

重振卷积网络进行图像恢复

Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll, *Fellow, IEEE*

摘要：图像恢复旨在从损坏的图像中重建高质量的图像，在许多场景中发挥着重要作用。近年来，由于其强大的远程像素交互建模能力，图像恢复领域发生了从卷积神经网络（CNN）到基于 Transformer 的模型的范式转变。在本文中，我们探索了 CNN 在图像恢复方面的潜力，并表明所提出的简单卷积网络架构（称为 ConvIR）的性能可以与 Transformer 对应产品相当或更好。通过重新审视先进图像恢复算法的特点，我们发现了导致恢复模型性能提升的几个关键因素。这促使我们开发一种基于廉价卷积算子的新型图像恢复网络。综合实验表明，我们的 ConvIR 在 20 个基准数据集的 5 个代表性图像恢复任务（包括图像去雾、图像运动/散焦去模糊、图像去雨和图像去雪）中提供了最先进的性能和低计算复杂度。

索引术语——卷积神经网络、频率调制、图像恢复、表示学习

1 简介

AS 最基本的视觉任务之一，图像恢复的目的是从退化的图像中恢复干净的图像，在遥感、无人系统、摄影和医学成像中发挥着重要作用[1]-[3]。由于这种反问题的不稳定性，许多传统算法都是基于手工设计的特征来减少解空间，这对于现实场景来说是不切实际的[4]。随着深度学习的发展，流形 CNN 基于巧妙模块或借用单元的框架已经被提出，例如编码器-解码器架构[5]、[6]、扩张卷积[7]-[9]、密集连接[10]和注意机制[11]，[12]。近年来，我们见证了从基于 CNN 的架构到 Transformer 模型的范式转变[13]、[14]。这些模型通过提供远程像素交互和输入特征的自适应能力，显著提高了图像恢复的最先进性能。尽管有一些补救措施[15]-[17]，降低图像恢复的自注意力的复杂性仍然是一个不小的问题。本文的主要目标是开发一种基于 CNN 的高效且有效的图像恢复架构，它可以性能与 Transformer 模型相当或更好。通过深入研究以前的先进图像恢复方法，我们总结了成功的图像恢复模型具有的几个关键因素：(a) 多尺度表示学习。最近的深度架构采用单个编码器-解码器[6]、[18]、[19]或多级范式[9]、[12]、[20]来学习多尺度特征表示，这有助于消除退化模糊不同的

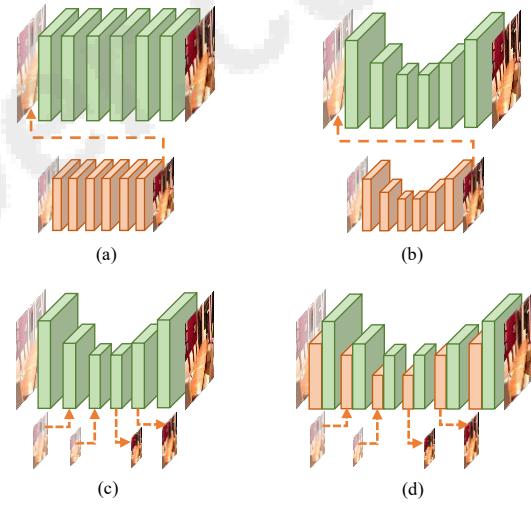


图 1. 多尺度架构之间的比较。 (a) 多级各向同性设计[9], [25]。 (b) 多级U形设计[26]。 (c) 单U形设计[6], [18]。 (d) 我们的网络模仿 U 形管道的多级设计。

尺寸。(b) 空间注意力。空间注意力有助于模型关注重要区域，这对于处理空间变化的模糊很有用[11]、[21]、[22]。(c) 频率调制。频率调制操作通过减少清晰图像对和退化图像对之间的频率差异，是对空间特征细化的有力补充[23]、[24]。(d) 低计算复杂度。这对于图像恢复至关重要，图像恢复通常涉及高分辨率图像。考虑到上述分析，我们重新思考卷积网络的设计，并开发一种高效且有效的图像恢复架构。首先，针对多尺度学习，我们回顾了图 1 中几种代表性的多尺度架构，并提出了

· Y. Cui 和 A. Knoll 就职于慕尼黑工业大学计算、信息和技术学院，邮编：85748 Garching, 慕尼黑, 德国。电子邮件： {yuning.cui, knoll}@in.tum.de · 任伟、曹X，光明区公共路66号中山大学深圳校区网络科学与技术学院中国广东省深圳市 518107。电子邮件： {renwq3, caoxiaochun}@mail.sysu.edu.cn。

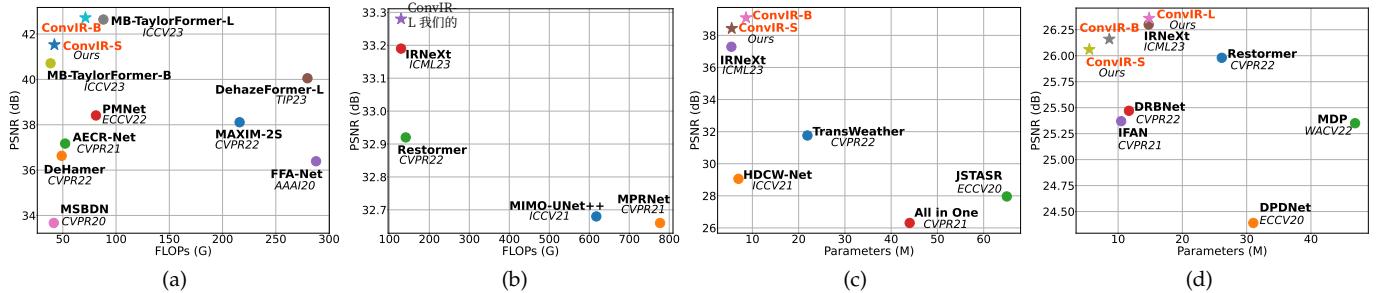


图 2. 针对四种图像恢复任务所提出的模型和最先进算法之间的比较。(a) 用于图像去雾的 SOTS-Indoor [27] 数据集上的 PSNR 与 FLOPs; (b) GoPro [25] 图像运动去模糊数据集上的 PSNR 与 FLOPs; (c) PSNR 与 CSD [28] 图像除雪数据集上的参数对比; (d) 用于图像散焦去模糊的 PSNR 与 DPDD [29] 数据集上的参数。

模仿单个 U 形网络中的多级机制。具体来说，对于 U 形架构的每个尺度，我们将特征图下采样为不同的尺寸，以便模型可以以从粗到细的方式处理模糊。其次，在注意力机制设计方面，我们开发了一种多形状注意力模块，该模块不仅可以在多种形状（即正方形和双向矩形区域）中进行有效的信息聚合，而且可以自适应输入特征。此外，该模块采用扩张卷积的思想来扩大感知领域。第三，由于形状和退化图像对之间存在显着的频率差异，以及不同光谱分量的重要性不同[30]，我们通过重新校准获得的高通滤波器的权重来强调用于高质量恢复的信息频率分量。多形状注意力模块的注意力图。最后，我们将上述模块插入到卷积U形主干中以建立我们的ConvIR，它获得了与Transformer模型相当或更好的性能。对于图像去雾，我们的模型优于之前最先进的方法（MB-TaylorFormer- L [31]）在 SOTS-Indoor [27] 和 SOTS-Outdoor [27] 基准上分别提高了 0.08 dB 和 1.33 dB PSNR，效率更高，如图 2 (a) 所示。对于运动模糊去除任务，我们的方法在 GoPro [25] 数据集上比最近基于 Transformer 的 PromptRestorer [32] 实现了 0.22 dB PSNR 的性能增益。此外，我们的模型显示了除雪任务的潜力，并且在 CSD [28] 数据集上的 PSNR 优于 TransWeather [33] 7.34 dB。此外，在用于散焦去模糊的 DPDD [29] 数据集上，我们的网络在组合类别中获得了 26.36 dB PSNR，比 Restormer [16] 提高了 0.38 dB PSNR。总的来说，本文的主要贡献可以总结如下：

- 大量实验表明，我们的模型在 20 个数据集上针对五种典型图像恢复任务（即图像去雾、图像运动/散焦去模糊、图像去雨和图像去雪）提供了最先进的性能。

这项研究是我们会议论文[34]的延伸。与初步版本相比，本文的主要改进是：a) 架构改进。与会议论文[34]中的单形状注意力机制相比，我们通过额外结合双向矩形注意单元开发了多形状注意（MSA）模块。此外，我们将不同速率的扩张操作注入到所提出的MSA中以扩大感受野。架构修改提高了模型的性能，同时所需的计算开销可以忽略不计。例如，我们的模型在 GoPro [25] 数据集上比 IRNeXt [34] 产生了 0.09 dB PSNR 的性能增益，仅额外增加了 0.07M 参数和 0.01G FLOP。b) 实验。我们的模型扩展到夜间（NHR [35], GTA5 [36]）和遥感（SateHaze1k-Thin/Moderate/Thick [37]）图像去雾问题并实现了最先进的性能。此外，我们在更合成和真实的数据集上评估我们的模型，例如 Haze4K [38]、O-HAZE [39] 和 I-Haze [40]。此外，我们对我们方法的关键组成部分进行了更多的消融研究。c) 模型多样性。为了提高部署便利性并展示我们设计的有效性，我们提供了三个版本的模型，即ConvIR-S（小）、ConvIR-B（基础）和ConvIR-L（大），以进行全面比较针对不同问题使用最先进的算法。在会议版本[34]中，我们主要采用ConvIR-L进行图像去模糊任务，采用ConvIR-S进行图像去雾和去雪。值得一提的是，我们的小版本在图像散焦去模糊的 DPDD [29] 数据集上仍然优于基于 StrongTransformer 的 Restormer [16]，参数较少，如图 2 (d) 所示。此外，ConvIR-B 在 SOTS-Indoor [27] 数据集上实现了最先进的性能，与最近基于 Transformer 的 MB-TaylorFormer-L [31] 相比，PSNR 的性能增益提高了 0.08 dB，复杂性降低了 19%，如图2(a)所示。

- 我们确定了成功的图像恢复方法所具有的特性，并提出了一种新型卷积模型 ConvIR，它通过将多阶段机制合到 U 形网络中来增强多尺度表示学习。
- 我们提出了一种有效的内容感知扩张多形状注意模块，可以通过重新权衡高通滤波器的权重来强调信息频率分量。

2 相关工作

2.1 图像恢复

作为一项长期存在的任务，图像恢复为可见性和下游高级任务（例如目标检测[41]和分割[42]）提供高质量的图像。CNN 多年来已成为该领域的主流，并取得了许多关于各种恢复任务的成功案例[25]、[29]、[43]、[44]。为了提高性能，人们开发了许多高级模块，并从其他领域借用了这些模块，以增强这些基于 CNN 的框架的能力[45]。例如，编码器-解码器架构在高效的层次表示学习中很流行[5]、[6]、[46]。多阶段范式[26]、[47]和多块学习方法[12]、[21]、[48]用于以从粗到细的方式恢复干净图像。扩张卷积有助于提取多尺度特征并捕获大感受野[7]、[23]。最近，Transformer 模型已被引入图像恢复中，并显着提高了各种图像恢复的最先进性能。由于自注意力强大的能力来建模远程依赖性[13]。例如，郭等人。[49]首次将Transformer引入图像去雾。Chen 等人。[50]提出了一种用于除雪的多尺度投影变换器。然而，Transformer 模型的关键组成部分（自注意力）相对于输入大小具有二次复杂度。一些补救措施通过减少注意力操作的大小[51]或切换操作维度来缓解这个问题。梁等人。[52] 和 Tsai 等人。[17]分别计算局部窗口和条带区域内的自注意力。扎米尔等人。[16]通过跨通道而不是空间维度应用自注意力来寻求解决方案。尽管如此，如何降低 Transformer 的复杂度仍然是该领域实际应用中的一个棘手问题。

2.2 注意力机制

在注意力机制在分类和检测等高级任务中取得成功的推动下，人们提出了各种注意力模块来关注图像恢复的基本内容[20]、[21]、[53]。例如，秦等人。[11]结合了通道注意力和像素注意力机制来进行图像去雾，以不平等对待不同的特征和像素。扎米尔等人。[12]设计了一个监督注意力模块来控制阶段之间的信号流。我们的注意力模块模仿深度卷积[54]来进行信息聚合，它具有自注意力的内容感知特性，同时保持计算效率。与我们的模块最相关的工作是学习用于恢复的动态滤波器的方法[6]、[55]、[56]。我们没有直接将学习到的注意力权重应用于输入特征，而是提前执行滤波器调制，通过重新调整注意力图中高通滤波器的重要性来强调特征中的信息频谱分量。此外，我们的模块提供多形状表示学习，并且不会产生与它们一样多的注意力权重，从而导致参数更少和复杂性更低。

