arXiv:2305.05146v1 [cs.CV] 9 May 2023

一种用于精准图像复原的山形单阶段网络

Hu Gao[1], Jing Yang[1], Ying Zhang[1], Ning Wang[1], Jingfan Yang[1], Depeng Dang[1*]

[1]北京师范大学人工智能学院，北京，中国，北京市，100000。

*通讯作者。电子邮箱：ddepeng@bnu.edu.cn；贡献作者：gao h@mail.bnu.edu.cn；yangjing2019@mail.bnu.edu.cn；202120081049@mail.bnu.edu.cn；wangningbnu@mail.bnu.edu.cn；yangjingfan@mail.bnu.edu.cn；

摘要

图像复原旨在从损坏的输入图像中获取高质量图像，例如进行去模糊和去雨。在图像复原中，通常需要在空间细节与上下文信息之间保持复杂的平衡。尽管多阶段网络能够最优地平衡这些相互竞争的目标并取得显著效果，但也会增加系统复杂度。本文提出一种基于简洁 U-Net架构的山形单阶段设计，通过移除或替换不必要的非线性激活函数，以较低的系统复杂度实现上述平衡。具体而言，我们提出一种特征融合中间件（FFM）机制，作为编码器-解码器架构层级之间的信息交换组件。该机制将上层信息无缝整合至相邻的下层，并依次向下传递至最低层。最后，所有信息都融合至原始图像分辨率处理层。这既保留了空间细节，又整合了上下文信息，从而确保实现高质量图像复原。此外，我们提出一种多头注意力中间块（MHAMB），作为编码器与解码器之间的桥梁，以捕获更多全局信息，并突破卷积神经网络的感受野限制。大量实验表明，我们的方法 M3SNet 在多种图像复原任务（如图像去雨和去模糊）中，计算成本不到以往最先进模型的一半，性能却优于这些模型。

关键词： 图像复原，单阶段，特征融合中间层，多头注意力中间块

1 引言

图像退化是图像采集过程中常见的问题，成因多种多样，例如相机的局限性、环境条件以及人为因素。比如，光圈较小、传感器尺寸较小且动态范围有限的智能手机相机，可能会因身体移动造成设备抖动，从而拍摄出模糊且有噪声的图像。同样，在恶劣天气条件下拍摄的图像也可能受到雾和雨的影响。大多数经典图像复原任务都可以表述为:

$$\mathbf{L} = \mathbf{D}(\mathbf{H}) + \gamma \tag{1}$$

其中，$\mathbf{L}$ 表示观测到的低质量图像，$\mathbf{H}$ 表示与之对应的高质量图像，而$\mathbf{D}(\cdot)$ γ分别表示成像和传输过程中的退化函数与噪声。当$\mathbf{D}(\cdot)$发生变化时，这一公式可表示不同的图像复原任务。例如，若$\mathbf{D}(\mathbf{H})$ $= \mathbf{H}t - At + A$，则对应去雨或去雾任务，其中 A 是全局大气光



图 1：M3SNet 在各种图像复原任务上的可视化结果。左：退化图像。右：M3SNet 的预测结果。从上到下分别为图像去模糊和图像去雨任务。

而 t 是定义如下的传输矩阵：

$$t = e^{-\alpha d} \tag{2}$$

其中， α 是大气散射系数， d 是物体与相机之间的距离。

图像复原旨在从退化图像中恢复高质量清晰图像 $\mathbf{H}$ 。$\mathbf{L}$由于对于任何原始输入都存在许多候选结果，因此这是一个高度病态的问题。为了将无限的可行候选空间限制在自然图像范围内，传统方法 [1–7] 会针对给定类型的复原问题，明确设计适当的先验，例如领域相关先验和任务相关先验。然后，通过求解最大后验（MAP）问题，可以获得潜在的高质量图像：

$$\hat{\mathbf{H}} = \arg\max_{\mathbf{H}} \log P(\mathbf{L}|\mathbf{H}) + \log P(\mathbf{H}) \tag{3}$$

其中， $P(\mathbf{L}|\mathbf{H})$表示给定清晰图像$\mathbf{H}$时观测到退化图像$\mathbf{L}$ 的概率，而 $P(\mathbf{H})$表示清晰图像$\mathbf{H}$的先验分布。这也可以表示为约束最大似然估计：

$$\hat{\mathbf{H}} = \arg\min_{\mathbf{H}} \|\mathbf{L} - \mathbf{D}(\mathbf{H})\|^2 + \lambda\Psi(\mathbf{H}) \tag{4}$$

其中，保真项 $|\mathbf{L} - \mathbf{D}(\mathbf{H})|^2$ 是似然 $P(\mathbf{L}|\mathbf{H})$的近似，而正则化项 $\lambda\Psi(\mathbf{H})$表示潜在图像$\mathbf{H}$的先验或对解的约束。其目的是在考虑施加于解的先验知识或约束的同时，表达重建图像与原始输入之间的保真度

尽管为图像复原设计有效的先验具有挑战性，而且这些先验可能缺乏泛化能力。借助大规模数据，卷积神经网络（CNN） [8–14]和 Transformer [15–18] 等深度模型已成为首选，因为它们能够通过捕捉自然图像统计特性，隐式学习更具泛化性的先验，并在图像复原方面达到最先进水平（SOTA）。这些深度学习模型相较于传统复原方法的性能提升，主要归功于其模型设计，其中包含许多用于图像复原的网络模块和功能单元，例如递归残差学习 [19], 、Transformer [15, 16, 18], 、编码器-解码器 [12, 13, 20],、多尺度模型 [21, 22], 和生成模型 [23–25]。

然而，这些用于低层视觉问题的模型大多基于单阶段设计，忽略了空间细节与上下文信息之间的交互。为了解决这一局限， [8–11] 提出了一种多阶段架构：首先通过编码器-解码器架构学习上下文特征，随后将其与高分辨率分支融合，以保留局部信息。尽管该方法性能良好，但它需要在后续阶段对前一阶段的结果进行细化，导致系统复杂度较高。

根据所呈现的信息，一个自然而然的问题是：是否可以采用单阶段架构来降低系统复杂度?

