

引文格式:

孙有晓, 冯锐成, 关伟鹏, 等. 基于图像超分辨率的多尺度特征交互传播研究 [J]. 集成技术, 2022, 11(2): 41-54.

Sun YX, Feng RC, Guan WP, et al. Research on multi-scale feature propagation and communication based on image super resolution [J]. Journal of Integration Technology, 2022, 11(2): 41-54.

基于图像超分辨率的多尺度特征交互传播研究

孙有晓^{1,2,3} 冯锐成² 关伟鹏² 乔宇² 董超^{2*}

井焜⁴ 刘辰飞⁴ 许野平⁴ 陈英鹏⁴ 周卫东¹

¹(山东大学微电子学院 济南 250101)

²(中国科学院深圳先进技术研究院 深圳 518055)

³(山东中科先进技术研究院有限公司 济南 250101)

⁴(神思电子技术股份有限公司 济南 250101)

摘要 该文构建了一种可对不同形式的多尺度结构进行归纳的统一框架,并基于该框架系统地探究了多尺度卷积的两个因素——特征传播和跨尺度交互,提出了简单而有效的多尺度卷积单元——多尺度-跨尺度-权重共享的卷积(MS³-Conv)网络。实验结果表明,与基于标准卷积的网络相比,基于MS³-Conv的网络可使用较少的参数和较低的计算成本实现更好的图像重建性能。除了定量分析,该文也对重建图像进行可视化分析,证明了MS³-Conv网络能更好地恢复高频细节。

关键词 深度学习; 多尺度卷积; 超分辨率

中图分类号 TP 399 **文献标志码** A **doi**: 10.12146/j.issn.2095-3135.20211026002

Research on Multi-scale Feature Propagation and Communication Based on Image Super Resolution

SUN Youxiao^{1,2,3} FENG Ruicheng² GUAN Weipeng² QIAO Yu² DONG Chao^{2*}

JING Kun⁴ LIU Chenfei⁴ XU Yeping⁴ CHEN Yingpeng⁴ ZHOU Weidong¹

¹(School of Microelectronics, Shandong University, Jinan 250101, China)

²(Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

³(Shandong Institute of Advanced Technology Co., Ltd., Chinese Academy of Sciences, Jinan 250101, China)

⁴(Synthesis Electronics Technology Co., Ltd., Jinan 250101, China)

*Corresponding Author: chao.dong@siat.ac.cn

收稿日期: 2021-10-26 修回日期: 2021-12-16

基金项目: 山东省发展和改革委员会新旧动能转换重大工程重大课题攻关项目; 山东省自然科学基金联合基金重点项目(ZR2020LZH009)

作者简介: 孙有晓, 硕士研究生, 研究方向为图像处理; 冯锐成, 博士, 研究方向为图像处理; 关伟鹏, 博士研究生, 研究方向为视频图像超分与深度学习; 乔宇, 研究员, 研究方向为计算机视觉; 董超(通讯作者), 副研究员, 博士研究生导师, 研究方向为计算机视觉, E-mail: chao.dong@siat.ac.cn; 井焜, 博士, 应用研究员, 研究方向为人工智能; 刘辰飞, 高级工程师, 研究方向为人工智能与图像处理; 许野平, 应用研究员, 研究方向为人工智能与图像处理; 陈英鹏, 硕士研究生, 研究方向为人工智能与图像处理; 周卫东, 教授, 博士研究生导师, 研究方向为深度学习和信号处理。

Abstract This paper presents a unified framework for various multi-scale structures. With this framework, two factors of multi-scale convolution, i.e. feature propagation and cross-scale communication are explored. A generic and efficient multi-scale convolution unit named Multi-Scale cross-Scale Share-weights Convolution (MS^3 -Conv) is proposed. Experimental results showed that, the proposed MS^3 -Conv can achieve better super resolution performance than conventional convolution methods with less parameters and computational cost. By observation of the visual quality, results also showed that the MS^3 -Conv outperform in the reconstruction of high-frequency image details.

Keywords deep learning ; multi-scale convolution; super-resolution

Funding This work is supported by Major Project of Shandong Development and Reform Commission, and Key Program of Joint Fund of Shandong Provincial Nature Science Foundation (ZR2020LZH009)

1 引言

图像超分辨率(Image Super Resolution, SR)本质也是多尺度问题,重建得到的高分辨(High-Resolution, HR)图像,可以看作是通过扩大低分辨率(Low-Resolution, LR)图像的尺度得到的。随着深度学习在SR^[1-2]中的开创性探索,多尺度技术以多种方式应用于图像超分辨率领域。在深度学习方法取得巨大成功之前,多尺度技术最早应用在基于图像自相似(Self-Similarity)的超分辨率算法中,其中图像金字塔里的相似块是由LR图像逐阶下采样得到的。

目前,多尺度技术被广泛应用于深度学习任务中。本文将多尺度技术的应用概括为以下几方面:(1)多尺度技术在计算机视觉任务中应用广泛。例如,基于多尺度中间特征的Lap-SRN^[3]和U-Net^[4]网络结构。FPN^[5]和PSP^[6]在网络末端融合不同深度的特征,主要应用于目标检测和分割任务中。MSR-GCN^[7]提出多尺度残差图卷积网络,用于端到端的人体姿势预测。主要应用于图像分类识别的MSDNet^[8]提出了包含多个分支(每个分支都有对应的空间分辨率)的网络架构,实现了计算资源的合理分配。HRNets^[9]通过在并行的多分辨率子网络上交互信息来进行多尺度特

征的融合操作,实现用高分辨表征来估计姿态关键点。同样应用于姿态估计的Hourglass^[10]通过短连接将降采样到升采样过程中的底层特征与原尺度的特征进行组合,以获取多尺度特征信息。Big-Little Net^[11]和Elastic^[12]亦采用类似的设计,使用不同分支处理不同的尺度特征,从而捕获跨尺度的信息。PMRN^[13]改进了多尺度残差块,用小滤波器的组合代替较大的滤波器,逐步探索每层信息,能够有效地恢复结构纹理。SRNTT^[14]以多尺度的方式将匹配的特征自适应地迁移到重建图像。Multigrid CNNs^[15]提出了一种多网格金字塔特征表示,并定义了可以在整个网络中集成的多重网格卷积(Multi-grid Convolution, MG-Conv)算子。类似于MG-Conv算子,Octave Convolution(Oct-Conv)^[16]在不同的空间分辨率下对特征图进行卷积,对低频分量进行了压缩以减少空间冗余。(2)多尺度技术能够很好地融合到神经网络的注意力机制中。例如,主要应用在图像识别分类任务中的MViT^[17]将多尺度特征层次结构的核心思想与Transformer模型相结合。EPSANet^[18]在通道注意力的基础上,引入多尺度思想并构建了EPSA模块,相比其他单尺度注意力机制,EPSANet能够提取更加丰富的多尺度特征图空间信息,实现跨维度通道注意力特征的交

