

一种多尺度的图像动态场景盲去模糊网络^{*}

唐述, 万盛道, 谢显中, 杨书丽, 黄容, 顾佳, 郑万鹏



(重庆邮电大学 计算机网络和通信技术重庆市重点实验室, 重庆 400065)

通信作者: 唐述, E-mail: tangshu@cqupt.edu.cn

摘要: 近几年, 基于卷积神经网络 (convolutional neural network, CNN) 的单幅图像动态场景盲去模糊 (single image dynamic scene blind deblurring, SIDSBD) 方法已经取得了巨大的进步. 其成功主要是源于多尺度模型或者多块模型、编解码器架构的设计和残差块结构的设计 3 个方面. 基于此, 提出了一种新的多尺度卷积神经网络 (multi-scale convolutional neural network, MSCNN) 来进一步开发多尺度模型、编解码器架构和残差块结构的优势, 以实现更高质量的动态场景盲去模糊. 首先, 受到空间金字塔池化 (spatial pyramid pooling, SPP) 和多块模型的启发, 提出了一种分等级的多块通道注意力机制 (hierarchical multi-patch channel attention, HMPCA). 提出的 HMPCA 通过利用特征图的全局特征统计量和局部特征统计量来自适应地对特征图进行逐通道的权重赋值. 因为利用了局部信息, 因此 HMPCA 可以被认为是增加了通道方向的感受野, 也正因如此, 提出的 HMPCA 能够进一步增强网络的表达能力. 其次, 不同于现有的多尺度模型, 发展出了一种新的多尺度模型, 该模型中的每个尺度是由多个编码器和多个解码器构成的. 因为 HMPCA, 使得同一尺度内的编码器和解码器并不完全相同, 因此提出的多尺度模型可以被看作是增加了编解码器的深度, 因此能够提升每一个尺度的去模糊性能, 最终实现更高质量的动态场景盲去模糊. 大量的实验结果表明: 提出的方法较近几年的一些成功的 SIDSBD 方法相比, 能够复原出更高质量的去模糊图像, 在客观的评价指标和主观的视觉效果上均有显著的改进.

关键词: 卷积神经网络; 动态场景盲去模糊; 多尺度模型; 通道注意力机制; 空间金字塔池化

中图法分类号: TP391

中文引用格式: 唐述, 万盛道, 谢显中, 杨书丽, 黄容, 顾佳, 郑万鹏. 一种多尺度的图像动态场景盲去模糊网络. 软件学报, 2022, 33(9): 3498–3511. <http://www.jos.org.cn/1000-9825/6297.htm>

英文引用格式: Tang S, Wan SD, Xie XZ, Yang SL, Huang R, Gu J, Zheng WP. Multi-scale Image Blind Deblurring Network for Dynamic Scenes. Ruan Jian Xue Bao/Journal of Software, 2022, 33(9): 3498–3511 (in Chinese). <http://www.jos.org.cn/1000-9825/6297.htm>

Multi-scale Image Blind Deblurring Network for Dynamic Scenes

TANG Shu, WAN Sheng-Dao, XIE Xian-Zhong, YANG Shu-Li, HUANG Rong, GU Jia, ZHENG Wan-Peng

(Chongqing Key Laboratory of Computer Network and Communications Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: Recently, the convolutional neural network (CNN) based single-image dynamic scene blind deblurring (SIDSBD) methods have made significant progress. Their success mainly stems from the multi-scale/multi-patch model and the design of the encoder-decoder architecture and the residual block structure. In this paper, a novel multi-scale CNN (MSCNN) is proposed to further exploit the advantages of the multi-scale model, the encoder-decoder architecture, and the residual block structure, which can achieve higher-quality SIDSBD. First, inspired by the spatial pyramid pooling (SPP) and the multi-patch model, this study put forward a hierarchical multi-patch

* 基金项目: 国家自然科学基金 (61601070, 61501074); 重庆市教委科学技术研究重点项目 (KJZD-K201800603); 重庆市教委科学技术研究重大项目 (KJZD-M201900602); 重庆市基础研究与前沿探索项目 (cstc2018jcyjAX0432); 重庆市技术创新与应用发展专项面上项目 (cstc2020jcsx-msxmX0135)

收稿时间: 2020-08-05; 修改时间: 2020-09-19, 2020-10-20; 采用时间: 2020-12-30; jos 在线出版时间: 2022-07-15

channel attention (HMPCA) strategy to perform adaptive weight assignment for feature images channel-wise by using the global and local feature statistics. The proposed HMPCA uses local information, which can be considered to enlarge the receptive field in the channel direction and thus can enhance the representational ability of the network. Then, different from existing multi-scale models, a novel multi-scale model is built, in which each scale consists of multiple encoders and decoders. Because of the HMPCA, the encoders and decoders at the same scale are not exactly the same. The proposed multi-scale model can be regarded to increase the depth of the encoder-decoder architecture, thus able to improve the deblurring performance of each scale and finally achieve higher-quality blind deblurring for dynamic scenes. Extensive experiments comparing the proposed SIDSBD method with state-of-the-art ones demonstrate the superiority of the method in terms of both qualitative evaluation and quantitative metrics.

Key words: convolutional neural network (CNN); blind deblurring for dynamic scene; multi-scale model; channel attention; spatial pyramid pooling (SPP)

1 引言

由于成像系统自身的物理局限性和复杂多变的拍摄环境,不可避免的会导致拍摄到的图像出现模糊.在众多的不同类型的模糊图像中,自然动态场景的模糊图像因为其最接近人们真实的日常生活,因此其对应的盲去模糊问题成为了人们研究的热点.迄今为止,单幅动态场景模糊图像的盲去模糊方法可以分为:基于建模最优化的方法和基于学习的方法两大类.2010年,Harmeling等人提出了一种有效的滤波器流方法来模拟由于相机小尺度的面内平移、旋转和倾斜而造成的非均匀模糊^[1].同年,Gupta等人提出了一种基于投影运动路径模糊模型的运动密度函数来估计由于相机面内平移和旋转所造成的空间变化模糊^[2].2011年,Hirsch等人将投影运动路径模糊模型和有效的滤波器流框架结合到一起,提出了一种去除相机抖动的非均匀盲去模糊方法^[3].2012年,Xu等人将场景的深度考虑到动态场景的盲去模糊中,提出了一种深度感知的非均匀盲去模糊方法^[4].2013年,Xu等人提出了一种类L0范数的图像先验稀疏约束来实现单幅图像的盲去模糊^[5].2014年,Hu等人同样考虑了场景的深度,提出了一种联合估计场景深度和相机抖动的动态场景非均匀盲去模糊方法^[6].然而,因为在真实的日常生活中,造成图像模糊的原因非常复杂,目前还无法用数学模型来准确模拟其模糊的过程,因此导致现有的基于建模最优化的方法仅能去除一些简单的静态场景的相机抖动模糊(场景内物体静止,模糊仅由相机的抖动造成),无法适用于复杂的动态场景盲去模糊问题:场景内多个刚性物体的运动、多个非刚性物体的运动和相机抖动同时存在的情况,同时还考虑了场景的深度.

近几年,随着深度学习、神经网络的迅猛发展,基于端到端学习的方法已经被成功的运用到动态场景的盲去模糊中,并已取得了前所未有的成功,尤其是基于卷积神经网络(convolutional neural network, CNN)的动态场景盲去模糊方法最具代表性,复原效果也最为显著.2017年,Xu等人提出了一种生成对抗网络,其能够直接从一幅模糊的低分辨率图像中复原出其对应的清晰高分辨率图像^[7].同年,Nah等人提出了一种由粗到精的多尺度CNN来实现端到端的动态场景盲去模糊.提出的网络由3个尺度组成,每个尺度都包含了19个残差块^[8].为了获得较好的动态场景盲去模糊结果,Nah等人的残差块去除了传统残差块中的批量正则化单元和线性整流单元.2018年,Kupyn等人针对动态场景模糊图像的盲复原问题,提出了一种端到端的去模糊生成对抗网络.该网络的生成器是由9个残差块和一个全局的跳连接组成,同时,Kupyn等人还利用了一种基于VGG19(visual geometry group, VGG)的内容损失来提升图像复原的效果^[9].2018年,Tao等人将Nah等人的残差块^[8]引入到编解码器中,提出了一种多尺度循环网络(scale-recurrent network, SRN)^[10].Tao等人将残差块和编解码器结合到一起,进一步提升了SIDSBD的性能.2018年,Zhang等人认为大多数的基于深度学习的方法并没有完全利用模糊图像中的分级特征,因此,Zhang等人提出了一种残差密集网络和一种局部融合模块来实现更有效的特征提取^[11].同年,Zhang等人根据动态场景模糊图像中的模糊是空间变化的这一特性,提出了一种空间变化的循环神经网络(recurrent neural network, RNN).Zhang等人的网络针对图像的不同区域学习不同的权重,以达到复原动态场景模糊图像的目的^[12].在Zhang等人的RNN中,采用了与SRN^[10]相同的残差块和编解码器结构.2019年,Gao等人基于SRN^[10]的基础上分析了模糊图像在不同尺度上的特性,同时受到传统建模方法求解过程中对参数设置的启发,提出了一种参数选择性共享策略和一种嵌套的跳连接结构来实现动态场景模糊图像的清晰化盲复原^[13].Gao等人的方法将基