2.3 谱网络

由于清晰和退化图像对的光谱特征存在很大差异[30]、[57]，因此在恢复问题的传统算法中广泛采用频率处理[58]、[59]。最近，研究人员将基于频率的模块纳入 CNN 和 Transformer 模型中，以弥补频谱差距。例如，毛等人。[57]支持基于傅里叶变换和CNN的运动去模糊的低频和高频学习。邹等人。[23]提出了一种基于小波的重建模块来恢复更多的高频细节。于等人。[60]通过重新审视频域中的雾度退化，在幅度谱的指导下重建相位分量。周等人。[61]将基于傅里叶的一般先验纳入空间交互和通道演化中。陈等人。[28]提出了一种基于双树复小波变换的分层网络除雪[62]。这些算法的常见做法是首先通过小波和傅里叶变换将空间特征变换到频域，然后利用卷积来调制得到的频谱。ConvIR不是遵循上述变换-CNN-逆变换的范式，而是进行滤波器调制使用轻量级注意力参数来计算注意力权重。因此，滤波器对于信息频率信号的重要性就被提升了。此外，我们将精细滤波器应用于空间特征，而无需使用任何现有的变换工具（例如傅里叶变换和小波变换）将这些特征变换到谱域，从而节省了计算开销。

3 方法

在本节中，我们首先描述我们网络的整体管道。然后，我们介绍核心组件：多尺度模块（MSM）和多形状注意力（MSA）。损失函数在最后部分介绍。

3.1 整体架构

如图 3 (a) 所示，所提出的网络采用 U 形架构进行图像恢复。具体来说，给定任何退化图像 $I \in R^{3 \times H \times W}$ ，ConvIR 首先应用 3×3 卷积层来生成大小为 $C \times H \times W$ 的浅层特征，其中 C 表示通道数， $H \times W$ 表示空间位置。然后，浅层特征通过三个 CNNBlock 产生深层特征。每个 CNNBlock 包含多个残差块，最后一个残差块插入了我们的 MSM，如图 3 (c) 所示。在此过程中，通道扩展，而空间分辨率降低。此外，遵循先前的算法[5]、[47]、[57]，将多个下采样的退化图像合并到主路径中，以更好地处理不同模糊级别的图像。具体来说，Convs 通过逐渐增加通道数来从下采样的退化图像中提取特征。随后，将提取的特征与主路径中的特征连接起来，然后进行卷积以减少通道数量。接下来，将深度特征输入到另外三个 CNN 块中以恢复高分辨率特征。期间

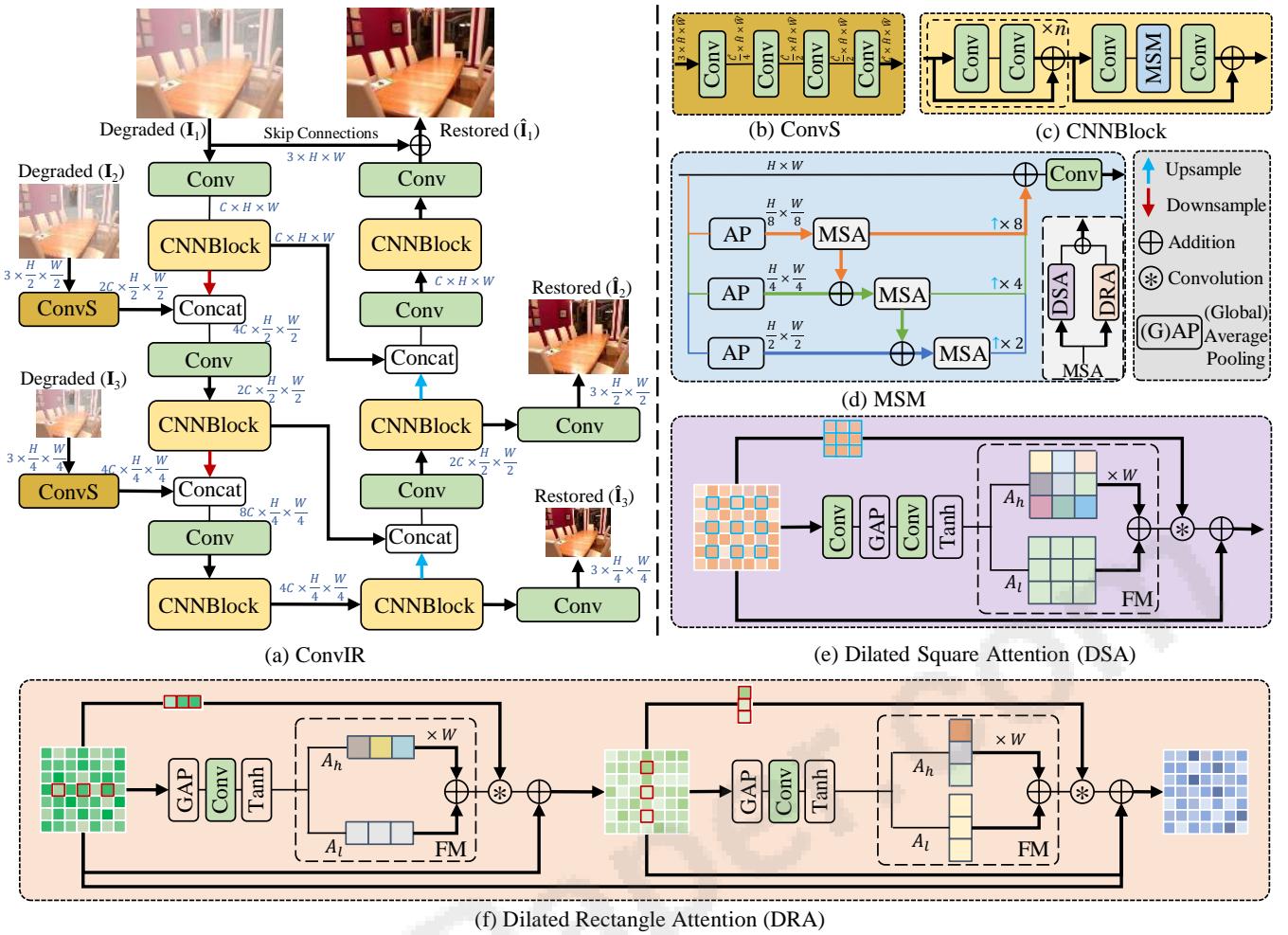


图 3. 所提出的 ConvIR 的架构。(a) ConvIR 由六个 CNNBlock 组成，并采用多输入和多输出策略进行图像恢复。(b) ConvS 从低分辨率退化图像中提取浅层特征，包括一系列内核大小为 3×3 、 1×1 、 3×3 和 1×1 的卷积，并逐渐将通道数从 3 增加到目标数量。(c) CNNBlock 包含多个残差块，并将所提出的多尺度模块 (MSM) 插入到最后一个残差块中。(d) MSM 在 U 形网络的每个尺度上提供多尺度表示学习。主要算子多形状注意力 (MSA)，并行包含扩张方形注意力 (DSA) 和扩张矩形注意力 (DRA)。(e) DSA 基于滤波器调制 (FM) 在扩张的方形区域内执行信息聚合。(f) DRA 基于 FM 采集两个方向矩形内的信号。

训练时，采用多输出策略，其中在解码器的前两个 CNN 块之后使用 3×3 卷积和图像级跳跃连接来预测低分辨率干净图像，为简单起见，图 3 (a) 中省略了这些连接。此外，解码器特征与编码器特征连接以帮助恢复，随后使用 1×1 卷积层将通道数量减半。残留的干净图像是通过 3×3 卷积生成的，其中添加了降级的输入以输出最终的恢复图像。接下来，我们详细介绍建议的模块：MSM 和 MSA。

多级网络并在单个 U 形框架的各个尺度中实现，如图 1 (d) 所示。MSM 的架构如图 3 (d) 所示。对于输入张量 $X \in R^{H \times W}$ ，为了清楚起见，通道维度被忽略，我们的 MSM 利用具有不同下采样比率的平均池 (AP) 算子将 X 转换为不同的特征空间。在每个分支中，MSA 后得到的特征通过加法运算合并到下一个分支中。这样，MSM 可以模仿多级网络逐步消除退化。最后，将所有分支的输出统一为原始输入大小并相加。在 ConvIR 中，我们凭经验采用三个分支加上恒等连接，其中下采样率设置为 $\{8, 4, 2\}$ 。对于第 i ($i \in \{1, 2, 3\}$) 分支 (除了恒等路径)，输出特征可以通过以下方式获得：

3.2 Multi-Scale Module (MSM)

单一编码器-解码器范式通常应用于最近的深度恢复架构中，以有效地学习分层表示。然而，这些作品中的尺度数量仅限于处理不同尺寸的退化模糊。为了增强多尺度学习并在每个尺度上以从粗到细的方式去除模糊，我们模仿

$$\hat{X}_i = \text{MSA}(\text{AP}_{2^{4-i}}(X) + \hat{X}_{i-1} \uparrow_2) \uparrow_{2^{4-i}}, \quad (1)$$

其中 $X_0=0$ ； $\text{AP } 2^{4-i}$ 表示平均池化，下采样率为 2^{4-i} ； $\uparrow 2$ 代表双线性

上采样率为2进行插值。综上所述，MSM的整个过程可以形式化地表示为：

$$\hat{\mathbf{X}} = \text{Conv}_{3 \times 3} \left(\sum_{i=1}^3 \mathbf{x}_i + \mathbf{X} \right), \quad (2)$$

其中 Conv 3×3 表示内核大小为 3×3 的卷积。

3.3 Multi-Shape Attention (MSA)

为了促进多尺度学习，我们的目标是设计一个插入 MSM 每个分支的高效模块来细化特征。配备自注意力机制的 Transformer 模型在各种图像恢复任务中取得了良好的性能 [13]、[14]。然而，尽管有一些补救措施 [15]、[17]，自注意力的二次复杂性问题仍然难以解决。另一方面，卷积算子具有静态滤波器，无法处理空间变化的退化模糊 [16]。在这项工作中，我们通过结合自注意力和卷积算子的优点提出了 MSA。我们的 MSA 继承了前者的感知特性并保持了后者的效率特征。此外，我们的 MSA 涉及不同形状和不同速率的膨胀机制的算子，以增强多形状和多尺度表示学习并扩大感受野。如图 3 (d) 所示，MSA 由并行的 Dilated Square Attention (DSA) 和 Dilated Rectangle Attention (DRA) 组成。接下来，我们描述 DSA 和 DRA。

3.3.1 扩张平方注意力 (DSA) 如图 3 (e) 所示，DSA 利用简单的卷积块来生成注意力权重，该权重适应输入特征，然后使用卷积运算进行聚合。在典型的自注意力中，Softmax 用于标准化注意力权重。然而，所得的和一权重可以被视为低通滤波器的内核 [63]，这不适合图像恢复，因为清晰图像和退化图像之间的显着差异主要在于高频成分 [30], [57]。我们从两个方面解决了注意权重生成步骤中的上述问题：(i) 绕过低频率的限制- 具有双曲正切函数 (Tanh) 的通滤波器，以及 (ii) 通过提出的滤波器调制 (FM) 提高高通滤波器注意力权重的重要性。Tanh 的利用。我们用 Tanh 替代 Softmax。这个方案有两个优点。首先，我们避开低通滤波器的限制。其次，由于 Tanh 将注意力权重投影到 $(-1, 1)$ ，负权重扫描有助于抑制执行信息聚合时的有害像素。形式上，给定 $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$ ，注意力权重生成过程可以表示为：

$$\mathbf{A}^{DSA} = \text{Tanh}(\text{Conv}_{1 \times 1}(\text{GAP}(\text{Conv}_{3 \times 3}(\mathbf{X})))), \quad (3)$$

其中 GAP 表示全局平均池化，Tanh 表示双曲正切函数。为了在注意力权重的复杂性和多样性之间取得更好的权衡，我们不是为每个通道产生注意力权重 [6], [55]，而是对组中的输入特征施加注意力权重。在每个特征组中，注意力权重在通道和空间维度上共享。

A DSA $\in \mathbb{R}^{G \times K \times K}$ ，其中 G 是组数，K 2 是积分的绝对区域大小。FM。除了使用 Tanh 之外，我们还建议重新权衡注意力图中高通滤波器的比率，以使网络能够更多地关注信息丰富的频率分量。为此，如图 3 (e) 所示，我们首先将注意力图 A DSA 分解为低通/高通滤波器，然后使用可训练的通道参数重新权衡高通滤波器。因此，重新组装的滤波器变得自适应以强调有用频率。在实践中，由于其易于实现，我们将低通滤波器称为仅保留输入的直流分量的特定滤波器，该滤波器可以从 A DSA 中提取经过：

$$\mathbf{A}_l^{DSA} = \frac{1}{K^2} \mathbf{E}, \quad (4)$$

其中 $\mathbf{E} \in \mathbb{R}^{G \times K \times K}$ 与 A DSA 具有相同的形状，所有值为 1。有关方程的更多详细信息，请参阅附录。4. 那么，高通滤波器可以认为是低通滤波器的互补部分：

$$\mathbf{A}_h^{DSA} = \mathbf{A}^{DSA} - \mathbf{A}_l^{DSA}. \quad (5)$$

接下来，调制后的注意力图可以通过以下方式获得：

$$\tilde{\mathbf{A}}^{DSA} = \mathbf{A}_l^{DSA} + W \mathbf{A}_h^{DSA}, \quad (6)$$