互。DMSANet^[19]聚合不同尺度的特征, 并使用空间和通道注意力模块来自适应地将局部特征与全局依赖关系相结合。TTSR^[20]利用跨层级的特征融合机制来增强网络的特征表示能力。基于非局部注意力机制的RCAN^[21]、SAN^[22]网络只是利用了单一尺度特征, 而CSNLN^[23]提出了跨尺度非局部注意力模块, 利用自然图像中广泛存在的跨尺度特征相似性来提高重建性能。(3) 由于深度学习网络模型需要大量的数据样本进行训练才能达到较好的泛化能力, 因此可以通过数据增强来增加训练样本的数量以及多样性。一般在数据增强任务中, 常见的操作有图像旋转、镜像变换以及图像缩放。其中, 图像缩放操作通过设置不同的尺度缩放因子(0.6~0.9)来调整训练数据集的图像大小比例就是利用多尺度技术。

虽然多尺度技术在图像超分辨率领域应用广泛, 但很少有文献深入研究多尺度卷积本身。因此, 本文系统地研究了SR中多尺度卷积的设计和使用, 提出一种对不同形式的多尺度结构进行归纳的新框架。具体来说, 在统一框架中现有的不同多尺度结构可以写成形式相同但转换函数不同的表达式。如公式(1), 使用统一框架表述U-Net、Octave卷积和Multi-grid卷积。该公式表明, 多尺度卷积的性能取决于两个因素——特征传播和跨尺度通信。其中, 特征传播决定每个尺度中信息转化的方式, 跨尺度通信控制跨尺度的信息流动。在此基础上, 本文调整转换函数以探究最有效的结构设计。图1(c)展示了改进多尺度

卷积形式的过程。多尺度卷积(MS-Conv)将输入图像的特征信息划分成两个不同尺度的分支并单向传播; 多尺度+跨尺度卷积(MS²-Conv)增加了跨尺度的通信路径, 使信息可以在不同尺度之间转换; 为了进一步减少参数量和计算成本, 高低尺度的滤波器采用权重共享的策略, 使用较小尺度的跨尺度滤波器, 并将这种卷积命名为多尺度+跨尺度+共享权重卷积(Multi-Scale+cross-Scale+Share-weight Convolution, MS³-Conv)。MS³-Conv与基于标准卷积的网络的性能指标值(Peak Signal to Noise Ratio, PSNR)几乎相同, 但其计算复杂度、总参数量仅分别为后者的67%和75%, MS³-Conv与基于标准卷积的网络的对比验证了该多尺度卷积结构不同部件和策略的有效性。本文的贡献主要有3个方面:(1)对理解和设计多尺度卷积网络作出统一解释。在统一解释多尺度卷积网络这个框架中, 现有的多尺度结构可以使用相同的形式表述, 但转换函数不同。(2)基于这个统一解释多尺度卷积网络的框架, 本文对多尺度卷积的形式进行了系统的研究, 并提出一种计算速度更快和重建性能更好的新型高效多尺度结构。(3)本文全面研究了多尺度卷积网络的重建图像在不同区域的视觉质量差异。通过大量实验证明多尺度网络更优于重建高频细节。

2 多尺度网络

本文提出归纳不同形式的多尺度卷积的统一

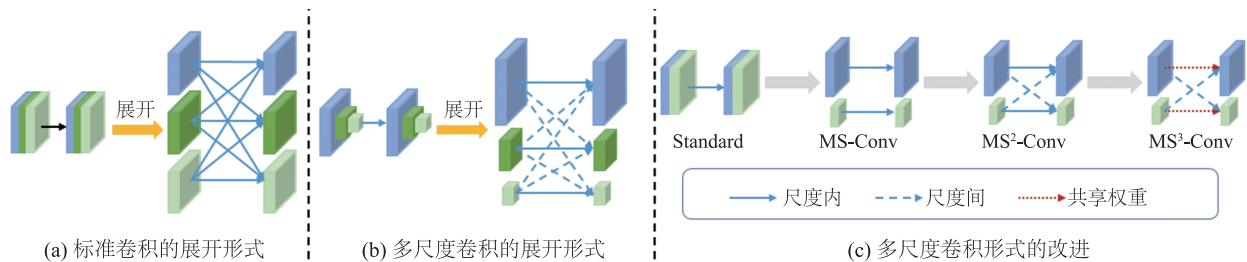


图1 标准卷积与多尺度卷积的结构比较

Fig. 1 The structural comparison of standard convolution and multi-scale convolution

框架,在该框架下提出了“多尺度+跨尺度+共享权重”卷积(MS³-Conv),并阐述了多尺度与权重的重要作用。

2.1 多尺度卷积的表示

本文首先提出了归纳不同形式的多尺度卷积的统一框架。为了研究基于不同形式的多尺度卷积网络的关系,将多尺度特征图看作是多个不同尺度的并行分支,如图1(b)。本文提出了基于多尺度特征表示与权重共享卷积层的网络模型,并将其分为3个部分:将传统卷积网络的第一层改为多尺度特征分离的卷积层,最后一层改为多尺度特征融合的卷积层,其他所有层改为多尺度特征映射与交互的卷积层。基于多尺度特征表示与权重共享卷积层的网络将输入图像的特征信息划分为更高(更精细)尺度和更低(更粗略)尺度的特征,分别定义为 X_H 和 X_L 。将采集到的信息通过转换函数 $f(\cdot)$ 映射到相应的输出分支,每个分支的输出特征 (Y_H, Y_L) 可以分解为四项。其中, f_{HH} 和 f_{LL} 为尺度内的信息转换; f_{HL} 和 f_{LH} 为尺度间的信息转换。然后将转换后的特征信息相加,即 $Y_H = f_{HH}(X_H) + f_{LH}(X_L)$ 。多尺度卷积可以表述如下:

$$\begin{bmatrix} Y_H \\ Y_L \end{bmatrix} = \begin{bmatrix} f_{HH} & f_{LH} \\ f_{HL} & f_{LL} \end{bmatrix} \begin{bmatrix} X_H \\ X_L \end{bmatrix} \quad (1)$$

公式(1)概括了输入特征的转换融合过程。如图1(b)所示,多尺度特征表示与权重共享卷积层的展开形式类似于具有尺度间和尺度内转换的多尺度分支全连接网络。如图1(a)所示,左侧标准卷积等效于右侧多分支的卷积,其中, X_H 和 X_L 的尺度相同。在多尺度特征表示与权重共享卷积层的网络中,通过尺度间转换与上采样或下采样算子相结合来匹配融合特征的空间分辨率。同时,公式(1)可以很容易地扩展到具有2个以上分支(尺度)并行的网络。

对于U-Net网络模型,先将U-Net结构中的多尺度卷积归纳,其中,特征图按比例因子2进

行下采样。转换矩阵定义为:

$$f = \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & W_{LL} \end{bmatrix} \quad (2)$$

其中, I 为恒等映射; W_{LL} 为低尺度内部转换的卷积。这表明高分辨率信息可以通过短连接传播,低分辨率尺度通过卷积滤波器进行转换,不同尺度的分支之间并没有实现信息转换。

对于Oct-Conv网络模型,Octave多尺度卷积的转换矩阵可以表示为:

$$f = \begin{bmatrix} W_{HH} & \uparrow \circ W_{LH} \\ W_{HL} \circ \downarrow_{avg} & W_{LL} \end{bmatrix} \quad (3)$$