于 CNN 的动态场景盲去模糊方法的性能向前提升了一大步. 2019 年, Zhang 等人^[14]提出了一种深度的多块分级网络 (deep multi-patch hierarchical network, DMPHN) 来实现端到端的动态场景盲去模糊. DMPHN 由多个水平组成, 每个水平由多个残差块和一对编解码器组成, 与多尺度模型^[10,12,13]不同, DMPHN 的每个水平的输入图像都具有相同的空间分辨率, 因此 DMPHN 不仅能够在同一水平内通过残差块学习到有效的图像特征, 还能够在相邻水平间也实现类似残差的操作. DMPHN 同样使用了与 SRN^[10]相同的残差块. 2020 年, Zhang 等人提出了一种双重生对抗网络模型来实现动态场景的盲去模糊^[15]. 双重生对抗网络模型由一个学习模糊生成对抗网络和一个学习去模糊生成对抗网络组成, 其中, 学习模糊生成对抗网络用来学习清晰图像模糊的过程, 其结果则用于指导学习去模糊生成对抗网络的复原.

通过以上的分析可知, 基于端到端的 CNN 动态场景盲去模糊方法因为其强大的学习能力, 能够直接学习到模糊图像与对应清晰图像之间的映射关系, 从而完全避免了对动态场景图像模糊过程的数学建模问题以及对应的最优化求解问题, 因此与基于建模最优化的方法相比, 基于端到端的 CNN 能够更好地适用于复杂的动态场景盲去模糊. 同时, 我们进一步对现有的一些较为成功的基于 CNN 的动态场景盲去模糊方法进行深入的研究和分析发现, 这类方法的成功主要是源于: 多尺度模型^[8,10,12,13]或者多块模型^[14], 编解码器架构的设计^[10,14]和残差块结构的设计^[8,10,14]3 个方面. 因此, 本文提出了一种新的多尺度卷积神经网络 (multi-scale convolutional neural network, MSCNN) 来进一步开发多尺度模型、编解码器架构和残差块结构的优势, 以实现更高质量的动态场景盲去模糊. 首先, 受到空间金字塔池化 (spatial pyramid pooling, SPP)^[16]和 DMPHN^[14]的启发, 本文提出了一种分等级的多块通道注意力机制 (hierarchical multi-patch channel attention, HMPCA). 提出的 HMPCA 分别利用了特征图的全局特征统计量和局部特征统计量来自适应地对同一尺度内的特征图分别进行逐通道的权重赋值. 对于每个尺度内的多个编解码器而言, HMPCA 将会在更靠前的编码器和解码器上考虑更多的局部信息, 因此可以被认为是增加了通道方向的感受野, 也因此能够进一步提升网络的表达能力. 其次, 不同于现有的多尺度模型^[8,10,12,13], 本文提出的 MSCNN 的每个尺度由多个编码器和多个解码器构成 (如图 1 所示, 每个尺度由 3 个编码器和 3 个解码器构成), 并且因为 HMPCA, 使得同一尺度内的所有编码器和解码器并不完全相同, 这就意味着一幅模糊图像在同一个尺度内将会被不同的编码器和解码器连续编解码多次. 因此, 本文提出的 MSCNN 可以被看作是增加了编解码器的深度, 因此能够提升每一个尺度的去模糊性能, 最终实现更高质量的动态场景盲去模糊. 大量的实验结果表明: 本文提出的方法较近几年的一些成功的动态场景盲去模糊方法相比, 能够复原出更高质量的去模糊图像, 在客观的评价指标和主观的视觉效果上均有显著的改进. 接下来我们就将对本文提出的 MSCNN 进行详细的论述.

2 提出的多尺度卷积神经网络

本文提出的 MSCNN 的网络结构如图 1 所示: 网络由 3 个尺度组成, 每一个尺度均由 3 个编码器和 3 个解码器组成. 其中, $\mathbf{B}_i (i \in \{1, 2, 3\})$ 分别表示输入的 3 个尺度的模糊图像, $\mathbf{S}_i (i \in \{1, 2, 3\})$ 则分别表示 3 个尺度的输出结果. \mathbf{B}_1 是原始分辨率的模糊图像, 而 \mathbf{B}_2 和 \mathbf{B}_3 分别是 \mathbf{B}_1 下采样 2 倍和下采样 4 倍的结果. \mathbf{S}_{3up} 和 \mathbf{S}_{2up} 分别是 \mathbf{S}_3 和 \mathbf{S}_2 上采样 2 倍的结果, \mathbf{S}_1 是我们最终的原始分辨率的复原结果图像. 我们定义在第 i 个尺度的第 j 个编码器的输入和输出分别为 $\mathbf{EN_I}_{i,j}$ 、 $\mathbf{EN_O}_{i,j}$, 第 i 个尺度的第 j 个解码器的输入和输出分别为 $\mathbf{DE_I}_{i,j}$ 、 $\mathbf{DE_O}_{i,j}$. 根据上面的定义, 我们可以得到每个尺度的输入, 也即是每个尺度中第 1 个编码器的输入 $\mathbf{EN_I}_{i,1}$ 为:

$$\mathbf{EN_I}_{i,1} = \mathbf{B}_i \oplus \mathbf{S}_{(i+1)up} \quad (1)$$

其中, \oplus 表示通道方向的拼接操作, 在第 3 个尺度, $\mathbf{S}_{4up} = \mathbf{B}_3$. 对于每个尺度内第 2 个和第 3 个编码器的输入为:

$$\mathbf{EN_I}_{i,j} = \mathbf{DE_O}_{i,j-1} + \mathbf{B}_i \quad (2)$$

其中, $i \in \{1, 2, 3\}$, $j \in \{2, 3\}$. 然后, 每个尺度内解码器的输入分别为:

$$\mathbf{DE_I}_{i,1} = \mathbf{EN_O}_{i,1} \quad (3)$$

$$\mathbf{DE_I}_{i,2} = \mathbf{EN_O}_{i,2} + \mathbf{EN_O}_{i,1} \quad (4)$$

$$\mathbf{DE_I}_{i,3} = \mathbf{EN_O}_{i,3} + \mathbf{EN_O}_{i,2} + \mathbf{EN_O}_{i,1} \quad (5)$$

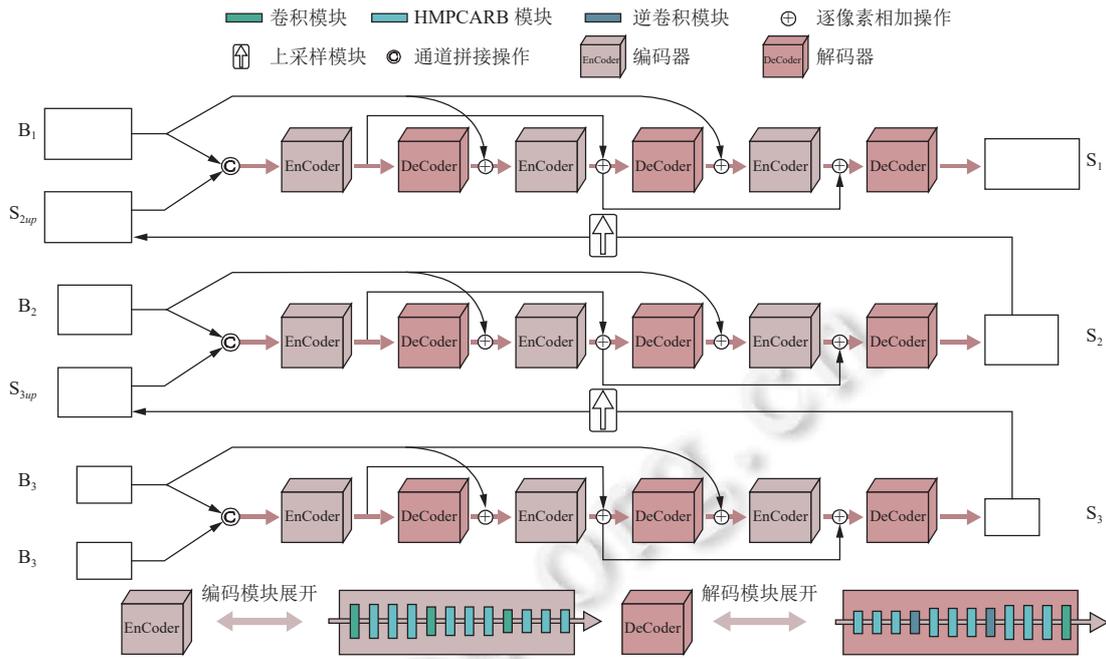


图1 提出的多尺度卷积神经网络 (MSCNN)

