

高效多尺度卷积自注意力单幅图像除雨方法

李民瑶¹, 王鑫^{1,2,3}, 颜靖柯³, 覃琴³

(1. 桂林电子科技大学 计算机与信息安全学院, 广西 桂林 541004;

2. 电子科技大学 信息与软件工程学院, 成都 610000;

3. 桂林电子科技大学 海洋工程学院, 广西 北海 536000)

摘要: 雨天作为较常见的一种自然天气情况, 会极大地影响户外视觉系统所拍摄到的图像和视频数据的成像质量并制约后续高级计算机视觉任务的性能; 针对目前除雨算法存在伪影残留、细节丢失等问题, 为了充分提取图像特征, 有效去除雨条纹, 提高除雨效率, 提出一种新颖的单阶段深度学习除雨方法; 采用高效卷积和跨尺度自注意力相结合的方式, 弥补纯卷积网络无法满足的全局特征建模能力; 嵌入多尺度空间特征融合模块, 有效增加网络的感受野, 增强网络对不同分布的雨条纹特征的学习能力; 设计了一种混合损失函数, 利用各损失函数的优势来弥补单一损失函数表现出来的缺陷; 经过在不同类型数据集上的大量实验证明, 该算法不仅能够有效去除雨条纹, 充分保留背景细节, 而且处理速度也有显著的提升。

关键词: 图像除雨; 深度学习; 卷积神经网络; 自注意力机制; 多尺度网络

Efficient Multi-scale Convolution Self-attention Single Image Rain Removal Method

LI Minyao¹, WANG Xin^{1,2,3}, YAN Jingke³, QIN Qin³

(1. School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China;

2. School of Information and Software Engineering, University of Electronic Science and

Technology of China, Chengdu 610000, China;

3. School of Marine Engineering, Guilin University of Electronic Technology, Beihai 536000, China)

Abstract: As a relatively common natural weather condition, rainy days will significantly affect the imaging quality of images and video data captured by outdoor vision systems, and restrict the performance of subsequent advanced computer vision tasks. Aimed at the problems of artifact residues, details lost in the algorithm for removing rain, In order to fully extract image features, effectively remove rain streaks, and improve the efficiency of rain removal, a novel single-stage deep learning rain removal method is proposed to solve the problems of artifacts and loss of details in current rain removal algorithms. The combination of efficient convolution and cross-scale self-attention was used to make up for the global feature modeling capabilities that pure convolutional networks cannot meet. Embedding multi-scale spatial feature fusion modules to increase the receptive field of the network effectively and enhance the network's ability to learn rain streak features of different distributions. A hybrid loss function was designed to use the advantages of each loss function to make up for the shortcomings of a single loss function. A large number of experiments on different types of datasets have proved that the algorithm can effectively remove the rain streaks, fully retain the background details, and has a significant increase in processing speed.

Keywords: image rain removal; deep learning; convolutional neural network; self-attention mechanism; multi-scale network

0 引言

随着智能科学技术的不断创新以及软硬件技术的快速发展, 人们慢慢步入智慧信息化的新时代, 生活生产中的设备逐渐智能化, 计算机视觉应用日趋广泛, 例如车牌识

别自动抬杆系统、视频监控系统、智能自动驾驶系统, 然而, 其中大部分视觉系统是在室外环境下进行图像数据的采集, 因此将不可避免地受到诸如雨、雾、雪等恶劣天气的影响。雨天作为较常见的一种自然天气情况, 会极大地影响户外视觉系统的成像质量, 并制约后续高级计算机视

收稿日期: 2021-12-13; 修回日期: 2021-12-22。

基金项目: 广西自然科学基金面上项目(2019GXNSFAA245053); 广西科技重大专项(AA19254016); 广西硕士研究生创新项目(YCSW2021174); 海洋强国战略下广西海洋文化译介研究(2021KY0184); 北海市科技规划项目(202082033)。

作者简介: 李民瑶(1996-), 男, 江西赣州人, 硕士研究生, 主要从事图像处理、深度学习方向的研究。

王鑫(1976-), 男, 陕西西安人, 博士研究生, 教授, 硕士生导师, 主要从事图像处理、网络信息安全、物联网及数据挖掘方向的研究。

通讯作者: 覃琴(1985-), 女, 广西桂林人, 硕士研究生, 高级实验师, 主要从事图像处理、物联网、无线传感器网络方向的研究。

引用格式: 李民瑶, 王鑫, 颜靖柯, 等. 高效多尺度卷积自注意力单幅图像除雨方法[J]. 计算机测量与控制, 2022, 30(6): 190-196, 202.

