

基于深度学习的单幅图像超分辨率重建算法综述

李佳星¹ 赵勇先^{2,3} 王京华^{1,4}

摘 要 单幅图像超分辨率 (Single image super-resolution, SISR) 重建是计算机视觉领域上的一个重要问题, 在安防视频监控、飞机航拍以及卫星遥感等方面具有重要的研究意义和应用价值. 近年来, 深度学习在图像分类、检测、识别等诸多领域中取得了突破性进展, 也推动着图像超分辨率重建技术的发展. 本文首先介绍单幅图像超分辨率重建的常用公共图像数据集; 然后, 重点阐述基于深度学习的单幅图像超分辨率重建方向的创新与进展; 最后, 讨论了单幅图像超分辨率重建方向上存在的困难和挑战, 并对未来的发展趋势进行了思考与展望.

关键词 单幅图像超分辨率, 计算机视觉, 深度学习, 神经网络

引用格式 李佳星, 赵勇先, 王京华. 基于深度学习的单幅图像超分辨率重建算法综述. 自动化学报, 2021, 47(10): 2341–2363

DOI 10.16383/j.aas.c190859

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

LI Jia-Xing¹ ZHAO Yong-Xian^{2,3} WANG Jing-Hua^{1,4}

Abstract Single image super-resolution (SISR) reconstruction is an important problem in the field of computer vision. It has important research significance and application value in security video surveillance, aircraft aerial photography and satellite remote sensing. In recent years, deep learning has made a breakthrough in many fields such as image classification, detection and recognition, and promoted the development of image super-resolution reconstruction technology. This paper first introduces the common public image datasets for single image super-resolution reconstruction. Then, the innovation and progress of single image super-resolution reconstruction based on deep learning are emphasized. Finally, the difficulties and challenges in the single image super-resolution reconstruction are discussed, and the future development trend is discussed.

Key words Single image super-resolution (SISR), computer vision, deep learning, neural network

Citation Li Jia-Xing, Zhao Yong-Xian, Wang Jing-Hua. A review of single image super-resolution reconstruction algorithms based on deep learning. *Acta Automatica Sinica*, 2021, 47(10): 2341–2363

单幅图像超分辨率 (Single image super-resolution, SISR) 重建是根据一张低分辨率 (Low resolution, LR) 图像恢复出高分辨率 (High resolution, HR) 图像的过程, 研究超分辨率的学者将图像重建

成 $\times 2$ 、 $\times 3$ 、 $\times 4$ 、 $\times 8$ 这 4 种尺度的较多, 其中 $\times 2$ 代表将图像的边长放大 2 倍, 即像素密度增加 4 倍, $\times 3$ 、 $\times 4$ 和 $\times 8$ 与其同理. 如何保证重建后图像的质量更接近 Ground truth 图像成为了研究热点, 其目标如下:

$$\hat{y} = \arg \min_y [L(F_{sr}(x), y)] + \lambda \Phi(y) \quad (1)$$

其中, x 为 LR 图像, y 为对应的 Ground truth 图像, $F_{sr}(x)$ 为运用某种算法重建后的 HR 图像, λ 为平衡参数, $\Phi(y)$ 为正则化项.

超分辨率重建在卫星、遥感、天文学、安防、生物医学等诸多领域以及恢复珍贵的历史图像资料上起到了非常重要的作用. 目前在计算机视觉上单幅图像超分辨率已经成为一个专门的学术问题, 吸引了国内外众多学者的关注与研究.

在深度学习未兴起前, 经典的单幅图像超分辨率算法占据主导地位, Lanczos 重采样^[1] 和双三次插值^[2] 得到了广泛的应用, 但采用插值方法有时会

收稿日期 2019-12-17 录用日期 2020-06-11

Manuscript received December 17, 2019; accepted June 11, 2020

国防基础科研计划 (JCKY2019411B001), “111” 计划 (D17017), 露泉创新基金 (LQ-2020-01) 资助

Supported by Defense Industrial Technology Development Program (JCKY2019411B001), the “111” Project of China (D17017) and Innovation Foundation of Luquan (LQ-2020-01)

本文责任编辑 白翔

Recommended by Associate Editor BAI Xiang

1. 长春理工大学机电工程学院 长春 130022 2. 中国科学院长春光学精密机械与物理研究所 长春 130033 3. 中国科学院大学北京 100049 4. 长春理工大学跨尺度微纳制造教育部重点实验室 长春 130022

1. College of Mechanical and Electric Engineering, Changchun University of Science and Technology, Changchun 130022 2. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033 3. University of Chinese Academy of Sciences, Beijing 100049 4. Ministry of Education Key Laboratory for Cross-Scale Micro and Nano Manufacturing, Changchun University of Science and Technology, Changchun 130022

导致图像边缘和细节模糊,因此其他传统算法也被相继提出^[3-5],有效地增强了图像的质量.经典的超分辨率重建算法需要很多先验知识,且要求研究者具有深厚的专业知识储备.随着深度学习的兴起,由于该技术不需要过多的先验知识,且重建后的图像质量优于传统算法,因此得到了广泛的关注. Dong等^[6]首先将卷积神经网络应用到图像超分辨率重建技术中,提出了超分辨率重建卷积神经网络 (Super-resolution convolutional neural network, SRCNN),图1为SRCNN的模型框架图,虽然只有三层神经网络,但相比于经典超分辨率算法,取得了显著的效果. SRCNN的出现,吸引了国内外学者将神经网络的各种变体应用到图像超分辨率研究中,包括卷积神经网络^[7]、对抗神经网络^[8]以及二者的结合^[9]等.

本文从超分辨率图像数据集、基于深度学习的单幅图像超分辨率重建的研究进展、图像质量评估准则、实验结果与分析以及存在的问题与挑战出发,对单幅图像超分辨率重建进行了全面综述,系统的回顾了单幅图像超分辨率重建技术的发展.

1 超分辨率图像数据集

为了方便比较算法的优越性,许多学者在提出超分辨率重建算法的同时会公布数据集,以供其他学者们使用. 目前国际上已经形成了专用于超分

率图像公共基准数据集,如使用较为广泛的 Set5^[10]、Set14^[11]、Urban100^[12]、General-100^[13]、BSDS300^[14]、BSDS500^[15]、Manga109^[16]、T91^[17]等,这些图像来源不同,有人物、动物、风景、建筑、生活上常见的景象以及虚拟合成的动漫图像等,且每个数据库图像的分辨率、质量各不相同,各个数据集的图片数量与格式也并不同,有JPG、PNG与BMP等图像格式,表1和表2列出了几种常用超分辨率图像数据集的详细信息. 除了以上那些图像数据集, DIV2K^[18]、L20^[19]、OutdoorScene^[20]、PIRM^[21]、ImageNet^[22]、MSCOCO^[23]、VOC2012^[24]、CelebA^[25]、LSUN^[26]、WED^[27]、Flickr2K^[28]、City100^[29]和SR-RAW^[30]等数据集也被应用在图像超分辨率重建中,大大扩充了数据集的数量与种类,有利于检验各个模型的泛化能力. 使用同样的基准数据集进行测试,得到的测试结果会更具有公平性与说服力,图2所示为超分辨率数据集的示例图像.

2 基于深度学习的单幅图像超分辨率重建的研究进展

分辨率越高,图像所包含的信息越多,将单幅低分辨率图像恢复成高分辨率图像是一个不适定问题,得到了广泛的关注. 传统的典型方法如基于空间域、频域、非均匀插值、迭代反投影、凸集投影法以及基于统计学习与基于字典学习的方法等,都为

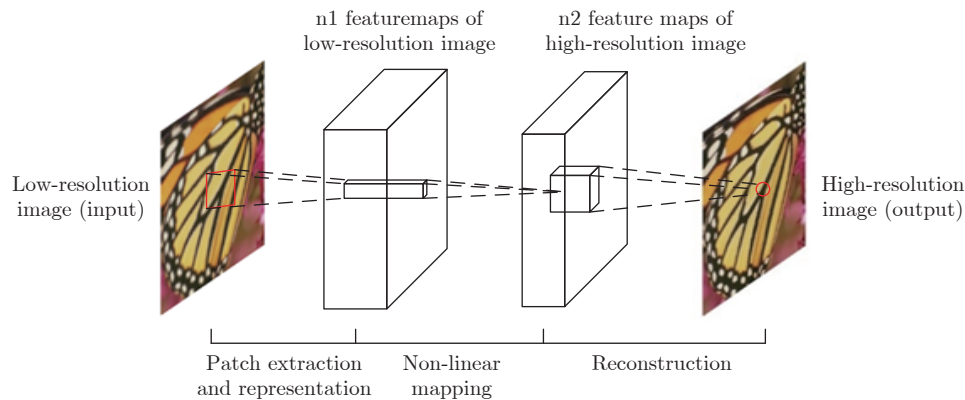


图1 SRCNN网络结构^[6]
Fig.1 The SRCNN network structure^[6]

表1 常用超分辨率训练数据集
Table 1 Widely used Super-resolution training datasets

数据集名称	图像数量	图像格式	图像描述	平均像素 ^[31]	平均分辨率
BSDS200 ^[15]	200	JPG	BSDS500的子集用于训练	154, 401	(432, 370)
T91 ^[17]	91	PNG	车、人脸、水果、花等	58, 853	(264, 204)
General-100 ^[13]	100	BMP	人物、动物、日常景象等	181, 108	(435, 381)

表 2 常用超分辨率测试数据集
Table 2 Widely used Super-resolution testing datasets

数据集名称	图像数量	图像格式	图像描述	平均像素	平均分辨率
Set14 ^[11]	14	PNG	人物、动物、自然景象	230, 203	(492, 446)
BSDS100 ^[15]	100	JPG	BSDS500的子集用于测试	154, 401	(432, 370)
Set5 ^[10]	5	PNG	人物、动物、昆虫等	113, 491	(313, 336)
Urban100 ^[12]	100	PNG	建筑物	774, 314	(984, 797)
Manga109 ^[16]	109	PNG	漫画	966, 11	(826, 1 169)



图 2 超分辨率数据集示例
Fig. 2 Examples of super-resolution datasets

这个领域做出了突破,但大多数算法都需要一些先验知识. 2016 年, SRCNN^[6] 的提出,很好地将神经网络学习迁移到超分辨率重建领域. 通过训练 LR-HR 图像对,得到一套神经网络模型,测试的结果优于经典算法,且无需过多的先验知识. 本部分从基于有监督学习、弱监督学习、无监督学习以及针对特定领域的图像超分辨率重建这 4 个模块出发,系统的阐述深度学习在超分辨率领域的发展进程与重大突破.

2.1 基于有监督学习的超分辨率重建

有监督学习是指利用一组带有标签的数据,学习从输入到输出的映射,然后将这种映射关系应用到未知数据上,达到分类或回归的目的. 由于有监督学习方法相对简单且优于多数传统算法,因此受到广大学者们的青睐. 对于超分辨率领域,大多数模型也都是基于有监督学习的方法. 本节从有监督学习角度出发,根据所使用的神经网络类型可以分为基于卷积神经网络、递归神经网络、循环神经网络和对抗神经网络;根据所使用的机制可分为基于

反馈机制和通道注意力机制;此外,还有基于损失函数类型、上采样层类型的不同,针对多尺度超分辨率重建的方法共分为 9 个方面,详细阐述有监督学习在超分辨率重建中的应用发展.

2.1.1 基于卷积神经网络的超分辨率重建

由于卷积神经网络在图像分类、识别等领域有着出色的表现,因此 Dong 等^[6]将卷积神经网络(Convolutional neural network, CNN)应用到超分辨率重建中,对采集到的 HR 图像以一定的采样因子下采样,得到的图像称为 LR 图像. 再利用双三次插值的方法重建成与原来 HR 图像同等尺寸大小,将其作为输入,经过构建的卷积层 Conv1+激活函数 Relu1—卷积层 Conv2+激活函数 Relu2—卷积层 Conv3 框架的学习,与相对应的 HR 图像求损失函数,使损失函数不断减小,期间通过反向传播调整各卷积层的权值,直至损失函数收敛,使重建后的 SR 图像质量逼近于 HR 图像质量,从而达到重建的目的. SRCNN 的提出,具有里程碑式的意义,但它也有一些不足,如太过依赖小图像区域的上下文信息、训练时收敛较慢、网络仅适用于单一

采样尺度,若换成另一尺度采样,则需要重新训练模型.针对这几种缺点,同年 Dong 等^[13]又提出了快速超分辨率重建卷积神经网络(Fast super-resolution convolutional neural network, FSRCNN),如图 3 所示,FSRCNN 在 SRCNN 基础上做了一些改进, SRCNN 的网络输入是将 LR 图像进行双三次插值放大为 Ground truth 图像尺寸,这样会增加需提取的图像特征信息、网络参数和算法时间复杂度,而 FSRCNN 网络的输入不需要将 LR 图像放大,仅将 LR 图像作为输入,网络的最后一层采用反卷积上采样层,此时再将图像放大为 Ground truth 图像尺寸,大大减少时间的消耗;如图 4 所示,如果训练不同上采样倍率的模型,只需将最后一层的上采样层微调;另一处改进是网络模型增加了网

络层数并使用了较小的卷积核,使网络更深,学习到的特征更多.