2.1 提出的 HMPCA 机制

2018年, Zhang 等人提出了一种通道注意力机制 (channel attention, CA) 来开发特征图之间的通道相关性, 增强了网络的表达能力, 提升了图像超分辨率重建的准确性^[17]. 因此, 本文在 SPP 和 DMPHN 的启发下, 结合 CA 的思想, 提出了一种分等级的多块通道注意力机制 (hierarchical multi-patch channel attention, HMPCA). 如图2-图4所示, 本文提出的 HMPCA 总共由3个模型组成: 提取全局特征统计量的全局注意力模型 HMPCA-global-model (图2)、提取局部特征统计量的二分块注意力模型 HMPCA-two-patches-model (图3) 和四分块注意力模型 HMPCA-four-patches-model (图4). 我们用 $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_c]$ 来表示 HMPCA 的输入, \mathbf{F} 是由 C 个特征图组成, 每个特征图的空间分辨率大小均为 $H \times W$. 如图2所示, HMPCA-global-model 首先将每个特征图进行全局平均池化操作:

$$z_g^c = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W f_c(h, w) \tag{6}$$

其中, $f_c(h, w)$ 表示第 c 个特征图 \mathbf{f}_c 的第 h 行和第 w 列所对应的元素的值, 显而易见, z_g^c 也就为 \mathbf{f}_c 上所有元素值的平均值, 即全局平均值. 因此, 对所有 C 个特征图分别进行全局平均池化操作后即可得到一个维度为 $1 \times 1 \times C$ 的向量 $\mathbf{Z}_g = (z_g^1, z_g^2, \dots, z_g^C)$.

接下来, 为了开发特征图之间全局信息的通道相关性, 我们采用了3个全连接层和两种非线性函数来对 \mathbf{Z}_g 进行处理:

$$\mathbf{S}_g = \sigma(\mathbf{W}_g^3 \delta(\mathbf{W}_g^2 \delta(\mathbf{W}_g^1 \mathbf{Z}_g + \mathbf{b}_g^1) + \mathbf{b}_g^2) + \mathbf{b}_g^3) \tag{7}$$

其中, \mathbf{S}_g 表示特征图的全局通道注意力权重, 即全局特征统计量. $\mathbf{W}_g^1 \in \mathbb{R}^{C \times C \times 1 \times 1}$ 和 $\mathbf{b}_g^1 \in \mathbb{R}^C$ 分别表示第1个全连接层的权重和偏移量. $\mathbf{W}_g^2 \in \mathbb{R}^{C \times rC \times 1 \times 1}$ 和 $\mathbf{b}_g^2 \in \mathbb{R}^C$ 分别表示第2个全连接层的权重和偏移量, 它们将输入的特征图的通道方向的维度压缩为 rC . $\mathbf{W}_g^3 \in \mathbb{R}^{rC \times C \times 1 \times 1}$ 和 $\mathbf{b}_g^3 \in \mathbb{R}^C$ 分别表示第3个全连接层的权重和偏移量, 它们又将输入的特征图的通道方向的维度扩展为 C . $\delta(\cdot)$ 表示 ReLU 激活函数, 而 $\sigma(\cdot)$ 表示 Sigmoid 激活函数. 最后, 我们利用计算得到的 \mathbf{S}_g 来对输入的特征图进行逐通道的权重赋值:

$$\mathbf{f}'_c = \mathbf{f}_c \times s_{g_c} \tag{8}$$

其中, s_{g_c} 为 \mathbf{S}_g 中的第 c 个元素, 表示第 c 个特征图的全局特征统计量.

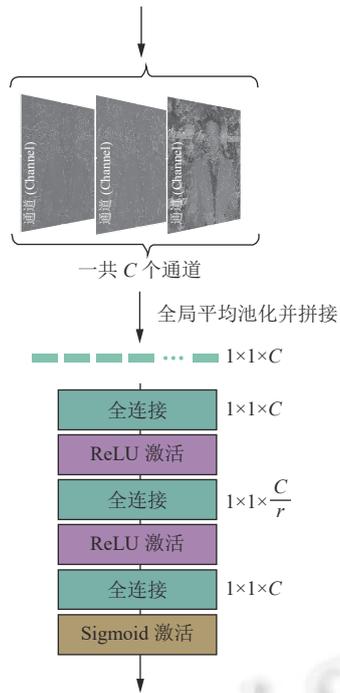


图2 全局注意力模型

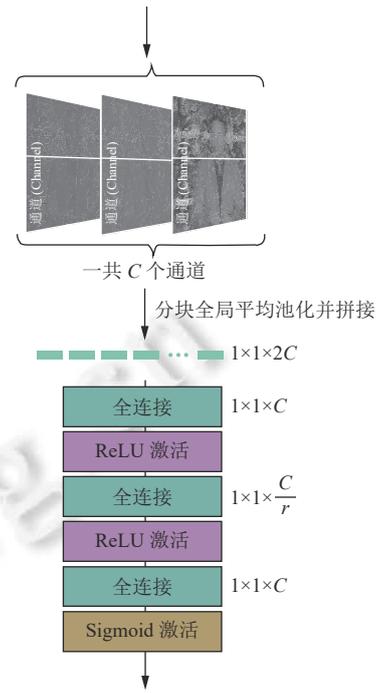


图3 二分块注意力模型

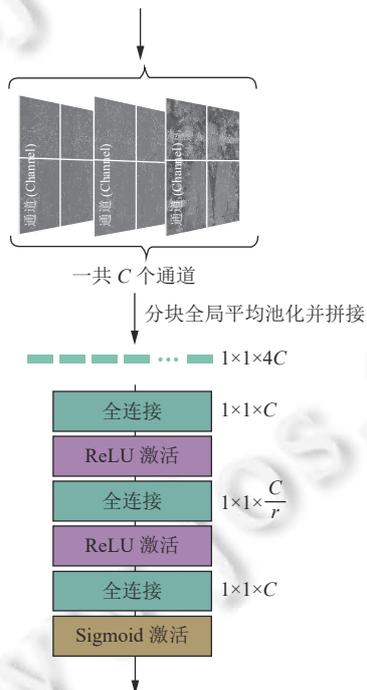


图4 四分块注意力模型

为了能够有效提取特征图的局部特征统计量,我们利用了 SPP 的分块全局平均池化.以图 3 中的 HMPCA-two-patches-model 为例,我们将每个特征图按行平分分为两个图像块,然后分别对它们进行平均池化操作:

$$z_{2l_1}^c = \frac{1}{\lfloor \frac{H}{2} \rfloor \times W} \sum_{h=1}^{\lfloor \frac{H}{2} \rfloor} \sum_{w=1}^W f_c(h, w) \quad (9)$$

$$z_{2l_2}^c = \frac{1}{\left(H - \lfloor \frac{H}{2} \rfloor\right) \times W} \sum_{h=\lfloor \frac{H}{2} \rfloor+1}^H \sum_{w=1}^W f_c(h, w) \quad (10)$$

