

图像超分辨率重建研究综述

唐艳秋¹⁾, 潘泓¹⁾, 朱亚平²⁾, 李新德¹⁾

(1. 东南大学自动化学院, 复杂工程系统测量与控制教育部重点实验室, 江苏南京 210096;

2. 中国传媒大学信息工程学院, 北京 100024)

摘 要 图像超分辨率重建(Super-resolution Reconstruction, SR)就是由一张或多张低分辨率图像得到高分辨率图像的过程。SR 技术自 1964 年出现以来, 不断发展, 在卫星成像、医学图像处理、人脸分析、安全监控和生物特征识别等领域被广泛应用。本文在回顾 SR 技术发展历史的基础上, 全面综述了 SR 技术在各个时期的代表性方法, 包括了经典的差值算法、早期基于重构的方法、近期基于浅层机器学习和基于深度学习的方法, 从信号处理、机器学习以及深度学习等不同角度描述了 SR 问题的解决思路, 并重点介绍了深度学习在图像超分辨率领域的最新工作进展。我们从模型类型、网络结构、信息传递方式、损失函数形式等方面对各种算法进行了详细分析和评述, 对比总结了其优缺点。最后探讨了图像超分辨率技术未来的发展方向。

关键词 图像超分辨率, 深度学习, 图像重建, 损失函数

中图法分类号 TP391

A Survey of Image Super-resolution Reconstruction

TANG Yan-qiu¹, PAN Hong¹, ZHU Ya-ping², Li Xin-de¹

(1. School of Automation, Southeast University, Nanjing, Jiangsu 210096, China

Key Laboratory of Measurement and Control of CSE, Ministry of Education, Southeast University, Nanjing, Jiangsu 210096, China;

2. School of Information Engineering, Communication University of China, Beijing 100024, China)

Abstract Image Super-resolution Reconstruction (SR) aims to obtain high-resolution images from one or more low-resolution images. Since emerging in 1964, SR has been developing continuously and widely applied in different fields, from satellite imaging to medical image processing, to facial image analysis, video surveillance and biometrics. This survey retrospects the history of SR technique and provides a comprehensive overview of representative SR methods, including classical interpolation-based methods, early construction-based algorithms and recent approaches based on shallow machine learning and deep learning. We describe typical solutions to SR tasks from perspective of signal processing, machine learning and deep learning, with an emphasis on the recent work and progress on deep learning-based approaches. We elaborate the details of various deep learning-based SR methods, including their strength and weakness, in terms of the deep learning model, architecture, message pass and network's loss function. Finally, we discuss the possible research directions on SR technique.

Key words Image super-resolution; deep learning; image reconstruction; loss function

1 引言

与普通图像相比, 高分辨率图像具有更密集的像素, 包含更多精细有效的视觉信息, 这些信息在医学影像处理^[1-3]、安全监控^[4-6]、遥感成像^[7-9]、视频转换^[10-12]等应用中至关重要, 因此获取高分辨率图像具有十分重要的意义。然而, 图像获取过程中, 图像质量会受到硬件条件限制、成像环境变化和图像传输信道干扰等因素影响。此外, 欠采样效应造成图像频谱混叠, 使获取的图像产生失真; 大气扰

动、拍摄图像与传感器之间的相对运动造成的图像变形、光学系统以及相对运动造成的模糊、下采样和系统噪声等, 都会导致图像质量出现不同程度的下降。采用增加传感器阵列采样密度的办法^[13]来提高图像分辨率和消除变形效应的方法往往面临昂贵的代价和技术困难, 因此通过图像处理、机器学习等技术设计相应算法来提高图像分辨率有着极大的现实意义和应用价值。

收稿日期: 年-月-日; 最终修改稿收到日期: 年-月-日 *投稿时不填写此项*。本课题得到国家自然科学基金(No.61671151, 61573097, 61703100, 91748106)、江苏省自然科学基金(No.BK20181265)、流程工业综合自动化国家重点实验室基金(No.PAL-N201704)、国家留学基金委(CSC)、江苏省青蓝工程人才项目、江苏省六大高峰人才项目和江苏省优势学科经费资助。唐艳秋, 女, 1995年生, 东南大学硕士研究生, 主要研究领域为机器学习、深度学习、图像处理。E-mail: 220171510@seu.edu.cn, 手机号: 18795968596。潘泓(通信作者), 男, 1974年生, 博士, 副研究员, 中国计算机学会(CCF)会员(会员号: A8758M), 主要研究领域为机器学习、深度学习、计算机视觉、医疗图像处理、机器视觉、多媒体信号处理。E-mail: enhpan@seu.edu.cn。朱亚平, 女, 1977年生, 博士, 教授, 主要研究领域为机器学习、深度学习、图像处理、模式识别。E-mail: zhuyaping@cuc.edu.cn。李新德, 男, 1975年生, 博士, 教授, 主要研究领域为机器感知与学习、数据挖掘、机器视觉、智能信息融合和不确定推理、智能机器人。E-mail: xindeli@seu.edu.cn

图像超分辨率重建 (Super-resolution Reconstruction, SR)是指采用信号/图像处理算法和机器学习技术,从同一场景中已有的一张或多张低分辨率 (Low-resolution, LR)图像构建高分辨率 (High-resolution, HR)图像的技术^[14]。图像超分辨率重建的概念最早在 1964 年和 1968 年分别由 Harris^[15]和 Goodman^[16]提出,他们的方法被称为 Harris-Goodman 频谱外推法。1984 年, Tsai 和 Huang^[17]等人提出频域中由低分辨率图像序列复原高分辨率图像的方法后, SR 技术逐渐引起人们注

意并被广泛研究。从 Tsai 和 Huang 开创性工作以来至今的三十多年中,人们已经提出了各种 SR 技术,包括了频域到空域的方法,以及从信号处理角度到机器学习角度的解决方案。2014 年左右,随着深度学习的兴起,人们将深度学习技术应用在 SR 问题中,提出了包括 SRCNN、VDSR、SRGAN 和 RDN 在内的一系列基于深度学习框架的 SR 方法,极大提升了 SR 算法的性能。图 1 展示了图像 SR 研究工作的发展路线。

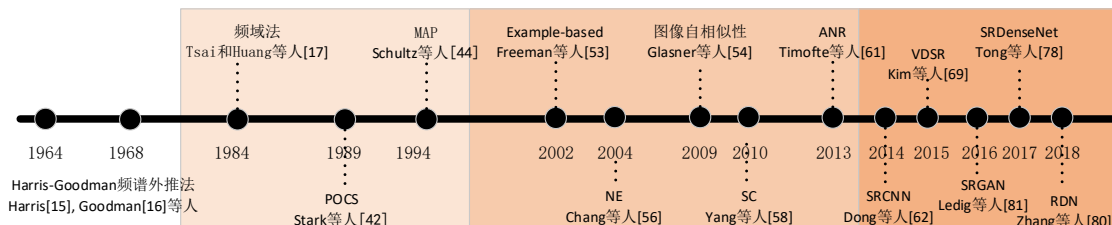


图 1. 图像 SR 方法的发展历程

传统图像超分辨率重建算法包括基于插值,基于重构和基于浅层结构的学习方法,这些方法普遍难以恢复丰富的高频信息,导致重建图像模糊且计算复杂、实时性不高等问题,不适用于大放大因子(如 $\times 4, \times 8$)的要求。为了解决这些问题,近年来,人们将深度学习用于图像超分辨率重建中: 1) 通过加深加宽深度网络、增加网络内部连接、加入图像边缘,梯度,类别等先验知识、优化感知损失函数等策略使网络更好地学习从低分辨率图像到高分辨率图像的映射关系,从而重建出高质量的图像细节; 2) 通过残差学习、参数共享、梯度裁剪等策略减少网络参数,避免梯度消失和梯度爆炸,加快网络收敛速度,实现实时高分辨率重建; 3) 通过逐级放大的方式,网络能实现大放大倍数的高分辨率重建。如今,基于深度学习的超分辨率重建已逐渐成为主流方法。

随着对 SR 问题研究的深入,大量方法不断涌现。一些学者相继对该领域工作进行了综述。早期综述文章^[18-20]主要总结了传统 SR 算法。文献^[20]根据算法的不同输入输出组合,将超分辨率问题分为基于重建的超分辨率、视频超分辨率、单帧图像超分辨率三大类,全面综述了该问题的常用算法及当时的最新研究成果,由于深度学习尚未用于 SR 领域,因此该综述没有涉及基于深度学习的 SR 算法。文献^[21]比较全面地介绍了现有图像超分辨率重建方法,但基于深度学习的方法仅介绍了代表性工作-SRCNN。文献^[22-24]对基于深度学习的 SR 技术进行了综述。文献^[22]介绍了图像和视频的超分辨率重建现有算法和常用数据库,但仅限于基于卷积神经网络(Convolutional Neural Network, CNN)的方法,且未系统归纳和对比相关的算法。文献^[23]从网络类型、网络结构、训练方法等方面详细分析

了现有技术的优势和不足,但涉及的算法较少,不够全面。文献^[24]从网络结构和优化目标出发进行了较为全面的阐述,分析整理了基于深度学习的图像 SR 技术,但分类不够明确。这些综述工作极大推动并促进了该领域的研究与发展。然而,随着时代的发展,超分辨率重建技术也在突飞猛进,近几年来该领域的研究成果可谓日新月异,各种优秀算法也是层出不穷。因此,对这些新方法和工作进行及时归纳总结,将会有效推动和促进计算机视觉领域的发展。本文在上述工作基础上,对图像超分辨率重建问题及其解决方法进行了全面系统地梳理和总结(整体框架如图 3 所示),并在传统的基于插值、基于重建和基于浅层结构的学习方法基础上,重点补充并更新了最新基于深度学习的 SR 领域的研究成果,希望为相关领域的后续研究提供有益参考。

本文后续安排如下:第 2 节给出图像超分辨率重建问题的数学描述以及常用的数据集和评价标准;第 3 节对基于图像的超分辨率方法进行了分类,并介绍了基于差值、基于重建以及基于学习(除深度学习外)的图像超分辨率重建方法;第 4 节详细讨论了各种基于深度学习的 SR 方法中网络结构、信息传递方式及损失函数构造;第 5 节系统比较了传统和基于深度学习的 SR 方法的性能指标;第 6 节给出了结论及展望。

2 图像超分辨率重建

2.1 问题描述

理想高分辨率图像一般经过几何运动变形、光学模糊、传感器模糊、降采样以及噪声污染等一系列过程得到低分辨率图像,这一过程称为图像退化

过程，如图 2 所示。

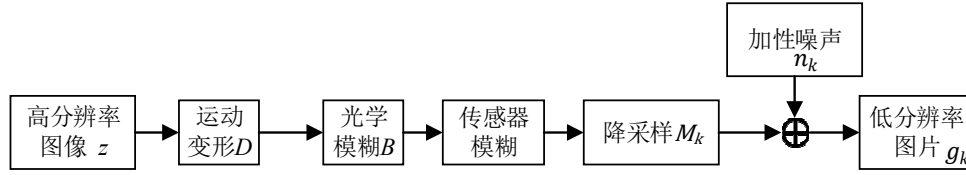


图 2. 图像退化过程

由图像的退化过程可以得到观测模型为：

$$g_k = DBM_k z + n_k \quad (1)$$

其中， g_k 表示观测得到的低分辨率图像， z 表示与 g_k 对应的待重建的高分辨率图像。 D, B, M_k 分别表示几何运动矩阵，模糊矩阵和降采样矩阵， n_k 是加性噪声。

图像高分辨率重建就是通过图像处理、机器学习等技术从观测到的低分辨率图像 g_k 或低分辨率图像序列 $\{g_k\}$ 中估计出高分辨率图像 z 的过程。根据不同的重建标准，从观测到的低分辨率图像映射得到的高分辨率图像不唯一，因此，图像超分辨率重建是一个一对多映射的病态问题。

2.2 常用数据集

图像超分辨率重建中常用的数据集包括 91-images^[62]，BSD100^[25]，Set5^[26]，Set14^[27] 和 Urban100^[28]。91-images 数据集包含 91 张图像，通常需要将图像分成图像块构成训练集；BSD100 数据集是伯克利大学 Computer Vision Group 提供的数据集，包含 100 张分辨率为 80×120 的自然图像，其拓展版本包括 BSD 300 和 BSD 500，分别包含 300 和 500 张图像，其中 200 张图像用于测试，其余图像用于训练和验证。Set5 和 Set14 分别包含 5 张和 14 张图像，是图像超分辨率重建领域最常用的测试集。Urban100 包含 100 张室外各种建筑物的图像，是公认的最具挑战性的数据集之一。表 1 总结了这些数据集。

表 1 图像超分辨率重建中常用的数据集

| 数据集 | 图像数量 | 图像分辨率 | 图像内容 |
|---------------------------|------|-------------------|-----------|
| 91-images ^[62] | 91 | 119×105 至 435×348 | 植物、水果、车辆等 |
| BSD 100 ^[25] | 100 | 80×120 | 风景、人物、动物等 |
| Set5 ^[26] | 5 | 256×256 至 512×512 | 人、动物 |
| Set14 ^[27] | 14 | 276×276 至 720×576 | 人、动物、植物 |
| Urban100 ^[28] | 100 | 256×161 至 256×256 | 城市各种建筑物 |

2.3 常用评价指标

重建高分辨率图像的质量可通过主观评价和客观评估得到。主观评价一般由观察者从视觉感官效果出发，对重建图像的优劣打分，最终评价结果受观察者个体影响较大，无统一标准。客观评估则通过计算量化评估指标衡量重建图像质量。常用的量化指标包括重建均方误差（Mean Squared Error, MSE）、峰值信噪比（Peak Noise to Signal Ratio, PNSR）和结构相似性（Structural Similarity, SSIM）。

1. MSE

重建均方误差定义了重建高分辨率图像与真实高分辨率图像之间的像素误差。

$$L_{MSE} = \|I_{est} - I_{HR}\|_2^2 \quad (2)$$

其中， I_{est} 表示重建图像， I_{HR} 表示原高分辨率图像， $\|\cdot\|_2^2$ 表示 L2 范数。

2. PNSR

峰值信噪比表示了图像最大功率和噪声功率的比值，可以用来度量重建高分辨率图像与真实高分辨率图像之间失真程度，单位为分贝(dB)。

$$PNSR = 10 \log_{10} \left(\frac{Peak^2}{MSE} \right) \quad (3)$$

其中，Peak 是图像像素可取的最大值，MSE 表示重建图像与真实高分辨率图像间像素的均方误差，见式 (2)。

3. SSIM

图像间的像素相似度不能准确反映重建效果的质量，因此，纽约大学计算视觉实验室提出了 SSIM 指标^[29]，从亮度、对比度和结构三方面度量图像间的相似性。SSIM 定义如式(4)所示，取值范围为[0,1]，值越大就表示两幅图像结构相似程度越高。

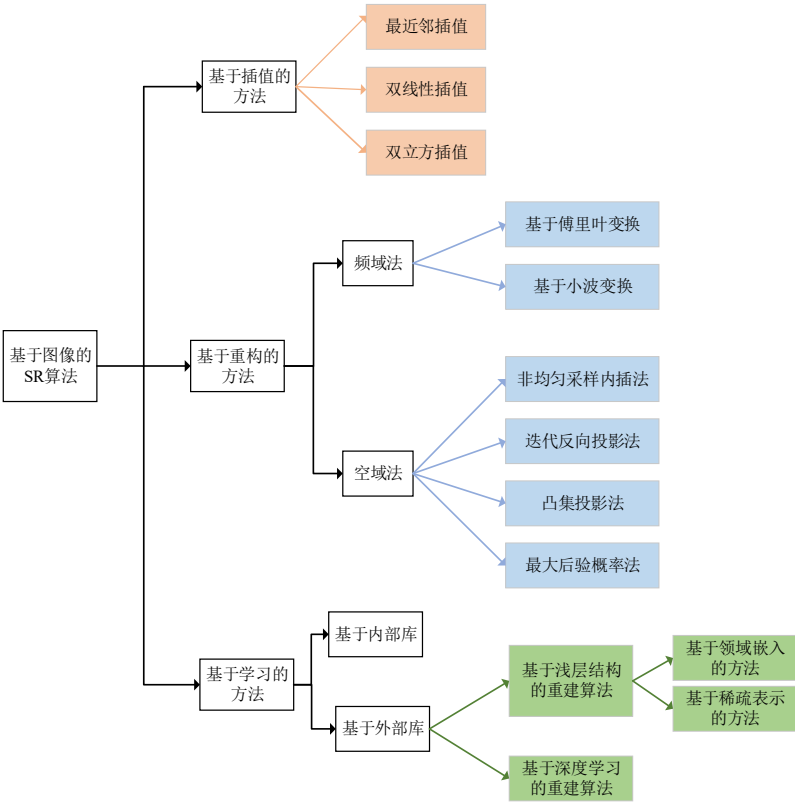
$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (4)$$