其中 W 表示通过反向传播直接优化的可学习参数，并初始化为 1。最后，我们通过卷积运算将所得的注意力权重应用于输入特征，其中输入特征的像素以扩张的方式采样以获得大的感受野。形式上，对于第 g 组中的每个通道，可以通过以下方式获得输出：

$$\hat{\mathbf{X}}_{g,h,w} = \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} \mathbf{X}_{g,(h-\lfloor \frac{K}{2} \rfloor+i)d, (w-\lfloor \frac{K}{2} \rfloor+j)d} \tilde{\mathbf{A}}_{g,i,j}^{DSA} + \mathbf{X}_{g,h,w}, \quad (7)$$

(7) 其中 g、h、w 分别是组的索引、高度和宽度。d 表示膨胀率。

3.3.2 扩张矩形注意力 (DRA) 除了 DSA 之外，我们还提出了 DRA，它在正交方向上集成矩形内的信息，以改进多形状表示学习。该架构如图 3 (f) 所示。与 DSA 类似，我们采用卷积网络来生成原始注意力权重，然后通过矩形 FM 重新评估其高通滤波器。随后，将调制的注意力权重施加到输入特征的扩张像素上以进行信息集成。以水平单元为例，将输入特征表示为 $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$ ，每组的输出特征可以通过以下形式得到：

$$\hat{\mathbf{X}}_{g,h,w}^H = \sum_{j=0}^{K-1} \mathbf{X}_{g,h,(w-\lfloor \frac{K}{2} \rfloor+j)d} \tilde{\mathbf{A}}_{g,j}^H + \mathbf{X}_{g,h,w}, \quad (8)$$

$$\tilde{\mathbf{A}}^H = \text{RFM}(\mathbf{A}^H), \quad (9)$$

$$\mathbf{A}^H = \text{Tanh}(\text{Conv}_{1 \times 1}(\text{GAP}(\mathbf{X}))) \in \mathbb{R}^{G \times K}, \quad (10)$$

其中 g、h、w 索引组、高度和宽度；d 为膨胀率；A H 表示注意力图

DRA 中的水平单位； RFM 意味着将 FM 应用于矩形注意力权重。卧式单位可

得出 $XH = H(X)$ 。同理，以 X 为输入，则输出

垂直单位可由 $XV = V(XH)$ 产生。最 $V^T = \mathcal{V}(\hat{\mathbf{X}}^H)$ 。Finally, the 后，DRA 的结果由 $X = XV + X$ 生成。

input, the output of the $\hat{\mathbf{X}}^V + \mathbf{X}$.

3.4 损失函数

由于我们在 MSA 中引入 FM，除了空间 L1 损失之外，我们还采用频谱 L1 损失来强调有用频率。双域损失函数由下式给出：

$$\mathcal{L}_{spatial} = \sum_{i=1}^3 \frac{1}{P_i} \|\hat{\mathbf{I}}_i - \mathbf{Y}_i\|_1, \quad (11)$$

$$\mathcal{L}_{frequency} = \sum_{i=1}^3 \frac{1}{S_i} \|[\mathcal{R}(\hat{\mathbf{I}}_i), \mathcal{I}(\hat{\mathbf{I}}_i)] - [\mathcal{R}(\mathbf{Y}_i), \mathcal{I}(\mathbf{Y}_i)]\|_1, \quad (12)$$

其中 i 是多个输出的索引，如图 3(a) 所示； $\hat{\mathbf{I}}$ 和 \mathbf{Y} 分别是恢复的图像和 groundtruth； P 和 S 表示归一化的总元素； $[\cdot, \cdot]$ 是串联运算符； R 和 I 是快速傅里叶变换产生的实部和虚部。最终的损失函数由下式得到：

$$L_{\text{总计}} = L_{\text{空间}} + \lambda L_{\text{频率}}, \quad (13)$$

其中 λ 设置为 0.1 以平衡双域训练。

4 实验

为了验证我们方法的有效性，我们在 20 个不同的数据集上针对五种图像恢复任务评估了 ConvIR：图像去雾、图像散焦去模糊、图像去雪、图像去雨和图像运动去模糊。在本节中，我们首先介绍实验设置。然后，我们展示我们的模型的结果，并将它们与最先进的方案进行定量和定性比较。最后，我们进行了广泛的消融研究，以验证我们提出的组件的功效。在表中，最佳结果和次佳结果分别以粗体和下划线表示。

4.1 实验设置

实施细节。我们针对不同的问题训练不同的模型。除非另有说明，所有实验均采用以下超参数。组数(G)和区域大小(K)分别设置为8和3。MSM的三个分支(从上到下)的扩张率分别为7、9和11。我们使用 Adam 优化器 [64] 训练模型，初始学习率为 $1e-4$ ，通过余弦退火 [65] 逐渐降低到 $1e-6$ 。对于数据增强，我们仅使用随机水平翻转。根据不同问题的复杂性，我们在实验中引入了三种ConvIR变体，通过改变CNNBlock中规则残差块的数量(n)来全面验证有效性。具体来说，我们在 ConvIR-S(小)、ConvIR-B(基础)和 ConvIR-L(大)中设置 $n=3$ 、 $n=7$ 和 $n=15$ 。所有模型均在 NVIDIA Tesla A100 上进行训练。

表 1 合成 SOTS [27] 数据集上的图像去雾比较。

| Methods | SOTS-Indoor | | SOTS-Outdoor | | Params (M) | FLOPs (G) |
|------------------------|--------------|--------------|--------------|--------------|------------|-----------|
| | PSNR | SSIM | PSNR | SSIM | | |
| GridDehazeNet [20] | 32.16 | 0.984 | 30.86 | 0.982 | 0.956 | 21.5 |
| MSBDN [10] | 33.67 | 0.985 | 33.48 | 0.982 | 31.35 | 41.54 |
| FFA-Net [11] | 36.39 | 0.989 | 33.57 | 0.984 | 4.456 | 287.8 |
| AECR-Net [68] | 37.17 | 0.990 | - | - | 2.611 | 52.2 |
| DeHamer [49] | 36.63 | 0.988 | 35.18 | 0.986 | 132.50 | 60.3 |
| DehazeFormer-L [51] | 40.05 | <u>0.996</u> | - | - | 25.44 | 279.7 |
| MAXIM [47] | 38.11 | 0.991 | 34.19 | 0.985 | 14.1 | 216 |
| PMNet [69] | 38.41 | 0.990 | 34.74 | 0.985 | 18.90 | 81.13 |
| MB-TaylorFormer-B [31] | 40.71 | 0.992 | 37.42 | 0.989 | 2.68 | 38.5 |
| MB-TaylorFormer-L [31] | <u>42.64</u> | 0.994 | <u>38.09</u> | 0.991 | 7.43 | 88.1 |
| ConvIR-S (Ours) | 41.53 | 0.996 | 37.95 | 0.994 | 5.53 | 42.1 |
| ConvIR-B (Ours) | 42.72 | 0.997 | 39.42 | 0.996 | 8.63 | 71.22 |

表 2 Haze4K [38] 数据集上的图像去雾结果。

| Methods | PSNR | SSIM | Params/M | FLOPs/G |
|------------------------|--------------|-------------|----------|---------|
| DehazeNet [43] | 19.12 | 0.84 | 0.01 | 0.58 |
| AOD-Net [70] | 17.15 | 0.83 | 0.002 | 0.12 |
| GridDehazeNet [20] | 23.29 | 0.93 | 0.956 | 21.5 |
| MSBDN [10] | 22.99 | 0.85 | 31.35 | 41.54 |
| FFA-Net [11] | 26.96 | 0.95 | 4.456 | 287.8 |
| DMT-Net [38] | 28.53 | 0.96 | - | - |
| PMNet [69] | 33.49 | <u>0.98</u> | 18.90 | 81.13 |
| FSNet [71] | 34.12 | 0.99 | 13.28 | 110.5 |
| ConvIR-S (Ours) | 33.36 | 0.99 | 5.53 | 42.1 |
| ConvIR-B (Ours) | 34.15 | 0.99 | 8.63 | 71.22 |
| ConvIR-L (Ours) | 34.50 | 0.99 | 14.83 | 129.34 |

GPU 与 PyTorch。附录中提供了数据集和特定训练配置的更多详细信息。评估指标。我们对所有数据集采用峰值信噪比(PSNR)和结构相似性(SSIM) [66] 指标。较高的 PSNR 和 SSIM 表明恢复的图像在像素内容和结构特性方面更接近目标图像。此外，平均绝对误差

(MAE) 和学习感知图像块相似度(LPIPS) [67] 被用于散焦去模糊，得分越低表明性能越好。除非另有说明，否则 FLOP 是在 256×256 个 patch 上测量的。

4.2 实验结果

4.2.1 图像去雾 对于这个问题，我们首先将我们的方法与表 1 中 SOTS [27] 数据集上最先进的算法进行比较。可以看出，我们的 ConvIR-B 在所有指标上都取得了最佳的精度结果。特别是，ConvIR-B 在 SOTS-Indoor [27] 和 SOTS-Outdoor [27] 数据集上的 PSNR 性能分别优于第一个基于 Transformer 的去雾算法 DeHamer [49]，参数减少了 93%，PSNR 分别提高了 6.09 dB 和 4.24 dB。我们的小型模型明显优于昂贵的基于 Transformer 的 DehazeFormer-L [51]，在 SOTS-Indoor 上的性能增益为 1.48 dB PSNR，复杂度降低了 85%。此外，我们的两个模型以相当的计算开销超越了最近的 MB-TaylorFormer [31] 的相应变体。此外，我们在表 2 中展示了对更现实的合成 Haze4K [38] 数据集的比较。

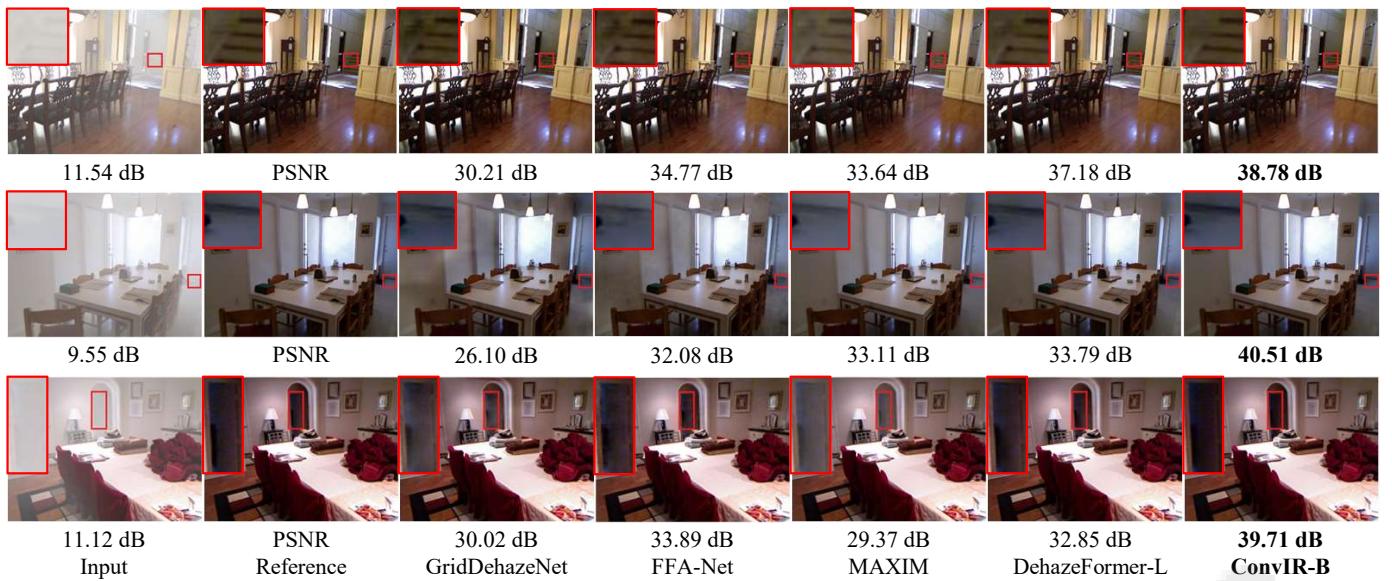


Fig. 4. Image dehazing comparisons on the SOTS-Indoor [27] dataset.

