

DOI:10.19651/j.cnki.emt.2208826

单图像超分辨率方法综述*

陈晓^{1,2} 荆茹韵¹

(1.南京信息工程大学电子与信息工程学院 南京 210044; 2.南京信息工程大学江苏省大气环境与装备技术协同创新中心 南京 210044)

摘要: 图像超分辨率是指从低分辨率图像生成对应的高分辨率图像。它在人脸识别、数字高清电视、视频通信等领域应用广泛。为了对图像超分辨率技术进行深入探索和总结,本文首先回顾了单图像超分辨率的历史,综述了基于非学习的超分辨率方法,其中展开介绍了基于插值和基于重建的方法,接着重点介绍了基于学习的方法,详细分析了基于深度学习的图像超分辨率,具体总结了 SRCNN、ESPCN、SRGAN 这三种图像超分辨率方法,并将其与递归结构、密集结构、注意力机制网络结构进行对比,之后分析了损失函数和上采样方式在图像超分辨率中的作用,介绍了常用数据集和图像评价指标,展示了图像超分辨率的可视化结果。最后,总结了现有单图像超分辨率方法的进展和不足。

关键词: 图像超分辨率;深度学习;图像处理;方法综述;递归;注意力机制

中图分类号: TP183 **文献标识码:** A **国家标准学科分类代码:** 250.2060

A review on single image super-resolution

Chen Xiao^{1,2} Jing Ruyun¹

(1. School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: Image super-resolution (SR) reconstruction refers to the generation of a corresponding high Resolution image from a low Resolution image. SR has important application value in monitoring, remote sensing, digital HD, video coding communication and other fields. In this paper, we first review the history of single image SR and summarize the non-learning based SR methods. Among them, interpolation and reconstruction based methods are introduced, and then learning-based methods are introduced, in which SR based on deep learning is analyzed such as SRCNN, ESPCN and SRGAN are summarized in detail, and compared with recursive structure, dense structure and attention mechanism network structure. Then the function of loss function and upsampling method in image SR is analyzed, common data sets and image evaluation indexes are introduced, and visualization results of image SR are displayed. Finally, the progress and deficiency of single image super-resolution technology are summarized.

Keywords: image super-resolution; deep learning; image processing; method review; recursion; attention mechanism

0 引言

在采集图像时,模糊、运动变换等会影响成像质量,数模转换时也难以会有失真和退化。这些限制会导致图像质量不佳、分辨率低。通过硬件手段改造图像获取设备来提高图像分辨率需要的技术水平和成本都较高,实际应用也会受到制约,因此,利用软件提高图像分辨率的方法应用而生。目前,图像超分辨率在人脸监控、数字高清电视、视频通信等领域有重要的应用价值。

Harris^[1]在1964年研究了光学成像系统分辨率的物理极限,由此奠定了图像超分辨率(super resolution, SR)的数学基础。随后的1984年, Tsai^[2]和Huang在Harris的研究基础上提出了一种图像SR方法——利用多幅低分辨率(low resolution, LR)图像通过频域变换得到一幅高分辨率(high resolution, HR)图像。根据当前已经存在的研究结果和相关方法的应用,早期图像SR可以分为基于插值、基于重建和基于学习的方法。后来由于更普遍退化模型的出现,图像SR从集中于频域转向了空间域的研究,尤其是