Kim 等^[7]提出了图 5 所示的超分辨率重建极深卷积神经网络(Very deep convolutional network for super-resolution, VDSR)模型,与 SRCNN 相比, VDSR 增加了神经网络的层数,可以提取更多的特征图,使重建后的图像细节更丰富,且随着层数的加深,感受野也随之变大,解决了 SRCNN 依赖小图像区域的上下文信息的问题;该模型采用残差学习的方法,对输入的 LR 图像进行双三次插值得到的图像添加到模型的最后一层上,使网络仅仅学习 HR 图像与 LR 图像的差异部分,有效地解决了收敛困难的问题;该模型的训练数据集并不单单使用一种采样因子的 LR 图像作为输入,而是包含

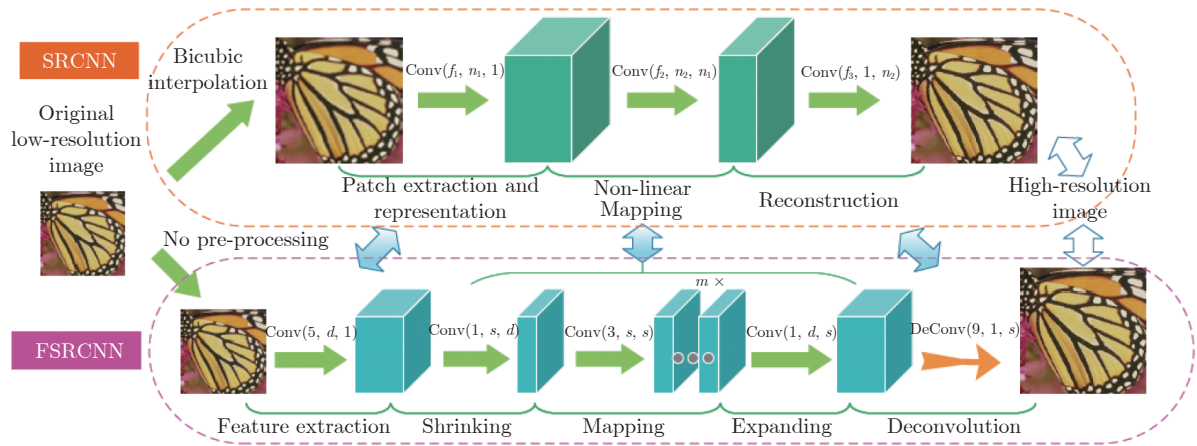


图 3 FSRCNN 网络结构与 SRCNN 网络结构的对比^[13]

Fig.3 Comparison of FSRCNN network structure and SRCNN network structure^[13]

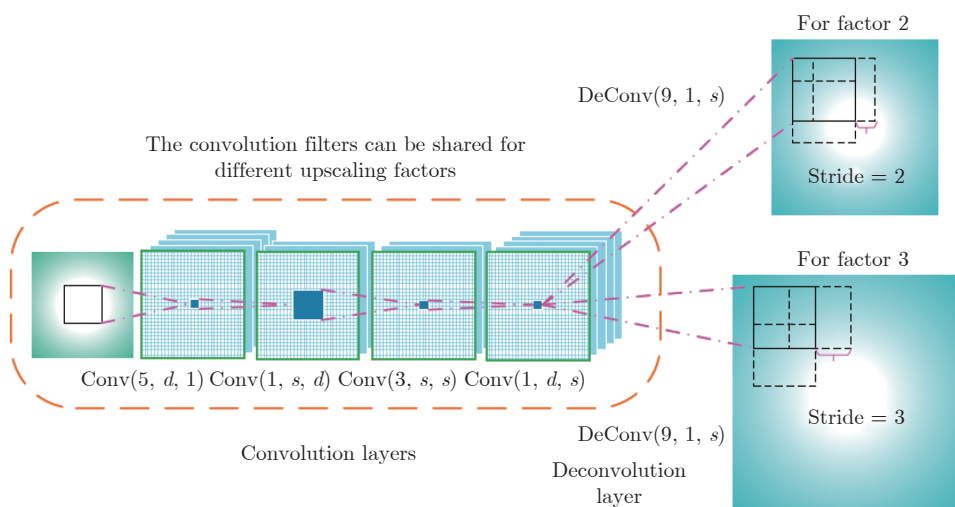


图 4 FSRCNN 网络卷积层与反卷积层的具体结构^[13]

Fig.4 The concrete structure of convolution layer and deconvolution layer of FSRCNN network^[13]

了多种采样因子的 LR 图像, 这样训练出来的网络可以测试多种采样因子的 LR 图像. Hu 等^[32]对 SR-CNN 提出了两点改进, 首先采用随机线性纠正单元 (Randomized rectified linear unit, RReLU) 去避免原有网络学习中对图像某些重要信息的过压缩, 然后用 Nesterov 加速梯度 (Nesterov's accelerated gradient, NAG) 方法去提升网络的收敛速度, 并且避免了网络在梯度更新的时候产生较大的震荡.

以往模型都是先利用双三次插值等方法将 LR 图像上采样到 HR 空间, 再在 HR 空间上提取特征, 从而重建成最后的 HR 图像, 这种方法会增加计算复杂度. 为此, Shi 等^[33]提出了图 6 所示的高效子像素卷积神经网络 (Efficient sub-pixel convolutional neural network, ESPCN) 模型, 该模型是第一个能够在单独的 K2 GPU 上实时处理 1080p 视频的卷积神经网络, 其核心思想是使用亚像素卷积层来替代反卷积层实现上采样操作, 网络的输入是未放大尺寸的 LR 图像, 提取特征也都是在低维空间下

进行, 通过三个卷积层后, 得到通道数为 r^2 的特征图 (Feature map), 其中 r 为上采样的倍率, 然后再将尺寸为 $H \times W \times r^2$ 的特征图重新排列成 $rH \times rW \times 1$ 的高分辨率图像. 若需要获得不同上采样倍率的图像, 只需改变 r 的值. 该模型的优点是运行时间比以往基于卷积神经网络的模型快一个数量级, 缺点是当改变倍率时, 需要重新训练模型.

Mao 等^[34]提出了图 7 所示的极深残差编解码网络 (Very deep residual encoder-decoder networks, RED-Net) 模型, 该网络模型由卷积层和反卷积层组成, 且卷积层与反卷积层呈对称分布, 即编码-解码结构. 卷积层起到提取特征与去噪的作用, 反卷积层先接收去噪后的特征图像, 再将其重建成高分辨率图像, 这样会使图像更为清晰. RED-Net 模型也采用了跳跃连接, 但与 ResNet^[35]不同, 该跳跃是从卷积层到对应的反卷积层之间的跳跃连接.

Tai 等^[36]提出了图 8 所示的 MemNet 网络, 与传统的神经网络如 VDSR^[7]、DRCN^[37]不同, 这是一

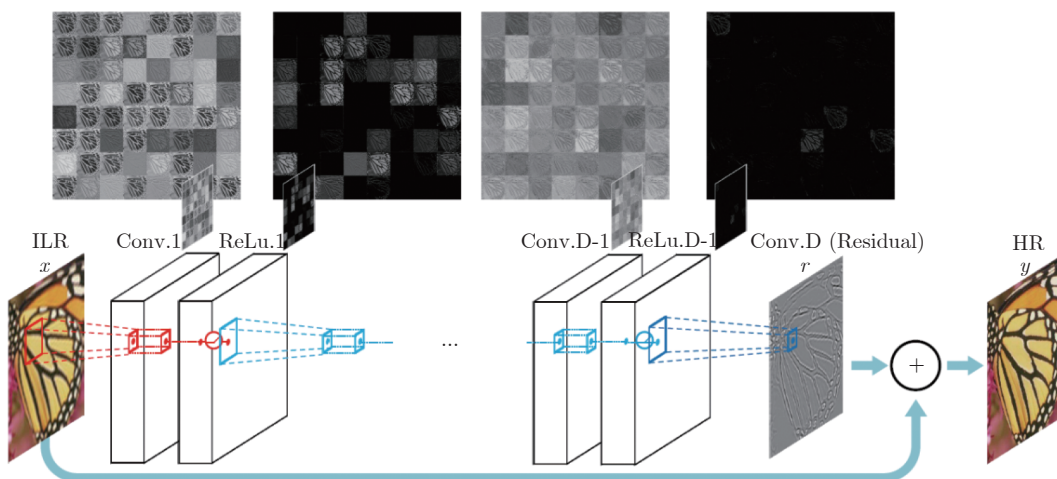


图 5 VDSR 网络结构^[7]

Fig.5 The VDSR network structure^[7]

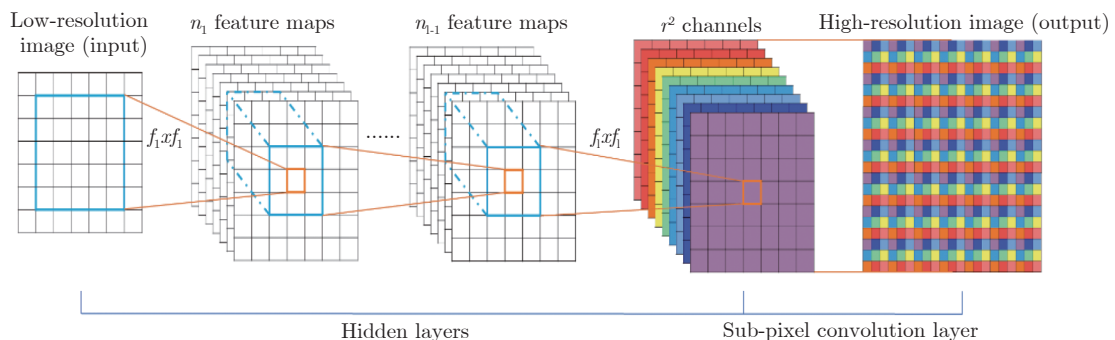
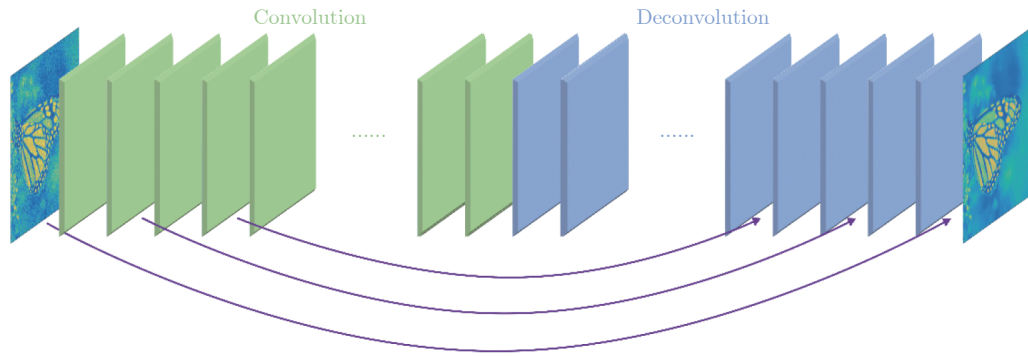
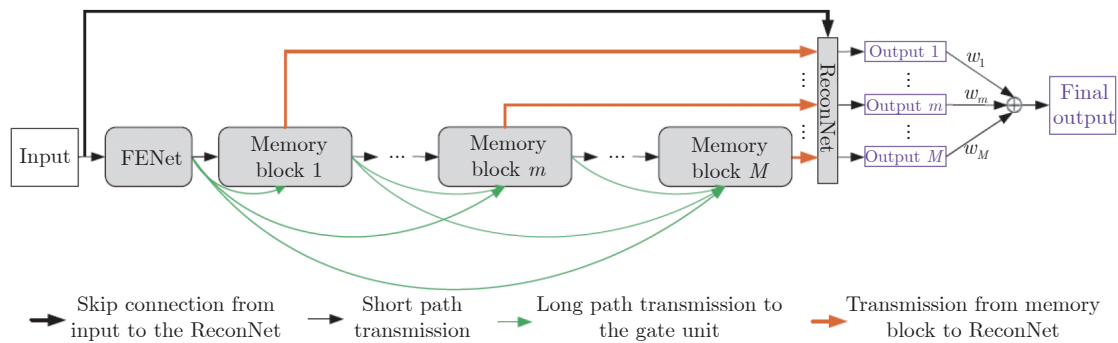


图 6 ESPCN 网络结构^[33]

Fig.6 The ESPCN network structure^[33]

图 7 RED-Net 结构^[34]Fig.7 The RED-Net structure^[34]图 8 MemNet 网络结构^[36]Fig.8 The MemNet network structure^[36]