其中, u_x 和 u_y 分别为 x 和 y 的平均值, 表示对亮度的估计; σ_x^2 和 σ_y^2 分别为 x 和 y 的方差, 表示对比度的估计; 分母 $(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)$ 为协方差, 表示对结构的估计; 参数 C_1 和 C_2 是常数。

3 图像超分辨率重建的主要方法

如图 3 所示, 按算法原理, 我们将基于图像的超分辨率重建方法分为基于插值的方法、基于

重构的方法和基于学习的方法三大类。其中, 基于重构的方法包括频域法及空域法, 基于学习的算法包括基于内部库和基于外部库两类问题, 而基于外部库的方法又可进一步分成基于浅层结构和基于深度学习的方法。目前, 基于深度学习的高分辨率重建算法是该领域的研究热点, 我们将在第四节中重点介绍。



3.1 基于插值的方法

基于插值的方法^[30-36]利用一些特定核函数确定的权值, 对某个待插值点周围的像素点进行加权平均来预测该像素点的值。对不同的核函数,

由于权重可事先计算得到, 重建时只需通过查表即可得到。传统的插值方法有最近邻插值^[31]、双线性插值^[32]和双立方插值^[33], 表 2 比较三种插值方法的优缺点。

表 2 基于插值的方法比较

| 方法 | 原理 | 优点 | 缺点 |
|-----------------------|----------------------------------|----------------------------------|-----------------------------|
| 最近邻插值 ^[31] | 待插值点的灰度值由其最近邻像素点的灰度值决定 | 计算简单, 插值速度快 | 重建图像易产生块效应, 图像边缘产生不同程度的锯齿效应 |
| 双线性插值 ^[32] | 分别在水平和垂直方向上进行线性插值 | 克服最近邻插值引起的图像边缘锯齿效应和块效应 | 未考虑边缘和图像的梯度变化, 重建图像边缘较为模糊 |
| 双立方插值 ^[33] | 利用待插值点周围上下左右四个邻域内的 16 个像素点进行三次插值 | 消除图像边缘锯齿效应和块效应, 重建图像的视觉效果好于前两种算法 | 计算量大, 运算时间长, 实时性差 |

在传统插值方法基础上, 不断有学者提出了新的插值算法。文献^[34]结合传统插值方法和小波变换, 利用 Sobel 算子设定插值点的边缘方向得到初始放大图像, 叠加小波变换提取的图像高频

成分, 经过小波逆变换得到高分辨率图像。该算法考虑到了全局相关性, 得到更加清晰的边缘信息, 使重建图像具有更高的信噪比和更好的视觉效果, Zhang 等人^[35]在基于定向滤波和数据融合

的基础上,提出了一种边缘引导非线性插值方法,该方法既保持边缘清晰度又减少了振铃伪影。Shi 等人^[36]提出了一种基于边缘的扩展方法,通过对由双线性或双三次插值得到的扩展图像,使用 Canny 边缘检测器来减少锯齿形边缘噪声,并修改边缘邻域中的像素值以减少重构模糊。

插值重建法原理简单,运算速度快,实时性好,但难以修复图像的高频细节,尤其当放大因子增加时,重建性能会急剧下降,出现高频细节丢失、边缘模糊等问题。

3.2 基于重建的方法

如前所述,基于重建的图像超分辨率需要求解式(1)所示的图像观测模型的逆问题(即由观测到的低分辨率图像重建高分辨率图像)。通常要对低分辨率图像序列进行亚像素精度对齐,得到高低分辨率图像间的运动偏移量,构建观测模型中的空间运动参数,并应用不同的约束条件(如平滑性、能量有限性和非负性等)求解高分辨图像。在基于重构的超分辨率图像重建方法中,主要可分为频域法和空域法两大类。

3.2.1 频域法

1984 年, Tsai 和 Huang^[17]首次提出了在频域内提高图像分辨率的思想,利用一个低分辨率图像序列重建一幅高分辨率图像。它的主要思想是假定低分辨率图像生成模型后,分别对低分辨率图像和原始高分辨率图像进行离散傅里叶变换和连续傅里叶变换,在频域中建立起二者之间的线性关系,从而重建出高分辨率图像。该方法假设图像中不存在运动模糊和观测噪声,同时忽略了光学系统的点扩散函数,因此只适合简单理想的图像退化模型。Kim 等人^[37]在超分辨率重建过程中考虑观测噪声和空间模糊等因素,将递归最小二乘法应用于高分辨率图像的重建。文献[38]使用离散余弦变换减少了运算量,并使用多通道自适应规整化参数来减轻高分辨率重建问题的病态性。

传统频域方法只能处理输入低分辨率图像之间存在全局整体运动的情况,为了更好地处理场景中存在相对物体运动的情况,文献[39-40]提出了基于小波变换的图像超分辨率方法,提高图像的局部质量。

频域重建方法算法简单,运行速度快,但该方法建立在整体平移运动的图像退化模型基础上,处理复杂退化模型的能力有限且难以加入先验知识。因此,该方法没有引起广泛关注。

3.2.2 空域法

空域法是对影响图像成像效果的空域因素(包括光学模糊、运动模糊等)进行建模。因此,基于空域的 SR 方法更接近实际应用情况。目前,常用的空域超分辨率图像重建方法有非均匀采样

内插法(Non-uniform interpolation, NUI)、迭代反投影法(Iterative back projection, IBP)、基于最大后验概率的方法(Maximum a posterior, MAP)、凸集投影法(Projection onto convex sets, POCS)和 MAP-POCS 混合算法等。

(1) 非均匀采样内插法(NUI)

非均匀采样内插法^[41]的本质仍是对未知像素点的拟合。它首先对低分辨率图像序列进行配准,并在目标图像特定位置的内容约束下,进行插值得到高分辨率图像。非均匀采样内插法简单直观,计算效率高,但是需要假定所有低分辨率图像和模糊特征服从相同的噪声分布,在图像复原阶段忽略了插值阶段的误差,并且难以加入先验约束知识,使得算法适应性不好,重建效果也不佳,因此应用较少。

(2) 凸集投影法(POCS)

凸集投影法^[42]的基本过程是利用图像的先验知识获得一定的约束条件(平滑性、非负性、能量有限性和目标图像峰值像素的约束等),把约束条件定义为一组约束凸集合,然后通过求解这些集合的交集获得高分辨率图像解空间。对于超分辨率图像空间中的任意一点,经过多次迭代运算,可以得到一个满足所有凸集约束条件的收敛解。

该算法的优点在于易加入先验知识,能较好地恢复图像的边缘和细节。但初始点的随机选择会导致解不唯一,且迭代过程收敛慢、计算量大。

(3) 迭代反投影法(IBP)

迭代反投影法^[43]首先通过对原始低分辨率图像 y 进行插值,得到预估的高分辨率图像 x_0 ,并利用先验知识建立观测模型,将 x_0 输入式(5)所示的观测模型中得到预估的低分辨率图像 y_0 。

$$y_0 = W_k x_0 + n_k \quad (5)$$

计算 y_0 与 y 之间的误差,用此误差反向投影来修正 x_0 得到新的高分辨率图像 x_1 :

$$x_1 = x_0 + H^{BP}(y - y_0) \quad (6)$$

其中, H^{BP} 表示后向投影算子。不断重复该迭代过程,直到重建误差满足要求。

该方法原理简单,收敛快,但是难以加入高分辨率图像的先验信息,反向投影算子 H^{BP} 较难选择且通常情况下解不唯一。由于每次迭代的误差均匀地累加到重建图像上,所以图像边缘存在一定程度的锯齿。

(4) 最大后验概率法(MAP)

最大后验概率方法是由 Schultz 等人^[44-45]提出的一种基于概率的算法,其基本思想是用 Markov 随机场对图像建模,在给定输入低分辨率图像序列的条件下,使重建高分辨率图像的后验概率最大化。

若将高分辨率图像 H 和低分辨率图像序列

L_1, L_2, \dots, L_k 视为两个随机过程, 则有:

$$\hat{H}_{map} = \arg \max_H [-\ln P(L_1, L_2, \dots, L_k | H) - \ln p(H)] \quad (7)$$

式中, \hat{H}_{map} 表示对高分辨率图像 H 的估计;

$P(L_1, L_2, \dots, L_k | H)$ 表示已知高分辨率图像为 H 的条件下, 低分辨率图像序列为 L_1, L_2, \dots, L_k 的概率; $p(H)$ 表示高分辨率图像 H 的先验概率; $\ln p(H)$ 体现图像的先验知识, 称为正则化项, 用来防止出现病态的重建结果。

基于 MAP 的高分辨率重建方法的研究主要集中在先验知识 (如边缘先验^[46]) 的加入、正则化项的形式 (Tikhonov 正则项^[47], 全变差正则化^[48], 双边全变差正则化^[49]等)、正则化项的系数选取方法 (利用 L 曲线^[50], U 曲线^[51]求解) 以及 MAP 与其他算法的结合 (如 MAP-POCS 算法^[52]) 等方面。

该方法在求解过程中加入先验约束, 可以确保解存在并且唯一, 收敛稳定性高, 但是收敛慢, 计算量大, 边缘保持能力不如 POCS 方法, 重构得到的高分辨率图像的细节易被平滑掉。

(5) MAP-POCS 算法

MAP-POCS 算法^[52]利用 POCS 易于加入先验约束条件的特点, 将 MAP 估计子作为一个凸集约束加入约束集合中, 构建新的凸集, 然后对 MAP 进行梯度下降迭代, 再进行数据一致性约束迭代求解出高分辨率图像。该算法结合了 MAP 和 POCS 两种算法的优点, 被认为是基于重构的超分辨率算法中最好的方法。

表 3 比较了基于频域和基于空域重构方法的特点和复杂度。不难看出, 与其它算法相比, MAP-POCS 算法具有更好的综合性能。

表 3 基于频域和空域的 SR 方法比较

| | 频域方法 | 空域方法 | | | | |
|------------|------|------|-----|------|-----|----------|
| | | NUI | IBP | POCS | MAP | MAP-POCS |
| 图像先验知识 | 难加入 | 难加入 | 难加入 | 易加入 | 易加入 | 易加入 |
| 解是否唯一 | 是 | 是 | 否 | 否 | 是 | 是 |
| 复杂退化模型处理能力 | 差 | 中等 | 中等 | 高 | 高 | 高 |
| 计算复杂度 | 小 | 中等 | 中等 | 大 | 大 | 大 |
| 边缘保持能力 | 差 | 差 | 差 | 较好 | 差 | 好 |
| 适应性 | 较差 | 差 | 中等 | 高 | 高 | 高 |

3.3 基于学习的方法

基于学习的方法也称为基于样例 (Example-based) 的图像 SR, 其基本思想是通过训练学习低分辨率图像与高分辨率图像之间的映射关系, 获得先验知识, 用于高分辨率图像的重建。

基于学习的高分辨率重建方法一般包含: 训练集构建、特征学习和高频细节重建等三个步骤。训练集构建主要通过对高分辨率图像进行降采样等降质操作得到对应的低分辨率图像; 特征学习利用机器学习算法学习低分辨率图像到高分辨率图像的映射关系; 高频细节重建则根据学习到的高-低分辨率图像映射关系, 从低分辨率图像中恢复出高分辨率图像。按照训练集样本的组成, 可以将基于学习的方法分为基于内部库和基于外部库的方法。在基于内部库的方法中, 训练样本集直接取自输入图像, 对输入图像分块并由输入图像块在不同分辨率下的图像块集合组成, 通过挖掘输入图像在不同尺度间的自相似性重建其高分辨率图像。而基于外部库的方法中, 训练样本集由同一尺寸的低分辨率-高分辨率图像对组成, 这些图像对来自外部数据库, 与输入图像无关。

3.3.1 基于内部库的方法

基于内部库的方法假设图像^[53-55]具有自相似性, 即图像的局部内容及结构在不同尺度间或同一尺度内反复出现。训练集样本通常由对输入图像作多尺度变换得到的不同分辨率下的图像组

成。

Freeman 等人^[53]认为两个相似的低分辨率图像块对应的高分辨率图像块也应相似, 两个相邻的高分辨率图像块重叠部分像素值应该一致。他们使用马尔科夫网络分别对高低分辨率图像对之间, 相邻的高分辨率图像块之间进行建模, 对输入的低分辨率图像搜索拼接得到对应的高分辨率图像。Glasner 等人^[54]将同一尺度下图像自相似性和跨尺度的图像相似性相结合, 通过搜索—粘贴的方式逐级放大图像, 填充高频细节, 实现高分辨率重建。在此基础上, Freedman 等人^[55]在图像块搜索阶段将全局搜索和局部搜索相结合, 加快搜索速度。为快速稳定地进行高分辨率重建, 这两种方法均从小的放大因子开始, 不断迭代至需要的放大倍数。

3.3.2 基于外部库的方法

基于外部库的方法主要是学习高低分辨率图像块之间的映射关系。训练集的样本由同一尺寸的低分辨率(LR)—高分辨率(HR)图像对构成。根据学习算法的结构层数, 可进一步分为基于浅层结构和基于深度学习的重建算法。基于浅层结构的重建算法通常包含 2-3 层网络结构, 其中邻域嵌入算法和基于稀疏表示的方法是其代表算法。

(1) 基于邻域嵌入 (Neighbor Embedding, NE) 的方法

基于邻域嵌入方法的基本思想是假设高分辨率图像与其对应的低分辨率图像块在特征空间中

具有相似的局部流形。利用邻域嵌入结合其它约束条件,如邻域嵌入结合最小二乘法 (Neighbor Embedding + Least Square, NE+LS), 邻域嵌入结合局部线性嵌入^[56](Neighbor Embedding + Locally Linear Embedding, NE+LLE), 以及邻域嵌入结合非负最小二乘法^[57] (Neighbor Embedding + Non-Negative Least Square, NE+NNLS)得到一组权值,使得低分辨率图像与其在样本中对应的 K 个最近邻图像块加权得到的低分辨率图像误差最小,对高分辨率图像块应用这组系数进行加权即可得到重建的高分辨率图像。

邻域嵌入方法简单直接,对于样本集的依赖性小,但是最近邻图像块数目 K 需人为选定,易导致过拟合或者欠拟合的现象,从而影响重建图像质量。

(2) 基于稀疏表示 (Sparse Coding, SC) 的方

法

基于稀疏表示的方法^[58]在 LR-HR 样本库中建立一对共享系数,并生成对应最小重建误差的过完备字典,对待重建的低分辨率图像利用 LR 字典计算其稀疏表示系数,将同样的系数应用于 HR 字典上生成 HR 特征,得到最终的重建高分辨率图像。

稀疏表示法是对基于领域嵌入方法的拓展,克服了后者需要确定最近邻图像块数目的缺点。基于稀疏表示的方法可扩展性较强,且对噪声不敏感,但图像的重建质量和计算复杂度均与字典大小相关。当建立的字典完备性不强时,图像边缘细节较模糊。