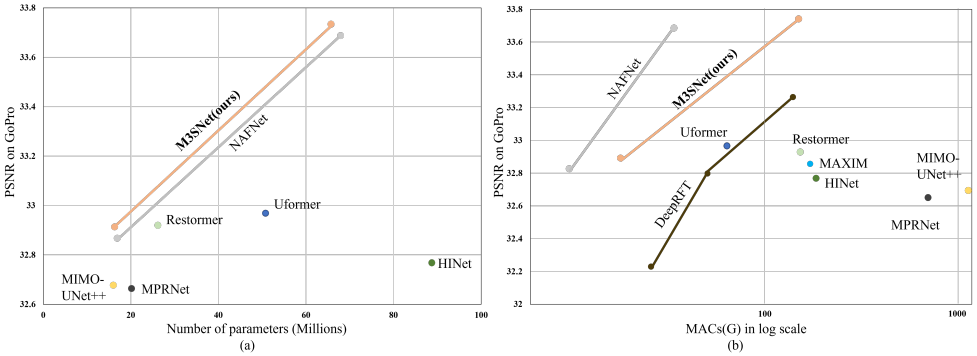

图 2：图像去模糊任务中的峰值信噪比与计算成本对比。在不同参数规模下，我们的模型均达到最先进水平。此外，我们的模型所需的乘加运算次数（MACs）相对较少。

并在保持最先进水平性能的同时，实现与多阶段架构相同的空间细节和上下文信息之间的平衡。为实现这一目标，我们提出了一种山形单阶段图像复原架构，称为 M3SNet，该架构包含几个关键组件。1）。我们采用 NAFNet [12] 作为基线架构，并着重修改网络模型，以实现多阶段功能。通过消除多阶段之间的信息传递，并从网络结构中移除非线性激活函数，我们能够降低系统的复杂度。2）。我们引入了一种特征融合中间件（FFM）机制，以促进不同层的编码器与解码器模块之间的多尺度信息融合，从而获取更多上下文信息。此外，该方法还支持对原始图像分辨率进行调整，进而有助于保留空间细节。3）。多头注意力中间模块（MHAMB）作为连接编码器与解码器的桥梁，突破了卷积神经网络的感受野限制，并能够捕获更多全局信息。

本文的主要贡献如下：

1. 一种新颖的单阶段方法，能够生成上下文丰富、空间精确的输出，其效果可媲美多阶段架构。我们的架构采用单阶段设计，避免了在不同阶段之间传递信息的需求，因此降低了系统复杂度。2. 一种特征融合中间件机制，可在多个尺度间交换信息，同时保留从输入图像到输出图像的精细细节。3. 多头注意力中间块能够捕获更多全局信息。4. 我们展示了 M3SNet 的通用性：在 6 个合成和真实世界数据集上的多种复原任务（图像去雨和去模糊）中取得新的最先进结果，同时保持较低的复杂度（见图2）。此外，我们还提供了详细分析、定性结果和泛化测试。

2 相关工作

图像退化是由相机设备和各种环境因素引起的常见现象。根据具体的退化现象，人们提出了不同的图像复原任务，例如去模糊和去雨。早期的图像复原工作主要依赖于人工构建先验知识，例如全变分和自相似性 [1–7]。随着深度学习的兴起，基于数据的方法（如卷积神经网络 [8, 26–30] 和 Transformer [16–18, 31, 32] ）凭借出色的性能，已成为图像复原的主流方法。根据架构设计，这些方法可分为单阶段或多阶段方法。

2.1 单阶段架构。

近年来，大多数图像复原研究都聚焦于单阶段架构。在这些架构中，主要包括基于编码器-解码器的 U-Net [8, 12, 13,18、25、33–35]以及双网络结构 [36–40] 。

编码器-解码器方法。近年来，编码器-解码器凭借捕获多尺度信息的能力，受到了图像复原领域研究人员的广泛关注。为了构建用于图像复原的有效且高效的 Transformer 架构， [18] 提出了一种新颖的局部增强窗口和多尺度复原调制器，以构建分层编码器-解码器网络。 [30] 利用选择性核特征融合，实现不同尺度间的信息交换以及基于注意力的信息聚合。 [22] 将强化-操作-相减增强策略融入解码器，开发了一种简单而有效的增强解码器，以渐进式复原无雾图像。通过消除或替换非线性激活函数， [12] 建立了一个简单的基线，在减少计算资源需求的同时取得了可衡量的效果。

双网络方法。 双网络架构由两个并行分支构成，分别从输入中估计目标信号的结构分量和细节分量。随后，根据特定的任务制定模块将这些分量组合起来，重建最终结果。该架构最初由 [41] 提出，此后启发了许多后续研究，包括图像去雾 [36–38], 、图像去雨 [42], 、图像去噪 [39],以及图像超分辨率/去模糊 [40]。事实证明，双网络方法能够有效解决各种低层视觉问题：它有助于更好地分离结构信息和细节信息，从而提升准确性和计算效率。此外，该架构具有灵活性，可适用于不同类型的数据，因此成为图像复原领域的热门选择。

尽管这些网络已经取得了显著成就，但在图像恢复过程中，如何有效平衡保留空间细节与上下文信息这两个相互竞争的目标，仍然是一项挑战。

2.2 多阶段架构。

研究表明，在高级视觉问题中，多阶段网络比其单阶段对应网络更为有效 [43–46]。近年来，一些研究尝试 [8, 10, 11, 47–50] 将多阶段网络应用于图像复原。这些方法旨在将图像复原过程分解为若干易于处理的阶段，使轻量级子网络能够逐步复原出清晰图像。这种方法有助于各阶段的子网络分别捕获空间细节和上下文信息。为了避免在每个阶段使用相同的子网络可能产生的次优结果，研究人员提出了一种监督注意力机制，并采用不同的子网络结构 [8]。此外， [49] 还提出了一种新颖的自监督事件引导深层分层多块网络，通过由细到粗的层次化局部表征来处理模糊图像和视频。然而，由于后续阶段需要对前一阶段的结果进行细化，这种方法增加了系统的复杂性。

3 方法

我们的主要目标是构建一种单阶段网络架构，在使用较少计算资源的同时，平衡对空间细节和上下文信息的需求，从而高效应对图像复原这一具有挑战性的任务。如图3所示，M3SNet基于U-Net架构构建。从图中可以明显看出，与传统U-Net网络不同，我们将其架构反转，并引入了两个关键组件：(a) 特征融合中间件（FFM）和(b) 多头注意力中间块（MHAMB）。该模型的架构呈山形，我们将图像复原过程比作登山。

整体流程。 给定退化图像$\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$，M3SNet首先应用一个$3\times 3$卷积层来提取浅层特征图$\mathbf{F_0} \in \mathbb{R}^{H\times W\times C}$($H, W, C$分别表示特征图的高度、宽度和通道数)。接下来，这些浅层特征$\mathbf{F_0}$ 经过 $4-level$ 编码器-解码器和一个多头注意力中间块，得到深层特征$\mathbf{F_{DF}} \in \mathbb{R}^{H\times W\times C}$。每一层都在编码器和解码器之间包含多个特征融合中间件，以捕获多尺度信息并保留空间细节。最后，我们对深层特征$\mathbf{F_{DF}}$ 应用卷积，生成残差图像$\mathbf{R} \in R^{H\times W\times 3}$，并将其与退化图像相加ˆ，得到复原图像：$\mathbf{I} = \mathbf{R} + \mathbf{I}$。我们使用峰值信噪比损失优化所提出的网络：

$$PSNR = 10 \cdot log_{10} \cdot \frac{(2^n-1)^2}{||\hat{\mathbf{I}} - \dot{\mathbf{I}}||^2} \tag{5}$$