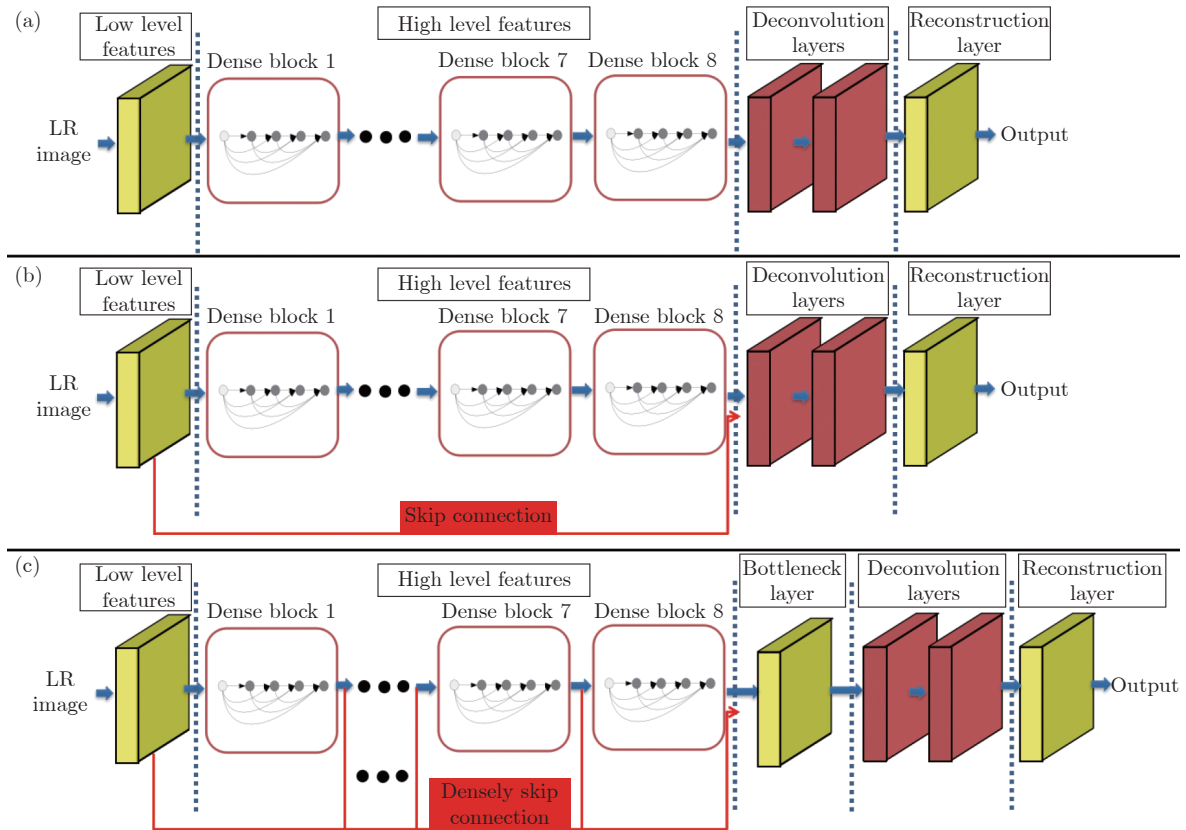
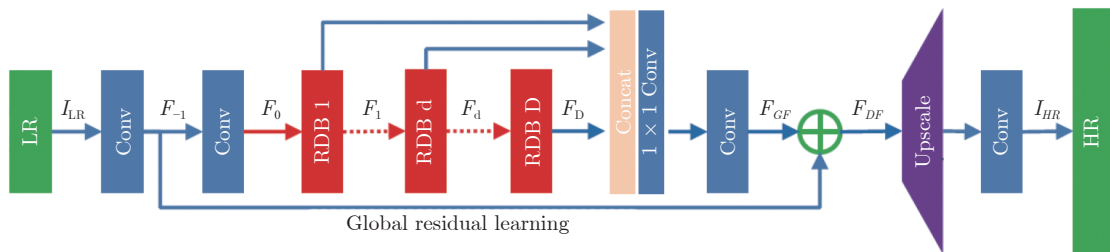
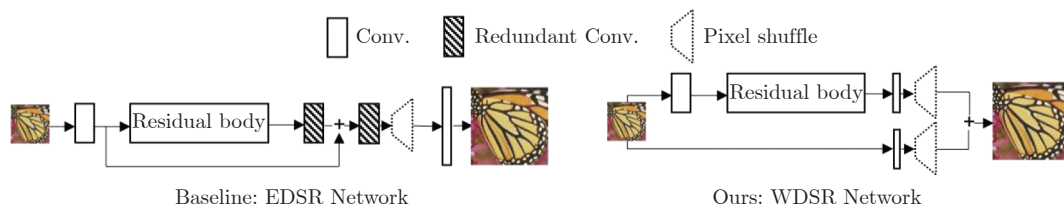
个长期记忆的模型. 传统的神经网络基本上都是单向传播, 当层数非常多的时候, 越靠后的层接收到的信号越微弱, MemNet 网络由特征提取网络 (Feature extraction net, FENet)、若干个记忆模块 (Memory block) 与重建网络 (Reconstruction net, ReconNet) 组成, 其中 FENet 由卷积层构成; 与常规神经网络不同, 每个 Memory block 都参与了最终输出, 即每个记忆模块不仅与下一个记忆模块连接, 而且直接连接 ReconNet 层; ReconNet 是个卷积核为 1×1 的卷积层, 负责将所有记忆模块的存储信息分别输出, 所有输出分别赋予不同的权重, 进行相加得到最后的输出结果.

与 ResNet 网络^[35] 直接将输入与最后一层串联起来不同, DenseNet^[38] 是将每一层与其他层都串联起来, 这样更能保留原始图像的特征信息, 即使传递到后面的层, 图像信息也不易丢失, 并且很好地解决了梯度消失的问题. 考虑到 DenseNet 的优点, Tong 等^[39] 提出图 9 所示的 SRDenseNet 网络, 该网络由低水平特征提取层、8 个密集模块 (Dense block)、2 个反卷积上采样层和 1 个重构层组成, 其中每个密集模块就是采用了 DenseNet 的结构.

Zhang 等^[40] 将残差模块 (Residual block) 与密

集模块结合起来, 形成了残差密集模块 (Residual dense block, RDB), 即图 10 所示的 RDN 网络模型. 该网络模型主要由 4 个模块组成, 分别为浅特征提取网络 (Shallow feature extraction net, SFENet)、残差密集模块、密集特征融合 (Dense feature fusion, DFF) 和上采样网络 (Up-sampling net, UPNet), 其中 SFENet 包含 2 个卷积层, 具有提取输入图像特征的作用; RDB 模块去掉了 DenseNet 每个模块中的批正则化和池化层, 上一个 RDB 模块与下一个 RDB 模块中的 1 到 d 层做了局部特征融合 (Local feature fusion, LFF), 每一层与每一模块连接都很紧密, 更好的保证了信息流的贯通; DFF 通过 Concat 层将每个 RDB 模块得到的特征图都串联起来; UPNet 实现将特征图放大的操作.

Yu 等^[41] 在 2018 年提出了 WDSR 模型, 并取得了当年 NTIRE 竞赛^[42] 的冠军. 如图 11 所示, WDSR 模型网络由卷积模块、残差模块和像素重组 (Pixel shuffle) 模块三部分组成, 其中卷积模块、残差模块与常规神经网络相同, Pixel shuffle 采用了 ESPCN 模型^[33] 提出的亚像素层来实现上采样的操作. 与 2017 年 NTIRE 竞赛^[28] 的冠军 EDSR 模型^[43] 相比, WDSR 模型一方面去除了很多冗余的卷积层,

图 9 SRDenseNet 网络结构^[38]Fig.9 The SRDenseNet network structure^[38]图 10 RDN 网络结构^[40]Fig.10 The RDN network structure^[40]图 11 EDSR 与 WDSR 网络结构的对比^[41]Fig.11 Comparison of EDSR network structure and SRCNN network structure^[41]

Yu 等认为残差模块已经涵盖了这些卷积层的效果, 去除了这些冗余层会使计算更快, 并且效果也没有下降; 另一方面改进了残差模块. Yu 等提出了 2 个

版本的 WDSR, 分别是图 12 所示的 WDSR-A 与 WDSR-B^[41], 2 个版本只在残差模块上有区别. 与 EDSR 模型的残差模块相比, EDSR 的 Relu 激活

函数是在两个卷积运算中间,而且卷积核的个数较少;而 WDSR-A 是在不增加计算开销的前提下,增加 Relu 激活函数前面的卷积核个数以增加特征图的宽度,这样效果会更好. WDSR-B 在 WDSR-A 的基础上,将 Relu 后的大卷积核拆分成两个小卷积核,既减少了模型参数,又可以在同样计算开销的前提下获得更多未使用激活函数前的特征图.除此之外, WDSR 还采用了权重归一化 (Weight normalization) 的策略,实验表明权重归一化可以使用比以往模型高 10 倍的学习率,而且能得到更高的测试准确率.

许多基于卷积神经网络的 SISR 模型使用失真导向 (Distortion-oriented) 的损失函数,此类模型很难恢复真实图像纹理和细节,生成的图像看起来较模糊,因此恢复真实纹理和细节在图像超分辨率

域仍然是一项挑战,目前这方面的工作有 SRGAN^[8]、EnhanceNet^[44] 和 SFTGAN^[20] 等,但是在生成图像时通常会产生伪影,整幅图像看起来总有一些不自然. Soh 等^[45] 对此类问题提出了 FRSR 模型和图 13 所示的 NatSR 模型,可生成真实纹理和自然细节,获得高视觉质量.该模型添加了图 14 所示的自然流形鉴别器 (Natural manifold discriminator, NMD) 与基于不规则残差学习的卷积神经网络结构.从图像评价指标上看, NatSR 算法并不是最好的,但在视觉效果上 NatSR 表现更好.

2.1.2 基于递归神经网络的超分辨率重建

在超分辨率问题中,为了使生成的图像质量高,多数卷积神经网络不使用池化层,但随着网络的加深,将会增加更多的参数,容易使网络过拟合;且模型过大,难以存储和重现.为了解决以上问题,考虑

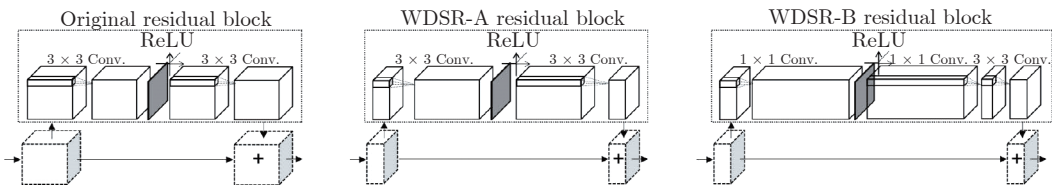


图 12 三种残差模块的对比^[41]
Fig.12 Comparison of three residual blocks^[41]

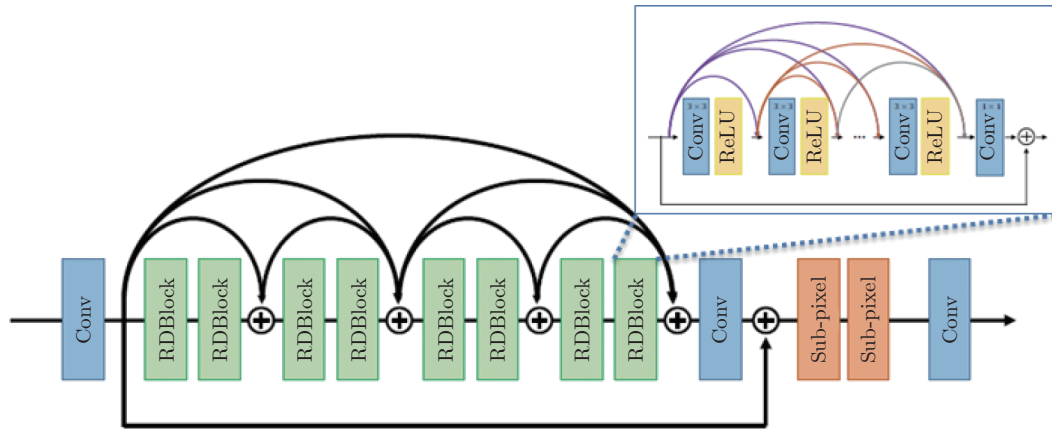


图 13 NatSR 网络结构^[45]
Fig.13 The NatSR network structure^[45]

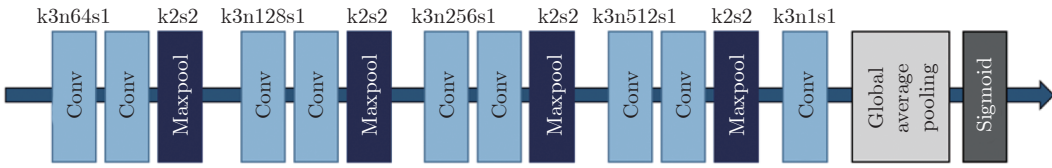


图 14 NMD 结构^[45]
Fig.14 The NMD structure^[45]

到递归神经网络是多次递归其模型中的一层或几层神经网络, 可有效减少网络的参数, 因此 Kim 等^[37] 采用了递归神经网络作为其基本网络, 提出了 DRCN 模型, 其使用的递归层都是监督式的; 并采用了跳跃连接结构, 有效地解决了梯度消失/爆炸的问题. Tai 等^[46] 提出的 DRRN 延续了 VDSR^[7] 认为的网络层数越多、超分辨率重建的效果越好的观点, 将网络层数增加到 52 层, 为了减少网络参数, 也采取了递归操作与残差学习的方式. 在网络结构上与 DRCN^[37] 有所不同, 采用间隔层权重共享的操作, 而不像 DRCN^[37] 的 Inference network 里每一层都是权重共享. 与具有 20 层深度的 VDSR 模型相比, 效果虽然有所提升, 但提升幅度并不大; 相比较于同期的 LapSRN^[47], 网络的效果有待优化, 同时每个递归单元只有 2 个卷积层确实有些不足. Zhou 等^[48] 将递归神经网络与残差学习相结合, 提出一种简洁紧凑型递归残差网络结构, 在重建出同等质量超分辨率图像的前提下, 模型参数数量及计算复杂度分别仅为 VDSR 方法的 1/10 和 $1/(2n^2)$.