觉任务的性能表现, 因此, 从有雨图像中去除雨噪声并恢复清晰的背景是一个重要的研究方向, 对图像除雨问题的研究在生活、生产中的应用具有重要价值。

在单幅图像除雨问题中, 通常把有雨图像看作是由雨条纹层和背景层叠加而成, 从有雨图像中分离雨条纹层和背景层是一个逆问题。传统方法主要分为基于先验的方法^[1-3]和基于滤波器^[4-7]的方法, 其中基于先验的方法利用先验信息约束逆问题的解空间, 通过设计优化算法对模型求解得到无雨图像。基于滤波器的方法将有雨图像分解为高频部分和低频部分, 利用雨条纹的物理特性设计滤波器对有雨图像进行滤波操作得到无雨图像。由于人工设计的局限性, 传统方法只能提取较为简单的特征, 而真实有雨图像中存在不同方向和密度的雨条纹, 其特征分布更为复杂。

最近, 基于深度学习的方法在单幅图像除雨领域中被大量应用, 证明了深度学习的方法在从有雨图像中恢复干净背景图像方面的优势。例如 Yang^[8] 等提出一个新的雨模型, 将全球大气光和大气透射率考虑其中, 同时根据输入的雨图生成二进制雨条纹特征图用于雨噪声的检测, 构建了一个循环雨水检测和清除的网络, 该方法也可以循环逐步地消除雨水条纹和雨条纹累积, 即使在大雨的情况下也表现优良。为了解决大雨累积产生的薄雾现象, 作者进一步提出了增强版本 JORDER_E, 其中包括额外的雨雾去除和细节保留的步骤^[9]。为了充分的学习随机分布的雨条纹特征, Li 等^[10] 提出了一种基于压缩和激励 (SE) 模块的上下文信息融合网络 (RESCAN), 通过 SE 模块为每个通道分配不同大小的权重 α 值来学习不同分布的雨水特征。由于现有的基于深度学习的除雨方法依赖大量有雨/无雨图像对数据集来学习雨水特征, 然而现有的数据集大部分的合成的数据集, 这些数据集不能充分的表征真实世界的降雨分布特征。为了解决这个问题, Wang 等^[11] 半自动地构建一个涵盖大范围自然降雨场景的大范围有雨/无雨图像对数据集, 并提出了一种空间注意力网络 (SPANet) 去除雨条纹。由于基于深度学习的除雨网络越来越复杂, 缺乏足够的可解释性, 并且没有与雨条纹内在的物理结构相结合。针对这个问题, Wang 等^[12] 基于卷积字典学习, 利用近端梯度下降技术设计了一种雨卷积字典网络 (RCD-Net), 该方法虽然达到了很好的除雨效果, 但经过实验证明, 该方法的处理速度较慢。现有的除雨网络通常基于单一的网络结构来训练而没有考虑跨尺度信息的关联, 这可能会导致信息的丢失, 因此, Wang 等^[13] 提出了一种跨尺度的多子网络结构 (DCSFN) 来解决这个问题, 子网络通过门控循环单元和内部尺度连接块来融合多尺度特征, 从而提高对雨条纹的表示能力。最近, Zamir 等^[14] 提出了一个可适用于图像除雨、去模糊和去噪的综合性多阶段图像复原网络 (MPRNet), 该方法首先使用编码器-解码器架构学习上下文特征, 然后将它们与保留了局部信息的高分辨率分支相结合。在每个阶段, 作者引入了一种新颖的逐

像素自适应设计, 利用原位监督注意力来重新加权局部特征, 实现了不同阶段之间的信息交换。上述基于深度学习的除雨方法^[8-9] 大多将单幅图像除雨问题看成是一个多阶段的任务, 并使用递归结构进行雨条纹特征提取和清除, 这导致网络层次的加深和参数数量的剧增, 带来巨大的计算开销。另外, 现有基于卷积神经网络的除雨算法以有限的感受野对有雨图像以逐块的方式进行操作, 因此, 在较大区域中缺少空间上下文信息的依赖, 事实上, 这些上下文信息对于图像除雨问题有积极的作用^[15]。

为了解决上述局限性, 本文设计了一种新颖的单阶段高效除雨网络, 网络整体为编码-解码结构, 其中, 在编码过程中引入卷积自注意力机制, 提供跨尺度信息交互, 丰富网络的全局特征建模能力, 然后将高效的卷积操作和自注意力机制进行合理地组合, 使除雨模型兼具局部和全局的特征提取能力和推理速度, 以期能够快速应对不同场景下的降雨情况; 另外, 通过嵌入多尺度空间特征融合模块, 以不同的感受野提取图像的特征信息, 增强了网络模型对不同尺度特征的感知能力; 最后, 设计了一种全面的混合损失函数作为目标优化函数, 该损失函数不仅考虑了图片的亮度、颜色、结构等基本物理特性, 同时还考虑了不同分辨率下图像之间的差异, 使得模型能够取得良好除雨效果的同时, 保留更多的背景细节, 达到更好的视觉效果。

1 方法

当前基于深度学习的除雨方法大多将除雨问题分为两个阶段: 雨条纹检测阶段和去除阶段, 而这样会直接导致两个局限, 其一是由于多阶段而带来的网络深度加深, 导致参数量增多、网络不易收敛等问题; 其二是此类网络通常包含递归结构, 导致较高的计算开销, 训练速度慢等问题。

基于上述局限, 本文提出一种高效的多尺度自注意力单阶段除雨网络, 网络模型的整体框架如图 1 所示。整体网络是一个编码器-解码器结构, 由多个高效卷积块及其变体 (融合自注意力) 和多尺度特征融合模块组成的串行网络。

在下文中, 将对网络模型的各个模块进行详细的描述。

1.1 卷积自注意力机制

随着 Transformer 这一类自注意力模型在自然语言处理^[16-17] 领域中的突出表现, 许多国内外研究者也试图将自注意力机制引入计算机视觉^[18] 领域之中。在图像除雨任务中, 许多研究者致力于通过设计一些精细的分支模块如 SE^[19] 模块和 SAM^[20] 模块等来提高卷积神经网络对图像特征的感知能力, 然而, 卷积操作中固有的对局部信息处理的归纳偏置, 制约了卷积神经网络对全局特征的建模能力, 同时也制约了基于卷积神经网络的除雨方法的性能提升。受 Transformer 的启发, 本文引入一种具有跨尺度信息交互能力的自注意力机制 (CoaT, Conv-attentional transformer), 这是在原始 Transformer 模型基础之上的一种改进。