˙其中，$\mathbf{I}$ 表示真实图像。

3.1 特征融合中间件（FFM）

通过在初始阶段引入编码器和解码器网络，并在最后阶段采用以原始图像输入分辨率运行的网络，该

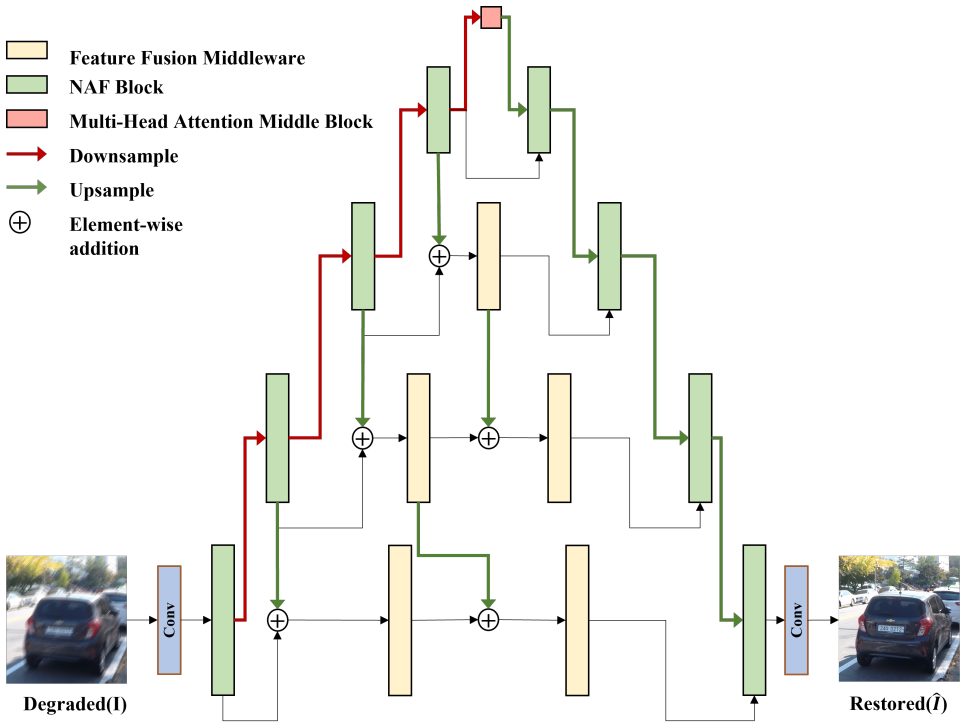

图 3：用于图像复原的 M3SNet 架构。

多阶段网络能够生成空间细节准确、上下文信息可靠的高质量图像。然而，这一过程的后续阶段需要修正前一阶段的结果，这会使系统略微复杂一些。单阶段网络的复杂度相对较低，但可能难以有效平衡空间细节与上下文信息。因此，我们正在探索一种用于特征融合的中间件机制，使简单的单阶段 U-Net架构能够实现与多阶段架构相同的功能，同时不牺牲表达的流畅性，也不损害原意。

如图所示 4(a)，特征融合中间件（FFM）是一个集成了上采样和特征融合的无非线性激活块（NAFBlock）。我们在编码器-解码器架构层级之间引入了 FFM，以将上层信息整合到相邻的下层中。这种整合按顺序进行，从最高层到最低层，直至所有信息都融合到原始图像分辨率处理层级。该方法增强了网络捕获和融合多尺度特征的能力，涵盖低层的简单模式（如角点或边缘/颜色连接），以及高层的复杂特征（如显著变化和特定物体）。因此，这种结构在整合上下文信息的同时保留了空间细节，最终确保实现高质量图像复原。形式上，令 $\mathbf{FE_i} \in \mathbb{R}^{\frac{H}{i^2} \times \frac{W}{i^2} \times i^2 C}$ 表示第 i 层编码器（$i = 1, 2, 3, 4$）的输出。每一层的特征融合信息 $FFM_{s,i}$ 定义为

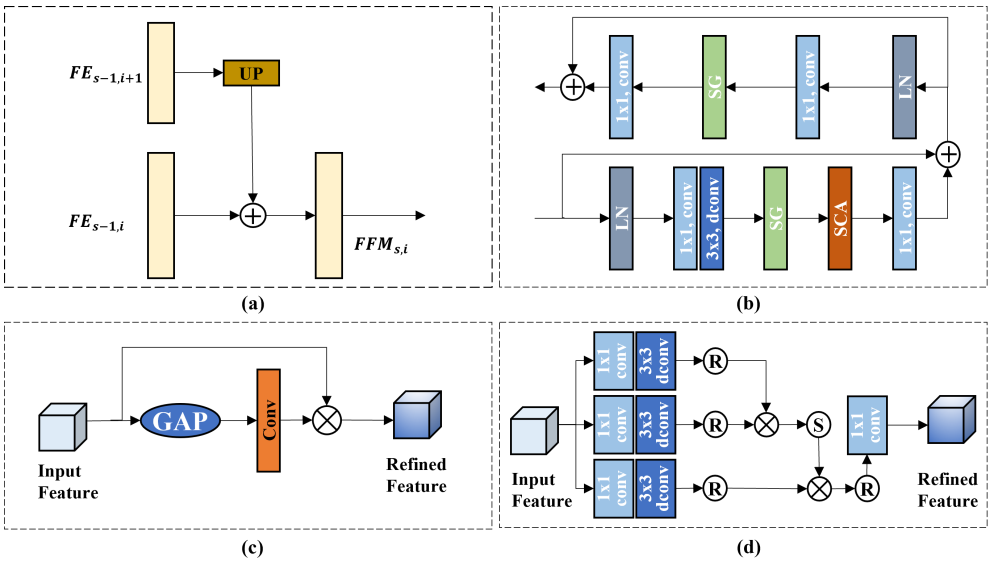

图4：(a) 特征融合中间件（FFM），能够在保留精细细节的同时实现多尺度间的信息交换。(b) 非线性激活自由块（NAFBlock）的架构。(c) 简化通道注意力（SCA）。(d) 多头注意力中间块（MHAMB），能够捕获更多全局信息。

as:

$$FFM_{1,i} = H_{Naf_{1,i}}(FE_i \oplus UP(FE_{i+1})) \tag{6}$$

$$FFM_{s,i} = H_{Naf_{s,i}}(UP(FFM_{s-1,i+1}) \oplus FFM_{s-1,i}) \tag{7}$$

其中，$\oplus$表示逐元素相加，$UP(\cdot)$表示上采样操作，而 $H_{Naf_{s,i}}(\cdot)$表示第 i 层中的第 s 个FFM。

这种设计有两大优势。首先，特征融合中间件整合多尺度信息，使网络模型能够捕获丰富的上下文信息。其次，1_{th} 层中的特征融合中间件直接在原始图像分辨率下运行，不采用任何下采样操作，因此网络模型能够获取高分辨率特征的详细空间信息。

