 ReLU 的 SSA 会触发信息丢失，通常需要额外的技术来放松稀疏性，这违背了使用 SSA 优于 DSA 的动机。简单地应用基于 ReLU 的 SSA 会给管道带来过度稀疏性，即学习到的特征表示包含的信息不足以用于后续过程。相反，使用基于 softmax 的 DSA 会无意中在不相关区域引入噪声交互，从而给恢复高质量图像带来挑战。因此，我们不是偏爱一种范式，而是提出一种双分支自注意力机制作为基本组成部分，具有自适应注意力分数，以利用两种范式的优势。式中的注意力矩阵。(3) 可以进一步更新为：

$$A = (w_1 * SSA + w_2 * DSA)V, \quad (6)$$

其中  $w_1, w_2 \in \mathbb{R}^1$  是用于自适应调制二分支的两个归一化权重， $*$  表示乘法运算。更具体地说，它可以通过以下方式计算：

$$w_n = \frac{e^{a_n}}{\sum_{i=1}^N e^{a_i}}, n = \{1, 2\} \quad (7)$$

其中  $\{a_1, a_2\}$  是可学习参数，以两个分支中的 1 个开头。这种设计确保了更好的



表 1.SPAD [78] 去除雨纹的定量比较。

SPAD [78]		
Method	PSNR ↑	SSIM ↑
DDN [13]	36.16	0.9463
RESCAN [33]	38.11	0.9797
PReNet [63]	40.16	0.9816
RCDNet [75]	43.36	0.9831
SPDNet [94]	43.55	0.9875
SPAIR [57]	44.10	0.9872
DualGCN [14]	44.18	0.9902
SEIDNet [39]	44.96	0.9911
MPRNet [99]	45.00	0.9897
Fu <i>et al.</i> [15]	45.03	0.9907
Restormer [100]	46.25	0.9911
SCD-Former [19]	46.89	0.9941
IDT [88]	47.34	0.9929
Uformer [82]	47.84	0.9925
DRSformer [8]	48.53	0.9924
AST-B (Ours)	<u>49.51</u>	<u>0.9942</u>
AST-B+ (Ours)	<b>49.72</b>	<b>0.9944</b>

表 2. 针对雨滴去除的 AGAN 数据 [58] 的模型效率分析。

AGAN-Data [58]		
Method	PSNR ↑	SSIM ↑
Eigen's [12]	21.31	0.757
Pix2pix [26]	27.20	0.836
Uformer [82]	29.42	0.906
WeatherDiff <sub>128</sub> [54]	29.66	0.923
TransWeather [72]	30.17	0.916
WeatherDiff <sub>64</sub> [54]	30.71	0.931
TKL&MR [7]	30.99	0.927
All-in-One [32]	31.12	0.927
DuRN [44]	31.24	0.926
CCN [61]	31.34	0.929
Quan's [62]	31.37	0.918
AttenGAN [58]	31.59	0.917
IDT [88]	31.87	0.931
MAXIM-2S [71]	31.87	<u>0.935</u>
AWRCF [93]	31.93	0.931
AST-B (Ours)	<u>32.32</u>	<u>0.935</u>
AST-B+ (Ours)	<b>32.45</b>	<b>0.937</b>

表 3. Dense-Haze [1] 真实除雾的定量比较。

Dense-Haze [1]		
Method	PSNR ↑	SSIM ↑
RIDCP[87]	8.09	0.42
DCP [20]	10.06	0.39
SGID [2]	13.09	0.52
D4[91]	13.12	0.53
AOD-Net [30]	13.14	0.41
GridDehazeNet [43]	13.31	0.37
DA-Dehaze [65]	13.98	0.37
FFA-Net [59]	14.39	0.45
Uformer [82]	15.22	0.43
Restormer[100]	15.78	0.55
AECR-Net [86]	15.80	0.47
Fourmer[116]	15.95	0.49
DehazeFormer-S [69]	16.29	0.51
DeHamer [18]	16.62	<u>0.56</u>
MB-TaylorFormer-B [60]	16.66	<u>0.56</u>
AST-B (Ours)	<u>17.12</u>	0.55
AST-B+ (Ours)	<b>17.27</b>	<b>0.57</b>

在可以过滤掉的不相关区域的噪声交互和可以利用足够的信息特征之间进行权衡。换句话说，该模型能够控制特定任务的输入标记的稀疏程度。特征细化前馈网络。常规 FFN [73] 单独处理每个像素位置的信息，这在通过自注意力机制改进特征表示方面发挥着至关重要的作用。因此，设计一个有效的 FFN 来增强特征，从而提高潜在的高质量图像恢复至关重要。当采用ASSA作为基本组件来去除空间域中的冗余信息时，通道中仍然存在冗余。为了克服这个问题，我们开发了FRFN，以增强和简化范式执行特征转换。具体来说，我们通过引入 PConv 操作 [5] 来增强特征中的信息元素，并引入门机制来减少冗余信息的处理负担，从而构建 FRFN。FRFN可以表示为：

$$\begin{aligned}
\hat{X}' &= GELU(W_1 PConv(\hat{X})), [\hat{X}'_1, \hat{X}'_2] = \hat{X}', \\
\hat{X}'_r &= \hat{X}'_1 \otimes F(DWConv(R(\hat{X}'_2))), \\
\hat{X}'_{out} &= GELU(W_2 \hat{X}'_r),
\end{aligned} \quad (8)$$

其中W<sub>1</sub>和W<sub>2</sub>表示线性投影；[ , ]指通道方式切片操作；R(·)和F(·)说明了将序列输入转换为2D特征图以及相反的Re-shape和Flatten操作，这对于将局部性引入架构至关重要[34]；PConv(·)和DWConv(·)分别指部分卷积[5]和深度卷积[22]操作；⊗表示矩阵乘法。总体而言，FRFN能够通过从矩阵中提取那些代表性特征来增强特征表示。