2.1.3 基于循环神经网络的超分辨率重建

目前图像超分辨率重建主要朝着两个方向发展, 一是网络越来越深, 但是随着网络层数的增加也会引起模型参数过多等问题; 二是残差学习, 包括全局残差学习、局部残差学习以及二者相结合的形式, 虽然 DRRN、DRCN 将递归神经网络与残差学习相结合, 但仍会导致重建速度变慢. Han 等^[49] 引入了循环神经网络 RNN, 提出基于双循环网络状态的 DSRN 模型. 该模型通过延迟反馈机制, 在两个方向 (LR 到 HR 和 HR 到 LR) 之间交换循环信号, 并将这两种状态预测的特征联合起来进行最终的预测, 取得了不错的效果.

2.1.4 基于对抗神经网络的超分辨率重建

基于卷积、递归和循环这三种神经网络都是以峰值信噪比作为导向的 SISR 模型, 其网络生成的图像虽然峰值信噪比很高, 但感观质量较差. 为生成符合人眼感观的图像, Ledig 等^[8] 提出了 SRGAN 模型, 网络主体采用对抗神经网络 (Generative adversarial net, GAN)^[50], 损失函数采用感知损失与对抗损失之和. 虽然峰值信噪比不是最高, 但是 SRGAN 模型产生出的图像更加自然清晰, 更符合人眼的视觉效果. 尽管 SRGAN 取得了很好的视觉效果, 但随着网络的加深, 批归一化层可能会使图像出现伪影. Wang 等^[51] 对 SRGAN 模型进行了改进, 提出了 ESRGAN 模型. ESRGAN 在 SRGAN^[8] 的基础上去掉了批归一化层, 并引入残差密集模块 (Residual-in-residual dense block) 作为网络的基本单元,

且采用 RaGAN^[52] 的思想, 将判别器由原来预测图像的绝对真实性改为了相对真实性, 即判别器接收到生成器传来的图像时, 不再单单预测这幅图像是否真实的概率, 而是预测这幅图像比虚假图像更真实的概率. SRGAN^[8] 是使用激活后的特征求感知损失, 而 ESRGAN 是使用激活前的特征求感知损失. 这样做克服了两个缺点: 1) 激活后的特征是稀疏的, 而使用激活前的特征求感知损失, 会得到更多的信息; 2) 若使用激活后的特征作为输入, 则重建的图像与真实图像对比会出现亮度不一致的情况. Wang 等^[20] 提出了 SFTGAN 模型, 在模型中新添加了空间特征转换层 (Spatial feature transform, SFT), 将有效的先验信息与神经网络相结合进行端到端的训练, SFT 层通过对模型中的每个中间特征进行空间仿射变换来学习参数对, 从而自适应地调节输出. 与 SRGAN^[8] 和 EnhanceNet^[44] 模型相比, 重建后的图像在视觉上显得更加自然.

2.1.5 基于反馈机制的超分辨率重建

根据应用的网络类型分类可以分为基于卷积、递归、循环和对抗神经网络, 相关方法前文已经介绍. 有学者从另一个角度出发, 在神经网络中添加反馈机制和注意力机制, 更有助于提升图像质量. 对此, Haris 等^[53] 将传统算法迭代反投影 (Iterative back projection, IBP) 与深度神经网络相结合, 提出了 DBPN 模型. 考虑到近年来提出的超分辨率模型中, 没有很好地解决 LR 图像与 HR 图像的相互依赖关系, Haris 等认为超分辨率网络模型缺乏反馈机制, 不能像人类视觉系统那样利用反馈机制来指导任务, 为此提出一种上采样和下采样交替式进行的网络, 为每个阶段的错误映射提供反馈机制, 思想上很像传统的迭代反投影算法, 因而取名为 DBPN 网络, 该模型成为了 $\times 8$ 尺度上的先进算法. Li 等^[54] 提出了一种图像超分辨率反馈网络 SRF-BN, 高阶层的信息通过反馈连接自顶向下的提供反馈流. 同时, 这种具有反馈连接的递归结构具有较强的重建能力, 且只需要较少的参数. 该模型还采用了一种基于课程的训练策略, 将多张越来越难重构的 HR 图像作为连续迭代的目标输入网络. 该方法使网络能够逐步地学习复杂的退化模型, 而其他方法只能学习退化模型中的一种模式.

2.1.6 基于通道注意力机制的超分辨率重建

大多数基于神经网络的超分辨率 (Super-resolution, SR) 方法没有充分利用原始 LR 图像的信息, 从而得到的结果不太理想, 且多数基于神经网络的模型主要专注于设计更深或是更宽的网络, 以学习更具有判别力的高层特征, 却很少发掘层间特征的

内在相关性,从而阻碍了网络的特征提取能力.为此,Zhang等^[55]将通道注意力机制(Channel attention, CA)和残差块相结合,提出了RCAN网络模型,RCAN延续了神经网络深度越深,效果越好的理念.该模型主要由4个部分组成,分别是浅层特征提取模块、残差中的残差(Residual in residual, RIR)深度特征提取模块、上采样模块和重构图像模块.其中除了RIR模块之外,其余模块与大多数SR网络相同,RIR模块由残差组(Residual group, RG)与一个长跳跃连接组成,每个RG由残差通道注意力模块(Residual channel attention block, RCAB)与短跳跃连接组成,由于RIR模块的优势,可使神经网络的深度超过400层.此外,Zhang等还指出低分辨率图像存在大量低频信息,这些信息可以通过长跳跃连接直接传到网络的最后一层,使网络重点学习高频信息,并减轻了网络的学习负担.

Dai等^[56]提出了一个二阶注意力网络SAN,该模型利用二阶通道注意力机制SOCA来进行相关性学习;同时利用非局部增强残差组(Non-locally enhanced residual group, NLRG)来捕获长距离的空间内容信息.Zhou等^[57]提出了基于特征融合注意网络的单幅图像超分辨率方法,网络模型主要包括特征融合子网络和特征注意子网络.特征融合子网络可以更好地融合不同深度的特征信息,以及增加跨通道的学习能力;特征注意子网络则着重关注高频信息,以增强边缘和纹理.

2.1.7 基于损失函数的改进策略

除了在网络结构进行改进外,所使用的损失函数不同,生成的图像质量也有所不同.常用的损失函数有像素损失、内容损失、对抗损失、纹理损失、总变差损失以及上下文损失等.

1) 像素损失(Pixel loss)

目前多数SR模型均采用像素损失,像素损失一般分为L1损失和L2损失,其表达式分别如下:

$$L_{pixel_l1} = \frac{1}{CHW} \sum_{k=1}^C \sum_{i=1}^H \sum_{j=1}^W |\hat{y}_{i,j,k} - y_{i,j,k}| \quad (2)$$

$$L_{pixel_l2} = \frac{1}{CHW} \sum_{k=1}^C \sum_{i=1}^H \sum_{j=1}^W (\hat{y}_{i,j,k} - y_{i,j,k})^2 \quad (3)$$

其中, C 为图像的通道数,一般为3; H 为图像的高度; W 为图像的宽度; $\hat{y}_{i,j,k}$ 为生成的高分辨率图像的每个像素点; $y_{i,j,k}$ 为Ground truth图像的每个像素点.

虽然在峰值信噪比上取得了显著的效果,但是在主观视觉上发现只运用像素损失生成的图像缺乏

高频信息.为此,Johnson等^[58]用感知损失来代替像素损失,感知损失由内容损失与Gatys等^[59]提出的风格损失组成.

2) 内容损失(Content loss)

采用预训练好的卷积神经网络提取生成图像与Ground truth图像的高阶特征,再逐层求平方差,将所有层加起来,即为内容损失.内容损失表达式如下:

$$L_{content} = \frac{1}{C_l H_l W_l} \times \sqrt{\sum_{k=1}^{C_l} \sum_{i=1}^{H_l} \sum_{j=1}^{W_l} (\phi^{(l)}(\hat{y}_{i,j,k}) - \phi^{(l)}(y_{i,j,k}))^2} \quad (4)$$

其中, C_l 为特征图通道数; H_l 为特征图的高度; W_l 为特征图的宽度; $\phi^{(l)}(\hat{y}_{i,j,k})$ 为使用预训练模型提取生成的高分辨率图像的第 l 层特征图的每个像素点; $\phi^{(l)}(y_{i,j,k})$ 为使用预训练模型提取Ground truth图像的第 l 层特征图的每个像素点.

3) 对抗损失(Adversarial loss)

在超分辨率领域,采用对抗性学习是很简单的,在这种情况下,只需要将SR模型作为一个生成器,另外定义一个鉴别器来判断输入图像是否生成.Ledig等^[8]首先在SRGAN模型中引入了基于交叉熵的对抗损失,具体表达式如下:

$$L_{gan_ce_g} = -\log D(\hat{y}) \quad (5)$$

$$L_{gan_ce_d} = -\log D(y) - \log(1 - D(\hat{y})) \quad (6)$$

其中, $L_{gan_ce_g}$ 为SR模型生成器的对抗损失; $D(\cdot)$ 代表鉴别器; $L_{gan_ce_d}$ 是鉴别器的对抗损失.Wang等^[60]和Yuan等^[61]在网络中引入了基于最小平方差(Least mean error)的对抗损失^[62],其表达式如下:

$$L_{gan_ls_g} = (D(\hat{y}) - 1)^2 \quad (7)$$

$$L_{gan_ls_d} = (D(\hat{y}))^2 + (D(y) - 1)^2 \quad (8)$$

4) 纹理损失(Texture loss)

Gatys等^[63-64]将纹理损失引入到超分辨率重建中,纹理损失利用Gram矩阵 $G^{(l)} \in \mathbf{R}^{C_l \times C_l}$,Gram矩阵表达式如下:

$$G_{i,j}^{(l)}(I) = F_i^{(l)}(I) \cdot F_j^{(l)}(I) \quad (9)$$

其中, $F_i^{(l)}(I)$ 是图像 I 的第 l 层第 i 个通道的特征图; $F_j^{(l)}(I)$ 是图像 I 的第 l 层第 j 个通道的特征图.纹理损失表达式如下:

$$L_{texture} = \frac{1}{C_l^2} \sqrt{\sum_{i=1}^{C_l} \sum_{j=1}^{C_l} (G_{i,j}^{(l)}(\hat{y}) - G_{i,j}^{(l)}(y))^2} \quad (10)$$

5) 总变差损失 (Total variation loss)

为了抑制生成图像 $\hat{y}$ 中的噪声, Ledig 等^[8] 和 Yuan 等^[61] 将总变差损失^[65] 引入超分辨率模型中, 总变差损失为图像相邻像素之间的绝对差值的总和. 其表达式如下:

$$L_{tv} = \frac{1}{CHW} \sum_{k=1}^C \sum_{i=1}^H \sum_{j=1}^W \times \sqrt{(\hat{y}_{i,j+1,k} - \hat{y}_{i,j,k})^2 + (\hat{y}_{i+1,j,k} - \hat{y}_{i,j,k})^2} \quad (11)$$

6) 上下文损失 (Contextual loss)

为了生成逼真的图像, 很多模型采用对抗损失, 但这种模型很难训练. Mechrez 等^[66] 采用上下文损失来取代对抗损失, 与 L1 损失和 L2 损失相结合, 得到了不错的结果.

鉴于每种损失函数都有其各自的侧重点, 将多种损失函数联合起来训练网络, 得到的模型生成的图像无论在客观评价或主观视觉上都会取得不错的效果. 现在的多数模型都是采用多种损失函数联合训练的方式, 如 Sajjadi 等^[44] 提出的 EnhanceNet 模型, 受到 VGG 网络^[67] 的启发, 卷积核全部采用 3×3 的大小, 目的是在保持定量参数的情况下可以构建更深的网络. 损失函数由感知损失、对抗损失和纹理损失组成, 能够产生逼真的纹理.

2.1.8 基于上采样层的改进方法

在网络结构的改进中, 除了可以应用不同的神经网络类型以及结合多种损失函数进行监督外, 还可以对其上采样模块进行改进. 将 LR 图像上采样

为 HR 图像, 上采样模块是必不可少的. 如图 15 所示, 根据上采样模块所在位置的不同, 分为预上采样 (Pre-upscaling)、后上采样 (Post-upscaling)、渐进上采样 (Progressive upscaling) 和迭代上下采样 (Iterative up-and-down sampling) 这 4 种框架. 文献 [6, 36, 37, 46, 68] 均采用预上采样框架; 文献 [8, 38, 43, 49] 均采用后上采样框架; 文献 [47, 69] 采用渐进上采样框架; 文献 [53] 采用迭代上采样框架. 根据上采样层所使用的方法不同, 又可分为插值上采样、反卷积上采样以及亚像素卷积这三种方法, 其中插值上采样可实现任意倍率的放大, 但会带来一定程度的平滑和模糊效果; 反卷积上采样只能实现整数倍率的超分辨率重建; 亚像素卷积的优点是拥有较大的感受野, 但其同样只能实现整数倍率的超分辨率重建. Kim 等^[70] 就当前流行的上采样方法进行了改进, 提出了 EUSR 模型. 该模型采用了增强上采样模块 (Enhanced upscaling module, EUM) 作为上采样层, EUM 利用非线性操作和残差学习, 在 2018 年 NTIRE 超分辨率挑战^[42] 的结构相似度^[71] 评价指标上取得了第 9 名的好成绩.