其中, $\uparrow \circ W_{LH}$ 为最近邻上采样 \uparrow 和卷积函数 W_{LH} 组合成的不同尺度之间转换的卷积操作; \downarrow_{avg} 表示步幅为2的平均池化操作; W_{HH} 为高尺度内部转换的卷积; W_{LL} 为低尺度内部转换的卷积。Octave卷积通过尺度内和尺度间转换函数将输入特征划分为较高(较精细)尺度和较低(较粗略)的特征,以减小内存占用量和计算成本。

MG-Conv网络模型利用多尺度特征,将特征图 X 划分为3个尺度,即 $X = \{X_H; X_M; X_L\}$ 。MG-Conv实现转换和融合的过程表示为:

$$\begin{bmatrix} Y_H \\ Y_M \\ Y_L \end{bmatrix} = \begin{bmatrix} W_{HH} & \uparrow \circ W_{MH} & \mathbf{0} \\ W_{HM} \circ \downarrow_{max} & W_{MM} & \uparrow \circ W_{LM} \\ \mathbf{0} & W_{ML} \circ \downarrow_{max} & W_{LL} \end{bmatrix} \begin{bmatrix} X_H \\ X_M \\ X_L \end{bmatrix} \quad (4)$$

其中, \downarrow_{max} 为最大池化操作,使高(精细)尺度和低(粗略)尺度进行横向通信; W_{HH} , W_{MM} 和 W_{LL} 表示尺度内部转换的卷积; W_{HM} , W_{MH} , W_{LM} 和 W_{ML} 均表示尺度间转换的卷积。值得注意的是,特征信息仅在相邻尺度之间进行传播,如 X_H 并不会被转换聚合到 Y_L ,反之亦然。

2.2 多尺度权重共享卷积

本文先研究最简单的多尺度卷积(MS-Conv)。MS-Conv将输入图像的特征信息划分成两个不同尺度的分支并单向传播。这种网络结构可以降低计算复杂度,但会导致重建性能下降。为了缓解这种情况,本文增加了跨尺度的通信路

径, 使信息可以在不同尺度之间转换。将改进的卷积命名为多尺度+跨尺度卷积(MS^2 -Conv)。为减少网络参数量和计算成本, 进一步提出了更高效的多尺度卷积, 并称为多尺度+跨尺度+共享权重卷积(MS^3 -Conv)。关键思想是保持尺度间的通信路径—— $f_{HL}, f_{LH} \neq 0$, 尺度内采用共享权重策略转换信息—— $f_{HH} = f_{LL}$ 。通过大量实验证明了跨尺度通信可以显著提高重建性能。其中, 采用共享权重策略是受 TridentNet^[24]的启发, 它构建了一个并行的多尺度分支结构, 并且不同的尺度使用相同的转换参数。此外, 跨尺度通信路径使用的卷积大小为 1×1 , 而不是 3×3 , 这样可以进一步减少冗余的参数。综上所述, MS^3 -Conv 的转换矩阵可表示为:

$$f = \begin{bmatrix} W_{IS} & \uparrow \circ W_{LH} \\ W_{HL} \circ \downarrow_{avg} & W_{IS} \end{bmatrix} \quad (5)$$

其中, W_{IS} 为用于特征传播的 3×3 卷积; W_{HL} 和 W_{LH} 是用于跨尺度通信的 1×1 卷积。图 2 展示了 MS^3 -Conv 如何替换 SRResNet 网络模型中的标准卷积。

MS^3 -Conv 和 Oct-Conv 之间的主要差异有:

(1) 对于尺度内转换, MS^3 -Conv 每个尺度都共享参数。高/低尺度的划分比例固定为 0.5。(2) 对于尺度间的通信, MS^3 -Conv 采用两个 1×1 卷

积核, 而 Oct-Conv 使用 3×3 卷积核。不同于 TridentNet 构建多尺度的分支块并将它们集成到骨干网络中, MS^3 -Conv 可以作为通用的“即插即用”卷积单元, 很容易替换标准卷积融入到大多数网络结构中。另外, MS^3 -Conv 采用卷积进行尺度间通信, 而 TridentNet 没有通信路径, 只是通过非极大值抑制来合并不同尺度分支的输出。

2.3 多尺度与权重的作用

多尺度表征一直被应用于捕捉空间长程依赖和空间冗余。为了研究多尺度卷积对空间冗余的影响, 本文将多尺度卷积与空洞卷积^[25]联系在一起, 通过空洞卷积(利用稀疏卷积核之间的空间来扩大感受野), 使网络的不同分支可以具有相同的结构和不同的感受野。

如图 3 所示, 探究了简单的多尺度表示形式 $D_2 \rightarrow W_{d=1} \rightarrow U_2$, 其中, D_2 和 U_2 分别表示缩放因子为 2 的最近邻下采样和上采样; $W_{d=1}$ 表示扩张率为 1 的卷积滤波器。这个函数还可以重新定义为 $W_{d=2} \rightarrow D_2 \rightarrow U_2$, 其中, $D_2 \rightarrow U_2$ 本质上是一个损失函数, 并且该函数不会改变空间特征图的分辨率。在这种情况下, 该多尺度表示形式是 $W_{d=2}$ (扩大感受野)和 $D_2 \rightarrow U_2$ (有损下/上采样)的组合。本文还探究了参数共享但扩张率不同的两个多尺度分支, 并定义为 $W_{d=2}^s$ 。考虑了常用多