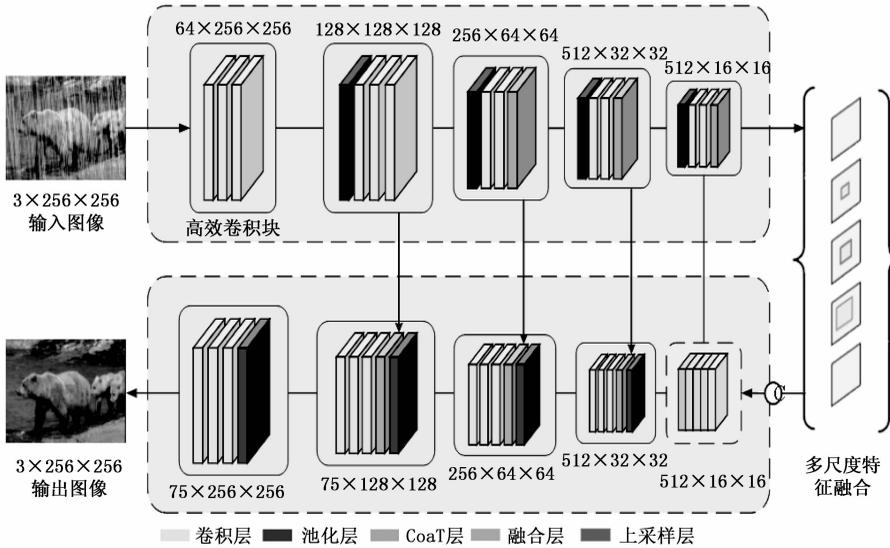


图 1 网络结构

简单来说,原始的 Transformer 将一个输入向量 $\mathbf{X} \in R^{N \times C}$ 转化为一系列的向量表示序列 $[x_1, \dots, x_N]$, 自注意力机制首先将这些向量表示序列经过线性变换 W^Q 、 W^K 和 $W^V \in R^C \times C$ 映射成相应的查询向量 Q , 键向量 K 和值向量 V , 对 Q , K 做归一化操作, 即除以 $\sqrt{d_k}$, 然后使用 $softmax$ 函数激活, 最后与值向量 V 点乘得到注意力 $Attention(X)$, 其公式如下:

$$Attention(\mathbf{X}) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

从公式 (1) 中可看出, 原始自注意力机制的空间复杂度为 $O(N^2)$, 时间复杂度为 $O(N^2C)$, 在处理高分辨率图像使计算量太大。因此, 受自注意力线性化的启发, Xu 等^[21]对此进行了改进, 使用两个函数 $\varphi(\cdot)$, $\psi(\cdot)$: $R^{N \times C} \rightarrow R^{N \times C}$ 将其分解得到:

$$FactorAtt(x) = \varphi(Q)(\psi(K)^T V) \quad (2)$$

根据 LambdaNets^[22], 以 φ 为恒等函数, ψ 为 $softmax$, 同时重新加入缩放因子 $1/\sqrt{d_k}$ 得到分解注意力机制:

$$FactorAtt(x) = \frac{Q}{\sqrt{d_k}}(softmax(K)^T V) \quad (3)$$

虽然该分解注意力机制减少了原始自注意力的所带来的计算负担, 但是原始 Transformer 仅由线性层和自注意力模块组成, 缺乏捕捉局部特征的能力, 当特征图中存在相似的局部特征时, 无法计算它们之间的差异。例如, 在图像除雨问题中, 在有雨图像的背景层中可能存在与雨条纹结构相似的特征, 当使用原始 Transformer 进行特征提取时, 自注意力机制会把雨条纹和类似雨条纹的背景结构都视为雨条纹, 这样在除雨的过程中就会导致在把雨条纹去除的同时, 背景图案也遭到了破坏。因此, 位置编码在 Transformer 中显得极为重要。

为了解决上述问题, 通过一个窗口大小为 M 的编码窗口, 整合一个相对位置编码为 $P = \{p_i, i = (M-1)/2,$

$\dots, (M-1)/2\}$, 得到相对注意力图 $EV \in R^{C \times C}$, 此时视输入向量序列为一维的, 有:

$$FactorAtt(x) = \frac{Q}{\sqrt{d_k}}(softmax(K)^T V) + EV \quad (4)$$

在图像任务中, 编码矩阵 $E \in R^{N \times N}$, 设通道数为 c , 对于每一个通道, 有:

$$E_{ij}^c = (i, j)q_i^c p_{j-i}^c, EV = \sum_j E_{ij}^c v_j^c \quad (5)$$

由公式 (5) 可知, EV 的计算仍旧需要 $O(N^2)$ 空间复杂度和 $O(N^2C)$ 时间复杂度, 因此, 将 EV 项简化为 \widehat{EV} , 在视觉任务中使用二维深度卷积来计算 \widehat{EV} :

$$\widehat{EV} = Q^{\circ} DepthwiseCon2D(P, V) \quad (6)$$

其中: \circ 表示对应矩阵位置的元素直接相乘。

因此, 加入相对位置编码的分解注意力机制计算公式为:

$$ConvAtt(x) = \frac{Q}{\sqrt{d_k}}(softmax(K)^T V) + \widehat{EV} \quad (7)$$

加入位置编码后的分解注意力机制将不仅拥有对全局特征的建模能力, 而且还对局部相似特征敏感, 有利于雨噪声的清除与背景细节纹理的保留。

1.2 卷积与 CoaT 有效结合

目前, 在基于深度学习的除雨方法中, 大多数方法都采用纯卷积神经网络的形式, 这些方法的局限性在于, 虽然卷积神经网络在局部特征建模上有良好的性能, 但在图像全局区域上, 无法对相似特征进行关联。虽然有些研究人员通过空洞卷积的方式来扩大卷积神经网络的感受野, 以获得更完整的特征表示, 但这也仅仅限于有限的区域。感受野是卷积和自注意力的最关键的区别之一, 更大的感受野提供了很多的上下文信息, 全局感受野是自注意力引入计算机视觉领域的关键因素, 但原始自注意力的空间复杂度和时间复杂度大, 尤其是对于网络的输入为高分辨图像时, 计算成本是无法接受的, 而本文引入的 CoaT 自注意力机制经过优化后, 其空间复杂度和时间复杂度低于原始 Transformer 自注意力机制。因此, 本文接下来通过探索高效卷积和 CoaT 自注意力机制合理的结合方式, 以得到一个相对良好的图像除雨方法, 在速度和精度上达到平衡。

对于如何将卷积与自注意力有效结合这一问题上, Dai 等^[23]也对此进行了研究, 他们的研究表明: 随着数据集增大, 模型的卷积阶段 (C) 越多, 模型之间的泛化能力的差距越小, 但是, 也并非简单地使用更多的 CoaT (T), 就能拥有更好的视觉处理能力。

受上述研究启发, 本文设计了组合 A: Conv-Conv-

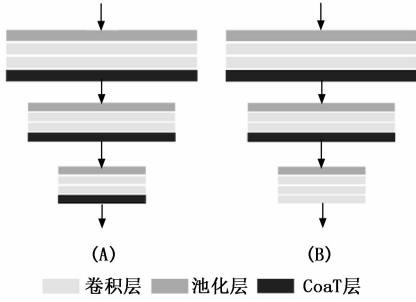


图 2 组合 A 和组合 B

Transformer-Transformer-Transformer (C-C-T-T-T) 和组合 B: Conv-Conv-Transformer-Transformer-Conv (C-C-T-T-C) 两种组合方案 (如图 2 所示), 来探究跨尺度卷积自注意力模块与普通卷积相结合的数量对所提网络模型性能的影响, 以及该模块在不同分辨率图像中对特征提取的有效性, 然后在数据集 Rain100H 上训练了这两种不同的组合方案, 并比较了这两种方案在 Rain100H 数据集上的评估指标 SSIM、PSNR、参数量和计算量。如表 1 所示, 使用组合 A 与组合 B 这两种方案在 Rain100H 数据集上训练后得到的评估指标数值相当, 组合 A 稍微领先, 即除雨效果差距不大, 但是组合 A 中包含更少的卷积块, 这意味着更少的参数量和计算消耗。因此, 本文采用组合 A 的方案将卷积块与 CoaT 结合。

表 1 不同算法在合成数据集上的评估指标对比

数据集	Rain100L		Rain100H	
	SSIM	PSNR	SSIM	PSNR
RESCAN(2018)	0.981 0	37.45	0.845 9	26.85
SPANet(2019)	0.963 2	34.56	0.821 3	24.63
RCDNet(2020)	0.982 6	40.00	0.889 3	31.16
DCSFN(2020)	0.982 1	37.60	0.887 0	27.76
JORDER-E(2020)	0.983 4	38.59	0.893 5	30.62
MPRNet(2021)	0.965 3	36.10	0.903 2	30.46
本文算法	0.979 5	39.32	0.918 5	31.41

1.3 多尺度空间特征融合

利用膨胀卷积来实现多尺度特征提取的方法在图像处理^[24]领域中十分常见。在图像除雨算法中, 通常会对图像进行若干次下采样操作, 以丢弃图像中的冗余信息, 但同时也会导致部分有效信息丢失, 使得无法准确定位雨条纹的位置, 会出现雨条纹清除不彻底、图像背景结构不完整等问题。为了缓解编码阶段中下采样过程会出现图像部分特征丢失的问题, 本文设计了一个多尺度空间特征融合块嵌入在编码阶段的最后一个阶段, 从多个尺度聚合上下文信息, 以充分学习不同尺寸的雨条纹特征, 使得模型能够应对真实环境下各种复杂的降雨情形。其具体结构如图 3 所示。

