基于深度学习的单幅图片超分辨率重构研究进展

张 宁[1, 2] 王永成[1] 张 欣[1, 2] 徐东东[1, 2]

摘 要 图像超分辨率重构技术是一种以一幅或同一场景中的多幅低分辨率图像为输入, 结合图像的先验知识重构出一幅高分辨率图像的技术. 这一技术能够在不改变现有硬件设备的前提下, 有效提高图像分辨率. 深度学习近年来在图像领域发展迅猛, 它的引入为单幅图片超分辨率重构带来了新的发展前景. 本文主要对当前基于深度学习的单幅图片超分辨率重构方法的研究现状和发展趋势进行总结梳理: 首先根据不同的网络基础对十几种基于深度学习的单幅图片超分辨率重构的网络模型进行分类介绍, 分析这些模型在网络结构、输入信息、损失函数、放大因子以及评价指标等方面的差异; 然后给出它们的实验结果, 并对实验结果及存在的问题进行总结与分析; 最后给出基于深度学习的单幅图片超分辨率重构方法的未来发展方向和存在的挑战.

关键词 深度学习, 单幅图片超分辨率, 卷积神经网络, 生成对抗网络

引用格式 张宁, 王永成, 张欣, 徐东东. 基于深度学习的单幅图片超分辨率重构研究进展. 自动化学报, 2020, **46**(12): 2479−2499

DOI 10.16383/j.aas.c190031

A Review of Single Image Super-resolution Based on Deep Learning

ZHANG Ning[1, 2] WANG Yong-Cheng[1] ZHANG Xin[1, 2] XU Dong-Dong[1, 2]

Abstract Super-resolution (SR) refers to an estimation of high resolution (HR) image from one or more low resolution (LR) observations of the same scene, usually employing digital image processing and machine learning techniques. This technique can effectively improve image resolution without upgrading hardware devices. In recent years, deep learning has developed rapidly in the image field, and it has brought promising prospects for single-image super-resolution (SISR). This paper summarizes the research status and development tendency of the current SISR methods based on deep learning. First, we introduce a series of networks characteristics for SISR, and analysis of these networks in the structure, input, loss function, scale factors and evaluation criterion are given. Then according to the experimental results, we discuss the existing problems and solutions. Finally, the future development and challenges of the SISR methods based on deep learning are presented.

Key words Deep learning, single-image super-resolution (SISR), convolutional neural network (CNN), generative adversarial network

Citation Zhang Ning, Wang Yong-Cheng, Zhang Xin, Xu Dong-Dong. A review of single image super-resolution based on deep learning. *Acta Automatica Sinica*, 2020, **46**(12): 2479−2499

图像分辨率是衡量图像质量的一项重要指标. 高分辨率图像拥有更高的像素密度, 更多的细节信息, 不但能大大提高人们的视觉体验, 还能对其包含的信息进一步挖掘和利用. 然而, 由于信号传输带宽及成像传感器等的限制, 成像设备获取到的图像通常具有较低的分辨率. 成像设备精度的提高, 图像尺寸的缩小, 单位面积内像素数量的增加, 都受到了当前制造水平的制约, 从硬件上提高其分辨率较为困难[1], 且耗时较长, 成本极高; 同时, 成像系统会受到噪声、模糊等干扰, 获取的图像质量下降, 分辨率受损. 图像超分辨率 (Image super-resolution, SR) 重构技术是指利用一幅或者多幅低分辨率图片通过软件处理生成一幅具有较高分辨率图像的技术. 采用图像超分辨率重构技术提高图像分辨率具有成本低、周期短等优点, 因此成为图像领域的一个研究热点. 图像超分辨率重构技术自提出以来, 备受学者关注, 其技术发展大致可分为三个阶段: 20 世纪 60 年代, Harris[2] 和 Goodman[3] 首次提出 "图像超分辨率"这一概念, 但是由于当时没有实际方法实现, 超分辨率重构只停留在理论研究阶段; 直至 1984 年, Tsai 等[4] 利用多幅低分辨率图片通过傅里叶变换域处理获得了一幅高分辨率图像, 这

收稿日期 2019-01-10 录用日期 2019-06-06
Manuscript received January 10, 2019; accepted June 6, 2019
国家自然科学基金 (11703027) 资助
Supported by National Natural Science Foundation of China (11703027)
本文责任编委 刘青山
Recommended by Associate Editor LIU Qing-Shan
1. 中国科学院长春光学精密机械与物理研究所 长春 130033 2. 中国科学院大学 北京 100049
1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033 2. University of Chinese Academy of Sciences, Beijing 100049

是利用软件技术获取重构图像的思想首次成功应用于实践, 自此开启了图像超分辨率重构的实践发展阶段; 2014 年, Dong 等[5] 提出 SRCNN (Super-resolution convolutional neural network) 结构, 将深度学习引入单幅图片超分辨率重构领域, 为图像超分辨率重构技术的发展开辟了新的研究思路. 本文结构安排如下: 第 1 节总体介绍图像超分辨率重构技术与单幅超分辨率重构方法; 第 2 节具体介绍基于深度学习的单幅图片超分辨率重构的网络模型; 第 3 节对第 2 节提出的模型进行了分析与总结; 第 4 节给出结论与展望.

1 单幅图片超分辨率重构技术与方法

图像超分辨率重构根据输入信息不同可以分为单幅图片超分辨率 (Single image super-resolution, SISR) 重构和多幅图片超分辨率 (Multi-frame super-resolution) 重构; 根据处理域不同可分为频域法和空域法, 空域法中根据实现方法又可分为基于插值、重构和学习三种方法[6], 上述重构方法之间的关系如图 1 所示.

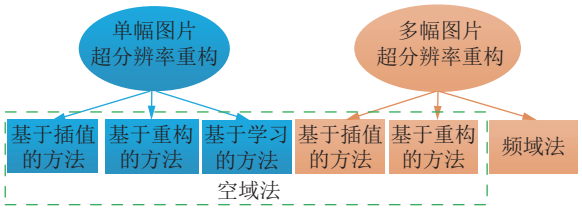


图 1 图像超分辨率重构方法分类

Fig. 1 Taxonomy of SR techniques

由于获取多幅包含亚像素信息的低分辨率图片较为困难, 对于单幅图片超分辨率重构的研究尤为重要. 目前常用的单幅图片超分辨率重构方法主要有插值法、重构法和学习法.

插值法如最近邻插值, 双线性插值, 双三次插值法, 它们操作简单, 易于实现, 但是插值的过程中只是简单地利用了待插值像素点附近的有限多个像素点信息, 单纯增加了图像的像素个数, 并不能增强图像的高频细节信息, 获取的高分辨率图像通常会出现块效应、模糊效应及混叠效应等[7].

重构法是学习法出现之前主流的图像超分辨率重构方法, 包括迭代反投影法[8]、KK[9]、全变分正则法[10]、解卷积法[11] 等. 基于重构的单幅图片超分辨率重构方法通常需要明确的先验信息对重构结果进行约束[12], 例如噪声扰动的形式, 能量函数等; 或者进行迭代计算来逼近原始高分辨率图像. 因此, 基于重构的方法通常计算量大, 求解困难, 耗时较长.

学习法可以分为基于浅层学习[13] 与基于深度学习两种. 浅层学习方法主要是根据经典的机器学习算法衍生的. 经典的机器学习是利用底层算法获取数据的一部分特征, 通常采用手动或半自动的方式选择特征并进行参数调整, 耗时耗力且需要对相应领域有专门的知识了解. 基于浅层学习的超分辨率重构方法利用机器学习算法局部地估计输出高分辨率图片的细节[12], 如基于像素的统计学方法[14−15], 基于例子的方法包括最近邻嵌入[16]、稀疏表示[17] 等. 深度学习属于机器学习范畴, 但其近年来受到的关注要远远高于经典机器学习方法. 它通过深层网络自动地学习输入数据的抽象特征, 并且通过反向传播算法来调整网络参数. 与此同时, 深度学习可以利用加深或者加宽网络结构学习更为复杂的映射关系以及处理大量或者高维的数据.

2017 年 NTIRE (New trends in image restoration and enhancement workshop)[18] 图像超分辨率挑战大赛上, 完成挑战的 20 支参赛队伍中, 只有一支队伍没有采用深度学习策略, 他们取得的重构效果并不理想, 且单幅图片重构时间远远多于其他基于深度学习的方法. 这一结果表明, 基于深度学习的图像超分辨率算法已成为单幅图片超分辨率重构主流算法, 也是未来图像超分辨率技术发展的必然趋势.

文献 [19] 对于传统图像超分辨率重构算法进行了全面、细致的总结, 文献 [20] 对于除深度学习以外的单幅图片超分辨率重构技术进行了总结. 本文主要对自 SRCNN 模型提出以来的基于深度学习的单幅图片超分辨率重构算法进行介绍, 分析这些模型在网络结构、输入信息、损失函数、放大因子以及评价指标等方面的差异, 给出它们的实验结果, 并对实验结果及存在的问题进行分析与总结, 最终根据分析结果给出基于深度学习的单幅图片超分辨率重构技术的发展趋势.

2 基于深度学习的单幅图片超分辨率重构方法

深度学习在图像领域的发展过程中具有重要作用. 随着神经网络在图像分类领域的发展, 图像超分辨率重构方法也受经典的神经网络模型启发并取得了重大进步. 这些方法以标准的卷积神经网络、残差网络 ResNet[21]、生成对抗网络 (Generative adversarial network, GAN)[22] 及其他网络结构为基础, 并针对图像超分辨率重构的特定需求, 发展形成了一系列重构效果优异的网络模型. 图 2 根据不同基础网络结构, 给出了本节介绍的基于深度学习的单幅图片超分辨率重构方法的分类图.

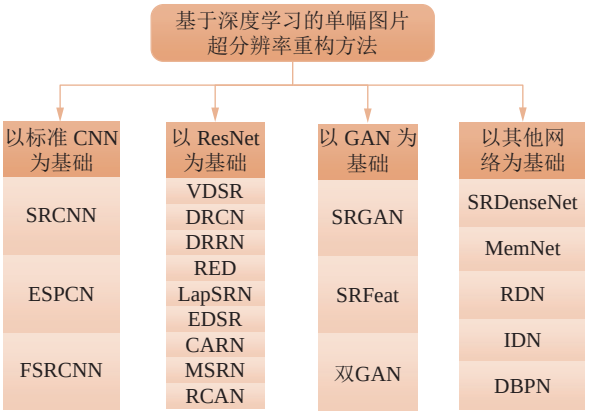


图 2　SISR 模型分类图

Fig. 2　Taxonomy of SISR methods

2012 年, AlexNet[23] 的引入使得图像分类效果有了大幅提升, 同时掀起了深度学习的浪潮, 超分辨率重构技术引入神经网络之初, 即采用了标准的卷积神经网络, 构建了 SRCNN 等经典网络模型; 随着卷积神经网络的发展, ResNet、DenseNet[24] 等网络模型以及跳跃连接、递归监督等策略解决了深层网络的难以训练的问题, 超分辨率重构网络模型也随之过渡到了深度神经网络模型阶段; GAN 模型的提出为图片生成提供了思路, 同时也为生成高分辨率图像提供了模型基础, 将 GAN 应用至图像超分辨率重构技术中, 获得了较高的图像视觉质量, 成为了基于深度学习的单幅图片超分辨率重构的重要组成部分; 同时, 一些模型采用深度学习策略如密集连接、记忆机制、蒸馏机制、迭代投影等, 以实现更好的重构效果. 接下来对这些模型进行具体介绍.

2.1　基于标准卷积神经网络结构的 SR 模型

2.1.1　SRCNN 模型

香港中文大学的 Dong 等[25] 首次将卷积神经网络引入图像超分辨率, 构建了 SRCNN 网络模型. 该网络仅由三层构成, 分为特征提取层、非线性映射层以及重构层, 如图 3 所示. 利用一层卷积网络, 将输入的低分辨率 (Low resolution, LR) 图像中的 n_1 个特征提取出来, 通过非线性映射层将 n_1 个特征映射到 n_2 个特征的特征空间, 作为高分辨率 (High resolution, HR) 图像的特征, 最后通过重构层利用 n_2 个特征重构出 HR 图像.

SRCNN 网络结构简单, 作者通过与当时流行的稀疏表示[17] 方法进行对比, 获得二者之间的差别与联系, SRCNN 取得的优于所有非深度学习方法的效果, 充分体现了深度学习在单幅图片超分辨率重构领域的前景.

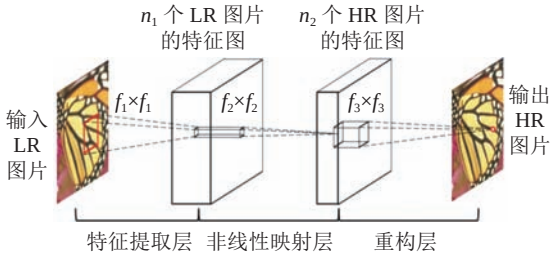


图 3　SRCNN 网络模型图

Fig. 3　The network structure of SRCNN

同时, Dong 等[25] 还从网络结构、输入数据以及评价指标上对 SRCNN 模型进行了分析. 网络结构包括卷积核数目、卷积核大小、网络深度; 输入数据从不同的彩色空间出发, 分析彩色空间不同、处理通道不同以及是否进行预训练三个方面对重构效果的影响, 最终得出只对 YC_bC_r 空间中的 Y 通道操作获得的效果最佳的结论, 并成为后续研究的基准实验条件; 评价指标考察了峰值信噪比 (Peak signal-to-noise ratio, PSNR)、结构相似度 (Structural similarity, SSIM) 以及其他指标, 并以两者为主要评价指标对重构结果进行了评价.