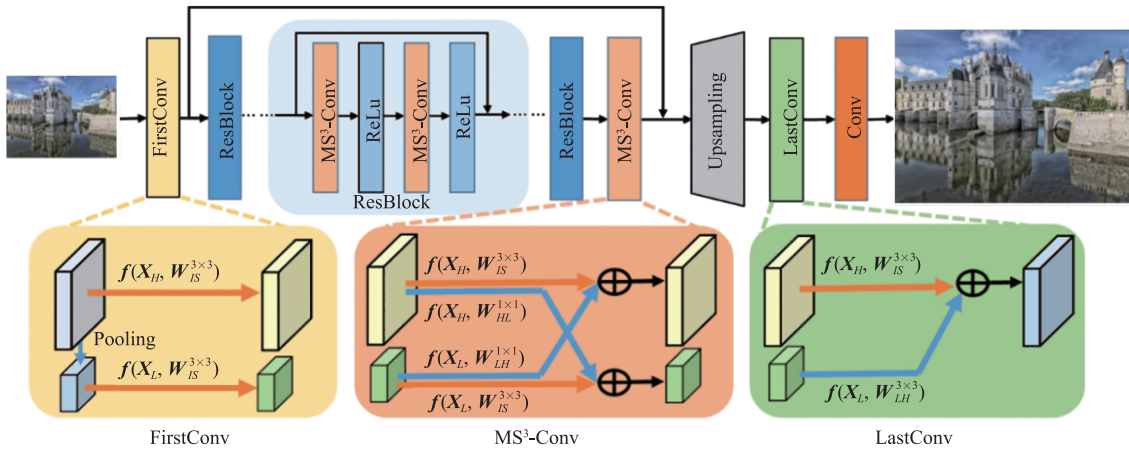
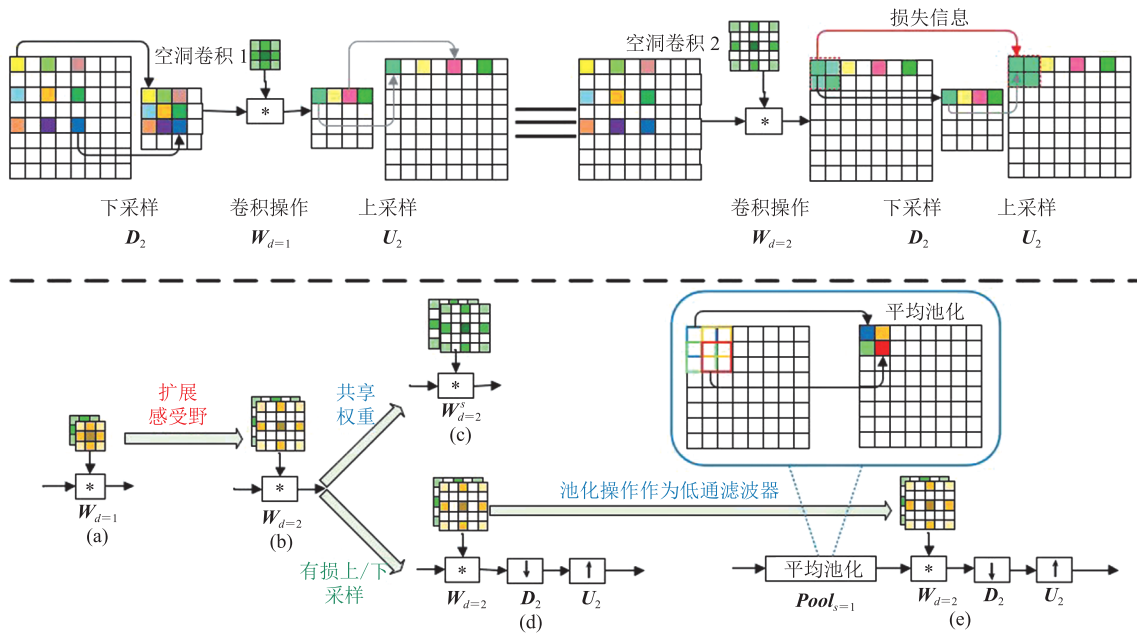


图 2 基于 MS^3 -Conv 的 SRResNet 网络模型

Fig. 2 SRResNet network model based on MS^3 -Conv



(上)下采样特征图进行卷积操作和扩张率为2的卷积操作；(下)不同多尺度卷积的关系

图3 不同形式的多尺度结构的比较

Fig. 3 Comparison of different forms of multi-scale structures

尺度结构 $Pool_{s=2} \rightarrow W_{d=2} \rightarrow D_2$ ，其中， $Pool_{s=2}$ 表示步幅为 2 的平均池化操作。类似于上式，这个函数可以改写为 $Pool_{s=1} \rightarrow W_{d=2} \rightarrow D_2 \rightarrow U_2$ ，其中，卷积滤波器之前的平均池化操作可以看作是低通滤波器，用来补偿下/上采样的损失信息。以上 5 种情况的预实验见表 1，实验表明，感受野大小和有损下采样对于 SR 不是至关重要的。本文推测这可能和图像的空间冗余有关。此外，相比每个分支采用共享权重策略的转换，更多的

表征能力(参数)只会带来微小的改进。然而，通过相同尺度间卷积核的权值共享的形式可以降低参数量，同时不同尺度的信息经过相同的非线性映射实现不同尺度特征的信息融合，保留不同尺度特征的多样性，进而能够在参数量与运算量降低的同时，保持了相同的重建性能值。因此，可合理地将卷积滤波器分配到低分辨率特征映射以降低计算复杂度，并使用不同尺度的共享权重分支来减少内存开销。

表 1 在 BSD100 数据集上的预实验

Table 1 Pilot experiments on BSD100 dataset

函数	图	PSNR (dB)	备注
$W_{d=1}$	3(下)(a)	27.58	
$W_{d=2}$	3(下)(b)	27.57	扩大感受野
$W_{d=2}^s$	3(下)(c)	27.56	共享权重
$W_{d=2} \rightarrow D_2 \rightarrow U_2$	3(下)(d)	27.56	损失函数
$Pool_{s=1} \rightarrow W_{d=2} \rightarrow D_2 \rightarrow U_2$	3(下)(e)	27.57	低通滤波器；补偿

3 实验设置

本文将 SRResNet^[26] 作为基准网络，它包含 16 个残差块，没有批归一化层。所有卷积层都有 64 个滤波器。为了衡量网络计算复杂度，采用常用指标——每秒浮点运算次数 (Floating-Point Operations Per second, FLOPs) 来计算乘法和加法操作的次数。实验模型采用 $\times 4$ 的缩放因子，使用 Pytorch^[27] 框架和单个 NVIDIA Titan

Xp GPU 进行训练。所有训练/测试模型都是基于以上相同条件来实现的。与 Lim 等^[28]实验一样, 本文使用 DIV2K 数据集^[29]中的 800 个图像对作为训练集。裁剪步幅设置为 240, 将 HR 图像裁剪成一组大小为 480×480 子图像。在训练期间, 先从 480×480 的子图像中随机裁剪出 128×128 的图像块作为真实数据, 然后使用双立方插值下采样得到 LR 图像。此外, 对原训练集随机水平/垂直翻转和 90° 旋转以增强数据。测试数据集使用 Set5^[30]、Set14^[31]、BSD100^[32]、Urban100^[33]、DIV2K 数据集。使用 YCbCr 空间的 Y 通道(即亮度)的 PSNR 数值作为性能评价指标。batch size 设置为 16, 采用 ADAM^[34]优化器来训练模型, 并设置相关参数 $\beta_1=0.9$, $\beta_2=0.999$, $\epsilon=10^{-8}$ 。初始学习率设置为 2×10^{-4} , 总的迭代次数是 10^6 , 接下来的学习率每迭代 2.5^5 次衰减一半。本文没有使用 Lim 等^[28]实验中的 ℓ_2 损失, 而是使用了 ℓ_1 损失。

4 实验与分析

4.1 多尺度卷积的有效性

首先分析比较了基于多尺度卷积与标准卷积网络的图像重建性能。为了将多尺度卷积集成到基准网络模型中, 使用 MS-Conv 替换对应的标准卷积层。如图 1(c), 为了公平比较, 网络的主拓扑和其他结构的参数保持一致。每个分支(尺度)都采用一个通道数为 32、卷积核大小为 3×3 的卷积算子。它们在 DIV2K 测试集上的 PSNR 值如表 2 所示, 将一半的特征图压缩到较低尺度时, 计算复杂度从 42.76 G 下降到 16.70 G, 但这样直接压缩会导致 PSNR 下降 0.16 dB。这说明 MS-Conv 可以实现重建性能和复杂性之间的平衡。在一些强调处理速度和低延时的实际应用中, 牺牲一定的重建性能是可接受的。其中, 表 2 中的参数量(Parameters, Params.)的单位是兆(M)。