NAFBlock。NAFBlock 是 U-Net 网络的一种变体，它通过替换或移除非线性激活函数来简化系统。图 4(b) 展示了如何获得输出Y （输入为 X ），使用层归一化、卷积、简单门控和简化通道注意力。表示如下：

$$X_1 = X + C_1(SCA(SG(C_3(C_1(LN(X)))))) \tag{8}$$

$$Y = X_1 + C_1(SG(C_1(LN(X_1)))) \tag{9}$$

$$SG = X_{f1} \times X_{f2} \tag{10}$$

其中 C_1是 1×1卷积， C_3是 3×3深度卷积，GAP 是自适应平均池化， X_{f1}, $X_{f2} \in R^{H \times W \times \frac{C}{2}}$通过将 X_{f3}划分为通道维度得到，并且 $SCA(\cdot)$ 如图 4(c) 所示。

最后，通过这一单阶段架构获得深度特征$\mathbf{F_{DF}}$，如下所示：

$$FD_i = H_{Naf_{\text{-}1,i}}(FD_{i+1} + FFM_{\text{-}1,i}) \tag{11}$$

$$F_{DF} = FD_1 \tag{12}$$

其中， FD_i 是第 i级解码器的输出，而-1表示这是该级别最后一个特征融合中间件。

3.2 多头注意力中间块（MHAMB）

Transformer 模型 [16–18] 因具备捕获全局信息的能力，在图像复原任务中日益受到关注，近年来其使用频率不断提高便是明证。正如公式 15所示，全局范围内的计算复杂度与 token 数量呈二次关系，因此不适用于表示高分辨率图像。

$$\mathcal{O}_{MSA} = 4hwC^2 + 2(hw)^2C \tag{13}$$

为缓解这一问题， [16, 17, 32] 等人提出了多种降低复杂度的方法。本文提出一种多头注意力中间块（MHAMB），作为编码器-解码器的桥梁，如图 4(d)所示。
MHAMB 利用全局自注意力处理并整合由编码器最后一层生成的特征图信息。该方法在处理大图像时尤其高效，因为卷积会对空间进行下采样，而注意力机制则能有效处理较低分辨率，从而取得更好的性能。对于编码器最后一层的输出FE_4，我们的 MHAMB 首先应用一个 1×1 卷积和一个 3×3 深度卷积，以生成 $query$、key和 $value$矩阵 $\mathbf{Q} \in \mathbb{R}^{H \times W \times C}$、$\mathbf{K} \in \mathbb{R}^{H \times W \times C}$和$\mathbf{V} \in \mathbb{R}^{H \times W \times C}$，如下所示：

$$\begin{aligned} Q &= C_3(C_1(FE_4)) \\ K &= C_3(C_1(FE_4)) \\ V &= C_3(C_1(FE_4)) \end{aligned} \tag{14}$$

然后，我们将$\mathbf{Q}$、$\mathbf{K}$、$\mathbf{V}$重塑为$\hat{}\mathbf{Q} \in \mathbb{R}^{(HW) \times \frac{C}{h} \times h}$、$\hat{}\mathbf{K} \in \mathbb{R}^{(HW) \times \frac{C}{h} \times h}$、$\hat{}\mathbf{V} \in \mathbb{R}^{(HW) \times \frac{C}{h} \times h}$，其中 h 是头的数量。接下来，注意力矩阵由自注意力机制计算如下：

$$Attention(\hat{Q}, \hat{K}, \hat{V}) = SoftMax(\frac{\hat{Q}\hat{K}}{\beta})\hat{V} \tag{15}$$

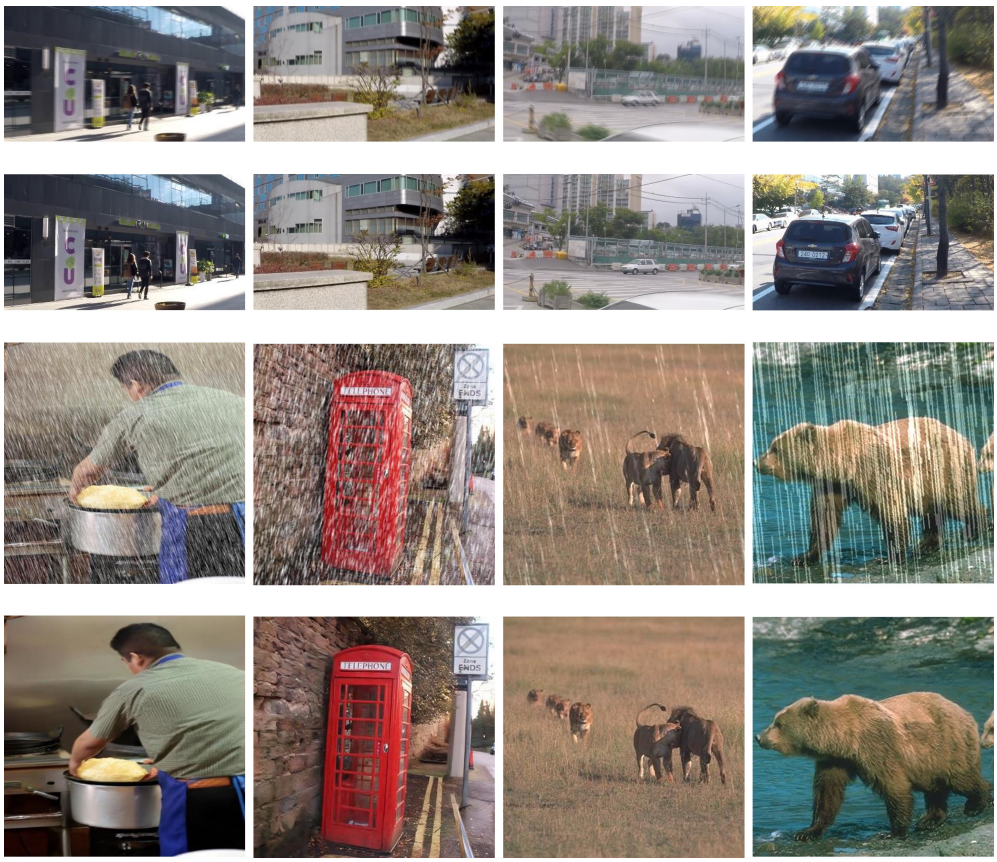

图5：M3SNet 在各种图像复原任务上的可视化结果。对于每组图像对，上方为退化图像，下方为 M3SNet 的预测结果。

其中，β 是学习缩放参数，用于在应用 softmax函数之前调整 Q 和 K 的点积ˆ ˆ的幅值。最后，我们将注意力矩阵重塑回其原始维度 $\mathbb{R}^{H\times W\times C}$，并应用 1×1卷积。随后，将所得输出与 FE_4 相加，并传递给解码器。这使我们能够捕获更多信息并提升模型的整体性能，正如以下实验所示。

4 实验

我们在图像复原的两项任务的基准数据集上评估了所提出的 M3SNet，包括 (a) 图像去雨和 (b) 图像去模糊。图5 展示了我们的方法预测的一些图像。我们的模型复原出了更清晰、接近真实值的图像。