简化信息流，同时简化冗余信息。它还为模型提供了清除通道维度上无信息特征的机会。

#### 4. 实验

在本节中，我们评估 AST 在各种图像恢复任务上的性能，例如雨纹、雾霾和雨滴去除。还进行了消融研究以证明所提出模块的有效性。

##### 4.1. 实验设置

实施细节。在默认设置下，AST对于编码器和解码器部分都包含N<sub>1</sub>=4个阶段，并且在瓶颈中发展一个阶段。我们通过改变嵌入维度C和Transformer块（编码器和解码器共享相同的N<sub>2</sub>块，而瓶颈包括N<sub>3</sub>块）来构建普通模型的两个变体，称为AST-T和AST-B。对于AST-T，我们将C设置为16，N<sub>2</sub>和N<sub>3</sub>设置为[2,2,2,2]和2，而对于AST-B，我们将C设置为32，N<sub>2</sub>和N<sub>3</sub>设置为[1,2,8,8]和2。默认的分割窗口大小为8，并且它们在Transformer块中共享每个头的相同尺寸，遵循[82]中的方法。我们采用具有默认设置的AdamW优化器[47]来训练我们的模型。学习率最初设置为0.0002，然后使用余弦衰减策略逐渐降低到0.000001[48]。我们随机使用旋转和翻转操作策略进行增强。使用渐进式学习策略来节省时间，类似于[70, 100]。评估指标。为了评估恢复性能，我们采用PSNR和SSIM指标[81]。此外，NIQE[52]被用作非参考指标。值得注意的是，对于去雨，按照现有的工作[75, 82]，PSNR/SSIM分数在Y通道上计算

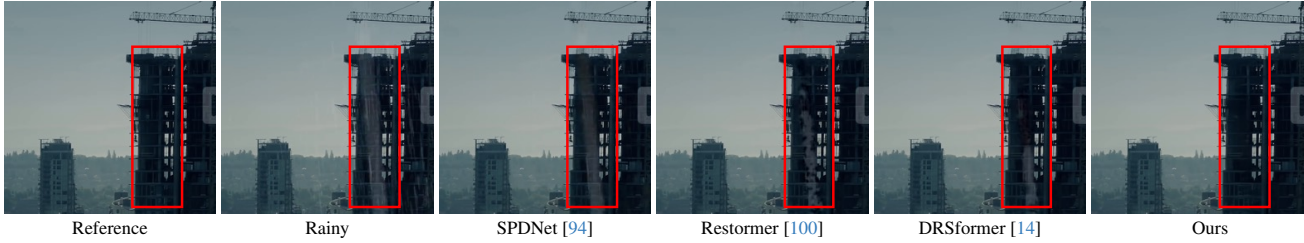


Figure 3. Qualitative comparisons on SPAD [78] for real rain removal.

YCbCr 空间。当使用几何自集成策略[38]时，我们用“+”符号表示该方法。表中的最佳和第二好分数已突出显示并带有下划线。

#### 4.2.雨纹去除

我们在 SPAD 基准 [78] 上进行了去雨实验，并将 AST 的性能与 5 种最先进的算法进行了比较，包括 DDN [13]、RES-CAN [33]、PReNet [63]、RCDNet [75]、SPDNet [94]、SPAIR [57]、DualGCN [14]、SEIDNet [39]、MPRNet [99]、Fu 等人。[15]、Restormer [100]、SCD-Former [19]、IDT [88]、Uformer [82] 和 DRSformer [8]。在选项卡中。如图 1 所示，与之前最好的基于 CNN 的方法 Fu 等人相比，AST-B 在 PSNR 指标方面实现了 4.48 dB 的增益。[15]，与之前最好的基于 Transformer 的模型 DRSformer [8] 相比，提高了 0.98 dB。我们在图 2 中进行了视觉比较。3，其中 AST-B 可以更成功地去除真实的雨条纹，同时保留结构内容。

#### 4.3.去除雨滴

我们在 AGAN-Data [58] 基准上进行雨滴去除实验，并将我们的 AST 与各种最先进的去雨滴方法进行比较，包括 Eigen 的 [12]、Pix2pix [26]、Uformer [82]、WeatherDiff 128 [54]、TransWeather [72]、WeatherDiff 64 [54]、TKL&MR [7]、All-in-One [32]、DuRN [44]、CCN [61]、Quan's [62]、AttenGAN [58]，IDT [88]，MAXIM-2S [71] 和 AWRCP [93]。在选项卡中。2、AST-B 的 PSNR 性能优于之前的最佳方法 AWRCP [93] 0.39 dB，并且优于基于并发扩散的方法 WeatherDiff 128 [54] 2.66 dB。

#### 4.4.真正的除雾

我们在Dense-Haze基准[1]上进行真实去雾评估，并将 AST 与十五种最先进的去雾作品进行比较，包括 RIDCP[87]、DCP[20]、SGID[2]、D4[91]，AOD-Net [30]，Grid-DehazeNet [43]，DA-Dehaze [65]，FFA-Net [59]，Uformer [82]，Restormer [100]，AECR-Net [86]，Fourmer [116]、DehazeFormer-S [69]、DeHamer [18] 和 MB-TaylorForm [60]。在选项卡中。3、AST-B 在所考虑的 PSNR 指标中获得最佳值

表 4. 不同自注意力机制的消融研究。

Models	Swin SA [37]	Top-k SA [8]	Condensed SA [114]	ASSA Ours
Params	6.65	6.67	6.07	6.65
FLOPs	13.32	13.59	11.46	13.35
PSNR	44.47	44.67	44.94	45.43

表 5. 与标准自注意力机制和相应稀疏版本的比较。

	Method	PSNR
(1)	Standard Local Self-Attention [82]	45.09
	Sparse Local Self-Attention	44.58
(2)	Standard Channel Self-Attention [100]	44.91
	Sparse Channel Self-Attention	44.45