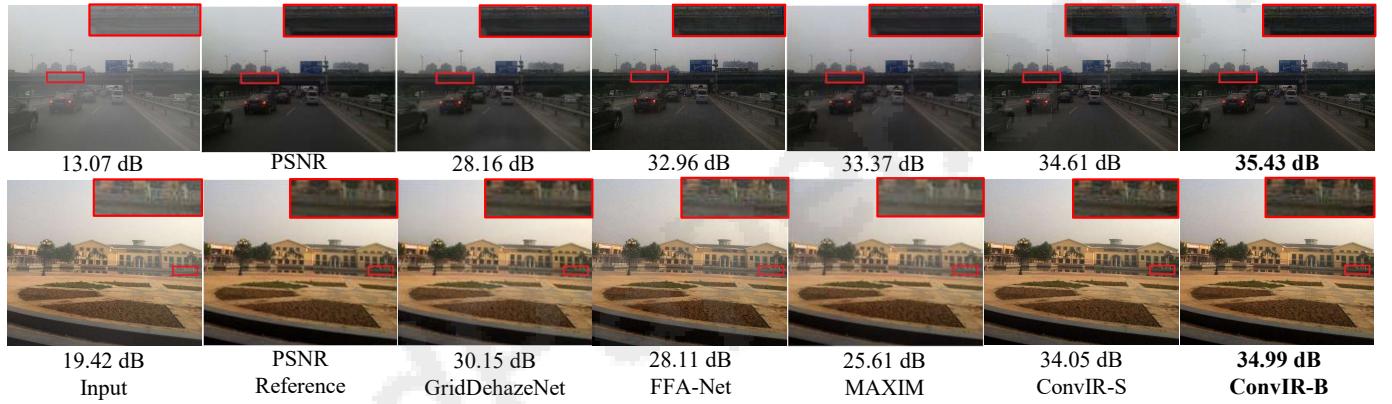


图 5. SOTS-Outdoor [27] 数据集上的图像去雾比较。

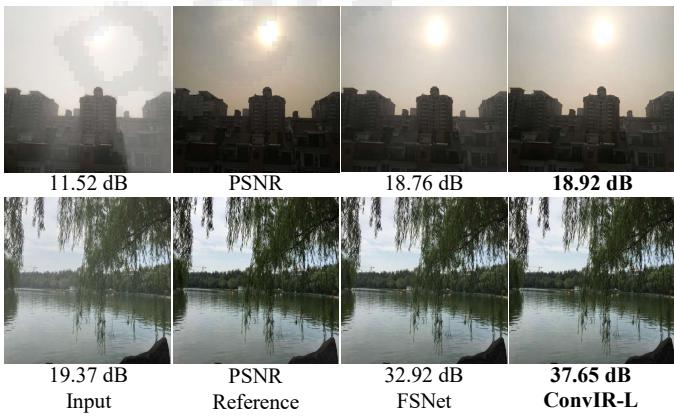


图 6. Haze4K [38] 数据集上的图像去雾比较。

比最近的 FSNet [71] 好 0.03 dB PSNR，只有 65% 的参数和 64% 的 FLOP。为了进一步证明我们模型的优越性，我们展示了我们的大型模型的结果，该模型的计算成本与 FSNet 相当。ConvIR-L 比 FSNet [71] 产生了 0.38 dB PSNR 的显着性能增益。

TABLE 3
Image dehazing comparisons on four real-world datasets:
Dense-Haze [72], NH-HAZE [73], O-HAZE [39], and I-Haze [40].

| Methods | Dense-Haze | | NH-HAZE | | O-HAZE | | I-Haze | |
|------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| GridDehazeNet [20] | 13.31 | 0.368 | 13.80 | 0.537 | 18.92 | 0.672 | 18.73 | 0.769 |
| MSBDN [10] | 15.13 | 0.555 | 19.23 | 0.706 | 24.36 | 0.749 | 19.62 | 0.618 |
| FFA-Net [11] | 15.70 | 0.549 | 19.87 | 0.692 | 22.12 | 0.770 | 19.72 | 0.733 |
| DeHamer [49] | 16.62 | 0.560 | 20.66 | 0.684 | 25.11 | 0.777 | - | - |
| PMNet [69] | 16.79 | 0.510 | 20.42 | 0.730 | - | - | - | - |
| MB-TaylorFormer-B [31] | 16.66 | 0.560 | - | - | 25.05 | 0.788 | - | - |
| MB-TaylorFormer-L [31] | 16.64 | 0.566 | - | - | 25.31 | 0.782 | - | - |
| ConvIR-S (Ours) | 17.45 | 0.648 | 20.65 | 0.807 | 25.25 | 0.784 | 21.95 | 0.888 |
| ConvIR-B (Ours) | 16.86 | 0.621 | 20.66 | 0.802 | 25.36 | 0.780 | 22.44 | 0.887 |

这些合成日间数据集 SOTS-Indoor [27]、SOTS-Outdoor [27] 和 Haze4K [38] 的视觉比较分别如图 4、图 5 和图 6 所示。我们的模型生成的无雾图像在视觉上更忠实于地面实况图像。我们进一步将我们的模型与四个真实世界数据集（即 Dense-Haze [72]、NH）上最先进的方案进行广泛比较-HAZE [73]、O-HAZE [39] 和 I-Haze [40]。表 3 显示最佳结果主要由



图 7. NHR [35] 数据集上的夜间图像去雾比较。

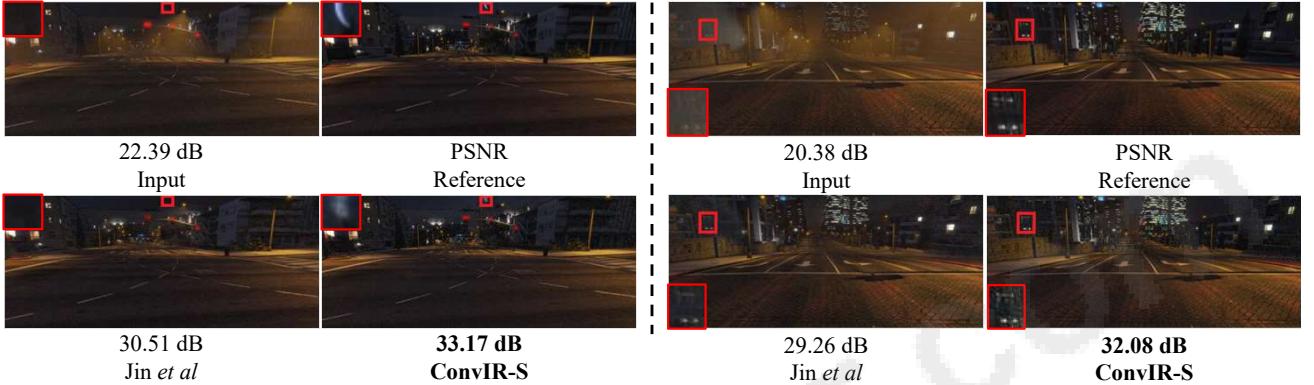


图 8. GTA5 [36] 数据集上的夜间图像去雾比较。

表 4 遥感数据集上的图像去雾比较：SateHaze1k-Thin、
SateHaze1k-Moderate 和 SateHaze1k-Thick [37]。† 表示专
门为遥感设计的方法。

| Methods | Thin | | Moderate | | Thick | |
|-------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| AOD-Net [70] | 19.54 | 0.854 | 20.10 | 0.885 | 15.92 | 0.731 |
| H2RL-Net [†] [74] | 20.91 | 0.880 | 22.34 | 0.906 | 17.41 | 0.768 |
| FCFT-Net [†] [75] | 23.59 | 0.913 | 22.88 | 0.927 | 20.03 | 0.816 |
| Uformer [15] | 22.82 | 0.907 | 24.47 | 0.939 | 20.36 | 0.815 |
| C ² PNet [76] | 19.62 | 0.880 | 24.79 | 0.940 | 16.83 | 0.790 |
| Restormer [16] | 23.08 | 0.912 | 24.73 | 0.933 | 18.58 | 0.762 |
| Trinity-Net [†] [77] | 21.55 | 0.884 | 23.35 | 0.895 | 20.97 | 0.823 |
| UMWTransformer [78] | 24.29 | 0.919 | 26.65 | 0.946 | 20.07 | 0.825 |
| FocalNet [46] | 24.16 | 0.916 | 25.99 | 0.947 | 21.69 | 0.847 |
| ConvIR-S (Ours) | 25.11 | 0.978 | 26.79 | 0.978 | 22.65 | 0.950 |

我们的模型。具体来说，我们的基本版本 ConvIR-B 在 Dense-Haze [72] 和 O-HAZE [39] 数据集上比最近的 MB-TaylorFormer-L [31] 分别获得了 0.22 dB 和 0.05 dB PSNR 的性能增益。值得注意的是，ConvIR-S 的性能优于 MB-TaylorFormer-L，在 Dense-Haze 上获得了 0.81 dB PSNR 的显着性能增益，而复杂度只有一半。因为图像去雾在遥感中起着至关重要的作用。我们提供了遥感 SateHaze1k [37] 数据集的实验结果。这些模型在其三个子集上分别进行训练和测试。表 4 显示我们的模型在所有指标上都表现最佳。特别是，我们的小模型显着优于一般的图像恢复方法[46]和遥感方法[77]

表 5 NHR [35] 数据集上的夜间图像去雾比较。† 表示专门为
夜间除雾设计的方法。

| Methods | PSNR | SSIM |
|-----------------------------|--------------|--------------|
| NDIM [†] [80] | 14.31 | 0.526 |
| GS [†] [81] | 17.32 | 0.629 |
| MRPF [†] [82] | 16.95 | 0.667 |
| MRPT [†] [82] | 19.93 | 0.777 |
| OSFD [†] [35] | 21.32 | 0.804 |
| HCD [83] | 23.43 | 0.953 |
| FocalNet [46] | 25.35 | 0.969 |
| Jin et al [†] [79] | 26.56 | 0.890 |
| ConvIR-S (Ours) | 28.85 | 0.981 |
| ConvIR-B (Ours) | 29.49 | 0.983 |

厚电平的 PSNR 分别为 0.96 dB 和 1.68 dB。

此外，我们在两个夜间数据集 NHR [35] 和 GTA5 [36] 上进行了实验。NHR [35] 的定量比较如表 5 所示。我们的 ConvIR-B 和 ConvIR-S 分别获得了最好和第二好的结果。特别是，ConvIR-S 的性能优于最新的算法 [79]，使用的参数减少了 3.8 倍，PSNR 显著提高了 2.29 dB。图 7 说明我们的 ConvIR-S 比其他算法恢复了更清晰的白天图像。此外，我们还提供了另一个夜间除雾数据集 GTA5 的结果，其地面实况图像位于夜间场景中。表 6 显示我们的两个版本优于算法 [79]，该算法是专门为夜间除雾而设计的。图 8 表明我们的模型在夜间场景中具有鲁棒性。

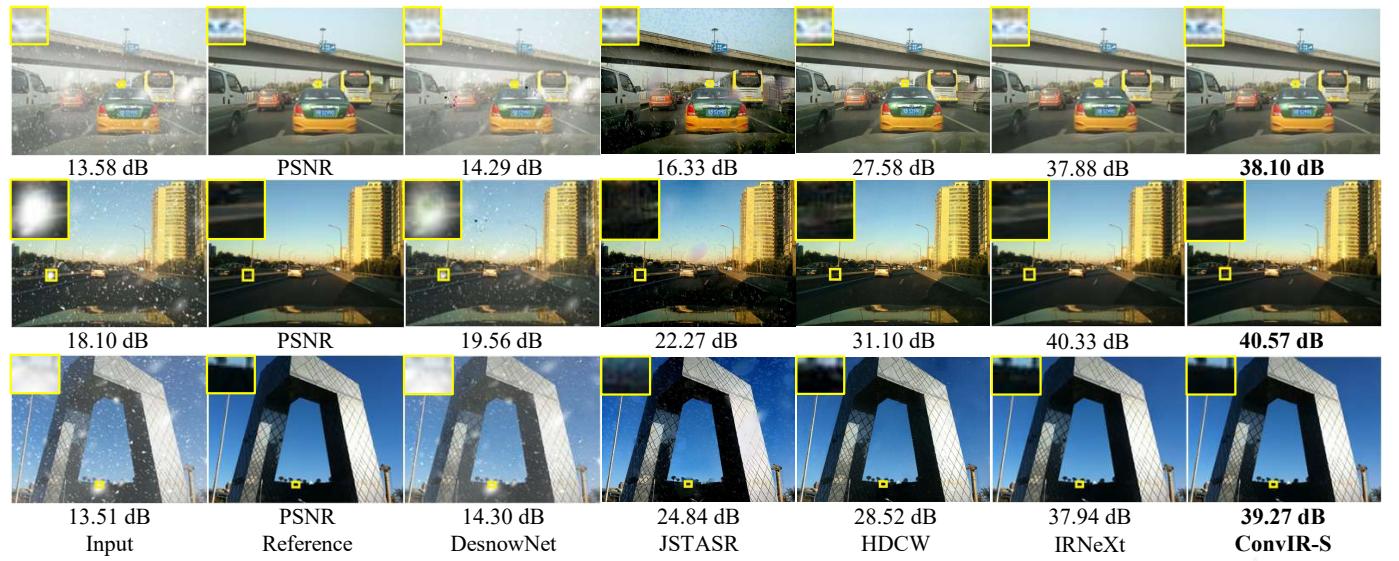


图 9. CSD [28] 数据集上的图像除雪比较。

表 6 GTA 5 [36] 数据集上的夜间图像去雾比较。† 表示专门为夜间除雾设计的方法。

| Methods | PSNR | SSIM |
|--|--------------|--------------|
| GS [†] [81] | 21.02 | 0.639 |
| MRP [†] [82] | 20.92 | 0.646 |
| Ancuti <i>et al.</i> [†] [84] | 20.59 | 0.623 |
| Yan <i>et al.</i> [†] [36] | 27.00 | 0.850 |
| CycleGAN [85] | 21.75 | 0.696 |
| Jin <i>et al.</i> [†] [79] | 30.38 | 0.904 |
| ConvIR-S (Ours) | 31.68 | 0.917 |
| ConvIR-B (Ours) | 31.83 | 0.921 |