SRCNN 虽然取得了优于传统方法的效果, 但是它也存在着不足之处. 首先, 输入为经过插值后的 LR 图像, 不仅增加了计算量, 更会引入不必要的噪声与重影, 从而影响重构效果; 其次, 该网络并没有随着网络深度增加而获得重构效果的提升, 模型具有很大的改进空间; 最后, 该模型收敛速度较慢, 且训练同一个模型只针对一个尺度放大因子, 要实现×2、×3、×4 不同放大倍数需要不同的训练过程.

针对 SRCNN 模型存在的问题, Dong 等[26] 对其进行改进, 提出 Fast-SRCNN (FSRCNN) 模型. FSRCNN 首先引入解卷积层, 以解决 SRCNN 输入经过插值的 LR 图像的问题, 减少了计算量和输入误差且能够通过改变解卷积层实现不同尺度放大; 同时采用收缩策略, 将特征维度降低; 最后采用小卷积核, 减少计算量, 加深网络. FSRCNN 比 SRCNN 提高了近 40 倍训练速度并且重构效果更优.

尽管 SRCNN 网络模型存在不足之处, 它的提出仍然可以视为单幅图片超分辨率重构的方法的里程碑. 它的模型简单, 效果优于传统方法, 虽然研究者们又提出了很多新的网络模型, 但是 SRCNN 仍旧作为优秀的基准实验, 指导着基于深度学习的单幅图片超分辨率重构方法的改进, 并且为具体的应用提供了借鉴方法. 如 Luo 等[27] 通过改进 SRCNN 网络模型, 对卫星视频序列进行超分辨率重构, 并

取得了良好的效果; Ducournau 等[28] 使用中间层参数微调后的 SRCNN 网络模型对卫星获取的海表面温度 (Sea surface temperature, SST) 图进行超分辨率重构, 以便用于后续数据处理; Rasti 等[29] 利用 SRCNN提高人脸分辨率后提升了识别准确率; Zhang 等[30]将 SRCNN 应用于提高红外热成像图像分辨率等. 这些具体应用体现出 SRCNN 的提出具有重要意义与实用价值.

2.1.2 ESPCN 模型

Twitter 的 Shi 等[31] 提出了一种基于像素重排的 ESPCN (Efficient sub-pixel convolutional neural network) 网络模型, 该模型的核心为亚像素卷积, 即通过卷积后得到的特征 (Feature maps) 进行像素排列, 以填充原始低分辨率图像的亚像素信息. 网络结构如图 4 所示.

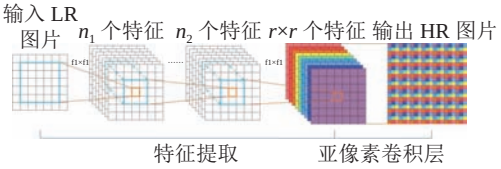


图 4 ESPCN 网络模型图

Fig. 4 The network structure of ESPCN

整个网络由特征提取部分和亚像素卷积部分构成, 特征提取部分通过不同的卷积层提取输入大小为 $W \times H \times C$ 的低分辨率图片的特征, 其中, W 和 H 为低分辨率图片的尺寸, C 为通道数, 最后一层卷积层输出的特征为 $r \times r$ 个, r 代表尺度放大因子; 亚像素卷积层是将 $r \times r$ 个通道的特征进行像素重排, 从而得到输出大小为 $rW \times rH \times C$ 的高分辨率图片.

与 FSRCNN 使用的解卷积层不同, ESPCN 模型的提出提供了一种新的图像重构策略, 这一策略与图像插值十分类似, 因此该模型可以看作通过卷积神经网络学习由低分辨率图像到高分辨率图像之间的插值函数, 插值方式不是简单地利用邻域像素计算, 所以重构效果远远优于插值法. 同时, 通过设置特征通道数目, 能够实现 $\times 2$、$\times 3$、$\times 4$ 不同尺度因子放大. 这一结果不仅证明了深度学习中的卷积神经网络可以学习任何复杂的映射关系, 也为图像超分辨率重构提供了一种可以灵活调整放大倍数的重构策略, 具有重要的意义.

2.1.3 小结

第 2.1 节介绍了三种以标准卷积神经网络构建的单幅图片超分辨率重构网络模型, 这三种网络结构简单并且取得了优良的重构效果. 表 1 从输入图像类型、网络层数、损失函数、评价指标以及放大因子这些方面对三种网络模型进行了对比总结.

表 1 三种网络模型对比

Table 1 Comparison of the above three models

网络模型	输入图像	网络层数	损失函数	评价指标	放大因子
SRCNN	ILR	3	L_2 范数	PSNR, SSIM, IFC	2, 3, 4
FSRCNN	LR	$8+m$	L_2 范数	PSNR, SSIM	2, 3, 4
ESPCN	LR	4	L_2 范数	PSNR, SSIM	2, 3, 4

表 1 中, ILR 表示插值 (Interpolated) 后的 LR 图片, m 为可调参数. SRCNN 模型首次将深度学习引入图像超分辨率重构领域, 通过多种结构调整及参数设置实验, 分析了各种网络参数以及输入数据对重构效果的影响, 其结论具有很强的参考性, 如输入图片的通道选取, 评价指标的选取; 但是它的一些结论也有待分析, 如卷积核大小、网络深度的影响等. 基于这些结论, 作者又给出了改进版本 FSRCNN 模型. 而 ESPCN 模型同样具有简单的网络结构, 其核心思想亚像素卷积层为高分辨率图像的重构做出了巨大贡献, 当前大部分基于深度学习的单幅图片超分辨率重构方法网络中的图像重构均依赖于亚像素卷积和解卷积两种方法[32], 可见其深刻的影响.

2.2 基于残差网络结构的 SR 模型

2.2.1 VDSR 模型

首尔国立大学计算机视觉实验室的 Kim 等[33] 首先将学习残差的思想引入超分辨率重构方法中, 采用 3×3 的小卷积核构建了网络层数深达 20 层的 VDSR (Very deep convolutional networks for super-resolution) 模型, 增加了感受野. 网络结构如图 5 所示.

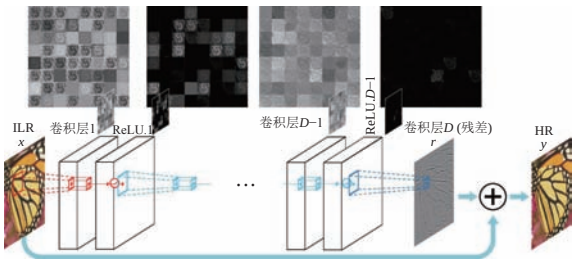


图 5 VDSR 网络模型图

Fig. 5 The network structure of VDSR

输入插值后的低分辨率图像, 网络学习其与高分辨率图像之间的残差, 最后在重建层加入一个跳跃连接, 利用低分辨率图片与残差重构出高分辨率图片.

VDSR 模型引入残差学习的思想主要有两方面原因: 首先, 网络加深存在梯度消失/爆炸的问题, 而学习残差可以减轻深度网络训练中此类问题带来的影响; 更重要地, 低分辨率图像与高分辨率图像之间具有大量相似的低频信息, 而网络学习残差能够避免重复学习低频信息, 只学习高频细节信息, 从而降低了计算成本, 加快了收敛速度, 节省了运算时间.

然而 VDSR 模型仅在整个网络中加入一个跳跃连接, 并没有很有效地减轻梯度消失的问题, 因此在训练过程中, 使用较高的学习率加快学习速度使网络尽可能收敛, 然后采用自适应梯度裁剪策略解决梯度消失/爆炸问题.

VDSR 模型是图像超分辨率领域首个深层网络, 通过采用多层小卷积核, 既能够减少参数量, 又能够加深网络, 增加感受野, 学习更好的特征. 它的重构效果优于 SRCNN 模型, 证明了随着网络加深, 模型具有更好的表达能力, 与深层次网络在图像其他领域的应用结论一致.

2.2.2　DRCN 模型

DRCN (Deeply-recursive convolutional network) 模型是 Kim 等[34] 于 VDSR 模型提出同年又提出的一个深度超分辨率重构网络模型. DRCN 模型的核心思想是递归训练, 如图 6 所示.

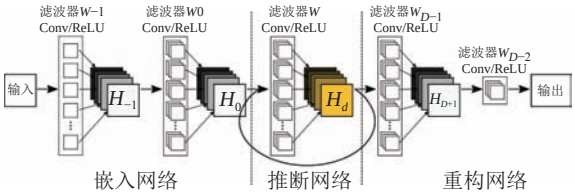


图 6　DRCN 网络模型图

Fig. 6　The network structure of DRCN

网络分为三个部分, 嵌入网络、推断网络和重构网络. 嵌入网络的目的是将低分辨率图片的浅层特征提取出来作为递归网络的输入; 推断网络由 16 个递归单元组成, 数据循环地通过该单元多次, 每个单元具有相同的参数, 等效于同一组参数网络的串联, 但是递归神经网络容易过拟合或者梯度消失/爆炸, 因此在递归单元与重构网络之间加入了跳跃连接, 同时采用递归监督策略; 重构网络接收嵌入网络以及每一个递归单元的输入, 加权平均后得到重构图片, 如图 7 所示.

具体地, H_1 和 H_D 是 D 个共享参数的卷积层, 将 D 个卷积层的每一层结果都通过相同的重构网络, 在重构网络中通过跳跃连接与原始图像相加, 得到 D 个输出重构结果, 这些结果在训练的同时

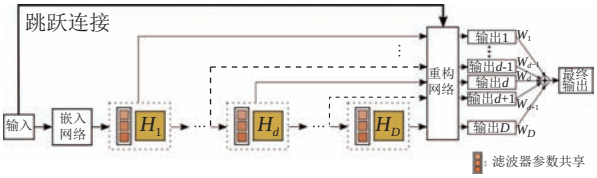


图 7　DRCN 模型中的跳跃连接和递归监督

Fig. 7　The skip-connection and recursive-supervision in DRCN

被监督, 避免梯度消失/爆炸的问题. 最后将 D 个重构结果进行加权平均, 得到最终输出的高分辨率图片.

DRCN 网络模型的特征包括: 引入了递归单元, 加深了网络结构; 每个递归单元之间参数共享, 减少了参数量; 加入跳跃连接不仅利用了从浅层到深层的特征, 还能够对其进行监督训练, 解决梯度消失/爆炸的问题. 由重构结果可知, 这一网络模型尤其对具有纹理的图像重构效果更好, 充分体现了其重构有效性.

2.2.3　DRRN 模型

DRRN (Deep recursive residual network) 是南京理工大学的 Tai 等[35] 受 DRCN 模型的启发提出的. 同样采用递归单元, 但是 DRRN 采用了嵌套结构, 如图 8 所示.

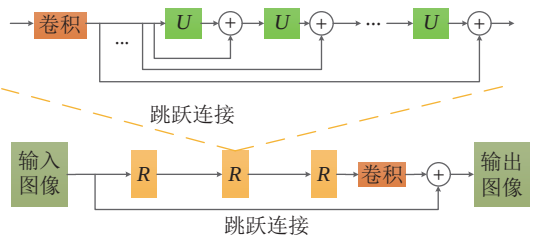
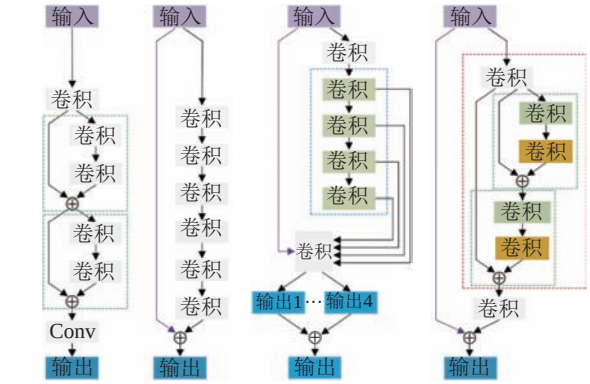


图 8　DRRN 网络模型图

Fig. 8　The network structure of DRRN

整个网络由 3 个 (可调整数目) 递归单元 R 构成, 每个递归单元中又嵌套 8 个 (仍旧可调整数目) 子递归单元 U, 参数共享且加入局部跳跃连接, 最外层又加入一个全局跳跃连接. 每个子递归单元由两个卷积层 (包括卷积、BN、ReLU) 构成, 网络深度达到 52 层.

VDSR, DRCN, DRRN 三种模型之间既有区别又有联系. 首先三者均以残差网络为原型发展而来, 利用残差学习和跳跃连接的思想, 既能进行稀疏残差学习, 减少参数量, 又能加深网络, 解决梯度消失/爆炸问题; 更重要地, 由三者之间网络结构对比可以发现, VDSR 和 DRCN 模型是 DRRN 模型的特例, 当 DRRN 模型的子递归单元数目 U 为 0 时, 就与 VDSR模型等价, 如图 9 所示.



(a) 残差网络 (b) VDSR 模型 (c) DRCN 模型 (d) DRRN 模型
(a) ResNet (b) VDSR (c) DRCN (d) DRRN

图 9 四种网络模型对比
Fig. 9 The comparison of the above four models

以 ResNet 为基础, VDSR 模型基于全局残差学习, DRCN 模型既包括全局残差学习, 又包括单一权重递归的局部残差学习, 而 DRRN 模型则为全局残差学习与多权重递归的局部残差学习相结合. 这几种网络模型均通过增加网络深度取得了良好的重构效果.