在 ESPCNN^[33]、EDSR^[43]、RDN^[40] 与 RCAN^[55] 等先进方法中, 通常是在网络末端放大特征图, 但是以上方法需要对每一种缩放因子设计一个特定的上采样模块, 并且上采样模块通常只适用于正整数的缩放因子, 这些缺点限制了 SISR 在现实中的应用. Hu 等^[72] 提出了 Meta-SR 模型, 该模型引进了 Meta-Upscale 模块来取代传统的上采样模块. 受元学习 (Meta-learning) 的启发, 对于任意的缩放因

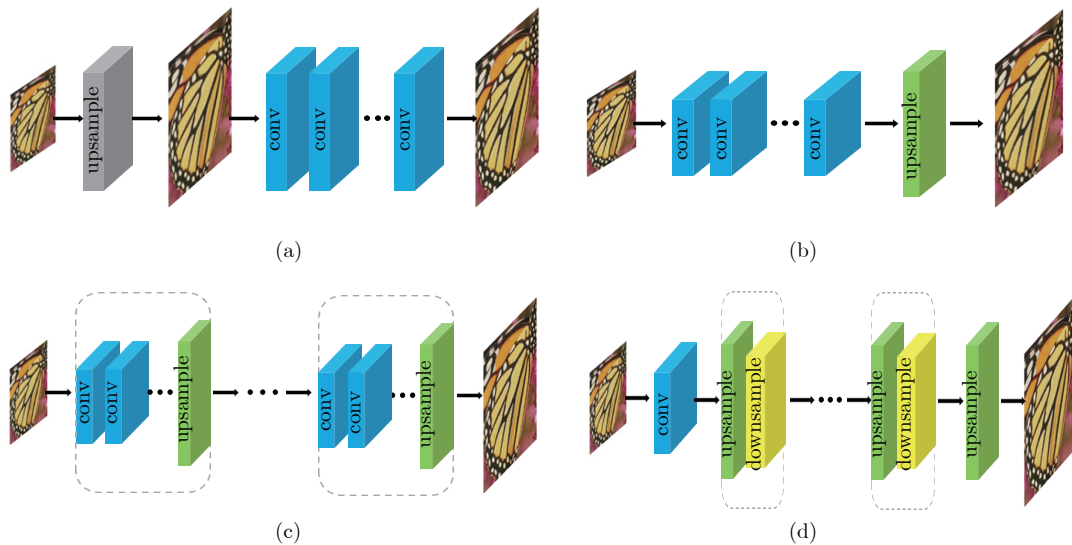


图 15 主流上采样方法 ((a) 预上采样 SR 网络; (b) 后上采样 SR 网络; (c) 渐进上采样 SR 网络; (d) 迭代上下采样 SR 网络)

Fig.15 Mainstream upsampling methods ((a) Pre-upscaling SR network; (b) Post-upscaling SR network; (c) Progressive upscaling SR network; (d) Iterative up-and-down sampling network)

子, Meta-Upscale 模块都能够动态地预测上采样滤波器的权重, 使得单个模型便可以实现任意缩放因子 (包含非整数) 情况下的超分辨率重建。

2.1.9 针对多尺度超分辨率重建

针对多数算法仅有一次上采样的操作, 而无法同时得到多种分辨率图像, Lai 等^[47] 提出级联金字塔结构 LapSRN. LapSRN 网络有两个分支, 一个是特征提取分支, 一个是图像重构分支. 该模型是逐步学习, 输出不同尺度的残差, 得到对应尺度的重构结果, 而不是像其他模型只有一个输出; 与以往的 L2 损失函数不同, 该模型得到的每种分辨率都有一个对应的损失函数, 总的损失函数即为多种分辨率的损失函数之和。

Lim 等^[43] 分别针对单一尺度超分辨率与多尺度超分辨率提出 EDSR 网络与 MDSR 网络, 这两种模型框架主要采用 He 等^[35] 提出的 ResNet 网络, 但与之不同的是, 去掉了其中的批归一化层. 批归一化层的主要作用是加快训练速度, 改善数据变化大因而模型难收敛的问题. 由于框架本身学习的是残差图像, 需要学习的特征减少, 很大程度上提高了训练速度, 因此批归一化层在 ResNet 网络中并不是十分重要, 且批归一化层对每一批数据都要进行归一化处理, 使得 GPU 可用显存减少, 移除批归一化层可起到节约显存的功效. MDSR 网络模型的设计具有权重共享的功能, 既能达到训练多个分辨率的目的, 又能使整套模型的计算量减少, 节约存储空间, Timofte 等凭借该模型赢得了 2017 年 NTIRE 竞赛^[28] 的冠军。

2.2 基于弱监督学习的超分辨率重建

在超分辨率重建领域大多数模型都是基于有监督学习的方法, 鲜少有学者提出弱监督或无监督学习模型. 由于收集相同场景的 LR-HR 图像对是困难的, 因此有监督学习并不适用于现实场景的应用,

弱监督或无监督学习的研究变得极为重要。

Bulat 等^[73] 针对大多数模型是从 HR 图像下采样得到对应的 LR 图像, 并不能真正代表真实场景的 LR 图像的情况, 提出了新的退化模型. 受 CycleGAN^[74] 的启发, Bulat 等将整个网络模型分为 2 段, 一段是用不成对的 LR-HR 图像训练得到由 HR 图像到 LR 图像的 GAN 退化模型, 取代多数模型采用的双三次下采样方法; 另一段是用上一段网络生成的 LR 图像作为输入, 再用成对的 LR-HR 图像训练 GAN 超分辨率网络。

受 CycleGAN^[74] 与 WESPE^[9] 的启发, Yuan 等^[61] 提出了一种弱监督学习模型 CinCGAN, 如图 16 所示该模型由 4 个生成器、2 个鉴别器以及 2 个 CycleGAN 组成, 其中生成器与鉴别器的结构如图 17 所示, 第一个 CycleGAN 用于有噪声的 LR 图像到真实的 Clean LR 图像的映射, 第二个 CycleGAN 的结构与第一个相同, 不同点在于第二个 CycleGAN 是学习 Clean LR 图像到 Clean HR 图像的映射, 且 LR 图像与 HR 图像是不成对的。

2.3 基于无监督学习的超分辨率重建

文献 [75-76] 证明了单张图像内部块之间的信息熵小于多张图像块之间的信息熵, 这就意味着图像内部块比多张图像块拥有着更强的预测能力. 基于此, Shocher 等^[77] 认为仅靠图像的内部块递归循环^[78] 就足以重建高分辨率图像, 提出了图 18 所示的 ZSSR 模型, 这也是第一个无监督学习的 SR 方法. ZSSR 模型将待测试的低分辨率图像下采样后的图像作为输入, 而目标图像则是待测试的低分辨率图像; 网络采用轻量级全卷积结构, 再将待测试的低分辨率图像作为输入得到高分辨率图像. ZSSR 模型由于不依赖外部图像数据集, 因此能处理如不用双三次退化得到的低分辨率图像以及有模糊、噪声、伪影等非理想条件下的图像, 且不需要预训练,

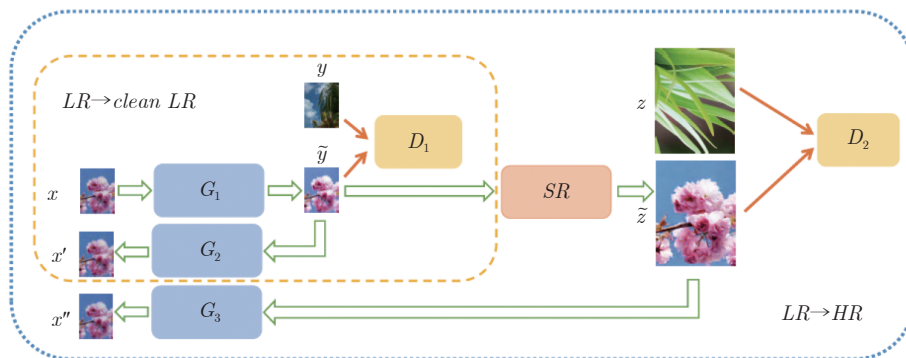


图 16 CinCGAN 网络结构^[61]

Fig.16 The CinCGAN network structure^[61]

2.4 针对特定领域的超分辨率重建

2.4.1 针对盲超分辨率 (Blind Super-resolution) 图像重建的方法

绝大多数图像超分辨率算法都是假设图像退化过程中的下采样模糊核是预先定义或已知的, 如双三次下采样. 此类算法在实际应用时, 如果模糊核是复杂或未知时, 图像超分辨率效果会大大降低. 当模糊核不匹配时, 重建后的超分辨率图像会过于模糊或产生类似图 19 所示的振铃效应; 只有模糊核匹配时, 才会得到较好的图像超分辨率效果, 因此在真实场景中准确估计出模糊核是必要的.

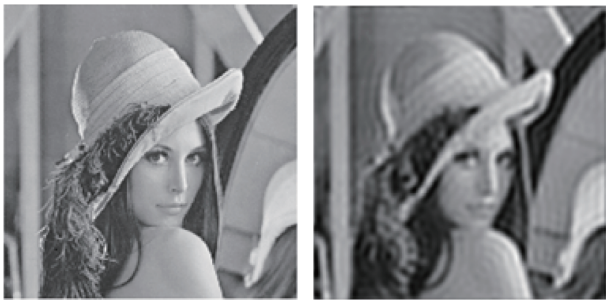


图 19 振铃效应
Fig.19 Ringing effect

Zhang 等^[79]认为采用双三次下采样得到的低分辨率图像不能够代表现实生活中的低分辨率图像, 测试时向网络里直接输入真实的低分辨率图像, 得到的效果往往不是很好. 因此 Zhang 等考虑了噪声和模糊这两个因素, 在退化过程中采用一种维度拉伸的策略, 将这两个关键因素作为网络输入, 提出了 SRMD 模型.

Gu 等^[80]提出了 SFTMD 模型与迭代核校正 (Iterative kernel correction, IKC) 的方法. 为了正确估计出模糊核, Gu 等提出校正函数, 因为一次校正可能不充分或校正过度, Gu 等采用逐步校正模糊核的方式去尽可能地逼近真实核函数. 在模型的搭建上, 考虑到 SRMD^[79]工作存在如下问题: 1) 模糊核没有包含图像的全部信息; 2) 模糊核信息只在

网络的首层考虑到, 网络的深层部分并没有使用到模糊核信息. 为解决上述问题, Gu 等在 SRResNet^[8]网络中增加了空间特征变换层.

Zhang 等^[81]提出一种即插即用图像恢复的 DPSR 模型, 通过设计一个新的 SISR 退化模型来代替盲超分辨率的模糊核估计, 该模型引入能量函数来优化新的退化模型, 采用变量分割的方法实现即插即用, 具有较高的灵活性, 并且采用合成的图片和真实的图片来对算法进行评估, 以实现盲超分辨率的处理.

2.4.2 面向实际场景的图像超分辨率重建的方法

多数超分辨率重建模型都是先获得高分辨率图像, 再通过双三次、高斯或池化等下采样操作得到对应的低分辨率图像, 以此构成图像对. 该方法的优势是省时省力、数据易得, 但在放大远处物体时却丧失了原始数据的准确细节, 直接从 HR 图像上通过算法获得的 LR 图像终究不是真实情况下的输入, 所以在现实中的泛化能力有限. 由于缺乏真实的训练数据, 现有的超分辨率方法在实际场景中表现不佳. 为了解决此类问题, Xu 等^[82]通过模拟图 20 所示的数码相机成像, 构建了图 21 所示的双卷积神经网络来生成真实的训练数据, 其中 X_{raw} 为低分辨率 Raw 格式的图像, X_{ref} 为通过 Dcraw^[83]软件生成的低分辨率彩色图像, 图 22 所示为图像恢复 (Image restoration) 分支, 将学习了空间颜色转换后的图像与 Raw 格式的图像进行特征融合, 以实现更有效的色彩校正. 实验表明使用 Raw 格式的图像更能恢复精确的细节和清晰的轮廓, 在实际场景中表现很好.

同样对于实际场景来说, 计算消耗的问题也不容小觑, 且随着网络深度的增加, 需要更多的训练技巧, 否则训练过程中的数值就会变得不稳定^[43, 84]. He 等^[85]针对此类问题将常微分方程 (Ordinary differential equation, ODE) 与神经网络相结合, 衍生出 LF-block 和 RK-block 两种网络结构, 相对应于常微分方程中的 Leapfrog 方法和 Runge-Kutta 方法.