最先进的方法。与之前最好的基于 CNN 的方法 ARCT-Net [86] 相比，我们的 AST-B 的 PSNR 增益为 1.37 dB。此外，与最近基于 Transformer 的方法相比，我们的 AST-B 至少实现了 0.46 dB 的改进 [60,69,116]。

#### 4.5.分析与讨论

探索最有用的信息并减少 Transformer 架构中的冗余，可以为各种图像恢复任务提供良好的结果。在这里，我们对 AST 进行了更深入的分析，并说明了所提出模块的有效性。对于消融研究，我们在 SPAD [78] 数据集上训练去雨模型 AST-T。为了公平比较，所有模型都在  $128 \times 128$  图像块上训练 10 个 epoch，并计算输入大小为  $256 \times 256$  的 FLOP。亚萨。为了研究 ASSA 组件的有效性，我们将其替换为现有的有效注意力机制：（1）Swin 自注意力

（SwinSA）[37]，（2）Top-k 自注意力（Top-k SA）[8]，以及（3）压缩自注意力（Condensed SA）[114]。我们在表中显示了定量结果。4. 与 Swin SA 相比，ASSA 在 PSNR 方面提供了 0.96 dB 的有利增益，但复杂性略有增加（0.03G Flops）。此外，与提出清除令牌和冗余信息之间的噪声交互的紧密相关方法相比，我们的 ASSA 设计比 Top-k SA [8] 获得了 0.76 dB 的性能改进，比 Condensed SA 获得了 0.49 dB 的性能改进[114]。自适应架构设计的有效性。所提出的自适应架构设计用于减少

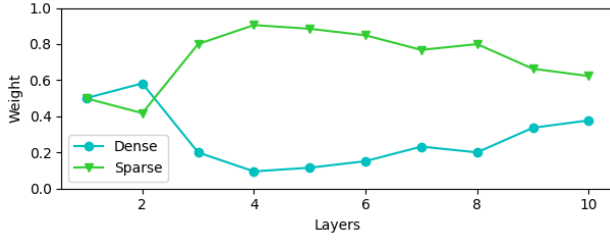


Figure 4. Learned weights for sparse and dense branches.

噪声代表特征和冗余信息，同时适当保留信息特征。为了研究配备基于 ReLU 的稀疏注意力的图像恢复模型是否会遇到 NLP 领域类似的性能下降现象，我们首先基于两种主流范式构建了两个版本的稀疏自注意力机制：（1）局部自注意力[82]和（2）通道自注意力[100]。如表所示。5、直接用基于 ReLU 的稀疏自注意力替换标准的基于 softmax 的密集自注意力会导致 LocalSelf-Attention 和 Channel Self-Attention 的性能分别显著下降 0.51 dB 和 0.46 dB。由于基于 ReLU 的稀疏自注意力（SSA）过于稀疏的问题，导致性能下降，因此我们计算注意力层的熵，类似于[16]，以衡量注意力集中度。具体来说，注意力熵定义为：

$$Entropy_{Att} = -\frac{1}{H} \sum_h \frac{1}{L} \sum_{ij} Att_{h,lij} * \log Att_{h,lij}, \quad (9)$$

其中  $Att_{h,lij}$  表示查询 token  $i$  对层  $l \in L$  的头  $h \in H$  的关键 token  $j$  的注意力得分。较低的熵表示平均注意力趋于集中，而较高的熵表示注意力趋于集中。更加分散。如选项卡所示。如图 6 所示，基于 softmax 的密集自注意力（DSA）得分最高，而 SSA 得分最低。换句话说，DSA 更均匀地从源标记中提取特征，这可能会引入不相关区域的噪声交互。SSA 专注于一些太稀疏而无法涵盖必要关系的标记。相反，我们的方法达到了一个折衷方案，即可以充分探索信息上下文，同时忽略冗余特征，从而带来明显的性能提升。然后，我们展示了使用所提出的自适应两分支架构设计的必要性和优越性，即标准密集自注意力和相应的稀疏版本，通过对表中的训练模型变体进行实验来缓解挑战。7. 与配备 DSA 的模型相比，直接应用 SSA 的性能并不令人满意。特别是，与自适应激活相比，例如 ACON 和 Meta-

Table 6. Entropy analysis of different self-attention mechanisms.

Structure	DSA	SSA	ASSA (Ours)
Entropy	3.733	1.543	3.134
PSNR	45.09	44.58	45.43

Table 7. Ablation study on various activation choices in the self-attention mechanism.

Type	Dense	Sparse		Adaptive		
Variety	softmax	ReLU <sup>2</sup>	StarReLU [96]	ACON [51]	Meta-ACON [51]	ASSA(Ours)
PSNR	45.09	44.58	45.30	43.23	43.67	45.43
$\Delta$	-0.34	-0.85	-0.13	-2.20	-1.76	-

表 8. 特征细化前馈网络替代方案的消融研究。

Models	FFN [11]	DFN [34]	GDFN [100]	LeFF [82]	FRFN Ours
Params	7.77	7.92	6.49	7.92	6.65
FLOPs	15.25	16.20	13.19	16.30	13.35
PSNR	44.13	43.46	44.66	44.77	45.43