2.2.4 RED 模型

南京大学的 Mao 等[36] 提出了一种对称的编解码网络结构 RED (Residual encoder-decoder networks) 用于图像复原技术, 包括单幅图片超分辨率重构. 该网络主要包含两个部分: 卷积部分与解卷积部分. 卷积层用于提取低分辨率图片的抽象特征, 解卷积层用于放大特征尺寸并且恢复图像细节. 两部分之间隔层通过一个跳跃连接相连, 密集的跳跃连接既能够传递提取的特征信息, 又可以缓解梯度消失/爆炸的问题. 网络结构如图 10 所示.

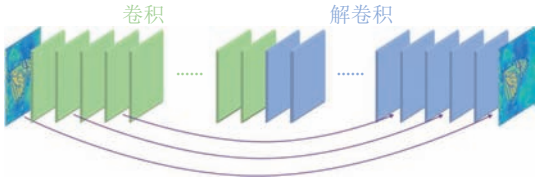


图 10 RED 网络模型图
Fig. 10 The network structure of RED

RED 网络不仅可以处理单幅图片超分辨率重构, 还能进行图像去噪, JPEG 图像去块以及图像修复, 是一个综合应用的网络, 且无论是重构效果还是其他方面的应用都取得了当时最好的结果.

2.2.5 LapSRN 模型

加州大学默赛德分校的 Lai 等[37] 提出了一种结合金字塔结构进行单幅图片超分辨率重构的 LapSRN (Laplacian pyramid super-resolution network) 模型, 该模型能够逐步重构不同放大倍数的高分辨率图片, 不仅能够完成×2 和×4 倍数的超分辨率重构, 而且在×8 放大因子取得了良好的效果, 网络结构如图 11 所示. 整个网络特征提取过程与图像重构过程共同进行, 首先输入 LR 图像经过一系列卷积层进行特征提取, 学习的是残差, 然后经过一层解卷积进行尺度放大, 得到×2 倍数的残差图片, 并与经过同倍数插值的原始图片相加, 得到×2 倍数的高分辨率重构图像; 特征提取模块继续进行残差学习, 经过解卷积层后再次放大两倍, 得到×4 倍数的残差图片, 同时图像重构模块由重构出的超分辨率图像继续插值, 二者相加得到×4 倍数的高分辨率重构图像. 以此类推, 可以得到放大倍数为 2^n 的重构图片.

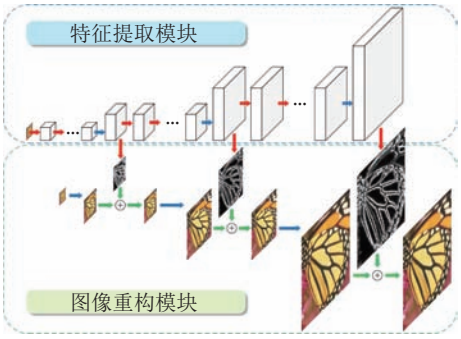


图 11 LapSRN 网络模型图
Fig. 11 The network structure of LapSRN

同时, 该网络采用了一种有别于 L_2 损失的损失函数, Charbonnier 损失函数. 表达式为

$$L(\hat{y},y;\theta)=\frac{1}{N}\sum_{i=1}^{N}\sum_{s=1}^{L}\rho\left(\hat{y}_s^{(i)}-y_s^{(i)}\right)=\frac{1}{N}\sum_{i=1}^{N}\sum_{s=1}^{L}\rho\left(\left(\hat{y}_s^{(i)}-x_s^{(i)}\right)-r_s^{(i)}\right) \quad (1)$$

其中, $\hat{y}$ 表示重构图像, y 表示原始低分辨率图像, θ 为网络参数, N 为总像素数, s 为放大倍数, x 为插值后的低分辨率图像, r 为残差图像.

$$\rho(x)=\sqrt{x^2+\epsilon^2} \quad (2)$$

式 (2) 即为 Charbonnier 惩罚函数, 其中 ϵ 取 10^{-3}. 在 LapSRN 模型之前的模型采用 L_2 范数, 它能获取最小均方误差, 从而得到较高的 PSNR 值, 但是视觉上容易产生平滑效果. 而 Charbonnier 损失函数类似一种 L_1 范数, 采用 L_1 范数能够获取较为显著的细节信息.

LapSRN 模型能够处理不同尺度的超分辨率图

像, 尤其是在×8 放大因子能够取得良好的重构效果. 另一个重要贡献是采用了区别于 L_2 范数的 Charbonnier 损失函数, 取得了良好的视觉效果.

2.2.6　EDSR 模型

2017 年 NTIRE 图像超分辨率重构大赛中, 首尔大学 SNU CVLab[38] 团队获得了冠军, 他们提出了 EDSR (Enhanced deep residual networks for super-resolution) 模型. 该模型在残差网络的基础上进行了改进, 去掉了批归一化 (Batch normalization, BN) 层. BN 层用于解决高级图像问题如图像分类、检测等比较有效, 而图像超分辨率重构属于底层问题, 去掉 BN 层可以节省一半的内存和计算时间, 且重构效果良好. 结构如图 12 所示.

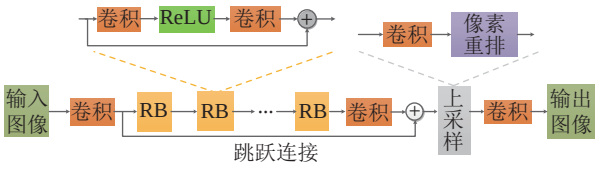


图 12　EDSR 网络模型图

Fig. 12　The network structure of EDSR

网络由一系列残差块 RB 级联而成, 每个残差块内部不包括 BN 层, 上采样过程通过像素重排实现, 且×4 倍数结果可由×2 倍数实现, 节省了一部分内存. 同时整个网络采用了 L_1 范数作为损失函数.

此外, EDSR 模型还提出了一种数据增强策略以提高重构效果, 称为 Geometric self-ensemble. 具体做法是对 LR 图片进行几何变换, 输入到网络中获取高分辨率图片, 再对其进行相应的逆变换, 所有结果进行平均作为最终的重构结果. 这种策略能够使得网络从不同的 “视角”学习高低分辨率图片之间的映射关系, 从而提升重构效果.

EDSR 模型在 2017 年 NTIRE 图像超分辨率重构大赛中不仅获得了最好的重构效果, 也具有较快的重构速度, 还提出了一种新的数据增强策略, 是当之无愧的冠军.

2.2.7　CARN 模型

随着深度学习的发展, 能够在移动端进行处理的轻量级网络逐步进入人们的视线. 既能获取相对好的重构效果, 又能构建参数少的轻量级网络成为单幅图片超分辨率重构技术的另一个发展方向. 为减少神经网络的参数量与计算复杂度, 各领域学者们致力于网络轻量化研究, 如分组卷积[23] 的引入以及 MobileNet[39] 的提出. 受此启发, 韩国亚洲大学的 Ahn 等[40] 提出了轻量级图片超分辨率重构神经网络模型 CARN (Cascading residual network). CARN 模型同样在 ResNet 网络基础上进行改进, 如图 13 所示.

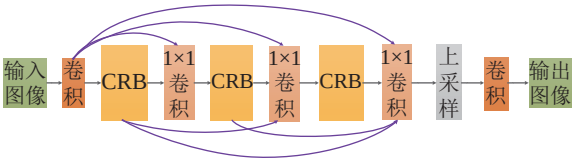


图 13　CARN 网络模型图

Fig. 13　The network structure of CARN

网络采用密集级联, 堆叠多个级联残差块 CRB (Cascading residual blocks), 并用 1×1 卷积核对特征进行融合, 分为 CARN 和 CARN-M 两种模型, 后者在 CRB 中采用分组卷积, 且每个 CRB 参数共享, 构成递归单元, 减少了参数, 从而形成了轻量级网络模型. CARN-M 模型与 CARN 相比参数量减少至将近 1/5, 但是其重构图像 PSNR 值仅比 CARN 模型降低 0.26～0.68 dB 左右. 这就实现了简化网络与保留重构效果之间的相对平衡.

2.2.8　MSRN 模型

2018 年 ECCV (European conference on computer vision) 会议上, 来自华东师范大学的 Li 等[32] 提出了 MSRN (Multi-scale residual network) 网络模型. 该模型主要具有以下两个特点: 1) 提出一种多尺度残差模块 MSRB (Multi-scale residual block); 2) 通过采用 ESPCN 模型中的通道像素重排方法实现不同放大倍数的重构图片. 网络模型如图 14 所示.

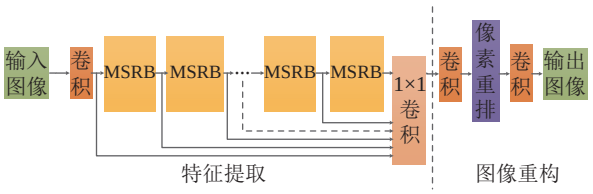


图 14　MSRN 网络模型图

Fig. 14　The network structure of MSRN

输入图像经过一层卷积层提取特征后, 输入一系列 MSRB 中, 每个 MSRB 的特征输出最终通过 1×1 卷积核进行特征降维与融合, 改变特征通道数目以用于图像重构. MSRB 是网络的核心, 它类似于 GoogLeNet[41] 中的 Inception 结构, 从不同的卷积核尺度提取不同的特征, 并且将这些特征融合, 如图 15 所示.

MSRN 模型指出, 其他很多模型的训练都或多或少采用了一定的训练技巧, 因此在训练复现的过

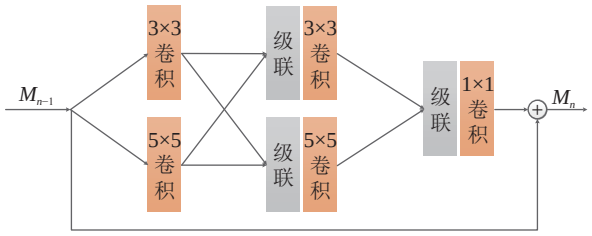


图 15 MSRB 结构

Fig. 15 The structure of multi-scale residual block

程中, 尽管参数设置相同, 却总是难以达到模型给定的参数指标. 所以 MSRN 模型在训练过程没有采用任何技巧, 只在基础网络结构上进行了改进; 通过 1×1 卷积实现通道数任意降维, 可以保证各种尺度的放大倍数; 数据增强只采用了尺度变换、水平翻转与旋转; 损失函数采用 L_1 范数, 这样的实验设置保证了复现时训练可以达到论文中给定的重构结果.

2.2.9 RCAN 模型

RCAN (Residual channel attention networks) 是美国东北大学的 Zhang 等[42] 提出的一个基于通道注意机制的网络模型. 网络模型如图 16 所示.

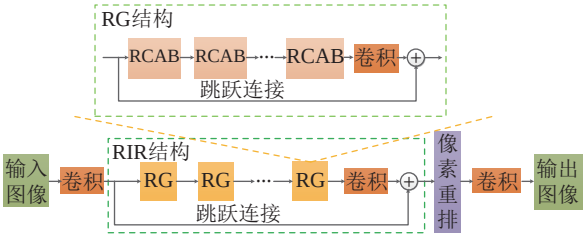


图 16 RCAN 网络模型图

Fig. 16 The network structure of RCAN

该模型的基本单元是 RIR (Residual in residual) 结构, RIR 结构由一系列残差模组 (Residual group, RG) 堆叠而成, RG 结构又由通道注意残差块 (Residual channel attention block, RCAB) 堆叠而成, 因此称为 Residual in residual结构.

Zhang 等[42] 指出, 经过神经网络不同卷积核提取的特征对于图像超分辨率重构的贡献不同, 因此每个通道特征所占的比重应当有所不同, 基于此提出了通道注意机制 CA(Channel attention), 如图 17 所示.

输入为 $H \times W \times C$, 经过一个全局池化 (Global pooling) 函数 H_{GP}, 输出为维度为 $1 \times 1 \times C$ 且具有全局特征的 z_c, 即

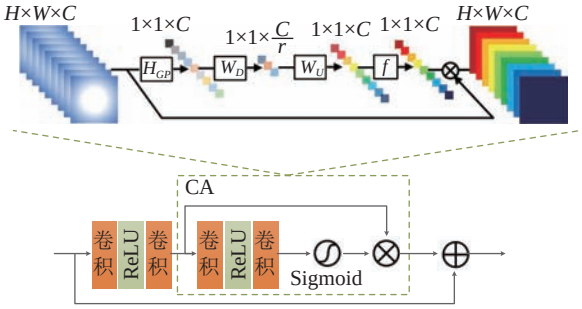


图 17 RCAB 结构图

Fig. 17 The structure of residual channel attention block

$$z_c = H_{GP}(x_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i,j) \tag{3}$$

然后经过由 ReLU 和 Sigmoid 单元构成的门控单元, 从而得到各个通道具有不同比重的通道系数 s, 即

$$s = f\left(W_U \delta\left(W_D z\right)\right) \tag{4}$$

最后利用该系数与原始输入信息进行通道对应相乘, 得到经过通道注意机制后的特征.

通道注意机制 CA 是一种对于提取特征进行选择的新方法, 这一思想可以采用不同的方法实现, 也可以应用于基于深度学习进行数据处理的其他领域, 具有重要的研究意义与应用价值.