在该模块中采用 5 个并行的卷积操作对输入特征进行处理。首先使用一个 1×1 卷积对输入的特征图进行降维,

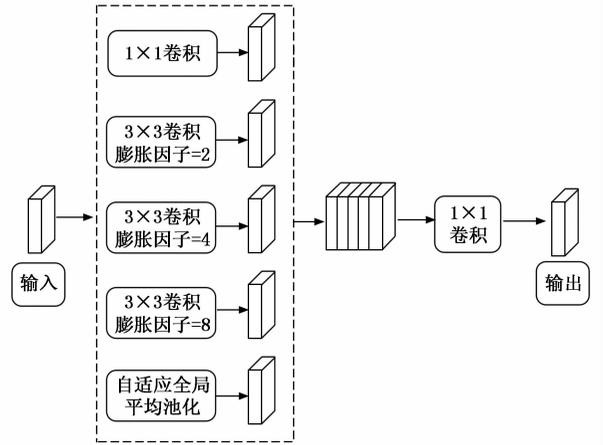


图 3 多尺度特征融合模块

然后对 3 个 3×3 卷积分别设置不同的膨胀因子 2、4、8, 以 3 种不同的感受野对图像进行特征提取, 提高模型对不同尺寸雨条纹的感知能力, 其次使用一个自适应平均池化操作降低信息冗余, 最后, 使用一个 1×1 卷积降低通道数, 将 5 个不同尺度的特征图融合在一起, 图像中的有效信息将得到充分的学习。另外, 使用跳跃连接 (如图 1 中的长箭头所示) 将编码阶段相同尺度的特征图进行融合来指导上采样过程, 以捕获多尺度下的细粒度语义和粗粒度语义, 这样有利于保留更多的细节纹理。

1.4 损失函数

在神经网络中, 损失函数的作用是在神经网络的前向传播过程中计算每次迭代的预测值与真实值之间的差距, 通过反向传播指导网络更新各层参数, 从而使得网络朝着预期方向进行训练。由此可见, 一个合适的损失函数在神经网络中有着至关重要的作用。目前, 大多数基于深度学习的单幅图像除雨方法均使用单一的函数作为损失函数, 比如均方误差 (MSE, mean squared error)、平均绝对误差 (MAE, mean absolute error)、多尺度结构相似 (MS-SSIM, mutil-scale structural similarity) 和总变差 (TV, total variation) 损失函数等, 表达式分别如下:

$$L_{MSE}(P) = \frac{1}{N} \sum_{p \in P} (\hat{y}(P) - y(P))^2 \quad (8)$$

$$L_{MAE}(P) = \frac{1}{N} \sum_{p \in P} |\hat{y}(P) - y(P)| \quad (9)$$

$$L_{MS-SSIM}(P) = 1 - (MS - SSIM(\hat{p})) \quad (10)$$

$$L_{TV}(P) = \sum_{i,j} ((p_{i,j+1} - p_{i,j})^2 + (p_{i+1,j} - p_{i,j})^2)^{\beta/2} \quad (11)$$

其中: P 表示一块像素区域, \hat{p} 表示像素区域 P 中的一个像素点, p 表示像素区域 P 中的一个像素点, N 表示像素区域 P 中的像素点数量, \hat{y} 表示预测图像, y 表示真实图像, i 和 j 表示像素点 p 的坐标。

然而, MSE 和 MAE 这两个损失函数都是基于逐像素进行差异比较, 并没有考虑人类视觉系统 (HVS) 对视觉场景的感知特点, 因此基于逐像素的差异比较损失函数并

不能准确的衡量某些失真类型区域；MS-SSIM 损失函数对一致性偏差不是特别敏感，容易导致图像亮度的变化和颜色的偏差；而 TV 损失函数通过计算相邻像素之间的差异来约束图像的平滑度，可用于解决除雨后图像中雨条纹残留带来的伪影问题，但是不适合单独使用。因此，本文设计了一个如公式 (13) 所示的混合型损失函数，利用各损失函数的优势来弥补单一损失函数表现出来的缺陷，增强了网络对图像细节恢复的能力。

首先，考虑到 MS-SSIM 损失函数能够根据 HVS 的敏感性在不同的尺度上计算 SSIM 的权重，即从多个尺度来感知图像质量的差异，这对于除雨任务是有利的，但是对于图像的亮度和颜色处理欠佳，而 MAE 损失函数不会过度惩罚较大的误差，即能够保留颜色和亮度，这两个损失函数的优缺点正好互补。因此，将 MAE 损失函数与 MS-SSIM 损失函数按照一定的权重混合，公式 (12) 如下：

$$L_{MS-SSIM-MAE} = \alpha \cdot L_{MS-SSIM} + (1 - \alpha) \cdot G_{\alpha_c}^N \cdot L_{MAE} \quad (12)$$

其中：根据 zhao 等^[25]经验，将 α 设置为 0.84。

在有雨图像中，大多数雨条纹和背景细节纹理存在高频区域，而 MAE 损失函数对于图像的高频部分会给予较大的权重，这将导致在保留细节的同时也会使得雨条纹部分残留，因此将 MSE 损失函数和 TV 损失函数按一定权重混合来去除雨条纹伪影。最终的混合损失函数如下：

$$L_{Mix} = L_{MS-SSIM-MAE} + \mu \cdot L_{MSE} + \lambda \cdot L_{TV} \quad (13)$$

其中： μ 和 λ 为惩罚因子，根据实验逐步调整，得到的值分别为 0.3 和 2×10^{-8} 。

2 实验与结果分析

实验环境为 Intel Core i7-9700K 处理器，Ubuntu 18.04 操作系统，使用英伟达 RTX3070、显存为 8 G 的 GPU 并行加速网络训练，所使用的编程语言为 Python，并采用 Facebook 人工智能研究院推出的 PyTorch 深度学习框架来搭建网络。所有实验均在此环境上进行。