ACON [51]，我们的 ASSA 仍然可以实现最大的性能增益（45.43 dB）。我们最终在图 4 中可视化每个 SSA 和 DSA 分支的学习权重。正如预期的那样，模型首先平等地对待两个分支，以确保足够的性能信息，并且随着层的加深以更好地聚合特征，更加关注稀疏分支。我们注意到，学习到的权重起到了软选择的作用，从而使模型能够适应两个分支的影响。FRFN 的有效性。特征图通常具有高通道维度，尤其是在深层，并且并非所有特征通道都包含恢复清晰图像的关键信息。简单地对所有通道应用相同的特征转换可能会导致多余的冗余信息。在实践中，增强信息渠道以进一步推进特征表示学习是一项艰巨的任务。为了证明我们的 FRFN 的效果，我们首先将其与四个变体进行比较，包括（1）普通前馈网络（FFN）[11]，（2）配备深度卷积的前馈网络（DFN）[34]，（3）门控 Dconv 前馈网络（GDFN）[100]，以及（4）局部增强前馈网络（LeFF）。表中列出了定量比较。8. 我们的 FRFN 实现了最好的 PSNR 值，参数和 FLOPs 稍微多一些。换句话说，FRFN 可以选择更有用的信息并减少冗余特征，从而比其他考虑的设计更好地与我们提出的 ASSA 设计配合。尽管 GDFN [100] 利用像我们这样的门控机制来控制信息流，但 FRFN 执行微妙的增强和简化特征转换，以帮助选择信息最多的特征。结果，FRFN 比 GDFN 实现了 0.77 dB 的 PSNR 增益。我们还在表 1 中进行了消融研究。9 调查 FRFN 的影响。与基线模型相比 (a)



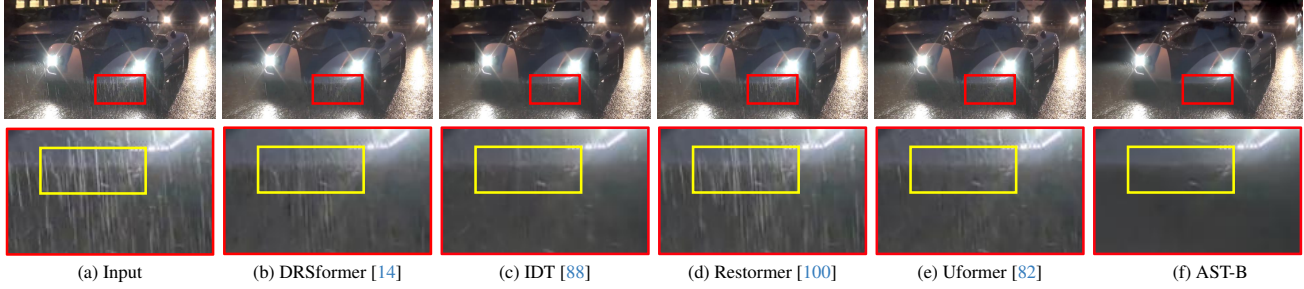


图 5. 实际除雨的互联网数据 [78] 的定性比较。

表 9. 用于图像去雨的 SPAD 上的 FRFN 消融研究。

	Enhance	Ease	DWConv	Params	FLOPs	PSNR	Latency
(a)			✓	8.06	17.16	45.02	15.55
(b)	✓		✓	8.26	17.55	45.16	17.03
(c)		✓	✓	6.45	12.96	45.30	19.01
(d)	✓	✓	✓	6.65	13.35	45.43	19.75

Table 10. Results of no-reference assessment metric NIQE for deraining task under the real-world scenario.

Methods	Input	Uformer [82]	Restormer [100]	IDT [88]	DRSformer [8]	AST-B Ours
NIQE ↓	6.274	5.749	6.162	6.079	5.994	5.493

引入了具有深度卷积层的局部性，继现有工作[82]之后，我们的 FRFN (d) 通过设计增强和简化方案提供了性能优势 (0.41 dB)。具体来说，使用 PConv 算子 [5] 增强有价值的信息，并使用门控机制缓解隐藏在特征图中的冗余，分别产生 0.28 dB 和 0.14 dB 的改进。PConv 仅对部分通道进行卷积，可以将其视为稀疏操作以选择有用的通道。通过这种方式，引导网络集中于重要特征，增强提取信息特征的能力。这些结果证明了我们 FRFN 的设计贡献以及增强和简化方案。感知质量评估。继[8]之后，我们从互联网数据[78]基准中随机选择了20张真实世界的雨天图像来进行评估。如选项卡中所示。如图 10 所示，AST-B 实现了最低的 NIQE 值，这意味着在现实环境下比考虑的方法具有更好的感知质量。此外，如图5所示的定性比较，AST-B消除了雨纹退化并产生了视觉上忠实的结果，这表明它有能力处理看不见的真实退化。

## 5. 结论

这项工作的目标是通过自适应学习信息最丰富的表示并消除特征中的噪声信息，从降级版本中恢复清晰的图像。虽然我们引入了 NLP 中基于 ReLU 的稀疏自注意力 (SSA) 来消除不相关标记之间的噪声交互，而不是直接使用它



图 6. 错误修复的示例。AST 的典型故障可以在严重退化的现实场景中找到。

作为一个基本组成部分，我们的目标是首先防止由于基于 ReLU 的 SSA 的熵较小而导致的信息丢失。为了有效地实现这一目标，我们探索了一种自适应架构设计，确保必要的信息在另一个密集分支的帮助下向前流动。此外，我们提出了一种 FRFN 来通过增强和简化方案执行特征转换，其中可以学习判别性特征表示以提高高质量图像重建。我们的 AST 优于采用选择操作（例如，Top-K 选择和稀疏通道 SA）或将特征投影到超像素空间（例如，Condensed SA）以减轻冗余的相关基线，而最终，它实现了良好的效果多项降解去除任务的结果。

局限性。未来的工作可以集中在当前的局限性（例如，为具有各种退化的低质量图像开发统一模型），以及该特定于任务的模型提供的机会（例如，注入先验，如暗通道先验）用于图像去雾和 Retinexmodel 之前用于消除低光条件）。图 6 展示了一个失败案例，其中 AST 难以处理严重退化的场景。

致谢本文工作得到了天津市自然科学基金项目 (No.20JCJC00020)、国家自然科学基金项目 (Nos. U22B2049, 62302240)、中央高校基本科研业务费专项资金、南开超级计算中心的资助大学 (NKSC)。