2.2.10 小结

ResNet 的提出使得卷积神经网络能够构建的更深、学到的特征更多, 在图像分类中取得了令人瞩目的成果. 同样地, 在单幅图片超分辨率重构领域, 残差网络由于其可构建深度网络, 也受到许多相关学者的青睐. 第 2.2 节介绍了 9 种以残差网络为基础的单幅图片超分辨率重构网络模型, 这些模型各有优点, 充分利用了残差网络的优势并结合自己的创新点, 获得了重构效果上的极大的进步. 表 2 从输入图像类型、网络层数、损失函数、评价指标以及放大因子这些方面给出了这些模型的对比总结. 由表 2 可以看出随着时间的推移, 模型的发展也越来越有效, 输入图片由插值后的 LR 图像逐步变为原始 LR 图像, 减少了训练过程的计算量; 网络层数逐步加深, 网络结构也逐步精炼; 同时, 损失函数也逐渐由 L_2 范数演变为 L_1 范数, 因为 L_1 范数损失得到的重构结果细节更加突出, 更加符合人的视觉主观评价; 评价指标不再只由 PSNR 与 SSIM决定, 尤其是 ECCV18 会议开始利用重构后的图像进行分类, 通过评价分类效果间接对重构效果进行评价.

2.3 基于生成对抗网络结构的 SR 模型

2.3.1 SRGAN 模型

2014 年, Goodfellow 等[22] 提出了生成对抗网

表 2　基于残差网络的 9 种模型对比

Table 2　Comparison of the nine models based on ResNet

网络模型	输入图像	网络层数	损失函数	评价指标	放大因子
VDSR	ILR	20	L_2 范数	PSNR, SSIM	2, 3, 4
DRCN	ILR	16 (Recursions)	L_2 范数	PSNR, SSIM	2, 3, 4
DRRN	ILR	52	L_2 范数	PSNR, SSIM,IFC	2, 3, 4
RED	ILR	30	L_2 范数	PSNR, SSIM	2, 3, 4
LapSRN	LR	27	Charbonnier	PSNR, SSIM, IFC	2, 4, 8
EDSR	LR	32 (Blocks)	L_1 范数	PSNR, SSIM	2, 3, 4
CARN	LR	32	L_1 范数	PSNR, SSIM, 分类效果	2, 3, 4
MSRN	LR	8 (Blocks)	L_1 范数	PSNR, SSIM, 分类效果	2, 3, 4, 8
RCAN	LR	20 (Blocks)	L_1 范数	PSNR, SSIM, 分类效果	2, 3, 4, 8

络 (Generative adversarial network, GAN), GAN 的提出是图片生成技术的良好开端. 对于单幅图片超分辨率重构而言, 利用丢失了高频细节信息的低分辨率图像去推断原始高分辨率图像, 无异于一种图片细节生成技术, 因此, GAN 的提出也为图像超分辨率重构技术提供了一种思路.

Twitter 的 Ledig 等[43] 将生成对抗网络引入图像超分辨率重构领域, 提出了 SRGAN (Generative adversarial network for image super-resolution) 网络模型. 该模型包括生成器与判别器两个部分, 如图 18和图 19 所示.

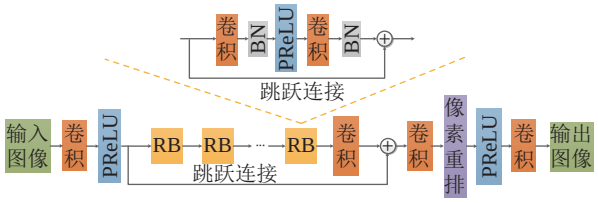


图 18　SRGAN 网络的生成器

Fig. 18　The generator of SRGAN

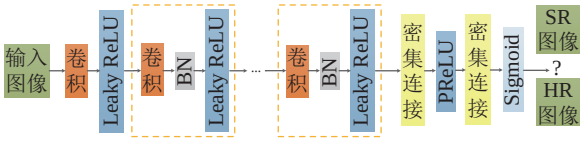


图 19　SRGAN 网络的判别器

Fig. 19　The discriminator of SRGAN

生成器以残差网络为基础, 为防止生成对抗网络训练崩溃, 对生成器进行了预训练. 预训练过程与普通模型相同, 此时损失函数为 VGG 损失, 它由生成器生成图片的特征表达与原始高分辨率图像的特征表达之间的欧氏距离定义, 即

$$L_{\mathrm{VGG}}=\frac{1}{W\times H}\sum_{i=1}^{W}\sum_{j=1}^{H}\left(\phi\left(I^{\mathrm{HR}}\right)-\phi\left(G_{\theta_G}\left(I^{\mathrm{LR}}\right)\right)\right) \tag{5}$$

其中, $W\times H$ 为图片尺寸, $\phi(\cdot)$ 为 VGG 函数, G_{θ_G} 为生成器. 预训练结束后, 将重构图片输入鉴别器, 然后加入对抗损失, 对生成器和鉴别器交替训练. 对抗损失为

$$L_{\mathrm{Gen}}=-\sum_{i=1}^{N}\log D_{\theta_D}\left(G_{\theta_G}\left(I^{\mathrm{LR}}\right)\right) \tag{6}$$

其中, D_{θ_D} 为鉴别器, N 表示像素总数. 整个网络的损失函数为 VGG 损失和对抗损失的加权和, 即

$$L=L_{\mathrm{VGG}}+10^{-3}L_{\mathrm{Gen}} \tag{7}$$

同时, SRGAN 模型提出了一种主观评价指标. 生成对抗网络的重构结果在评价指标 PSNR 值上并没有取得明显优势. PSNR 值受均方误差影响, 但是 SRGAN 的损失函数未采用均方误差, 且经过对抗训练后, 重构图片的细节更加突出, 符合人的视觉感知, 因此, SRGAN 模型提出了主观评价指标 MOS (Mean opinion score). 通过 26 位评价者对不同网络模型重构的结果进行打分评价, 分数从 1 到 5 代表由坏到好, 最终的 MOS 是由评估者评分的算术平均值来确定的. 经过主观评价, SRGAN 取得了最优的重构效果. 这一结果表明, PSNR 和 SSIM 两项评价指标对于图像质量的评价具有一定的局限性, 图像超分辨率重构效果不应只依赖于这两项指标的提升. 但是主观评价需要消耗大量的人力财力, 评价者易受主观思想的影响, 且需要具有专门领域知识的人进行评价, 因此图像质量评价的指标选取仍有很大的研究空间.

2.3.2　SRFeat 模型

SRFeat (Single image super-resolution with feature discrimination) 是 Park 等[44] 提出的一种基于生成对抗网络的单幅图片超分辨率重构模型, 该

网络的核心是增加了一个特征判别器, 不仅使重构的图片与高分辨率图片进行对抗训练, 还要对提取的特征与高分辨率图片的特征进行对抗训练. SRFeat 模型与 SRGAN 模型结构相似, 但是 SRFeat 模型的生成器网络中加入了 1×1 卷积, 对提取到的特征进行了融合与调整, 结构如图 20 所示.

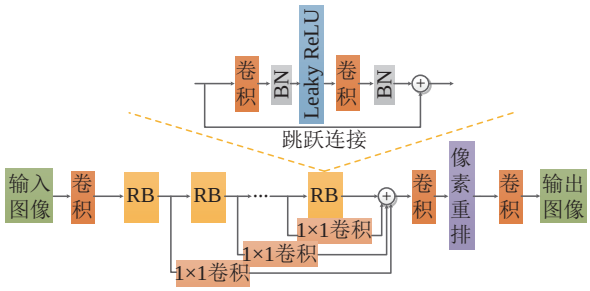


图 20 SRFeat 网络模型图

Fig. 20 The network structure of SRFeat

SRFeat 模型的生成器同样需要进行预训练, 而它的判别器与 SRGAN 模型的判别器结构相同, 将其中一个置于中间层, 另一个置于重构输出层, 损失函数不仅包括预训练生成网络的损失, 还包括特征对抗损失与图片对抗损失.

生成对抗网络重构出的图片通常含有很多额外的细节信息, 并不都是我们需要的细节, 因此, 为了使重构图片尽可能与原始高分辨率图片接近, SRFeat 模型通过在特征域加入判别器使得重构图像的特征也尽可能与高分辨率图像特征相同, 这就约束了生成图片的细节信息, 避免产生不必要的噪声. 在两个判别器的共同约束下, SRFeat 模型的重构结果与 SRGAN 相比更加理想.

2.3.3 双 GAN 模型

英国诺丁汉大学的 Bulat 等[45] 提出利用两个 GAN 网络来进行人脸超分辨率重构的网络模型. 尽管这一模型是具体应用于人脸图像超分辨率重构的, 但是它的思想可以推广到其他的图像超分辨率重构应用方面. Bulat 等[45] 指出, 当前提出的单幅图片超分辨率重构方法在研究阶段能够取得良好的效果, 因为低分辨率图像通常是由计算机进行人为下采样获取的. 但是实际应用过程中, 图像降质的过程十分复杂, 利用人工构建的高低分辨率图像对进行训练通常难以获取理想的重构效果. 因此, 作者提出先学习图像的降质过程, 再进行图像超分辨率重构的思想. 该模型首先利用一个 GAN 网络学习由高分辨率图像降质为低分辨率图像的过程, 以生成低分辨率图像, 然后利用生成的低分辨率图像与原始高分辨率构成图像对, 利用另外的 GAN 网络进行超分辨率重构. 网络结构如图 21.

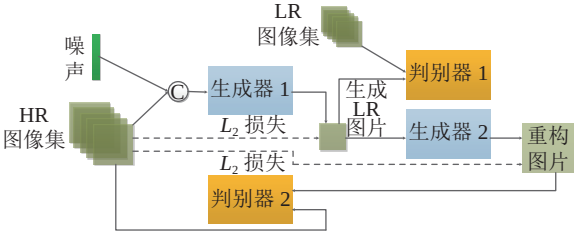


图 21 双 GAN 网络模型图

Fig. 21 The network structure of Double-GAN

将高分辨率图像增加噪声后输入 High-to-low 的 GAN 网络生成器中, 与低分辨率图像 (不需要与输入高分辨率图像对应) 对抗训练, 获取经过神经网络学习后降质的低分辨率图像; 然后将生成的低分辨率图像输入到 Low-to-high 的 GAN 网络生成器中, 再通过与原始高分辨率图像对抗训练得到重构图像.

该网络将真实图片的降质过程作为图像超分辨率重构研究的一部分, 能够将实验研究与实践结合, 提高图像超分辨率重构的实际应用价值. 真实图像的降质过程也可以不局限于利用生成对抗网络进行学习, 如何用神经网络学习实际获取低分辨率图片的过程也是图像超分辨率重构技术的重难点问题之一.

2.3.4 小结

当前单幅图片超分辨率重构的另一种发展方向是对于细节信息的追求, 而生成对抗网络能够通过对抗训练产生图片的细节信息, 因此成为图像超分辨率重构的重要基础模型. 但是由于生成对抗网络的不稳定性, 通常需要对生成器进行预训练. 表 3 对第 2.3 节介绍的 3 种网络模型从输入图像类型、网络层数、损失函数、评价指标以及放大因子这些

表 3 基于生成对抗网络的 3 种模型对比

Table 3 Comparison of the three models based on GAN

网络模型	输入图像	网络层数	损失函数	评价指标	放大因子
SRGAN	LR	16 (Blocks)	VGG	PSNR, SSIM, MOS	2, 3, 4
SRFeat	LR	16 (Blocks)	VGG	PSNR, SSIM, 分类效果	2, 3, 4
双GAN	LR	16 (Blocks)	L_2 范数	PSNR	2, 3, 4

方面进行了总结对比.

基于生成对抗网络的生成器模型均以残差网络为基础, 判别器的加入使得训练过程更加逼近原始高分辨率图片, 尽管 GAN 网络能够产生视觉效果良好的高频细节信息, 但是随着细节的增加, 使用 PSNR, SSIM 作为评价指标也受到了挑战. SRGAN 模型引入了主观评价指标 MOS, 但是如何确定重构细节信息的有效性仍有待研究. 通过引入主观评价, 将主客观评价进行融合, 或者通过卷积神将网络构建一个能够模拟主观评价的评价系统是两种可行的解决方法.

2.4　基于其他网络结构的 SR 模型

2.4.1　SRDenseNet 模型

SRDenseNet (Super-resolution using dense skip connections) 是 Tong 等[46] 在 DenseNet 的基础上加入跳跃连接构建的网络模型. DenseNet 与 ResNet不同, 它是通过对每一层的特征充分利用来解决网络加深时梯度消失/爆炸的问题. SRDenseNet 不仅像 DenseNet 那样充分利用特征, 还加入了跳跃连接以传递之前的信息. 结构如图 22所示. 网络共堆叠8 个 DB (Dense block), 每个 DB 内部密集连接, 充分利用了内部特征; 同时外部也密集连接, 又充分利用了前面 DB 的特征. 所有的特征都会与损失函数连接, 从而减轻梯度消失/爆炸问题, 经过 1×1 卷积进行降维, 然后重构图像.

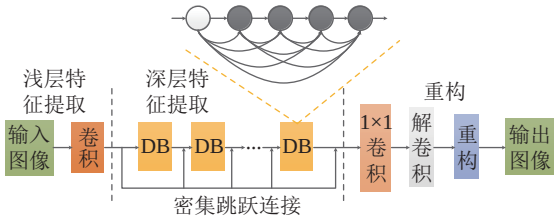


图 22　SRDenseNet 网络模型图

Fig. 22　The network structure of SRDenseNet

SRDenseNet 本身以 DenseNet 为基础, 又加入了跳跃连接, 整个网络包括了 69 个权重层和 68 个激活层, 以极深的网络进行特征提取, 获取了较好的重构效果.

2.4.2　MemNet 模型

MemNet (Memory network) 是南京理工大学的 Tai 等[47] 提出的针对图像去噪、超分辨率重构等图像复原问题的综合网络模型. MemNet 将门控单元引入到递归网络中, 使得先提取到的特征有权重地输入到循环的 MB (Memory block) 中. 网络结构如图 23 所示.