收稿日期:2022-01-11

* 基金项目:江苏省333高层次人才培养工程(2625)项目资助

近年基于学习的图像 SR 方法不断发展,解决了基于重建的方法在放大倍数上的局限性。而随着卷积神经网络的提出,基于深度学习的图像 SR 可以将图像的深层特征从训练集中自适应地学习到,从而建立高低分辨率图像之间的映射关系,且可以保留图像更多的高频特征信息,因此在计算机视觉领域得到了广泛普及。最初基于卷积神经网络的 SRCNN^[3]、SCN^[4]、FSRCNN^[5] 算法虽然结构简单,但无法保留更多的高频细节,之后的深度残差网络 VDSR^[6]、LapSRN^[7]、EDSR^[8]、递归神经网络 DRCN^[9] 和 DRRN^[10],解决了多尺度 SR 需求,且在参数较少时也能有不错的重建性能,再后来的密集连接网络 SRDenseNet^[11] 和 RDN^[12]、生成对抗网络 SRGAN^[13]、RFB-ESRGAN^[14] 和注意力机制网络 SAN^[15] 等,在大尺度放大倍数时,表现出逐渐提升的图像 SR 性能,且涌现出更多新的神经网络结构,如 ADDRSR^[16]、CROSS-MPI^[17]、SRWARP^[18] 等,使得超分辨网络在算法能耗、速度、无法解决任意倍数等问题上得到了改善。

本文综述了单图像超分辨率方法,从介绍基于非学习的图像 SR 算法入手,分别总结基于插值、基于重建算法之间的优缺点,之后,重点会对基于学习中深度学习的图像 SR 算法进行分类,对比不同方法间的差别与联系,阐述常用数据集和对图片的评价和指标,从而进行对于图像 SR 技术拓展应用,发现现有的不足之处。

1 图像超分辨率的概念

图像超分辨率是指用信号处理方法从低分辨率图像生成对应的高分辨率图像。图像分辨率表示单位长度的像素数。单位长度的像素数越多,图像清晰度越好,图中存储的信息量越大。提高图像分辨率最直接的方法是增加图像传感器的像素密度或尺寸。像素密度一方面受制造工艺限制,另一方面像素密度增加导致图像传感器单个像素感光元件接收到的光通量减小,加性噪声等因素使图像信噪比下降、成像质量恶化。增加图像传感器的尺寸会增加成本,而且会受到应用场景和环境等因素的限制。为了克服这些缺陷,图像超分辨率用信号处理的方法从 LR 图像得到 HR 图像,从而提高图像的分辨率。

如果图像做 N 倍($\times N$)超分辨率,则得到的高分辨率图像的像素是原图像的 N^2 倍,也就是说平均一个原像素要生成 N 个新像素,多出 $N-1$ 个像素。由于多出的每个像素的值都有多种可能,对于同一张低分辨率图像,存在多张可能的高分辨率图像。从数学求解的角度看,图像超分辨率就是已知部分数据求全部数据,是一个无中生有的过程,是一个极具挑战性的病态逆问题。通常在求解高分辨率图像时会加一个先验约束条件来输出一个合理的图像。

目前的单图像超分辨率方法主要分为基于插值、基于重建和基于学习的方法。先验约束条件来自于图像的先验知识。前两种方法先验约束条件来自于图像自身。而在基

于学习的方法中,先验约束条件从成对出现的低-高分辨率图像的实例中学到。

2 基于非学习的图像超分辨率

2.1 基于插值的方法

原始图像之间相邻像素存在相关性,基于插值的方法就有效利用了这之间的相关性完成图像 SR。该方法在训练样本不足时,也能拥有较好的效果。通常可细分为最近邻插值法、双线性插值法和双三次插值法。最近邻插值可以描述为当图片放大时,缺少的像素直接照搬旁边最近的与它颜色相似的像素来生成新的图像,该方法生成的图像会产生锯齿,重建效果一般。双线性插值法不直接使用相邻的像素值,而是在缺少像素点的四邻域内的两个方向上分别做线性插值,该方法比最近邻插值法计算量大,结果通常不是线性的,但没有像最近邻插值法的锯齿,效果较好。但由于它的低通滤波特性,无法保留原图像更多的高频信息,所以图像的轮廓会略显模糊。双三次插值法是二维空间中最常用的插值方法最,某个点的像素值由空间矩形网格周围 16 个采样点像素值加权得到。在二维图像空间中,通常需要使用两个多项式插值三次函数,因此,该方法是基于插值的 SR 方法中效果较好的,也经常用它来与深度学习的方法做对比。