使用真实的传感器数据进行训练的主要障碍是

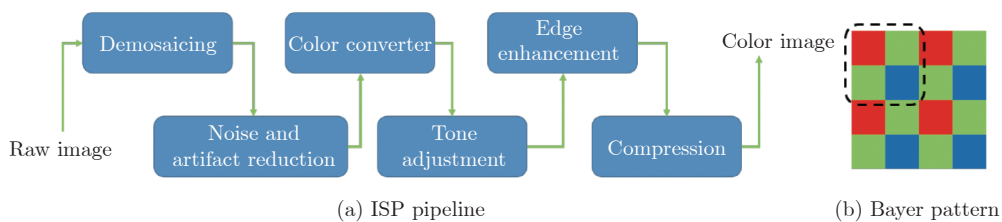
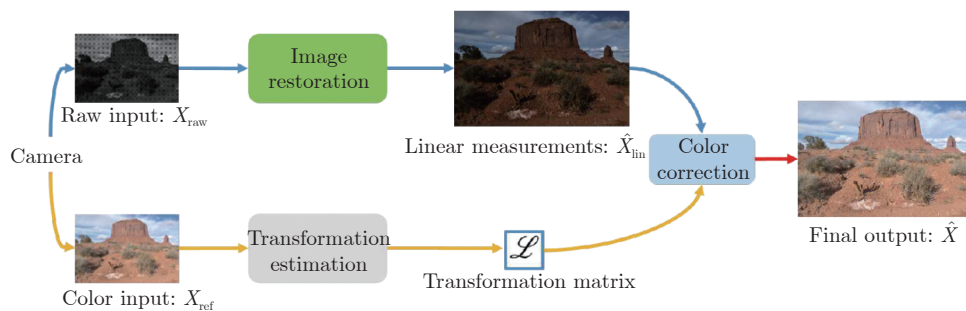
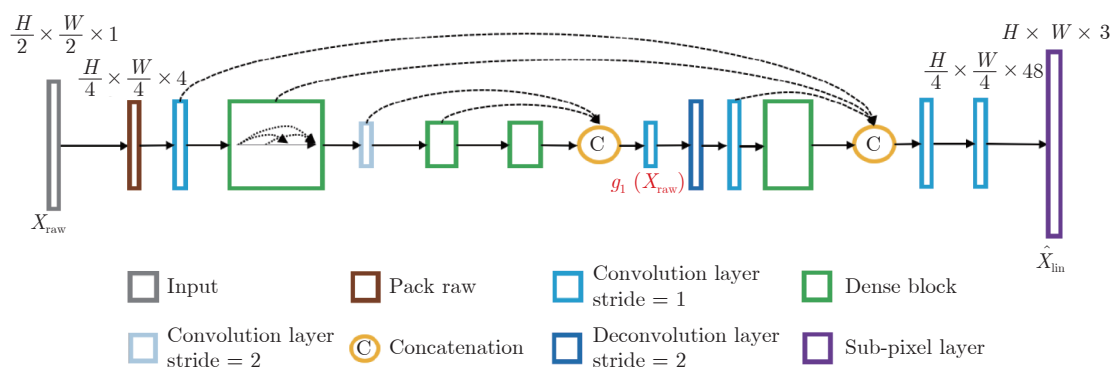


图 20 数码相机成像原理^[82]
Fig.20 Principle of digital camera imaging^[82]

图 21 双卷积神经网络^[82]Fig.21 A dual convolutional neural network^[82]图 22 图像恢复分支^[82]Fig.22 The image restoration branch^[82]

缺少 Ground truth 图像, 为此, Chen 等^[29] 提出了 CameraSR 模型, 提供了全新的获得图像对的思路, 即 R-V 退化模型, 其中 R 代表分辨率, V 代表视野域。视野域小, 目标物体具有更大的分辨率; 视野域大, 目标物体的分辨率很低。受 Ignatov 等^[30] 的启发, 采用 DSLR 拍摄含有目标物体的 HR 图像, 采用智能手机拍摄含有目标物体的 LR 图像。通过 SIFT 关键点^[86] 匹配与 RANSAC^[87] 估计单应性, 就可获得一个 LR-HR 图像对, 这种获得图像对的方式是从真实场景中得到的, 训练出的模型泛化能力更强。Zhang 等^[88] 展示了如何通过光学变焦获得 Ground truth 图像, 并提供了一个真实的数据集 SR-RAW。SR-RAW 是用变焦镜头拍摄的, 长焦距拍摄的图像可作为短焦距拍摄的光学 Ground truth 图像。Zhang 等还提出一种新的上下文双边损失 (Contextual bilateral loss, CoBi) 来处理轻微失调的图像对, CoBi 考虑了局部上下文相似性, 其训练后的网络在图像放大倍数为 4 倍和 8 倍上实现了最先进的性能。

2.4.3 针对人脸超分辨率重建的问题

针对人脸超分辨率问题, Chen 等^[89] 提出了 FS-RNet 和 FSRGAN 模型, 首先构建一个粗糙的 SR

网络生成粗糙的高分辨率图像; 之后将粗糙的高分辨率图像送入 2 个分支网络, 其中一个是精细的 SR 网络, 用来提取图像特征, 另一个是先验信息估计网络, 用来估计人脸特征点和分割信息; 最后将 2 个分支结果汇入一个细粒度解码器进行重构信息, 该模型在人脸超分辨率重建上取得了不错的结果。

Bulat 等^[90] 提出了 Super-FAN 模型, 该模型是第一个既提高面部分辨率又具有检测面部特征功能的端到端系统, 其核心在于通过热像图回归和优化热像损失函数, 利用集成子网络进行面部对齐, 将结构信息合并到基于 GAN 的超分辨率算法中, 该模型在现实世界中的低分辨率人脸图像上也表现出了良好的效果。

3 图像质量评估准则 (Image quality assessment, IQA)

为了方便评估超分辨率模型产生出的图像质量, 有必要运用图像质量评估准则对图像质量作出准确的评价。图像质量评估准则主要分为客观评价方法、主观评价方法以及其他评价准则, 客观评价方法如峰值信噪比与结构相似性^[71] 等, 通过公式计

算出数值比较生成图像质量的优劣,该方法可以对图像进行定量分析,但缺点是峰值信噪比与结构相似性很大的图像有时也会出现纹理模糊,甚至出现伪影,因此单靠客观评价方法不能完全判断出图像质量的好坏.而主观评价方法是在视觉上感知图像质量的好坏,该方法具有直观性,更符合我们的需求,但缺点是使用此类方法费时、代价昂贵且不方便.将客观评价方法与主观评价方法相结合更能有效检验算法的可靠性与准确性.本节将主要介绍几种最常用的 IQA 方法,还会介绍几种近年来使用相对较多的评价指标.

3.1 客观评价方法

1) 峰值信噪比 (Peak signal-to-noise ratio, PSNR)

峰值信噪比是一个表示信号最大可能功率和影响表示精度的破坏性噪声功率比值的工程术语,常用作图像压缩等领域中图像重建质量的评价方法,其简单地通过均方差 (Mean squared error, MSE) 进行定义,在超分辨率重建中,用于检测重建后的图像 x 与真实图像 y 的相似度. PSNR 越大代表图像质量越好,一般 PSNR 的范围在 20 到 40 之间.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y(i) - x(i))^2 \quad (12)$$

$$PSNR = 10 \cdot \lg \left(\frac{L^2}{MSE} \right) \quad (13)$$

其中, N 为图像的像素个数; y 为真实图像; x 为模型重建后的图像; L 为图像的最大像素值,一般为 255.

2) 结构相似性 (Structural similarity index, SSIM)

结构相似性是 Wang 等^[71] 根据人类视觉系统 (Human visual system, HVS) 提出的一种基于结构信息退化的质量评估框架,是全参考 (Full-reference) 质量评估的一种.与 PSNR 不同, SSIM 考虑了图像的结构信息、亮度与对比度,用图像均值作为亮度的估计、标准差作为对比度的估计、协方差作为结构相似程度的度量,反映了物体结构的属性.

给定两个图像 x 和 y , 两张图像的结构相似性可按照以下方式求出:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (14)$$

$$C_1 = (k_1 l)^2 \quad (15)$$

$$C_2 = (k_2 l)^2 \quad (16)$$

其中, μ_x 是 x 的平均值, μ_y 是 y 的平均值, σ_x^2 是 x 的方差, σ_y^2 是 y 的方差, σ_{xy} 是 x 和 y 的协方差, C_1 和 C_2 是用来维持稳定的常数, l 是像素值的动态范围, $k_1 = 0.01$, $k_2 = 0.03$. 结构相似性的范围为 0 到 1, 当两张图像一模一样时, SSIM 的值等于 1.

3) 信息保真度准则 (Information fidelity criterion, IFC) 与视觉信息保真度 (Visual information fidelity, VIF)

Sheikh 等^[91-92] 提出了信息保真度准则与视觉信息保真度,通过计算待评图像与参考图像之间的信息来衡量待评图像的质量优劣,其判断方法与 PSNR 相似.

3.2 主观评价方法

为了更加准确的评判图像的质量,除了以上 4 种常见的客观评价准则,其他的评价准则也相继被应用到超分辨率重建领域中.在国际标准中,平均主观意见分 (Mean opinion score, MOS) 最开始是用来评价经过压缩后的语音质量的准则,后来逐渐被应用到评估图像的质量.如表 3 所示, MOS 取值范围是 1 ~ 5 分, 1 分表示图像的质量最差, 5 分表示图像的质量最好.与 PSNR 这种客观评价方法不同, MOS 是一种主观评价方法.

表 3 MOS 评估准则
Table 3 The MOS assessment

分数	绝对评估	相对评估
1	图像质量非常差	该组中最差
2	图像质量较差	差于该组中平均水平
3	图像质量一般	该组中的平均水平
4	图像质量较好	好于该组中的平均水平
5	图像质量非常好	该组中最好

3.3 其他评价准则

此外还有 NQM^[93]、MS-SSIM^[94]、VSNR^[95]、FSIM^[96]、文献 [97]、NIQE^[98] 和 Ma' s score^[99] 等评价准则被应用到图像质量评估上,多种准则的结合使用更能检验出图像质量的好坏.

4 实验结果对比与分析

4.1 图像质量评估准则的对比

目前几乎所有的超分辨率重建算法都采用 PSNR 与 SSIM 来检测模型产生的图像质量,表 4 和表 5 分别为部分网络模型在基准数据集 Set5、Set14 的平均 PSNR 与 SSIM 的对比,作为常用的两种客观评价方法,由于计算简单且物理意义明确,因此应

表 4 部分网络模型在基准数据集 Set5、Set14 的平均 PSNR 对比

Table 4 The average PSNR comparison of some network models on the Set5 and Set14 benchmark datasets

方法	Set5				Set14			
	×2	×3	×4	×8	×2	×3	×4	×8
Bicubic ^[2]	33.66	30.39	28.42	24.39	30.23	27.54	26.00	23.19
SRCNN ^[6]	36.66	32.75	30.49	25.33	32.45	29.30	27.50	23.85
VDSR ^[7]	37.10	32.89	30.84	25.72	32.97	29.77	28.03	24.21
ESPCN ^[43]	—	33.13	30.90	—	—	29.49	27.73	—
SRGAN ^[8]	—	—	30.64	—	—	—	26.92	—
LapSRN ^[47]	37.52	33.82	31.54	26.14	33.08	29.87	28.19	24.44
SRDenseNet ^[38]	—	—	32.02	—	—	—	28.50	—
EDSR ^[43]	38.20	34.76	32.62	26.96	34.02	30.66	28.94	24.91
EnhanceNet ^[44]	—	—	31.74	—	—	—	28.42	—
DBPN ^[53]	38.09	—	32.47	27.21	33.85	—	28.82	25.13
RCAN ^[55]	38.33	34.85	32.73	27.47	34.23	30.76	28.98	25.40
SRMD ^[61]	37.79	34.12	31.96	—	33.32	30.04	28.35	—
ZSSR ^[77]	37.37	33.42	31.13	—	33.00	29.80	28.01	—
Meta-SR ^[72]	—	—	—	—	34.04	30.55	28.84	—
OISR ^[85]	38.12	34.56	32.33	—	33.80	30.46	28.73	—

表 5 部分网络模型在基准数据集 Set5、Set14 的平均 SSIM 对比

Table 5 The comparison of average SSIM of partial network models on the Set5 and Set14 benchmark datasets

方法	Set5				Set14			
	×2	×3	×4	×8	×2	×3	×4	×8
Bicubic ^[2]	0.9299	0.8682	0.8104	0.657	0.8687	0.7736	0.7019	0.568
SRCNN ^[6]	0.9542	0.9090	0.8628	0.689	0.9067	0.8215	0.7513	0.593
VDSR ^[7]	0.9587	0.9213	0.8838	0.711	0.9124	0.8314	0.7674	0.609
FSRCNN ^[13]	0.9558	0.9140	0.8657	0.682	0.9088	0.8242	0.7535	0.592
SRGAN ^[8]	—	—	0.8472	—	—	—	0.7397	—
LapSRN ^[47]	0.959	0.9227	0.885	0.738	0.913	0.8320	0.772	0.623
SRDenseNet ^[38]	—	—	0.8934	—	—	—	0.7782	—
EDSR ^[43]	0.9606	0.9290	0.8984	0.775	0.9204	0.8481	0.7901	0.640
MemNet ^[36]	0.9597	0.9248	0.8893	0.7414	0.9142	0.8350	0.7723	0.6199
DBPN ^[53]	0.960	—	0.898	0.784	0.919	—	0.786	0.648
RCAN ^[55]	0.9617	0.9305	0.9013	0.7913	0.9225	0.8494	0.7910	0.6553
SRMD ^[61]	0.960	0.925	0.893	—	0.915	0.837	0.777	—
ZSSR ^[77]	0.9570	0.9188	0.8796	—	0.9108	0.8304	0.7651	—
Meta-SR ^[72]	—	—	—	—	0.9213	0.8466	0.7872	—
OISR ^[85]	0.9609	0.9284	0.8968	—	0.9196	0.8450	0.7845	—

用最为广泛, 但并不能很好的代表感知视觉质量. 表 6 为部分网络模型在基准数据集 Set5、Set14 和 BSDS100 的×4 尺度上的 MOS 对比, 相比于 PSNR 和 SSIM, MOS 更能代表感知视觉质量的好坏. 可以看出, 在基准数据集 Set5、Set14 上采样因子为×4 时, RCAN 在客观评价指标 PSNR 和 SSIM 上表现最好; 但在主观评价指标 MOS 上 SRGAN

表现最好.