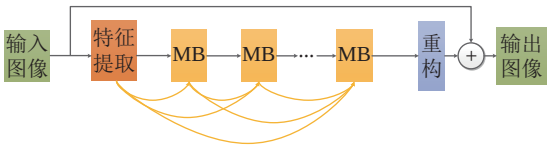


图 23　MemNet 网络模型图

Fig. 23　The network structure of MemNet

图 23 中, 每个 MB 中包含残差块和门控单元, 残差块主要利用的是当前记忆单元中的特征信息, 门控单元利用的是前面所有记忆单元和当前单元的特征, 从而构成短期记忆 (Short memory) 和长期记忆 (Long memory), 因此称为 MemNet. 门控单元的引入既能够传递之前的特征信息用于后面的图像复原, 又能够解决梯度消失/爆炸的问题.

2.4.3　RDN 模型

美国东北大学的 Zhang 等[48] 提出了一种结合 DenseNet 和 ResNet 的网络模型 RDN (Residual dense network). RDN 网络结构充分的利用了卷积神经网络提取的分层特征, 包含 4 个模块, 分别为浅层特征提取、RDB (Residual dense blocks)、DFF (Dense feature fusion) 以及上采样模块. 网络结构如图 24所示.

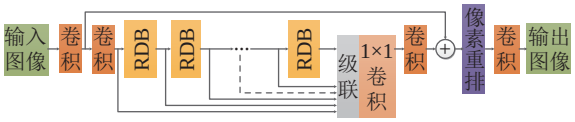


图 24　RDN 网络模型图

Fig. 24　The network structure of RDN

图 24 中 RDB 模块的结构如图 25 所示, 内部进行了密集连接, 并通过级联与 1×1 卷积实现了局部特征融合, 每个 RDB 的输出不仅接收内部的局部融合特征, 还接收前一个 RDB 的输出; 每个 RDB 之间输出也通过级联与 1×1 卷积实现了全局特征融合; 特征重构部分采用 ESPCN 模型的通道像素重排策略.

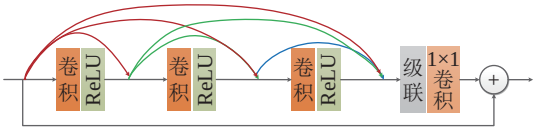


图 25　RDB 结构图

Fig. 25　The structure of residual dense block

RDN 与 SRDenseNet, MemNet 模型之间具有相似的结构, 但是又存在不同之处. 表 4 对与 RDN 模型结构相似的网络进行了对比.

表 4 3 种网络模型对比

Table 4 Comparison of the three models

网络模型	递归单元	密集连接	特征融合	重构效果
SRResNet	RB	无	无	细节明显
DenseNet	DB	无	所有DB之后	–
SRDenseNet	DB	DB之间	所有DB之后	较好
MemNet	MB	MB之间	无	较好
RDN	RDB	RDB内部	RDB内部和所有RDB之后	好

由表 4 可知, RDN 模型既吸取了 SRResNet 模型与 SRDenseNet 模型的长处, 又对二者进行了改进, 取得了良好的重构效果.

2.4.4 IDN 模型

西安电子科技大学的 Hui 等[49] 针对卷积神经网络计算量大、运行内存高的问题提出了一种 IDN (Information distillation network) 模型. 网络包含 3 个部分, 特征提取部分 FBlock (Feature extraction block), 堆叠的信息提取部分 DBlock (Information distillation block) 和重构部分 RBlock (Reconstruction block). 结构如图 26 所示.

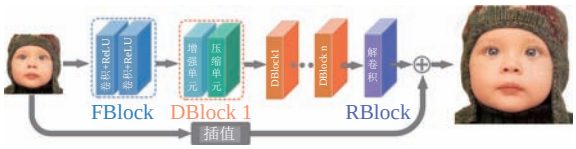


图 26 IDN 网络模型图

Fig. 26 The network structure of IDN

每个信息提取部分 DBlock 包含增强单元和压缩单元, 增强单元的结构如图 27 所示.

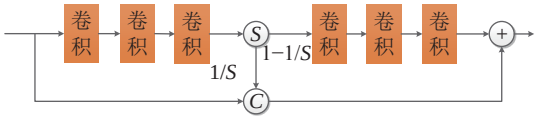


图 27 DBlock 结构图

Fig. 27 The structure of information distillation blocks

图 27 中 S 表示将特征分组 (Slice), C 表示特征级联. 将前面输入的特征提取出 $1/s$ 直接级联到输出端, 剩下的 $(1-1/s)$ 继续输入到后面的网络中, 达到信息增强的效果. 而压缩单元的压缩功能由 1×1 卷积实现.

IDN 模型通过对 DBlock 中的浅层特征进行提取并与深层特征级联达到增强信息传递的目的, 使得细节信息被更好地提取出来, 又通过 1×1 卷积将其压缩, 节省了运行内存和计算量. 整个 IDN 模型只包含 4 个 DBlock, 但是训练结果优于 DRRN 模型、MemNet 模型等深层次网络; 而重构时间约 0.01 s, 速度约为 MemNet 模型的 1000 倍.

2.4.5 DBPN 模型

丰田技术研究院的 Haris 等[50] 根据迭代反投影方法构建的卷积神经网络模型 DBPN (Deep back-projection networks) 是 2018 年 NTIRE 图像超分辨率重构挑战大赛的冠军. 与迭代反投影法不同, 它迭代投影的对象不是高、低分辨率图像而是它们的特征 (Feature maps). 结构如图 28 所示.

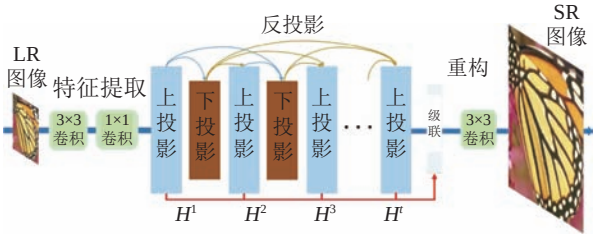


图 28 DBPN 网络模型图

Fig. 28 The network structure of DBPN

其核心是迭代上投影和下投影模块, 两个模块分别如图 29 和图 30 所示.

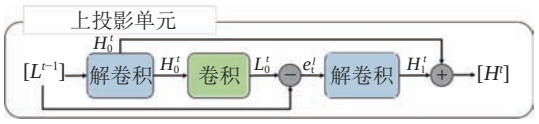


图 29 上投影单元结构

Fig. 29 The structure of up-projection unit

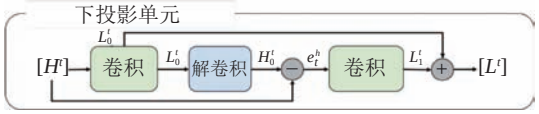


图 30 下投影单元结构

Fig. 30 The structure of down-projection unit

在上投影单元中, 输入低分辨率图像特征 L^{t-1}, 经过上采样得到高分辨率特征 H_0^t, 再通过卷积进行下采样, 与 L^{t-1} 求误差, 再上采样至与高分辨率图像特征同尺寸并与 H_0^t 相加, 得到输出 H^t, 下投影单元同理. 将上下投影单元交替堆叠, 且所有输出密集连接, 并用跳跃连接将每个上投影单元的输出连接到后面的上投影单元及级联层, 用于最终的图像重构.

DBPN 网络借鉴迭代反投影法的思想, 通过对特征的不断迭代, 提取了图像的细节信息用于重构, 迭代过程使神经网络能够更好地受特征约束, 拟合高低分辨率图像之间的映射关系, 在×2、×4 以及×8

倍数的重构结果均十分理想, 且参数量远少于 2017 年冠军 EDSR 模型.

2.4.6　小结

随着图像领域各种基础网络模型的提出, 单幅图片超分辨率重构方法也取得了相应的进展, 第 2.4 节共介绍了 5 种基于不同网络结构的单幅图片超分辨率重构模型, 这些网络模型在输入图像、网络层数、损失函数、评价指标、放大因子方等面的对比见表 5. 其中, RDN 模型结合了 SRDenseNet 与 MemNet模型的优点并加以改进, 取得了良好的重构效果; IDN 模型网络轻便、参数量少、训练速度快, 不要求较高计算性能的设备来处理; DBPN 模型通过反复迭代能够实现×8 倍数的重构, 且重构效果良好.

3　分析与总结

3.1　结果与分析

自 2014 年 SRCNN 模型提出以来, 经过短短几年, 基于深度学习的单幅图片超分辨率重构方法蓬勃发展, 重构结果不断取得突破. 本文介绍了多种基于标准卷积神经网络、残差网络、生成对抗网络以及其他网络等网络模型为基础构建的单幅图片超分辨率重构模型, 表 6 ~ 9 分别给出了这些网络模型在基准测试集 Set5、Set14、BSD100、Urban100 以及 Manga109 上的×2、×3、×4和×8 放大因子重构结果的 PSNR/SSIM 值. 表中未给出基于生成对抗网络模型的 PSNR/SSIM 值, 因为这些模型重构效果注重细节, 不适合用 PSNR/SSIM 进行评价. 同时, 未给出没有在基准测试集上测试的部分网络结果. 重构效果 PSNR/SSIM 值最高的结果用粗体字表示. 表 10 给出了本文介绍的所有模型的网络基础、模型框架、网络结构设计特点, 并对不同网络采用的实验平台及训练测试时间进行了汇总, 以供参考. 由表 6 ~ 10 可知, 应用卷积神经网络解决单幅图片超分辨率重构问题取得了良好的重构结果: ESPCN 模型的通道重排策略及 FSRCNN 的

表 5　基于其他网络的 5 种模型对比

Table 5　Comparison of the five models based on other networks

网络模型	输入图像	网络层数	损失函数	评价指标	放大因子
SRDenseNet	LR	8 (Blocks)	L_2 范数	PSNR, SSIM	4
MemNet	ILR	80	L_2 范数	PSNR, SSIM	2, 3, 4
RDN	LR	20 (Blocks)	L_1 范数	PSNR, SSIM	2, 3, 4
IDN	LR	4 (Blocks)	L_1 范数	PSNR, SSIM, IFC	2, 3, 4
DBPN	LR	2/4/6 (Units)	L_2 范数	PSNR, SSIM	2, 4, 8

表 6　各个网络模型在 Set5、Set14、BSD100、Urban100 和 Manga109 测试集上×2 倍数重构结果 (单位: dB/-)

Table 6　Quantitative results of the SR models on Set5, Set14, BSD100, Urban100 and Manga109 with scale factor ×2 (Unit: dB/-)

放大尺度	网络模型	Set5 (PSNR/SSIM)	Set14 (PSNR/SSIM)	BSD100 (PSNR/SSIM)	Urban100 (PSNR/SSIM)	Manga109 (PSNR/SSIM)
×2	SRCNN[5]	33.66/0.9542	32.45/0.9067	31.36/0.8879	29.50/0.8946	35.60/0.9663
	FSRCNN[26]	37.05/0.9560	32.66/0.9090	31.53/0.8920	29.88/0.9020	36.67/0.9694
	ESPCN[31]	37.00/0.9559	32.75/0.9098	31.51/0.8939	29.87/0.9065	36.21/0.9694
	VDSR[33]	37.53/0.9588	33.03/0.9124	31.90/0.8960	30.76/0.9140	37.22/0.9729
	DRCN[34]	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133	37.63/0.9723
	DRRN[35]	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188	37.60/0.9736
	RED[36]	37.66/0.9599	32.94/0.9144	31.99/0.8974	–	–
	LapSRN[37]	37.52/0.9590	33.08/0.9130	31.08/0.8950	30.41/0.9100	37.27/0.9855
	EDSR[38]	38.11/0.9602	33.92/0.9195	32.32/0.9013	32.93/0.9351	39.10/0.9773
	CARN-M[40]	37.53/0.9583	33.26/0.9141	31.92/0.8960	30.83/0.9233	–
	MSRN[32]	38.08/0.9605	33.74/0.9170	32.23/0.9013	32.22/0.9326	38.82/**0.9868**
	RCAN[42]	**38.33**/**0.9617**	**34.23**/**0.9225**	**32.46**/**0.9031**	**33.54**/**0.9399**	**39.61**/0.9788
	MemNet[47]	37.78/0.9597	33.28/0.9142	32.08/0.8978	31.31/0.9195	37.72/0.9740
	RDN[48]	38.24/0.9614	34.01/0.9212	32.34/0.9017	32.89/0.9353	39.18/0.9780
	IDN[49]	37.83/0.9600	33.30/0.9148	32.08/0.8985	31.27/0.9196	–
	DBPN[50]	38.09/0.9600	33.85/0.9190	32.27/0.9000	32.55/0.9324	38.89/0.9775

表 7 各个网络模型在 Set5、Set14、BSD100、Urban100 和 Manga109 测试集上×3 倍数重构结果 (单位: dB/-)
Table 7 Quantitative results of the SR models on Set5, Set14, BSD100, Urban100 and Manga109 with scale factor ×3 (Unit: dB/-)