2.1 参数设置

本文网络中所有融合模块的卷积核大小均为 1×1 ，其余卷积核大小为 3×3 ，采用多尺度特征融合模块，其中包含 3 个膨胀卷积块，膨胀因子分别为 2、4、8，采用 ReLU 函数作为非线性激活单元，使用 Adam 优化器对网络进行优化，学习率初始化为 0.000 2，每 50 个 epochs 更新一次学习率，学习率衰减策略为衰减率与迭代次数呈反比的固定步长衰减策略。

2.2 合成数据集实验对比

为了验证本文的方法在除雨效果方面的提升，将其在两个比较流行的公开合成数据集上进行训练和测试，数据集分别为 Yang 等^[8]提供的 Rain100L 和 Rain100H 数据集。其中，Rain100L 数据集包含 200 张有雨图像，其中包含的有雨图像的类型单一，因此是相对容易的数据集；Rain100H 是一个包含 1 800 张有雨图像的综合数据集，其中包含多种方向和密集程度不一的雨条纹，是一个具有挑战性的数据集，保证了实验的说服力。同时为了全面的验

证本文所提方法的性能，将其与目前比较先进的几种基于深度学习的除雨方法进行了比较，具体为：MPRNet (2021)、RCDNet (2020)、JORDER-E (2020)、DCSFN (2020)、SPANet (2019)、RESCAN (2018)。为了体现对比的公平性，所参与对比的算法均使用对应论文中提供的源代码和最优参数，并在本文使用的数据集上重新训练。

图 4 展示了不同算法在合成数据集 Rain100H 上的主观实验结果，从图中可以看出，本文算法可以有效地去除不同方向和密集程度不同的雨条纹，并且生成了接近真实的无雨图像，同时保留了大部分的细节。相比之下，其他方法生成的图像较平滑，甚至会破坏背景内容，图 4 (a) 女孩头发纹理几乎消失，图 4 (b) 十字架建筑模糊不清，在所对比的 3 种方法中，只有 MPRNet 算法保留了部分细节。DCSFN 算法使用 SSIM 损失作为损失函数，使得物体结构的边界明显，导致雨条纹的残留，如图 4 (b) 中 DCSFN 输出图像的天空可以明显看到雨条纹的伪影；RCDNet 算法使用 MSE 损失函数，导致算法对图像中的平滑区域惩罚力度较大，使图像变得模糊。本文算法输出的图像不仅将雨条纹去除，而且女性脸部自然、女孩头发纹理清晰、十字架完整，充分保留了背景细节，这表明将卷积的局部特征建模能力和自注意力的全局特征建模能力进行合理地结合对除雨任务有益，同时也验证了所提混合损失函数的有效性。

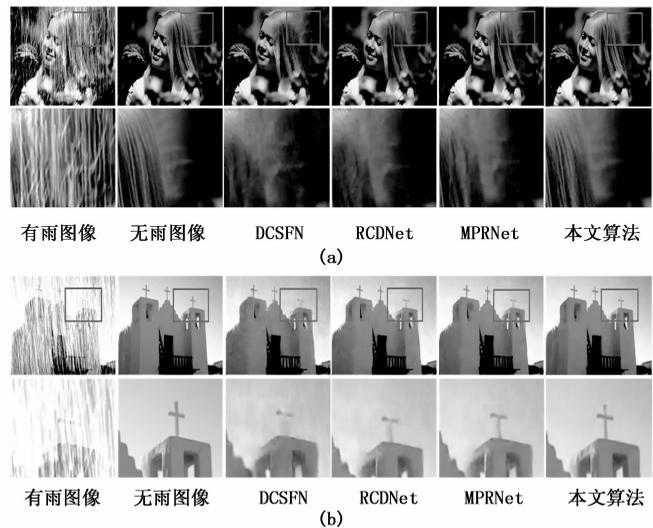


图 4 不同算法在合成数据集 Rain100H 上的主观实验结果

除了对比各算法的主观效果外，为了在数据上体现本文所提算法带来的性能提升，本文采用结构相似性 SSIM 和峰值信噪比 PSNR 两种图像质量评估指标来对各算法进行客观评价。其中 SSIM 值越接近 1，表明两幅图像的相似度越高；PSNR 值越大，表明图像的失真越少。表 2 给出了各算法在不同数据集上的 SSIM 值和 PSNR 值。由表 2 可知，本文算法与目前一些先进的算法对比具有优势。单从图像质量评估指标来看，虽然在 Rain100L 数据集上，本文方法在 PSNR 指标上略逊色于 RCDNet，SSIM 指标略

低于 JORDER-E, 但在 Rain100H 数据集上, 我们的方法占据了领先地位, 与最新的 MPRNet 算法对比, SSIM 指标提升了 0.015 3, PSNR 指标提高了 0.95 dB, 可能的原因是 Transformer 中较弱的归纳偏置依赖更大的训练集, 因此本文方法在较大的 Rain100H 数据集上具有更好的表现。

表 2 组合 A 和组合 B 在 Rain100H 数据集上的评估指标对比

数据集	Rain100H			
	评估指标	SSIM	PSNR/dB	参数量/M
组合 A	0.918 5	31.41	33.64	52.17
组合 B	0.916 0	31.35	34.94	52.51

另外, 为了体现本文所提算法的除雨效率, 还对比了在数据集 Rain100H 上不同算法平均处理一张有雨图像的运行时间及评估指标 PSNR 和 SSIM。对比结果如图 5 所示, 由图可知, 本文算法在 PSNR 指标上与 RCDNet 接近, 但是处理速度快 50 倍左右; 与最新算法 MPRNet 相比, 不仅取得了更好的除雨效果, 而且除雨速度显著提升。