## References

- [1] Codruta O Ancuti, Cosmin Ancuti 和 RaduTimofte. 浓雾: 浓雾和无雾图像的图像去雾基准. ICIP, 2019. 5, 6[2] 白浩然, 潘金山, 项新光, 唐金辉. 利用渐进式特征融合的自引导图像去雾. 提示, 31: 1217-1229, 2022. 5, 6[3] 皮埃尔·夏博尼耶, 劳尔·布兰克-费罗, 吉尔斯·奥伯特和米歇尔·巴洛. 用于计算成像的两种确定性半二次正则化算法. ICIP, 1994.4[4] 陈汉庭, 王云鹤, 郭天宇, 徐昌, 邓一平, 刘振华, 马思伟, 徐春晶, 徐超, 高文. 预先训练的图像处理变压器. InCVPR, 2021. 3[5] Jierun Chen, Shiu-hong Kao, Hao He, Weipeng Zhuo, SongWen, Chul-Ho Lee 和 S-H Gary Chan. 跑步, 不要走路: 追求更高的失败率以获得更快的神经网络. 在 CVPR, 2023 年. 5, 8[6] 陈良宇, 褚晓杰, 张翔宇, 孙健. 图像恢复的简单基线. ECCV, 2022. 1[7] 陈伟廷, 黄志凯, 蔡正哲, 杨浩翔, 丁建军和郭赛彦. 通过两阶段知识学习和多对比正则化学习多重不利天气去除: 走向统一模型. CVPR, 2022. 5, 6[8] 陈翔, 李浩, 李明强, 潘金山. 学习稀疏变换网络以实现有效的图像去雨. CVPR, 2023 年. 2, 3, 5, 6, 8[9] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-WonJung 和 Sung-Jea Ko. 重新思考单图像去模糊中从粗到细的方法. ICCV, 2021 年. 2[10] 邓鑫和皮尔·路易吉·德拉戈蒂 (Pier Luigi Dragotti). 用于多模态图像恢复和融合的深度卷积神经网络. TPAMI, 43(10):3333-3348, 2021. 2[11] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, 穆斯塔法·德哈尼 (Mostafa Dehghani)、马蒂亚斯·明德勒 (Matthias Minderer)、格奥尔格·海戈尔德 (Georg Heigold)、西尔万·盖利 (Sylvain Gelly)、雅各布·乌什科雷特 (Jakob Uszkoreit) 和尼尔·霍尔斯基 (Neil Houlsby). 一张图像相当于 16x16 个单词: 用于大规模图像识别的 Transformers. ICLR, 2021 年. 2, 4, 7[12] David Eigen, Dilip Krishnan 和 Rob Fergus. 恢复透过被泥土或雨水覆盖的窗户拍摄的图像. 载于 ICCV, 2013. 5, 6[13] 付雪阳, 黄家斌, 曾德禄, 黄跃, 丁兴浩, 约翰·佩斯利. 通过深度细节网络从单幅图像中去除雨水. CVPR, 2017. 5, 6[14] 付雪阳, 齐齐, 查正军, 朱玉瑞, 丁兴浩. 通过双图卷积网络去除雨纹. AAAI, 2021. 2, 5, 6, 8[15] 付雪阳, 肖杰, 朱玉瑞, 刘爱萍, 吴峰, 查正军. 使用超图卷积网络进行连续图像去雨. TPAMI, 45 (8) : 9534-9551, 2023. 5, 6[16] 哈米德雷扎·加德尔和克里斯托夫·蒙兹. 神经机器翻译中的attention关注什么? arXiv预印本 arXiv:1710.03348, 2017. 7
- [17] Shuhang Gu, Yawei Li, Luc Van Gool, and Radu Timofte. 用于快速图像去噪的自引导网络. 在ICCV, 2019. 2[18] 郭春乐, 严启新, Saeed Anwar, 丛润民, 任文琪, 李崇义. 具有传输感知 3D 位置嵌入的图像去雾变压器. 在 CVPR, 2022 年. 2, 5, 6 [19] 郭云, 肖雪耀, 常毅, 邓淑敏, 严鲁新. 从天空到地面: 真正除雨的大规模基准和简单基线. InICCV, 2023. 5, 6[20] 何凯明, 孙健, 唐晓鸥. 使用暗通道先验的单图像去雾. TPAMI, 33(12):2341-2353, 2010. 2, 5, 6[21] 何凯明, 张翔宇, 任少清, 孙健. 图像识别中的深度残差学习. 在 CVPR, 2016 年. 2[22] Andrew G Howard, 朱梦龙, 陈波, DmitryKalenichenko, 王伟军, Tobias Weyand, Marco And-dreetto 和 Hartwig Adam. Mobilenets: 用于移动视觉应用的高效卷积神经网络. arXivpreprint arXiv:1704.04861, 2017. 5[23] Jiri Hron, Yasaman Bahri, Jascha Sohl-Dickstein 和 Ro-man Novak. 无限注意力: 用于深度注意力网络的 Nngp 和 ntk. ICML, 2020. 2[24]Weizhe Hua, Zihang Dai, Hanxiao Liu, and Quoc Le.线性时间内的变压器质量. ICML, 2022 年. 3[25] 高黄, 庄刘, Laurens Van Der Maaten 和 Kil-ian Q Weinberger. 密集连接的卷积网络. CVPR, 2017. 2[26] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou 和 Alexei Aefros. 使用条件对抗网络进行图像到图像的翻译. CVPR, 2017. 5, 6[27] Kevin Jarrett, Koray Kavukcuoglu, Marc' Aurelio Ranzato 和 Yann LeCun. 物体识别的最佳多级架构是什么? ICCV, 2009. 2[28]贾国立, 杨巨峰. S 2-ver: 半监督视觉情感识别. ECCV, 2022. 2[29]Orest Kupyn, Tetiana Martyniuk, Junru Wu, 和Zhangyang Wang. Deblurgan-v2: 去模糊 (数量级) 更快更好. ICCV, 2019. 2[30]李博一, 彭秀莲, 王张扬, 徐吉正, 冯丹. Aod-net: 一体化除雾网络. 在ICCV, 2017. 5, 6[31] 李崇义, 郭春乐, 韩令浩, 蒋军, 程明明, 顾金伟, 陈昌来. 使用深度学习的低光图像和视频增强: 调查. TPAMI, 44(12): 9396-9416, 2022. 2[32]李若腾, Robby T. Tan 和 Loong-Fah Cheong. 总而言之, 使用建筑搜索来消除恶劣天气. InCVPR, 2020. 5, 6[33] 李夏, 吴建龙, 林周辰, 刘红, 查宏斌. 用于单图像去雨的循环挤压和激励上下文聚合网络. ECCV, 2018. 5, 6[34]李亚伟, 张凯, 曹杰章, Radu Timofte, LucVan Gool. Localvit: 为视觉变革者带来本地化. arXiv 预印本 arXiv:2104.05707, 2021. 5, 7