放大尺度	网络模型	Set5 (PSNR/SSIM)	Set14 (PSNR/SSIM)	BSD100 (PSNR/SSIM)	Urban100 (PSNR/SSIM)	Manga109 (PSNR/SSIM)
	SRCNN[5]	32.75/0.9090	29.30/0.8215	28.41/0.7863	26.24/0.7989	30.48/0.9117
	FSRCNN[26]	33.18/0.9140	29.37/0.8240	28.53/0.7910	26.43/0.8080	31.10/0.9210
	ESPCN[31]	33.02/0.9135	29.49/0.8271	28.50/0.7937	26.41/0.8161	30.79/0.9181
	VDSR[33]	33.68/0.9201	29.86/0.8312	28.83/0.7966	27.15/0.8315	32.01/0.9310
	DRCN[34]	33.85/0.9215	29.89/0.8317	28.81/0.7954	27.16/0.8311	32.31/0.9328
	DRRN[35]	34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378	32.42/0.9359
×3	RED[36]	33.82/0.9230	29.61/0.8341	28.93/0.7994	–	–
	EDSR[38]	34.65/0.9280	30.52/0.8462	29.25/0.8093	28.80/0.8653	34.17/0.9476
	CARN-M[40]	33.99/0.9236	30.08/0.8367	28.91/0.8000	26.86/0.8263	–
	MSRN[32]	34.38/0.9262	30.34/0.8395	29.08/0.8041	28.08/0.5554	33.44/0.9427
	RCAN[42]	**34.85/0.9305**	**30.76/0.8494**	**29.39/0.8122**	**29.31/0.8736**	**34.76/0.9513**
	MemNet[47]	34.09/0.9248	30.00/0.8350	28.96/0.8001	27.56/0.8376	32.51/0.9369
	RDN[48]	34.71/0.9296	30.57/0.8468	29.26/0.8093	28.80/0.8653	34.13/0.9484
	IDN[49]	34.11/0.9253	29.99/0.8354	28.95/0.8031	27.42/0.8359	–

表 8 各个网络模型在 Set5、Set14、BSD100、Urban100 和 Manga109 测试集上×4 倍数重构结果 (单位: dB/-)
Table 8 Quantitative results of the SR models on Set5, Set14, BSD100, Urban100 and Manga109 with scale factor ×4 (Unit: dB/-)

放大尺度	网络模型	Set5 (PSNR/SSIM)	Set14 (PSNR/SSIM)	BSD100 (PSNR/SSIM)	Urban100 (PSNR/SSIM)	Manga109 (PSNR/SSIM)
	SRCNN[5]	30.48/0.8628	27.50/0.7513	26.90/0.7101	24.52/0.7221	27.58/0.8555
	FSRCNN[26]	30.72/0.8660	27.61/0.7550	26.98/0.7150	24.62/0.7280	27.90/0.8610
	ESPCN[31]	30.66/0.8646	27.71/0.7562	26.98/0.7124	24.60/0.7360	27.70/0.8560
	VDSR[33]	31.35/0.8830	28.02/0.7680	27.29/0.7251	25.18/0.7540	28.83/0.8870
	DRCN[34]	31.56/0.8810	28.15/0.7627	27.24/0.7150	25.15/0.7530	28.98/0.8816
	DRRN[35]	31.68/0.8888	28.21/0.7721	27.38/0.7284	25.44/0.7638	29.19/0.8914
	RED[36]	31.51/0.8869	27.86/0.7718	27.40/0.7290	–	–
	LapSRN[37]	31.54/0.8850	28.19/0.7720	27.32/0.7270	25.27/0.7560	29.09/0.8900
×4	EDSR[38]	32.46/0.8968	28.80/0.7876	27.71/0.7420	26.64/0.8033	31.02/0.9148
	CARN-M[40]	31.92/0.8903	28.42/0.7762	27.44/0.7304	25.63/0.7688	–
	MSRN[32]	32.07/0.8903	28.60/0.7751	27.52/0.7273	26.04/0.7896	30.17/0.9034
	RCAN[42]	**32.73/0.9013**	**28.98/0.7910**	**27.85/0.7455**	**27.10/0.8142**	**31.65/0.9208**
	SRDenseNet[46]	32.02/0.8934	28.50/0.7782	27.53/0.7337	26.05/0.7819	–
	MemNet[47]	31.74/0.8893	29.26/0.7723	27.40/0.7281	25.50/0.7630	29.42/0.8942
	RDN[48]	32.47/0.8990	28.81/0.7871	27.72/0.7419	26.61/0.8028	31.00/0.9151
	IDN[49]	31.82/0.8930	28.25/0.7730	27.41/0.7297	25.41/0.7632	–
	DBPN[50]	32.47/0.8980	28.82/0.7860	27.72/0.7400	26.38/0.7946	30.91/0.9137

解卷积策略能够实现图像在网络的最后阶段插值, 从而减少网络模型的训练数据量, 减轻由预插值图像带来的重构效果上的不良影响; 应用 ResNet 和 DenseNet 网络模型可以加深网络, 充分学习原始低分辨率图像的特征, 同时采用递归监督、金字塔结构、特征融合、密集连接、通道注意机制等网络结构设计, 减轻梯度消失/爆炸问题, 加强特征的提取与传递, 提高了重构效果, 尤其是 ECCV18 提出的 RCAN 模型取得的重构效果指标结果最好, 充分说明了通道注意机制引入的有效性; RDN、IDN、MemNet 以及 DBPN 模型则分别采用深度学习实现对低分辨率图像的特征提取与传递, 以达到增强重构效果的目的. 总之, 单幅图片包含的先验信息要远低于多幅图片[51], 而应用深度学习的方法, 能够充分利

表 9　各个网络模型在 Set5、Set14、BSD100、Urban100 和 Manga109 测试集上×8 倍数重构结果 (单位: dB/-)
Table 9　Quantitative results of the SR models on Set5, Set14, BSD100, Urban100 and Manga109 with scale factor ×8 (Unit: dB/-)

放大尺度	网络模型	Set5 (PSNR/SSIM)	Set14 (PSNR/SSIM)	BSD100 (PSNR/SSIM)	Urban100 (PSNR/SSIM)	Manga109 (PSNR/SSIM)
×8	LapSRN[37]	26.14/0.7380	24.44/0.6230	24.54/0.5860	21.81/0.5810	23.39/0.7350
	MSRN[32]	26.59/0.7254	24.88/0.5961	24.70/0.5410	22.37/0.5977	24.28/0.7517
	RCAN[49]	**27.47/0.7913**	**25.40/0.6553**	**25.05/0.6077**	**23.22/0.6524**	**25.58/0.8092**
	DBPN[50]	27.12/0.7840	25.13/0.6480	24.88/0.6010	22.73/0.6312	25.14/0.7987

表 10　各个网络模型的网络基础、模型框架、网络设计、实验平台及运行时间总结
Table 10　Summary of the SR models in network basics, frameworks, network design, platform and training/testing time

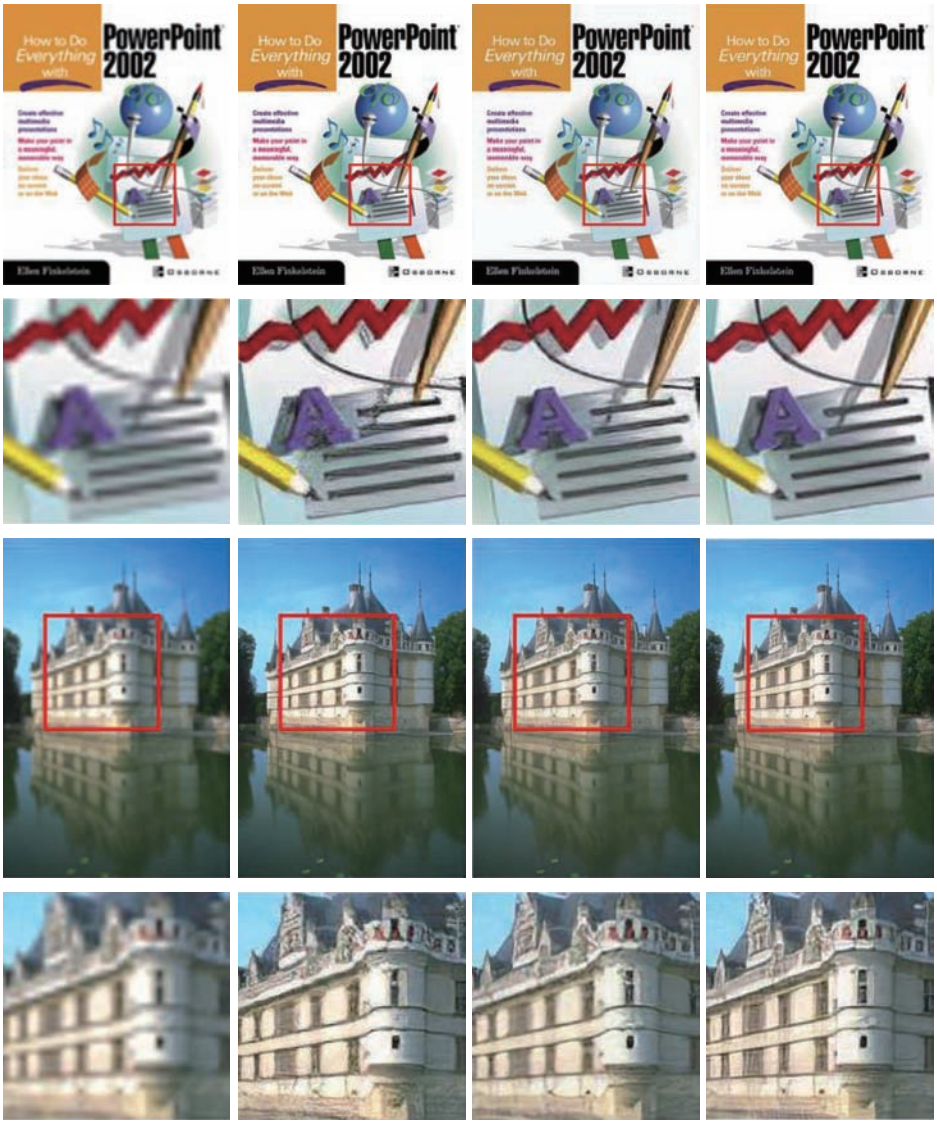
网络模型	网络基础	模型框架	结构设计特点	实验平台	训练/测试时间
SRCNN[5]	CNN	预插值	经典 CNN 结构	CPU	–
FSRCNN[26]	CNN	后插值 (解卷积)	压缩模块	i7 CPU	0.4 s (测试)
ESPCN[31]	CNN	后插值 (亚像素卷积)	亚像素卷积	K2 GPU	4.7 ms (测试)
VDSR[33]	ResNet	预插值	残差学习, 自适应梯度裁剪	Titan Z GPU	4 h (训练)
DRCN[34]	ResNet	预插值	递归结构, 跳跃连接	Titan X GPU	6 d (训练)
DRRN[35]	ResNet	预插值	递归结构, 残差学习	Titan X GPU×2	4 d/0.25 s
RED[36]	ResNet	逐步插值	解卷积–反卷积, 跳跃连接	Titan X GPU	3.17 s (测试)
LapSRN[37]	ResNet	逐步插值	金字塔结构, 特征–图像双通道	Titan X GPU	0.02 s (测试)
EDSR[38]	ResNet	后插值 (亚像素卷积)	去 BN 层, Self-ensemble	Titan X GPU	8 d (训练)
CARN[40]	ResNet	后插值 (亚像素卷积)	递归结构, 残差学习, 分组卷积	–	–
MSRN[32]	ResNet	后插值 (亚像素卷积)	多尺度特征提取, 残差学习	Titan Xp GPU	–
RCAN[42]	ResNet	后插值 (亚像素卷积)	递归结构, 残差学习, 通道注意机制	Titan Xp GPU	–
SRGAN[43]	GAN	后插值 (亚像素卷积)	生成器预训练	Telsa M40 GPU	–
SRFeat[44]	GAN	后插值 (亚像素卷积)	特征判别器, 图像判别器	Titan Xp GPU	–
双GAN[45]	GAN	–	两个 GAN 网络构成图像降质与重构闭合回路	–	–
SRDenseNet[46]	其他	后插值 (解卷积)	密集连接, 跳跃连接	Titan X GPU	36.8 ms (测试)
MemNet[47]	其他	预插值	记忆单元, 跳跃连接	Telsa P40 GPU	5 d/0.85 s
RDN[48]	其他	后插值 (解卷积)	密集连接, 残差学习	Titan Xp GPU	1 d (训练)
IDN[49]	其他	后插值 (解卷积)	蒸馏机制	Titan X GPU	1 d (训练)
DBPN[50]	其他	迭代插值	上、下投影单元	Titan X GPU	4 d (训练)

用具有相同降质模型的不同图片样本之间的先验信息, 且获取的重构效果优于非深度学习方法, 根据表 6, 最好的结果在 Manga109 数据集上×2 倍数的 PSNR 值达到了 39.61 dB (高分辨率图片 40 dB); 根据表 10, 在重构速度上, 一旦网络训练学习完成, 重构一幅图片的速度最高可达到 4.7 ms.

根据本文给出的超分辨率重构网络模型的重构结果, 当前基于深度学习的单幅图片超分辨率重构方法发展趋势包含以下四个方面:

1) 追求好的重构效果. 由表 6～9 可知, 以残差网络 (ResNet) 为基础的 RCAN 模型取得的重构效果最佳; 且以残差网络为基础的超分辨率重构模型众多, 辅以各种加强特征提取与传递的模块和监督策略, 取得了较好的重构效果. 因此, 在重构效果方面, 对于特征的提取与传递、梯度监督的策略进行研究, 并以残差网络构建新的超分辨率重构模型具有一定的可行性.