如图 23 所示, 在主观视觉与 PSNR 方面上, 分别运用 Bicubic、SRCNN、EDSR、RCAN、EnhanceNet、SRGAN 以及 ESRGAN 模型在 Set14 和 BSDS100 数据集上进行实验, 可以看出 RCAN 模型的 PSNR 最大; 但在主观视觉上, ESRGAN 生成的图像最清晰, 无论从颜色、纹理分布更接近 HR

表 6 部分网络模型在基准数据集 Set5、Set14 和 BSDS100 的 $\times 4$ 尺度上的 MOS 对比

Table 6 The MOS comparison of some network models at $\times 4$ of the benchmark datasets Set5, Set14 and BSDS100

方法	Set5	Set14	BSDS100
Bicubic ^[2]	1.97	1.80	1.47
SRCNN ^[6]	2.57	2.26	1.87
ESPCN ^[33]	2.89	2.52	2.01
DRCN ^[37]	3.26	2.84	2.12
SRResNet ^[8]	3.37	2.98	2.29
SRGAN ^[8]	3.58	3.72	3.56
HR	4.32	4.32	4.46

图像, RCAN 较为模糊, 这也说明了传统客观评价标准的局限性.

4.2 运行时间的对比

除了图像质量评估外, 运行时间也是评价网络模型性能优劣的重要指标之一. 运行时间指测试时重建一张高分辨率图像的时间. 如表 7 所示, 其中“—”代表未发现此项.

5 思考与展望

虽然深度学习的出现, 使基于单幅图像超分辨率重建算法性能得到准确率与速度的提升, 但目前仍然存在很多不足:

1) 基于无监督学习的单幅图像超分辨率重建算法的研究问题

文献 [68, 77] 虽然提出了基于无监督学习的超分辨率重建的算法, 但这方面可供参考的文献仍然很少, 且目前获取 HR-LR 图像对的主流方法依然包含很多主观因素, 不能很好地代替现实数据集. 若采用无监督学习的方法, 无需大量样本进行训练, 也可节约获取大量数据集的时间, 直接使用现实图像进行训练与测试, 不依靠外部数据集, 更能提高模型的泛化能力, 因此研究基于无监督学习的单幅图像超分辨率重建算法很有必要.

2) 盲超分辨率解决的问题

盲超分辨率是指从高分辨率图像到低分辨率图像下采样的尺度未知, 需要预先估计模糊核. 在模糊核不匹配时, 模型得到的 SR 图像会过于模糊或产生振铃效应, 只有在模糊核匹配时, 才会有较好的图像超分辨率效果^[100], 因此估计退化模糊核是一个关键步骤. 文献 [79–81] 提出了关于此方面的解决方法, 但盲超分辨率问题仍然有很大的上升空间, 如何构建深度学习模型来解决盲超分辨率依旧是未来值得关注的问题.

3) 使用一种模型可实现任意倍率缩放的超分辨率算法的问题

文献 [72] 提出的 Meta-SR 网络模型, 仅使用一种模型就实现了任意倍率的缩放 (不限于整数), 使重建图像的尺寸大小不再拘泥于输入图像的如 $\times 2$ 、 $\times 3$ 、 $\times 4$ 、 $\times 8$ 、 $\times 16$ 等整数倍率, 还可实现 $\times 1.2$ 、 $\times 1.5$ 等小数倍率的图像重建, 这是一个很有价值的创新, 解决了多年来被人们忽视但却很重要的问题, 未来在此方向还会有更多的学者进行探索与创新.

4) 应用到现实场景中的问题

虽然从深度学习出现之后, 在单幅图像超分辨率算法上得到了很大的进步与改良. 但许多算法模型应用到现实场景中, 往往得不到预期的效果, 甚至有时比传统算法得到的图像质量还要差. 如何提升模型的泛化能力, 将其早日应用到实践当中是当前最为重要的问题. 未来的发展趋势将会是传统算法与神经网络的结合, 既能发挥传统算法先验知识的重要性, 又能实现端到端高层语义特征的学习, 更好的解决现实场景中图像超分辨率重建的问题.

6 结论

超分辨率重建在计算机视觉领域具有重要的研究意义和应用价值, 本文首先介绍单幅图像超分辨率重建的常用公共图像数据集; 然后, 重点阐述基于深度学习的单幅图像超分辨率重建方向的创新与进展; 最后, 讨论了单幅图像超分辨率重建方向上存在的困难和挑战, 并对未来的发展趋势进行了思考与展望. 总体而言, 深度学习为解决单幅图像超分辨率重建问题上提供了新的技术, 取得了较为显著的成果, 但将其应用到现实场景中仍存在很多问题, 如现实生活中得到的低分辨率图像并不是从高分辨率图像双三次下采样得到的, 而许多基于深度学习的超分辨率模型训练时都是将双三次下采样得到的低分辨率图像作为输入, 测试时若将现实生活中的低分辨率图像作为输入, 重建的效果往往不是很好等等诸多问题. 说明目前深度学习技术还不够成熟, 有很大的上升空间, 但其前景可期, 且图像超分辨率重建技术具有很大的实用价值, 因此基于深度学习的单幅图像超分辨率重建依然是一个亟待研究的重点方向.

致谢

感谢山东大学贾晔焯教授在论文修改过程中的建议.

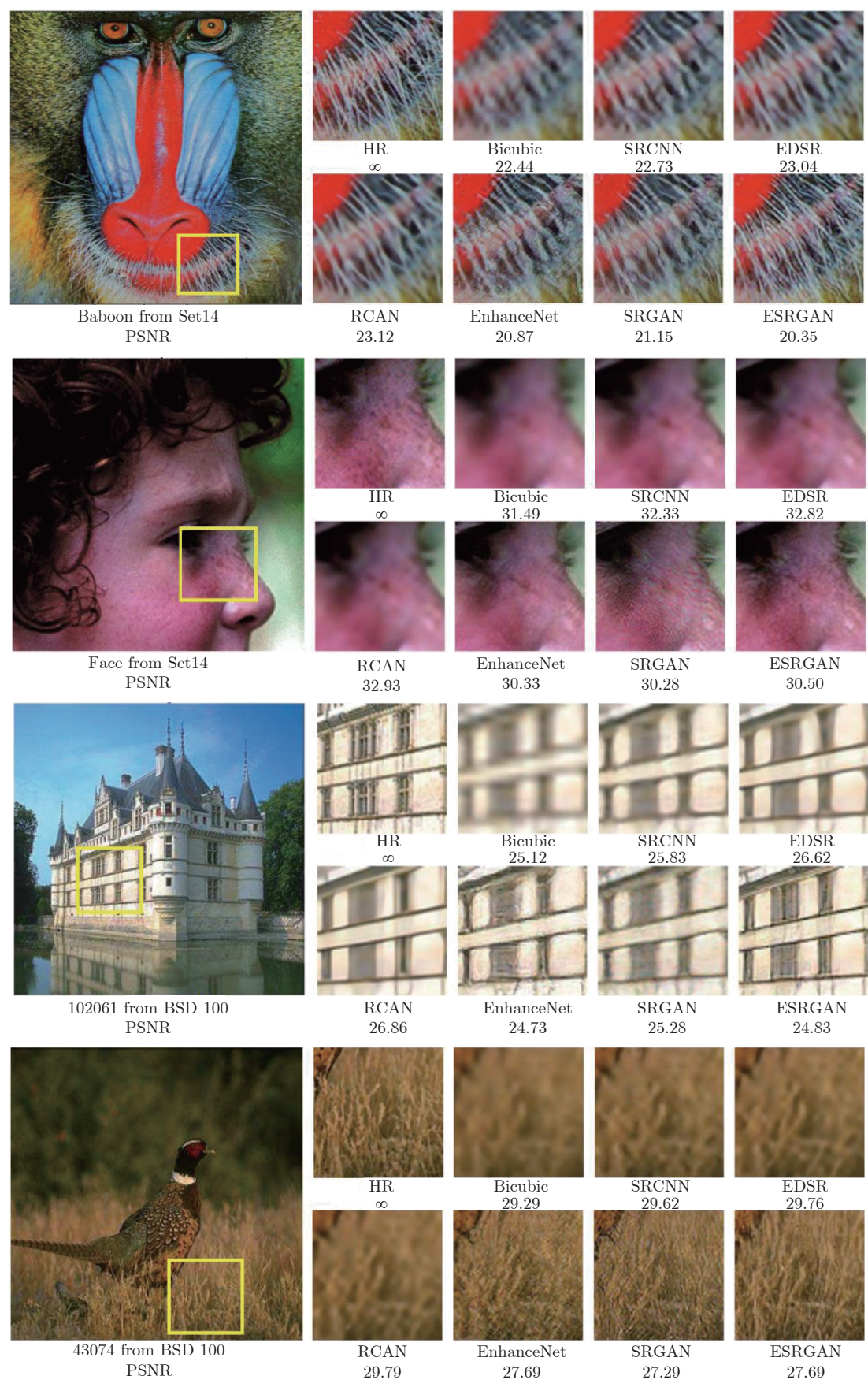


图 23 部分模型的主观视觉与 PSNR 的比较

Fig. 23 Comparison of subjective vision and PSNR of partial models

表 7 部分网络模型在各测试数据集上的运行时间对比
Table 7 The comparison of running time of partial network models on each testing datasets

方法	深度学习框架	CPU/GPU	测试数据集	上采样因子	运行时间 (s)
SRCNN ^[6]	Caffe	CPU	Set5	×3	2.23
VDSR ^[7]	MatConvNet	CPU	Set5	×3	0.13
ESPCN ^[33]	Theano	CPU	Set14	×3	0.26
FSRCNN ^[13]	Caffe	CPU	Set14	×3	0.061
LapSRN ^[47]	MatConvNet	GPU	Set14	×4	0.04
MemNet ^[36]	Caffe	GPU	Set5	×3	0.4
EnhanceNet ^[44]	Tensorflow	GPU	Set5	×4	0.009
MS-LapSRN ^[69]	MatConvNet	GPU	Urban100	×4	0.06
ZSSR ^[77]	—	GPU	BSDS100	×2	9
Meta-SR ^[72]	—	GPU	BSDS100	×2	0.033

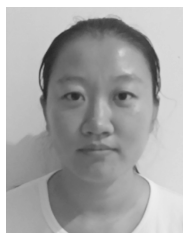
References

- Duchon C E. Lanczos filtering in one and two dimensions. *Journal of Applied Meteorology*, 1979, **18**(8): 1016–1022
- Keys R. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1981, **29**(6): 1153–1160
- Freeman W T, Jones T R, Pasztor E C. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 2002, **22**(2): 56–65
- Chang H, Yeung D Y, Xiong Y M. Super-resolution through neighbor embedding. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). Washington, DC, USA: IEEE, 2004.
- Sun J, Xu Z B, Shum H Y. Image super-resolution using gradient profile prior. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Anchorage, AK, USA: IEEE, 2008. 1–8
- Dong C, Loy C C, He K M, Tang X O. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(2): 295–307
- 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 (CVPR). Las Vegas, NV, USA: IEEE, 2016. 1646–1654
- Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017. 105–114
- Ignatov A, Kobyshev N, Timofte R, Vanhoey K, Van Gool L. WESPE: Weakly supervised photo enhancer for digital cameras. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Salt Lake City, USA: IEEE, 2018.
- Bevilacqua M, Roumy A, Guillemot C, Morel M L A. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: Proceedings of the 2012 British Machine Vision Conference. Surrey, UK: BMVA Press, 2012. 135.1–135.10
- Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations. In: Proceedings of the 7th International Conference on Curves and Surfaces. Avignon, France: Springer, 2010. 711–730
- Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015. 5197–5206
- Dong C, Loy C C, Tang X O. Accelerating the super-resolution convolutional neural network. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016. 391–407
- Martin D, Fowlkes C, Tal D, Malik J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV). Vancouver, BC, Canada: IEEE, 2001. 416–423
- Arbeláez P, Maire M, Fowlkes C, Malik J. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(5): 898–916
- Fujimoto A, Ogawa T, Yamamoto K, Matsui Y, Aizawa K. Manga109 dataset and creation of metadata. In: Proceedings of the 1st International Workshop on coMics ANalysis, Processing and Understanding. Cancun, Mexico: ACM, 2016. 1–5
- 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
- Agustsson E, Timofte R. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, HI, USA: IEEE, 2017. 1122–1131
- Timofte R, Rothe R, Van Gool L. Seven ways to improve example-based single image super resolution. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016. 1865–1873
- Wang X T, Yu K, Dong C, Loy C C. Recovering realistic texture in image super-resolution by deep spatial feature transform. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE, 2018. 606–615
- Blau Y, Mechrez R, Timofte R, Michaeli T, Zelnik-Manor L. The 2018 PIRM challenge on perceptual image super-resolution. In: Proceedings of the 2018 European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018. 334–355
- Deng J, Dong W, Socher R, Li L J, Li K, Li F F. ImageNet: A large-scale hierarchical image database. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Miami, FL, USA: IEEE, 2009. 248–255
- Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft COCO: Common objects in context. In: Proceedings of the 13th European Conference on Computer Vision