其中, $\lfloor \cdot \rfloor$ 表示向下取整. 如公式 (9) 和公式 (10) 所示, 对每幅特征图分别进行分块和平均池化操作, 我们将得到两个维度为 $1 \times 1 \times C$ 的向量, 并将这两个向量拼接成一个维度为 $1 \times 1 \times 2C$ 的向量 $\mathbf{Z}_{2l} = (z_{2l_1}^1, z_{2l_1}^2, \dots, z_{2l_1}^C, z_{2l_2}^1, z_{2l_2}^2, \dots, z_{2l_2}^C)$ (如图 3 所示). 同理, 对于 HMPCA-four-patches-model 而言, 将每个特征图按行和列平分为 4 个图像块, 再对每个图像块进行平均池化操作, 那么我们可以得到一个维度为 $1 \times 1 \times 4C$ 的向量 $\mathbf{Z}_{4l} = (z_{4l_1}^1, z_{4l_1}^2, \dots, z_{4l_1}^C, z_{4l_2}^1, z_{4l_2}^2, \dots, z_{4l_2}^C, \dots, z_{4l_4}^1, \dots, z_{4l_4}^C)$ (如图 4 所示). 采用与公式 (7) 相似的方法, 我们再次利用了 3 个全连接层, 以及 ReLU 和 Sigmoid 两种激活函数, 并结合 \mathbf{Z}_{kl} ($k \in \{2, 4\}$) 来开发特征图之间局部信息的通道相关性:

$$\mathbf{S}_{kl} = \sigma(\mathbf{W}_{kl}^3 \delta(\mathbf{W}_{kl}^2 \delta(\mathbf{W}_{kl}^1 \mathbf{Z}_{kl} + \mathbf{b}_{kl}^1) + \mathbf{b}_{kl}^2) + \mathbf{b}_{kl}^3) \quad (11)$$

其中, $\mathbf{W}_{kl}^1 \in \mathbb{R}^{kC \times C \times 1 \times 1}$ 和 $\mathbf{b}_{kl}^1 \in \mathbb{R}^C$ 分别表示第 1 个全连接层的权重和偏移量. $\mathbf{W}_{kl}^2 \in \mathbb{R}^{C \times rC \times 1 \times 1}$ 和 $\mathbf{b}_{kl}^2 \in \mathbb{R}^{rC}$ 分别表示第 2 个全连接层的权重和偏移量. $\mathbf{W}_{kl}^3 \in \mathbb{R}^{rC \times C \times 1 \times 1}$ 和 $\mathbf{b}_{kl}^3 \in \mathbb{R}^C$ 分别表示第 3 个全连接层的权重和偏移量. 在本文中, 取 $r = \frac{1}{16}$. 由公式 (11) 可知, 经过最后一层全连接的输出, 得到的 $\mathbf{S}_{kl} = (s_{kl}^1, s_{kl}^2, \dots, s_{kl}^C)$ 是一个维度为 $1 \times 1 \times C$ 的向量, 其中的每个元素为对应的特征图的局部注意力权重, 即局部特征统计量:

$$\mathbf{f}'_c = \mathbf{f}_c \times s_{kl}^c \quad (12)$$

其中, s_{kl}^c 为 \mathbf{S}_{kl} 中的第 c 个元素, 表示第 c 个特征图的局部特征统计量. 在得到了特征图的全局特征统计量和局部特征统计量之后, 我们对每个尺度上越靠前的编码器和解码器考虑更多的局部信息, 因此, 我们给每个尺度的第 1 个编码器和第 1 个解码器中的残差块分配 HMPCA-four-patches-model, 每个尺度的第 2 个编码器和第 2 个解码器中的残差块分配 HMPCA-two-patches-model, 每个尺度的最后一个编码器和最后一个解码器中的残差块分配 HMPCA-global-model. 由此可见, 本文提出的 HMPCA 结合了 SPP、DMPHN 和 CA 的思想, 在每个尺度上都同时考虑了特征图的全局特征统计量和局部特征统计量, 因此可以被认为是增加了特征图通道方向的感受野, 也因此能够进一步增强网络的盲去模糊表达能力. 接下来, 我们就将提出的 HMPCA 引入到编解码器中的残差块中, 创建出一种分等级的多块通道注意力残差块 (hierarchical multi-patch channel attention residual block, HMPCARB).

2.2 HMPCARB

残差块第一次是由 He 等人^[18]提出, 残差块用公式表示如下:

$$\mathbf{Y} = \mathbf{X} + F(\mathbf{X}) \quad (13)$$

其中, $F(\cdot)$ 表示卷积+ReLU 激活+卷积的函数操作, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_C]$ 表示输入的 C 幅特征图, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_C]$ 表示输出的 C 幅特征图. 在本文中, 我们将 HMPCA 机制引入到传统残差块里面, 创建出 HMPCARB, 其结构如图 5 所示. 因为我们的 3 种 HMPCA 机制: 图 2 所示的全局注意力模型, 图 3 所示的二分块注意力模型和图 4 所示的四分块注意力模型, 都是以相同的方式引入到传统的残差块里面. 因此接下来我们只详细介绍如何将 HMPCA-global-model 引入到传统的残差块里面, 而对于 HMPCA-two-patches-model 和 HMPCA-four-patches-model, 因为其引入方式类似, 故不再赘述.

首先将输入 \mathbf{X} 经过两层卷积和 ReLU 激活得到输出特征图 $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_C]$:

$$\mathbf{F} = F(\mathbf{X}) \quad (14)$$

然后使用 HMPCA-global-model 计算 \mathbf{F} 里面每张特征图的全局特征统计量, 如公式 (6) 和公式 (7) 所示, 接着再使用得到的全局特征统计量来逐通道的对特征图 \mathbf{F} 进行权重赋值:

$$\mathbf{F}' = \mathbf{F} \odot \mathbf{S}_g \quad (15)$$

那么, 图 5 所示的 HMPCARB 的输出为:

$$\mathbf{Y}' = \mathbf{X} + \mathbf{F}' \tag{16}$$

其中, \odot 表示点乘 (即 \mathbf{S}_g 中每个元素与 \mathbf{F} 中的每幅特征图对应相乘), $\mathbf{F}' = [\mathbf{f}'_1, \mathbf{f}'_2, \dots, \mathbf{f}'_C]$ 表示被 HMPCA 逐通道权重赋值后的特征图. 因此, 经过本文提出的 HMPCA 机制, 对于每组特征图我们都可以得到一个维度为 $1 \times 1 \times C$ 的向量, 其中的每个元素都与其相应的特征地图一一对应, 再经过公式 (15) 和公式 (16) 便可得到我们的 HMPCARB 模块 (如图 5 所示).

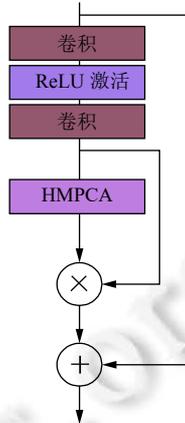


图 5 分等级的多块通道注意力残差块

2.3 MSCNN 的网络架构

如图 1 所示, 本文提出的 MSCNN 共由 3 个尺度组成, 每个尺度层分别包含有 3 个编码器和 3 个解码器. 由前面的分析可知, 因为 HMPCA, 同一尺度内的所有编码器和解码器并不完全相同 (会对同一尺度内不同编码器和不同解码器的残差块引入不同的多块通道注意力模型, 如图 2-图 5 所示), 这就意味着一幅模糊图像在同一个尺度内将会被不同的编码器和解码器连续编解码多次. 在每一个尺度内, 每个编码器的第一个卷积模块、每个解码器的最后一个卷积模块和 HMPCARB 中的卷积模块的卷积步长为 1, 其余卷积模块的卷积步长均为 2. 这些卷积步长为 2 的卷积模块会将先前层输出的特征图的通道数量加倍, 但同时也会将特征图的空间分辨率减半. 因为编码器和解码器是对称的, 所以, 解码器中的逆卷积模块则用来将先前层输出的特征图的通道数量减半, 同时将其空间分辨率加倍. 在损失函数的计算方面, 本文采用了多尺度模型所通用的是多尺度平方误差损失函数 (multi-scale mean square error, MSMSE):

$$Loss = \sum_{i=1}^3 \frac{1}{N_i} \|\mathbf{S}_i - \mathbf{GT}_i\|_2^2 \tag{17}$$

其中, $\mathbf{S}_i (i \in \{1, 2, 3\})$ 表示第 i 个尺度的输出结果, $\mathbf{GT}_i (i \in \{1, 2, 3\})$ 表示第 i 个尺度对应的真实清晰图像, \mathbf{GT}_2 和 \mathbf{GT}_3 分别是由 \mathbf{GT}_1 下采样 2 倍和 4 倍得到. 在模型的参数共享方面, 不同尺度相同序号的编码器参数共享, 不同尺度相同序号的解码器参数共享, 而同一尺度内的编解码器参数不共享, 即: $EnCoder_{i_1, j}$ 与 $EnCoder_{i_2, j}$ 的参数共享, 以及 $DeCoder_{i_1, j}$ 与 $DeCoder_{i_2, j}$ 的参数共享, 其中 $j \in \{1, 2, 3\}$, $i_1 \in \{1, 2, 3\}$, $i_2 \in \{1, 2, 3\}$ 且 $i_1 \neq i_2$. 而 $EnCoder_{i, j_1}$ 与 $EnCoder_{i, j_2}$ 的参数不共享, 以及 $DeCoder_{i, j_1}$ 与 $DeCoder_{i, j_2}$ 的参数不共享, 其中 $i \in \{1, 2, 3\}$, $j_1 \in \{1, 2, 3\}$, $j_2 \in \{1, 2, 3\}$ 且 $j_1 \neq j_2$.