2.2 基于重建的方法

先验知识是非常复杂的,但是图像 SR 却必须要以它来作为约束条件来生成 HR 图像。先验知识提高了线性模型的性能使得基于重建的方法在生成清晰锐利的细节方面更具有优势。基于重建的方法对 LR 图像的获取过程建模,构造 HR 图像的先验约束条件,把 SR 就转变为含约束条件的代价函数优化,把病态问题转化为良态问题。通常把人为定义的关于 HR 图像的平滑约束,作为先验知识。

1) 迭代反向投影法

迭代反向投影法(IBP)是在 1991 年提出的,就是将 LR 图像进行一个反向投影,建立 LR 和 HR 图像之间映射的操作。尽管会存在一定的误差集合,但是最后会得到与对应的 HR 图像。后来,Tom 等引入改进的运动补偿方法,将其运用于基于 IBP 算法^[19]的单图像 SR 中,算法不仅直观简单,性能也有了显著提高,更是将其推广到视频序列超分辨重建上,实现更广泛的应用。

2) 凸集投影法

尽管 IBP 算法优点明显,但也存在难以利用先验信息,问题解不唯一的缺点。在 1989 年,Stark 等^[20]和 Oskoui 把凸集投影法(POCS)应用在超分辨率中,但是使用的运动模型却并未考虑到运动模糊所带来的影响,效果有些不大理想。而后在 1998 年,Patti 等^[21]提出了可以包含加性噪声的 POCS 方法,不仅解决了运动模糊的问题,还考虑到了混叠和传感器模糊对 SR 的限制。在 POCS 算法中,一个约束凸集可以由 SR 非常有限的解区域中一个如非负性、

能量有界性等类似的限制条件来定义,然后对带有 HR 图像性质的强制约束集求交集,迭代得到解空间。POCS 克服了 IBP 算法未加入先验信息的弱点,保持了 HR 图像的边缘和细节特征,对图像 SR 的发展有一定的促进作用。

3) 自适应滤波方法

POCS 算法在加入先验信息的同时,存在着收敛稳定性低、运算量大的缺点,虽然加入松弛投影算子可以改善,但图像边缘和细节就会无法保持。1999 年 Elad 等^[22]提出了基于递归最陡下降法和递归最小均方法来实现 SR。2000 年 Alam 等^[23]对滤波做了改进。2007 年 Naratanan 等^[24]提出了基于自适应滤波法,把相邻的 LR 图像进行像素加权,从而实现图像 SR,获得相应的 HR 图像。该方法实现了对图像边缘和细节特征的保留,但不能包含先验知识,尤其不含有非线性先验知识。

3 基于学习的图像超分辨率

基于学习的图像超分辨率方法通过学习获得先验知识。它利用图像在高频细节上的相似性,通过学习样本图像获得 HR 与 LR 图像之间的映射关系,从而生成 HR 图像。

3.1 建立训练样本

基于学习的图像超分辨率方法需要一个训练样本图像库。图像样本库中 LR 图像由 HR 图像经过退化而成,退化过程可以用以下公式表示:

$$I_x = D(I_y; \delta) \quad (1)$$

在这个公式中, D 所代表的是退化函数, I_y 可以理解为对应的 HR 图像, δ 是表示整个退化过程中像缩放因子或噪声这样的参数。一般来说,退化过程 D 是未知的,只提供 LR 图像 I_x , 所以在这种情况下,需要从 LR 图像 I_x 中恢复出与对应 HR 图像 I_y 尽可能相似的超分辨率图像 \hat{I}_y , 该过程可以表示为:

$$\hat{I}_y = F(I_x; \theta) \quad (2)$$

其中, F 表示超分辨模型, θ 表示模型的参数。

退化过程可能会受到例如各向异性退化、散斑噪声、压缩伪影和传感器噪声等不利影响,所以过程处于未知状态。在对图像 SR 时,如果可以知道图像经历怎样的过程导致退化,图像恢复就会变得相对容易。因此建立相关的图像退化模型,模拟整个退化过程是至关重要的。由于线性系统利于求解图像恢复问题,运算简单,因此线性非时移系统经常被用于图像退化模型中,尤其是需要的退化图像不太复杂时,线性非时移系统模型可极大地简化计算。而非线性时移系统尽管能更加准确地描述图像恢复问题的本质,但其在数学上很难求解,通常只会在对图像退化提出非常精确要求时使用。除了线非时移退化模型以外,模糊退化和随机噪声退化模型也经常用来进行图像 SR 之前的退化工作。