表 1：图像去雨数据集描述。

数据集	训练样本	测试样本	测试集重命名
雨14000 [51]	11200	0	-
雨1800 [52]	1800	0	-
雨12 [56]	12	0	-
雨800 [53]	700	98	Test100
雨1200 [54]	0	1200	Test1200
Rain100H [52]	0	100	Rain100H
Rain100L [52]	0	100	Rain100L

表 2：图像去模糊数据集描述。

数据集	训练样本	测试样本	测试集重命名
GoPro [55]	2130	1111	GoPro
HIDE [57]	0	2025	HIDE

4.1 数据集与评估方案

我们使用峰值信噪比和结构相似性指数作为质量评估指标。为了报告各方法相对于表现最佳方法的误差降低幅度，我们将峰值信噪比转换为均方根误差（$RMSE \propto \sqrt{10^{-PSNR/10}}$），并将结构相似性指数转换为 DSSIM（DSSIM =(1 - SSIM)/2）。

图像去雨。 如表 1 所示，我们的去雨模型使用从多个数据集收集的 13,712 对干净图像与雨天图像进行训练 [51–54]。我们在多个测试集上评估模型性能，包括 Test 100 [53],Test1200 [54],Rain100H [52], 和 Rain100L [52]。

图像去模糊。如表 2 所示，为了进行图像去模糊，我们使用 GoPro [55] 数据集，其中包含 2,103 对用于训练的图像对和 1,111 对用于评估的图像对。此外，我们将使用 GoPro 训练的模型直接应用于 HIDE 数据集的测试图像，以评估我们方法的泛化能力。HIDE 数据集专为感知人体的运动去模糊而设计，其测试集包含 2,025 张图像。

4.2 实现细节

我们训练所提出的 M3SNet 时不进行任何预训练，并针对不同的图像复原任务分别训练模型。对于网络的每个层级，我们采用以下模块配置：编码器使用 $[1, 1, 1, 28]$ 个模块，解码器使用 $[1, 1, 1, 1]$ 个模块，FFM 使用 $[2, 2, 1, 0]$ 个模块。此外，在编码器和解码器之间的桥接部分使用一个 MHAMB。我们使用 Adam [58] 优化器($\beta_1 = 0.9$, $\beta_2 = 0.999$)和 PSNR 损失训练模型 5×10^5 次迭代，初始学习率为 1×10^{-3}，并通过余弦退火 [59]逐渐降低至 1×10^{-7}。我们从训练图像中提取大小为 256×256 的图像块，并将批量大小设为 32。我们采用 TLC [60]，以解决在图像块上训练、在完整图像上测试所导致的性能下降问题。数据增强方面，我们进行水平和垂直翻转。



图 6：**图像去雨**示例。M3SNet 的输出在两个图像样本中均未出现雨纹痕迹。M3SNet 还恢复出了细节最丰富的图像。

4.3 图像去雨结果

在我们的图像去雨任务中，我们使用 YCbCr颜色空间中的 Y 通道计算峰值信噪比/结构相似性指数分数，这与此前的工作一致，例如[8, 61, 62]。如表 3 所示，我们的方法已被证明能够显著且稳定地优于现有方法。具体而言，与最佳的基于卷积神经网络的方法 SPAIR [62] 相比，我们的方法在所有数据集上的平均性能提升了 0.93 dB，错误率降低了 10.2%。此外，在 Rain100L 等单个数据集上，我们的方法较 HINet [9] 最高可提升 2.76 dB。

除了定量评估之外，图6 还展示了定性结果，证明我们的 M3SNet 能够有效去除各种方向和程度的雨纹，同时保留图像的结构内容。

4.4 图像去模糊结果

GoPro [55] 和 HIDE [57] 数据集上的图像去模糊方法性能评估见表 4。我们的 M3SNet 优于其他方法，在所有数据集上的平均性能提升了 0.09 dB [55, 57]。与我们的基线网络 NAFNet 相比 [12], , 在宽度为 32 和 64 时，我们分别提升了 0.04 dB 和 0.05 dB。值得注意的是，尽管我们的网络仅在 GoPro 数据集上训练，但在 HIDE 数据集上仍取得了最先进的结果（峰值信噪比为 31.49 dB）。这展现了其出色的泛化能力。

图 7 展示了我们的方法生成的一些去模糊图像。与其他方法相比，我们的模型恢复出的图像更加清晰，也更接近真实值。

Table 3: Image deraining results. The best and second best scores are **highlighted** and underlined. Our M3SNet is better than the state-of-the-art by 0.93 dB.

Methods	Test100 [63]		Test1200 [61]		Rain100H [64]		Rain100L [64]		Average			
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑		SSIM ↑	
DerainNet [65]	22.77	0.810	23.38	0.835	14.92	0.592	27.03	0.884	22.48	(73.0%)	0.796	(63.7%)
SEMI [66]	22.35	0.788	26.05	0.822	16.56	0.486	25.03	0.842	22.88	(71.7%)	0.744	(71.1%)
DIDMDN [54]	22.56	0.818	29.65	0.901	17.35	0.524	25.23	0.741	24.58	(65.6%)	0.770	(67.8%)
UMRL [67]	24.41	0.829	30.55	0.910	26.01	0.832	29.18	0.923	28.02	(48.8%)	0.880	(38.3%)
RESCAN [11]	25.00	0.835	30.51	0.882	26.36	0.786	29.80	0.881	28.59	(45.4%)	0.857	(48.3%)
PreNet [10]	24.81	0.851	31.36	0.911	26.77	0.858	32.44	0.950	29.42	(39.9%)	0.897	(28.2%)
MSPFN [61]	27.50	0.876	32.39	0.916	28.66	0.860	32.40	0.933	30.75	(29.9%)	0.903	(23.7%)
MPRNet [8]	30.27	0.897	32.91	0.916	30.41	0.890	36.40	0.965	32.73	(12.0%)	0.921	(6.3%)
SPAIR [62]	30.35	0.909	33.04	0.922	30.95	0.892	36.93	0.969	32.91	(10.2%)	**0.926**	(0.0%)
HINet [9]	30.29	0.906	33.05	0.919	30.65	0.894	37.28	0.970	32.81	(11.2%)	0.922	(5.1%)
M3SNet-32	31.29	0.903	33.46	0.924	30.64	0.892	39.62	0.984	33.75	(1.0%)	**0.926**	(0.0%)
M3SNet-64	31.25	0.901	**33.52**	0.925	30.54	0.889	**40.04**	**0.985**	**33.84**	(0.0%)	0.925	(1.3%)



图 7： **图像去模糊**在 GoPro 数据集上的示例 [55]。与当前最先进的方法相比，我们的 M3SNet 能够恢复出更清晰、感知上更忠实的图像。

4.5 消融研究

我们在图像去模糊（GoPro [55]）任务上开展消融研究，以分析模型各组件的影响。值得注意的是，我们采用 NAFNet-32 [12] 作为基线网络，通过添加所提出的组件来展示其有效性。表 5 展示了 M3SNet 的有效性。接下来，我们将介绍各组件的影响。

