结合多尺度特征融合和多输入多输出编-解码器的去模糊算法

赵 倩，周冬明*，杨 浩，王长城，李 森

(云南大学 信息学院，云南 昆明 650504)

摘 要：针对相机抖动、拍摄物体快速运动以及低快门速度等因素造成的图像非均匀模糊，提出一种结合多尺度特征融合和多输入多输出编-解码器的去模糊算法。首先使用多尺度特征提取模块来提取较小尺度模糊图像的初始特征，该模块使用扩张卷积来以较少的参数量获得更大的感受野。其次，通过特征注意力模块来自适应地学习不同尺度特征中的有效信息，该模块利用小尺度图像的特征来生成注意图，能够有效地减少冗余特征。最后，使用多尺度特征渐进融合模块逐步融合不同尺度的特征，使得不同尺度特征信息能够进行互补。相比以往的使用多个子网堆叠的多尺度方法，文中使用单个网络就能提取多尺度特征，从而降低了训练难度。为了评估网络的去模糊效果和泛化性能，提出的算法在基准数据集 GoPro、HIDE 和真实数据集 RealBlur 上均进行了测试。在 GoPro 和 HIDE 数据集上的峰值信噪比值分别为 31.73 dB 和 29.39 dB，结构相似度值分别为 0.951 和 0.923，其结果均高于目前先进的去模糊算法，并且在真实数据集 RealBlur 上也取得了最佳效果。实验结果表明，提出的去模糊算法相比现有算法去模糊更为彻底，能有效地复原图像的边缘轮廓和纹理细节信息，并且能够提升后续高级计算机视觉任务的鲁棒性。

关键词：图像去模糊； 图像恢复； 深度学习； 多输入多输出； 多尺度网络

中图分类号：TP391.4 **文献标志码：**A **DOI：**10.3788/IRLA20220018

Image deblurring via multi-scale feature fusion and multi-input multi-output encoder-decoder

Zhao Qian, Zhou Dongming*, Yang Hao, Wang Changcheng, Li Miao

(School of Information Science & Engineering, Yunnan University, Kunming 650504, China)

Abstract: A deblurring method combining multi-scale feature fusion and a multi-input multi-output encoder-decoder is proposed for non-uniform blurred images caused by camera shake, fast motion of the captured object, and low shutter speed. Firstly, the initial features of smaller-scale blurred images are extracted using a multi-scale feature extraction module, which uses dilated convolution to obtain a larger receptive field with a smaller number of parameters. Second, the feature attention module is used to adaptively learn useful information from different scale features, which can effectively reduce redundant features by using features of small-scale images to generate attention maps. Finally, the multi-scale feature progressive fusion module is applied to gradually fuse features at different scales, making the information of different scale features to complement each other. Compared with

收稿日期:2022-01-06； **修订日期:**2022-03-04

基金项目:国家自然科学基金 (62066047, 61966037)

作者简介:赵倩，女，硕士生，主要从事基于深度学习的图像处理方面的研究。

导师(通讯作者)简介:周冬明，男，教授，博士生导师，博士，主要从事基于深度学习的图像处理、基于机器学习的生物信息处理等方面的研究。

recent multi-scale methods that use multiple subnets stacked on top of each other, we use a single network to extract multi-scale features, thus reducing the training difficulty. To evaluate the deblurring effect and generalization performance of the network, the proposed method is tested on both the benchmark datasets GoPro, HIDE, and the real dataset RealBlur. The peak signal-to-noise ratio values of 31.73 dB and 29.39 dB and the structural similarity values of 0.951 and 0.923 on the GoPro and HIDE datasets, respectively. The deblurring performance is higher than that of recent state-of-the-art deblurring methods, and it also has better performance on the RealBlur dataset containing real scenarios. The experimental results demonstrate that the proposed method is more effective than recent deblurring methods, can effectively restore the edge contour and texture detail information of images. In addition, our method can improve the robustness of subsequent high-level computer vision tasks.

Key words: image deblurring; image restoration; deep learning; multi-input multi-output; multi-scale networks

0 引 言

在图像拍摄过程中，由于相机抖动、拍摄对象快速移动或失焦等原因导致的图像质量大幅衰减的现象称为图像退化，将退化图像恢复为清晰图像的技术称为图像恢复技术，图像去模糊技术属于图像复原技术的一种。模糊图像会严重影响后续计算机视觉任务的性能，通常在目标检测[1−2]、图像分割[3]和目标跟踪[4]等高级计算机视觉任务中大多数都假设输入图像是无模糊的，一旦输入的图像是模糊的，这些任务往往无法准确检测或分割图像中的模糊对象。由于图像去模糊技术能显著提高输入模糊图像的后续计算机视觉任务的性能，图像去模糊技术受到了国内外学者的广泛关注[5−6]。传统的去模糊方法[7−12]通常对模糊核做出假设，对不同类型的均匀、非均匀以及深度感知模糊进行建模并施加各种约束条件，利用图像的先验信息求解模糊核，最后从给定的模糊图像中恢复出对应的清晰图像。尽管传统的方法易于实现，但这些方法大多对模糊模型的假设比较简单，不能很好地去除真实世界中复杂的非均匀模糊。此外，传统的去模糊方法计算推理复杂且大多需要多次迭代来优化参数，使得图片处理时间过长，从而限制了算法的实际应用。随着深入学习研究的深入，许多基于深度学习的去模糊算法[13−17]不断地被提出，此类方法不依赖于自然图像的先验知识，能够以端到端的方式学习模糊图像和对应的清晰图像之间的非线性映射关系，从而能更好地处理动态场景中的非均匀模糊。早期的深度学习方法[14]主要使用单一尺度的网络架构，但由于单尺度网络感受野较小，在编码上下文信息方面效率较低难以提取更全面的全局特征和局部特征。为此 Nah 等人[5]提出一种基于多尺度的去模糊网络，该网络由多个子网络组成，每个子网络输入一张缩小的图像，并以“从粗到细”的方式逐渐恢复清晰的图像。由于多尺度的方法被证明是有效的，许多基于多尺度的去模糊方法逐渐被提出[16−17]，然而这些多尺度方法都是将多个子网络堆叠到一起，使得网络训练更加复杂并且运行时间更长。因此文中提出一种结合多尺度特征融合和多输入多输出编-解码器的去模糊算法，不同于 Nah 等人[5]的多个子网输入不同尺度图像的方法，文中的多尺度特征能够输入到单一的编码器中，同时解码器能输出多张不同尺度的清晰图像，网络更加简单并且能有效去除图像模糊。具体来说，文中贡献如下：

(1) 提出一种结合多尺度特征融合和多输入多输出编-解码器的去模糊算法。相比其他堆叠多个子网的多尺度模糊方法，网络复杂度较低。

(2) 为了有效利用多尺度特征信息，分别提出了多尺度特征提取模块 (Multi-scale feature extraction module, MFEM) 和多尺度特征渐进融合模块。此外，文中基于 SAM[18] 设计了一个特征注意力模块 (Feature attention module, FAM) 来增强或抑制不同尺度的特征信息，从而提高网络学习并区分特征的能力。

(3) 文中利用峰值信噪比 (Peak signal to noise ratio, PSNR) 和结构相似性 (Structural similarity, SSIM) 对所提出的网络进行量化评估。大量的实验结果表明，在合成数据集 GoPro 和 HIDE 中，文中方法相较其他基