目前的研究工作很多是直接退化过程表示为图像的下采样操作,从而建立退化映射过程,如下:

$$D(I_y; \delta) = (I_y) \downarrow_s, \{s\} \subset \delta \quad (3)$$

式中: \downarrow_s 所代表的其实是具有缩放因子 s 的下采样操作,双三次插值下采样是常用手法。但也有些更复杂的卷积和下采样的组合来演示整个退化过程,如下:

$$D(I_y; \delta) = (I_y \otimes k) \downarrow_s + n_s, \{k, s, s\} \subset \delta \quad (4)$$

式中: $I_y \otimes k$ 所代表的是 HR 图像和模糊核 k 之间存在的卷积, n_s 是代表在这当中的一些具有标准偏差 s 的加性高斯白噪声,该退化过程已被证实比式(3)的模型更接近真实情况,且对 SR 过程更加有利。

综合以上内容,超分辨重建过程可被描述为:

$$\hat{\theta} = \operatorname{argmin} L(\hat{I}_y, I_y) + \lambda \Phi(\theta) \quad (5)$$

其中, $L(\hat{I}_y, I_y)$ 表示生成的 HR 图像 \hat{I}_y 与原始图像 I_y 之间的误差,用来作为损失函数使用, $\Phi(\theta)$ 表示正则项, λ 表示权衡函数,即正则项的系数。

3.2 基于传统学习的图像超分辨率

为了研究 HR 与 LR 图像之间关系,需要利用一定数量的样本图像来进行训练,同时还可以将研究结果应用在低分辨率图像生成高分辨率图像的过程中。在深度学习用于图像重建领域之前,基于学习的方法由于具有较快的运算速度和出色的重建性能而备受关注,是 SR 领域较为主流的方法之一。

1) K 邻近方法

2000 年, Freeman 等^[25]研究出了基于学习的图像 SR。他创造了在 NN 算法之上加入马尔可夫网络的算法。通过这个网络模型来学习样本,并从中学习 LR 和 HR 图像之间的联系。通过该网络的学习,可以使图像细节信息被预估到,从而建立 LR 和 HR 图像之间的映射关系,实现图像 SR。该方法不仅提高了 NN 算法的鲁棒性,还使图像中的高频信息得以提取出来,并且能够实现放大尺度为 4 倍的图像效果。

这种方法在提出之前还未使用基于学习的方法来实现超分辨率。基于学习的方法相比于基于重建的方法,能够更好地得到丰富的高频信息,并且在方法四倍之后,依然能够得到高清图片。但 KNN 算法对图像噪声和训练样本的条件比较高,给前期数据处理带来了一定的难度。

2) 幻想脸法和图像类推法

在基于学习的图像 SR 领域, Baker 等^[26]于 1999 年提出幻想脸法,在最大后验概率的网络框架下,对人脸图像加以金字塔结构的梯度分布先验信息,按照像素进行重建。之后, Hertzman^[27]又提出了图像类推的 SR 方法,本质上是一个基于多尺度的自回归模型,也就是将所提供的图片来作为训练样本。把其中的一幅图片作为另外一个版本图片的滤波版本,最后将学习到的新知识应用到新的图片中,最终会得到一个比较相似的滤波图像。

4 基于深度学习的图像超分辨率

在卷积神经网络出现之前,图像SR问题的关键在于LR图像和HR图像之间共享特征信息。之前的方法对于这些信息的利用还不够完全,导致图像轮廓纹理的表达并不是很充分。卷积神经网络开始应用于计算机的视觉领域之后,各种基于深度学习的SR方法不断涌现。CNN模型由特征提取、非线性映射和特征重建三部分组成,卷积在这三部分起到了重要的作用,CNN的优点是在数学上表达简单,可以通过联合训练完成对图像的SR。由此可以看出,深度学习是比较适合在单幅图像超分辨率重建领域发挥一定应用的。