- [35] 李亚伟, 范雨辰, 向晓宇, Demandolx, Rakesh Ranjan, Radu Timofte, Luc Gool. 用于图像恢复的图像层次结构的高效且明确的建模. CVPR, 2023 年. 3[36] 李志远、Srinadh Bhojanapalli、Manzil Zaheer、SashankReddi 和 Sanjiv Kumar. 使用尺度不变架构对神经网络进行鲁棒训练. ICML, 2022.3[37]梁景云, 曹杰章, 孙国磊, 张凯, LucVan Gool, Radu Timofte. Swinir: 使用 swin 变压器进行图像恢复. ICCV 研讨会, 2021. 1, 3, 6[38] Bee Lim、Sanghyun Son、Heewon Kim、Seungjun Nah 和 Kyoung Mu Lee. 用于单图像超分辨率的增强型深度残差网络. CVPR 研讨会, 2017. 6[39] 林迪, 王鑫, 沉佳, 张仁杰, 刘若男, 王妙辉, 谢五元, 郭庆, 李平. 图像除雨的生成状态估计和信息解耦. 见 NeurIPS, 2022. 2, 5, 6[40] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang. 用于图像恢复的非本地循环网络. 见 NeurIPS, 2018. 2[41] 刘娜, 李伟, 王银健, 陶然, 杜谦, Jo-celyn Chanussot. 高光谱图像恢复综述: 从低秩张量逼近的角度看. SCIS, 66(4):140302, 2023. 2[42] 刘鑫, 杨巨峰. 用于对应修剪的渐进邻居一致性挖掘. CVPR, 2023. 2[43]Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen.GridDehazeNet: 基于注意力的多尺度网络进行图像去雾. 载于ICCV, 2019. 5, 6[44] 刘星, 菅沼正德, 孙准, 冈谷隆之. 双残差网络利用成对操作的潜力进行图像恢复. CVPR, 2019. 2,5, 6[45] 刘鑫, 肖国宝, 陈日清, 马佳一. Pgfnet: 用于二视图对应学习的偏好引导过滤网络. TIP, 32:1367–1378, 2023. 2[46] 刘泽, 林雨桐, 曹悦, 胡涵, 魏逸轩, 张正, 林书豪, 郭百宁. Swin 变压器: 使用移动窗口的分层视觉变压器. InICCV, 2021. 3, 4[47] Ilya Loshchilov 和 Frank Hutter. 解耦权重衰减正则化. arXiv 预印本 arXiv:1711.05101, 2017. 5[48] Ilya Loshchilov 和 Frank Hutter. Sgdr: 带有热重启的随机梯度下降. ICLR, 2017. 5[49] Andreas Lugmayr、Martin Danelljan、Radu Timofte、Kang-wook Kim、Younggeun Kim、Jae-young Lee、Zechao Li、Jinshan Pan、Dongseok Shim、Ki-Ung Song、Jinhui Tang、Cong Wang和赵志豪. 2022 年所有关于学习超分辨率空间的挑战. 2022 年 CVPR 研讨会. 2[50]罗方舟, 吴晓林, 郭彦辉. 用于参数图像恢复问题的功能神经网络. In NeurIPS, 2021. 2[51] 马宁宁, 张翔宇, 刘明, 孙健. 是否激活: 学习定制激活. 在 CVPR, 2021 年. 7
- [52] Anish Mittal、Rajiv Soundararajan 和 Alan C Bovik. 制作“完全盲”图像质量分析器. SPL,20(3):209–212, 2012. 5[53] 牛奔, 温伟雷, 任文琪, 张向德, 杨连平, 王淑珍, 张凯浩, 曹晓春, 沉海峰. 通过整体注意力网络的单图像超分辨率. ECCV, 2020 年. 2[54] Ozan Özdenizci 和 Robert Legenstein. 使用基于补丁的去噪扩散模型在恶劣天气条件下恢复视力. TPAMI, 45(8):10346–10357, 2023. 2, 5,6[55] 潘金山, 孙德清, Hanspeter Pfister, 杨明轩. 通过暗通道先验去模糊图像. TPAMI, 40(10):2315–2328, 2017. 1[56] Namuk Park 和 Songkuk Kim. 视觉转换器如何工作? ICLR, 2022 年. 4[57] Kuldeep Purohit、Maitreya Suin、AN Rajagopalan 和 Vishnu Naresh Boddeti. 使用失真引导网络进行空间自适应图像恢复. ICCV, 2021. 5,6[58] 钱瑞, Robby T Tan, 杨文瀚, 苏家俊, 刘家英. 专注的生成对抗网络, 用于从单个图像中去除雨滴. CVPR, 2018. 5, 6[59] 徐勤, 王志林, 白远超, 谢晓东, 贾慧珠. Pfa-net: 用于单图像去雾的特征融合注意网络. AAAI, 2020. 5, 6[60] 邱宇伟, 张凯浩, 王晨曦, 罗文瀚, 李洪东, 金志. Mb-taylorformer: 通过泰勒公式扩展的多分支高效变压器, 用于图像去雾. 载于ICCV, 2023. 2, 5, 6[61] 全锐杰, 于欣, 梁远志, 杨易. 一次性去除雨滴和雨痕. CVPR, 2021. 5,6[62] 全宇辉, 邓士杰, 陈一新, 季辉. 深度学习通过窗户看到雨滴. InICCV, 2019. 5, 6[63] 任东伟, 左王猛, 胡清华, 朱鹏飞, 孟德宇. 渐进式图像去雨网络: 更好、更简单的基线. CVPR, 2019 年. 2,5,6[64] Olaf Ronneberger、Philipp Fischer 和 Thomas Brox. U-net: 用于生物医学图像分割的卷积网络. 载于MICCAI, 2015. 2[65]邵元杰, 李乐仁汉, 任文琪, 高长新, 桑农. 图像去雾的域适应. InCVPR, 2020. 5, 6[66] 沉凯, 郭俊良, 谭旭, 唐思良, 王锐, 边江. Transformer.arXiv 中关于 relu 和 softmax 的研究 预印本 arXiv:2302.06461, 2023. 2, 4[67] David R So, Wojciech Mańke, Hanxiao Liu, Zihang Dai, Noam Shazeer, 和 Quoc V Le. 入门: 寻找用于语言建模的高效转换器. arXiv preprint arXiv:2109.08668, 2021. 2[68] 宋西斌, 周定福, 李伟, 戴玉超, 沉哲伦, 张良军, 李洪东. Tusr-net: 通过自正则化和像素关注的双特征进行三重展开单图像去雾. 提示, 32: 1231–1244, 2023. 2