准方法具有更高的 PSNR 和 SSIM。在真实数据集 RealBlur-R 和 RealBlur-J 上的结果表明，文中方法具有更好的泛化性和鲁棒性。

(4) 为了进一步评估文中算法在后续高级计算机视觉任务上的应用价值，使用预先训练的 YOLOv4[1] 对模糊图像和去模糊后的图像进行目标检测。结果表明文中算法能够有效提升后续高级计算机视觉任务的性能。

1 相关工作

Yuan 等人[6] 使用模糊图像以及对应的同一场景下包含噪声的清晰图像来估计模糊核，通过利用噪声图像中清晰的细节信息来较好地估计初始核，并提出了残差反卷积来减少图像反卷积固有的振铃伪影。但由于对模糊核的估计过于单一且假设模糊核是空间不变的，使得该方法不能很好地处理真实相机抖动造成的非均匀模糊。为此，Whyte 等人[8] 提出了一个参数化几何模型，该模型能有效处理由于相机旋转引起相机抖动产生的非均匀图像模糊。但该方法忽略了相机平移造成的图像模糊，仅对相机旋转建立了几何模型。针对相机平移造成的运动模糊，Xu 等人[9] 提出了一种基于 L0 稀疏表示的运动去模糊方法。该方法在优化过程中不需要额外的滤波，仅需要少量的迭代就能够收敛。Hu 等人[11] 考虑到模糊图像中的光线条纹包含丰富的模糊信息，提出了一种利用光线条纹进行建模的弱光图像去模糊方法，该方法通过检测模糊图像中有用的光条纹来估计模糊核从而去除模糊。Pan 等人[12] 观察到模糊图像的暗通道具有更小的稀疏性，提出一种基于暗通道先验的图像盲去模糊方法，该方法不需要任何复杂的模糊核估计就能够去除非均匀模糊。

近年来，由于深度神经网络具有强大的特征学习以及非线性建模能力，其在目标检测，目标分割，图像恢复等计算机视觉任务中得到广泛应用。Sun 等人[13] 首先将卷积神经网络 (CNN) 应用到图像去模糊领域，该方法通过预测小图像块上运动模糊的概率分布来估计非均匀运动模糊核，并利用小图像块的先验来去除运动模糊。然而，使用具有均匀运动模糊的小图像块训练 CNN 忽略了较大区域上模糊图像和运动模糊核的映射关系，因此该方法去模糊性能不佳。之后，Gong 等人[15] 将整张图片的运动模糊表示为像素方向的线性运动模糊，再通过 CNN 直接估计模糊图像中的运动流来恢复出清晰图像。总的来说，早期基于 CNN 的去模糊方法大多通过利用 CNN 来估计模糊核从而复原图像，然而当模糊核估计不准确时，这些方法很难实现理想的去模糊效果。因此，最近的去模糊方法大多以端到端的方式直接训练无核估计的网络来复原图像。Nah 等人[5] 提出了一个基于高斯金字塔结构的多尺度卷积神经网络，这种由粗到细的网络结构能够充分提取图像的多尺度特征来恢复清晰图像。但使用独立的子网络分别训练每个尺度的图像，使得网络总体参数量较大且训练困难。为此，Tao 等人[19] 在不同的多尺度特征提取子网络中共享参数，在提升网络去模糊效果的同时，减少了网络参数并降低了运行时间。但该方法忽略了图像特征的尺度变化特性，所有子网共享参数的方式可能会丢失多尺度特征信息，使得网络不能有效的复原图像细节信息。Gao 等人在参考文献 [20] 的基础上提出了一种有效的参数选择性共享网络，并在网络的非线性变换模块中引入了一种新的嵌套跳跃连接结构来代替简单地堆叠卷积块来提高网络性能。Kupyn 等人[21−22] 将生成对抗网络 (GAN) 应用于图像去模糊，先后提出了 DeblurGAN、DeblurGAN-v2 使用生成对抗网络将模糊图像直接映射到清晰图像来去除模糊，然而这两种方法很难复原复杂场景下的非均匀动态模糊。Zhang 等人[23] 提出了一个端到端的多层次去模糊网络，该网络的每个子网能够提取由不同分割方式产生的小图像块的细节特征，并逐层融合提取到的特征信息来复原图像。Cai 等人[24] 将暗通道和亮通道先验信息嵌入神经网络中来聚合通道特征，并对网络进行稀疏正则化操作来提高网络性能。但先验知识的加入提高了模型建模的复杂度，在真实模糊场景下模型的泛化性能不佳。为此，Zhang 等人[25] 考虑到目前去模糊数据集大多为合成数据集，使用生成对抗网络生成真实的模糊图像，从而提高网络模型在真实场景下的去模糊性能。Park 等人[26] 提出了一种基于多时相递归神经网络的单幅图像去模糊算法，该算法首先将深度模糊分解成一系列轻度模糊，然后再以迭代的方式逐步去除模糊。Zou 等人[27] 提出一种基于小波变换的扩张网络去模糊算法，该算法使用具有不同扩张

率的扩张卷积来获得具有不同感受野的特征，并利用小波变换模块来恢复图像的纹理细节信息。

2 文中方法

2.1 网络整体架构

文中所提出的整体网络架构如图 1 所示，该网络包含四个子模块，分别为编码网络块 (Encode Block, EB)、解码网络块 (Decode Block, DB)、多尺度特征提取模块 (Multi-scale Feature Extraction Module, MFEM) 和多尺度特征渐进融合模块 (Progressive Multi-scale Feature Fusion Module, PMFM)。其中第一个编码网络块 EB_1 由一个卷积层和一个残差组 (Residual Group, RG) 构成，残差组由多个残差块 (Residual Block, RB) 堆叠而成，每个残差块包含两个 3×3 卷积。第二个编码网络块 EB_2 和第三个编码网络块 EB_3 在 EB_1 的基础上增加了特征注意力模块来自适应地学习不同尺度的有用特征并减少冗余特征。与第一个编码网络块 EB_1 对应的是第一个解码网络块 DB_1，该网络块由一个卷积层和一个 RG 组成，第二个解码网络块 DB_2 和第三个解码网络块 DB_3 在 DB_1 的基础上增加了转置卷积，利用转置卷积来进行上采样操作从而恢复特征图大小。

在编码阶段，首先将输入模糊图像 B_1、B_2 和 B_3 的分辨率大小缩放为 256×256、128×128 和 64×64。把分辨率大小为 256×256 的模糊图像 B_1 作为第一个编码网络块 EB_1 的输入，利用多尺度特征提取模块来提取分辨率大小为 128×128 的模糊图像 B_2 和分辨率大小为 64×64 的模糊图像 B_3 中的多尺度特征，然后再将提取到的特征分别输入到第二个编码网络块 EB_2 和第三个编码网络块 EB_3 中的特征注意力模块中来增强或抑制不同尺度的特征信息，使得更多有用的特征输入到下一模块。在每个编-解码网络块间使用多尺度特征渐进融合模块逐步融合 EB_1、EB_2 和 EB_3 提取到的多尺度特征，将其输入到解码网络块中在不同尺度清晰图像的监督下逐步复原图像。

在解码阶段，将多尺度特征渐进融合模块提取到的特征信息分别与第二个解码网络块 DB_2 和第三个解码网络块 DB_3 的输出特征进行融合，使得网络能够充分学习不同特征间的互补信息，减少图像细节信息的丢失。在每个解码网络块中，使用与输入模糊图像分辨率大小相同的清晰图像对不同尺度的输出图像进行监督。解码阶段中图像恢复过程可表示为：