4.1 经典的基于深度学习的图像超分辨率模型

1) SRCNN

SRCNN(super-resolution convolutional neural network)是基于深度学习图像SR的开山之作。2014年,Dong等^[3]首次根据三层的卷积神经网络去模拟低分辨率图像和高分辨率图像之间的非线性映射。具体的SRCNN网络模型如图1所示。该网络由特征提取层、非线性映射层和特征重建层组成,LR图像会在提取特征层中经卷积操作,提取出多个图像块,将这些图像块经过一定的操作之后,就可以形成非线性映射到另一个特征矩阵上,构成非线性映射层,最后再经过一个卷积层,通过融合将非线性映射层的特征矩阵还原成HR图像。该过程可以用以下公式表示:

$$F_1(I_x) = \max(W_1 * I_x + b_1, 0) \quad (6)$$

$$F_2(I_x) = \max(W_2 * F_1(I_x) + b_2, 0) \quad (7)$$

$$F_3(I_x) = W_3 * F_2(I_x) + b_3 \quad (8)$$

由此公式中可以看出, F 代表卷积层的操作, W 和 b 所代表的其实就是在卷积层中的权重和偏置, $*$ 代表的是卷积操作, $\max(\#, 0)$ 可以当作 ReLU 激活函数使用。

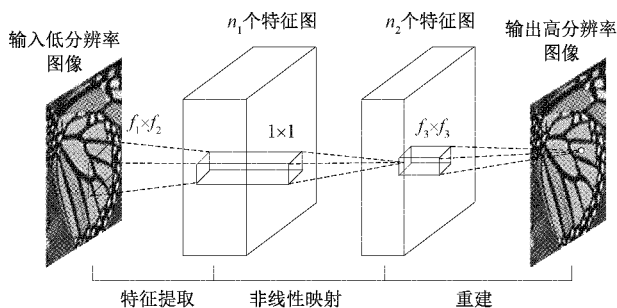


图1 SRCNN网络结构^[3]

SRCNN结构简单,也比之前方法的效果好,但由于训练层数少,训练速度慢,视野不够宽阔等问题导致在放大倍数增加时,效果不及预期。因此,需要研究更深的网络模型来应对这些问题。

2) ESPCN

ESPCN^[28] (efficient sub-pixel convolutional neural network)是一个比SRCNN更深的无需进行上采样就可以

将LR图像通过卷积层实现SR的网络结构,如图2所示。ESPCN算法中存在的卷积与SRCNN中的卷积层不同,它是亚像素卷积。通俗点来理解就是LR图像经过3个卷积层之后,可以形成一个特征图像,这个特征图像是可以由通道数 r^2 来表示的,并且它和最开始的LR图像的大小是相同的,这就代表每个特征图像中的像素通道都可以排列成 $r \times r$ 的区域,它可以和HR图像中一个 $r \times r$ 大小的图像块形成对应效果,之后,大小与 $H \times W \times r^2$ 相同的特征图像会重新进行一次排列,最后排列成 $rW \times rW \times 1$ 的HR图像。

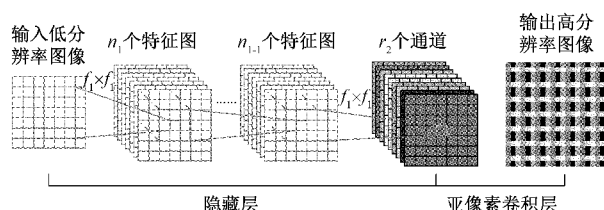


图2 ESPCN网络结构^[28]

亚像素卷积的实质其实没有相关的卷积操作,它仅仅只是通过使用亚像素卷积,来达到图像从LR到HR放大的一个过程,在这过程中,插值函数也被应用到了前面的卷积层^[29]中,这些是可以自动学习的。

该方法并没有真正用到卷积操作,所以叫亚像素卷积,由于特征提取层和非线性映射层的卷积是对LR图像的,在对图像大小做变换时只在最后一层,计算量得到了极大降低,效率高。但由于训练过程中使用的均方误差作为损失函数,无法保留图像原有的高频信息。