- (ECCV). Zurich, Switzerland: Springer, 2014. 740–755
- 24 Everingham M, Eslami S M A, Van Gool L, Williams C K I, Winn J, Zisserman A. The PASCAL visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 2015, **111**(1): 98–136
 - 25 Liu Z W, Luo P, Wang X G, Tang X O. Deep learning face attributes in the wild. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 3730–3738
 - 26 Yu F, Zhang Y D, Song S R, Seff A, Xiao J X. Construction of a large-scale image dataset using deep learning with humans in the loop. arXiv preprint arXiv: 1506.03365, 2016.
 - 27 Ma K D, Duanmu Z F, Wu Q B, Wang Z, Yong H W, Li H L, et al. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 2017, **26**(2): 1004–1016
 - 28 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 (CVPRW). Honolulu, HI, USA: IEEE, 2017. 1110–1121
 - 29 Chen C, Xiong Z W, Tian X M, Zha Z J, Wu F. Camera lens super-resolution. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019. 1652–1660
 - 30 Ignatov A, Kobyshev N, Timofte R, Vanhoey K, Van Gool L. DSLR-quality photos on mobile devices with deep convolutional networks. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 3297–3305
 - 31 Wang Z H, Chen J, Hoi S C H. Deep learning for image super-resolution: A survey. arXiv preprint arXiv: 1902.06068, 2020.
 - 32 Hu Chang-Sheng, Zhan Shu, Wu Cong-Zhong. Image super-resolution based on deep learning features. *Acta Automatica Sinica*, 2017, **43**(5): 814–821
(胡长胜, 詹曙, 吴丛中. 基于深度特征学习的图像超分辨率重建. *自动化学报*, 2017, **43**(5): 814–821)
 - 33 Shi W Z, Caballero J, Huszár 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 (CVPR). Las Vegas, NV, USA: IEEE, 2016. 1874–1883
 - 34 Mao X J, Shen C H, Yang Y B. Image restoration using convolutional auto-encoders with symmetric skip connections. arXiv preprint arXiv: 1606.08921, 2016.
 - 35 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 (CVPR). Las Vegas, NV, USA: IEEE, 2016. 770–778
 - 36 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 (ICCV). Venice, Italy: IEEE, 2017. 4549–4557
 - 37 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 (CVPR). Las Vegas, NV, USA: IEEE, 2016. 1637–1645
 - 38 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 (CVPR). Honolulu, HI, USA: IEEE, 2017. 2261–2269
 - 39 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 (ICCV). Venice, Italy: IEEE, 2017. 4809–4817
 - 40 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 (CVPR). Salt Lake City, USA: IEEE, 2018. 2472–2481
 - 41 Yu J H, Fan Y C, Yang J C, Xu N, Wang Z W, Wang X C, et al. Wide activation for efficient and accurate image super-resolution. arXiv preprint arXiv: 1808.08718, 2018.
 - 42 Timofte R, Gu S H, Wu J Q, Van Gool L, Zhang L, Yang M H, et al. NTIRE 2018 challenge on single image super-resolution: Methods and results. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, USA: IEEE, 2018.
 - 43 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 (CVPRW). Honolulu, HI, USA: IEEE, 2017. 1132–1140
 - 44 Sajjadi M S M, Schölkopf B, Hirsch M. EnhanceNet: Single image super-resolution through automated texture synthesis. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 4501–4510
 - 45 Soh J W, Park G Y, Jo J, Cho N I. Natural and realistic single image super-resolution with explicit natural manifold discrimination. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019. 8114–8123
 - 46 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 (CVPR). Honolulu, HI, USA: IEEE, 2017. 2790–2798
 - 47 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 (CVPR). Honolulu, HI, USA: IEEE, 2017. 5835–5843
 - 48 Zhou Deng-Wen, Zhao Li-Juan, Duan Ran, Chai Xiao-Liang. Image super-resolution based on recursive residual networks. *Acta Automatica Sinica*, 2019, **45**(6): 1157–1165
(周登文, 赵丽娟, 段然, 柴晓亮. 基于递归残差网络的图像超分辨率重建. *自动化学报*, 2019, **45**(6): 1157–1165)
 - 49 Han W, Chang S Y, Liu D, Yu M, Witbrock M, Huang T S. Image super-resolution via dual-state recurrent networks. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE, 2018. 1654–1663
 - 50 Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge, MA, USA: MIT Press, 2014. 2672–2680
 - 51 Wang X T, Yu K, Wu S X, Gu J J, Liu Y H, Dong C, et al. ESRGAN: Enhanced super-resolution generative adversarial networks. In: Proceedings of the 2018 European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018. 63–79
 - 52 Jolicœur-Martineau A. The relativistic discriminator: A key element missing from standard GAN. arXiv preprint arXiv: 1807.00734, 2018.
 - 53 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 (CVPR). Salt Lake City, USA: IEEE, 2018. 1664–1673
 - 54 Li Z, Yang J L, Liu Z, Yang X M, Jeon G, Wu W. Feedback network for image super-resolution. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Re-

- cognition (CVPR). Long Beach, CA, USA: IEEE, 2019. 3862–3871
- 55 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 the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018. 286–301
 - 56 Dai T, Cai J R, Zhang Y B, Xia S T, Zhang L. Second-order attention network for single image super-resolution. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019. 11057–11066
 - 57 Zhou Deng-Wen, Ma Lu-Yao, Tian Jin-Yue, Sun Xiu-Xiu. Feature fusion attention network for image super-resolution [Online], *Acta Automatica Sinica*, available: <https://doi.org/10.16383/j.aas.c190428>, November 7, 2019 (周登文, 马路遥, 田金月, 孙秀秀. 基于特征融合注意网络的图像超分辨率重建 [Online], *自动化学报*, 获取自: <https://doi.org/10.16383/j.aas.c190428>, 2019年11月7日)
 - 58 Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution. In: Proceedings of the 14th European Conference on Computer Vision (ECCV). Amsterdam, The Netherlands: Springer, 2016. 694–711
 - 59 Gatys L A, Ecker A S, Bethge M. A neural algorithm of artistic style. arXiv preprint arXiv: 1508.06576, 2015.
 - 60 Wang Y F, Perazzi F, McWilliams B, Sorkine-Hornung A, Sorkine-Hornung O, Schroers C. A fully progressive approach to single-image super-resolution. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Salt Lake City, USA: IEEE, 2018.
 - 61 Yuan Y, Liu S Y, Zhang J W, Zhang Y B, Dong C, Lin L. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Salt Lake City, USA: IEEE, 2018.
 - 62 Mao X D, Li Q, Xie H R, Lau R Y K, Wang Z, Smolley S P. Least squares generative adversarial networks. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 2813–2821
 - 63 Gatys L A, Ecker A S, Bethge M. Texture synthesis using convolutional neural networks. arXiv preprint arXiv: 1505.07376, 2015.
 - 64 Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016. 2414–2423
 - 65 Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 1992, **60**(1-4): 259–268
 - 66 Mechrez R, Talmi I, Shama F, Zelnik-Manor L. Maintaining natural image statistics with the contextual loss. In: Proceedings of the 14th Asian Conference on Computer Vision (ACCV). Perth, Australia: Springer, 2018. 427–443
 - 67 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 1409.1556, 2015.
 - 68 Lempitsky V, Vedaldi A, Ulyanov D. Deep image prior. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE, 2018. 9446–9454
 - 69 Lai W S, Huang J B, Ahuja N, Yang M H. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, **41**(11): 2599–2613
 - 70 Kim J H, Lee J S. Deep residual network with enhanced up-scaling module for super-resolution. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Salt Lake City, USA: IEEE, 2018.
 - 71 Wang Z, Bovik A C, Sheikh H R, Simoncelli E P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004, **13**(4): 600–612
 - 72 Hu X C, Mu H Y, Zhang X Y, Wang Z L, Tan T N, Sun J. Meta-SR: A magnification-arbitrary network for super-resolution. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019. 1575–1584
 - 73 Bulat A, Yang J, Tzimiropoulos G. To learn image super-resolution, use a GAN to learn how to do image degradation first. In: Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018. 187–202
 - 74 Zhu J Y, Park T, Isola P, Efros A A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 2242–2251
 - 75 Zontak M, Irani M. Internal statistics of a single natural image. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, CO, USA: IEEE, 2011. 977–984
 - 76 Mosseri I, Zontak M, Irani M. Combining the power of internal and external denoising. In: Proceedings of the 2013 IEEE International Conference on Computational Photography (ICCP). Cambridge, MA, USA: IEEE, 2013. 1–9
 - 77 Shocher A, Cohen N, Irani M. Zero-shot super-resolution using deep internal learning. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE, 2018. 3118–3126
 - 78 Glasner D, Bagon S, Irani M. Super-resolution from a single image. In: Proceedings of the 12th International Conference on Computer Vision (ICCV). Kyoto, Japan: IEEE, 2009. 349–356
 - 79 Zhang K, Zuo W M, Zhang L. Learning a single convolutional super-resolution network for multiple degradations. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE, 2018. 3262–3271
 - 80 Gu J J, Lu H N, Zuo W M, Dong C. Blind super-resolution with iterative kernel correction. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019. 1604–1613
 - 81 Zhang K, Zuo W M, Zhang L. Deep plug-and-play super-resolution for arbitrary blur kernels. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019. 1671–1681
 - 82 Xu X Y, Ma Y R, Sun W X. Towards real scene super-resolution with raw images. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019. 1723–1731
 - 83 Coffin D. Dcraw: Decoding raw digital photos in linux [Online], available: <http://www.cybercom.net/dciffin/dcraw/>
 - 84 Szegedy C, Ioffe S, Vanhoucke V, Alemi A A. Inception-v4, inception-ResNet and the impact of residual connections on learning. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI). San Francisco, USA: AAAI Press, 2017. 4278–4284
 - 85 He X Y, Mo Z T, Wang P S, Liu Y, Yang M Y, Cheng J. ODE-inspired network design for single image super-resolution. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019. 1732–1741

- 86 Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, **60**(2): 91–110
- 87 Fischler M A, Bolles R C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 1981, **24**(6): 381–395
- 88 Zhang X E, Chen Q F, Ng R, Koltun V. Zoom to learn, learn to zoom. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019. 3757–3765
- 89 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 (CVPR). Salt Lake City, USA: IEEE, 2018. 2492–2501
- 90 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 (CVPR). Salt Lake City, USA: IEEE, 2018. 109–117
- 91 Sheikh H R, Bovik A C, de Veciana G. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, 2005, **14**(12): 2117–2128
- 92 Sheikh H R, Bovik A C. Image information and visual quality. *IEEE Transactions on Image Processing*, 2006, **15**(2): 430–444
- 93 Damera-Venkata N, Kite T D, Geisler W S, Evans B L, Bovik A C. Image quality assessment based on a degradation model. *IEEE Transactions on Image Processing*, 2000, **9**(4): 636–650
- 94 Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment. In: Proceedings of the 37th Asilomar Conference on Signals, Systems & Computers. Pacific Grove, CA, USA: IEEE, 2003. 1398–1402
- 95 Chandler D M, Hemami S S. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Processing*, 2007, **16**(9): 2284–2298
- 96 Zhang L, Zhang L, Mou X Q, Zhang D. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 2011, **20**(8): 2378–2386
- 97 Blau Y, Michaeli T. The perception-distortion tradeoff. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE, 2018. 6228–6237
- 98 Mittal A, Soundararajan R, Bovik A C. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, 2013, **20**(3): 209–212
- 99 Ma C, Yang C Y, Yang X K, Yang M H. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, 2017, **158**: 1–16
- 100 Michaeli T, Irani M. Nonparametric blind super-resolution. In: Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, NSW, Australia: IEEE, 2013. 945–952



李佳星 长春理工大学机电工程学院硕士研究生. 主要研究方向为计算机视觉, 深度学习, 单幅图像超分辨率重建.

E-mail: jiaxing4912@163.com

(LI Jia-Xing Master student at the College of Mechanical and Electric Engineering, Changchun University of Science and Technology. Her research interest covers computer vision, deep learning, and single image super-resolution reconstruction.)



赵勇先 中国科学院大学机械电子工程专业硕士研究生. 2020 年获得长春理工大学学士学位. 主要研究生方向为计算机视觉和深度学习.

E-mail: zyx_19980824@163.com

(ZHAO Yong-Xian Master student at the University of Chinese Academy of Sciences, majoring in Mechanical and Electronic Engineering. He received his bachelor degree from Changchun University of Science and Technology in 2020. His research interest covers computer vision and deep learning.)



王京华 长春理工大学机电工程学院讲师. 2010 年获得哈尔滨工业大学博士学位. 主要研究方向为移动机器人, 无人系统, 智能控制. 本文通信作者.

E-mail: hit1920s@163.com

(WANG Jing-Hua Lecturer at the College of Mechanical and Electric Engineering, Changchun University of Science and Technology. He received his Ph. D. degrees from Harbin Institute of Technology in 2010. His research interest covers mobile robots, unmanned systems and intelligent control. Corresponding author of this paper.)