2.4 本文提出的 MSCNN 与 SRN 和 DMPHN 的不同之处

Tao 等人提出的 SRN^[10]和 Zhang 等人提出的 DMPHN^[14]是与本文提出的 MSCNN 最接近的两种方法. 虽然相近, 但实质上还是有许多的不同之处. 首先, 对于 SRN 而言, 虽然本文提出的 MSCNN 与 Tao 等人的 SRN 都是工作在“由粗到精”的 3 个尺度的基础上的, 即: 首先将模糊图像分别进行 2 倍和 4 倍的下采样, 形成具有 3 个尺度层的图像金字塔; 然后从最低尺度层 (最粗的分辨率层) 开始复原, 并将复原结果上采样 2 倍之后作为一个尺度层 (较精细的分辨率层) 的初始值来指导较精细层的复原; 如此反复直到复原出最高尺度层 (原始分辨率层) 的图像为止. 但是本文提出的 MSCNN 与 SRN 还是存在有两个主要的不同之处: (1) SRN 的每个尺度

仅由一个编码器和一个解码器构成,而本文提出的 MSCNN 的每个尺度是由多个编解码器组成,且同一尺度内的所有编码器和解码器并不完全相同(因为不同编解码器使用的 HMPCA 模块不相同),因此本文提出的 MSCNN 可以被看作是增加了编解码器的深度,这与简单的通过增加卷积模块的数量来增加网络的深度是完全不同的,因为如果只是简单地增加一个编码器和一个解码器内卷积模块的数量,那么图像也只会编解码一次,而不会被编解码多次;(2)本文提出了一种 HMPCA,改进了 SRN 中的残差块结构,进一步增强了网络的盲去模糊的表达能。

其次,对于 Zhang 等人提出的 DMPHN^[14]而言,其与本文提出的 MSCNN 同样存在有两个主要的区别:首先, DMPHN 并没有采用多尺度的模型,而是采用的一种“由精到粗”的分块模型,且与 SRN 相似,在 DMPHN 的每个层级内仅包含有一个编码器和一个解码器;其次, DMPHN 采用了与 SRN 相同的残差块,同样没有引入任何的注意力机制.图 6 所示是本文提出的 MSCNN 与 SRN^[10]和 DMPHN^[14]在网络结构上的比较.由图 6 可知,本文提出的 MSCNN 与 SRN 和 DMPHN 在网络的结构上还是存在有较大差别的。

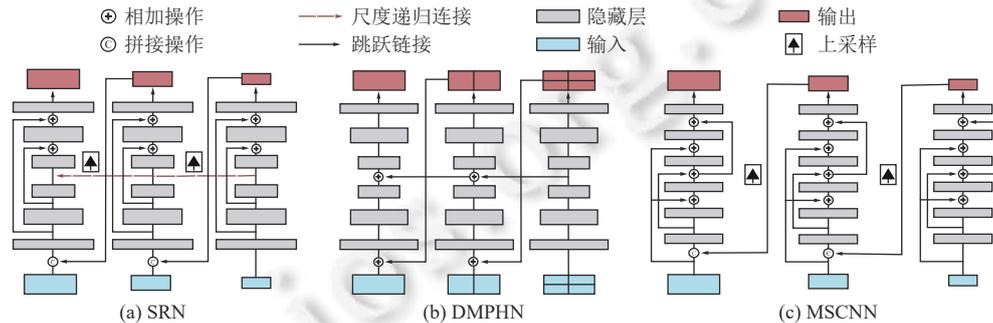


图 6 网络结构对比

3 实验及分析

本节我们会先介绍本文提出方法的实现细节,包括使用的训练数据、模型超参数等,然后对本文提出方法的可行性和性能进行验证。

3.1 数据集以及超参数的设置

(1) 数据集. 本文所使用的数据集为 Nah 等人提供的 GoPro 动态场景模糊图像数据集^[8],这个数据集被作者划分为训练集和测试集两部分,其中训练集一共含有 2 103 张动态场景模糊图像和其对应的真实清晰图像,测试集被作者划分为 11 个子测试集,共含有 1 111 张动态场景的模糊图像和与之对应的清晰图像.为了保证实验的公平性,在本文的实验中,所有基于学习的 SIDSBD 方法均是用 GoPro 的 2 103 张训练集进行训练,然后用 1 111 张测试集进行测试。

(2) 模型训练. 我们的网络是在一台具有 NVIDIA 2080TI 和 Intel Core I7-7820X 的台式电脑上训练的,深度学习框架使用的是 PyTorch. 在训练我们模型的时候,我们设置网络输入的 batch size 为 4,优化器使用的 Adam,它的两个超参 β_1 和 β_2 分别是 0.9 和 0.999. 初始学习率是 $1E-4$,并且一共训练了 2 400 个 epoch. 图像在输入神经网络之前会先进行预处理:将其像素值归一化到 $[0, 1]$ 之间,然后减去 0.5 以保证所有输入网络的图像的像素值范围在 $[-0.5, 0.5]$ 之间. 如果没有特别说明的话,默认情况下 MSCNN 中所有的卷积核大小均为 3×3 . 为了验证本文提出的 MSCNN 的性能,我们采用了图像去模糊领域通用的峰值信噪比 (peak signal-to-noise ratio, PSNR) 和结构相似度 (structural similarity index metric, SSIM) 两种客观评价指标来定量的评价本文提出方法的性能。

3.2 消融实验

如前所述,本文提出的 MSCNN 的创新点主要在两个方面:(1)提出了一种分等级的多块通道注意力机制 (HMPCA) 来进一步提升网络的表达能力;(2)不同于现有的多尺度模型,本文提出的 MSCNN 的每个尺度由多个不同的编码器和多个不同的解码器构成,因此一幅模糊图像在同一个尺度内将会被不同的编码器和解码器连续编

解码多次,进而能够提升每一个尺度的去模糊性能,最终实现更高质量的动态场景盲去模糊.因此,接下来,在本小节中,我们将通过消融实验来验证本文提出的创新点的有效性.首先,为了验证本文提出的同一尺度内多编码器和多解码器的模型的有效性,我们分别将每个尺度内的编解码数量降为 1 个编码器+1 个解码器(命名为 MSCNN-1-Codec 模型)和 2 个编码器+2 个解码器(命名为 MSCNN-2-Codecs 模型),以及分别将每个尺度内的编解码数量增加为 4 个编码器+4 个解码器(命名为 MSCNN-4-Codecs 模型)和 5 个编码器+5 个解码器(命名为 MSCNN-5-Codecs 模型)这 4 种情况,其余的网络结构和参数不变. MSCNN-1-Codec、MSCNN-2-Codecs、MSCNN-4-Codecs、MSCNN-5-Codecs 和本文提出的 MSCNN,5 种模型在 1 111 张 Gopro 测试集上的平均 PSNR 值和平均 SSIM 值如表 1 所示.如表 1 所示,本文提出的 MSCNN 较 MSCNN-1-Codec 和 MSCNN-2-Codecs 相比,平均的 PSNR 和 SSIM 均提升较大,PSNR 分别提升了 1.25 dB 和 0.49 dB,平均的 SSIM 也分别提升了 0.0217 和 0.0058.而同时,因为同一尺度内编解码器数量的增加,MSCNN 的模型大小也比 MSCNN-1-Codec 和 MSCNN-2-Codecs 分别增加了 4.8 M 和 3 M,所需的平均运行时间也分别增加了 0.24 s 和 0.16 s.另一方面,虽然我们将每个尺度内的编解码器分别增加到 4 个(MSCNN-4-Codecs)和 5 个(MSCNN-5-Codecs),但是 MSCNN-4-Codecs 和 MSCNN-5-Codecs 的平均 PSNR 仅比 MSCNN 高出了 0.09 dB 和 0.19 dB,平均 SSIM 也仅提升了 0.000 8 和 0.001 7.然而, MSCNN-4-Codecs 和 MSCNN-5-Codecs 的模型大小和平均运行时间却远高于 MSCNN,分别增加了 7.4 M 和 13.5 M,以及分别增加了 0.39 s 和 0.71 s,远高于 MSCNN 比 MSCNN-1-Codec 和 MSCNN-2-Codecs 的增量,如表 1 所示.因此,综合考虑网络的去模糊性能和运行效率,本文决定采用图 1 所示的 MSCNN:每个尺度内 3 个编码器+3 个解码器的网络设置.