FFM 我们通过将所提出的特征融合中间件添加到 NAFNet-32 [12]中，验证其有效性。如表 5 所示，在编码器-解码器的每个层级中添加 FFM 均能提升性能（32.90 dB），优于基线网络。这表明，我们提出的 FFM 有助于模型捕获更多多尺度信息，同时保留空间细节信息。

MHAMB如表 5 所示，在顶层添加多头注意力中间块（MHAMB）以捕获更多全局信息后，峰值信噪比仅从 30.87 dB 略微提升至 30.88 dB。

同时添加这两个组件后，性能显著提升，峰值信噪比从 32.87 dB 提高到 32.91 dB。

4.6 资源高效

为了实现更高的准确率，深度学习模型变得日益复杂。然而，更大的模型需要更多资源，在某些情况下可能并不实用。因此，我们需要设计能够实现高准确率的轻量级图像复原模型。在我们的工作中，我们设计了一种山形单阶段网络。该架构在空间细节

表4：图像去模糊结果。所提出的 M3SNet 仅在 GoPro 数据集上训练，但在两个数据集效果的平均值上，比最先进水平高出 0.09 dB。

方法	GoPro [55] 峰值信噪比 ↑	GoPro [55] 结构相似性指数 ↑	HIDE [57] 峰值信噪比 ↑	HIDE [57] 结构相似性指数 ↑	平均值 峰值信噪比 ↑		平均值 结构相似性指数 ↑	
DeblurGAN [23]	28.70	0.858	24.51	0.871	26.61	(49.9%)	0.865	(70.4%)
不 *etal.* [55]	29.08	0.914	25.73	0.874	27.41	(45.1%)	0.894	(62.3%)
DeblurGAN-v2 [25]	29.55	0.934	26.61	0.875	28.08	(40.7%)	0.905	(57.9%)
SRN [47]	30.26	0.934	28.36	0.915	29.31	(31.7%)	0.925	(46.7%)
Gao *etal.* [68]	30.90	0.935	29.11	0.913	30.01	(26.0%)	0.924	(47.4%)
DBGAN [24]	31.10	0.942	28.94	0.915	30.02	(25.9%)	0.929	(43.7%)
MT-RNN [69]	31.15	0.945	29.15	0.918	30.15	(24.8%)	0.932	(41.2%)
DMPHN [50]	31.20	0.940	29.09	0.924	30.15	(24.8%)	0.932	(41.2%)
苏因 *etal.* [70]	31.85	0.948	29.98	0.930	30.92	(17.8%)	0.939	(34.4%)
SPAIR [62]	32.06	0.953	30.29	0.931	31.18	(15.3%)	0.942	(21.0%)
MIMO-UNet++ [33]	32.45	0.957	29.99	0.930	31.22	(14.9%)	0.944	(28.6%)
MPRNet [8]	32.66	0.959	30.96	0.939	31.81	(8.9%)	0.949	(21.6%)
MPRNet-local [8]	33.31	0.964	31.19	0.945	32.25	(4.2%)	0.955	(11.1%)
Restormer [16]	32.92	0.961	31.22	0.942	32.07	(6.1%)	0.952	(16.7%)
Restormer-local [16]	33.57	0.966	**31.49**	0.945	32.53	(1.0%)	0.956	(9.1%)
Uformer [18]	32.97	**0.967**	30.83	**0.952**	31.90	(8.0%)	**0.960**	(0.0%)
HINet [9]	32.71	-	-	-	-	-		
HINet-local [9]	33.08	0.962	-	-	-	-		
NAFNet-32 [12]	32.87	0.961	-	-	-	-		
NAFNet-64 [12]	33.69	**0.967**	-	-	-	-		
M3SNet-32（我们的方法）	32.91	0.965	30.92	0.948	31.92	(7.8%)	0.957	(6.9%)
M3SNet-64（我们的方法）	**33.74**	**0.967**	**31.49**	0.951	**32.62**	(0.0%)	0.959	(2.4%)

表5：对所提出的 M3SNet 各个组件进行的消融研究。

	NAFNet-32	+FFM	+ MHAMB	+ FFM 与 MHAMB
PSNR	32.87	32.90	32.88	32.91
SSIM	0.9606	0.9632	0.9606	0.9644

与上下文信息之间实现了平衡，同时尽可能减少复原图像所需的计算资源。

如表 6 所示，我们的 M3SNet 性能优于其他模型。尽管参数量比 MIMO-unet ++ [33] 高出 0.6M，我们提出的 M3SNet-32 仍实现了更好的性能，同时显著减少了计算资源消耗，其乘加运算次数约为 MIMO-unet++ 的 1/40。虽然我们模型的乘加运算次数高于 Uformer [18] 和 NAFNet [12]，但模型参数量更少，性能仍然更好。综合考虑模型参数量、乘加运算次数和性能等因素，我们的模型是最佳选择。

表6：模型计算复杂度评估。评估在英伟达 1060 图形处理器上进行，输入尺寸为256 × 256。

方法	PSNR	参数(M)	乘加运算次数(G)
MIMO-UNet++ [33]	32.68	16.1	1235
MPRNet [8]	32.66	20.1	778.2
MPRNet-local [8]	33.31	20.1	778.2
HINet [9]	32.77	88.7	170.7
Restormer [16]	32.92	26.13	140
Restormer-local [16]	33.57	26.13	140
Uformer [18]	32.97	50.88	89.5
NAFNet-32 [12]	32.87	17.1	32
NAFNet-64 [12]	33.69	68	65
M3SNet-32（我们的方法）	32.91	16.7	37
M3SNet-64（我们的方法）	33.74	66.3	146

5 结论

本文提出了一种具有山形结构的单阶段网络，该网络能够有效捕获多尺度特征信息，同时最大限度地减少图像复原所需的计算资源。我们的设计遵循一项核心原则，即在图像复原过程中平衡上下文信息与空间细节这两个相互竞争的目标。为此，我们提出了一种特征融合中间件机制，使编码器-解码器架构不同层级之间能够实现无缝的信息交换。该方法将上层信息与相邻的下层信息平滑融合，并最终将所有信息整合到原始图像分辨率的处理层级。为了克服卷积神经网络感受野的局限性并捕获更多全局信息，我们采用多头注意力中间块作为编码器-解码器架构的桥梁。此外，为了保持计算效率并控制模型规模，我们替换或移除了非线性激活函数，转而采用乘法运算。我们在多个基准数据集上开展了大量实验，结果表明，M3SNet模型在仅使用较少计算资源的情况下，显著优于现有方法。