表 2 标准卷积与多尺度卷积在 DIV2K 数据集上的性能对比

Table 2 Comparison of the standard and multi-scale convolution on DIV2K

卷积形式	转换矩阵	FLOPs (G)	Params. (M)	PSNR (dB)
标准卷积		42.76	1.59	30.47
MS-Conv	$\begin{bmatrix} W_{HL} & 0 \\ 0 & W_{LH} \end{bmatrix}$	16.70	0.82	30.31
MS ² -Conv w/o LH	$\begin{bmatrix} W_{HL} & 0 \\ f_{HL} & W_{LH} \end{bmatrix}$	18.74	1.21	30.33
MS ² -Conv w/o HL	$\begin{bmatrix} W_{HL} & f_{LH} \\ 0 & W_{LH} \end{bmatrix}$	18.74	1.21	30.29
MS ² -Conv	$\begin{bmatrix} W_{HL} & f_{LH} \\ f_{HL} & W_{LH} \end{bmatrix}$	20.78	1.59	30.38

注: “MS²-Conv w/o LH” “MS²-Conv w/o HL” 分别表示没有低到高、高到低的跨尺度通信路径

4.2 跨尺度通信的有效性

为了进一步提高性能, 本文评估跨尺度通信(MS²-Conv)的有效性。参考 Oct-Conv, 本文将转换函数 f_{HL} 和 f_{LH} 分别设置为 $\uparrow \circ W_{LH}$ 和 $W_{HL} \circ \downarrow_{avg}$ 。类似于特征传播卷积, W_{LH} 和 W_{HL} 是有 32 通道且卷积核大小为 3×3 的卷积层, 通过移除任意一条通信路径来测试单向跨尺度通信的有效性。在表 2 中, 移除任意一条跨尺度通信路径(高到低和低到高), 结果都会导致重建性能下降, 这表明了跨尺度通信路径的重要性。表 2 中的 MS²-Conv w/o LH 和 MS²-Conv w/o HL 分别与 MS-Conv 比较, 可以观察到单一通信路径网络和没有通信路径网络的重建性能指标数值近似相同(差异小于 0.02 dB), 这表明双向跨尺度连接对控制信息流动是至关重要的。此外, MS²-Conv 的 PSNR 比 MS-Conv 提高 0.07 dB, FLOPs 大约增加 4 G。同时, 本文还研究了网络复杂性和跨尺度通信对重建性能的作用, 以基于 MS-Conv 的网络作为参照网络, 比较了增加复杂度的两种情形: 第一种是逐渐增加网络深度; 第二种是固定网络深度并逐步将 MS-Conv 替换为 MS²-Conv,

如图 1(c)。图 4 在 BSD100 数据集上测试比较了这两种网络的性能指标——PSNR 和 FLOPs。可以观察到它们的重建指标数值差异相对较小，但增加跨尺度通信路径的重建性能指标数值的变化趋势更加平稳。这些结果表明跨尺度的通信路径能促进信息传播。

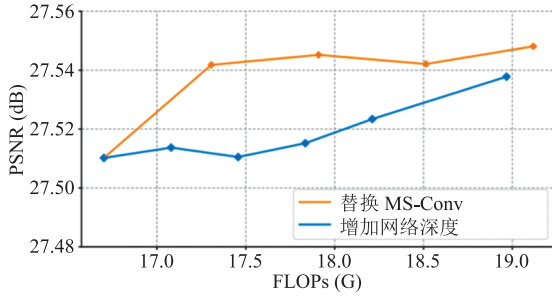


图 4 增加网络复杂性的两种情形

Fig. 4 Two cases of increasing the complexity of a network

4.3 权重共享的有效性

为了进一步减少参数量，探究了跨尺度通信的滤波器采用共享权重策略的有效性，并提出改进的卷积—— MS^3 -Conv。将 MS^3 -Conv 与其他形式的多尺度卷积进行比较，且评估了一系列不同深度网络的效率。其中， MS^3 -Conv 的尺度间通信路径的滤波器尺寸为 1×1 。 MS^3 -Conv-L 的尺度间通信路径的滤波器尺寸为 3×3 。为了评估滤波器大小对重建性能的影响，本文也对比了 MS^3 -Conv 和 MS^3 -Conv-L 的重建性能。图 5 展示了不同网络结构的性能变化，对比的指标包括 PSNR 和 Params.、PSNR 和 FLOPs。

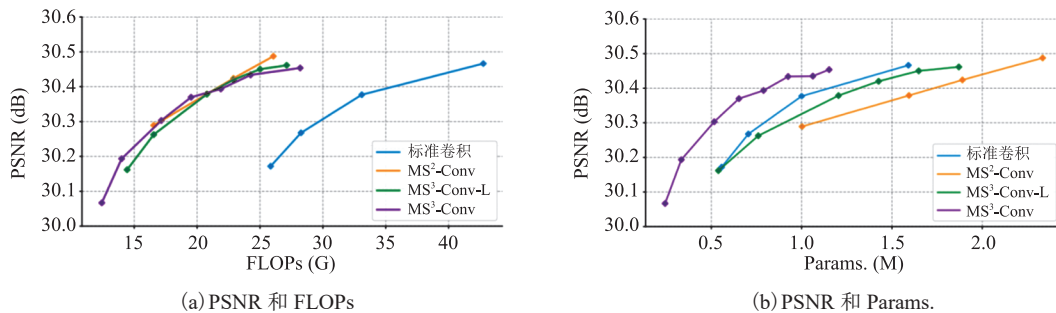


图 5 多尺度网络与标准卷积网络结构的性能比较

Fig. 5 The performance comparison of multi-scale networks and standard convolutional networks

如图 5(a)所示，当增加网络层数时， MS^2 -Conv、 MS^3 -Conv 和 MS^3 -Conv-L 的图像重建性能指标值的变化趋势相似，且所有基于多尺度卷积的网络的重建性能始终优于标准卷积网络。特别是，基于多尺度卷积的网络的计算复杂度仅为标准卷积网络的 61%，而其图像重建性能与标准卷积网络几乎相同，由此说明多尺度卷积可以充分利用特征表示。此外，在计算量大约 26 G FLOPs 的情况下，所有基于多尺度卷积的网络比标准卷积网络提升了 0.3 dB，这表明多尺度卷积的计算效率更高。从图 5(b)可以观察到，紫色曲线保持在绿色和黄色曲线上方， MS^3 -Conv 比 MS^2 -Conv 和 MS^3 -Conv-L 的内存效率更高。为达到标准卷积网络的重建性能指标值， MS^3 -Conv、 MS^3 -Conv-L 和 MS^2 -Conv 分别只需要 1.15 M、1.87 M 和 2.33 M 参数量，这表明增加尺度间通信路径的滤波器尺寸，只能得到较小的性能提升。值得注意的是， MS^3 -Conv-L 和 MS^2 -Conv 比标准卷积的参数量更多，这呼应了本文降低内存成本的操作。

4.4 多分支网络的有效性

为了研究 MS^3 -Conv 并行分支的数量对图像重建性能的影响(类似于 HRNet)，本文通过在不同尺度的并行分支之间交换特征信息来实现多尺度通信。由于额外的通信路径，具有更多分支模型的复杂度略高于两个分支的 MS^3 -Conv。表 3 中的结果表明，基于 MS^3 -Conv 网络的重建性能