表 7 CSD [28]、SRRS [86] 和 Snow100K [87] 数据集上的图像除雪比较。

| Methods | CSD | | SRRS | | Snow100K | | Params (M) | FLOPs (G) |
|------------------------|--------------|-------------|--------------|-------------|--------------|-------------|-------------|--------------|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | | |
| DesnowNet [87] | 20.13 | 0.81 | 20.38 | 0.84 | 30.50 | 0.94 | 15.6 | 1.7K |
| All in One [88] | 26.31 | 0.87 | 24.98 | 0.88 | 26.07 | 0.88 | 44 | 12.26 |
| JSTASR [86] | 27.96 | 0.88 | 25.82 | 0.89 | 23.12 | 0.86 | 65 | - |
| HDCW-Net [28] | 29.06 | 0.91 | 27.78 | 0.92 | 31.54 | 0.95 | 6.99 | 9.78 |
| SMGARN [89] | 31.93 | 0.95 | 29.14 | 0.94 | 31.92 | 0.93 | 6.86 | 450.3 |
| TransWeather [33] | 31.76 | 0.93 | 28.29 | 0.92 | 31.82 | 0.93 | 21.9 | 5.64 |
| MSP-Former [50] | 33.75 | 0.96 | 30.76 | 0.95 | 33.43 | 0.96 | 2.83 | 4.42 |
| FocalNet [46] | 37.18 | 0.99 | 31.34 | 0.98 | 33.53 | 0.95 | 3.74 | 30.63 |
| IRNeXt [34] | 37.29 | 0.99 | 31.91 | 0.98 | 33.61 | 0.95 | 5.46 | 42.09 |
| ConvIR-S (Ours) | 38.43 | 0.99 | 32.25 | 0.98 | 33.79 | 0.95 | 5.53 | 42.1 |
| ConvIR-B (Ours) | 39.10 | 0.99 | 32.39 | 0.98 | 33.92 | 0.96 | 8.63 | 71.22 |

4.2.2 Image Desnowing

我们比较了三个广泛采用的数据集的除雪性能：CSD [28]、SRRS [86] 和 Snow100K [87]。表 7 显示我们的 ConvIR-B 在所有指标上都产生了最佳结果。特别是，在计算开销相当的情况下，ConvIR-B 在 SRRS 和 Snow100K 上的 PSNR 方面分别优于最近的 IRNeXt [34] 0.48 dB 和 0.31 dB。在包含更复杂的雪景的具有挑战性的 CSD 数据集上，我们的 ConvIR-B 的优势

表 8 Test100 [90] 和 Test2800 [91] 上的图像除雨比较。

| Methods | Test100 | | Test2800 | |
|------------------------|--------------|--------------|--------------|--------------|
| | PSNR | SSIM | PSNR | SSIM |
| DerainNet [92] | 22.77 | 0.810 | 24.31 | 0.861 |
| SEMI [93] | 22.35 | 0.788 | 24.43 | 0.782 |
| UMRL [94] | 24.41 | 0.829 | 29.97 | 0.905 |
| RESCAN [95] | 25.00 | 0.835 | 31.29 | 0.904 |
| PreNet [96] | 24.81 | 0.851 | 31.75 | 0.916 |
| MSPFN [97] | 27.50 | 0.876 | 32.82 | 0.930 |
| MPRNet [12] | 30.27 | 0.897 | 33.64 | 0.938 |
| FSNet [71] | 31.05 | 0.919 | 33.64 | 0.936 |
| ConvIR-L (Ours) | 31.40 | 0.919 | 33.73 | 0.937 |

变得更加明显，展示了我们的网络在除雪方面的卓越能力。值得一提的是，我们的小模型在所有指标上都优于基于 Transformer 的 Tran-sWeather [33]，同时消耗的参数少了 75%。此外，与专门为除雪而设计的 MSP-Former [50] 相比，我们的小模型在 CSD 数据集上显示出 4.68 dBPSNR 的显著性能提升。图 9 显示我们的模型在消除积雪退化和恢复更多方面比竞争对手更有效没有明显伪影的详细轮廓，例如第二张图像中的道路分隔线。

4.2.3 Image Deraining

我们通过在复合数据集 [91]、[99]–[101] 上训练模型来进行图像去雨实验。Test100 [90] 和 Test2800 [91] 的评估结果如表 8 所示。我们的模型在 Test100 [91] 上的 PSNR 分别显示超过基于 CNN 的 FSNet [71] 和 MPRNet [12] 0.35 dB 和 1.13 dB PSNR。90] 数据集。与其他最先进的方案相比，我们的模型的优越性也可以在 Test2800 [91] 数据集上找到。Test100 [90] 数据集上的视觉比较如图 10 所示。正如我们所看到的，我们的方法生成



图 10. Test100 [90] 数据集上的图像去雨比较。

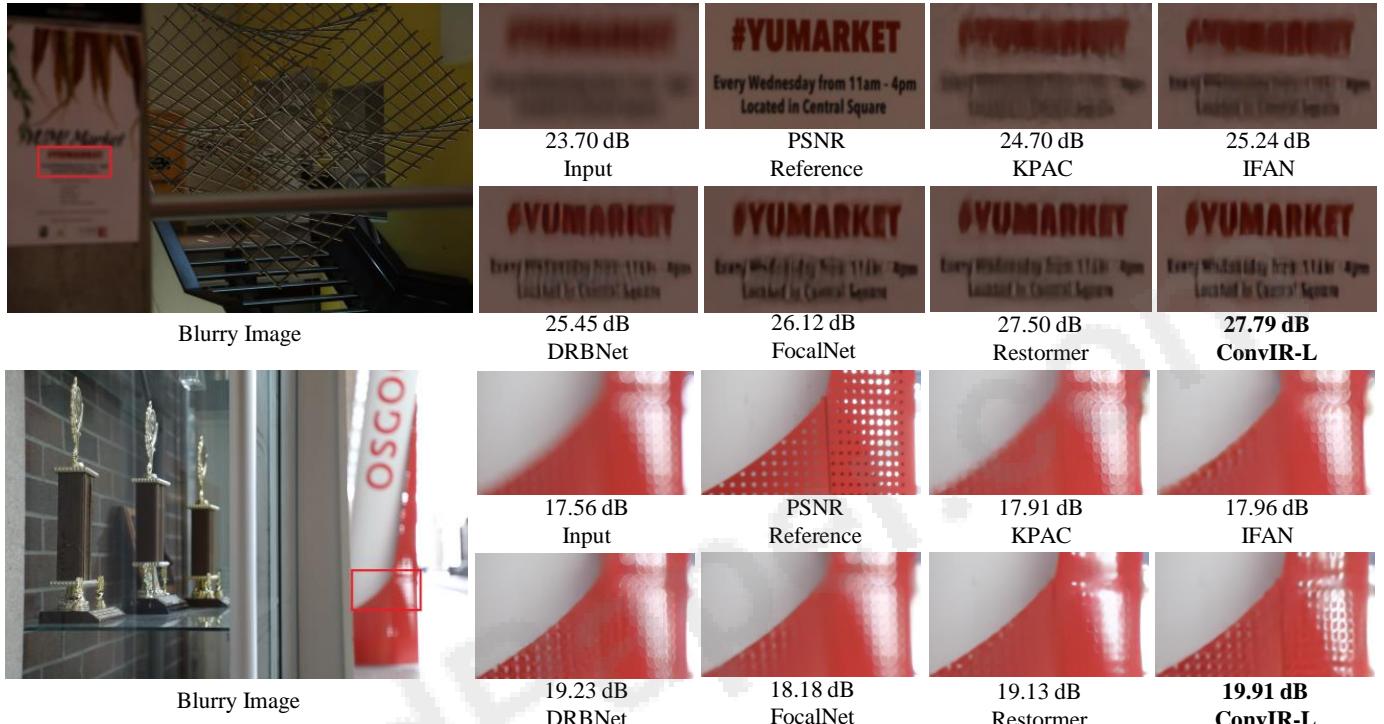


图 11. DPDD [29] 数据集上的图像散焦去模糊比较。

表 9 DPDD [29] 数据集上的图像散焦去模糊比较。分辨率为 $3 \times 720 \times 1280$ 下测量。

| Methods | Indoor Scenes | | | | Outdoor Scenes | | | | Combined | | | | Params (M) | FLOPs (G) |
|------------------------|---------------|--------------|--------------|--------------|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|------------|-----------|
| | PSNR↑ | SSIM↑ | MAE↓ | LPIPS↓ | PSNR↑ | SSIM↑ | MAE↓ | LPIPS↓ | PSNR↑ | SSIM↑ | MAE↓ | LPIPS↓ | | |
| DPDNet [29] | 26.54 | 0.816 | 0.031 | 0.239 | 22.25 | 0.682 | 0.056 | 0.313 | 24.34 | 0.747 | 0.044 | 0.277 | 31.03 | 770 |
| KPAC [7] | 27.97 | 0.852 | 0.026 | 0.182 | 22.62 | 0.701 | 0.053 | 0.269 | 25.22 | 0.774 | 0.040 | 0.227 | 2.06 | 113 |
| MDP [98] | 28.02 | 0.841 | 0.027 | - | 22.82 | 0.690 | 0.052 | - | 25.35 | 0.763 | 0.040 | - | 46.86 | 1898 |
| IFAN [6] | 28.11 | 0.861 | 0.026 | 0.179 | 22.76 | 0.720 | 0.052 | 0.254 | 25.37 | 0.789 | 0.039 | 0.217 | 10.48 | 363 |
| DRBNet [19] | | | | | | | | | 25.73 | 0.791 | - | 0.183 | 11.69 | 693 |
| FocalNet [46] | 29.10 | 0.876 | 0.024 | 0.173 | 23.41 | 0.743 | 0.049 | 0.246 | 26.18 | 0.808 | 0.037 | 0.210 | 12.82 | 1376 |
| Restormer [16] | 28.87 | 0.882 | 0.025 | 0.145 | 23.24 | 0.743 | 0.050 | 0.209 | 25.98 | 0.811 | 0.038 | 0.178 | 26.16 | 1983 |
| IRNeXt [34] | 29.22 | 0.879 | 0.024 | 0.167 | 23.53 | 0.752 | 0.049 | 0.244 | 26.30 | 0.814 | 0.037 | 0.206 | 14.76 | 1778 |
| ConvIR-S (Ours) | 28.95 | 0.877 | 0.024 | 0.158 | 23.32 | 0.747 | 0.050 | 0.221 | 26.06 | 0.810 | 0.037 | 0.190 | 5.53 | 579 |
| ConvIR-B (Ours) | 29.06 | 0.879 | 0.024 | 0.156 | 23.42 | 0.752 | 0.049 | 0.219 | 26.16 | 0.814 | 0.037 | 0.188 | 8.63 | 979 |
| ConvIR-L (Ours) | 29.37 | 0.887 | 0.023 | 0.143 | 23.51 | 0.757 | 0.049 | 0.203 | 26.36 | 0.820 | 0.036 | 0.174 | 14.83 | 1778 |

通过更好地消除雨天退化和恢复色彩，获得比其他竞争对手更高质量的图像。

4.2.4 Image Defocus Deblurring

我们使用我们的三个变体在广泛使用的 DPDD [29] 数据集上进行图像散焦去模糊实验，以与最先进的方法进行全面比较。

图像保真度得分如表 9 所示。我们的ConvIR-L 在组合类别中实现了 26.36 dB PSNR，比具有相似参数和 FLOPs 的最新基于 CNN 的 IRNeXt [34] 算法高 0.06 dB。与强大的 Transformer 相比基于 Restormer [16]，我们的大型模型在 PSNR 方面获得了 0.38 dB 的显着平均性能增益。值得一提的是



图 12. GoPro [25] 数据集上的图像运动去模糊比较。

表 10 GoPro 上的图像运动去模糊结果 [25]。使用 NVIDIA Tesla V100 GPU 以同步方式通过 `torch.cuda.synchronize()` 在 $3 \times 720 \times 1280$ 图像上测试推理时间。

| Methods | PSNR | SSIM | FLOPs (G) | Params (M) | Time (s) | Memory (G) |
|------------------------|--------------|--------------|-----------|------------|----------|------------|
| MIMO-UNet++ [5] | 32.68 | 0.959 | 617.64 | 16.1 | 1.277 | 10.395 |
| MPRNet [12] | 32.66 | 0.959 | 777.01 | 20.1 | 1.148 | 10.415 |
| MAXIM-3S [47] | 32.86 | 0.961 | 119.5 | 22.2 | - | - |
| Restormer [16] | 32.92 | 0.961 | 140.99 | 26.1 | 1.218 | 12.333 |
| Stripformer [17] | 33.08 | 0.962 | 170.46 | 20.0 | 1.054 | 12.149 |
| PromptRestorer [32] | 33.06 | 0.962 | - | - | - | - |
| IRNeXt [34] | 33.19 | 0.963 | 129.33 | 14.76 | 0.291 | 6.865 |
| ConvIR-L (Ours) | 33.28 | 0.963 | 129.34 | 14.83 | 0.323 | 6.867 |

尽管仅使用 21% 的参数和 29% 的 FLOP，我们的小版本仍然优于 Restormer，组合场景的 PSNR 增益为 0.08 dB。视觉比较如图 11 所示。可见，我们的方法从硬散焦退化中恢复了更多结构细节，比如海报上的文字。