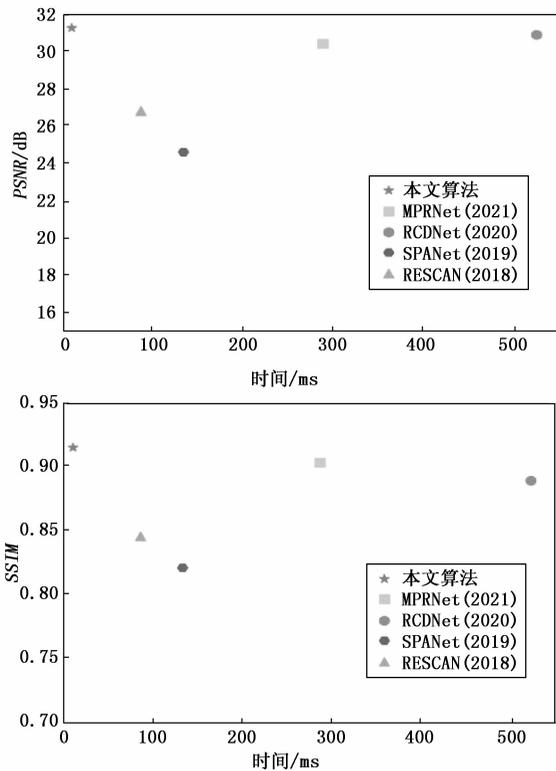


图 5 不同算法 Rain100H 上的平均运行时间及评估

2.3 仿真数据集实验对比

为了验证本文所提算法的泛化能力, 将其与最新算法 MPRNet (2021) 在接近真实的有雨数据集 SPA 上进行了比较, 其对比结果如图 6 所示, 两种算法对接近真实的有雨图像的除雨效果相当, 但是本文方法保留了更多的细节, 如图 6 (a) 中, MPRNet 算法把原图背景中的白色长条物去除了一个, 造成图像信息的丢失, 而在经过本文算法得到的除雨图像中, 其得到完整的保留。实验证明: 本文算

法的泛化能力较强, 不仅能够有效去除合成有雨图像上不同分布的雨条纹, 并且对于接近真实的降雨分布图像也有很好的清除效果。

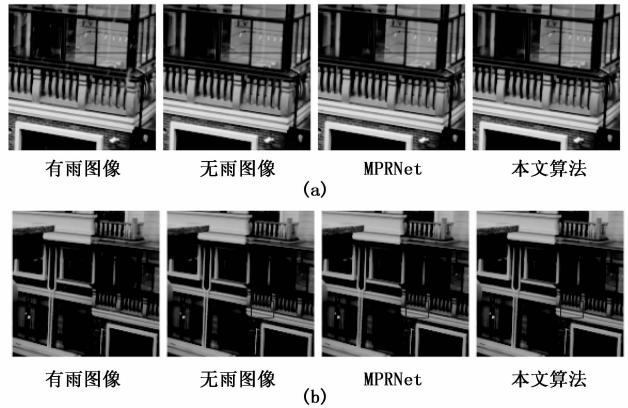


图 6 不同算法在仿真数据集 SPA 上的主观实验结果

2.4 真实数据集实验对比

为了更加全面地评估所提算法的性能, 我们收集了一些真实世界中由手机拍摄的有雨图像数据集, 包括大雨和小雨。然后, 使用在合成数据集上训练的模型用于测试, 并与最新算法 MPRNet (2021) 对比。如图 7 所示, 本文算法去除了大部分的雨痕, 并且除雨效果优于 MPRNet。具体地, 在图 7 (a), 人物和雨伞上的雨水被完全去除, 而 MPRNet 处理的图像中还残留很多雨滴。在图 7 (b) 中, 本文算法去除了图中明显的雨条纹, 而 MPRNet 仅去除了部分轻微的雨痕, 而粗雨痕没有被去除。在真实世界数据集上的实验证明, 本文算法的泛化能力较强并且可以很好地应用于自然环境中的有雨图像。

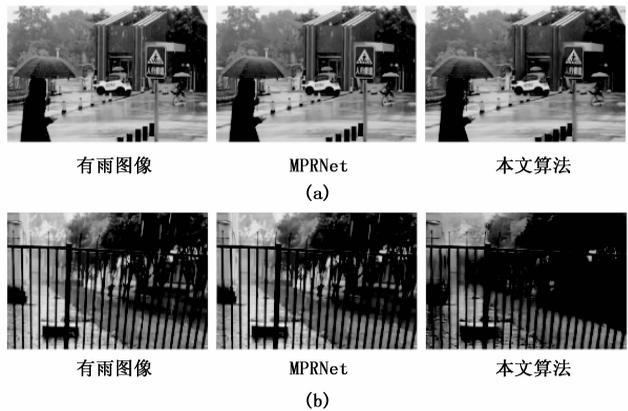


图 7 不同算法在真实数据集上的主观实验结果

2.5 消融实验

为了验证 CoaT 自注意力模块与普通卷积的结合对除雨任务的有效性, 以及结合方式对所提网络模型性能的影响, 设计了两种结合方案如图 2 所示, 其中组合 A: C-C-T-T-T 表示在编码阶段的第 3、4、5 个卷积块中嵌入 CoaT, 组合 B: C-C-T-T-C 表示在编码阶段的第 3、4 个卷积块中嵌入 CoaT, 然后在数据集 Rain100H 上进行消融实验, 结果如