基于外部库的方法还包括基于深度学习的高分辨率重建方法,这类方法将放在下一章重点讨论。

表 4 常用的基于浅层结构的 SR 方法

| 算法名称 | 基本原理 | 特点 |
|---|-------------------------------------|--|
| Self-Ex ^[59] | 基于图像的自相似性 | 利用图像内部信息,不需要建立外部库,但搜索时运算量大 |
| NE+ ^[56-57] | 基于高低分辨率图像对具有相似的局部流形 | 需要建立 LR-HR 样本库,学习低分辨率图像到高分辨率图像的映射关系,能获取更多高频信息,生成的高分辨率图像细节更丰富 |
| SC ^[58] | 基于高低分辨率图像对具有相同的稀疏表示系数 | |
| K-SVD ^[60] | 利用 K-SVD 方法进行高效的过完备字典建立,是对 SC 方法的改进 | |
| ANR(Anchored Neighborhood Regression) ^[61] | 稀疏表示和邻域嵌入方法的结合,利用全局协同编码加快运行速度 | |

4 基于深度学习的图像超分辨率重建

深度学习方法是近十年来人工智能领域取得的重要突破之一。它在计算机视觉、图像与视频分析、自然语言处理和多媒体信号处理等领域取得了巨大成功。它可以通过建立强有力的深层模型,重复利用中间层的计算结果减少模型参数,从像素级原始数据到抽象的语义概念逐层提取信息,得到图像全局特征和目标上下文信息。2014 年 Dong 等人^[62]提出一个含三个卷积层的网络-SRCNN(Super-resolution CNN),首次将深度学习用于图像 SR 领域,此后,基于深度学习的 SR 网络不断出现并改进。如图 4 所示,我们根据网络类型将基于深度学习的图像 SR 方法分为基于卷积神经网络(CNN)的 SR 方法和基于其它深度学习模型的 SR 方法。其中使用 CNN 结构的 SR 方法,根据网络层数可分为基于浅层网络(3 层及以下的 CNN)、基于深层网络(4-10 层的 CNN)和基于极深层网络(10 层以上的 CNN)的方法;根据组成网络的基本模块结构差异,基于极深层网络的方法又可进一步分为基于残差结构和基于密集连接的两大类,有些网络如 MemNet^[79], RDN^[80]则同时借鉴了这两种结构,而基于残差结构的 SR 方法中又包含了一类很特殊的方法,即基于生成对抗网络(Generative Adversarial Net, GAN)的 SR 方法。

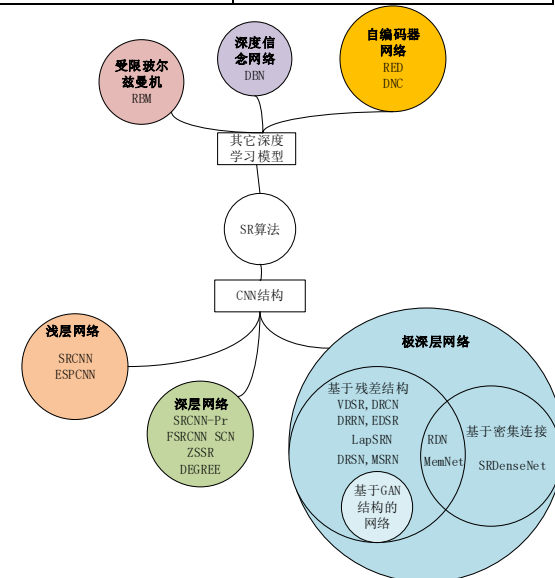


图 4 基于深度学习的图像超分辨率重建方法分类

4.1 基于浅层网络的图像超分辨率重建

(1) SRCNN

SRCNN^[62]借鉴了基于稀疏编码的图像超分辨率思想,通过卷积操作完成图像块的提取、特征非线性映射和重建过程三个步骤,其网络结构如图 5 所示。首先通过双三次线性插值将低分辨率图像插值到高分辨率图像尺寸作为输入。利用第一层卷积对图像密集采样,得到一系列图像块,每个图像块表示成一个高维向量,这些高维向量

组成 n_1 个特征图；第二层卷积将表示低分辨率图像的 n_1 个特征图非线性映射到表示高分率图像的 n_2 个特征图；第三层卷积采用全连接方式将预测得到的高分辨率图像块整合生成一张高分率图像。

SRCNN 将真实高分率图像和重建高分率图像间的均方误差 (MSE) 作为损失函数 (如式 (8) 所示)。

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n \|F(Y_i; \theta) - X_i\|^2 \quad (8)$$

式中, $L(\theta)$ 表示损失函数, θ 表示网络优化参数, n 表示训练样本集大小, Y_i 表示训练集中第 i 幅低分辨率图像, $F(Y_i; \theta)$ 表示由 Y_i 映射得到的高分辨率图像, X_i 表示第 i 幅真实高分率图像。

训练时, 采用动量梯度下降法(Momentum)最小化损失函数 $L(\theta)$, 进一步提升训练速度。当放大因子为 2,3,4 时, 该方法在测试集 Set5, Set14 上的重建图像的平均峰值信噪比(PSNR)比三次插值[33], SC[58], ANR[59]等传统方法高至少 0.3dB, 平均重建速度较最快的传统方法提升两倍左右。SRCNN 在图像重建速度和重建质量上均优于传统方法, 是深度学习在图像高分率重建问题任务中首次应用, 但该网络存在以下问题: 1) 只适用于单放大因子的图像高分率重建, 针对不同的放大因子, 网络模型需要重新训练; 2) 由于输入需要先插值至目标尺寸大小, 整个图像重建过程均在高分率空间中进行, 导致计算量大, 训练速度慢; 3) 网络比较浅, 卷积核感受野较小 (13×13) 不能很好地利用图像上下文信息。

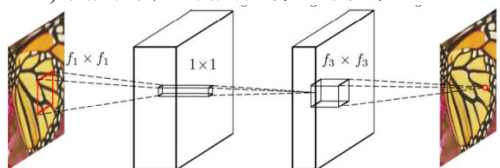


图 5. SRCNN 网络结构图[62]

(2) ESPCNN (Efficient Sub-Pixel Convolutional Neural Network)

为提高 SRCNN 的计算效率, 文献[63]提出了一种直接在低分辨率图像尺寸上提取特征, 计算得到高分率图像的高效方法—ESPCNN。如图 6 所示, 该网络主要有两部分组成: 1) 隐藏层: 原始低分辨率图像通过由两个卷积层组成的隐藏层后, 得到 $r^2 \times H \times W$ (r 是放大因子; H, W 分别是目标图像的高度和宽度) 的特征图像。2) 亚像素卷积层: 将每个像素的 r^2 个通道重新排列成一个 $r \times r$ 的区域, 对应于高分率图像中的一个 $r \times r$

大小的子块, 从而特征图像被重新排列成 $1 \times rH \times rW$ 大小的高分辨率图像。

由于 ESPCNN 只在最后一层对图像大小做变换, 前面的卷积运算都在低分辨率图像上进行, 因此运算效率较高。在 Set14 上进行放大三倍的图像 SR 测试, 每幅图像的平均重建时间是 4.7ms, 而 SRCNN 为 0.18s。ESPCNN 网络可以用于图像和视频的实时重建, 重建质量和 SRCNN 相当。

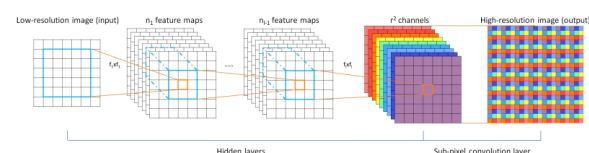


图 6. ESPCNN 网络结构图[63]

4.2 基于深层网络的图像超分辨率重建

(1) FSRCNN (Fast Super-Resolution by CNN)

针对 SRCNN 算法的缺点, FSRCNN[64]进行了相应的改进, 其改进主要体现在三个方面: 1). 为了减少计算量, 加快训练速度, FSRCNN 取消了对低分辨率图像进行双三次插值放大尺寸的步骤, 而是直接将原始的低分辨率图像输入到网络中; 2). 使用更小的卷积核和更多的映射层改变特征维数, 实现多通道的非线性映射和特征融合; 3). 通过微调最后一层反卷积层来训练不同上采样因子的模型。

图 7 比较了 SRCNN 和 FSRCNN 的网络结构, 由图可见, FSRCNN 的网络结构可以分为五个部分: 1). 特征提取: 直接对原始低分辨率图像进行操作; 2). 收缩: 使用 1×1 卷积核进行降维, 减少网络参数, 降低计算复杂度; 3). 非线性映射: 采用两个串联的 3×3 卷积核替代 SRCNN 中一个 5×5 卷积核, 减少运算参数; 4). 扩张: 由于低维度特征带来的重建效果不佳, 因此用 1×1 卷积核进行扩维, 该步骤是对收缩的逆操作; 5). 反卷积层: 实现图像放大。

FSRCNN 重建图像质量较 SRCNN 略好, 其主要优势在于运行速度十分快, 当放大因子为 2,

3, 4 时, FSRCNN 在测试集 Set5^[26], Set14^[27], BSD200^[25]上重建一幅高分辨率图像所需平均时间仅为 SRCNN 的十分之一, 可以达到毫秒级, 可以用于实时图像高分辨率重建。

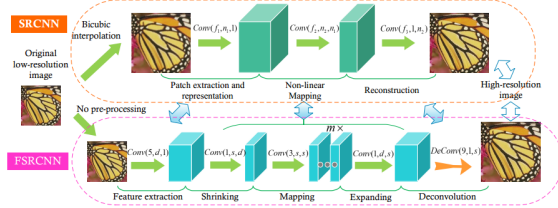


图 7. FSRCNN 网络结构图[64]

(2) SRCNN-Pr

SRCNN-Pr 是 Liang 等人^[65]提出的利用图像边缘先验信息和不同放大尺度间的图像自相似特性进行图像多尺度放大的网络, 其网络结构如图 8 所示。输入的低分辨率图像同时经过多个放大因子对应的网络实现多任务学习, 图像在不同尺度间的自相似性使得网络的第二层卷积间可以进行权值共享, 缓解参数过多的问题。网络末端的卷积层用于提取重建图像的边缘特征。在 SRCNN 的损失函数的基础上, 通过添加重建图像的梯度信息和目标 HR 的梯度信息一致的约束条件来引入图像边缘梯度先验。整个网络损失函数由重建损失 $Loss_s$ 和梯度信息损失 $Loss_{pr}$ 两部分组成, 如式 (9) 所示, 其中 α 是权重系数。

$$Loss = Loss_s + \alpha Loss_{pr} \quad (9)$$

当放大因子为 3 时, 在 Set5 测试集上, 该方法训练速度比 SRCNN 快, 重建图像的 PNSR 值较 SRCNN 高 0.01~0.02db, 将先验知识和深度学习特征提取相结合是 SRCNN-Pr 方法的创新之处, 但领域先验知识对重建图像的质量提升有限, 如何实现高质量的大放大倍数图像重建是 SRCNN-Pr 需要解决的问题。

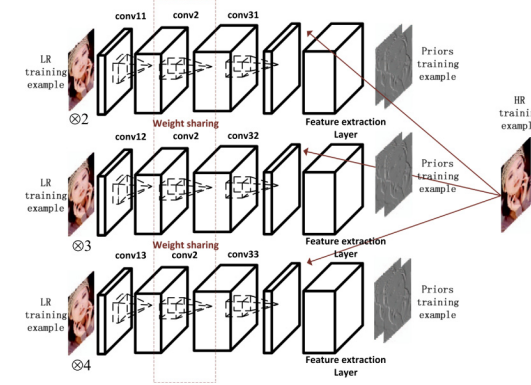
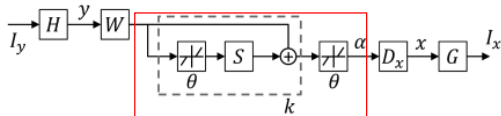


图 8. SRCNN-Pr 网络结构图[65]

(3) SCN(Sparse Coding Based Network)

SCN^[66]是 Wang 等人提出的一种通过神经网络实现基于稀疏编码的高分辨率图像重建网络。该方法主要思想是使用 LISTA (Learned Iterative Shrinkage and Thresholding Algorithm)将基于稀疏表示的 SR 方法中的稀疏表示、映射和重建三个模块纳入同一个稀疏编码网络 SCN 中, 并通过级联 SCN 可实现任意尺度的图像放大。SCN 网络结构如图 9 所示, 它将低分辨率图像 I_l 通过卷积层 H 分割为图像块并提取其特征 y , y 经过 LISTA 网络 (图中红色框部分) 可以得到 I_l 的稀疏编码 α , α 通过由高分辨率图像字典 D_x 对应的卷积层得到高分辨率图像块 x , 卷积层 G 将得到的高分辨率图像块 x 连接重建出一幅高分辨率图像 I_x 。

SCN 的思想与传统的基于稀疏编码的高分辨率重建方法完全一致, 但使用深度网络实现并进行训练, 避免了传统稀疏编码方法需要大量优化计算的缺点。相比 SRCNN 这类完全基于数据驱动的网络而言, 稀疏编码网络的每一层都具有很清楚的物理意义, 可解释性强, 通过 LISTA 子网络结构加入图像的先验知识可以使得重建图像有更好的视觉效果。文献[66]结果表明, 当放大因子为 2, 3, 4 时, SCN 的 PNSR 值较传统的基于稀疏表示的方法高 0.3~1.6dB, 较 SRCNN 方法提高 0.2~0.5dB。

图 9. SCN 网络结构图^[66]

(4) ZSSR ("Zero-Shot" Super-Resolution)

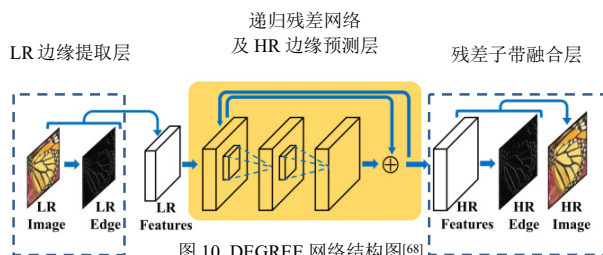
Shocher 等人^[67]提出了一种利用图像自相似性,进行复杂退化情况下的高分辨率图像重建的自监督式 CNN 网络-ZSSR。该网络不需要预训练,在测试时,利用输入图像不同分辨率下的图像构成训练集进行网络训练,学习如何从特定的低分辨率图像中恢复出高分辨率图像。该方法对输入图像训练网络模型,测试图像是模型从未遇到过的样本,属于针对特定图像(image-specific)的零样本重建方法。在理想退化情况下,该方法得到的重建图像质量和当时最好的 SR 方法 EDSR 相当,复杂退化情况下,重建图像 PNSR 值比 EDSR 高 1~2db。由于 ZSSR 不需要预训练过程,网络规模小,其训练和测试时间总和与 SRCNN 测试时间相当,运行速度较 EDSR 稍快。

(5) DEGREE (Deep Edge Guided Recurrent Residual network)

Yang 等人^[68]提出一种加入边缘先验的深度残差网络-DEGREE。如图 10 所示,该网络包含四个部分,1). LR 边缘提取层:提取图像边缘作为先验信息,然后将低分辨率图像及其边缘特征

图作为网络输入;2). 递归残差网络:将图像信号分解为边缘、纹理等成分,作为不同的子频带;3). HR 边缘预测层:实现各成分的重建;4). 残差子带融合层:将各子频带的重建信息融合生成高分辨率图像。

网络的损失函数由边缘损失和重建损失组成,均采用均方差形式。放大因子为 2, 3, 4 时,重建图像的 PNSR 值较 SRCNN 提升了 0.4~1.1db,将自然图像先验信息嵌入到图像高分辨率重建中,可以更好地恢复细节,获得更好的视觉效果,但网络复杂,运行速度慢,在大放大因子下,重建效果不好。

图 10. DEGREE 网络结构图^[68]

在 SRCNN 网络基础上,各种后续方法分别从提高运算速度,提升大放大因子下重建图像质量,加入图像先验知识提升重建图像质量等方面着手试图解决图像高分辨率重建中面临的难点和问题,并取得了较好的效果。表 5 总结对比了上述各种方法。