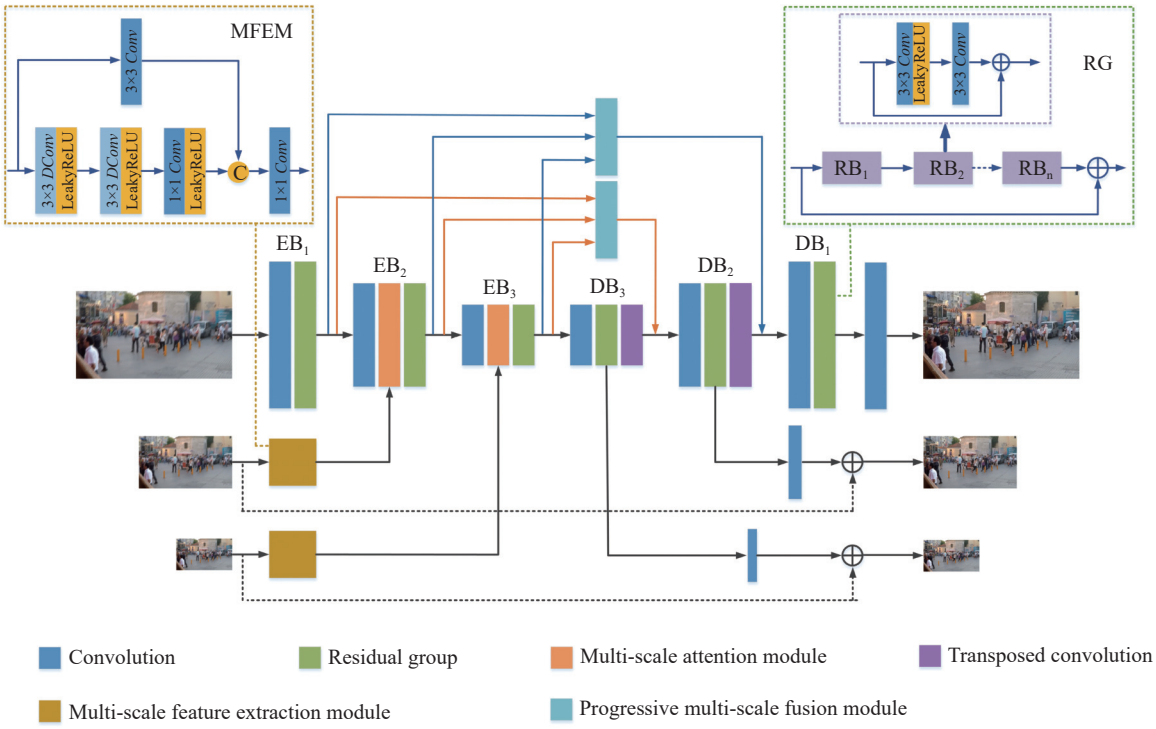


图 1 算法的网络架构

Fig.1 Network architecture of algorithm

$$I_n=\begin{cases} Conv(\mathrm{DB}_n(\mathrm{MFFM}_n^{\mathrm{out}},\mathrm{DB}_{n+1}^{\mathrm{out}}))+\mathrm{B}_n, & n=1,2 \\ Conv(\mathrm{DB}_n(\mathrm{EB}_n^{\mathrm{out}}))+\mathrm{B}_n, & n=3 \end{cases} \tag{1}$$

式中：$\mathrm{MFFM}_n^{\mathrm{out}}$为第 n 个多尺度特征融合模块的输出；$\mathrm{DB}_{n+1}^{\mathrm{out}}$和$\mathrm{EB}_n^{\mathrm{out}}$分别表示第 n+1 个解码网络块和第 n 个编码网络块的输出；DB_n表示在第 n 个解码网络块上进行解码操作；$Conv$为卷积操作，将特征图通道数恢复至 3 通道来得到清晰图像。

2.2 多尺度特征提取模块

为了提取不同分辨率大小的模糊图像 B_2 和 B_3 中的初始特征，文中设计了一个多尺度特征提取模块。如图 2 所示，该模块通过两条支路并行提取特征信息。其中一条支路中首先使用两个扩张卷积来捕捉模糊图像中每个像素点的全局特征和局部特征。扩张卷积通过在标准卷积的卷积核中插入空洞，以此来扩大感受野，相比普通卷积能够以较少的参数量获得更大的感受野。其次再将提取到的特征输入一个普通 1×1 卷积进行特征细化，其中卷积层后的激活函数均为 LeakyReLU 激活函数。此外由于扩张卷积的卷积结果之间没有相关性，这会导致部分局部信息的丢失，为了减少传输过程中细节信息的丢失，MFEM 在另一条支路中使用一个 3×3 卷积来提取更精细的特征，在不增加过多计算量的情况下同时获得全局和局部的特征信息。

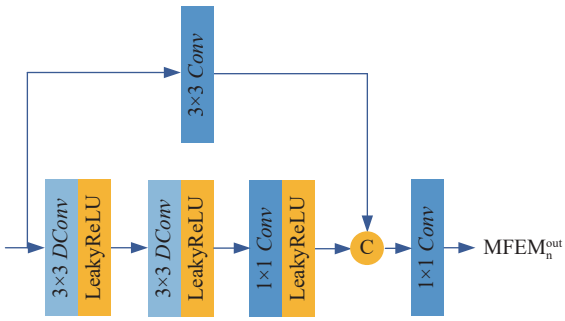


图 2 多尺度特征提取模块

Fig.2 Multi-scale feature extraction module

最后将两条支路的特征图进行拼接，并使用一个 1×1 卷积融合特征信息并调整通道数。该模块可以描述为：

$$F_1=\delta(Conv_3(\delta(DConv_3(\delta(DConv_3(\mathrm{B}_n)))))) \tag{2}$$

$$F_2=Conv_3(\mathrm{B}_n) \tag{3}$$

$$\mathrm{MFEM}_n^{\mathrm{out}}=Conv_1(Concat(F_1,F_2)) \tag{4}$$

式中：B_n 为不同分辨率大小的模糊图像且 n 的取值为 2 或 3；F_1 为 B_n 通过第一条支路后输出的特征图；F_2 为 B_n 通过第二条支路后输出的特征图；$\mathrm{MFFM}_n^{\mathrm{out}}$为多尺度特征提取模块的输出；$Conv_n$ 表示卷积核大小 $n\times n$ 的普通卷积；$DConv_3$ 表示扩张卷积，卷积核大小为 3×3 且扩张率为 3；δ 表示 LeakyReLU 激活函数。

2.3 特征注意力模块

为了能够自适应地学习不同尺度特征中的有用信息，受 SAM[18] 启发，文中设计了一个特征注意力模块。与 SAM 不同，文中设计的 FAM 不需要使用清晰图像进行监督来生成注意力图，降低了网络的复杂度。FAM 利用 EB 和 MFEM 的输出生成注意力图，为不同的特征图在空间上和通道上赋予不同的权重，从而增强或抑制不同尺度的特征信息，以此来提高网络学习并区分特征的能力。该模块能有效的减少冗余特征，使得更多有用的特征传播到下一模块中。