[69] 宋宇达, 何竹清, 钱慧, 杜. 用于单图像去雾的视觉转换器. TIP, 32:1927–1941, 2023. 5, 6[70] Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: 用于快速图像去模糊的带状变压器. ECCV, 2022. 5[71] 屠正中, 侯赛因·塔勒比, 张涵, 杨峰, 佩曼·米兰法尔, 艾伦·博维克, 李银晓. Maxim: 用于图像处理的多轴 mlp. CVPR, 2022. 2, 5, 6[72] Jeya Maria Jose Valanarasu, Rajeev Yasarla 和 Vishal M. Patel. Transweather: 基于 Transformer 恢复因恶劣天气条件而损坏的图像. 在 CVPR, 2022 年. 5, 6[73] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser 和 Illia Polosukhin. 您所需要的就是关注. 在 NeurIPS, 2017 年. 1, 3, 4, 5[74] 王聪, 潘金山, 王伟, 董江鑫, 王梦竹, 鞠亚坤, 陈俊阳, 吴晓明. Promptrestorer: 一种具有退化感知的提示图像恢复方法. 见 NeurIPS, 2023. 3[75] 王宏, 谢齐, 赵倩, 孟德宇. 一种模型驱动的单图像除雨深度神经网络. CVPR, 2020. 5, 6[76] Lijuan Wang, Guolijia, Ning Jiang, Haiying Wu, and Jufeng Yang. 轻松: 通过情感模糊敏感的合作网络进行强大的面部表情识别. In ACM MM, 2022. 2[77] 王皮超, 王雪, 王帆, 林明, 常树宁, 李浩, 金荣. Kvt: 用于增强视觉变压器的 k-nn 注意力. ECCV, 2022. 1[78] 王天宇, 杨欣, 徐科, 陈绍哲, 张强, 刘伟生. 使用高质量真实降雨数据集进行空间注意力单图像除雨. 2019 年 CVPR. 2, 5, 6, 8[79] 王文海, 谢恩泽, 李翔, 范登平, 宋凯涛, 丁亮, 路童, 罗平, 邵凌. Pvt2: 使用金字塔视觉变压器改进基线. CVMJ, 8 (3) : 415–424, 2022. 3[80] Yingqian Wang, Longguang Wang, Zhengyu Liang, Jun-gang Yang, Radu Timofte 和 Yulan Guo. 2023 年光场图像超分辨率挑战: 数据集、方法和结果. arXiv 预印本 arXiv:2304.10415, 2023. 2[81] Zhou Wang, Alan C Bovik, Hamid R Sheikh 和 Eero P Simoncelli. 图像质量评估: 从错误可见性到结构相似性. TIP, 13(4):600–612, 2004. 5[82] 王振东, 村晓东, 包建民, 周文刚, 刘建壮, 李厚强. Uformer: 用于图像恢复的通用 U 形变压器. 在 CVPR, 2022 年. 1, 2, 3, 4, 5, 6, 7, 8[83] 温昌松, 贾国立, 杨巨峰. Dip: 用于讽刺检测的双重不一致感知网络. In CVPR, 2023. 2[84] 文昌松, 张鑫, 姚星旭, 杨巨峰. 序数标签分布学习. 在 ICCV 中, 2023 年. 2