表 5 基于浅层及深层 CNN 的图像高分辨率重建方法对比

| 算法 | 网络层数 | 主要思想 | 领域先验知识及其加入形式 | 损失函数 |
|----------|------|-------------|---------------------|--------------------|
| SRCNN | 3 | SR 与 CNN 结合 | -- | 重建 MSE 损失 |
| ESPCNN | 3 | 逐级放大重建 | -- | 重建 MSE 损失 |
| FSRCNN | 8 | 加快运算速度 | -- | 重建 MSE 损失 |
| SRCNN-Pr | 4 | 加入梯度先验 | 梯度先验:边缘提取层+梯度信息损失函数 | 重建 MSE 损失+梯度信息损失函数 |
| SCN | 7 | SC 方法网络化 | 稀疏先验:LISTA 子网络 | 重建 MSE 损失 |
| ZSSR | 8 | 适应复杂退化模型 | 图像自相似性:全连接层 | 重建误差(L1 范数) |
| DEGREE | 10 | 加入边缘先验 | 边缘先验:边缘提取层+边缘信息损失函数 | 重构 MSE 损失 |

4.3 基于极深层网络的图像超分辨率重建方法

一般而言,网络越深越宽,其感受野越大,能利用的上下文信息越多,学习能力及层次特征表征能力越强,因此,随着深度学习在图像超分辨率重建领域的发展,越来越多学者尝试在加深

网络的同时使用跳连接(skip connection)和密集连接(dense connection)来增强网络的特征学习能力,应用残差学习,梯度裁剪,参数共享等策略解决深度网络参数量大,不易收敛以及梯度爆

炸/梯度消失等问题。

4.3.1 基于残差结构的超分辨率重建方法

(1) VDSR(Very Deep Convolutional Networks for SR)

VDSR^[69]是第一个将残差结构用于图像 SR 问题的深度网络。该方法认为低分辨率图像携带的低频信息与高分辨率图像的低频信息相近,因此只需要学习高分辨率图像和低分辨率图像之间的高频残差信息即可。VDSR 网络结构如图 11 所示,它将插值后得到具有目标尺寸大小的低分辨率图像作为网络输入,再将这个图像与网络学到的残差相加得到最终的高分辨率重建图像。

VDSR 的创新之处在于:1)网络深度被大大增加,整个网络共包含 20 层,同时越深的网络层拥有更大的感受野。感受野由 SRCNN 中的 13×13 变为 41×41 。感受野增大意味着网络能利用更多的上下文信息来预测图像细节。2)将残差学习思想融入 SR 问题,残差图像比较稀疏,大部分值都为 0 或者比较小,极大减轻了网络计算负担,加快了收敛速度。VDSR 还应用了自适应梯度裁剪(Adjustable Gradient Clipping),将梯度限制在某一范围,也一定程度上加快了收敛过程,避免了梯度爆炸问题。3)VDSR 在每次卷积前都对图像进行补 0 操作,保证了所有特征图和最终的输出图像在尺寸上保持一致,使得重建图像边缘更清晰。4)VDSR 将不同放大倍数的图像混合在一起训练,使得模型可以解决不同放大倍数的图像超分辨率问题。

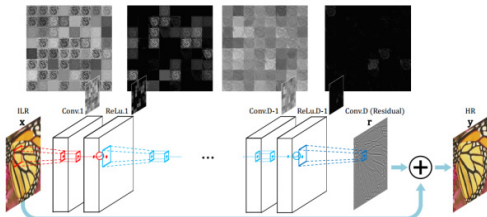


图 11. VDSR 网络结构图^[69]

(2) DRCN (Deeply-Recursive Convolutional Network)

Kim 等人进一步提出了一个与 VDSR 类似的网络—DRCN^[70],其网络结构如图 12 所示。DRCN 将递归神经网络(Recursive Neural Network)结构和残差学习思想应用于图像 SR 中。DRCN 分为三个模块:1)嵌入网络:用于特征提取;2)推理网络:使用递归结构加强层间的信息传递和上下文信息联系,学习高分辨率图像与低分辨率图像的差,恢复图像的高频部分;3)重建网络:将每一层的卷积结果和输入图像相加可得一个重建结果,对所有卷积做相同操作总共可得多个重建结果,再把它们加权平均可以得到最终的输出。在网络结构上,DRCN 中递归层之间参数共享,

可有效减少网络参数,同时用递归监督方法缓解梯度消失/爆炸问题,在训练过程中,将每一个递归层输出的误差和总输出的误差之和作为网络优化的目标函数。

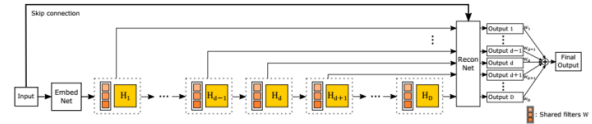


图 12. DRCN 网络结构图^[69]

(3) DRRN (Deep Recursive Residual Network for SR)

VDSR 的基本思想来源于全局残差学习,DRCN 将全局残差学习与单权重的递归学习以及多目标优化结合起来,Tai 等人^[71]受到 VDSR 和 DRCN 的启发,提出了一种结合多路径模式的局部残差学习和全局残差学习以及多权重递归学习的 SR 方法—DRRN。DRRN 一共 52 层,相较于 VDSR 的 20 层网络结构,网络规模更大,计算更复杂,重建图像质量仅有小幅度的提升。

(4) EDSR/MDSR (Enhanced Deep Residual Networks for SR)

文献 [72]去除残差块中的批归一化 (Batch Normalization, BN)层建立起一个用于单尺度高分辨率重建的网络模型—EDSR(图 13(a)所示)和多尺度高分辨率重建的网络模型—MDSR(图 13(b)所示)。

EDSR 在训练时,先训练低倍数的上采样模型,接着用训练低倍数上采样模型得到的参数来初始化高倍数的上采样模型,这样能减少高倍数上采样模型的训练时间,同时训练结果也更好。MDSR 是一个能同时进行不同上采样倍数放大的网络,MDSR 的中间部分和 EDSR 一样,只是在网络前面添加了不同放大倍数的预训练模型来匹配不同放大倍数的情况。在网络最后,将不同倍数的上采样结构平行排列获得不同倍数的输出结果。

在 EDSR 和 MDSR 网络架构中,由于去掉了 BN 层,因此,减少了内存使用,提升了训练速度。同时,这两个网络模型充分利用图像不同尺度间的内部关系,用低倍数的上采样模型作为预训练模型来训练高倍数的上采样模型,减少了模型参数,加快了训练,并且提升了图像的重建质量。

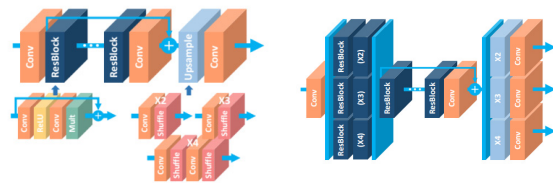


图 13. EDSR 和 MDSR 网络结构图^[72]

(5) LapSRN (Laplacian Pyramid Networks for SR)

SRCNN 的输入层中,需要使用插值将图像放

大到理想分辨率尺寸,这样增加了额外的计算开销,也会导致可见的重建伪影。而在 ESPCNN^[63]使用了亚像素卷积层来替换插值,但简单的浅层网络并不能很好地学到低分辨率图像到高分辨率图像的复杂映射。文献 [73]提出了一个基于拉普拉斯金字塔模型的网络结构 LapSRN。如图 14 所示, LapSRN 通过残差学习和逐级放大操作恢复高分辨率图像。每个放大模块完成两倍上采样操作,这些放大模块结构相同,模块间参数共享,模块级联完成逐级放大,在做高倍上采样时,也能得到中间低倍上采样结果的输出。

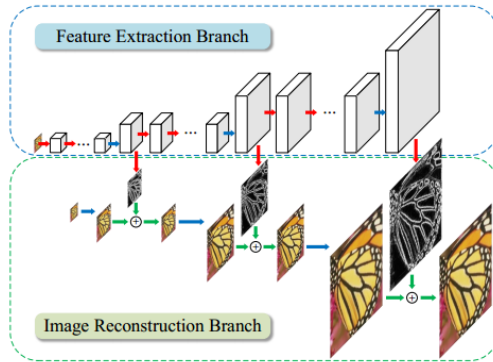


图 14. LapSRN 网络结构^[73]

为了避免 L2 型损失函数带来的边缘平滑和模糊问题,文中使用式 (10) 所示的 charbonnier 损失函数。

$$L(\hat{y}, y, \theta) = \frac{1}{N} \sum_{i=1}^N \sum_{s=1}^L \rho((\hat{y}_s^{(i)} - x_s^{(i)}) - r_s^{(i)}) \quad (10)$$

其中, N 是训练集中的图像数目; L 表示总的级数; ρ 表示 Charbonnier 惩罚函数; s 是级联的级数(取 1, 2, 3); $\hat{y}_s^{(i)}$ 表示第 s 级第 i 张重建图像; $x_s^{(i)}$ 表示第 s 级的第 i 张低分辨率图像; $r_s^{(i)}$ 表示学习到的第 s 级的第 i 张高分辨率图像和低分辨率图像之间的残差。

(6) DSRN (the Dual-State Recurrent Network)

Liao 和 Poggio 等人^[74]证明了一个权重共享的残差网络等同于一个浅层循环神经网络 (Recurrent Neural Network, RNN), 如图 16 (a) (b) (c) 所示。基于此, Han 等人^[75]发现包含递归结构和权重共享的残差网络, 如 DRCN 和 DRRN, 可以统一等效成一个展开的单状态 RNN 网络, 如图 15 (d) (e) 所示。

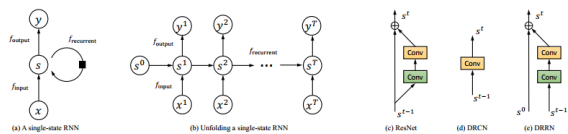


图 15. 残差网络等效于浅层循环神经网络示意图

(a) 单状态 RNN; (b) 展开的单状态 RNN; (c) 残差网络; (d) DRCN 网络; (e) DRRN 网络

单状态循环网络只将 LR 空间信息传递到 HR

空间,为了能同时利用 LR 和 HR 空间的特征,文献^[75]提出一种双状态循环网络—DSRN, 等效于 T 个参数共享的残差块级联。DSRN 的网络结构如图 16 所示, 图中箭头表示两个状态 (s_h : 高分辨率状态和 s_l : 低分辨率状态) 间的传递函数, 表 6 给出了该网络传递函数作用及其实现方式。可以看出, 该网络信息流动方向包括 LR-HR(黄色)、HR-LR(绿色)、HR-HR(橙色)和 LR-LR(蓝色), 这种设计加强了不同分辨率之间信息共享。DSRN 仍采用重建图像和真实图像的均方误差作为损失函数, T 个残差块的输出分别和输入低分辨率图像相加得到 T 个重建结果, 再将它们的均值作为最终输出, 这一点与 DRCN 类似。实验结果表明, 使用参数共享的网络比不使用的重建图像 PSNR 值高 0.2db, 双状态循环网络比单状态循环网络高出 0.3db 左右, 且重建图像有更好的视觉效果, 这也表明加强不同分辨率图像之间的信息传递, 充分利用不同分辨率图像之间的分层特征有助于更好地恢复图像的高频细节, 这一点和文献^[76,78,80]中所提出的算法核心思想相吻合, 只是实现方式不同。

表 6 DSRN 网络中传递函数及其作用和实现方式

| 传递函数 | 作用 | 实现方式 |
|--------------|-------------------------|------------|
| f_{input} | 将输入低分辨率图像变换到特征空间 | 含跳连接的两层卷积层 |
| f_{up} | 将信息从 LR 特征空间传递到 HR 特征空间 | 反卷积层 |
| f_{down} | 将信息从 HR 特征空间传递到 LR 特征空间 | 卷积层 |
| f_{hr} | 将信息从 HR 特征空间传递到 HR 特征空间 | 标准残差模块 |
| f_{lr} | 将信息从 LR 特征空间传递到 LR 特征空间 | 标准残差模块 |
| f_{output} | 将预测的 HR 特征转换为 HR 图像 | 一层卷积层 |

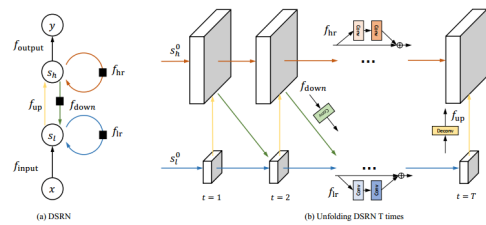


图 16. DSRN 网络结构图及其展开形式^[75]

(8) MSRN (multi-scale residual network)

文献^[76]提出了一种可以利用多尺度下图像分层特征, 进行任意尺度放大的图像 SR 网络—MSRN。如图 17 所示, MSRN 网络结构与 MemNet^[79]类似, 只是基本构成模块和模块内部连接不同。网络主要分为两部分: 1) 特征提取部分: 特征提取模块由多尺度残差模块 MSRB (multi-scale residual block) 和分层特征融合 HFFS (hierarchical feature fusion structure)两部分组成。MSRB 主要完成多尺度特征融合, 每个 MSRB 单

元使用局部残差学习减少计算复杂度,利用 1×1 卷积核组成瓶颈层提取各层特征中的有用信息,实现层次特征融合;2)重建部分:主要完成图像的放大, pixel shuffle 层使该网络能完成图像任意尺度的放大。当使用 8 个 MSRB 模块时,实验效果和 EDSR 方法相当,但网络参数只有后者的 1/7。

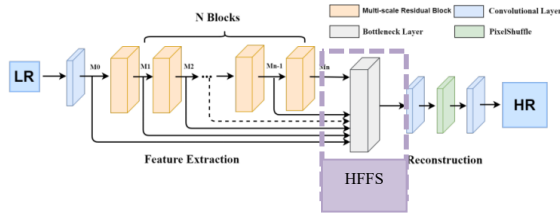


图 17. MSRN 网络结构图[76]

4.3.2 基于密集连接的网络

2016 年, Huang 等人^[77]提出密集连接网络 DenseNet (Densely Connected Convolutional Networks), DenseNet 的基本思路与残差网络 ResNet 一致,但它的组成单元是密集块(dense block)。图 18 比较了残差块和密集块两种结构的差异。DenseNet 将密集块中每一层的特征都输入给之后的所有层,使所有层的特征都串接起来,而不是像 ResNet 那样直接相加。这种结构能减轻整个网络梯度消失的问题、具有加强特征传播、支持特征复用和减少参数数量的优点。DenseNet 的出现为 SR 问题的研究注入了新活力。

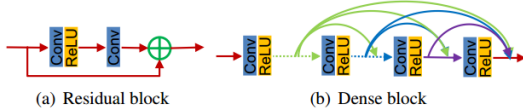


图 18. 残差块结构和密集块结构的比较

(1) SRDenseNet (Image Super-Resolution Network Using Dense Skip Connections)

SRDenseNet^[78]首先将密集块结构应用到 SR 问题上。SRDenseNet 可以分成四个部分,其网络结构如图 19 所示。首先由一个卷积层提取低分辨率图像的低层特征,然后多个密集块通过密集跳连接,将前面密集块的特征传递到后面密集块,逐级学习高层的特征,并由反卷积层对图像进行上采样,最后通过一个卷积层生成高分辨率输出。该网络结构的优点在于,通过不同层之间的密集跳连接,实现不同深度层次互补特征的共享,从而进一步提升重建图像质量。

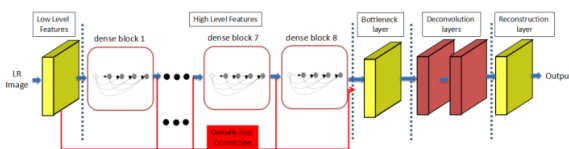


图 19. SRDenseNet 网络结构[78]

(2) MemNet (deep persistent memory network)

MemNet 模型^[79]是一个由递归单元和门控单元组成的内存块作为基本单元的深度持续记忆网

络,在网络结构上,与 DRCN 网络十分相似,在连接上,它采用的是全局密集连接,相较于 SRDenseNet 的局部密集连接,MemNet 考虑了前面状态 / 层对后续状态 / 层的影响,主要解决了深度网络中的长期依赖问题 (long-term dependency problem)。MemNet 的缺点在于前面层到后面层的直接连接,限制了长期连接的能力,使得局部信息不能完全被利用。此外,MemNet 网络在高分辨率空间中提取特征,增加了计算复杂度。

(3) RDN (Residual Dense Network for Image Super-Resolution)