如图 3 所示，FAM 有两个输入$\mathrm{EB}_n^{\mathrm{out}\downarrow}$和$\mathrm{MFFM}_n^{\mathrm{out}}$，其中当 n=1 时，$\mathrm{EB}_1^{\mathrm{out}\downarrow}$为第一个编码网络块输出特征通过一个步长为 2 的卷积下采样操作后得到的特征图，该特征图大小为 128×128，$\mathrm{MFFM}_1^{\mathrm{out}}$为 MFEM 提取到的特征图大小为 128×128 的多尺度特征。当 n=2 时，$\mathrm{EB}_2^{\mathrm{out}\downarrow}$为第一个编码网络块输出特征通过一个步长为 2 的卷积下采样操作后得到的特征图，该特征图大小为 64×64，$\mathrm{MFFM}_2^{\mathrm{out}}$为 MFEM 提取到的特征图大小为 64×64 的多尺度特征。F_x 是$\mathrm{EB}_n^{\mathrm{out}\downarrow}$和$\mathrm{MFFM}_n^{\mathrm{out}}$进行逐元素相加后得到的特征图，将其通过 1×1 卷积后并输入到 Sigmoid 激活函数中得到特征注意力图，利用特征注意力图中 0~1 之间的元素数值大小来表示不同信息在空间上和通道上的权重值，元素数值越大说明其对应位置特征信息获得了更高的关注，使得

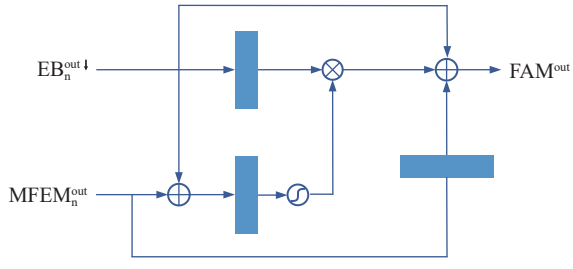


图 3 特征注意力模块

Fig.3 Feature attention module

网络能够自适应地对特征图 F_x 中的关键信息进行学习。

将特征注意力图与$\mathrm{EB}_n^{\mathrm{out}\downarrow}$经过 1×1 卷积后得到的特征进行逐元素相乘并输出特征 F_y, 以此来重新校准$\mathrm{EB}_n^{\mathrm{out}\downarrow}$中的特征信息。最后, 将 F_y、经过卷积操作后的多尺度特征$\mathrm{MFFM}_n^{\mathrm{out}}$和$\mathrm{EB}_n^{\mathrm{out}\downarrow}$进行逐元素相加得特征注意力模块的输出 $\mathrm{FAM}^{\mathrm{out}}$, 该模块可描述为:

$$F_x = \mathrm{EB}_n^{\mathrm{out}\downarrow} \oplus \mathrm{MFEM}_n^{\mathrm{out}} \tag{5}$$

$$F_y = \mathrm{Sig}(Conv(F_x)) \otimes Conv\left(\mathrm{EB}_n^{\mathrm{out}\downarrow}\right) \tag{6}$$

$$\mathrm{FAM}^{\mathrm{out}} = F_y \oplus \mathrm{EB}_n^{\mathrm{out}\downarrow} \oplus Conv(\mathrm{MFEM}^{\mathrm{out}}) \tag{7}$$

式中: ⊕表示特征图逐元素相加; ⊗表示特征图逐元素相乘; Sig 表示 Sigmoid 激活函数。

2.4 多尺度特征渐进融合模块

在传统的编-解码网络中, 通常利用跳跃连接简单地将编码网络块中的单一尺度特征信息传递到对应的解码网络块中, 使得解码网络块不能充分利用特征图的不同尺度特征信息, 且网络间信息流不够灵活。为此, 文中提出一个多尺度特征渐进融合模块(Progressive Multi-scale Feature Fusion Module, PMFM), 对第一个编码网络块的输出$\mathrm{EB}_1^{\mathrm{out}\downarrow}$、第二个编码网络块的输出$\mathrm{EB}_2^{\mathrm{out}\downarrow}$和第三个编码网络块的输出$\mathrm{EB}_3^{\mathrm{out}\downarrow}$逐步进行融合。该模块能够让信息流在不同尺度之间进行交互, 在从小尺度图像中捕获更多的上下文信息的同时, 又能从大尺度图像中学习更多的细节特征。使得小尺度特征图信息能够被充分利用并对大尺度特征图信息进行互补, 从而有效的融合不同尺度信息。与特征图简单拼接或逐元素相加的特征融合方式不同, 文中提出的 PMFM 能够将小尺度特征图逐渐的融合到大尺度特征图中从而能够有效利用不同尺度的特征信息。如图 4 所示, 该模块有 3 个输入即$\mathrm{EB}_1^{\mathrm{out}\downarrow}$、$\mathrm{EB}_2^{\mathrm{out}\downarrow}$和$\mathrm{EB}_3^{\mathrm{out}\downarrow}$, 其特征图大小分别为 256×256、128×128、64×64。

在该模块中, 将$\mathrm{EB}_1^{\mathrm{out}\downarrow}$通过一个 1×1 的卷积进行特征细化, 并将细化后得到的特征与$\mathrm{EB}_2^{\mathrm{out}\downarrow}$经过双线性上采样后得到的特征图进行拼接操作得融合后的特征图 F_1, 然后将 F_1 通过 1×1 卷积后与$\mathrm{EB}_3^{\mathrm{out}\downarrow}$经过双线性上采样后得到的特征图进行拼接得融合后的特征图 F_2。并且为了减少在特征融合时较大尺寸特征图中高频细节信息的丢失, 将第一个编码网络块提取

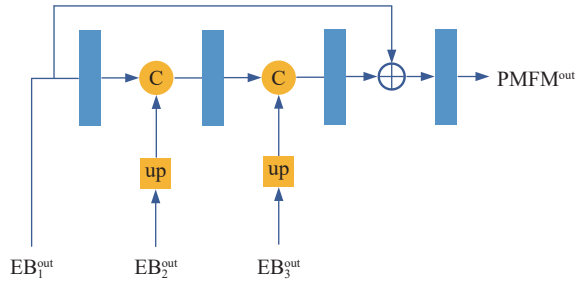


图 4 多尺度特征渐进融合模块

Fig.4 Progressive multi-scale feature fusion module

到的高分辨率特征信息$\mathrm{EB}_1^{\mathrm{out}\downarrow}$与特征图 F_2 进行逐元素相加得到特征图 F_3。最后, 使用 1×1 卷积对特征图 F_3 进行信息整合并输出 $\mathrm{PMFM}^{\mathrm{out}}$, 具体流程如下:

$$F_1 = Concat(Conv(\mathrm{EB}_1^{\mathrm{out}}), \mathrm{up}(\mathrm{EB}_2^{\mathrm{out}})) \tag{8}$$

$$F_2 = Concat(Conv(F_1), \mathrm{up}(\mathrm{EB}_3^{\mathrm{out}})) \tag{9}$$

$$F_2 = Conv(F_2) \oplus \mathrm{EB}_1^{\mathrm{out}} \tag{10}$$

$$\mathrm{PMFM}^{out} = Conv(F_3) \tag{11}$$

式中: *Conv* 为卷积操作, 卷积核大小均为 1×1; *Concat* 表示对特征图在通道上进行拼接; up 为双线性上采样操作; $\mathrm{PMFM}^{\mathrm{out}}$ 为多尺度特征融合模块的输出。