3) SRGAN

生成对抗网络^[13] (generative adversarial network, GAN)的出现很好地克服了高频信息丢失的问题,生成器和判别器组成了整个GAN网络,如图3所示。该网络通过训练一个生成函数 G 来建立LR图像和HR图像间的映射关系,可以描述为以下公式:

$$\hat{\theta}_G = \operatorname{argmin}_{\theta_G} \frac{1}{N} \sum_{n=1}^N L^{SR}(G_{\theta_G}(I_n^{LR}), I_n^{HR}) \quad (9)$$

其中, G_{θ_G} 是作为参数为 θ_G 的前馈CNN的一个生成器网络, $\theta_G = \{W_{1,L}; b_{1,L}\}$ 分别代表在L层卷积网络上的权重和偏置,可以通过对损失函数 L^{SR} 的优化得到。

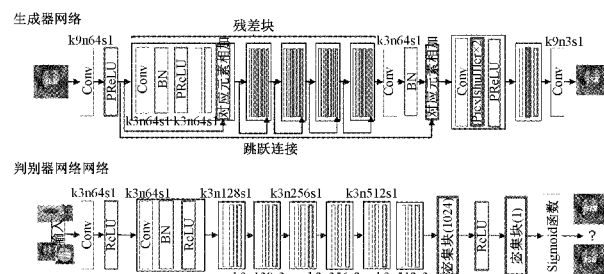


图3 SRGAN网络结构^[13]

这个网络最核心的就是二元零和博弈的对抗过程,生成器的主要作用是用来生成 HR 图像,判别器用来对生成图像和原始图像进行判别。

GAN 很好的解决了图像失真的问题,尽可能保留了图像的高频信息,使得生成的 HR 图像更加逼真,但随着网络深度的不断增加,导致参数的增加,从而不得不扩大训练集。深度过大的网络不仅计算量大,而且硬件耗费巨大,性能也会受到影响。因此,其他改进的网络结构不断涌现。

4) 其他网络结构

除 SRCNN、ESPCN 这样的线性深度结构和 GAN 这样的生成对抗网络以外,递归网络、密集连接网络、注意力机制网络^[30]也有了很大进展。基于递归结构的网络将递归学习引入到图像 SR 领域中。以 Kim 等^[9]为代表的研究者提出 DRCN(deep recursive convolutional network),采用循环监督和跳层连接的方式,用相同的循环层来代替不同的卷积层,可以在提高网络深度的条件下不引入更多的参数,不仅增加了网络的感受野,还能缩短模型的训练时间。之后,Tai 等^[10]又 DRCN 进行了改进,提出了 DRRN(deep recursive residual network),不但借鉴了 DRCN 中卷积层的递归思想,还借鉴了 VDSR(very deep convolutional network)中的全局残差连接,综合提出了局部残差连接的递归思想,在参数更少的情况下依旧能保持较高的重建性能。

近年来,受到 Huang 等^[31]提出的基于密集块的 DenseNet(dense convolutional network)启发,密集连接出现在了图像 SR 中。Zhang 等^[32]结合残差递归模块和密集连接模块,总结出了 RDN(residual dense network)算法,这个算法主要设计出残差密集模块,融合了密集连接层的特点和局部的一些特征,使得图像的特征信息可以得到充分利用,训练过程速度明显提升,性能也优于之前的残差递归方法。

上述所说的网络结构中,所有空间位置和图像通道的 ReLU 函数权重相同,如果可以选择出对图像 SR 相对更重要的激活值,并给予更大权重,重建性能将会得到很大提升。因此,基于注意力机制的网络结构可以实现这一想法。Zhang 等^[33]提出 RCAN(residual channel attention network)网络结构,将通道注意力机制加入残差网络,构建残差注意力模块,给图像中包含关键信息的通道赋予更多权重来提高网络的学习能力,实验表明,RCAN 的网络深度提升明显,SR 效果也比单纯的残差网络好很多。在这之后,Dai 等^[15]又提出 SAN(Second-order attention network)模型,利用二阶通道注意力模块中的二阶特征统计量自适应地实现对 LR 图像的重建。具体的网络结构分类如图 4 所示。