考虑到之前的网络 (VDSR^[69], DRCN^[70], MemNet^[79]等)没有充分利用低分辨率图像的分层特征,文献 [80]提出了一种结合局部残差学习和全局残差学习,能充分利用所有层携带信息的网络-RDN。该网络结构如图 20 所示,主要有四个部分组成:1)浅层特征提取模块(蓝色虚线框部分):对于输入的低分辨率图像,使用两个卷积层提取浅层特征;2)局部特征提取模块(红色框部分):然后使用一系列串联的一系列残差密集块 (Residual dense block, RDB, 结构如图 21 所示)提取丰富的局部特征,允许前一个 RDB 单元的输出生直接连接到后面所有的 RDB 单元建立连续记忆机制,最大程度地保留前面一层的信息,传递层间信息。在 RDB 单元内部包含局部特征融合模块(绿色虚线框部分),利用 RDB 的局部特征融合去除局部特征中的冗余信息,学习更多有效的特征,增强网络表达能力;3)全局特征融合模块(黄色虚线框部分):对获得的密集局部特征进行全局融合,从而自适应地学习全局密集分层特征。前面所有的操作均在低分辨率空间中进行,使得计算复杂度减小,运算速度加快;4)图像重建和放大(紫色虚线框部分)。

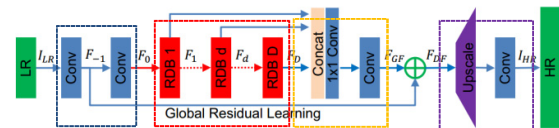


图 20. RDN 网络结构图[80]

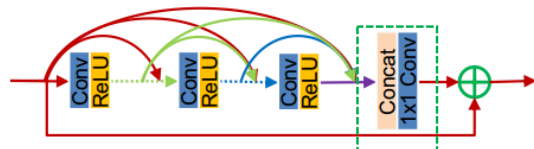


图 21. RDB 单元结构图[80]

由于 BN 层会消耗大量内存单元,增加计算复杂度,池化层会丢弃像素级信息。因此,RDN 中取消了 BN 层和池化层。该网络能对不同退化模型(如模糊下采样,降采样混合高斯噪声和双三次降采样)得到的低分辨率图像进行高分辨率重建,得到不错的重建效果。

4.3.3 基于残差结构与密集连接网络的比较

在基于残差结构的 DRCN^[70]和基于密集连接的 SRDenseNet^[78]模型基础上,各种新的网络结构不断被提出来。总体而言,相关工作集中在以下两个方面:1)前期工作主要是通过增加网络深度提升性能。研究重点在如何在增加网络深度的同时,减少网络参数,加快网络训练速度。这类方法(如 VDSR^[69],DRCN^[70])网络层间信息流动性差,没有充分挖掘图像不同层次的特征,但图像重建速度和质量比 SRCNN 好;2)第二类方法在增加网络深度的同时,从网络内部模块间的连接、模块内部各层的连接方式(跳连接、密集连接和密集跳连接)和不同层次特征的融合方式(局部

和全局残差学习)着手,加强网络内部信息的流动性,充分利用不同尺寸、不同层次的图像信息,获取高频特征用于高分辨率图像重建。典型例子包括 RDN 的基本模块内部密集连接,模块间局部信息和全局信息融合。这类网络结构较复杂,参数较多,计算复杂度高。表 7 从网络输入、包含的基本模块、模块间的连接关系、层间信息流动性和网络复杂度等方面,综合比较了基于残差学习的典型网络和基于密集连接的系列网络。可以看出,从 VDSR,DRRN,DRCN,SRDenseNet,MemNet 到 RDN,网络结构越来越复杂,网络内部层间的信息流动性逐步提高,不同层次间的信息被充分利用,同时重建图像质量也逐步提升。

表 7 基于残差学习与基于密集连接类型的网络比较

| 算法 | 输入 | 基本模块 | 模块间连接方式 | 残差学习 | 层间信息流动性 | 网络复杂度 |
|----------------------------|----------|------|--------------------|------|---------|--------------|
| VDSR ^[69] | LR+双三次插值 | 卷积层 | 串联 | 是 | 差 | 一般 (20 层) |
| DRCN ^[70] | LR+双三次插值 | 卷积层 | 串联 | 是 | 差 | 一般 (20 层) |
| DRRN ^[71] | LR+双三次插值 | 残差单元 | 串联 | 是 | 差 | 较复杂 (52 层) |
| SRDenseNet ^[78] | LR+双三次插值 | 密集块 | 密集跳连接 | 否 | 较好 | 较复杂 (68 层) |
| MemNet ^[76] | LR+双三次插值 | 内存块 | 密集连接 | 是 | 好 | 复杂 (80 层) |
| RDN ^[80] | LR | RDB | 级联 (concatenation) | 是 | 很好 | 十分复杂 (149 层) |

4.3.4 基于 GAN 的超分辨率重建方法

基于 GAN 的超分辨率重建方法中,网络内部都使用了残差模块,因此我们将基于 GAN 的 SR 方法统一并入基于残差结构的 SR 方法中。

从 Ledig^[81]等人提出将生成对抗网络应用于图像 SR 问题以来,基于 GAN 的图像 SR 方法层出不穷。根据算法研究内容,我们将其分为四大类(如图 22 所示):1)对应大放大因子下图像 SR 的 GAN 网络;2)加入先验知识的 GAN 网络;3)学习复杂退化模型的 GAN 网络;4)实现传统 SR 方法的 GAN 网络。

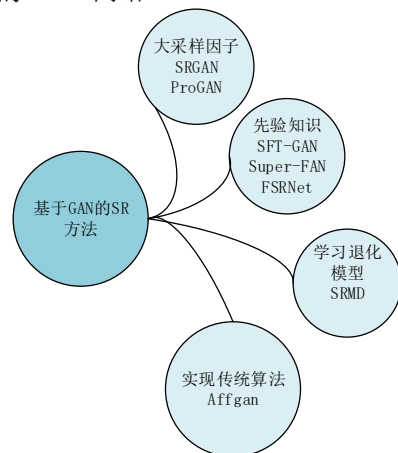


图 22. 基于 GAN 的 SR 方法分类

基于 GAN 的 SR 方法较其它 CNN 结构的 SR 方法最大不同点在于损失函数的设计。因此,我们先介绍 GAN 网络中常用的损失函数,然后再介绍基于 GAN 的各类 SR 模型。

4.3.4.1 基于 GAN 的 SR 方法中的常用损失函数

(i). GAN 的损失函数

GAN 的损失函数 $L_{GAN}(G)$ 由生成器的生成损失和判别器的对抗损失两部分组成,其表达式如式 11 所示,其中 $L_{GAN}(G)$ 和 $L_{GAN}(D)$ 分别对应生成损失和对抗损失。

$$L_{GAN} = \underbrace{E_{x \sim P_g} [1 - \log D(x)]}_{L_{GAN}(G)} + \underbrace{(-E_{x \sim P_r} [\log D(x)] - E_{x \sim P_g} [1 - \log D(x)])}_{L_{GAN}(D)} \quad (11)$$

Arjovsky 等人^[82]对 GAN 的损失函数进行改进,得到了 WGAN,其损失函数如式(12)所示,WGAN 解决了训练不稳定的问题。在基于 GAN 的 SR 问题中,文献[88,93]使用 WGAN 损失函数代替原始 GAN 损失函数。

$$L_{WGAN} = \underbrace{-E_{x \sim P_g} [\log D(x)]}_{L_{WGAN}(G)} + \underbrace{(E_{x \sim P_g} [\log D(x)] - E_{x \sim P_r} [\log D(x)])}_{L_{WGAN}(D)} \quad (12)$$

(ii). 基于像素的重建损失

$$L_{MSE} = \|I_{est} - I_{HR}\|_2^2 \quad (13)$$

其中, I_{est} 表示重建图像, I_{HR} 表示真实高分辨率图像。式 (13) 所示的损失函数就是式 (2) 定义的图像均方误差 (MSE) 函数。

(iii). GAN 的感知损失

$$L_p = \|\phi(I) - \phi(I_{HR})\|_2^2 \quad (14)$$

均方误差并不能反映输出图像和真实图像间的感知差异, 易使重建图像过于平滑, 缺少高频细节。图像特征可视化^[83]、纹理合成^[84]和图像风格转换^[85]等领域的研究表明, 使用感知损失函数能有效捕捉更高层次的图像特征差异, 从而生成高质量的重建图像。Johnson 等人^[86]首次在 SR 中提出使用感知损失代替基于像素的均方误差函数, 使生成图像和真实图像在语义和风格上更相似。如式 (14) 所示, 感知损失对应图像 I (可以是低分辨率图像 I_{LR} 或重建图像 I_{est}) 通过深度卷积神经网络提取到的图像高层次特征 $\Phi(I)$ 与网络提取到的目标高分辨率图像特征 $\Phi(I_{HR})$ 之间的均方误差。研究表明, 该方法重建出的图像较基于像素均方误差的方法在量化指标 (如 PNSR, SSIM) 上略微有所下降, 但重建图像包含更丰富真实的高频纹理细节, 并带来视觉逼真度的提升。

上述三类损失函数是基于 GAN 的 SR 方法中最常见的损失函数。具体而言, 各种基于 GAN 的 SR 方法对损失函数的选择和组合不尽相同, 且每项损失函数的表达式也略有差别。比如 SRGAN^[81] 方法对经典 GAN 模型的对抗损失函数进行了简化, 再结合式 (14) 表示的感知损失函数作为其感知损失函数。Enhancenet^[87] 中对感知损失进行数学运算得到纹理匹配损失, 结合感知损失和 GAN 对抗损失作为网络优化目标。Super-FAN^[93] 和 FSRNet^[94] 的损失函数中分别加入了表示先验知识的热图损失项和先验损失项。FSR-GAN^[88] 网络中的感知损失函数对应低分辨率图像 I_{LR} 与重建高分辨率图像 I_{HR} 之间的感知损失。SRFeat^[89] 网络中先利用基于像素的损失来预训练生成器, 然后利用重建图像 I_{est} 与真实图像 I_{HR} 间的感知损失、特征判别器的对抗损失以及 GAN 对抗损失的线性组合作为整个网络的损失函数训练 GAN 网络。表 8 总结了各种基于 GAN 结构的 SR 方法的损失函数形式。

表 8 基于 GAN 的 SR 方法损失函数对比

| 算法 | 损失函数 | 算法目标 |
|-----------------------------|---|-----------------------------|
| SRGAN ^[81] | 感知损失和对抗损失加权组合 | 解决大上采样因子的 SR 问题 |
| Enhancement ^[87] | 感知损失+GAN 对抗损失+纹理匹配损失 | 加入纹理损失恢复真实的图像细节, 减少棋盘格效应 |
| ProGAN ^[91] | 感知损失+GAN 损失 | 处理 4 倍、8 倍高分辨率重建放大 |
| SFT-GAN ^[92] | 感知损失+对抗损失 | 加入类别先验知识, 恢复真实的高频信息 |
| Super-FAN ^[93] | 像素损失、感知损失、热图损失和 WGAN 损失加权组合 | 同时实现人脸图像的高分辨率重建和人脸特征点定位 |
| FSRNet ^[94] | 先验损失、GAN 损失和感知损失加权组合 | 加入人脸特征的先验知识, 提升人脸图像高分辨率重建质量 |
| FSR-GAN ^[88] | WGAN 生成损失+低分辨率图像与高分辨率图像间的感知损失函数 | 改进损失函数, 提升图像超分辨率网络性能 |
| SRFeat ^[89] | 第一阶段: 基于像素的损失 第二阶段: 重建图像与目标图像间的感知损失、GAN 对抗损失和特征对抗损失的线性组合 | 加入特征判别器, 提升重建图像质量 |
| SRMD ^[96] | GAN 损失和基于像素的损失线性加权 | 解决复杂退化模型 |
| Affgan ^[97] | GAN 损失 | 实现传统算法 |

4.3.4.2 基于 GAN 超分辨率重建方法的网络模型

(1). 实现大放大因子下图像超分辨率的 GAN 模型

大部分传统重建方法和基于深度学习的重建方法通常在网络前端或末端进行放大, 这种重建方式适合于较小的放大因子 (如放大因子为 2-3), 当放大因子提高到 4 倍, 8 倍时, 重建效果不尽如人意。人们提出了各种实现大放大因子下的图像 SR 的网络模型^[73,81,91], 其中, 基于 GAN 的代表性工作包括 SRGAN 和 ProGAN。

a). SRGAN(Generative Adversarial Network for SR)

Ledig^[81]等人提出了生成对抗网络 SRGAN 来恢复 4 倍上采样因子图像的高频细节。SRGAN 由 G (generator) 网和 D (Discriminator) 网组成, 网络结构如图 23 所示。G 网是一个残差网络, 通过低分辨率图像生成高分辨率图像, 由 D 网判断输入图像是由 G 网生成的, 还是数据库中的真实图像。当 G 网和 D 网达到博弈平衡的时候, 该网络即可用于 SR。

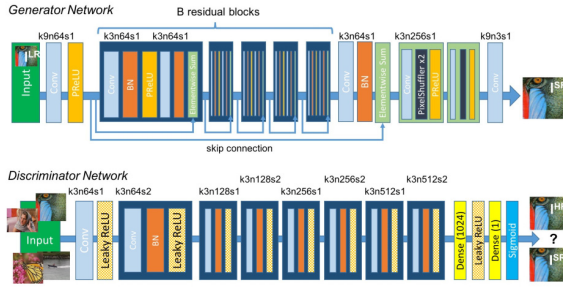


图 23. SRGAN 网络结构[81]

SRGAN 使用式(15)所示的感知损失函数 l^{SR} , 感知损失由内容损失 l_X^{SR} 和对抗损失 l_{Gen}^{SR} 两部分组成。内容损失部分与文献[86]中的特征重建损失完全一致。对抗损失关注纹理细节, 使得生成网络的输出更接近自然图像。

$$l^{SR} = l_X^{SR} + 10^{-3} l_{Gen}^{SR} \quad (15)$$

训练时, SRGAN 将真实高分辨率图像 I^{HR} 和 G 网络生成的高分辨率图像 $G_{\theta_G}(I^{LR})$ 输入到一个预训练好的 VGG19 网络中, 计算这两幅图像经过 VGG19 网络每一层输出特征的欧式距离之和作为内容损失 l_{VGG}^{SR} 。

$$l_{VGG}^{SR} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{j=1}^{H_{i,j}} \left[\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y} \right]^2 \quad (16)$$

其中, $W_{i,j}, H_{i,j}$ 分别表示 VGG 网络中特征图谱的宽度和高度, $\phi(I^{HR})$ 和 $\phi(G_{\theta_G}(I^{LR}))$ 分别表示 I^{HR} 和 $G_{\theta_G}(I^{LR})$ 对应的 VGG19 网络特征图。

优化对抗损失 l_{Gen}^{SR} 的作用是使判别器将生成器生成的图像 $G_{\theta_G}(I^{LR})$ 判定为真实图像的概率

$D_{\theta_D}(G_{\theta_G}(I^{LR}))$ 最小。

$$l_{Gen}^{SR} = \sum_{n=1}^N -\log D_{\theta_D}(G_{\theta_G}(I^{LR})) \quad (17)$$

在 Set5、Set14 和 BSD100 测试集上的实验结果表明 SRGAN 可以解决 4 倍上采样因子时的图像 SR 问题, 其重建图像的平均 PNSR 值较 SRCNN 方法降低 1db 左右, 较 DRCN 方法降低 1.3~2db 左右, 但图像的纹理细节更丰富逼真, 视觉效果更好。该方法的缺点在于网络结构复杂, 需要训练两个网络, 训练过程较长。

b). ProGAN(A Fully Progressive Approach to Single-Image Super-Resolution)

图像 SR 过程是图像恢复高频细节并放大的过程, 图像的放大主要有三类方法: i) 直接放大法, 包括直接对低分辨率图像上采样[62]或在网络最后一层使用反卷积[64]或者亚像素卷积进行放大