2) 追求丰富明显的细节. 生成对抗网络 (GAN) 能够通过对抗训练产生细节丰富的图片, 这一特点为单幅图片超分辨率重构提供了思路, 图 31[44] 是双三次插值、EnhanceNet[52] 模型、SRGAN 模型与 SRFeat 模型重构效果, 可以看出, 后三者在细节方面重构效果细节逐步明显, 模糊现象减少, 且由于 SRFeat 模型中两个判别器分别约束特征空间与图像空间, 生成的图片视觉效果更佳. 此外, 基于生成对抗网络的超分辨率重构模型重构出的图像由客观评价指标 PSNR 和 SSIM 衡量并不合理, SRGAN 提出了一种主观评价指标 MOS, 但是该指标需要



(a) 双三次插值
(a) Bicubic

(b) EnhanceNet
(b) EnhanceNet

(c) SRGAN
(c) SRGAN

(d) SRFeat
(d) SRFeat

图 31 基于生成对抗网络×4 放大倍数的单幅图片超分辨率重构结果

Fig. 31 Qualitative comparison of GAN-based SR methods at scaling factor 4

大量的专业人员进行评价, 且评价结果具有很强的主观性. 针对这类重构图像进行评价具有重要的研究意义, 可从主观与客观评价结合入手, 也可以从其他评价指标入手, 或是利用神经网络构建主观评价体系等.

3) 追求轻量化网络. 随着重构结果的指标越来越难提高, 单幅图片超分辨率重构网络越来越注重在追求好的重构结果的同时构建轻量级网络. 但是随着网络参数减少, 重构结果自然会下降, 实现二者间的平衡是单幅图片超分辨率重构网络的另一种发展方向. IDN 模型首先提出通过优化网络结构以节省内存空间和计算量; 图 32[40] 给出部分模型之间重构效果、计算量以及参数量之间的关系图, 由该图结果可以看出, MDSR (EDSR) 模型效果最佳, 但是其参数量很大, 约为 CARN-M 模型的 6 倍, 但 PSNR值仅高 0.4 dB 左右. 由此可见, CARN-M 模型既能够实现较好的重构效果, 参数量和计算量也较少.

4) 追求较高的放大倍数. 单幅图片超分辨率重构技术属于软件技术, 它只能在一定程度上对图片进行重构, 目前大多数网络模型只能实现×2、×3和×4 倍数放大, LapSRN 模型的提出为 ×8 倍数的重构提供了基准结果, NTIRE2017 挑战赛上的 DBPN 模型通过迭代实现了×8 倍数放大, ECCV 2018 会议上, MSRN、RCAN 两种模型均实现了×8 倍数放大且效果越来越好. 对于单幅图片超分辨重

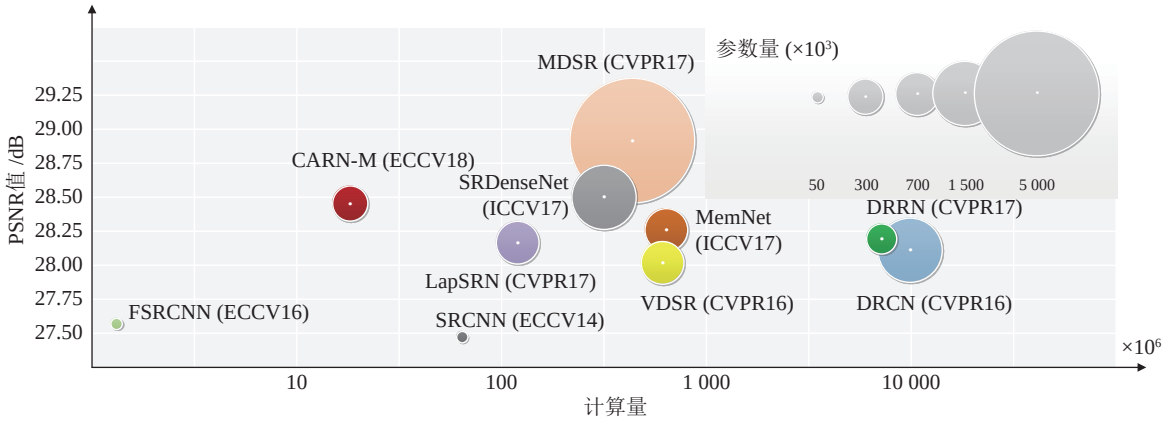


图 32　各个模型在 Set14 测试集×4 放大倍数的重构效果、计算量以及参数量之间的关系图

Fig. 32　Trade-off between performance vs. number of operations and parameters on Set14 ×4 dataset

构而言, 倍数越高重构越困难, 优化网络, 提高高倍数放大因子的重构效果, 也是超分辨率重构技术的未来发展需求之一.

5) 追求任意放大倍数. ESPCN 模型提出的亚像素卷积 (也称通道像素重排) 思想, 与解卷积层共同成为当前单幅图片超分辨率重构网络中的两大重构策略. 与解卷积不同, 亚像素卷积利用不同的特征数目 (Feature maps) 控制重构图像的大小, 即放大倍数与特征数相关. NTIRE2017 挑战赛中冠军 SNU CVLab 团队在 EDSR 模型基础上提出了 MDSR (Multi-scale deep super-resolution) 模型, 在特征提取网络之后通过亚像素卷积进行图像重构, 所有尺度的重构模型共享一个特征提取结构, 即通过一次学习能获取任意倍数的重构图像, 改变图像重构模块对于网络结构的影响不大, 通过卷积层即可实现, 但是重构结果与单一尺度模型相比存在差距, 因此, 在不追求训练学习时间的情况下, 研究者们通常选取单一尺度放大模型. 但是, 亚像素卷积能够通过改变特征数目来满足重构图像不同放大倍数的要求, 且具有较好的重构效果, 在此基础上, MSRN 和 RCAN模型均采用了这一策略, 实现了×2、×3、×4 和×8 倍数的放大, 与 LapSRN 和 DBPN 模型只能实现 2 的幂次倍数放大相比更具有灵活性.

3.2　关键问题总结

3.2.1　网络结构设计问题

一个好的超分辨率重构网络结构不仅能够获取较好的重构效果, 还能通过较少的内存和计算量提取尽量好的特征表示. 基于深度学习的单幅图片超分辨率重构方法紧跟深度学习在图像领域的发展, 充分利用了基础网络结构的优势, 但是, 针对图像超分辨率重构的特定应用, 应当更充分地研究如何使用合适深度的网络提取有利于重构的特征信息, 例如 RCAN 模型引入的通道注意机制, IDN 模型引入的蒸馏单元等, 这些特殊的网络设计使网络结构能够有效地选取有用的信息, 减少运行成本并且提高重构效果; 基于 GAN 网络的超分辨率重构模型生成的图像细节不可控, 如何对于生成器进行较强的约束, 以达到尽可能逼近原始高分辨率图像细节的目的也是一个不容忽视的问题; 无论是图像预插值还是解卷积和亚像素卷积, 都存在自身不可避免的问题, 如计算量大, 棋盘格效应以及重构效果不均匀等影响, 对于图像的重构阶段采用的技术有待提升.

3.2.2　网络训练及优化问题

卷积神经网络能够通过加深或者加宽网络拟合复杂的映射关系, 达到 “深度”学习的目的. 然而随着网络加深, 训练的难度也随之变大. SRCNN 模型增加到 4 层网络的时候重构效果就开始下降. ResNet、DenseNet 等网络模型的提出, 3×3、5×5 等小卷积核的引入, 特征信息的融合以及多种监督策略使得更深的网络模型能够应用于图像超分辨率, 深度网络训练难的问题逐步得到了缓解. EDSR 模型去掉了批归一化层 (BN), 减少运行内存和计算时间, 以用于构建更深的网络, 但是对于批归一化层的取舍缺乏充分的理论指导. 同时, 优化问题也是提高网络模型效果和效率的重要方面. 权重参数的初始化问题, PReLU、Leaky ReLU 等激活函数以及减少网络参数和计算量的分组卷积和 MobileNet 等思想的引入, 使得网络的训练过程得以优化. 数据增强策略如 EDSR 模型提出的 Self-ensemble, 通道随机重排[53] 等增强了对数据的利用, 尤其是对于重构图像的细节部分具有重要作用.

基于深度学习的单幅图片超分辨率重构方法的网络训练和优化方面近年来取得了一定的发展, 但是仍然具有很多值得探索改进的地方, 如提高深度网络的可学习性以及其特征的可选择性, 减少参数量, 数据增强等方面.

3.2.3 损失函数与评价指标的选取

损失函数可以看作是高分辨率图像与重构图像之间的约束, 能够指导模型进行优化. 基于深度学习的单幅图片超分辨率重构模型提出之初均以 L_2 范数作为损失函数, 目的是获取最小均方误差从而获得较高的 PSNR 值. 但是随着 PSNR 值越来越高, 研究者发现图像质量并未取得良好的视觉效果. 针对这一问题, 在损失函数的选取上, L_1 范数、VGG19 损失、纹理损失、对抗损失等损失函数逐渐被采用, 并且取得了一定的效果, 但是目前针对图像超分辨率重构的损失函数的选取没有完善的理论依据, 仍然是需要研究的重要问题之一.

评价指标与损失函数具有一定的联系, 好的评价指标能够有效地评价重构图像的质量, 从而对模型进行优化改进. 表 11 给出了常用评价指标的计算方法与优缺点, 其中 MAX_f 为重构图像 X 的最大信号值, MSE 为最小均方误差, μ 为均值, σ 为协方差, C 为常数. 客观评价指标 PSNR 和 SSIM 普遍被专家学者们认可, 但是它们对重构图像的评价并不全面, 主观评价指标评价结果可靠, 但是需要消耗巨大的人力财力. ECCV18 会议中, 多数研究者采用了重构的图像进行图像分类, 通过分类的准确率来间接评价重构图像的质量, 但是这种评价方法只能针对于具有明确物体的图像, 且需要一定的先验知识; SRGAN 模型提出的 MOS 评价指标是目前通用的主观评价指标, 尽管随着主观评价者数量的增加, 评价的可信度能逐渐提高, 但是 MOS 仍具有一些固有的缺陷, 比如, 每个评估者之间的主观差异, 以及评价分数的不连续性造成的评价误差等[54]; 通过神经网络学习人的主观评价, 构建主观评价系统是一项具有应用前景的研究, 但是对于图像主观评价系统的研究可以作为图像领域的一大分支, 不适合在图像超分辨率重构模型中具体研究. 评价指标的选取不仅指导图像超分辨率重构模型的优化改进, 还能应用于图像生成、图像融合、图像复原等领域的图像评价, 因此具有重要的研究意义与研究前景.

3.2.4 图像超分辨率重构的应用

超分辨率重构技术在遥感图像处理、医学诊断、安防监控等难以由低分辨率图像提取有用信息的领域都有广泛的应用. 但是当前的超分辨率重构技术多停留在研究阶段, 针对不同的应用领域, 对于超分辨率重构技术的要求各不相同.

遥感图像通常具有较低的空间分辨率, 难以获取高分辨率图像, 且图像降质过程复杂. Luo 等[27] 使用高分 2 号卫星图像作为高分辨率图像, 解决遥感图像高分辨率数据少的问题; Haut 等[55] 提出了一种无监督的生成网络模型直接生成高分辨率遥感图像; Liu 等[56] 采用卷积–解卷积对称结构处理对遥感图像进行多尺度处理获得重构图像, 并同时能对遥感图像进行彩色化操作.

人脸识别在安防监控领域具有重要的应用, 但是实际应用中, 获取的人脸图像由于成像设备和存储空间限制以及人的运动造成的模糊, 分辨率通常难以达到精确识别的要求. 人脸图像的超分辨率重构十分重要. 针对人脸图像的超分辨率重构通常需要结合先验信息, FSRNet[57] 利用了人脸特征的热度图对彩色人脸图片进行超分辨率重构; Super-FAN[58] 和 MTUN[59] 均引入了 FAN 网络以保证人脸特征的连续性; 双 GAN[45] 网络需要大量的高低分辨率人脸图像来拟合降质过程.

在实际应用中, 需要根据不同的应用场景对网络结构进行相应调整以满足不同的需求. 同时, 在不同应用领域中, 图像的降质过程存在很大的差异, 如何利用软件技术更好地拟合真实图片降质过程, 从而提高图像超分辨率重构技术在实际应用中的重构效果, 是图像超分辨率重构技术面临的一大难题. ECCV2018 提出双 GAN 模型学习人脸图像降质过程为这一实际应用问题提供了一定的思想, 但是该

表 11 常用图像质量评价指标的计算方法和优缺点总结

Table 11 Summary of evaluation metrics

评价指标	计算方法	优点	缺点
PSNR	$10\lg\frac{MAX_f}{MSE}$	能够衡量像素间损失, 是图像最常用的客观评价指标之一.	不能全面评价图像质量, 如PSNR值高不代表图像的视觉质量高.
SSIM	$\frac{(2\mu_x\mu_{x_0}+C_1)\times(2\sigma_{xx_0}+C_2)}{(\mu_x^2+\mu_{x_0}^2+C_1)\times(\sigma_x^2+\sigma_{x_0}^2+C_2)}$	能够衡量图片间的统计关系, 是图像最常用的客观评价指标之一.	不适用于整个图像评价, 只适用于图像的局部结构相似度评价.
MOS	由多位评价者对于重构结果进行评价, 分数从 1 到 5 代表由坏到好.	评价结果更符合人的视觉效果且随着评价者数目增加, 评价结果更加可靠.	耗时耗力, 成本较高, 评价者数量不多的情况下易受评价者主观影响, 且评分不连续易造成较大的误差.

思想在实际应用中有待完善.