始终优于单分支的标准网络, 大约增加 0.12 dB。另外, 当网络增加额外分支时, 相比二分支网络, 三分支网络和四分支网络并没有显著的改进 (小于 0.02 dB)。因此, 为了更好地平衡复杂性和性能, 本文将网络的尺度分支数量默认为 2。

表 3 在 DIV2K 数据集比较不同尺度数量的网络

Table 3 Comparison of networks with different numbers of scales on DIV2K

分支数量	FLOPs (G)	Params. (M)	PSNR (dB)
1	25.85	0.56	30.17
2	17.16	0.51	30.30
3	18.19	0.71	30.32
4	18.48	1.00	30.29

4.5 基于多尺度卷积的网络模型

为了体现本文提出的多尺度卷积的整体高效性, 将 MS³-Conv 应用于两个 SR 网络——CARN^[35]和 SRResNet 中, 构造出相应的多尺度网络模型。表 4 总结了基于多尺度卷积的网络模型在 SR 测试集上的测试结果。其中, 以 SRResNet 为骨干网络, 用 MS³-Conv 替换标准卷积可以减少 67% 的参数数量和 40% 的计算成本, 但这样会导致网络的图像重建性能显著下降。为了补偿重建性能的退化, 本文训练了一个更深的

网络, 表示为 MS³-Conv+, 仅使用 2/3 的计算复杂度和 3/4 的参数量就达到了标准网络的重建性能指标值。以 CARN 为骨干网络, MS³-Conv+ 相比标准卷积提升了较小的重建性能指标值 PSNR (在 Urban100 和 DIV2K 测试集上分别提升 0.01 dB 和 0.02 dB), 同时节省了 34% 的计算成本。有趣的是, 在大多数测试数据集上, 基于 MS³-Conv+ 卷积的 SRResNet 网络性能指标值略优于基于标准卷积的 CARN 网络 (两者在 DIV2K 测试集上的性能指标值分别为 30.46 dB 和 30.42 dB), 同时降低 20% 的计算成本。这意味着多尺度卷积可以在不改变网络拓扑的情况下提高网络性能。

5 结果可视化分析

本文在公开的标准数据集上比较了不同形式的多尺度卷积, 图 6 展示了一些具有代表性的定性结果, 并提供了 PSNR 和 SSIM 数值以供参考。对于图像 “img_093”, 基于标准卷积的网络复原的斑马条纹结构是错误的, 而所有基于多尺度卷积的网络的重构图像在高频细节方面都有较好的重建效果, 并趋于复原出正确的结构。对于图像 “img_038”, 标准卷积网络模型的复原图像有振铃伪影效果, 而基于 MS-Conv、MS²-

表 4 基于多尺度卷积的网络模型的定量结果

Table 4 Quantitative results of the network model based on multi-scale convolution

主干网络	卷积类型	FLOPs (G)	Params. (M)	PSNR (dB)				
				Set5	Set14	BSD100	Urban100	DIV2K
SRResNet	标准卷积	42.76	1.59	32.14	28.62	27.58	26.13	30.47
	MS ³ -Conv	17.16	0.52	32.01	28.48	27.51	25.84	30.30
	MS ³ -Conv+	28.98	1.19	32.07	28.60	27.57	26.12	30.46
CARN	标准卷积	35.51	1.15	32.15	28.58	27.57	26.03	30.42
	MS ³ -Conv	15.06	0.45	32.02	28.50	27.51	25.81	30.30
	MS ³ -Conv+	23.32	1.03	32.14	28.57	27.58	26.04	30.44

Conv 和 MS^3 -Conv 的网络模型重建结果有较好的视觉感知质量。除了基于不同形式的多尺度卷积的 SRResNet 网络模型, 本文还分析和比较了 MS^3 -Conv 应用于其他骨干网络模型的重构图像。如图 7 所示, 多尺度卷积的重构图像在高频细节方面具有更好的重建效果(如密集网格)。对于图像 “img_042” 和 “img_093”, 基于标准卷积的

两个网络模型都不能很好地恢复密集线条, 而对应的 MS^3 -Conv 网络模型都有较好的重建效果。对于图像 “img_098”, 观察到基于标准卷积的 SRResNet 网络无法完整地重建格子图形, 基于标准卷积的 CARN 网络重建的格子很模糊。相比之下, 对应的 MS^3 -Conv 网络可以减轻模糊效果并能够恢复更多高频细节。如图 8 所示, 在平坦

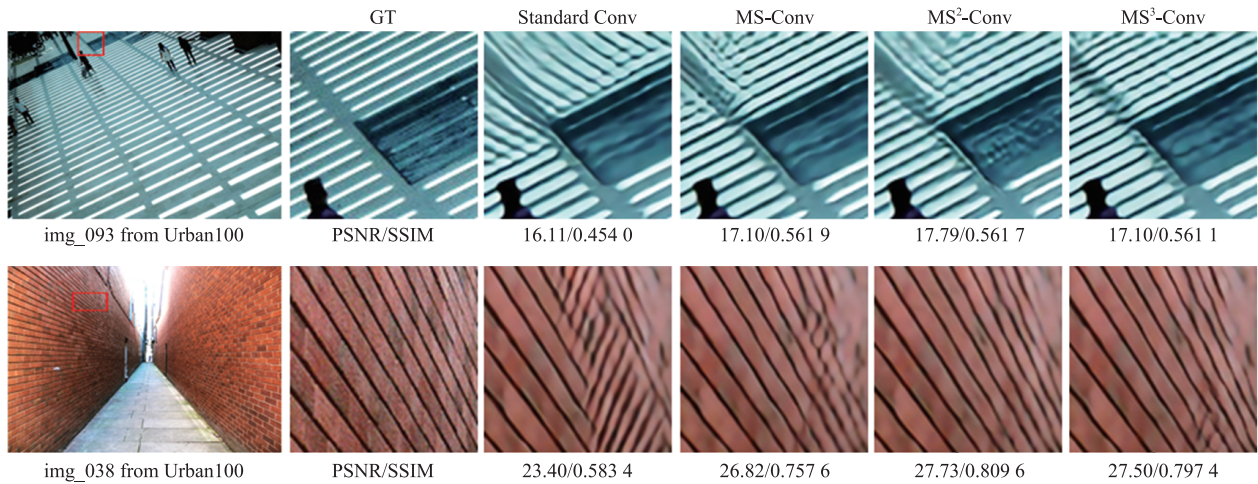


图 6 基于标准卷积和基于不同形式的多尺度卷积的 SRResNet 网络的定性比较

Fig. 6 Qualitative comparisons of standard convolution and variants of multi-scale convolution on SRResNet backbone