[63], 前者在高分辨率空间对图像进行重建, 计算量大, 后者是对前者的改进, 这两种方法通常仅适用于 2-3 倍的图像放大; ii) 迭代上采样下采样放大法。文献[90]基于 IBP 算法[43]的思想提出了一种图像 SR 网络—DBPN (Deep Back-Projection Networks)。该网络采用深度串联方式直接利用不同分辨率特征, 不断迭代地进行上采样和下采样操作作为每个阶段的映射误差提供误差反馈机制, 在上下采样阶段使用密集连接方式增加特征复用。即使只用浅层网络, DBPN 也能实现 8 倍放大的图像 SR; iii) 逐级放大法, 如 LapSRN 每次实现 2 倍采样放大, 直至实现 4 倍、8 倍高放大倍数的图像重建。文献[91]将这种逐级重建的思想和 GAN 联合起来, 提出了一种由渐进生成网络 (progressive generator) 和渐进判别网络 (progressive discriminator) 组成的图像多尺度逐级高分辨率重建网络—ProGAN。实验结果表明, 该网络能有效实现图像 4 倍、8 倍的 SR (现有 SRGAN 最多只能实现 4 倍放大), 同时 ProGAN 重建图像的 PNSR 值比同类逐级重建网络 LapSRN 平均高出 0.05~0.13db。

(2). 加入先验知识的 GAN

图像 SR 问题中可以利用图像的自相似性[65], 稀疏性[66]等领域知识和图像的边缘梯度信息[68]、类别信息[92]等先验知识指导网络设计, 使得重建图像具有丰富的高频细节和更好的视觉效果。基于领域先验知识的 GAN 结构 SR 方法包括:

a). SFT-GAN(GAN with Spatial Feature Transform for SR)

SRGAN 通过感知损失获得包含丰富细节的高分辨率重建图像, 但因为没有加入先验知识, 这些细节可能并不自然, 不真实。文献[92]提出了一种考虑类别先验的高分辨率重建网络, 其网络结构如图 24 所示。该网络也是一种基于残差学习的生成对抗网络, 生成器由条件网络和高分辨率重建网络组成。条件网络将表达语义先验类别的语义分割概率图作为输入, 实时生成空间特征变换层 (Spatial Feature Transform Layer, SFT layer) 共享的语义先验知识, 重建网络由 16 个残差块和空间特征变换层组成, 利用类别先验知识恢复出与所属语义类别特征一致的纹理。判别器不但要判断生成图像是否真实, 还需要预测输入的类别信息。SFT-GAN 也使用感知损失函数作为优化目标, 通过实验, 作者验证了加入类别先验信息可以恢复出更真实的图像纹理细节。

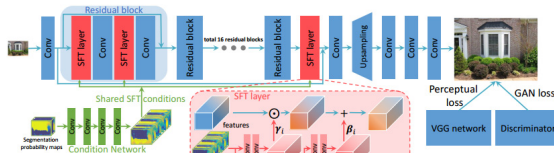


图 24. SFT-GAN 结构图[92]

b). Super-FAN 及 FSRNet (Face Super-Resolution Network with Facial Priors)

使用图像 SR 技术提升人脸图像分辨率是 SR 在生物信息处理领域中的具体应用之一。应用 GAN 结构，结合人脸先验信息和图像 SR 技术是当前人脸 SR 研究的热点之一。

文献[93]提出了首个同时进行人脸 SR 和特征点定位的端到端网络—Super-FAN。该算法的基本思想是通过热图回归(heatmap regression)将一个人脸对齐的子网络 FAN 集成到基于 GAN 的图像 SR 网络中，并优化新的热图损失(heatmap loss)，同时实现人脸 SR 和人脸特征点定位。Super-FAN 的网络结构如图 25 所示，它由生成器、判别器和人脸对齐子网络 FAN 组成，如式(18)所示，Super-FAN 的损失函数 l^{SR} 由基于像素的损失 l_{pixel} 、感知损失 $l_{feature}$ 、生成对抗损失 l_{WGAN} 和热图损失 $l_{heatmap}$ 线性组合而成。加入热图损失可以突出低分辨率人脸图像和高分辨率人脸图像的结构相似性，同时实现人脸特征定位。Super-FAN 能对除了某些极端情况下的任意姿态人脸图像进行对齐和高分辨率重建，且重建图像的质量优于 SRGAN 方法。

$$l^{SR} = \alpha l_{pixel} + \beta l_{feature} + \gamma l_{heatmap} + \delta l_{WGAN} \quad (18)$$

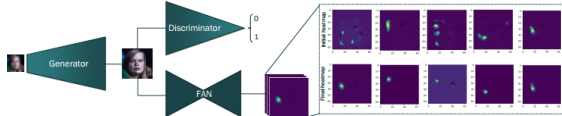


图 25. Super-FAN 网络结构图[93]

文献[94]将人脸先验知识加入网络，设计了人脸 SR 网络—FSRNet，该网络由生成器和判别器两部分组成，生成器网络如图 26 所示。FSRNet 中粗粒度网络恢复一个粗精度的高分辨率图像，然后将该图像分别送入细粒度网络的两个分支，第一个分支用于提取图像特征的细粒度图像超分辨率编码器，第二个分支是从粗粒度重建图像中估计人脸特征点和分割信息先验知识的先验估计网络，最后两个分支结果汇总到一个细粒度超分辨率解码网络中，重构出目标高分辨率图像。FSRNet 判别器网络结构和功能与 SRGAN 的判别器一致，但前者的损失函数 L_{FSRNet} （式(19)所示）在感知损失 l_p 和 GAN 损失 l_{GAN} 线性组合（权重系数分别为 γ_p 和 γ_{GAN} ）的基础上加入了先验损失

l_F 。人脸特征的先验知识能明显提升人脸图像 SR 质量。相比人脸特征点先验，分割先验知识包含更多信息，更能显著影响网络性能，而且全局分割图往往比局部分割图更有用。

$$L_{FSRNet} = l_F + \gamma_{GAN} l_{GAN} + \gamma_p l_p \quad (19)$$

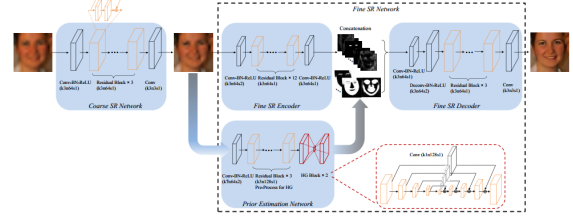


图 26. FSRNet 中生成器网络结构图[94]

上述两种人脸图像 SR 网络是目前人脸图像重建效果最好的两种方法，但得到的重建人脸图像的 PNSR 和 SSIM 值也比重建的自然图像低。这主要是由于在低分辨率情况下难以估计出人脸特征点，导致人脸对齐效果不够准确。相对于其它自然图像，人脸图像中各器官的相对位置关系和整体结构包含了丰富的先验信息，如何更有效地挖掘利用这些先验知识以重建出高质量的高分辨率人脸图是未来研究方向之一。

通过比较加入先验知识的各种 GAN 算法，我们可以看出图像先验知识通过网络，损失函数的形式加入到图像超分辨率重建网络中，各网络的损失函数也在 SRGAN 感知损失函数的基础上，综合了考虑基于像素、基于高层次特征和基于先验的各类损失，从而更好地提升了重建图像的质量。

(3). 学习复杂退化模型的 GAN

一般用于图像 SR 算法的低分辨率训练集，是通过高分辨率图像进行下采样得到的。由于忽略了传感器噪声、图像压缩和运动模糊等降质因素，这个过程可视为理想的退化模型。而自然图像的退化过程往往更复杂且未知，为了能适应复杂退化模型的 SR 问题，Shocher 等人[67]提出了一种不需要预训练，利用图像自相似性进行复杂退化情况下的高分辨率图像重建的轻量级网络。此外，文献[95]提出了一种维度拉伸策略，将退化过程的两个关键因素（即模糊核和噪声水平）作为网络输入，使得单个卷积超分辨率网络可以处理多个甚至是不均匀退化类型的超分辨网络模型。文献[96]使用一个基于 GAN 结构的 SR 方法解决该问题。该网络由一个学习高分辨率图像退化为低分辨率图像的 GAN (High-to-Low GAN) 和一个学习从低分辨率图像重建高分辨率图像的 GAN (Low-to-High GAN) 组成。首先利用不配对的高低分辨率图像训练 High-to-Low GAN，学习图像的退化过程。训练完成后，将低分辨率图像和与之

对应的高分辨率图像作为 Low-to-High GAN 网络的输入完成图像高分辨率重建。两个网络的损失函数均由 GAN 损失和基于像素的损失经过线性加权构成。虽然作者仅对人脸图像进行了实验,但该方法对经过复杂退化过程得到的自然图像同样有很好的重建效果,增加了 SR 技术应用于实际问题的可能性。

(4). 与传统超分辨率重建方法的结合的 GAN

传统 MAP 算法必须利用图像先验知识对图像分布建模,但图像分布往往不可预知,导致建模过程困难且建立的模型不准确使得重建图像质量不佳。文献[97]提出了一种将 GAN 模型与传统基于空域的 MAP 算法相结合的 SR 方法—AffGAN (affine projected GAN),将极大似然估计问题转化为求解重构图像的分布 q_θ 与真实 HR 图像分布 p_Y 的 KL 散度 $-KL[q_\theta||p_Y]$ 的极小值问题,并最小化 GAN 损失函数(最小化 $-KL[q_\theta||p_Y]$ 等同于最小化 GAN 损失函数)。训练过程中,通过对样本加入实例噪声增加 GAN 模型训练的稳定性。

4.4 基于其它深度学习模型的超分辨率重建方法

卷积神经网络被广泛应用于图像超分辨率重建问题中,与此同时,也有学者提出将受限玻尔兹曼机(Restricted Boltzmann Machine, RBM),深度信念网络(Deep Belief Network, DBN),自编码器(Auto Encoder, AE)等深度学习模型用于图像超分辨率重建问题。

文献[98]提出一种使用 RBM 训练稀疏字典,进行图像 SR 的方法。其中 RBM 的可见层由高低分辨率图像块构成,隐藏层代表高低分辨率图像字典对共享的稀疏系数,通过极大似然估计得到隐藏层与可见层间的连接参数,即字典元素。为加快训练速度,作者引入了对比散度算法,并通过增加 RBM 的隐藏层层数构成深度玻尔兹曼机(Deep Boltzmann Machine, DBM)。该方法的本质仍是稀疏表示^[58],只是将字典的构造和训练过程通过 RBM 完成,是传统方法和深度学习方法的结合。

DBN 由多层 RBM 组成。Zhou 等人^[99]利用 DBN 的层次结构学习低分辨率图像到高分辨率图像之间的非线性映射关系,通过训练和重建两个步骤完成图像高分辨率重建。在训练过程中,将 HR 图像分解成多个图像块,变换到 DCT 域,得到的二维 DCT 系数用于训练 DBN。重建时,将 LR 图像插值到目标尺寸后分解为图像块,变换到 DCT 域,利用训练得到的 DBN 来恢复丢失的高频信息。最后通过 DCT 逆变换获得重建图像。

文献 [100]提出了一种基于堆协同局部自编

码实现图像逐级放大的网络—DNC (Deep Network Cascade)。该网络由基本模块 SCLA (Stacked CLA) 级联而成,每个基本模块都由协同局部自编码器 CLA (collaborative local auto-encoder) 和非局部自相似性单元 NLSS (non-local self-similarity) 组成。该方法在多尺度图像块中进行非局部自相似搜索,增强输入图像块的高频纹理细节,抑制噪声,整合重叠图像块的兼容性。文献 [101]提出了由对称的卷积层—反卷积层构成的极深层残差自编码器网络 RED (very deep Residual Encoder-Decoder Networks)。RED 中每个卷积层和反卷积层间使用跳线连接,解决梯度消失问题。卷积层用来获取图像的抽象内容,反卷积层用来放大特征尺寸并且恢复图像细节。

5 讨论与分析

从首个基于深度学习的超分辨率重建网络—SRCNN 诞生以来,各种基于深度学习的超分辨率重建方法被不断提出,表 10 从网络输入、是否支持多尺度重建、是否包含残差学习、是否加入先验知识、网络训练策略、重建方式以及损失函数形式等方面综合比较了上面介绍的各种基于深度学习的代表性超分辨率重建方法。表中各种方法颜色与图 4 中颜色对应。表中符号说明如下:

① ‘BI(bicubic)’表示双三次插值,‘N’表示否,‘Y’表示是;

② Momentum, SGD, Adam, RMSprop, BFGS 分别代表动量梯度下降法、随机梯度下降法、自适应矩估计梯度下降法、基于加权平方和的自适应学习率梯度下降法和基于 BFGS 的拟牛顿法;

③ L2 表示 L2 范数损失函数;L1 表示 L1 范数损失函数;charbonnier 损失函数形式见式 (10)。

GAN 损失参见表 8。

通过表 10 对各种方法的比较,我们可以发现:

1). 在网络选择上,基于 CNN 的 SR 方法仍然是目前深度学习在 SR 应用中的主流方法,也有部分学者尝试使用其它深度学习模型,如受限玻尔兹曼机,深度置信网络,自编码器网络等进行图像超分辨率重建。

2). 从网络设计和训练策略来看,基于 CNN 的 SR 方法从 SRCNN^[62]开始,到基于残差结构的 VDSR 网络和基于密集连接的 SRDenseNet^[78]网络,再到基于 GAN 的 SR 方法,网络结构朝着更深、更宽、连接更复杂的趋势发展。深度网络能够充分利用图像上下文信息,扩大卷积感受野,深度挖掘图像信息,提取图像高层特征;复杂的连接形式能加强层间的信息传递和特征传播,充分利用图像的层次特征,但同时会带来训练慢,难收敛等问题。针对上述问题,通常可使用跳连接、密集连

表 10 基于深度学习的超分辨率重建方法

| 算法 | 输入 | 多尺度 | 残差学习 | 训练策略 | 先验知识 | 重建方式 | 损失函数 |
|-----------------------------|--------------------|----------------|------|------------------------|--------------|------|--------------------------|
| SRCNN ^[62] | LR+BI ^① | N ^① | N | Momentum ^② | -- | 直接 | L2 ^③ |
| ESPCNN ^[63] | LR | N | N | Momentum | -- | 直接 | L2 |
| FSRCNN ^[64] | LR | Y ^① | N | SGD ^② | -- | 直接 | L2 |
| SRCNN-Pr ^[65] | LR | Y | N | SGD | 梯度先验 自相似性 | 直接 | L2 |
| SCN ^[66] | LR+BI | Y | N | SGD | 稀疏先验 | 逐级 | L2 |
| DEGREE ^[68] | LR | N | Y | SGD 递归监督 | 边缘先验 | 直接 | L2 |
| ZSSR ^[67] | LR | N | N | Momentum | 自相似性 | 直接 | L1 ^③ |
| VDSR ^[69] | LR+BI | Y | Y | SGD 梯度裁剪 | -- | 直接 | L2 |
| DRCN ^[70] | LR+BI | N | N | SGD 递归监督 | -- | 直接 | L2 |
| DRRN ^[71] | LR+BI | N | Y | SGD 梯度裁剪 | -- | 直接 | L2 |
| EDSR ^[72] | LR | N | Y | Adam ^② | -- | 直接 | L1 |
| LapSRN ^[73] | LR | Y | Y | Momentum | -- | 逐级 | charbonnier ^③ |
| DSRN ^[75] | LR+BI | N | Y | SGD | -- | 直接 | L2 |
| MSRN ^[76] | LR+BI | Y | Y | Adam | -- | 直接 | L1 |
| SRDenseNet ^[78] | LR+BI | N | N | Adam | -- | 直接 | L1 |
| MemNet ^[79] | LR+BI | N | Y | SGD | -- | 直接 | L2 |
| RDN ^[80] | LR | N | Y | Adam | -- | 直接 | L2 |
| SRGAN ^[81] | LR | N | Y | Adam | -- | 直接 | L2 |
| ProGAN ^[91] | LR | Y | Y | Curriculum Learning | -- | 逐级 | L2 |
| Enhancenet ^[87] | LR | N | Y | Adam | -- | 直接 | L2 |
| SFT-GAN ^[92] | LR | N | Y | Adam | 类别先验 | 直接 | L2 |
| Super-FAN ^[93] | LR | N | Y | RMSprop ^② | 人脸特征点先验 | 直接 | L2 |
| FSRNet ^[94] | LR+BI | N | Y | RMSprop | 人脸特征及分割先验 | 直接 | L2 |
| SRFeat ^[89] | LR | N | Y | Adam | -- | 直接 | L2 |
| SRMD ^[96] | LR | N | Y | Adam | -- | 直接 | L2 |
| Affgan ^[97] | LR | N | Y | SGD | -- | 直接 | GAN 损失 ^③ |
| Gao et al. ^[98] | LR+BI | N | N | 对比散度 | 稀疏先验 | 直接 | -- |
| Zhai et al. ^[99] | LR+BI | N | N | SGD | -- | 直接 | -- |
| RED ^[100] | LR | N | Y | Adam | -- | 直接 | L2 |
| DNC ^[101] | LR | Y | N | BFGS | -- | 逐级 | L2 |