4 结论与展望

基于深度学习的图像超分辨率重构技术是当前单幅图片超分辨率重构的主流技术, 无论是在重构速度还是重构效果上均优于传统方法. 然而, 基于深度学习的单幅图片超分辨率重构也存在亟待解决的问题. 首先, 深度学习的理论普遍具有不可解释性, 对于网络的设计与调参往往需要有相关领域知识的人凭借经验来控制, 这大大限制了它的发展空间. 如何从基础理论上将网络的运行机制、原理解释清楚, 提出明确的理论指导, 是基于深度学习的超分辨率重构技术乃至整个深度学习技术面临的问题. 其次, 深度学习的网络模型通常缺乏一定的灵活性, 一旦图像的降质过程存在差异, 同一训练模型得到的重构效果就会受到影响. 如何将当前对于超分辨率重构技术的研究应到实际问题中, 增强网络模型的泛化能力, 是基于深度学习的图像超分辨率重构需要解决的问题之一. 最后, 无论是传统方法还是基于深度学习的方法, 对于较大的尺度放大因子 (如×8), 重构的效果都不是特别理想, 如何在高倍数放大因子的重构中取得良好的重构效果也是基于深度学习的图像超分辨率重构技术面临的挑战之一.

总之, 深度学习在单幅图片超分辨率重构领域已展现了巨大的潜力, 但是仍存在许多没有完善的解决方案的问题, 需要研究者们充分发挥创造力, 进行更多具有挑战性的研究工作.

References

1 Su Heng, Zhou Jie, Zhang Zhi-Hao. Survey of super-resolution image reconstruction methods. *Acta Automatica Sinica*, 2013, **39**(8): 1202−1213
(苏衡, 周杰, 张志浩. 超分辨率图像重建方法综述. 自动化学报, 2013, **39**(8): 1202−1213)

2 Harris J L. Diffraction and resolving power. *Journal of the Optical Society of America*, 1964, **54**(7): 931−936

3 Goodman J W. Introduction to Fourier Optics. New York: McGraw-Hill, 1968

4 Tsai R Y, Huang T S. Multiframe image restoration and registration. In: Advances in Computer Vision and Image Processing. Greenwich, CT, England: JAI Press, 1984. 317−339

5 Dong C, Loy C C, He K M, Tang X O. Learning a deep convolutional network for image super-resolution. In: Proceedings of ECCV 2014 European Conference on Computer Vision. Cham, Switzerland: Springer, 2014. 184−199

6 He Yang, Huang Wei, Wang Xin-Hua, Hao Jian-Kun. Super-resolution image reconstruction based on sparse threshold. *Chinese Optics*, 2016, **9**(5): 532−539
(何阳, 黄玮, 王新华, 郝建坤. 稀疏阈值的超分辨率图像重建. 中国光学, 2016, **9**(5): 532−539)

7 Li Fang-Biao. Research on Super-Resolution Reconstruction of Infrared Imaging System [Ph.D. dissertation], University of Chinese Academy of Sciences, China, 2018
(李方彪. 红外成像系统超分辨率重建技术研究 [博士学位论文], 中国科学院大学, 中国, 2018)

8 Irani M, Peleg S. Improving resolution by image registration. *Graphical Models and Image Processing*, 1991, **53**(3): 231−239

9 Kim K I, Kwon Y. Single-image super-resolution using sparse regression and natural image prior. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2010, **32**(6): 1127−1133

10 Aly H A, Dubois E. Image up-sampling using total-variation regularization with a new observation model. *IEEE Transactions on Image Processing*, 2005, **14**(10): 1647−1659

11 Shan Q, Li Z R, Jia J Y, Tang C K. Fast image/video upsampling. *ACM Transactions on Graphics*, 2008, **27**(5): 153

12 Hayat K. Super-resolution via deep learning. arXiv: 1706.09077, 2017

13 Sun Xu, Li Xiao-Guang, Li Jia-Feng, Zhuo Li. Review on deep learning based image super-resolution restoration algorithms. *Acta Automatica Sinica*, 2017, **43**(5): 697−709
(孙旭, 李晓光, 李嘉锋, 卓力. 基于深度学习的图像超分辨率复原研究进展. 自动化学报, 2017, **43**(5): 697−709)

14 He H, Siu W C. Single image super-resolution using Gaussian process regression. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Providence, RI, USA: IEEE, 2011. 449−456

15 Zhang K B, Gao X B, Tao D C, Li X L. Single image super-resolution with non-local means and steering kernel regression. *IEEE Transactions on Image Processing*, 2012, **21**(11): 4544−4556

16 Chan T M, Zhang J P, Pu J, Huang H. Neighbor embedding based super-resolution algorithm through edge detection and feature selection. *Pattern Recognition Letters*, 2009, **30**(5): 494−502

17 Yang J C, Wright J, Huang T S, Ma Y. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 2010, **19**(11): 2861−2873

18 Timofte R, Agustsson E, van Gool L, Yang M H, Zhang L, Lim B, et al. NTIRE 2017 challenge on single image super-resolution: Methods and results. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Hawaii, USA: IEEE, 2017. 1110−1121

19 Yue L W, Shen H F, Li J, Yuan Q Q, Zhang H Y, Zhang L P. Image super-resolution: The techniques, applications, and future. *Signal Processing*, 2016, **128**: 389−408

20 Yang C Y, Ma C, Yang M H. Single-image super-resolution: A benchmark. In: Proceedings of ECCV 2014 Conference on Computer Vision. Switzerland: Springer, 2014. 372−386

21 He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV,

USA: IEEE, 2016. 770−778

22 Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. In: Advances in Neural Information Processing Systems. Montreal, Quebec, Canada: Curran Associates, Inc., 2014. 1110−1121

23 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. Lake Tahoe, Nevada, USA: Curran Associates, Inc., 2012. 1097−1105

24 Huang G, Liu Z, van der Maaten L, Weinberger K Q. Densely connected convolutional networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017. 2261−2269

25 Dong C, Loy C C, He K M, Tang X O. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2016, **38**(2): 295−307

26 Dong C, Chen C L, Tang X O. Accelerating the super-resolution convolutional neural network. In: Proceedings of European Conference on Computer Vision. Amsterdam, Netherlands: Springer, 2016. 391−407

27 Luo Y M, Zhou L G, Wang S, Wang Z Y. Video satellite imagery super resolution via convolutional neural networks. *IEEE Geoscience & Remote Sensing Letters*, 2017, **14**(12): 2398−2402

28 Ducournau A, Fablet R. Deep learning for ocean remote sensing: An application of convolutional neural networks for super-resolution on satellite-derived SST data. In: Proceedings of the 9th IAPR Workshop on Pattern Recognition in Remote Sensing. Cancun, Mexico: IEEE, 2016. 1−6

29 Rasti P, Uiboupin T, Escalera S, Anbarjafari G. Convolutional neural network super resolution for face recognition in surveillance monitoring. In: Proceedings of the International Conference on Articulated Motion & Deformable Objects. Switzerland: Springer, 2016. 175−184

30 Zhang H Z, Casaseca-de-la-Higuera P, Luo C B, Wang Q, Kitchin M, Parmley A, et al. Systematic infrared image quality improvement using deep learning based techniques. In: Proceedings of the SPIE 10008, Remote Sensing Technologies and Applications in Urban Environments. SPIE, 2016.

31 Shi W Z, Caballero J, Huszar F, Totz J, Aitken A P, Bishop R, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016. 1874−1883

32 Li J C, Fang F M, Mei K F, Zhang G X. Multi-scale residual network for image super-resolution. In: Proceedings of 2018 ECCV European Conference on Computer Vision. Munich, Germany: Springer, 2018. 527−542

33 Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016. 1646−1654

34 Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016. 1637−1645

35 Tai Y, Yang J, Liu X M. Image super-resolution via deep recursive residual network. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017. 2790−2798

36 Mao X J, Shen C H, Yang Y B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Red Hook, NY, United States: Curran Associates Inc., 2016. 2810−2818

37 Lai W S, Huang J B, Ahuja N, Yang M H. Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017. 5835−5843

38 Lim B, Son S, Kim H, Nah S, Lee K M. Enhanced deep residual networks for single image super-resolution. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, HI, USA: IEEE, 2017. 1132−1140

39 Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861, 2017

40 Ahn N, Kang B, Sohn K A. Fast, accurate, and lightweight super-resolution with cascading residual network. In: Proceedings of 2018 ECCV European Conference on Computer Vision. Munich, Germany: Springer, 2018. 256−272

41 Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, NA, USA: IEEE, 2015. 1−9

42 Zhang Y L, Li K P, Li K, Wang L C, Zhong B N, Fu Y. Image super-resolution using very deep residual channel attention networks. In: Proceedings of 2018 ECCV European Conference on Computer Vision. Munich, Germany: Springer, 2018. 294−310

43 Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017. 105−114

44 Park S J, Son H, Cho S, Hong K S, Lee S. SRFeat: Single image super-resolution with feature discrimination. In: Proceedings of 2018 ECCV European Conference on Computer Vision. Munich, Germany: Springer, 2018. 455−471

45 Bulat A, Yang J, Georgios T. To learn image super-resolution, use a GAN to learn how to do image degradation first. In: Proceedings of 2018 ECCV European Conference on Computer Vision. Munich, Germany: Springer, 2018. 187−202

46 Tong T, Li G, Liu X J, Gao Q Q. Image super-resolution using dense skip connections. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 4809−4817

47 Tai Y, Yang J, Liu X M, Xu C Y. MemNet: A persistent memory network for image restoration. In: Proceedings of the

2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 4549−4557

48 Zhang Y L, Tian Y P, Kong Y, Zhong B N, Fu Y. Residual dense network for image super-resolution. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 2472−2481

49 Hui Z, Wang X M, Gao X B. Fast and accurate single image super-resolution via information distillation network. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 723−731

50 Haris M, Shakhnarovich G, Ukita N. Deep back-projection networks for super-resolution. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 1664−1673

51 Pan Zong-Xu, Yu Jing, Xiao Chuang-Bai, Sun Wei-Dong. Single-image super-resolution algorithm based on multi-scale nonlocal regularization. *Acta Automatica Sinica*, 2014, **40**(10): 2233−2244
(潘宗序, 禹晶, 肖创柏, 孙卫东. 基于多尺度非局部约束的单幅图像超分辨率算法. 自动化学报, 2014, **40**(10): 2233−2244)

52 Sajjadi M S M, Scholkopf B, Hirsch M. EnhanceNet: Single image super-resolution through automated texture synthesis. In: Proceedings of the 2016 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2016. 4501−4510

53 Bei Y J, Damian A, Hu S J, Menon S, Ravi N, Rudin C. New techniques for preserving global structure and denoising with low information loss in single-image super-resolution. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, USA: IEEE, 2018. 987−994

54 Wang Z H, Chen J, Hoi S C H. Deep learning for image super-resolution: A survey. arXiv:1902.06068, 2019

55 Haut J M, Fernandez-Beltran R, Paoletti M E, Plaza J, Plaza A, Pla F. A new deep generative network for unsupervised remote sensing single-image super-resolution. *IEEE Transactions on Geoscience and Remote sensing*, 2018, **56**(11): 6792−6810

56 Liu H, Fu Z L, Han J G, Shao L, Liu H S. Single satellite Imagery simultaneous super-resolution and colorization using multi-task deep neural networks. *Journal of Visual Communication & Image Representation*, 2018, **53**: 20−30

57 Chen Y, Tai Y, Liu X M, Shen C H, Yang J. FSRNet: End-to-end learning face super-resolution with facial priors. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: CVPR, 2018. 2492−2501

58 Bulat A, Tzimiropoulos G. Super-FAN: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 109−117

59 Yu X, Fernando B, Ghanem B, Poriklt F, Hartley R. Face super-resolution guided by facial component heatmaps. In: Proceedings of 2018 ECCV European Conference on Computer Vision. Munich, Germany: Springer, 2018. 219−235

张 宁 中国科学院大学电路与系统专业硕士研究生. 2017 年获东北大学学士学位. 主要研究方向为图像处理, 深度学习, 遥感图像超分辨率重构.
E-mail: neuq2013zn@163.com
(**ZHANG Ning** Master student in circuits and systems at University of Chinese Academy of Sciences. She received her bachelor degree from Northeastern University in 2017. Her research interest covers image processing, deep learning, and remote sensing image super-resolution.)

王永成 中国科学院长春光学精密机械与物理研究所研究员. 2003 年获吉林大学学士学位, 2010 年获中国科学院研究生院博士研究生学位. 主要研究方向为人工智能, 图像工程以及空间有效载荷嵌入式系统. 本文通信作者. E-mail: wyc_dyy@sina.com
(**WANG Yong-Cheng** Researcher of Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. He received his bachelor degree from Jilin University in 2003 and received his Ph.D. degree from Chinese Academy of Sciences in 2010. His research interest covers artificial intelligence, image engineering, and embedded system of space payload. Corresponding author of this paper.)

张 欣 中国科学院大学光学工程专业博士研究生. 2016 年获东北大学学士学位. 主要研究方向为深度学习和遥感图像分类识别.
E-mail: zhangxin162@mails.ucas.ac.cn
(**ZHANG Xin** Ph.D. candidate in optical engineering at University of Chinese Academy of Sciences. She received her bachelor degree from Northeastern University in 2016. Her research interest covers deep learning and remote sensing image classification.)

徐东东 中国科学院大学光学工程专业博士研究生. 2013 年获山东大学学士学位, 2015 年获哈尔滨工业大学硕士研究生学位. 主要研究方向为深度学习, 图像融合及嵌入式系统.
E-mail: sdwhxdd@126.com
(**XU Dong-Dong** Ph.D. candidate in optical engineering at University of Chinese Academy of Sciences. He received his bachelor degree from Shandong University in 2013 and master degree from Harbin Institute of Technology in 2015. His research interest covers deep learning, image fusion, and embedded system.)