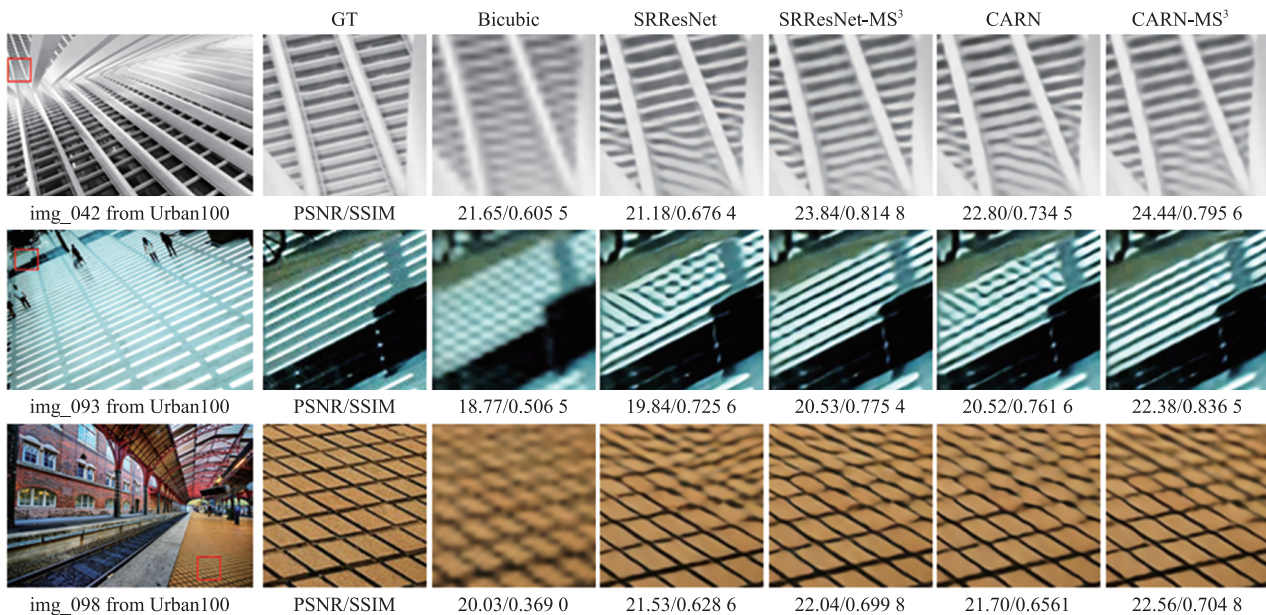


图 7 密集网格区域的定性比较

Fig. 7 Qualitative comparisons on dense-grid region

区域和平滑边缘, 可以观察到基于多尺度卷积与标准卷积的网络的 PSNR 数值有较大差距, 但无显著的视觉感知差异。对于图像 “img_0830”, 基于 MS³-Conv 的网络比基于标准卷积的网络的图像重建性能指标 PSNR 数值下降 3.4 dB, 但两者的视觉感知上无明显差异。这些基于区域的变化表明, 多尺度卷积与标准卷积对不同区域的重建能力不一致。如图 9 所示, 比较了多尺度图像的可视化效果。本文使用缩放因子 {0.8, 0.6, 0.5}

调整图像, 以对比这些网络模型的重建性能。基于 MS³-Conv 的 SRResNet 网络比基于标准卷积的 SRResNet 网络的图像重建性能更好(见黄色框)。基于标准卷积和基于 MS³-Conv 的 CARN 网络模型都能正确地重建出尺度因子为 0.8 和 0.6 的建筑物结构。对于较低的尺度(0.5)的图像, 标准卷积网络重建的图像丢失部分结构并出现混叠效应(见绿色框)。而基于 MS³-Conv 的 CARN 网络则可以缓解这种现象并恢复正确的结构。