4.2.5 图像运动去模糊

我们在广泛使用的合成 Go-Pro [25] 数据集和真实世界的 RSBlur [102] 数据集上评估我们的模型。表 10 列出了 GoPro [25] 数据集上准确性和计算成本的总体比较。与具有二次复杂度的 Transformer 模型 Restormer [16] 和 Stripformer [17] 相比，我们的网络、ConvIR-L、建立在高效的卷积网络上，并以更少的参数、更低的复杂性和更少的内存占用，分别获得了 0.36dB 和 0.20dB PSNR 的显著性能增益。此外，我们的模型运行速度比这两个模型快 3.77 倍和 3.26 倍。



图 13. 现实世界 RSBlur 的图像去模糊比较 [102]。

分别表明我们的模型在准确性和计算开销之间取得了更好的权衡。与更新的 Transformer 模型 PromptRestorer [32] 相比，我们的方法继续表现出优越的性能，得分提高了 0.22 dB。与基于 CNN 的方法（例如 MPRNet [12] 和 MIMO-UNet++ [5]）相比，我们的方法由于更先进的网络设计，例如经济高效的频率模块，在所有指标上都表现出优越性。此外，由于廉价的双向矩形注意单元和扩张操作，我们的模型在相似的计算开销下超越了 IRNeXt [34] 0.09 dB PSNR。GoPro [25] 数据集上的视觉比较如图 12 所示。我们的 ConvIR-L 产生通过消除大的运动模糊，比其他算法获得更视觉上令人愉悦的结果。此外，我们在表 11 中报告了真实世界 RSBlur [102] 的结果。可以看出，我们的方法优于基于 Transformer 的 Uformer [15] 和 Restormer [16] PSNR 分别提高了 0.08 dB 和 0.37 dB。图 13 中的视觉比较表明我们的模型生成的图像更接近参考，证明了我们的方法在现实场景中的鲁棒性。

表 11 现实世界 RSBlur [102] 数据集上的图像运动去模糊结果。

| Methods | PSNR | SSIM |
|------------------------|--------------|--------------|
| SRN-DeblurNet [103] | 32.53 | 0.840 |
| MIMO-UNet [5] | 32.73 | 0.846 |
| MIMO-UNet+ [5] | 33.37 | 0.856 |
| MPRNNet [12] | 33.61 | 0.861 |
| Restormer [16] | 33.69 | 0.863 |
| Uformer [15] | 33.98 | 0.866 |
| ConvIR-L (Ours) | 34.06 | 0.868 |

表 12 为获得更好性能而进行的分解消融研究。为了单独研究 MSM 的效果，我们在每个分支中部署 3×3 卷积来形成 MSM/Conv。这里，MSM 表示没有 MSA 的纯多尺度范式。DSA/Conv 表示 DSA 的降级版本，去掉了 FM 和扩张机制。

| Methods | a | b | c | d | e | f |
|--------------|-------|-------|-------|-------|-------|-------|
| Baseline | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| MSM/Conv | | ✓ | ✓ | ✓ | ✓ | ✓ |
| DSA/Conv | | | ✓ | ✓ | ✓ | ✓ |
| DSA/FM | | | | ✓ | ✓ | ✓ |
| DSA/Dilation | | | | | ✓ | ✓ |
| DRA | | | | | | ✓ |
| PSNR | 31.23 | 31.46 | 31.53 | 31.64 | 31.76 | 31.92 |
| Params/M | 6.90 | 8.45 | 8.55 | 8.56 | 8.56 | 8.63 |
| FLOPs/G | 66.32 | 71.17 | 71.19 | 71.19 | 71.19 | 71.22 |
| Time/s | 0.134 | 0.152 | 0.165 | 0.166 | 0.170 | 0.206 |

4.3 消融研究

遵循之前的方案[34]、[47]，我们使用 ConvIR-B 在 GoPro [25] 数据集上进行消融研究。基线模型是通过从我们的网络中删除 MSM 及其包含物而获得的。所有模型都训练了 1000 个 epoch。故障消融。我们通过将我们的组件连续应用到基线来执行分解消融。结果如表 12 所示。基线在 GoPro 上接收到 31.23 dB PSNR（表 12a）。在每个分支中部署仅有 3×3 卷积的 MSM 后，该模型实现了 0.23 dB 的改进（表 12b）。然后，我们用 DSA/Conv 替换这个 3×3 卷积，DSA/Conv 是 DSA 的降级版本，排除了 FM 和扩张机制，模型获得了 0.07 dB PSNR 的进一步提升性能（表 12c）。使用 FM 细化注意力图中的滤波器可带来 0.11 dB PSNR 的性能增益（表 12d）。使用扩张机制扩大感受野可将性能提升至 31.76 dB PSNR（表 12e）。另外部署矩形注意力，完整模型实现了最佳性能，比基线模型高 0.69 dB，并且仅引入了 1.73M 参数和 4.9G FLOPs。结果表明我们提出的组件的有效性。MSM 中的分支数量。分支数量在 MSM 由粗到细的机制中起着至关重要的作用。因此，我们通过改变分支数量来进行实验。表 13 显示采用更多分支会带来更好的性能。具体来说，当使用下采样率为 2 的单个分支时，该模型获得的 PSNR 增益为 0.22 dB

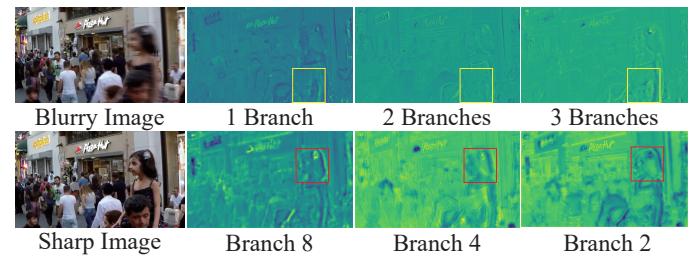


图 14. 中间特征图的可视化。前三个特征展示了在 MSM 中使用不同数量的分支的差异。相应的模型如表 13 所示。第二行的特征图展示了使用三个分支时具有不同下采样率的分支的结果。底线中的数字表示下采样率。

表 13 MSM 中的分支数量。该数字表示分支的下采样率。

| 2 | 4 | 8 | PSNR | Params/M | FLOPs/G |
|---|---|---|-------|----------|---------|
| ✓ | | | 31.23 | 6.90 | 66.32 |
| | ✓ | | 31.45 | 7.74 | 70.91 |
| | | ✓ | 31.62 | 8.18 | 71.15 |
| | | | 31.92 | 8.63 | 71.22 |

表 14 不同合并操作的消融研究。

| 池化类型 | 卷积 | 最大池化 | 平均池化 | Average Pooling |
|------|----|-------|-------|-----------------|
| PSNR | | 31.75 | 31.65 | 31.92 |

基线模型。当配备三个分支时，该模型通过产生 0.69 dB PSNR 改进，展示了所提出的从粗到精机制的有效性。我们进一步可视化图 14 中的中间特征，以研究我们从粗到精机制的效果-无性主义。我们首先展示从部署不同分支的模型的最后一个规模获得的特征。第一行的特征说明使用更多的分支可以恢复更多的高频信号以进行运动去模糊，例如黄色矩形中的轮廓。然后，我们绘制从利用三个分支的模型的每个分支获得的特征图。第二行显示模型逐渐恢复形状特征，例如红色区域中的面部。可视化展示了从粗到细机制的有效性。池化操作选择。我们研究了在 MSM 中使用不同池化技术的影响，即深度卷积、最大池化和平均池化。我们在所有变体中采用相同的下采样率。这三种操作具有相同的计算复杂度，而卷积则引入了额外的参数。结果如表 14 所示。平均池变体比其他两种变体取得了更好的结果。因此，我们选择平均池化作为默认配置。不同的激活函数。我们没有从 self-attention 继承 Softmax 来标准化注意力权重，而是在 MSA 中使用 Tanh。这允许网络避开低通滤波器并生成负权重

表 15 不同功能的消融研究。

| Methods | Softmax | Linear | Sigmoid | Tanh |
|---------|---------|--------|---------|-------|
| PSNR | 31.75 | 31.79 | 31.83 | 31.92 |

TABLE 16
Ablation studies for the number of groups in MSA.

| Methods | 2 | 4 | 8 | 16 |
|----------|-------|-------|-------|-------|
| PSNR | 31.80 | 31.79 | 31.92 | 31.85 |
| Params/M | 8.51 | 8.55 | 8.63 | 8.79 |

表17 扩张率的消融研究。(a,b,c)表示分别设置MSM的三个分支中a、b、c的膨胀率。

| Dilation Rates | (3,2,1) | (7,5,3) | (7,9,11) |
|----------------|---------|---------|----------|
| PSNR | 31.67 | 31.75 | 31.76 |

表 18 多输入和多输出机制的消融研究。

| Methods | a | b | c | d |
|-------------------------|-------|-------|-------|-------|
| Multi-output | ✓ | | | |
| Multi-input (3rd scale) | ✓ | ✓ | | |
| Multi-input (2nd scale) | ✓ | ✓ | ✓ | |
| PSNR | 31.23 | 30.77 | 30.39 | 28.97 |

对于在信息聚合过程中可能产生有害影响的像素。表 15 显示，与 Softmax 相比，线性投影和 Sigmoid 版本通过摆脱和为一的特性，分别实现了 0.04 dB 和 0.08 dB PSNR 的增益。Tanh 将注意力权重投射到 (-1, 1) 中，与基线模型相比，PSNR 显著提高了 0.17 dB。MSA 中的组数。在 MSA 中，我们学习分组注意力权重以进行信息聚合。为了研究注意力权重多样性的影响，我们通过改变组数进行实验。结果如表16所示。一般来说，随着组数的增加，性能有所提高。然而，第 8 组似乎已饱和，这种现象可能归因于过度拟合。扩张率。我们使用 MSA 中的扩张机制来扩大感受野。我们通过在 MSM 的三个分支中部署不同的膨胀率组合来进行实验。表 17 显示，增加扩张率可以通过感知更大的感受野获得更好的结果。最后，我们在模型中简单地选择 (7,9,11) 的组合。在这里，我们仅用一组有限的扩张率组合进行实验来验证我们设计的有效性，而不是详尽地搜索最佳选项。多输入和多输出策略的效果。我们研究了采用的多输入和多输出策略的效果通过从表 12 中的基线模型中逐渐删除这些技术来制定输出策略 [5]、[34]、[47]。表 18 显示该模型接收到 31.23 dB PSNR

表 19 使用相同数量参数 (0.09M) 的卷积模型和自注意力模型之间的总体比较。内存是训练时使用的内存。

| Methods | FLOPs/G | Memory/G | PSNR | Time/s |
|-----------|--------------|--------------|---------------|----------------|
| Conv | 6.19 | 14.3 | 25.86 | 0.019 |
| Attention | 8.28 (+2.09) | 43.7 (+29.4) | 25.85 (-0.01) | 0.085 (+0.066) |

使用多输入/输出机制。去除多输出方法后，性能下降至 30.77 dBPSNR。卸载多输入层会导致性能进一步下降。这些结果表明了多输入和多输出方案的有效性。

5 讨论

近年来，我们见证了从基于 CNN 的架构到具有二次复杂度的 Transformer 模型的范式转变。一些文献从不同的角度研究了卷积和 Transformer 模型之间的联系，包括通道混合、归一化、滤波器生成 [54] 和应用 [104] 以及频率偏好 [63]。在本研究中，我们只是因为自注意力的高度复杂性而激活了卷积网络。为了清楚地演示，我们进行了玩具实验。具体来说，我们通过分别仅部署五个残差卷积块和五个纯自注意力单元来构建两个微型模型 [105]。我们通过调整通道数来保持参数数量相等，以便进行公平比较。模型在 GoPro [25] 数据集上训练 100 个时期，初始学习率为 $16e-4$ ，批量大小为 64。获得的模型在使用 NVIDIA Tesla V100 GPU 的 GoPro [25] 测试集。表 19 显示，基于自注意力构建的模型在训练期间消耗更高的复杂度和内存占用。相比之下，卷积版本以更低的计算开销和更快的速度实现了相当的精度。因此，我们重新激活卷积网络以实现有效且高效的图像恢复。综合实验结果表明，使用正确的设计，卷积网络可以比精心设计的 Transformer 模型表现得更好或更好。我们希望这项研究能够激励研究人员进一步开发基于 CNN 的图像恢复模型的潜力。

六，结论

在这项研究中，我们分析了以前成功的图像恢复模型并确定了它们所拥有的良好特性。基于观察，我们提出了一种有效且高效的图像恢复卷积模型。在 20 个基准数据集上的大量实验结果表明，所提出的网络与 Transformer 模型相匹配，并在五个代表性图像恢复任务中实现了最先进的性能。我们的工作也有局限性。例如，我们仅在 MSM 中使用有限的膨胀率组合进行实验，以证明我们设计的有效性。有前途的方向包括学习最佳扩张率或

结合可变形算子[106]来捕获自适应和灵活的感受野。进一步的工作还可以涉及使用更便宜的替代品，例如 Ghost 模块[107]，来取代我们模型中的常规残差块，以实现轻量级设计。由于针对不同退化类型的自适应频率学习能力，我们的模型还具有用于一体式图像恢复任务的潜力。