接、残差学习、梯度裁剪和参数共享等策略减轻网络负担，加快学习速度，缓解梯度消失和爆炸。

3). 目前多数网络仅考虑简单图像退化模型，只有少数网络模，如 ZSSR^[67]，SFT-GAN^[92]，SRMD^[96]，考虑了存在光学模糊，运动模糊，传感器噪声等多种复杂自然退化因素下的图像超分辨率重建问题。

4). 在网络输入层，一般方法将低分辨率图像或对其上采样得到的目标尺寸大小的低分辨率图像作为网络输入。将后者作为输入会导致整个网络的运算在高分辨率空间进行，不仅增加了网络运算复杂度，也会使重建图像边缘模糊，丢失细节特征。

5). 一些网络能够得到多尺度、不同放大倍数下高分辨率重建图像，满足不同的放大需求。如 FSRCNN^[64]和 MSRN^[76]分别调整反卷积层和 pixel shuffle 层对重建图像的不同尺度放大；SRCNN-Pr^[65]和 VDSR^[69]采用并联结构分别训练不同放大因子对应的子网络，得到不同放大因子下重建图像；将多尺度 LR 图像进行作为输入，进行图像不同倍数的放大操作；SCN^[66]、LapSRN^[73]、

ProGAN^[91]和 DNC^[101]则通过级联结构采用逐级放大形式得到不同放大倍数的重建图像。

6). 领域先验知识可以指导设计基于深度学习超分辨率重建。如 SRCNN-Pr^[65]，LapSRN^[73]，SFT-GAN^[92]，DNC^[101]等方法利用图像不同尺度间的自相似性来恢复丢失的高频细节，实现图像的逐级高分辨率重建。而 SCN^[66]，SFT-GAN^[92]，Super-FAN^[93]和 FSRNet^[94]等方法通过在子网络构造中分别加入稀疏先验，类别先验及人脸特征先验知识，用来指导超分辨率重建。SRCNN-Pr^[65]则利用图像梯度先验知识，设计边缘损失和纹理损失函数，指导网络训练。

7). 融合领域先验知识和深度学习框架的超分辨重建方法比单纯使用基于数据驱动的深度学习超分辨率方法能获得更符合人眼视觉特性，更合理重建高分辨率图像。

8). 一般的深度学习方法仅能处理适用于图像 2-3 倍的重建，大上采样因子（如 4 倍，8 倍）的图像超分辨率重建主要通过逐级放大策略(如 LapSRN

[73] 和 ProGAN^[91]。

9). 损失函数的设计。大部分网络的损失函数是基于像素差异的 L1 型、L2 型损失函数。L1 型损失函数收敛性比 L2 型好, 针对 L2 型损失函数带来的振铃现象, LapSRN^[73]提出使用 charbonnier 损失函数代替 L2 型损失函数, 获得了更好的图像细节。基于像素差异的损失函数并不能反映重建图像和

真实图像间的主观感知差异, 基于 GAN 的 SR 方法中普遍结合感知损失函数, 生成高质量的重建图像并获得视觉效果的提升。

表 11 比较了放大因子分别为 2,3,4 的情况下, 一些典型图像超分辨率重建方法在 Set5, Set14, BSD100, Urban100 测试集中的实验结果, 实验结果均取自相关文献。

表 11 各种典型图像超分辨率重建方法实验结果 (PNSR/SSIM) 的比较

| | 数据集 | 评价指标 | Set5 | | | Set14 | | | BSD100 | | | Urban100 | | |
|--------|-------------------------|------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | 算法 | | ×2 | ×3 | ×4 | ×2 | ×3 | ×4 | ×2 | ×3 | ×4 | ×2 | ×3 | ×4 |
| 传统方法 | Bicubic ^[62] | PNSR | 33.66 | 30.39 | 28.42 | 30.24 | 27.55 | 26.00 | 29.56 | 27.21 | 25.96 | 26.88 | 24.46 | 23.14 |
| | | SSIM | 0.9299 | 0.8682 | 0.8104 | 0.8688 | 0.7742 | 0.7027 | 0.8431 | 0.7385 | 0.6675 | 0.8403 | 0.7349 | 0.6577 |
| | SelfEx ^[59] | PNSR | 36.49 | 32.58 | 30.31 | 32.22 | 29.16 | 27.40 | 31.18 | 28.29 | 26.84 | 29.54 | 26.44 | 24.79 |
| | | SSIM | 0.9537 | 0.9093 | 0.861 | 0.9034 | 0.8196 | 0.7518 | 0.8855 | 0.7840 | 0.7106 | 0.8967 | 0.8088 | 0.737 |
| | SC ^[58] | PNSR | -- | 31.42 | -- | -- | 28.31 | -- | -- | -- | -- | -- | -- | -- |
| | | SSIM | -- | 0.8821 | -- | -- | 0.7954 | -- | -- | -- | -- | -- | -- | -- |
| 深度学习方法 | NE+LLE ^[56] | PNSR | 35.77 | 31.84 | 29.61 | 31.76 | 28.60 | 26.81 | -- | -- | -- | -- | -- | -- |
| | | SSIM | 0.9490 | 0.8956 | 0.8402 | 0.8993 | 0.8076 | 0.7331 | -- | -- | -- | -- | -- | -- |
| | SRCNN ^[62] | PNSR | 36.66 | 32.75 | 30.48 | 32.42 | 29.28 | 27.49 | 31.36 | 28.41 | 26.90 | 29.50 | 26.24 | 24.52 |
| | | SSIM | 0.9542 | 0.9090 | 0.8628 | 0.9063 | 0.8209 | 0.7503 | 0.8879 | 0.7863 | 0.7101 | 0.8946 | 0.7989 | 0.7221 |
| | VDSR ^[69] | PNSR | 37.53 | 33.82 | 31.53 | 33.04 | 29.76 | 28.02 | 30.76/ | 27.14 | 25.18 | 30.76 | 27.14 | 25.18 |
| | | SSIM | 0.9587 | 0.9226 | 0.8854 | 0.9118 | 0.8311 | 0.7670 | 0.9140 | 0.8279 | 0.7524 | 0.9140 | 0.8279 | 0.7524 |
| | DRCN ^[70] | PNSR | 37.63 | 33.82 | 31.53 | 33.04 | 29.76/ | 28.02 | 31.85 | 28.80 | 27.23 | 30.76 | 27.14 | 25.18 |
| | | SSIM | 0.9588 | 0.9226 | 0.8854 | 0.9118 | 0.8311 | 0.7670 | 0.8942 | 0.7963 | 0.7233 | 0.9140 | 0.8279 | 0.7524 |
| | LapSRN ^[73] | PNSR | 37.52 | 33.82 | 31.54 | 33.08 | 29.79 | 28.19 | 31.80 | 28.82 | 27.32 | 30.41 | 27.07 | 25.21 |
| | | SSIM | 0.9591 | 0.9227 | 0.8855 | 0.9130 | 0.8320 | 0.7720 | 0.8950 | 0.7973 | 0.7280 | 0.9101 | 0.8272 | 0.7553 |
| | SRGAN ^[81] | PNSR | -- | -- | 32.05 | -- | -- | 28.53 | -- | -- | 27.57 | -- | -- | 26.07 |
| | | SSIM | -- | -- | 0.8910 | -- | -- | 0.7804 | -- | -- | 0.7354 | -- | -- | 0.7839 |
| | RED ^[100] | PNSR | 37.669 | 33.82 | 31.51 | 32.94 | 29.61 | 27.86 | 31.99 | 28.93 | 27.40 | -- | -- | -- |
| | | SSIM | 0.9599 | 0.9230 | 0.8869 | 0.9144 | 0.8341 | 0.7718 | 0.8974 | 0.7994 | 0.7290 | -- | -- | -- |
| | MDSR ^[72] | PNSR | 38.11 | 38.11 | 32.50 | 33.85 | 30.44 | 28.72 | 32.29 | 29.25 | 27.72 | 32.84 | 28.79 | 26.67 |
| | | SSIM | 0.9602 | 0.9602 | 0.8973 | 0.9198 | 0.8452 | 0.7857 | 0.9007 | 0.8091 | 0.7418 | 0.9347 | 0.8655 | 0.8041 |
| | RDN ^[80] | PNSR | 38.24 | 34.71 | 32.47 | 34.01 | 30.57 | 28.81 | 32.34 | 29.26 | 27.72 | 32.89 | 28.80 | 26.61 |
| | | SSIM | 0.9614 | 0.9296 | 0.8990 | 0.9212 | 0.8468 | 0.7871 | 0.9017 | 0.8093 | 0.7419 | 0.9353 | 0.8653 | 0.8028 |

其中, SC 和 NE+LLE 分别是基于稀疏编码和基于流行学习的传统图像超分辨率重建算法; SRCNN 是首个将深度学习应用于图像超分辨率重建中的方法; VDSR 首次将残差学习思想引入 SR 问题, 通过加深网络, 扩大图像感受野 (相比 SRCNN, 图像重建质量得到了提升)。DRCN 从训练策略上减少了网络参数, 提升了网络训练速度; LapSRN 提出了逐级放大重建高分辨率图像的思想, 在 4 倍上采样因子时, 重建性能较 SRCNN 和 VDSR 方法都好, 为实现大上采样因子的图像 SR 做了有益尝试。SRGAN 是第一个利用 GAN 结构结合感知损失函数解决 SR 问题的网络, 为后续基于 GAN 结构的 SR 方法提供了思路, RED 是一个结构简单、性能优良的自编码器网络。MDSR 利用图像自相似的领域先验知识实现图像多尺度 SR。

从实验结果来看, 基于深度学习的方法较基于流行学习和稀疏编码等传统方法在超分辨率图像重建质量上有较大提升。就基于深度学习的超分辨率方法而言, 图像自身的领域先验知识对提高重建图像质量具有十分重要的指导作用。RDN 是将密集连接、跳连接、局部特征融合和全局特征融合整合

起来, 形成了一个复杂结构的网络, 它充分利用图像各层次特征, 在多数情况下取得了最好的重建效果, 这说明网络越深, 网络内部信息流动性越好, 越有利于提升图像超分辨率重建质量。

6 总结与展望

图像超分辨率重建问题经历了基于插值、基于重构和基于学习的三个发展阶段。目前, 以基于深度学习为代表的基于学习的方法逐渐成为主流方法。基于深度学习的超分辨率方法在模型选择、网络结构设计、领域先验知识引入和损失函数优化等方面不断创新。尽管图像超分辨率重建研究取得了较大进展, 但仍存在以下问题需要解决。

1). 网络选择和设计。目前大部分基于深度学习的超分辨率重建方法采用 CNN 模型, 如何将其它深度学习模型应用于图像超分辨率重建值得探索。此外, 一般来说, 深度网络越深, 重建效果越好, 但也会导致训练困难, 不易收敛等问题。如何权衡网络的复杂性和重建图像质量的矛盾是需要进一步研究的问题。

2). 如何更有效地将特定应用场合中的领域知识、先验知识和深度学习框架结合起来, 提升现有基于深度学习的图像超分辨率方法的重建性能是值得探索的方向。领域和先验知识是人类对于相关领域经过长期研究、沉淀和积累得到的经过验证的正确知识和经验, 对该领域的后续研究具有非常重要的指导意义。而深度学习能充分利用大数据训练模型, 自动提取和选择超分辨率问题中有用的特征, 学习 LR 到 HR 的非线性映射。若能充分利用领域和先验知识指导深度学习模型, 不仅可以提取富含上下文信息的高层语义特征, 进而学习更复杂的 LR 到 HR 映射关系, 也能保证这种映射关系的合理性和鲁棒性, 从而进一步提高基于深度学习的超分辨率模型的重建性能和可解释性。

3). 针对复杂图像退化模型的超分辨率重建问题研究。研究过程中, 通常采用对高分辨率图像下采样得到低分辨率图像的方式模拟自然情况下图像退化过程, 而实际应用中的低分辨率图像通常存在运动变形、光学模糊、降采样以及噪声污染等多种复杂降质因素。如何学习得到一个可以处理复杂图像退化模型的网络是未来可研究的方向之一。

4). 大放大倍数的图像高分辨率重建。目前, 针对 2-3 倍放大因子的图像高分辨率重建取得不错的效果, 当放大因子为 4 或 8 时, 重建效果急剧下降, 现有方法还无法有效处理更大放大因子下的图像超分辨率任务。因此, 设计针对大放大倍数的图像超分辨率重建算法是亟待研究的方向。