表 2 所示, 组合 A 在 SSIM 和 PSNR 指标上略微领先组合 B, 说明嵌入更多的 CoaT 自注意力机制, 增强模型的全局特征建模能力, 能够带来更好的除雨效果; 其次, 组合 A 的参数量和计算量比组合 B 更少, 证明使用 CoaT 代替标准卷积可以减少模型的参数量, 降低计算消耗, 使模型的除雨效率更高。

3 结束语

针对单幅图像除雨问题, 本文设计了一种单阶段的高效多尺度特征融合除雨方法。首先, 该方法通过将高效卷积和跨尺度自注意力进行合理的结合, 提高了网络的全局特征建模能力, 从而获取更加有效的特征表示; 其次, 通过构造多尺度空间特征融合模块, 有效增加网络的感受野, 增强网络对不同分布的雨条纹特征的感知能力; 然后, 为了保留更多的背景细节, 使得去除雨水后的图像更加接近真实无雨图像, 设计了一种混合损失函数, 利用各损失函数的优势来弥补单一损失函数表现出来的缺陷, 解决了输出的无雨图像过于平滑、模糊、伪影残留和背景内容被破坏等问题。实验结果证明, 所提算法输出的无雨图像不仅在主观上达到了良好的视觉效果, 而且在客观评估指标上对比最新除雨算法也有所提升, 另外, 在除雨速度上也有显著的优势, 具有一定的实际应用价值。尽管所提方法能够在去除图像中的雨条纹的同时充分保留背景细节, 但是由于引入了自注意力机制, 需依赖更大的训练集, 这增加了网络的训练时间。在未来, 将进一步研究模型在较小数据集上的拟合问题, 并针对特定任务设计一种轻量化网络, 使得算法模型投入实际应用成为可能。

参考文献:

[1] MU P, CHEN J, LIU R, et al. Learning Bilevel layer priors for single image rain streaks removal [J]. *IEEE Signal Processing Letters*, 2018; 1-1.

[2] KANG L W, LIN C W, FU Y H. Automatic single-image-based rain streaks removal via image decomposition [J]. *IEEE Trans. Image Process*, 2012, 21 (4): 1742-1755.

[3] LEI Z, FU C W, LISCHINSKI D, et al. Joint bi-layer optimization for single-image rain streak removal [C] // *IEEE International Conference on Computer Vision*, IEEE, 2017; 2545-2553.

[4] JING X, WEI Z, PENG L, et al. Removing rain and snow in a single image using guided filter [C] // *2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE)*, 2012; 304-307.

[5] ZHENG X, LIAO Y, WEI G, et al. Single-image-based rain and snow removal using multi-guided filter [C] // *International Conference on Neural Information Processing*. Springer Berlin Heidelberg, 2013; 266-273.

[6] XU J, ZHAO W, LIU P, et al. An improved guidance image based method to remove rain and snow in a single image [J]. *Computer and Information Science*, 2012, 5 (3): 49.

[7] KIM J H, LEE C, SIM J Y, et al. Single-image deraining using an adaptive nonlocal means filter [C] // *IEEE International Conference on Image Processing*, IEEE, 2014; 914-917.

[8] YANG W, TAN R T, FENG J, et al. Deep joint rain detection and removal from a single image [C] // *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2017; 1357-1366.

[9] WENHAN Y, TAN R T, JIASHI F, et al. Joint rain detection and removal from a single image with contextualized deep networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42 (6): 1377-1393.

[10] XIA L, WU J, LIN Z, et al. Recurrent squeeze-and-excitation context aggregation net for single image deraining [M]. Springer, Cham, 2018; 254-269.

[11] WANG T, X YANG, XU K, et al. Spatial attentive single-image deraining with a high quality real rain dataset [C] // *IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, IEEE, 2019; 12270-12279.

[12] WANG H, XIE Q, ZHAO Q, et al. A model-driven deep neural network for single image rain removal [C] // *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2020; 3103-3112.

[13] WANG C, XING X, WU Y, et al. DCSFN: deep cross-scale fusion network for single image rain removal [C] // *Proceedings of the 28th ACM international Conference on Multimedia*, 2020; 1643-1651.

[14] ZAMIR S W, ARORA A, KHAN S, et al. Multi-stage progressive image restoration [J]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021; 14821-14831.

[15] HUANG D A, KANG L W, YANG M C, et al. Context-aware single image rain removal [C] // *2012 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, 2012; 164-169.

[16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] // *Advances in Neural Information Processing Systems*, 2017; 5998-6008.

[17] ZHANG Y, XU B, ZHAO T. Convolutional multi-head self-attention on memory for aspect sentiment classification [J]. *IEEE/CAA Journal of Automatica Sinica*, 2020, 7 (4): 1038-1044.

[18] QHA B, CHA B, XW A, et al. Facial expression recognition with grid-wise attention and visual transformer [J]. *Information Sciences*, 2021, 580; 35-54.

[19] Jie H, Li S, Gang S, et al. Squeeze-and-Excitation Networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017; 2011-2023.

[20] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C] // *European Conference on Computer Vision*. Springer, Cham, 2018; 3-19.

(下转第 202 页)