[85] 米切尔·沃茨曼、Jaehoon Lee、贾斯汀·吉尔默和西蒙·科恩布里斯. 在视觉变换器中用 relu 替换 softmax. arXiv 预印本 arXiv:2309.08586, 2023. 2[86] 吴海燕, 曲艳云, 林少辉, 周健, 乔睿智, 张志中, 谢媛, 马丽壮. 紧凑单图像去雾的对比学习. CVPR, 2021. 5, 6[87] 吴瑞琪, 段正鹏, 郭春乐, 柴智, 李崇义. Ridcp: 通过高质量码本先验恢复真实图像去雾. CVPR, 2023. 5, 6[88] 肖杰, 付雪阳, 刘爱萍, 吴峰, 查正军. 图像除雨变压器. TPAMI, 45(11): 12978–12995, 2022. 5, 6, 8[89] Tete Shaw, Mannat Singh, Eric Mintun, Trevor Darrell, Pi-otr Dollar 和 Ross Girshick. 早期的卷积有助于变形金刚看得更清楚. 见 NeurIPS, 2021. 4[90] 徐刚, 侯其斌, 程明明. 双频率变压器可实现 sdr 到 hdr 的高效转换. 和平号, 2024. 3[91] 杨阳, 王超跃, 刘日升, 张林, 郭晓杰, 陶大成. 通过密度和深度分解进行自我增强的不成对图像去雾. 在 CVPR, 2022 年. 5, 6[92] 杨益新, 潘金山, 彭中正, 杜晓宇, 陶竹林, 唐金辉. Bistnet: 语义图像先导引导的双向时间特征融合, 用于基于深度示例的视频着色. TPAMI, 2024. 2[93] 叶天, 陈思祥, 白金斌, 施军, 薛成浩, 姜静霞, 尹俊杰, 陈尔康, 刘云. 使用密码本先验去除不利天气. 在 ICCV, 2023 年. 2, 5, 6[94] 易乔思, 李俊成, 戴钦艳, 方发明, 张贵绪, 曾铁勇. 保留结构的除雨, 带有残留通道预先指导. ICCV, 2021. 5, 6[95] 于柯, 王新涛, 董超, 唐晓鸥, 陈变来. 路径恢复: 学习网络路径选择以进行图像恢复. TPAMI, 44 (10) : 7078–7092, 2022. 2[96] 于伟豪, 司晨阳, 周潘, 罗米, 周一晨, 冯嘉世, 颜水成, 王新超. Metaformer 视觉基线. TPAMI, 46 (2) : 896–912, 2023. 7[97] 袁莉, 刘欣怡, 于建南, 李艳峰. 一种基于改进 pointnet++ 的全牙分割模型. VI, 1(1):21, 2023. 2[98] 岳宗胜, 赵干, 张磊, 孟德宇. 对偶对抗网络: 走向真实世界的噪声去除和噪声一代. ECCV, 2020. 2[99] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang 和 Ling Shao. 多阶段渐进图像恢复. 在 CVPR, 2021 年. 5, 6[100] Syed Waqas Zamir, Aditya Arora, Salman Khan, Mu-nawar Hayat, Fahad Shahbaz Khan 和 Ming-Hsuan Yang. Restormer: 用于高分辨率图像恢复的高效转换器. CVPR, 2022 年. 1, 2, 3, 5, 6, 7, 8



[101] 翟英杰, 贾国立, 赖玉坤, 张, 杨巨峰, 陶大成。通过双边姿势和运动图卷积网络感知情绪的步态。TAFRC, 2024 年。2[102] 张彪、Ivan Titov 和 Rico Sennrich。线性单位的稀疏注意力。载于 EMNLP, 2021. 3, 4[103] 张凯, 李亚伟, 左旺猛, 张雷, LucVan Gool, Radu Timofte。使用深度降噪器先进行即插即用图像恢复。TPAMI, 44 (10) : 6360–6376, 2021. 1, 2[104] 张凯浩, 任文琪, 罗文瀚, 赖伟胜, Björn Stenger, 杨明轩, 李洪东。深度图像去模糊: 一项调查。IJCV, 130 (9) : 2103–2130, 2022. 2[105] 张玉伦, 李鲲鹏, 李凯, 王力辰, 钟必能, 付云。使用非常深的残差通道注意网络进行图像超分辨率。ECCV, 2018. 2[106] 张玉伦, 李鲲鹏, 李凯, 钟必能, 付云。用于图像恢复的剩余非局部注意网络。ICLR, 2019. 2[107] 张志成和杨巨峰。时间情感本地化: 聆听和观看未修剪的视频。InACMMM, 2022. 2[108] 张志成, 刘胜哲, 杨巨峰。多平面对象跟踪。ICCV, 2023. 2[109] 张志成, 王丽娟, 杨巨峰。通过跨模态时间擦除网络进行弱监督视频情感检测和预测。CVPR, 2023. 2[110] 张志成, 陈松, 王紫川, 杨巨峰。Planeseg: 构建一个用于增强平面区域分割的插件。TNNLS, 2024. 2[111] 张志成, 胡军耀, 程文涛, DandaPaudel, 杨巨峰。Extedm: 用于视频预测的分布外推扩散模型。CVPR, 2024. 2[112] 张志成、赵攀成、Eunil Park、JufengYang。Mart: 通过掩蔽时间分布蒸馏进行掩蔽情感表征学习。CVPR, 2024. 2[113] 赵广祥, 林俊阳, 张志远, 任旭安成, 苏奇, 孙旭。显式稀疏变换器: 通过显式选择集中注意力。arXiv 预印本 arXiv:1912.11637, 2019. 1[114] Haiyu Zhao, Yuanbiao Gou, Boyyun Li, Dezhong Peng, Jian Cheng Lv, and Xi Peng。全面细腻: 图像修复的高效转换器。在 CVPR, 2023 年。2, 3, 6[115] 赵攀成, 徐鹏, 秦鹏达, 范登平, 张志成, 贾国立, 周博文, 杨巨峰。湖红: 通过潜在背景知识检索生成伪装图像-增强扩散。在 CVPR, 2024 年。2[116] 周满, 黄杰, 郭春乐, 李崇义。Fourmer: 一种高效的图像恢复全局建模范式。In ICML, 2023. 2, 5, 6[117] Shihao Zhou, Mengxi Jiang, Qicong Wang, and Yunqi Lei。Towards locality similarity serving to 3d human 姿势估计。在 ACCV 中, 2020 年。2

[118] Shihao Zhou, Mengxi Jiang, Shanshan Cai, and Yunqi Lei。Dc-gnet: 用于 3D 人体形状重建的深度网格关系捕获图卷积网络。在 ACMMM, 2021 年。2