2.5 损失函数

在图像恢复任务中, 训练网络常用的损失函数是均方误差 (Mean Square Error, MSE) 损失函数。MSE 对网络输出图像和真实图像在对应像素点上计算差值并进行平方, 但由于平方操作通常会惩罚较大的误差值并容忍较小的误差值, 会导致输出结果过度平滑且图像边缘模糊。因此, 文中在训练网络时采用 L_1 损失和 SSIM 损失组成的混合损失, 总体损失函数 L 可以表示为:

$$L = \frac{1}{M}\sum_{i=1}^{M}\left\{\sum_{n=1}^{3} L_1(X_i, Y_i) + \lambda \sum_{n=1}^{3} \mathrm{SSIM}(X_i, Y_i)\right\} \tag{12}$$

式中: M 为训练集每个 batch 输入网络的图片数目; n 表示第 n 个尺度, 当 n 为 1 时, 表示对第一个尺度中分辨率大小为 256×256 的图像求损失; 当 n 为 2 时, 表示对第二个尺度中分辨率大小为 128×128 的图像求损失; 当 n=3 时, 表示对第三个尺度中分辨率大小为 64×64 的图像求损失; X 为文中网络输出的图像; Y 为对应的清晰图像; λ 为权重因子。

(1) L_1 损失

L_1 损失不会过度惩罚较大的误差值, 能够保留图

像结构和边缘信息，L_1 损失数学表达式为：

$$L_1 = \|X - Y\|_1 \tag{13}$$

(2) SSIM 损失

由于 SSIM 损失[28]是基于局部对比度、亮度和结构等局部图像特征来进行计算，所以使用 SSIM 损失来训练网络能获得更好的图像视觉效果，且能够较好地保留图像细节等高频信息，SSIM 损失可以表示为：

$$\mathrm{SSIM}(X,Y) = \frac{(2u_X u_Y + C_1)(2\sigma_{XY} + C_2)}{(u_X^2 + u_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)} \tag{14}$$

式中：u_X及u_Y、σ_X及σ_Y和σ_{XY}分别为文中网络输出的图像和对应的清晰图像的平均值、方差和协方差；C_1 和 C_2 均为常数使得整体公式稳定，文中取 C_1=0.01，C_2=0.03。

3 实验结果与分析

3.1 实验数据集

文中使用 GoPro 数据集来训练网络模型，该数据集由 Nah 等人[5]利用 GoPro 运动相机拍摄每秒 240 帧的视频序列，并对连续的多帧短曝光图像取平均来生成模糊图像。共包含 3214 对分辨率为 1280×720 的清晰图像和模糊图像，其中文中使用 2103 对图像进行训练，1111 对图像进行测试。为了评估网络的泛化能力，文中还使用 HIDE 数据集和 RealBlur 数据集对网络模型进行测试。HIDE 数据集由 Shen 等人[29]提出，该数据集主要包含多种场景下的人物运动模糊，共由 8422 对模糊图像和清晰图像组成，其中测试集由 2025 对的模糊图像和清晰图像组成。与 GoPro 和 HIDE 数据集不同，RealBlur 数据集[30]的图像对是在真实环境中采集的，包含 232 个不同场景的 4738 对图像，该数据集由共享相同图像内容的两个子集组成，其中一个子集 RealBlur-R 由相机原始图像组成，另一个子集 RealBlur-J 由经过相机处理后的 JPEG 图像组成。其中训练集包含 3758 对图像，测试集包含 980 对图像。

3.2 实验细节和参数设置

文中方法基于 PyTorch 框架实现，采用 NVIDIA GeForce RTX 3060 12 G GPU 对模型进行训练和测试。采用动量衰减指数 β_1=0.9，β_2=0.999 的 Adam 优化器更新网络参数，迭代次数为 1000 次。初始学习率设置为 10^{-4}，每迭代 200 次学习率减半。在每一轮训练迭代过程中，随机选取 6 张裁剪为 256 pixel×256 pixel 大小的图像作为网络输入，并通过随机旋转和垂直翻转的方式来增强数据。文中使用 GoPro 数据集对网络进行训练，并将训练好的模型在 GoPro 数据集、HIDE 数据集和真实数据集 RealBlur 上测试。

3.3 结果分析

3.3.1 量化评价

文中算法与经典去模糊算法以及目前基于深度学习的主流算法进行比较，如表 1 所示，经典算法有：Whyte 等人[8]、Xu 等人[9]和 Pan 等人[12]提出的算法，目前基于深度学习的去模糊算法有：DeblurGAN-v1[21]，DeblurGAN-v2[22]，SRN[19]，MT-RNN[26]，DBGAN[25]，DMPHN[23]以及 Nah 等人[5]和 Gao 等人[20]提出的算法。由于以上基于深度学习的去模糊算法均使用 GoPro 数据集对网络进行训练，所以文中直接使用作者公开发布的源代码对 GoPro 数据集、HIDE 数据集和 RealBlur 数据集进行测试，并采用峰值信噪比[31](Peak Signal-to-Noise Ratio, PSNR)、结构相似性[28](Structural Similarity, SSIM) 作为评价指标对所恢复的图像质量进行定量评价。由表 1 数据可知，文中算法在四个数据集上均取得了最佳效果，在 GoPro 数据集上 PSNR 为 31.73 dB，SSIM为 0.951，较 DMPHN[23]分

表 1 在各个数据集上的测试结果

Tab.1 Test results on various datasets

Method	GoPro		HIDE		RealBlur-R		RealBlur-J	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Xu et al.[9]	22.85	0.817	21.78	0.723	31.63	0.872	24.88	0.822
Whyte et al.[8]	24.47	0.843	22.81	0.735	30.56	0.854	25.92	0.844
Pan et al.[12]	24.73	0.876	23.92	0.763	32.92	0.891	25.79	0.854
DeblurGAN-v1[21]	25.64	0.859	23.96	0.809	34.28	0.932	27.01	0.865
Nah et al.[5]	27.83	0.915	25.73	0.874	33.92	0.947	27.11	0.876
DeblurGAN-v2[22]	29.08	0.918	27.51	0.884	34.16	0.942	27.17	0.877
SRN[19]	30.24	0.934	28.36	0.903	34.24	0.937	27.08	0.876
Gao et al.[20]	30.96	0.942	29.1	0.913	34.06	0.943	26.82	0.868
MT-RNN[26]	31.12	0.944	29.15	0.917	34.19	0.95	26.74	0.869
DBGAN[25]	31.18	0.946	28.94	0.915	32.99	0.926	24.87	0.821
DMPHN[23]	31.39	0.947	29.1	0.916	34.12	0.948	26.63	0.865
Ours	**31.73**	**0.951**	**29.39**	**0.923**	**34.35**	**0.951**	**27.19**	**0.878**

别提升了 0.34 dB 和 0.004, 在 HIDE 数据集上 PSNR 为 29.39 dB, SSIM 为 0.923, 与 MT-RNN[26] 相比 PSNR 和 SSIM 分别提升了 0.24 dB和 0.006。在真实数据集 RealBlur-J 和 RealBlur-R 上, 文中算法与其他基于深度学习的去模糊算法差异较小, 但仍达到了最佳效果, 与第二名 DeblurGAN-v2[22] 相比 PSNR 和 SSIM 分别提升了 0.02 dB 和 0.001。由四个数据集上的测试结果可知, 相比其他方法, 文中算法去模糊效果较好, 并具有更好的泛化能力和鲁棒性。