参考文献

[1] Rudin, L.I., Osher, S., Fatemi, E.: 基于非线性全变分的去噪算法. Physica D Nonlinear Phenomena (1992年)[2] Song, C.Z., Mumford, D.: 先验学习与吉布斯反应-扩散. IEEE模式分析与机器智能汇刊**19**(11), 1236–1250 (1997年)[3] Perona, P., Malik,J.: 使用各向异性扩散的尺度空间与边缘检测. IEEE模式分析与机器智能汇刊**12**(7), 629–639 (2002年)[4] Roth, S., Black, M.J.: 专家场：一种学习图像先验的框架. In: CVPR (2005年)

[5] Kim, K.I., Kwon, Y.：使用稀疏回归和自然图像先验进行单图像超分辨率。IEEE模式分析与机器智能汇刊 **32**(6), 1127 (2010)[6] Dong, W., Zhang, L., Shi, G., Wu, X.：通过自适应稀疏域选择和自适应正则化进行图像去模糊与超分辨率。IEEE图像处理汇刊 **20**(7), 1838– 1857 (2011)[7] He, K., Sun, J., Tang, X.：使用暗通道先验进行单幅图像去雾。IEEE模式分析与机器智能汇刊 (2011)[8] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.-H., Shao, L.：多阶段渐进式图像复原。载于：CVPR (2021)[9] Chen, L., Lu, X., Zhang, J., Chu, X., Chen, C.：Hinet：用于图像复原的半实例归一化网络。载于：CVPR (2021)[10] Ren, D., Zuo, W., Hu, Q., Zhu, P.F., Meng, D.：渐进式图像去雨网络：更优、更简单的基线。2019年 IEEE/CVF计算机视觉与模式识别会议 (CVPR), 3932–3941 (2019)[11] Li, X., Wu, J., Lin, Z., Liu, H., Zha, H.：用于单幅图像去雨的循环挤压-激励上下文聚合网络。载于：欧洲计算机视觉会议 (2018)[12] Chen, L., Chu, X., Zhang, X., Sun, J.：图像复原的简单基线。arXiv预印本 arXiv:2204.04676 (2022)[13] Chu, X., Chen, L., Yu, W.：Nafssr：使用NAFNet进行立体图像超分辨率。载于：CVPR (2022)[14] Pan, J., Sun, D., Zhang, J., Tang, J., Yang, J., Tai, Y.W., Yang, M.H.：用于低级视觉的双卷积神经网络。国际计算机视觉期刊 (2022)[15] Zhang, J., Zhang, Y., Gu, J., Zhang, Y., Kong, L., Yuan, X.：使用注意力可伸缩变换器实现精准图像复原。载于：国际学习表征会议 (2023)[16] Zamir, S.W.,Arora, A.,Khan, S., Hayat, M., Khan, F.S., Yang, M.-H.：Restormer：用于高分辨率图像复原的高效变换器。载于：CVPR (2022)[17] Tsai, F.-J., Peng, Y.-T., Lin, Y.-Y., Tsai, C.-C., Lin, C.-W.：Stripformer：用于快速图像去模糊的条带变换器。载于：ECCV (2022)[18] Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H.：Uformer：用于图像复原的通用U形变换器。载于：CVPR (2022)[19] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C.C., Qiao, Y., Tang, X.：Esrgan：增强型超分辨率生成对抗网络。载于：ECCV研讨会 (2018)

[20] Ronneberger, O.、P.Fischer、Brox, T.：U-Net：用于生物医学图像分割的卷积网络。载于：医学图像计算与计算机辅助干预（MICCAI）（2015年）[21] Kim, K.、Lee, S.、Cho, S.：Mssnet：用于单幅图像去模糊的多尺度阶段网络。arXiv **abs/2202.09652** （2022年）[22] Dong, H.、Pan, J.、Xiang, L.、Hu, Z.、Zhang, X.、Wang, F.、Yang, M.-H.：具有密集特征融合的多尺度增强去雾网络。载于：2020年IEEE/CVF计算机视觉与模式识别会议（CVPR），第2154–2164页（2020年）。https://doi.org/10.1109/CVPR42600.2020.00223[23] Kupyn, O.、Budzan, V.、Mykhailych, M.、Mishkin, D.、Matas, J.：DeblurGAN：使用条件对抗网络进行盲运动去模糊。2018年IEEE/CVF计算机视觉与模式识别会议，第8183–8192页（2017年）[24] Zhang, K.、Luo, W.、Zhong, Y.、Ma, L.、Stenger, B.、Liu, W.、Li, H.：通过真实模糊实现去模糊。2020年IEEE/CVF计算机视觉与模式识别会议（CVPR），第2734–2743页（2020年）[25] Kupyn, O.、Martyniuk, T.、Wu, J.、Wang, Z.：DeblurGAN-v2：更快、更优的去模糊（速度提升数个数量级）。2019年IEEE/CVF国际计算机视觉会议（ICCV），第8877–8886页（2019年）[26] Anwar, S.、Barnes, N.：密集残差拉普拉斯超分辨率。IEEE模式分析与机器智能汇刊（2020年）[27] Zhang, Y.、Li, K.、Li, K.、Wang, L.、Zhong, B.、Fu, Y.：使用超深残差通道注意力网络实现图像超分辨率。载于：ECCV（2018年）[28] Zhang, Y.、Tian, Y.、Kong, Y.、Zhong, B.、Fu, Y.：用于图像复原的残差密集网络。IEEE模式分析与机器智能汇刊（2020年）[29] Dudhane, A.、Zamir, S.W.、Khan, S.、Khan, F.S.、Yang, M.-H.：突发图像复原与增强。载于：CVPR（2022年）[30] Zamir, S.W.、Arora, A.、Khan, S.、Hayat, M.、Khan, F.S.、Yang, M.-H.、Shao, L.：学习增强特征以实现快速图像复原与增强。IEEE模式分析与机器智能汇刊（2022年）[31] Conde, M.V.、Choi, U.-J.、Burchi, M.、Timofte, R.：Swin2SR：用于压缩图像超分辨率与复原的Swinv2 Transformer。载于：欧洲计算机视觉会议（ECCV）研讨会论文集（2022年）[32] Liang, J.、Cao, J.、Sun, G.、Zhang, K.、Van Gool, L.、Timofte, R.：SwinIR：使用Swin Transformer进行图像复原。arXiv预印本 arXiv:2108.10257（2021年）[33] Cho, S.J.、Ji, S.W.、Hong, J.P.、Jung, S.W.、Ko, S.J.：重新思考单幅图像去模糊中的由粗到精方法。载于：ICCV（2021年）