表 1 消融实验结果

模型	PSNR (dB)	SSIM	模型大小 (M)	耗时 (s)
MSCNN-1-Codec	29.90	0.9227	3.9	0.24
MSCNN-2-Codecs	30.66	0.9386	5.7	0.32
MSCNN (本文)	31.15	0.9444	8.7	0.48
MSCNN-4-Codecs	31.24	0.9452	16.1	0.87
MSCNN-5-Codecs	31.34	0.9461	22.2	1.19
MSCNN-no-HMPCA	30.99	0.9428	7.8	0.33

接下来,为了验证本文提出的 HMPCA 的有效性,我们将 MSCNN 中的 HMPCA 机制去掉,采用 SRN^[10]的残差块来代替图 1 中的 HMPCARB,该模型被命名为 MSCNN-no-HMPCA.表 1 同样给出了 MSCNN-no-HMPCA 的平均 PSNR 值、平均 SSIM 值、模型大小和平均运行时间.如表 1 所示,本文提出的 MSCNN 较 MSCNN-no-HMPCA 相比,在模型大小增加不太大,运行时间增加不太多的情况下,平均的 PSNR 和平均 SSIM 却提升较大,分别提升了 0.16 dB 和 0.001 6.表 1 很好地证明了本文提出的 HMPCA 的有效性.

3.3 与前沿方法的比较实验

为了验证本文提出方法的优越性,在本小节,分别在 GoPro 数据集、视频去模糊数据集和真实的模糊图像上将本文提出的 MSCNN 与近几年提出的一些较为成功的单幅图像盲去模糊方法进行了比较: Xu 等人的方法^[5]、Nah 等人的方法^[8]、Kupyn 等人的方法^[9]、Zhang 等人的方法^[12]、Tao 等人的方法^[10]、Zhang 等人的方法^[14]和 Gao 等人的方法^[13].

(1) GoPro 数据集上的比较.表 2 所示是本文提出的 MSCNN 与 Xu 等人的方法^[5]、Nah 等人的方法^[8]、Kupyn 等人的方法^[9]、Zhang 等人的方法^[12]、Tao 等人的方法^[10]、Zhang 等人的方法^[14]和 Gao 等人的方法^[13]的客观评价指标比较.表 2 中的第 1 列是 GoPro 的 11 个子测试集和所有的 1 111 张测试图像,表 2 中第 2 列的 PSNR 值和 SSIM 值分别是在 GoPro 的 11 个子测试集上每个子测试集的平均 PSNR 值和平均 SSIM 值,和在 GoPro 的所有 1 111 张测试图像上的平均 PSNR 值和平均 SSIM 值,表 2 中第 3 列、第 4 列等的 PSNR 值和 SSIM 值同理.需要注意, Xu 等人的方法^[5]是基于建模最优化的方法,因此其不需要进行训练.对于 Zhang 等人的方法^[14],我们选用了他们的基本的 4 水平 DMPHN 模型(1-2-4-8,具体内容可参见文献[14]).如表 2 所示,仅有

Gao 等人的方法^[13]在其中的 3 个子测试集上的平均 PSNR 值和平均 SSIM 值略高于本文提出的 MSCNN, 除此之外, 在其余的 8 个子测试集上和所有的 1 111 张测试图像上, 本文提出的 MSCNN 均能获得最高的平均 PSNR 值和平均 SSIM 值. 表 2 很好地证明了在客观的评价指标上, 本文提出的 MSCNN 明显优于近几年提出的较为成功的 7 种单幅图像盲去模糊方法.

表 2 Gopro 测试集上的平均 PSNR 值和平均 SSIM 值对比结果

Dataset	Xu等人 ^[5]	Nah等人 ^[8]	Kupyn等人 ^[9]	Zhang等人 ^[12]	Tao等人 ^[10]	Zhang等人 ^[14]	Gao等人 ^[13]	MSCNN (本文)
GOPR0384_11_00	23.65 dB/ 0.8088	31.44 dB/ 0.9403	29.13 dB/ 0.9184	29.21 dB/ 0.9191	32.20 dB/ 0.9458	32.23 dB/ 0.9464	32.84 dB/ 0.9515	33.31 dB/ 0.9562
GOPR0384_11_05	23.66 dB/ 0.8498	31.47 dB/ 0.9547	27.98 dB/ 0.9262	28.35 dB/ 0.9302	31.78 dB/ 0.9555	32.04 dB/ 0.9588	32.22 dB/ 0.9597	32.68 dB/ 0.9634
GOPR0385_11_01	24.23 dB/ 0.8383	31.02 dB/ 0.9430	28.19 dB/ 0.9227	28.25 dB/ 0.9230	31.70 dB/ 0.9469	32.25 dB/ 0.9525	32.32 dB/ 0.9508	32.95 dB/ 0.9573
GOPR0396_11_00	19.64 dB/ 0.7458	31.56 dB/ 0.9726	29.58 dB/ 0.9627	28.78 dB/ 0.9566	31.79 dB/ 0.9732	31.09 dB/ 0.9693	32.09 dB/ 0.9745	32.47 dB/ 0.9761
GOPR0410_11_00	22.65 dB/ 0.6912	29.14 dB/ 0.8922	28.22 dB/ 0.8799	28.19 dB/ 0.8828	29.53 dB/ 0.9044	30.21 dB/ 0.9105	30.24 dB/ 0.9147	30.31 dB/ 0.9180
GOPR0854_11_00	21.10 dB/ 0.7003	26.24 dB/ 0.8494	24.95 dB/ 0.8317	25.33 dB/ 0.8426	28.20 dB/ 0.8953	28.45 dB/ 0.9007	29.12 dB/ 0.9107	28.90 dB/ 0.9100
GOPR0862_11_00	20.13 dB/ 0.6856	25.79 dB/ 0.8600	24.12 dB/ 0.8346	24.34 dB/ 0.8392	27.30 dB/ 0.8936	27.71 dB/ 0.9055	28.40 dB/ 0.9142	28.46 dB/ 0.9136
GOPR0868_11_00	20.80 dB/ 0.7988	26.52 dB/ 0.9169	24.63 dB/ 0.8955	24.81 dB/ 0.8980	28.63 dB/ 0.9442	28.68 dB/ 0.9459	29.50 dB/ 0.9522	29.77 dB/ 0.9550
GOPR0869_11_00	21.56 dB/ 0.7326	28.05 dB/ 0.9005	26.27 dB/ 0.8787	26.51 dB/ 0.8811	29.47 dB/ 0.9263	29.62 dB/ 0.9300	30.32 dB/ 0.9380	30.20 dB/ 0.9382
GOPR0871_11_00	21.15 dB/ 0.7421	26.94 dB/ 0.8903	25.64 dB/ 0.8814	25.51 dB/ 0.8795	28.62 dB/ 0.9202	29.07 dB/ 0.9263	29.55 dB/ 0.9317	29.68 dB/ 0.9337
GOPR0881_11_01	22.75 dB/ 0.8454	30.29 dB/ 0.9446	27.48 dB/ 0.9294	27.58 dB/ 0.9311	32.51 dB/ 0.9625	33.08 dB/ 0.9656	33.67 dB/ 0.9679	33.58 dB/ 0.9684
Mean	22.00 dB/ 0.7665	29.02 dB/ 0.9154	27.02 dB/ 0.8973	27.08 dB/ 0.8992	30.20 dB/ 0.9334	30.45 dB/ 0.9373	30.96 dB/ 0.9421	31.15 dB/ 0.9444