3.3.2 主观效果分析

除了通过评价指标 PSNR 和 SSIM 对文中算法进行定量分析, 文中还从 GoPro 数据集、HIDE 数据集、RealBlur-J 和 RealBlur-R 数据集随机选取不同场景的图像与目前主流算法进行视觉效果对比分析。图 5、6、7 分别展示了不同算法在 GoPro 数据集上、HIDE 数据集和 RealBlur 数据集的去模糊图像视觉效果。从图 5 中可以看出, Xu 等人[9] 和 Pan 等人[12] 提出的传统算法难以处理非均匀模糊, 使得重建后的图像中仍存在大量模糊。通过对比图 5 中第一张图片中的汽车后视镜可以得出, 文中算法对后视镜的边缘轮廓重建效果最好, DeblurGAN-v2[22] 去除模糊不彻底, 不能很好的恢复出主体轮廓, 而 DBGAN[25] 能恢复出后视镜轮廓, 但存在伪影无法重建图像细节。通过对比图 5 中第四张图片中的车牌数字可以看出, 在所有对

图 5 GoPro 数据集上各模型的主观效果对比

Fig.5 Comparison of the visual results of each model on the GoPro dataset

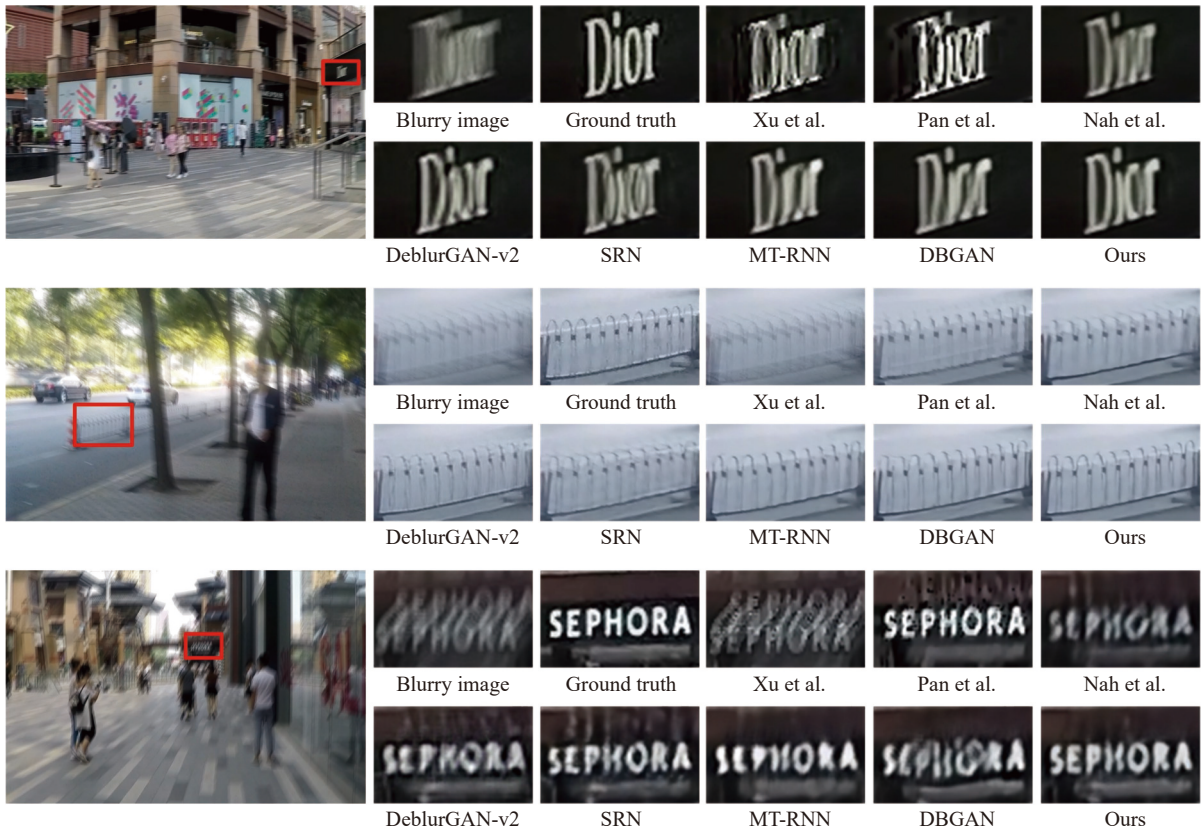


图 6 HIDE 数据集上各模型的主观效果对比

Fig.6 Comparison of the visual results of each model on the HIDE dataset

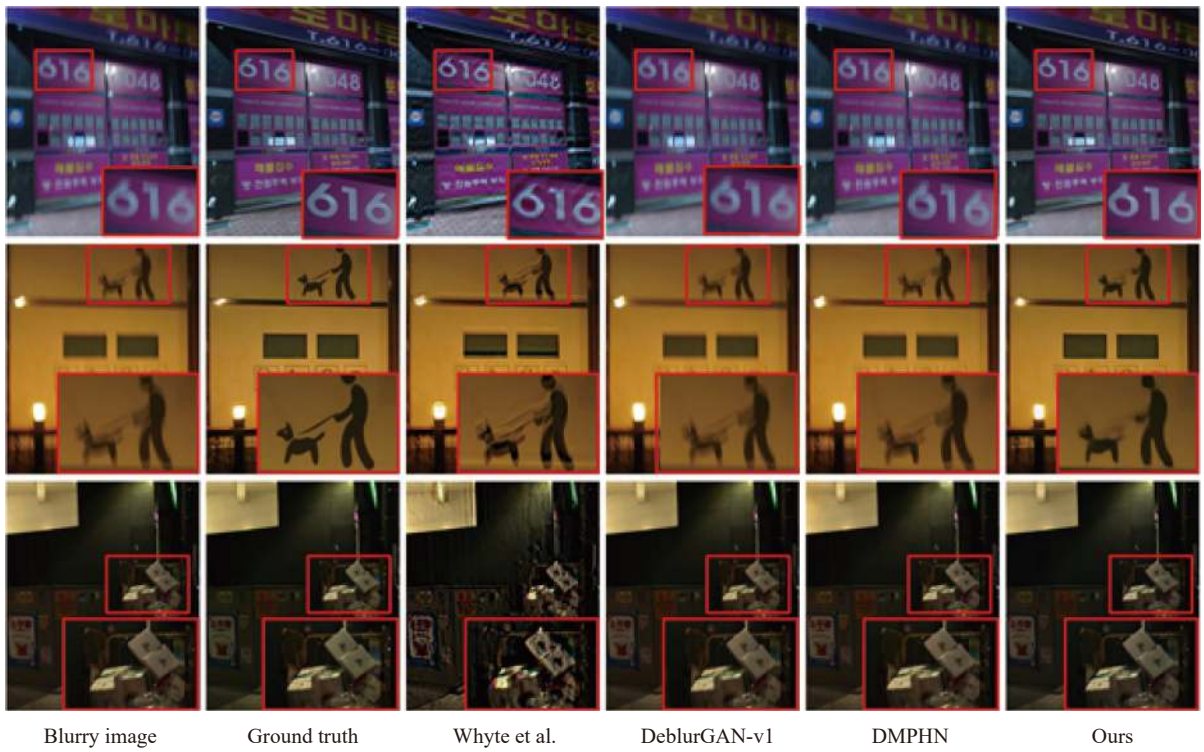


图 7 RealBlur 数据集上各模型的主观效果对比

Fig.7 Comparison of the visual results of each model on the RealBlur dataset

比方法中只有文中算法能清晰的恢复出数字, SRN[19] 和 MT-RNN[26] 等算法重建的图像边缘不够清晰, 无法看出清晰的数字。

通过对比图 6 中第二张图片的字母区域可以看出, 文中算法对字母区域的恢复最为清晰, Nah 等人[5] 提出的算法能去除部分模糊, 但不能清晰的恢复出每个字母, SRN[19] 相比 Nah 等人[5] 提出的算法去模糊效果明显提高, 但仍不能有效恢复出字母的边缘等高频信息。对比图 7 中的第一张图片的数字区域可以看出, 文中算法能很好地恢复出数字轮廓, 取得了较好的主观效果。