4.2 损失函数

图像 SR 常用损失函数是用来衡量生成的 HR 图像与原始 HR 图像之间的差异。在早期的图像研究中,逐像素损

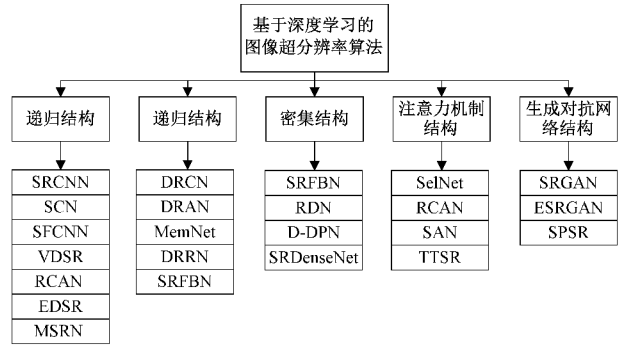


图 4 基于深度学习的图像超分辨率方法网络结构分类

失函数是常用的一种方法,也可以称之为均方误差(mean-square error, MSE),主要用来描述损失函数。但是后来经过研究表明该损失函数在测量图像感知质量时存在较大差异。因此,其他损失函数相继被提出,目前的感知损失函数、纹理损失函数、对抗损失函数等都可以用来测量生成的 HR 图像与原始 HR 图像间的差异。

使用逐像素损失进行衡量,具体可包括平均绝对误差 L_1 和均方误差 L_2 , 可以用以下公式表达:

$$L_{pixel_l1}(\hat{I}, I) = \frac{1}{hwc} \sum_{i,j,k} |\hat{I}_{i,j,k} - I_{i,j,k}| \quad (10)$$

$$L_{pixel_l2}(\hat{I}, I) = \frac{1}{hwc} \sum_{i,j,k} (\hat{I}_{i,j,k} - I_{i,j,k})^2 \quad (11)$$

其中, h, w, c 分别用来评价图像的高度、宽度和通道数, \hat{I} 表示生成的 HR 图像, I 表示原始输入的 LR 图像。 L_1 在性能和收敛性上都优于 L_2 , 并且像素损耗与 PSNR 存在很大的关联,因此像素损失函数便开始广泛使用。SRCNN 使用的就是 L_2 损失函数。像素损失函数并不会去考虑图像的质量,所以生成的 HR 图像基本都缺乏高频信息,这就导致图像的边缘非常的平滑。

在像素损失函数之后提出的感知损失函数、纹理损失函数、对抗损失函数等损失函数,都是在原有像素损失的基础上进行改进,克服了图像边缘问题,在用不同算法进行图像 SR 时,可以根据函数各自的特点运用到对应的网络结构中,提高图像效果。

4.3 上采样方式

直接去学习从低维空间到高维空间的映射,从理论上讲是很困难的,所以通过上采样算法来对高分辨率的图像采样,之后通过深度神经网络来进行精炼的过程,是一种非常简单的并且可行的解决方案。图像插值是对数字图像进行缩放,由于插值的方法众多,所以插值的结果与先进行哪个方向的插值无关;双三次插值使用的插值函数为三次多项式,也是进行图像两个方向的插值操作。这些方法简单易实现,在基于卷积神经网络的图像 SR 中广泛得到了应用。SRCNN 对图像先进行双三次插值建立 LR 与 HR 图像间端到端的线性映射,将 LR 图像上采样到 HR 图像,然后利用卷积神经网络对图像进行重建,恢复出更多

的图像信息。

除了基于插值的上采样方法以外,基于学习的采样方法用端到端的方式,引入转置卷积层和亚像素卷积层的概念,进行图像上采样。最经典的例子就是上述所介绍的ESPCN网络模型,运用亚像素卷积层进行上采样,实现图像的SR。这些基于学习的上采样框架从低维空间中提取高频信息来进行图像重建,避免了高维空间中的过度操作,已经成为图像SR领域应用最广泛的方法。