参考文献

- [1] Yang S, Sun Y, Chen Y, et al. Structural similarity regularized and sparse coding based super-resolution for medical images[J]. *Biomedical Signal Processing & Control*, 2012, 7(6):579-590.
- [2] Zhang M, Desrosiers C, Qu Q, et al. Medical image super-resolution with non-local embedding sparse representation and improved IBP[C]// *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2016:888-892.
- [3] Wang Y H, Li J B, Fu P. Medical Image Super-resolution Analysis with Sparse Representation[C]// *Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. IEEE Computer Society, 2012:106-109.
- [4] Jia Z, Wang H, Xiong Z, et al. Fast face hallucination with sparse representation for video surveillance[C]// *Pattern Recognition (ACPR)*, 2011 First Asian Conference on. IEEE, 2011: 179-183.
- [5] Liu H C, Li S T, Yin H T. Infrared surveillance image super resolution via group sparse representation[J]. *Optics Communications*, 2013, 289(4):45-52.
- [6] Uiboupin T, Rasti P, Anbarjafari G, et al. Facial image super resolution using sparse representation for improving face recognition in surveillance monitoring[C]// *Signal Processing and Communication Application Conference*. IEEE, 2016:437-440.
- [7] Hu M G, Wang J F, Ge Y. Super-resolution reconstruction of remote sensing images using multifractal analysis[J]. *Sensors*, 2009, 9(11):8669-8683.
- [8] 杨学峰. 遥感图像频域和空域超分辨重建技术研究[D]. 哈尔滨工业大学, 2011.
- [9] Hong Z H U, Weidong S, Hai T A N, et al. Remote Sensing Images Super Resolution Reconstruction Based on Multi-scale Detail Enhancement[J]. *Acta Geodaetica et Cartographica Sinica*, 2016, 45(9): 1081-1088.
- [10] Liao R, Tao X, Li R, et al. Video Super-Resolution via Deep Draft-Ensemble Learning[C]// *IEEE International Conference on Computer Vision*. IEEE, 2015:531-539.
- [11] Cheng M H, Lin N W, Hwang K S, et al. Fast video super-resolution using artificial neural networks[C]// *International Symposium on Communication Systems, Networks & Digital Signal Processing*. IEEE, 2012:1-4.
- [12] Chen J, Nunez-Yanez J, Achim A. Video Super-Resolution Using Generalized Gaussian Markov Random Fields[J]. *IEEE Signal Processing Letters*, 2011, 19(2):63-66.
- [13] 卓力, 王素玉, 李晓光. 图像/视频的超分辨率复原[J]. 人民邮电出版社, 2011: 102-105.
- [14] Tsai R Y. Multiframe image restoration and registration[J]. *Advance Computer Visual and Image Processing*, 1984, 1: 317-339.
- [15] Harris J L. Diffraction and resolving power[J]. *Journal of the Optical Society of America*, 1964, 54(7): 931-936.
- [16] Goodman J W, Cox M E. Introduction to Fourier Optics[M]. McGraw-Hill, 1968.
- [17] Tsai R Y, Huang T S. Multiframe image restoration and registration. *Advances in Computer Vision and Image Processing*, 1984, 1: 317-339.
- [18] Borman S, Stevenson R. Spatial Resolution Enhancement of Low-resolution Image Sequences: A Comprehensive Review with Directions for Future Research, Technical Report, Laboratory Image and Signal Analysis, University of Notre Dame, 1998.
- [19] Park S C, Min K P, Kang M G. Super-resolution image reconstruction: a technical overview[J]. *Signal Processing Magazine IEEE*, 2003, 20(3):21-36.
- [20] 苏衡, 周杰, 张志浩. 超分辨率图像重建方法综述[J]. *自动化学报*, 2013, 39(8):1202-1213.
- [21] Huang D, Liu H. A short survey of image super resolution algorithms[J]. *Journal of Computer Science Technology Updates*, 2015, 2(2): 19-29.
- [22] Hayat K. Super-resolution via deep learning[J]. *arXiv preprint arXiv:1706.09077*, 2017.
- [23] Sun X, Xiao-Guang L I, Jia-Feng L I, et al. Review on Deep Learning Based Image Super-resolution Restoration Algorithms[J]. *Acta Automatica Sinica*, 2017, 43(5):697-709.
- [24] Yang W, Zhang X, Tian Y, et al. Deep learning for single image super-resolution: A brief review[J]. *arXiv preprint arXiv:1808.03344*, 2018.
- [25] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]// *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. IEEE, 2001, 2: 416-423.
- [26] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding[J]. 2012.
- [27] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations[C] // *International conference on curves and surfaces*. Springer, Berlin, Heidelberg, 2010: 711-730.
- [28] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015: 5197-5206.
- [29] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. *IEEE transactions on image processing*, 2004, 13(4): 600-612.
- [30] Sigitani T, Iiguni Y, Maeda H. Image interpolation for progressive transmission by using radial basis function networks[J]. *IEEE Transactions on Neural Networks*, 2002, 10(2):381-390.
- [31] Schultz R R, Stevenson R L. A Bayesian approach to image expansion for improved definition[J]. *Image Processing IEEE Transactions on*, 1994, 3(3):233-242.
- [32] Hou H, Andrews H. Cubic splines for image interpolation and digital filtering[J]. *IEEE Transactions on Acoustics Speech & Signal Processing*, 1978, 26(6):508-517.
- [33] Li X, Orchard M T. New edge-directed interpolation[J]. *IEEE Trans Image Process*, 2001, 10(10):1521-1527.
- [34] 蔡念, 张海员, 张楠, 等. 基于小波的双线性插值误差补偿算法的图像放大[J]. *激光与红外*, 2010, 40(5):558-562.
- [35] Zhang L, Wu X. An edge-guided image interpolation algorithm via directional filtering and data fusion[J]. *IEEE Trans Image Process*, 2006, 15(8):2226-2238.
- [36] Shi H, Ward R. Canny edge based image expansion[C]// *Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on*. IEEE, 2002, 1: 1-1.

- [37] Kim S P, Bose N K, Valenzuela H M. Recursive reconstruction of high resolution image from noisy undersampled multiframes[J]. *Acoustics Speech & Signal Processing IEEE Transactions on*, 1990, 38(6):1013-1027.
- [38] Rhee S, Kang M G. Discrete cosine transform based regularized high-resolution image reconstruction algorithm[J]. *Optical Engineering*, 1999, 38(8):1348-1356.
- [39] Nguyen N, Milanfar P. An efficient wavelet-based algorithm for image superresolution[C]//Image Processing, 2000. Proceedings. 2000 International Conference on. IEEE, 2000, 2: 351-354.
- [40] Ji H, Fermüller C. Robust wavelet-based super-resolution reconstruction: theory and algorithm[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(4): 649-660.
- [41] Panagiotopoulou A, Anastassopoulos V. Super-resolution image reconstruction employing Kriging interpolation technique[C]//Systems, Signals and Image Processing, 2007 and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services. 14th International Workshop on. IEEE, 2007: 144-147.
- [42] Stark H, Oskoui P. High-resolution image recovery from image-plane arrays, using convex projections[J]. *Journal of the Optical Society of America A Optics & Image Science*, 1989, 6(11):1715.
- [43] Irani M, Peleg S. Super resolution from image sequences[C]//Pattern Recognition, 1990. Proceedings., 10th International Conference on. IEEE, 1990, 2: 115-120.
- [44] Schultz R R, Stevenson R L. A Bayesian approach to image expansion for improved definition[J]. *Image Processing IEEE Transactions on*, 1994, 3(3):233-242.
- [45] Schultz R R, Stevenson R L. Improved definition video frame enhancement[C]//Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on. IEEE, 1995, 4: 2169-2172.
- [46] Chantas G K, Galatsanos N P, Woods N A. Super-resolution based on fast registration and maximum a posteriori reconstruction[J]. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 2007, 16(7):1821.
- [47] Hennings-Yeomans P H, Baker S, Kumar B V K V. Simultaneous super-resolution and feature extraction for recognition of low-resolution faces[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008:1 - 8.
- [48] Babacan S D, Molina R, Katsaggelos A K. Total variation super resolution using a variational approach[C]// IEEE International Conference on Image Processing. IEEE, 2008:641-644.
- [49] Farsiu S, Robinson M D, Elad M, et al. Fast and robust multiframe super resolution[J]. *Image Processing IEEE Transactions on*, 2004, 13(10):1327-1344.
- [50] Bose N K, Lertrattanapanich S, Koo J. Advances in superresolution using L-curve[C]//IEEE International Symposium on Circuits and Systems. IEEE, 1999, 2001 (2): 433-436.
- [51] Yuan Q, Zhang L, Shen H, et al. Adaptive Multiple-Frame Image Super-Resolution Based on U-Curve[J]. *IEEE Transactions on Image Processing*, 2010, 19(12):3157-3170.
- [52] Elad M, Feuer A. Restoration of a single super-resolution image from several blurred, noisy, and undersampled measured images[J]. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 1997, 6(12):1646-58.
- [53] Freeman W T, Jones T R, Pasztor E C. Example-based super-resolution[J]. *Computer Graphics & Applications IEEE*, 2002, 22(2):56-65.
- [54] Glasner D, Bagon S, Irani M. Super-resolution from a single image[C]// IEEE, International Conference on Computer Vision. 2009:349-356.
- [55] Freedman G, Fattal R. Image and video upscaling from local self-examples[J]. *Acm Transactions on Graphics*, 2011, 30(2):1-11.
- [56] Chang H, Yeung D Y, Xiong Y. Super-resolution through neighbor embedding[C]//Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE, 2004, 1: 1-1.
- [57] M. Bevilacqua, A. Roumy, C. Guillemot, and M. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *British Machine Vision Conference, BMVC 2012, Surrey, UK, September 3-7, 2012*:1-10.
- [58] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation[J]. *IEEE Trans Image Process*, 2010, 19(11):2861-2873.
- [59] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 5197-5206.
- [60] Roman Zeyde, Michael Elad, Matan Protter. On single image scale-up using sparse-representations [C] // International Conference on Curves and Surfaces. Springer-Verlag, 2012:711-730.
- [61] Timofte R, De V, Gool L V. Anchored Neighborhood Regression for Fast Example-Based Super- Resolution[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2013:1920-1927.
- [62] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution[C]//European conference on computer vision. Springer, Cham, 2014: 184-199.
- [63] Shi W, Caballero J, Huszar F, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:1874-1883.
- [64] Dong C, Loy C C, Tang X. Accelerating the super-resolution convolutional neural network[C]//European Conference on Computer Vision. Springer, Cham, 2016: 391-407.
- [65] Liang Y, Wang J, Zhou S, et al. Incorporating image priors with deep convolutional neural networks for image super-resolution[J]. *Neurocomputing*, 2016, 194: 340-347.
- [66] Wang Z, Liu D, Yang J, et al. Deep networks for image super-resolution with sparse prior[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 370-378.
- [67] Shocher A, Cohen N, Irani M. "Zero-Shot" Super-Resolution using Deep Internal Learning[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 3118-3126.
- [68] Yang W, Feng J, Yang J, et al. Deep edge guided recurrent residual learning for image super-resolution[J]. *IEEE Transactions on Image Processing*, 2017, 26(12): 5895-5907.
- [69] Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1646-1654.
- [70] Kim J, Lee J K, Lee K M. Deeply-Recursive Convolutional Network for Image Super-Resolution [C]// Computer Vision and Pattern Recognition. IEEE, 2016:1637-1645.
- [71] Tai Y, Yang J, Liu X. Image Super-Resolution via Deep Recursive Residual Network[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017:2790-2798.
- [72] Lim B, Son S, Kim H, et al. Enhanced Deep Residual Networks for Single Image Super- Resolution[C]// Computer Vision and Pattern Recognition Workshops. IEEE, 2017:1132-1140.
- [73] Lai W S, Huang J B, Ahuja N, et al. Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017:5835-5843.
- [74] Liao Q, Poggio T. Bridging the gaps between residual learning, recurrent neural networks and visual cortex[J]. *arXiv preprint arXiv:1604.03640*, 2016.
- [75] Han W, Chang S, Liu D, et al. Image super-resolution via dual-state recurrent networks[C]//Proc. CVPR. 2018.
- [76] Li J, Fang F, Mei K, et al. Multi-scale Residual Network for Image Super-Resolution[C] // Proceedings of the European Conference on Computer Vision (ECCV). 2018: 517-532.
- [77] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.
- [78] Tong T, Li G, Liu X, et al. Image Super-Resolution Using Dense Skip Connections[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2017:4809-4817.
- [79] Tai Y, Yang J, Liu X, et al. MemNet: A Persistent Memory Network for Image Restoration[C]// IEEE International Conference on Computer Vision. IEEE, 2017:4549-4557.
- [80] Zhang Y, Tian Y, Kong Y, et al. Residual dense network for image super-resolution[C]//The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018.
- [81] Ledig C, Theis L, Huszar F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[J]. *arXiv preprint arXiv:1609.04802*, 2016.
- [82] Arjovsky M, Chintala S, Bottou L. Wasserstein gan[J]. *arXiv preprint arXiv:1701.07875*, 2017.
- [83] Simonyan, K., Vedaldi, A., Zisserman, A.: Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013
- [84] Gatys L, Ecker A S, Bethge M. Texture synthesis using convolutional neural networks[C]// Advances in Neural Information Processing Systems. 2015: 262-270.
- [85] Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. *arXiv preprint arXiv: 1508.06576*, 2015
- [86] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C] //European Conference on Computer Vision. Springer, Cham, 2016: 694-711.
- [87] Sajjadi M S M, Schölkopf B, Hirsch M. Enhancenet: Single image super-resolution through automated texture synthesis[C]//Computer

- Vision (ICCV), 2017 IEEE International Conference on. IEEE, 2017: 4501-4510.
- [88] Tan W, Yan B, Bare B. Feature Super-Resolution: Make Machine See More Clearly[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 3994-4002.
- [89] Park S J, Son H, Cho S, et al. SRFeat: Single Image Super-Resolution with Feature Discrimination [C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018: 439-455.
- [90] Haris M, Shakhnarovich G, Ukita N. Deep back-projection networks for super-resolution[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 1664-1673.
- [91] Wang Y, Perazzi F, McWilliams B, et al. A Fully Progressive Approach to Single-Image Super-Resolution[J]. arXiv preprint arXiv:1804.02900, 2018.
- [92] Wang X, Yu K, Dong C, et al. Recovering realistic texture in image super-resolution by deep spatial feature transform[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 606-615.
- [93] Bulat A, Tzimiropoulos G. Super-FAN: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 109-117.
- [94] Chen Y, Tai Y, Liu X, et al. FSRNet: End-to-End Learning Face Super-Resolution with Facial Priors[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2492-2501.
- [95] Zhang K, Zuo W, Zhang L. Learning a single convolutional super-resolution network for multiple degradations[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2018, 6.
- [96] Bulat A, Yang J, Tzimiropoulos G. To learn image super-resolution, use a GAN to learn how to do image degradation first[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 185-200.
- [97] Sønderby C K, Caballero J, Theis L, et al. Amortised map inference for image super-resolution[J]. arXiv preprint arXiv:1610.04490, 2016.
- [98] Gao J, Guo Y, Yin M. Restricted Boltzmann machine approach to couple dictionary training for image super-resolution[C]// IEEE International Conference on Image Processing. IEEE, 2014:499-503.
- [99] Zhou Y, Qu Y, Xie Y, et al. Image Super-Resolution Using Deep Belief Networks[C]// International Conference on Internet Multimedia Computing and Service. ACM, 2014:28.
- [100] Cui Z, Chang H, Shan S, et al. Deep Network Cascade for Image Super-resolution[M]// Computer Vision – ECCV 2014. Springer International Publishing, 2014:49-64.
- [101] Mao X J, Shen C, Yang Y B. Image restoration using convolutional auto-encoders with symmetric skip connections[J]. arXiv preprint arXiv:1606.08921, 2016.



TANG Yan-Qiu, born in 1995, Master Candidate. Her main research interests include machine learning, deep learning and image processing.

PAN Hong, born in 1974, Ph.D., Associate Research Professor. His main research interests include machine learning, deep learning, computer vision, medical image

processing, machine vision and multimedia signal processing.

ZHU Ya-Ping, born in 1977, Ph.D., Professor. Her main research interests include machine learning, deep learning, image processing and pattern recognition.

LI Xin-De, born in 1975, Ph.D., Professor. His main research interests include machine perception and learning, data mining, machine vision, intelligent information fusion and uncertainty inference, intelligent robots.

Background

Image Super-resolution Reconstruction (SR) is a classical challenging ill-posed problem, aiming to obtain high-resolution (HR) images from one or more low-resolution (LR) images. Since emerging in 1964, SR has been developing continuously and widely applied in different fields, from satellite imaging to medical image processing, to facial image analysis, video surveillance and biometrics.

The development of SR technique has experienced three stages: interpolation-based methods, reconstruction-based methods and learning-based methods that include machine learning-based and the recent deep learning-based methods. Previous methods for SR mainly have two drawbacks: 1). unclear definition of the mapping between the LR space and the HR space; 2). unable to learn a sophisticated high-dimensional mapping from massive raw data. Benefiting from the powerful mining of high-level representations that bridge the LR space and the HR space, recent deep learning-based SR methods have achieved

significant improvements, both quantitatively and qualitatively. When we apply deep learning algorithms to tackle a specific task, it is natural for us to consider both the universal deep learning approaches and the specific domain knowledge. Although many effective strategies like data preprocessing, model training techniques (such as regularization, cross-validation, parameter tuning etc.) are also quite important, integration of deep learning with domain knowledge in SR is the most essential factor to the performance improvement and the main source in the innovations of neural network architectures as well as optimization objectives for SR. In this survey, we first provide a comprehensive overview of typical SR methods, including interpolation-based methods, construction-based algorithms and recent approaches based on machine learning and deep learning. Then we elaborate a variety of deep learning-based solutions to SR tasks from the aspect of network architecture, message pass and optimization

of loss function, and provide a thorough analysis and comment on the pros and cons of all solutions. Finally, we highlight some future trends in the field of SR leveraging deep learning algorithms.

This work is supported by the National Natural Science Foundation of China under Grant (No. 61671151, 61573097, 61703100, 91748106), the Natural Science Foundation of Jiangsu Province under Grant (No.BK20181265), the Key Laboratory of Integrated Automation of Process Industry under Grant (PAL-N201704), Qing Lan Project of Jiangsu Province, Six Major Top-talent Plan of Jiangsu Province, and the Priority Academic Program Development of Jiangsu Higher Education Institutions. These research projects aim to tackle the critical concern of the notorious “black box” effect in the existing deep

learning models by integrating the task-specific constraints as well as the prior domain and experts’ knowledge from the traditional feature engineering and machine learning methods into the design of deep learning architectures. Our group has been working on machine learning, deep learning, computer vision and image processing for 10+ years and published 150+ papers relating with this research direction in prestigious international journals and conferences during the past couple of years. We hope this survey paper can provide readers an overall review of the retrospect and prospect of SR problem, and thus advancing the related research.