通过 3.3.1 节对表 1 中各方法进行定量对比分析, 以及 3.3.2 节对图 5、图 6 和图 7 的主观视觉效果对比分析的结果可知, 文中方法能够很好地处理非均匀模糊, 对图像边缘轮廓和细节等信息重建效果更好, 去模糊更为彻底。同时相比目前主流的去模糊方法, 文中方法在基准数据集 GoPro、HIDE 和真实数据集 RealBlur-R、RealBlur-J 上均取得了最佳效果, 具有更好的泛化能力和鲁棒性。

3.4 消融实验

3.4.1 编-解码器输入输出个数讨论

为了验证提取多尺度特征对网络去模糊性能提升的有效性, 文中在 GoPro 数据集上训练并测试了编-解码网络不同多输入多输出个数 N(尺度数) 的 PSNR 和 SSIM。当 N=1 时, 编-解码网络只输入输出一张分辨率大小为 256×256 的单一尺度图像; 当 N=2 时, 编-解码网络输入输出两张分辨率大小分别为 256×256 和 128×128 的图像; 当 N=3 时, 编-解码网络输入输出三张分辨率大小分别为 256×256、128×128 和 64×64 的图像。当 N=4 时, 编-解码网络输入输出四张分辨率大小分别为 256×256、128×128、64×64 和 32×32 的图像。测试结果如表 2 所示, 从表 2 中可以看出, 当使用单输入单输出编解码器时, PSNR 和 SSIM 值分别为 31.22 和 0.944。当 N=3 时, PSNR 和 SSIM 分别提升了 0.51 和 0.007, 证明了多尺度特征提取的有效性。当 N=4 时, PSNR 和 SSIM 相比 N=3 仅分别提升了 0.09 和 0.001, 这是因为输入图像的尺度过小只包含很少的信息, 这些特征信息对网络的去模糊性能提升较小。此外, 现有的多尺度去模糊方法[5, 20, 24] 都是三尺度结构网络。因此, 文中选择 N=3 作为最终网络模型的尺度数。

表 2 编-解码器不同输入输出个数的消融实验

Tab.2 Ablation study on different number of input and output numbers of encoder-decoder

N	PSNR	SSIM
1	31.22	0.944
2	31.58	0.949
3	31.73	0.951
4	31.85	0.952

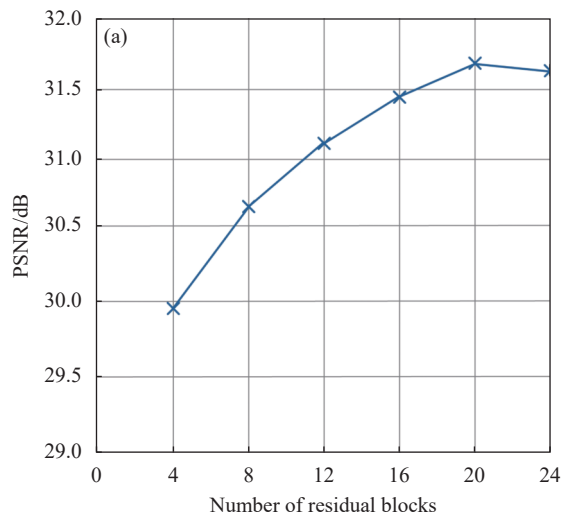
3.4.2 残差组层数分析

文中算法设计的编-解码网络块中的残差组由多个残差块堆叠而成, 为了评估残差块的个数对网络性能的影响, 在 GoPro 数据集上对残差块的个数 M 做了消融实验, M 的取值分别为 4, 8, 12, 16, 18, 20, 24。实验结果如表 3 所示, 当 M=4 时, PSNR 和 SSIM 值较低, 分别为 27.97 和 0.933。随着残差块数量的增多, PSNR 和 SSIM 值也随之提升, 当 M=20 时, PSNR 和 SSIM 值分别为 31.73 和 0.951。由图 8 可知, 当 M 大于 20 时, PSNR 和 SSIM 的增加速率减缓, 为了平衡参数量和去模糊性能, 文中取 M 值为 20。

表 3 不同残差块个数的消融实验

Tab.3 Ablation study on different number of residual blocks

M	4	8	12	16	20	24	28
PSNR	29.97	30.63	31.21	31.42	31.73	31.75	31.76
SSIM	0.933	0.939	0.944	0.948	0.951	0.951	0.951



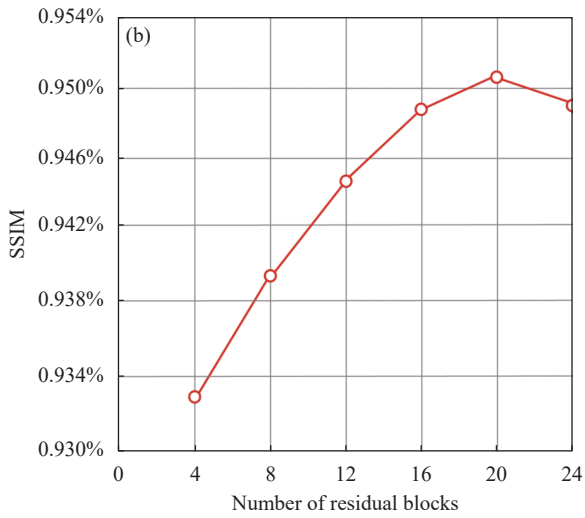


图 8 残差块个数分析

Fig.8 Analysis of the number of residual blocks

3.4.3 网络模块组合

文中算法在 GoPro 数据集进行了多种模块不同组合的消融实验。首先，评估了不同的特征融合方法对网络性能的影响，将文中所提出的多尺度特征融合模块 MPFM 与普通特征图通道拼接 (Concatenate) 和逐元素相加 (Sum) 的融合方法进行比较。如表 4 所示，使用逐级特征融合模块与使用逐元素相加的融合方法相比，PSNR 提高了 0.16，SSIM 提高了 0.006。与使用特征图拼接的融合方法相比，PSNR 提高了 0.08，SSIM 提高了 0.003。其次，为了验证提取多尺度特征提取模块 MFEM 对网络去模糊性能的影响，用一个 3×3 卷积来代替文中的多特征提取模块。从表 4 中可以看出，使用 MFEM 时 PSNR 提高了 0.09，SSIM 提高了 0.003。最后，为了验证特征注意力模块 FAM 的有效性，对 FAM 进行消融实验，由表 4 可知，去掉 FAM 模块后，PSNR 下降了 0.03，SSIM 下降了 0.001。图 9 为 FAM 所生成注意力图的可视化结果，从图中可以看出，相较于背景，人物运动造成的模糊区域获得了更高的权重，这说明 FAM 能够强调不同的局部特征并关注模糊程度较高的区域。