为了能够更加全面地证明本文提出方法的优越性, 除了客观的评价指标之外, 还在主观的视觉效果上将本文提出的 MSCNN 与文献 [5,8-10,12-14] 中的方法进行了比较. 如图 7 所示, 每组图像依次是: 模糊图像、原始的清晰图像、文献 [5]、文献 [8]、文献 [9]、文献 [12]、文献 [10]、文献 [14]、文献 [13] 和本文提出方法的复原结果. 可见, 文献 [5,8-10,12-14] 中的方法所得到的复原图像分别会存在有一种或者多种不同程度的瑕疵: 振铃效应、过度的平滑、色彩和形状的失真和变形等. 例如, 如图 7 中的第 (a) 幅图像所示, 文献 [5,8,9,12] 方法并不能彻底地去除模糊, 其复原图像仍然存在有不同程度的模糊影响. 而文献 [10,13] 和 [14] 方法所得到的复原图像会存在有一定程度的振铃效应 (详见图 7 中的第 (a) 幅图像和相应的放大区域). 如图 7 中的第 (b) 幅图像所示, 文献 [5,8-10,12-14] 方法的复原图像均存在有显而易见的振铃效应, 同时, 文献 [8,9] 的方法还会平滑掉较多的边缘和细节 (详见图 7 中的第 (b) 幅图像和相应的放大区域). 从图 7 中的第 (c) 幅图像和第 (d) 幅图像同样能够明显地看到文献 [5,8-10,12-14] 方法的瑕疵 (详见图 7 中的第 (c) 幅图像和第 (d) 幅图像, 以及相应的放大区域). 相比之下, 本文提出的 MSCNN 很明显能够有效去除各类瑕疵, 复原出更加锐化的边缘和细节, 得到更高质量的去模糊图像. 表 2 和图 7 从客观的评价指标和主观的视觉效果两个方面很好地证明了本文提出的 MSCNN 在 Gopro 数据集上的优越性.

在进行方法性能的比较实验之后, 本小节还将本文提出的 MSCNN 与 Tao 等人的方法^[10]、Zhang 等人的方法^[14]、Gao 等人的方法^[13]和 Zhang 等人的方法^[14]在模型大小和平均的运行时间上进行了比较 (GoPro 测试图像的图片分辨率为 720×1280). 如表 3 所示, 本文提出的 MSCNN 模型略大于 Tao 等人的方法^[10]和 Zhang 等人的方法^[14], 却远小于 Gao 等人的方法^[13]; 在平均运行时间上, 本文提出的 MSCNN 仅比 Zhang 等人的方法^[14]慢, 而比

Tao 等人的方法^[10]和 Gao 等人的方法^[13]都快;而在平均的 PSNR 值和 SSIM 值上均高于所有的方法.因此,表 3 很好地证明了虽然本文提出的方法能够获得最好的性能,但是却并不是以牺牲较多效率为代价的.

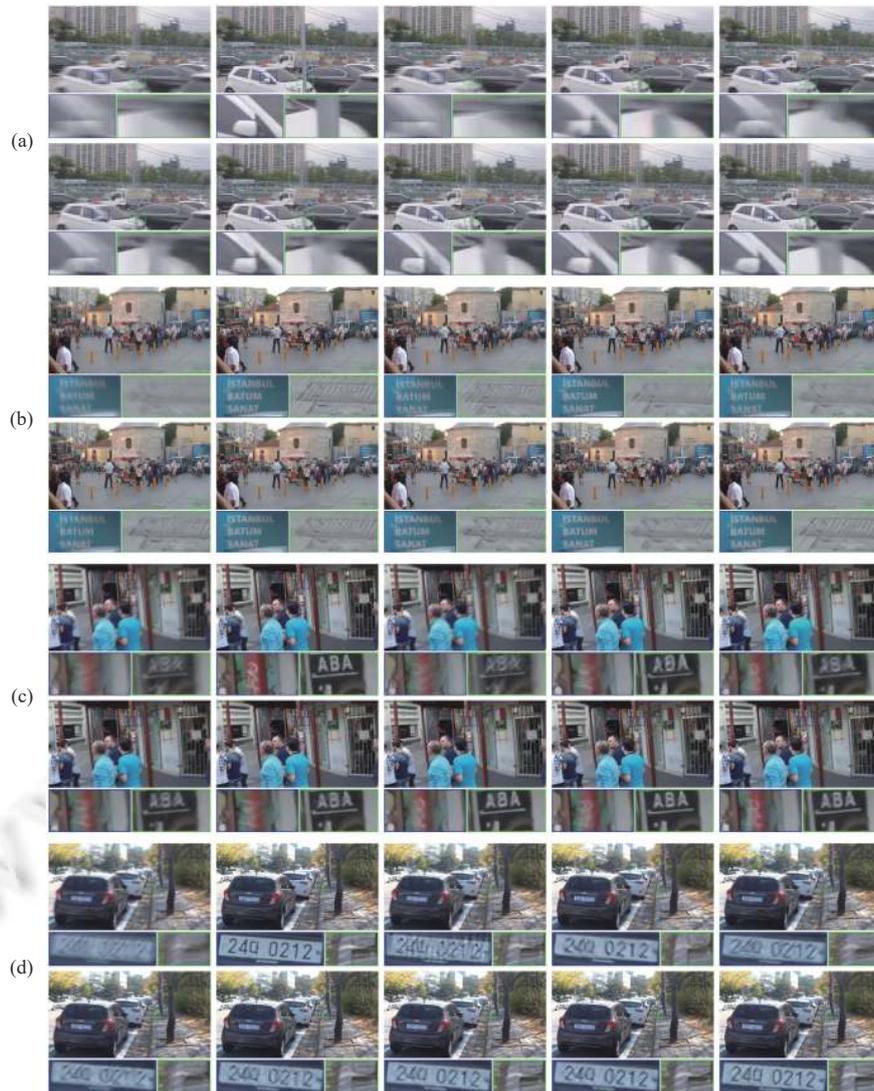


图 7 在 GoPro 测试集上的主观视觉效果图比较

表 3 在 1 111 幅 GoPro 测试集上的结果比较

方法	PSNR (dB)	SSIM	模型大小 (M)	耗时 (s)
Tao 等人 ^[10]	30.20	0.9334	8.1	0.79
Zhang 等人 ^[14]	30.45	0.9373	7.2	0.08
Gao 等人 ^[13]	30.96	0.9421	11.6	0.85
Zhang 等人 ^[15]	31.10	0.9424	—	—
MSCNN (本文)	31.15	0.9444	8.7	0.48

(2) 视频去模糊数据集上的比较.为了验证本文提出方法的泛化性,我们将 MSCNN 运用到 Su 等人提出的视频去模糊的数据集中^[19].表 4 所示是本文提出的 MSCNN 与 Tao 等人的方法^[10]、Zhang 等人的方法^[14]和 Gao 等

人的方法^[13]在视频去模糊数据集上的客观评价指标比较.如表4所示,本文提出的MSCNN在具有挑战性的视频去模糊数据集上也同样能获得最高的平均PSNR值和平均SSIM值,很好地证明了本文提出的MSCNN的泛化优越性.

表4 在视频去模糊数据集上的平均PSNR和平均SSIM值比较

数据集	Tao等人 ^[10]	Zhang等人 ^[14]	Gao等人 ^[13]	MSCNN(本文)
PSNR (dB)	29.57	28.12	29.80	29.91
SSIM	0.916 9	0.886 2	0.919 3	0.919 6

(3) 真实模糊图像上的比较.为了进一步验证本文提出方法的性能,将本文提出的MSCNN与Tao等人的方法^[10]、Zhang等人的方法^[14]和Gao等人的方法^[13]在真实的模糊图像上进行比较.如图8所示,在真实的模糊图像上,文献[10,13,14]中的方法所得到的复原图像也同样都会分别存在有一种或者多种不同程度的瑕疵:振铃效应、过度的平滑、色彩和形状的失真和变形等.而相比之下,本文提出的MSCNN在真实的模糊图像中也能够很明显地获得更高质量的复原图像:更少的振铃效应,几乎没有颜色的失真和形状的形变,同时具有更加锐化的边缘和更多的细节(如图8中的放大区域所示).图8很好地证明了本文提出的MSCNN在真实模糊图像上的优越性.