REFERENCES

- [1]Y. Lim、Y. Bliesener、S. Narayanan 和 K. S. Nayak, “使用卷积神经网络对螺旋实时核磁共振进行去模糊”，*医学磁共振*，卷。84, 没有。6, 第 3438–3452 页, 2020 年。[2]B. Rasti、Y. Chang、E. Dalsasso、L. Denis 和 P. Ghamisi, “遥感图像恢复：概述和工具箱”，*IEEE 地球科学与遥感杂志*，卷。10, 不。2, 第 201–230 页, 2021 年。[3]S. Zang、M. Ding、D. Smith、P. Tyler、T. Rakotoarivelio 和 M. A. Kaafar, “恶劣天气条件对自动驾驶汽车的影响：雨、雪、雾和冰雹如何影响自动驾驶汽车的性能”，*IEEE 车辆技术杂志*，卷。14, 没有。2, 第 103–111 页, 2019 年。[4]K. 张, 任, W. 罗, W.-S. 赖, B. 斯坦格, M.-H. 杨, H. 李, “深度图像去模糊：一项调查”，*国际计算机视觉杂志*，卷。130, 没有。9, 第 2103–2130 页, 2022 年。[5]S.-J. 曹, S.-W. 吉, J.-P. 洪, S.-W. 荣格和 S.-J. Ko, “重新思考单图像去模糊中从粗到细的方法”，载于 *IEEE 国际计算机视觉会议论文集*，2021 年，第 123 页。4641–4650。[6]J. Lee、H. Son、J. Rim、S. Cho 和 S. Lee, “用于单图像散焦去模糊的迭代滤波器自适应网络”，*《IEEE 计算机视觉和模式识别会议论文集》*，2021 年，第 2034–2042 页。[7]H. Son、J. Lee、S. Cho 和 S. Lee, “使用内核共享并行多孔卷积进行单图像散焦去模糊”，载于 *IEEE 国际计算机视觉会议论文集*，2021 年，第 2642–2650 页。[8]J. Li、W. Tan 和 B. Yan, “感知多样性运动去模糊与光全局上下文细化”，*IEEE 国际计算机视觉会议论文集*，2021 年，第 4116–4125 页。[9]W. 任, L. Ma, J. 张, J. Pan, X. Cao, W. Liu, 和 M.-H. Yang, “单图像去雾的门控融合网络”，载于 *IEEE 计算机视觉与模式识别会议论文集*，2018 年。[10]H. Dong, J. Pan, L. Shang, Z. Hu, X. Zhang, F. Wang 和 M.-H. Yang, “具有密集特征融合的多尺度增强去雾网络”，*IEEE 会议记录计算机视觉和模式识别*，2020。[11]X. Qin、Z. Wang、Y. Bai、X. Xie 和 H. Jia, “FFa-net: 用于单图像去雾的特征融合注意网络”，*《AAAI 人工智能会议论文集》*，2020 年，第 11 908 页–11 915。[12]S. W. Zamir、A. Arora、S. Khan、M. Hayat、F. S. Khan、M.-H. Yang 和 L. Shao, “多阶段渐进式图像恢复”，*IEEE 计算机视觉和模式识别会议论文集*，2021 年，第 14 821–14 831 页。[13]陈宏、王勇、郭涛、徐成、邓勇、刘子、马思成、徐成、C. Xu 和 W. Gau, “预训练图像处理变压器”，*《IEEE 计算机视觉和模式识别会议论文集》*，2021 年，第 12 299–12 310 页。[14]J. Liang、J. Cao、G. Sun、K. Zhang、L. Van Gool 和 R. Timofte, “Swinir: 使用 swin 变压器进行图像恢复”，载于 *IEEE 国际计算机视觉会议论文集*，2021 年，页。1833年–1844 年。[15]Z. Wang、X. Cun、J. Bao、W. Zhou、J. Liu 和 H. Li, “Uformer: 用于图像恢复的通用 U 形变压器”，发表于 *IEEE 计算机视觉会议论文集和模式识别*，2022 年，第 17 683–17 693 页。[16]S. W. Zamir、A. Arora、S. Khan、M. Hayat、F. S. Khan 和 M.-H. Yang, “Restormer: 用于高分辨率图像恢复的高效变压器”，载于 *IEEE 计算机视觉和计算机会议记录模式识别*，2022 年，第 5728–5739 页。[17]缩略词。蔡, Y.-T. 彭, Y.-Y. 林, C.-C. 蔡, C.-W. Lin, “Strip-former: 用于快速图像去模糊的带状变压器”，*《欧洲计算机视觉会议记录》*，2022 年。
- [18]陈东, 何明, 范强, 廖建, 张丽, 侯东, 袁丽, G. Hua, “用于图像去雾和去雨的门控上下文聚合网络”，*IEEE 计算机视觉应用冬季会议*，2019 年，第 1375–1383 页。[19]L. Ruan、B. Chen、J. Li 和 M. Lam, “学习使用光场生成和真实散焦图像进行去模糊”，*《IEEE 计算机视觉和模式识别会议论文集》*，2022 年，第 16 304–16 页 313。[20]X. Liu、Y. Ma、Z. Shi 和 J. Chen, “Griddehazenet: 用于图像去雾的基于注意力的多尺度网络”，*《IEEE 国际计算机视觉会议论文集》*，2019 年，第 7314–7323 页。[21]M. Suin、K. Purohit 和 A. N. Rajagopalan, “用于自适应运动去模糊的空间注意力补丁分层网络”，载于 *IEEE 计算机视觉和模式识别会议论文集*，2020 年。[22]Y. Cui、Y. Tao、W. Ren 和 A. Knoll, “图像去模糊的双域注意力”，*人工智能促进会*，2023 年。[23]W. Zou、M. Jiang、Y. Zhang、L. Chen、Z. Lu 和 Y. Wu, “Sdwnet: 用于图像去模糊的小波变换直扩张网络”，*IEEE 国际计算机会议论文集愿景*，2021 年，第 1895–1904 页。[24]Y. Cui、Y. Tao、Z. Bing、W. Ren、X. Gau、X. Cao、K. Huang 和 A. Knoll, “用于图像恢复的选择性频率网络”，*国际学习表示会议*，2023 年。[25]S. Nah、T. Hyun Kim 和 K. Mu Lee, “用于动态场景去模糊的深度多尺度卷积神经网络”，载于 *IEEE 计算机视觉和模式识别会议论文集*，2017 年。[26]H. Gau、X. Tao、X. Shen 和 J. Jia, “使用参数选择性共享和嵌套跳过连接进行动态场景去模糊”，载于 *IEEE 计算机视觉和模式识别会议论文集*，2019 年。[27]B. Li、W. Ren、D. Fu、D. Tao、D. Feng、W. Zeng 和 Z. Wang, “单图像去雾及其他基准测试”，*IEEE 图像处理交易*，卷。28, 没有。1, 第 492–505 页, 2018 年。[28]W.-T. 陈 H.-Y. 方, C.-L. 谢, C.-C. 蔡, 陈, J.-J. T.S.-Y. Kuo 等人, “所有雪被移除：使用分层双树复数小波表示和矛盾通道损失的单图像除雪算法”，*《IEEE 国际计算机视觉会议论文集》*，2021 年，第 4196–4205 页。[29]A. Abuolaim 和 M. S. Brown, “使用双像素数据进行散焦去模糊”，*欧洲计算机视觉会议记录*，2020 年，第 111–126 页。[30]K.-H. 刘, C.-H. 是的, J.-W. 钟和 C.-Y. Chang, “一种基于多尺度高频残差图像学习的运动去模糊方法”，*IEEE Access*，卷。8, 第 66 025–66 036 页, 2020 年。[31]Y. Qiu、K. Zhang、C. Wang、W. Luo、H. Li 和 Z. Jin, “Mb-taylorformer: 用于图像去雾的多分支高效变压器扩展 bytaylor 公式”，*《IEEE 国际会议论文集》计算机视觉*，2023 年，第 12 802–12 813 页。[32]C. Wang、J. Pan、W. Wang、J. Dong、M. Wang、Y. Ju、J. Chen 和 X.-M. 吴, “Promptrestorer: 一种具有退化感知的提示图像恢复方法”，载于 *第三十七届神经信息处理系统会议*，2023 年。[33]J. M. J. Valanarasu、R. Yasara 和 V. M. Patel, “Transweather: 基于 Transformer 的恶劣天气条件退化图像恢复”，载于 *IEEE 计算机视觉和模式识别会议论文集*，2022 年，第 2353–2363 页。[34]Y. Cui、W. Ren、S. Yang、X. Cao 和 A. Knoll, “Irnext: 重新思考用于图像恢复的卷积网络设计”，*《机器学习国际会议论文集》*，2023 年。[35]J. 张, Y. 曹, Z.-J. Zha 和 D. Tao, “利用综合基准进行夜间除雾”，*《ACM 国际多媒体会议记录》*，2020 年，第 2355–2363 页。[36]W. Yan、R. T. Tan 和 D. Dai, “使用高低频分解和灰度颜色网络进行夜间去雾”，*《欧洲计算机视觉会议记录》*，2020 年，第 473–488 页。[37]B. Huang、L. zhi、C. Yang、F. Sun 和 Y. Song, “基于条件生成对抗网络的 SAR 图像先验单卫星光学图像去雾”，发表于 *IEEE 计算机应用冬季会议论文集愿景*，2020，页。1806–1813。