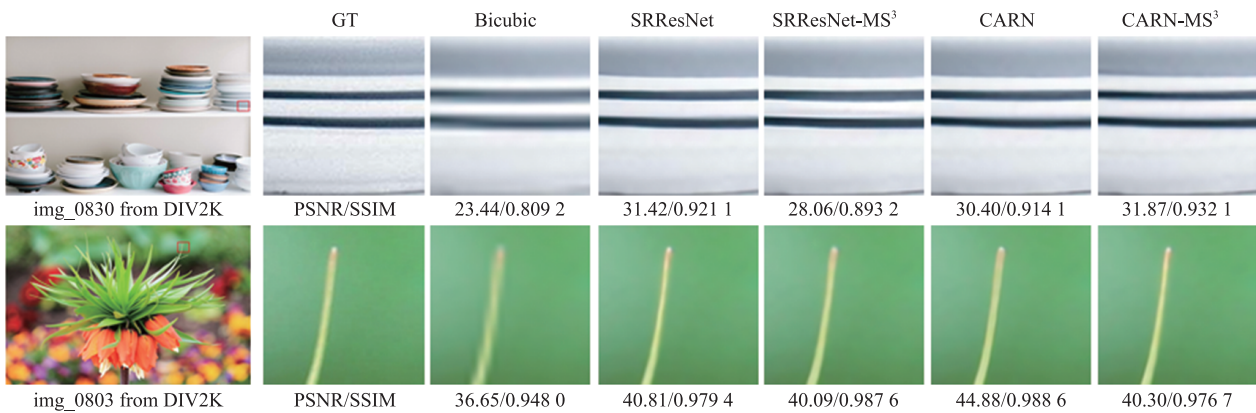


图 8 平坦区域的定性比较

Fig. 8 Qualitative comparisons on flat region

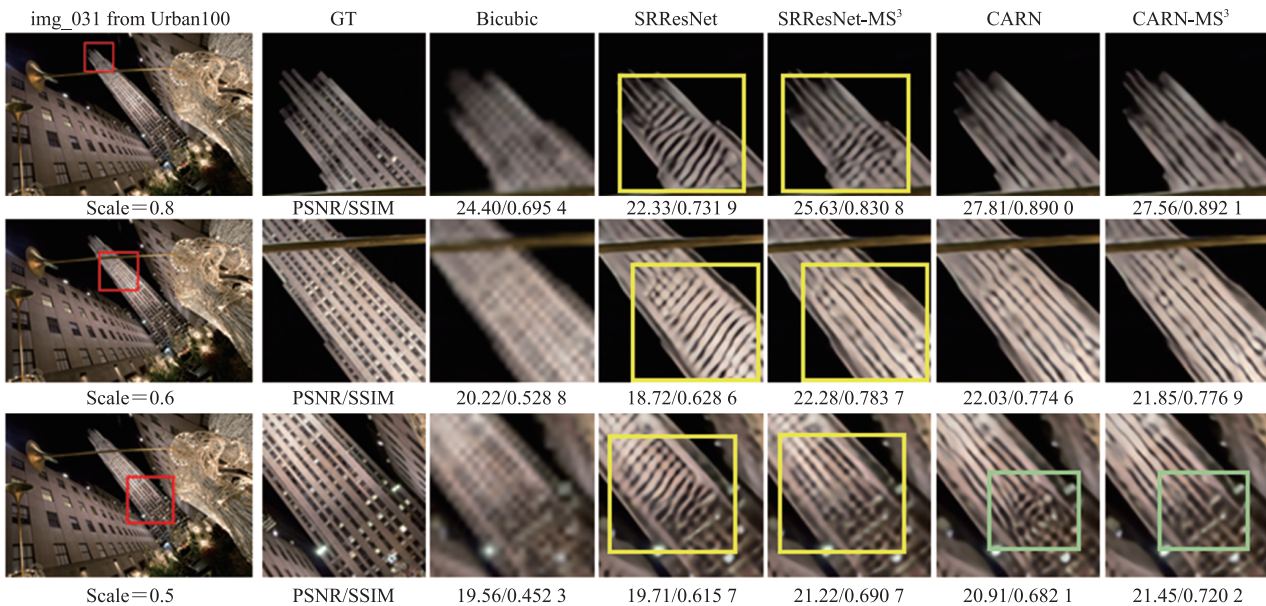


图 9 缩放图像的定性比较

Fig. 9 Qualitative comparisons on rescaling images

6 局限性讨论

在设计网络模型时,关于平衡网络性能与复杂度的灵活性方面,由于本文提出的多尺度卷积层网络使用了权重共享策略,所以灵活性稍差于 Octave 网络。在结果可视化方面,与标准卷积网络相比,基于本文提出的多尺度和权重共享卷积层网络的重构图像在高频细节方面都有较好的重建效果,能够减轻模糊现象且趋于恢复出正确的结构。然而,在平坦区域和平滑边缘,虽然基于多尺度卷积网络的重建性能指标值 PSNR 远高于基于标准卷积的网络,但两者并没有显著的视觉感知差异。这也是未来工作的重点,将进一步探究多尺度和权重共享卷积对不同区域的重构特点,探究如何减少平坦区域中的恢复错误的伪影现象,以此获得更好的视觉感知效果。本文对其他退化类型的数据以及真实场景下的数据并没有进行充分的测试与评估,故在未来的工作中,将进一步测试并评估基于本文提出的多尺度卷积网络对其他退化类型图像和真实场景图像的重建性能。

7 结论

本文深入研究基于超分辨率的多尺度特征交互传播,并提出了一种基于多尺度和权重共享卷积层的图像超分辨率方法,在相同尺度的卷积核之间,采用权重共享策略来减少参数量,使用不同尺度的信息经过相同的非线性映射实现不同尺度特征的信息交互融合,提取不同尺度特征,进而实现更好的图像重建性能。并通过大量的对比实验表明,本文提出的多尺度卷积能够解决网络参数量过大以及计算量大的技术问题,而且具有较强的灵活性和适配性,可以“即插即用”地应用到现有的网络结构中。

参 考 文 献

- [1] Dong C, Loy CC, He KM, et al. Learning a deep convolutional network for image super-resolution [C] // Proceedings of the European Conference on Computer Vision, 2014: 184-199.
- [2] Dong C, Loy CC, He KM, et al. Image super-resolution using deep convolutional networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(2): 295-307.
- [3] Lai WS, Huang JB, Ahuja N, et al. Deep laplacian pyramid networks for fast and accurate super-resolution [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 624-632.
- [4] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [C] // Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, 2015: 234-241.
- [5] Lin TY, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [6] Zhao HS, Shi JP, Qi XJ, et al. Pyramid scene parsing network [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881-2890.
- [7] Dang LW, Nie YW, Long CJ, et al. MSR-GCN: multi-scale residual graph convolution networks for human motion prediction [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 11467-11476.
- [8] Huang G, Chen DL, Li TH, et al. Multi-scale dense networks for resource efficient image classification [Z/OL]. arXiv Preprint, arXiv: 1703.09844, 2017.

- [9] Sun K, Zhao Y, Jiang BR, et al. High-resolution representations for labeling pixels and regions [Z/OL]. arXiv Preprint, arXiv: 1904.04514, 2019.
- [10] Newell A, Yang K, Deng J. Stacked hourglass networks for human pose estimation [C] // Proceedings of the European Conference on Computer Vision, 2016: 483-499.
- [11] Chen CF, Fan QF, Mallinar N, et al. Big-Little Net: an efficient multi-scale feature representation for visual and speech recognition [Z/OL]. arXiv Preprint, arXiv: 1807.03848, 2018.
- [12] Wang HY, Kembhavi A, Farhadi A, et al. Elastic: improving CNNs with dynamic scaling policies [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 2258-2267.
- [13] Liu LH, Hu XW, Zhu L, et al. Probabilistic multilayer regularization network for unsupervised 3D brain image registration [C] // Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, 2019: 346-354.
- [14] Zhang ZF, Wang ZW, Lin Z, et al. Image super-resolution by neural texture transfer [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 7982-7991.
- [15] Ke TW, Maire M, Yu SX. Multigrid neural architectures [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6665-6673.
- [16] Chen Y, Fan H, Xu B, et al. Drop an Octave: reducing spatial redundancy in convolutional neural networks with Octave Convolution [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3435-3444.
- [17] Fan HQ, Xiong B, Mangalam K, et al. Multiscale Vision Transformers [Z/OL]. arXiv Preprint, arXiv: 2104.11227, 2021.
- [18] Zhang H, Zu KK, Lu J, et al. Epsanet: an efficient Pyramid Split Attention block on convolutional neural network [Z/OL]. arXiv Preprint, arXiv: 2105.14447, 2021.
- [19] Sagar A. DMSANet: Dual Multi Scale Attention Network [Z/OL]. arXiv Preprint, arXiv: 2106.08382, 2021.
- [20] Yang FZ, Yang H, Fu JL, et al. Learning texture Transformer Network for image super-resolution [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 5791-5800.
- [21] Zhang YL, Li KP, Li K, et al. Image super-resolution using very deep residual channel attention networks [C] // Proceedings of the European Conference on Computer Vision, 2018: 286-301.
- [22] Cao ZJ, Long MS, Wang JM, et al. Partial transfer learning with selective adversarial networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2724-2732.
- [23] Mei YQ, Fan YC, Zhou YQ, et al. Image super-resolution with Cross-Scale Non-Local attention and exhaustive self-exemplars mining [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 5690-5699.
- [24] Li YH, Chen YT, Wang NY, et al. Scale-aware trident networks for object detection [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6054-6063.
- [25] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions [Z/OL]. arXiv Preprint, arXiv:

- 1511.07122, 2015.
- [26] Ledig C, Theis L, Huszar F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4681-4690.
- [27] Paszke A, Gross S, Chintala S, et al. Automatic differentiation in Pytorch [C] // Proceedings of the 31st Conference on Neural Information Processing Systems, 2017.
- [28] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017: 136-144.
- [29] Agustsson E, Timofte R. Ntire 2017 challenge on single image super-resolution: dataset and study [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017: 126-135.
- [30] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding [C] // Proceedings of the British Machine Vision Conference, 2012, <http://dx.doi.org/10.5244/C.26.135>.
- [31] Yang JC, Wright J, Huang TS, et al. Image super-resolution via sparse representation [J]. IEEE Transactions on Image Processing, 2010, 19(11): 2861-2873.
- [32] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics [C] // Proceedings of the Eighth IEEE International Conference on Computer Vision, 2001: 416-423.
- [33] Huang JB, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5197-5206.
- [34] Kinga DP, Ba JL. A method for stochastic optimization [Z/OL]. arXiv Preprint, arXiv: 1412.6980, 2015.
- [35] Ahn N, Kang B, Sohn K. Fast, accurate, and lightweight super-resolution with cascading residual network [C] // Proceedings of the European Conference on Computer Vision, 2018: 252-268.