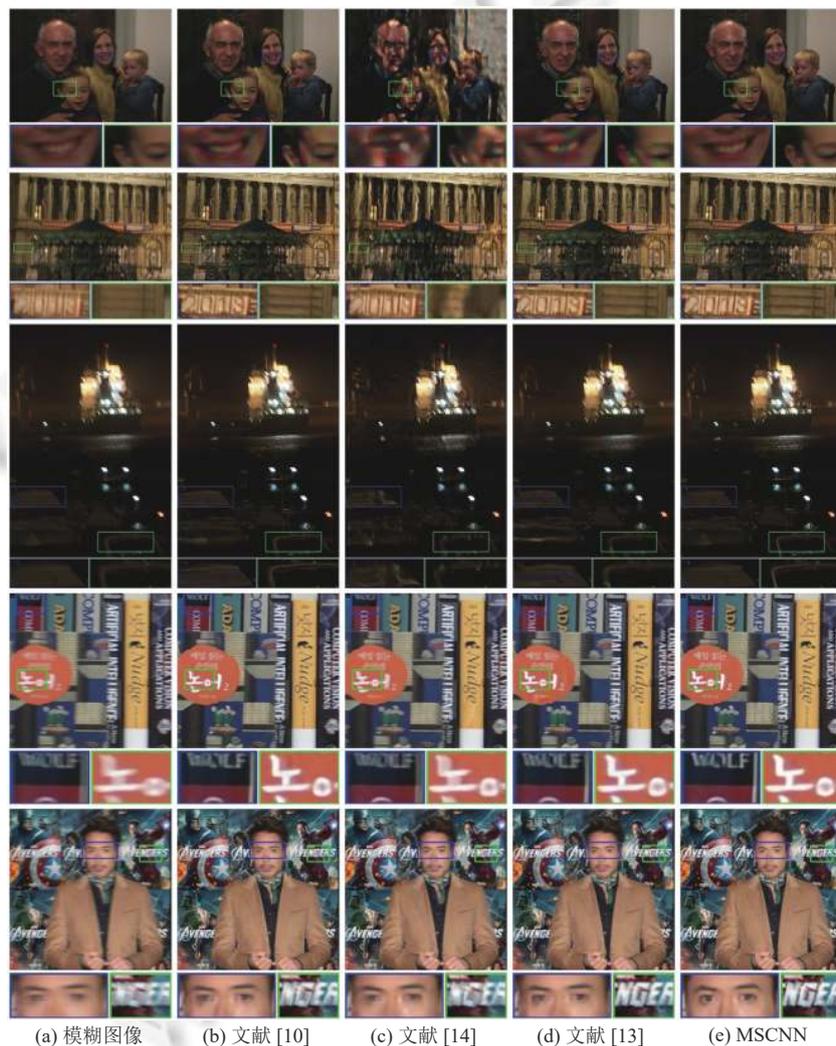


图8 真实模糊图像上的复原结果比较

4 结 论

本文提出了一种新的多尺度卷积神经网络 (MSCNN) 来实现高质量的单幅图像动态场景盲复原. 在本文中, 我们充分利用了编解码器这种结构, 提出了在一个尺度内利用多个不同的编码器和解码器的思想, 增加了编解码器的深度, 而这与简单的通过增加卷积模块的数量来增加网络的深度是完全不同的, 因此能够提升每一个尺度的去模糊性能. 除此之外, 本文还在 SPP 和 DMPHN 的启发下, 提出了一种 HMPCA 来增加通道方向的感受野, 从而进一步增强网络盲去模糊的表达能力. 结合多编解码器和 HMPCA, 本文提出的 MSCNN 最终实现了更高质量的动态场景盲去模糊. 在人造数据集和真实模糊图像上的大量实验结果表明: 本文提出的方法较近几年的一些成功的动态场景盲去模糊方法相比, 能够复原出更高质量的清晰图像, 在客观的评价指标和主观的视觉效果上均有显著的改进. 将本文提出的 MSCNN 扩展到视频去模糊领域是我们接下来研究工作的重点.

致谢 在此, 我们对对本文的工作给予帮助和支持的同行表示感谢, 尤其感谢 Xu Li, Nah Seungjun, Tao Xin 和 Zhang Hongguang 等人提供他们方法的程序源代码或者模型.

References:

- [1] Harmeling S, Hirsch M, Schölkopf B. Space-variant single-image blind deconvolution for removing camera shake. In: Proc. of the 23rd Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2010. 829–837.
- [2] Gupta A, Joshi N, Zitnick CL, Cohen M, Curless B. Single image deblurring using motion density functions. In: Proc. of the 11th European Conf. on Computer Vision. Heraklion: Springer, 2010. 171–184. [doi: 10.1007/978-3-642-15549-9_13]
- [3] Hirsch M, Schuler CJ, Harmeling S, Schölkopf B. Fast removal of non-uniform camera shake. In: Proc. of the 2011 Int'l Conf. on Computer Vision. Barcelona: IEEE, 2011. 463–470. [doi: 10.1109/ICCV.2011.6126276]
- [4] Xu L, Jia JY. Depth-aware motion deblurring. In: Proc. of the 2012 IEEE Int'l Conf. on Computational Photography. Seattle: IEEE, 2012. 1–8. [doi: 10.1109/ICCPHOT.2012.6215220]
- [5] Xu L, Zheng SC, Jia JY. Unnatural L0 sparse representation for natural image deblurring. In: Proc. of the 2013 IEEE Conf. on Computer Vision and Pattern Recognition. Portland: IEEE, 2013. 1107–1114. [doi: 10.1109/CVPR.2013.147]
- [6] Hu Z, Xu L, Yang MH. Joint depth estimation and camera shake removal from single blurry image. In: Proc. of the 2014 IEEE Conf. on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 2893–2900. [doi: 10.1109/CVPR.2014.370]
- [7] Xu XY, Sun DQ, Pan JS, Zhang YJ, Pfister H, Yang MH. Learning to super-resolve blurry face and text images. In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision. Venice: IEEE, 2017. 251–260. [doi: 10.1109/ICCV.2017.36]
- [8] Nah S, Kim TH, Lee KM. Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 257–265. [doi: 10.1109/CVPR.2017.35]
- [9] Kupyn O, Budzan V, Mykhailych M, Mishkin D, Matas J. DeblurGAN: Blind motion deblurring using conditional adversarial networks. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8183–8192. [doi: 10.1109/CVPR.2018.00854]
- [10] Tao X, Gao HY, Shen XY, Wang J, Jia JY. Scale-recurrent network for deep image deblurring. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8174–8182. [doi: 10.1109/CVPR.2018.00853]
- [11] Zhang YL, Tian YP, Kong Y, Zhong BN, Fu Y. Residual dense network for image restoration. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2021, 43(7): 2480–2495. [doi: 10.1109/TPAMI.2020.2968521]
- [12] Zhang JW, Pan JS, Ren J, Song YB, Bao LC, Lau RWH, Yang MH. Dynamic scene deblurring using spatially variant recurrent neural networks. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 2521–2529. [doi: 10.1109/CVPR.2018.00267]
- [13] Gao HY, Tao X, Shen XY, Jia JY. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 3843–3851. [doi: 10.1109/CVPR.2019.00397]
- [14] Zhang HG, Dai YC, Li HD, Koniusz P. Deep stacked hierarchical multi-patch network for image deblurring. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 5971–5979. [doi: 10.1109/CVPR.2019.00613]
- [15] Zhang KH, Luo WH, Zhong YR, Ma L, Stenger B, Liu W, Li HD. Deblurring by realistic blurring. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 2734–2743. [doi: 10.1109/CVPR42600.2020.00281]

- [16] He KM, Zhang XY, Ren SQ, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916. [doi: [10.1109/TPAMI.2015.2389824](https://doi.org/10.1109/TPAMI.2015.2389824)]
- [17] Zhang YL, Li KP, Li K, Wang LC, Zhong BN, Fu Y. Image super-resolution using very deep residual channel attention networks. In: Proc. of the 15th European Conf. on Computer Vision. Munich: Springer, 2018. 294–310. [doi: [10.1007/978-3-030-01234-2_18](https://doi.org/10.1007/978-3-030-01234-2_18)]
- [18] He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778. [doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90)]
- [19] Su SC, Delbracio M, Wang J, Sapiro G, Heidrich W, Wang O. Deep video deblurring for hand-held cameras. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 237–246. [doi: [10.1109/CVPR.2017.33](https://doi.org/10.1109/CVPR.2017.33)]



唐述(1981—)男,博士,副教授,CCF专业会员,主要研究领域为信息获取与处理,图像处理.



黄蓉(1994—)女,硕士,主要研究领域为图像处理,深度学习.



万盛道(1995—)男,硕士,主要研究领域为图像处理,深度学习.



顾佳(1996—)女,硕士,主要研究领域为图像处理.



谢显中(1965—)男,博士,教授,CCF高级会员,主要研究领域为无线通信技术,通信应用软件.



郑万鹏(1996—)男,硕士,主要研究领域为图像去模糊.



杨书丽(1995—)女,博士生,主要研究领域为图像处理,深度学习.