- [38] 刘Y., 朱L., 裴S.裴, H.付, J.秦, Q.张, L.万, W. Feng, “从合到成真：与未标记的真实数据合作进行图像去雾”，《ACM 国际多媒体会议记录》，2021 年，第 50–58 页。[39] C. O. Ancuti, C. Ancuti, R. Timofte 和 C. De Vleeschouwer, “O-haze: 具有真实雾霾和无雾户外图像的去雾基准”，载于 IEEE 计算机视觉和模式识别研讨会会议记录，2018 年，第 754–762 页。[40] C. Ancuti, C. O. Ancuti, R. Timofte 和 C. De Vleeschouwer, “I-haze: 具有真实雾霾和无雾室内图像的除雾基准”，智能视觉系统高级概念：第 19 届国际会议，ACIVS 2018，普瓦捷，法国，2018 年 9 月 24–27 日，2018 年第 19 期会议记录，第 620–631 页。[41] C. Li, H. Zhou, Y. Liu, C. Yang, Y. Xie, Z. Li 和 L. Zhu, “检测友好的去雾：真实世界模糊场景中的对象检测”，IEEE 模式分析和交易学报机器智能，2023。[42] S. Lee, T. Son 和 S. Kwak, “Fifo: 学习雾场景分割的雾不变特征”，《IEEE 计算机视觉和模式识别会议论文集》，2022 年，第 18 911–18 921 页。[43] B. Cai, X. Xu, K. Jia, C. Qing 和 D. Tao, “Dehazenet: 用于单图像除雾的端到端系统”，IEEE Transactions on Image Processing，卷。25, 没有。11, 第 5187–5198 页, 2016 年。[44] Y.-F. 刘, D.-W. 下巴, S.-C. 黄和 J.-N. Hwang, “Desnownet: 用于除雪的上下文感知深度网络”，IEEE 图像处理交易，卷。27, 没有。6, 第 3064–3073 页, 2018 年。[45] Y. Cui, W. Ren 和 A. Knoll, “用于图像恢复的全内核网络”，载于 AAAI 人工智能会议记录，第 1 卷。38, 没有。2, 2024 年, 第 1426–1434 页。[46] Y. Cui, W. Ren, X. Cao 和 A. Knoll, “图像恢复的焦点网络”，IEEE 国际计算机视觉会议论文集，2023 年，第 13 001–13 011 页。[47] Z. Tu, H. Talebi, H. 张、F. Yang, P. Milanfar, A. Bovik 和 Y. Li, “Maxim: 用于图像处理的多轴 MLP”，载于 IEEE 计算机视觉和会议会议记录模式识别，2022 年，第 5769–5780 页。[48] H. Zhang, Y. Dai, H. Li 和 P. Koniusz, “用于图像去模糊的深度堆叠分层多块网络”，载于 IEEE 计算机视觉和模式识别会议论文集，2019 年。[49] C.-L. Guo, Q. Yan, S. Anwar, R. Cong、W. Ren 和 C. Li, “具有传输感知 3d 位置嵌入的图像去雾变压器”，发表于 IEEE 计算机视觉和模式识别会议论文集，2022 年，第 5812–5820 页。[50] S. Chen, T. Ye, Y. Liu, T. Liao, J. Jiang, E. Chen 和 P. Chen, “Msp-former: 用于单图像除雪的多尺度投影变换器”，IEEE 国际声学会会议，语音和信号处理，2023 年，第 1–5 页。[51] Y. Song, Z. He, H. Qian 和 X. Du, “用于单图像去雾的视觉变换器”，arXiv 预印本 arXiv:2204.03883, 2022。[52] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin 和 B. Guo, “Swin 变压器：使用移位窗口的分层视觉变压器”，《IEEE 国际会议录》计算机视觉会议，2021 年，第 10 012–10 022 页。[53] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang 和 L. Shao, “学习丰富的特征以实现真实图像恢复和增强”，欧洲计算机视觉会议记录，2020 年，第 492–511 页。[54] Q. 韩, Z. 范, Q. 戴, L. 孙, M.-M. 程, J. 刘, 和 J. Wang, “论局部注意力与动态深度卷积之间的联系”，国际学习表征会议，2021 年。[55] S. Zhou, J. Zhang, J. Pan, H. Xie, W. Zuo 和 J. Ren, “用于视频去模糊的时空滤波器自适应网络”，发表于 IEEE 国际计算机视觉会议论文集，2019。[56] W. Wen, W. Ren, Y. Shi, Y. Nie, J. Zhang 和 X. Cao, “通过时空对齐网络实现视频超分辨率”，IEEE Transactions on Image Treatment, 卷。31, 第 1761–1773 页, 2022 年。[57] X. Mao, Y. Liu, W. Shen, Q. Li 和 Y. Wang, “单图像去模糊的深度残差傅里叶变换”，arXiv preprint arXiv:2111.11745, 2021。[58] B.K. Gunturk 和 X. Li, 图像恢复：基础知识和进展。CRC出版社，2012年。[59] M. R. Banham 和 A. K. Katsaggelos, “数字图像恢复”，IEEE 信号处理杂志，卷。14, 没有。2, 第 24–41 页, 1997 年。
- [60] H. Yu, N. Cheng, M. Zhou, J. Huang, Z. Xiao 和 F. Zhu, “图像去雾的频率和空间双重指导”，载于欧洲计算机视觉会议记录，2022 年，页数 181–198。[61] 周明, 黄杰, C.-L. 郭和 C. Li, “Fourmer: 一种高效的图像恢复全局建模范式”，国际机器学习会议，2023 年，第 42 589–42 601 页。[62] I. W. Selesnick, R. G. Baraniuk 和 N. C. Kingsbury, “双树复小波变换”，IEEE 信号处理杂志，卷。22, 没有。6, 第 123–151 页, 2005 年。[63] N. Park 和 S. Kim, “视觉变形器如何工作？”在国际学习表征会议，2022 年。[64] D. P. Kingma 和 J. Ba, “Adam: 一种随机优化方法”，arXiv 预印本 arXiv: 1412.6980, 2014 年。[65] I. Loshchilov 和 F. Hutter, “Sgdr: 热重启的随机梯度下降”，arXiv 预印本 arXiv: 1608.03983, 2016 年。[66] Z. Wang, A. C. Bovik, H. R. Sheikh 和 E. P. Simoncelli, “图像质量评估：从错误可见性到结构相似性”，IEEE 图像处理汇刊，卷。13, 没有。4, 第 600–612 页, 2004 年。[67] R. Zhang, P. Isola, A.A. Efros, E. Shechtman 和 O. Wang, “深度特征作为感知度量的不合理有效性”，载于 IEEE 计算机视觉和模式识别会议论文集，2018 年，第 586 页 - 595。[68] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhu, Y. Xie 和 L. Ma, “紧凑单图像去雾的对比学习”，载于 IEEE 计算机会议论文集视觉与模式识别，2021，第 10 551–10 560 页。[69] T. Ye, Y. 张, M. Jiang, L. Chen, Y. Liu, S. Chen 和 E. Chen, “图像去雾的感知和建模密度”，欧洲计算机视觉会议论文集，2022 年，第 130–145 页。[70] B. Li, X. Peng, Z. Wang, J. Xu 和 D. Feng, “Aod-net: 一体化去雾网络”，载于 IEEE 国际计算机视觉会议论文集，2017 年。[71] Y. Cui, W. Ren, X. Cao 和 A. Knoll, “通过频率选择进行图像恢复”，IEEE 模式分析和机器智能汇刊，2023 年。[72] C. O. Ancuti, C. Ancuti, M. Sbert 和 R. Timofte, “密集雾霾：使用密集雾霾和无雾图像进行图像去雾的基本”，IEEE 国际图像处理会议，2019 年，第 147 页。1014–1018。[73] C. O. Ancuti, C. Ancuti 和 R. Timofte, “Nh-haze: 具有非均匀有雾和无雾图像的图像去雾基准”，载于 IEEE 计算机视觉和模式识别研讨会会议记录，2020 年。[74] X. Chen, Y. Li, L. Dai, C. Kong, “单遥感卫星图像去雾的混合高分辨率学习”，IEEE 地球科学与遥感快报，卷。19, 第 1–5 页, 2021 年。[75] Y. Li 和 X. Chen, “一种用于遥感图像除雾的从粗到细的两级注意力网络”，IEEE 地球科学与遥感快报，卷。18, 没有。10, 第 1751–1755 页, 2020 年。[76] Y. Cheng, J. Zhan, S. He, J. Dong 和 Y. Du, “物理感知单图像去雾的课程对比正则化”，载于 IEEE 计算机视觉和模式识别会议论文集，2023 年，第 5785–5794 页。[77] K. Chi, Y. Yuan 和 Q. Wang, “Trinity-net: 基于梯度引导 swintransformer 的遥感图像去雾及其他技术”，IEEE 地球科学与遥感学报，2023 年。[78] A. Kulkarni, S. S. Phutke 和 S. Murala, “用于多天气图像恢复的统一变压器网络”，欧洲计算机视觉会议，2022 年，第 344–360 页。[79] Y. Jin, B. Lin, W. Yan, Y. Yuan, W. Ye 和 R. T. Tan, “使用引导 apsf 和梯度自适应卷积增强夜间雾霾图像的可见性”，《ACM 国际多媒体会议录》，2023 年，第 2446–2457 页。[80] J. Zhang, Y. Cao 和 Z. Wang, “基于新成像模型的夜间雾霾去除”，IEEE 国际图像处理会议，2014 年，第 4557–4561 页。[81] Y. Li, R. T. Tan 和 M. S. Brown, “利用发光和多种光色去除夜间雾霾”，《IEEE 国际计算机视觉会议记录》，2015 年，第 226–234 页。[82] J. 张, Y. 曹, S. Fang, Y. Kang 和 C. Wen Chen, “使用最大反射率先验快速去除夜间图像”，《IEEE 计算机视觉和模式识别会议论文集》，2017 年。

- [83] 王涛、陶刚、陆伟、张凯、罗伟、张晓、陆涛，“利用分层对比学习在雾霾天气中恢复视力”，模式识别，第1卷。145, p. 109956, 2024.[84] C. Ancuti, C. O. Ancuti, C. De Vleeschouwer 和 A. C. Bovik, “Night-time dehazing by fusion”, IEEE 国际图像处理会议, 2016 年, 第 2256–2260 页。[85] J.-Y. Zhu, T. Park, P. Isola 和 A. A. Efros, “使用循环一致对抗网络进行不成对的图像到图像转换”，载于 IEEE 国际计算机视觉会议论文集, 2017 年, 第 2223–2232 页。[86] W.-T. 陈 H.-Y. 方, J.J. 丁, C.-C. 蔡和 S.-Y. Kuo, “Jstasr: 基于改进的部分卷积和遮蔽效应去除的联合尺寸和透明度感知除雪算法”，欧洲计算机视觉会议记录, 2020 年, 第 754–770 页。[87] Y.-F. 刘, D.-W. 下巴, S.-C. 黄和 J.-N. Hwang, “Desnownet: 用于除雪的上下文感知深度网络”，IEEE 图像处理交易, 卷。27、没有。6, 第 3064–3073 页, 2018 年。[88] R. Li, R. T. Tan 和 L.-F. Cheong, “使用架构搜索一次性消除恶劣天气”，载于 IEEE 计算机视觉和模式识别会议论文集, 2020 年。[89] B. Cheng, J. Li, Y. Chen, S. Zhang 和 T. Zeng, “用于图像除雪的雪掩模引导自适应残差网络”，arXiv 预印本 arXiv: 2207.04754, 2022。[90] H. Zhang, V. Sindagi 和 V.M. Patel, “使用条件生成对抗网络进行图像去雨”，IEEE Transactionson 视频技术电路和系统, 卷。30, 没有。11, 第3943-3956页, 2019年。[91] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding 和 J. Paisley, “通过深度细节网络从单幅图像中去除雨水”，载于 IEEE 计算机视觉和模式识别会议论文集, 2017。[92] X. Fu, J. Huang, X. Ding, Y. Liao 和 J. Paisley, “Clearingthe skys: 用于单图像除雨的深层网络架构”，IEEE Transactions on Image Processing, 卷。26、没有。第 6 页。2944–2956, 2017。[93] W. Wei, D. Meng, Q. Zhao, Z. Xu 和 Y. Wu, “用于图像除雨的半监督迁移学习”，载于 IEEE 计算机视觉与模式识别会议论文集, 2019 年。[94] R. Yasarla 和 V. M. Patel, “不确定性引导的多尺度残差学习 - 使用循环旋转 cnn 进行单图像去雨”，载于 IEEE 计算机视觉和模式识别会议论文集, 2019 年。[95] X. Li, J. Wu, Z. Lin, H. Liu 和 H. Zha, “用于单图像去雨的循环挤压和激励上下文聚合网络”，欧洲计算机视觉会议记录, 2018 年。[96] D. Ren, W. Zuo, Q. Hu, P. Zhu 和 D. Meng, “渐进式图像去雨网络: 更好、更简单的基线”，载于 IEEE 计算机视觉和模式识别会议论文集, 2019 年。[97] K. Jiang, Z. Wang, P. Yi, C. Chen, B. Huang, Y. Luo, J. Ma 和 J. Jiang, “用于单图像脱雨的多尺度渐进融合网络”，载于 IEEE 计算机视觉和模式识别会议论文集, 2020 年。[98] A. Abuolaim, M. Afifi 和 M. S. Brown, “改进单图像散焦去模糊: 双像素图像如何帮助完成多任务学习”，《IEEE 计算机视觉应用冬季会议记录》，2022 年, 第 1231 页 1239。[99] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo 和 S. Yan, “单幅图像的深度联合训练检测和去除”，《IEEE 计算机视觉与模式识别会议论文集》，2017 年。[100] H. Zhang, V. Sindagi 和 V.M. Patel, “使用条件生成对抗网络进行图像去雨”，IEEE 视频技术电路与系统交易, 2019 年。[101] Y. Li, R. T. Tan, X. Guo, J. Lu 和 M.S. Brown, “使用层先验进行雨条纹去除”，载于 IEEE 计算机视觉和模式识别会议论文集, 2016 年。[102] J. Rim, G. Kim, J. Kim, J. Lee, S. Lee 和 S. Cho, “用于学习图像去模糊的真实模糊合成”，欧洲计算机视觉会议记录, 2022 年, 第 487–503 页。[103] X. Tao, H. Gao, X. Shen, J. Wang 和 J. Jia, “用于深度图像去模糊的尺度循环网络”，载于 IEEE 计算机视觉和模式识别会议论文集, 2018 年。[104] J. Zhou, P. Wang, F. Wang, Q. Liu, H. Li 和 R. Jin, “Elsa: 增强视觉变压器的局部自注意力”，arXiv preprintarXiv: 2112.12786, 2021。

- [105] X. Wang, R. Girshick, A. Gupta 和 K. He, “非局部神经网络”，载于 IEEE 计算机视觉和模式识别会议论文集, 2018 年, 第 7794–7803 页。[106] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu 和 Y. Wei, “可变形卷积网络”，IEEE 国际计算机视觉会议论文集, 2017 年, 第 764 页–773。[107] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu 和 C. Xu, “Ghostnet: 廉价操作带来更多功能”，载于 IEEE 计算机视觉和模式识别会议论文集, 2020 年, 第 1580–1589。



崔宇宁 (IEEE 学生会员) 获得工程学士学位2016年毕业于中南大学, 获工学硕士学位。2018年毕业于国防科技大学, 获博士学位。目前正在攻读博士学位。慕尼黑工业大学计算、信息和技术学院机器人、人工智能和实时系统系主任博士学位。他的研究兴趣在于图像修复。

任文琪 (IEEE 会员) 于2017年获得天津大学博士学位。2015年至2016年获得国家留学基金委支持, 与杨明湖三教授联合培养博士。加州大学默塞德分校电气工程与计算机科学系学生。现任中国深圳中山大学深圳校区网络科学与技术学院教授。他的研究兴趣包括图像处理和相关的高级视觉问题。2017年获得腾讯犀牛鸟精英研究生计划奖学金, 2018年获得MSRA星轨计划奖学金。

曹晓春 (IEEE 高级会员) 中国北京航空航天大学计算机科学专业和硕士学位, 以及博士学位。美国中佛罗里达大学计算机科学博士学位。曾任中国深圳中山大学深圳校区网络科学与技术学院教授。毕业后, 他在 ObjectVideo Inc. 工作了大约三年, 担任研究科学家。2008年至2012年, 任天津大学教授。他是 IET 院士。他是 IEEE 图像处理汇刊的编委会成员。他的论文被提名为中佛罗里达大学大学级优秀论文奖。

Alois Knoll (IEEE 院士) 于 1985 年在德国斯图加特大学获得电气/通信工程文凭 (硕士学位), 并于 1985 年在德国斯图加特大学获得博士学位。1988 年获得德国柏林工业大学计算机科学专业 (最高荣誉) 学位。1993 年之前, 他在柏林工业大学计算机科学系任教。1993 年, 他加入德国比勒费尔德大学, 担任教授, 并担任该系主任。2001 年之前在技术信息学研究组工作。2001 年起担任德国慕尼黑工业大学 (TUM) 信息学系教授。他的研究兴趣包括认知、医疗和基于传感器的机器人、多代理系统、数据融合、自适应系统、多媒体信息检索、应用于汽车软件和电动交通的嵌入式系统的模型驱动开发, 以及机器人和交通仿真系统。