表 4 不同模块组合的消融实验

Tab.4 Ablation study with different module combinations

Module	Combination of different modules				
MPFM	√	Sum	Concatenate	√	√
MFEM	√	√	√	×	√
FAM	√	√	√	√	×
PSNR	31.73	31.57	31.65	31.64	31.7
SSIM	0.951	0.945	0.948	0.948	0.95

图 9 FAM 注意力图可视化

Fig.9 Attention map visualization for FAM

3.5 评估文中算法对目标检测性能的影响

图像去模糊是一项基本的低级计算机视觉任务，其最终目标是服务于后续高级计算机视觉任务。由于相机抖动、物体快速运动以及低快门速度等因素造成的图像非均匀模糊，在很大程度上会降低高级计算机视觉任务的性能。然而现有的目标检测算法往往假设输入图像是无模糊的，使得这些算法无法精确的检测到模糊图像中的对象。为了评估文中算法在目标检测算法中的有效性，使用 YOLOv4[1] 对模糊图像和去模糊后的图像进行目标检测。如图 10 所示，未经文中算法处理的模糊图像识别率较低或无法识别出对象，而去模糊后的图像识别率显著提高，能够识别出更多对象。因此，文中算法能够通过有效去除模糊来增强目标检测算法的鲁棒性。

(a) 模糊图像
(a) Blurry image

(b) 去模糊图像
(b) Deblurred image

(c) 模糊图像目标检测结果
(c) Object detection result of blurry image

(d) 去模糊图像目标检测结果
(d) Object detection result of deblurred image

图 10 目标检测视觉效果对比

Fig.10 Comparison of visual results of target detection

4 结 论

针对现有的大多数去模糊算法仍存在去模糊不彻底且图像细节信息丢失等问题，文中提出一种结合多尺度特征融合和多输入多输出编-解码器的去模糊算法。首先通过一个基于扩张卷积多尺度特征提取模块来提取较小尺度图像的特征，然后通过特征注意力模块来为不同尺度的特征图在空间上和通道上赋予不同的权重，从而提高网络学习并区分特征的能力。提出了一个多尺度特征渐进融合模块不同尺度的特征逐步融合在一起，能够减少了网络传输过程中高频细节信息的丢失。此外，为了降低网络训练的复杂度，区别于堆叠多个编-解码子网来输入和输出多尺度图像的方式，文中网络模型使用单一编-解码结构，将多尺度图像输入输出到同一个编-解码器中，以“从粗到细”的方式逐步恢复清晰图像。实验结果表明，文中算法在基准数据集 GoPro 和 HIDE 以及真实数据集 RealBlur 上相较于目前先进的去模糊算法均取得了较好的客观评价和主观视觉效果，并且能够提升后续计算机视觉任务的性能。

参考文献：

[1] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[DB/OL]. (2020-04-23)[2022-01-06]. https://doi.org/10.48550/arXiv.2004.10934.

[2] Li Weipeng, Yang Xiaogang, Li Chuanxiang, et al. An improved semi-supervised transfer learning method for infrared object detection neural network [J]. ***Infrared and Laser Engineering***, 2021, 50(3): 20200511. (in Chinese)

[3] Zhang X, Xu H, Mo H, et al. Dcnas: Densely connected neural architecture search for semantic image segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 13956-13967.

[4] Wang Z, Zheng L, Liu Y, et al. Towards real-time multi-object tracking[C]//Computer Vision–ECCV 2020, 2020: 107-122.

[5] Nah S, Hyun Kim T, Mu Lee K. Deep multi-scale convolutional neural network for dynamic scene deblurring[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 3883-3891.

[6] Wu Di, Zhao Hongtian, Zheng Shibao. Motion deblurring method based on DenseNets [J]. ***Journal of Image and Graphics***, 2020, 25(5): 890-899. (in Chinese)

[7] Yuan L, Sun J, Quan L, et al. Image deblurring with blurred/noisy image pairs [J]. ***ACM Transactions on Graphics***, 2007, 26(3): 1-es.

[8] Whyte O, Sivic J, Zisserman A, et al. Non-uniform deblurring for shaken images [J]. ***International Journal of Computer Vision***, 2012, 98(2): 168-186.

[9] Xu L, Zheng S, Jia J. Unnatural l0 sparse representation for natural image deblurring[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 1107-1114.

[10] Wang Sha, Chen Yueting, Feng Huajun, et al. TwIST-TV regularization based image deblurring method [J]. ***Infrared and Laser Engineering***, 2014, 43(6): 2000-2006. (in Chinese)

[11] Hu Z, Cho S, Wang J, et al. Deblurring low-light images with light streaks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 3382-3389.

[12] Pan J, Sun D, Pfister H, et al. Blind image deblurring using dark channel prior[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1628-1636.

[13] Sun J, Cao W, Xu Z, et al. Learning a convolutional neural network for non-uniform motion blur removal[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 769-777.

[14] Mao X, Shen C, Yang Y B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections [J]. ***Advances in Neural Information Processing Systems***, 2016, 29: 2802-2810.

[15] Gong D, Yang J, Liu L, et al. From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2319-2328.

[16] Liu Pengfei, Zhao Huaici, Cao Feidao. Blind deblurring of noisy and blurry images of multi-scale convolutional neural network [J]. ***Infrared and Laser Engineering***, 2019, 48(4): 0426001. (in Chinese)

[17] Chen Qingjiang, Hu Qiannan, Li Jinyang. Image deblurring based on multi-scale alternating connection residual network [J]. ***Optics and Precision Engineering***, 2021, 29(7): 1686-1694. (in Chinese)

[18] Zamir S W, Arora A, Khan S, et al. Multi-stage progressive image restoration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 14821-14831.

[19] Tao X, Gao H, Shen X, et al. Scale-recurrent network for deep image deblurring[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8174-8182.

[20] Gao H, Tao X, Shen X, et al. Dynamic scene deblurring with parameter selective sharing and nested skip connections[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 3848-3856.

[21] Kupyn O, Budzan V, Mykhailych M, et al. Deblurgan: Blind motion deblurring using conditional adversarial networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8183-8192.

[22] Kupyn O, Martyniuk T, Wu J, et al. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 8878-8887.

[23] Zhang H, Dai Y, Li H, et al. Deep stacked hierarchical multi-patch network for image deblurring[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 5978-5986.

[24] Cai J, Zuo W, Zhang L. Dark and bright channel prior embedded network for dynamic scene deblurring [J]. ***IEEE Transactions on Image Processing***, 2020, 29: 6885-6897.

[25] Zhang K, Luo W, Zhong Y, et al. Deblurring by realistic blurring[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 2737-2746.

[26] Park D, Kang D U, Kim J, et al. Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training[C]//European Conference on Computer Vision, 2020: 327-343.

[27] Zou W, Jiang M, Zhang Y, et al. SDWNet: A straight dilated network with wavelet transformation for image deblurring[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 1895-1904.

[28] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity [J]. ***IEEE Transactions on Image Processing***, 2004, 13(4): 600-612.

[29] Shen Z, Wang W, Lu X, et al. Human-aware motion deblurring[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 5572-5581.

[30] Rim J, Lee H, Won J, et al. Real-world blur dataset for learning and benchmarking deblurring algorithms[C]//European Conference on Computer Vision, 2020: 184-201.

[31] Huynh-Thu Q, Ghanbari M. Scope of validity of PSNR in image/video quality assessment [J]. ***Electronics Letters***, 2008, 44(13): 800-801.