[34] Yue, Z.、Zhao, Q.、Zhang, L.、Meng, D.：双重对抗网络：面向真实世界的噪声去除与噪声生成。载于：ECCV（2020年）[35] Zhang, K.、Li, Y.、Zuo, W.、Zhang, L.、Gool, L.V.、Timofte, R.：利用深度去噪器先验实现即插即用的图像复原。IEEE模式分析与机器智能汇刊 **44**，第6360–6376页（2020年）[36] Zhu, H.、Xi, P.、Chandrasekhar, V.、Li, L.、Lim, J.H.：Dehazegan：当图像去雾遇上微分编程。载于：第二十七届国际人工智能联合会议 IJCAI-18（2018年）[37] Guo, T.、Li, X.、Cherukuri, V.、Monga, V.：用于去雾的稠密场景信息估计网络。载于：2019年IEEE/CVF计算机视觉与模式识别会议研讨会（CVPRW）（2019年）[38] Yang, A.、Wang, H.、Ji, Z.、Pang, Y.、Shao, L.：双路径网络中的双路径：用于单幅图像去雾。载于：第二十八届国际人工智能联合会议 IJCAI-19（2019年）[39] Tian, C.、Xu, Y.、Zuo, W.、Du, B.、Lin, C.-W.、Zhang, D.：用于图像去噪的双卷积神经网络的设计与训练。《知识库系统》**226**，第106949页（2021年）[40] Singh, V.、Ramnath, K.、Mittal, A.：细化高频信息以实现更清晰的超分辨率与去模糊。《计算机视觉与图像理解》（2020年）[41] Pan, J.、Liu, S.、Sun, D.、Zhang, J.、Liu, Y.、Ren, J.、Li, Z.、Tang, J.、Lu, H.、Tai, Y.W.a.：学习用于低级视觉的双卷积神经网络。载于：CVPR（2018年）[42] Siyuan, L.I.、Ren, W.、Zhang, J.、Yu, J.、Guo, X.：通过深度分解-组合网络实现快速单幅图像去雨。《计算机视觉与图像理解》（2018年）[43] Li, W.、Wang, Z.、Yin, B.、Peng, Q.、Du, Y.、Xiao, T.、Yu, G.、Lu, H.、Wei, Y.、Sun, J.：重新思考用于人体姿态估计的多阶段网络。**arXivabs/1901.00148**（2019年）[44] Cheng, B.、Chen, L.-C.、Wei, Y.、Zhu, Y.、Huang, Z.、Xiong, J.、Huang, T.、Hwu, W.-m.W.、Shi, H.：SPGNet：用于场景解析的语义预测引导。2019年IEEE/CVF国际计算机视觉会议（ICCV），第5217–5227页（2019年）[45] Ghosh, P.、Yao, Y.、Davis, L.S.、Divakaran, A.：用于动作分割的堆叠时空图卷积网络。2020年IEEE冬季计算机视觉应用会议（WACV），第565–574页（2018年）[46] Li, S.-J.、AbuFarha, Y.、Liu, Y.、Cheng, M.-M.、Gall, J.：Ms-TCN++：用于动作分割的多阶段时间卷积网络。IEEE Transactions on

Pattern Analysis and Machine Intelli 度 1 1 (2020年)htt ps://doi.org/10.1109/IEEE模式分析与机器智能汇刊.2020.3021756[47] Tao, X., Gao, H., Wang, Y., Shen, X., Wang, J., Jia, J.：用于深度图像去模糊的尺度循环网络。CVPR（2018年）[48] Fu, X., Liang, B., Huang, Y., Ding, X., Paisley, J.：用于图像去雨的轻量级金字塔网络。IEEE神经网络与学习系统汇刊（2018年）[49] Zhang, H., Zhang, L., Dai, Y., Li, H., Koniusz, P.：用于非均匀运动去模糊的事件引导式自监督多图块网络。《国际计算机视觉杂志》，1–18（2022年）[50] Zhang, H., Dai, Y., Li, H., Koniusz, P.：用于图像去模糊的深度堆叠层次多图块网络。载于：IEEE计算机视觉与模式识别会议（CVPR）（2019年）[51] Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.：通过深度细节网络去除单幅图像中的雨。载于：CVPR（2017）[52] Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.：从单幅图像中深度联合检测并去除雨。CVPR（2017）[53] Zhang, H., Sindagi, V.A., Patel, V.M.：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汇刊 **30**，3943–3956（2017）[54] Zhang, H., Patel, V.M.：使用多流密集网络进行密度感知单幅图像去雨。2018年IEEE/CVF计算机视觉与模式识别会议，695–704（2018年）[55] Nah, S., Kim, T.H., Lee, K.M.：用于动态场景去模糊的深度多尺度卷积神经网络。2017年IEEE计算机视觉与模式识别会议（CVPR），257–265（2016）[56] Li, Y., Tan, R.T., Guo, X., Lu, J., Brown, M.S.：利用层先验去除雨纹。载于：CVPR（2016）[57] Shen, Z., Wang, W., Lu, X., Shen, J., Ling, H., Xu, T., Shao, L.：感知人体的运动去模糊。2019年IEEE/CVF国际计算机视觉会议（ICCV），5571–5580（2019年）[58] Kingma, D., Ba, J.：Adam：一种随机优化方法。《计算机科学》（2014）[59] Loshchilov, I., Hutter, F.：SGDR：带热重启的随机梯度下降（2016）

[60] Chu, X., Chen, L., Chen, C., Lu, X.：通过重新审视全局信息聚合来改进图像复原。载于：ECCV（2021）[61] Jiang, K., Wang, Z., Yi, P., Chen, C., Huang, B., Luo, Y., Ma, J., Jiang, J.：用于单幅图像去雨的多尺度渐进融合网络。CVPR（2020）[62] Purohit, K., Suin, M., Rajagopalan, A.N., Boddeti, V.N.：使用失真引导网络进行空间自适应图像复原。CoRR **abs/2108.08617**（2021）[63] Zhang, H.,Sindagi, V.A.,Patel, V.M.：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汇刊 **30**，3943–3956（2017）[64] Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.：从单幅图像中进行深度联合雨检测与去除。2017年IEEE计算机视觉与模式识别会议（CVPR），1685–1694（2016）[65] Fu, X., Huang, J., Ding, X., Liao, Y., Paisley, J.W.：拨云见日：用于单幅图像去雨的深度网络架构。IEEE图像处理汇刊 **26**，2944–2956（2016）[66] Wei, W., Meng, D., Zhao, Q., Xu, Z.：用于单幅图像去雨的半监督卷积神经网络。arXiv **abs/1807.11078**（2018）[67] Yasarla, R., Patel, V.M.：利用循环移位卷积神经网络进行不确定性引导的多尺度残差学习，以实现单幅图像去雨。2019年IEEE/CVF计算机视觉与模式识别会议（CVPR），8397–8406（2019）[68] Gao, H., Tao, X., Shen, X., Jia, J.：通过参数选择性共享和嵌套跳跃连接实现动态场景去模糊。2019年IEEE/CVF计算机视觉与模式识别会议（CVPR），3843–3851（2019）[69]Park, D., Kang, D.U., Kim, J., Chun, S.Y.：利用增量时序训练的多时序循环神经网络，实现渐进式非均匀单幅图像去模糊。载于：ECCV（2019）[70] Suin, M., Purohit, K., Rajagopalan, A.N.：用于自适应运动去模糊的空间注意补丁层次网络。CVPR，3603–3612（2020）