最近的研究专注于放大尺度的研究,Zhou等^[17]提出的Cross-MPI算法,利用隐藏的场景结构进行高效的基于注意力的对应搜索进行图像SR,在8倍尺度上的性能有了明显提升。而如今更有网络在16倍尺度上拥有不错的重建效果,如RFB-ESRGAN^[14],其上采样方式结合最近邻插值和亚像素卷积法,改善网络空间和深度之间的信息交流,极大提高了图像重建后的放大倍数,是图像SR领域的又一次创新。

4.4 常用数据集

现有评价SR方法的图像数据集有很多都是根据深度的图像数据提取出来的,如Set5^[34],Set14^[35],BSD100^[36],Urban100^[37],Manga109^[38]等。上述一些数据集当中的图像数量和质量上存在较大差异,且在用不同方法进行SR时在不同放大倍数下SR图像的效果也不一样。同时,其中有些数据集仅提供HR图像,还需对次进行双三次插值操作来构建新的图像数据集,图5是上述所说的6个数据集中具有代表性的图像内容。

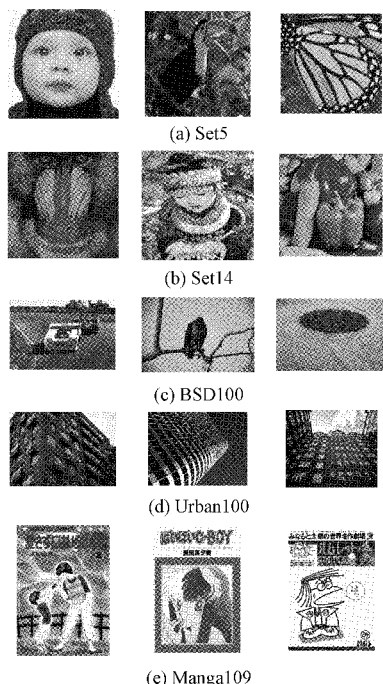


图5 5个超分辨率数据集的代表性图像

这里选取了其中Set5和MangoDB中的图像来进行双立方插值(Bicubic)和各种深度学习算法可视化效果的对比,如图6和7所示。

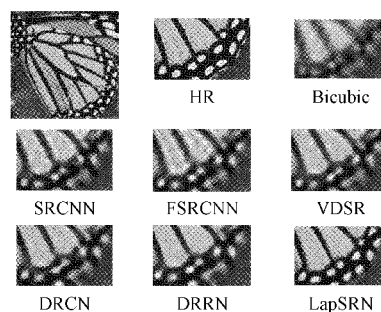


图6 各种算法在Set5上(butterfly)×8 SR的可视化结果

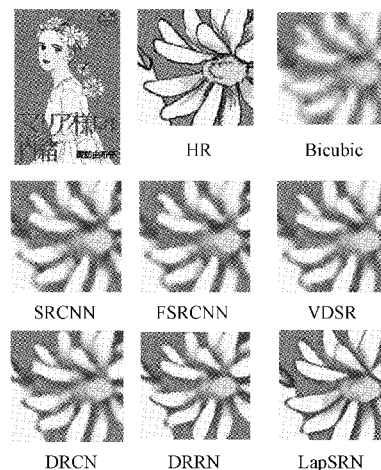


图7 各种算法在Manga109上(MariaSama.Niha.Naisyo)×8 SR的可视化结果

4.5 图像评价指标

常用的图像SR客观评价指标有峰值信噪比(PSNR)和结构相似性(SSIM)两个。

PSNR可以理解为此相关的两个数的比值(信号最大可能功率和影响它的表示精度的破坏性噪声功率的比值)。从以下公式中可以明显看出:

$$PSNR = 10\lg\left(\frac{x_{\max}^2}{MSE}\right) \quad (12)$$

式中: x_{\max} 所代表的是真实图像的最大像素值, MSE 表示最小化损失函数,可以对比为最大化PSNR指标。PSNR数值越大表示失真越小,在进行图像与影像的压缩中,如果可以将峰值信噪比的值控制在30~50 dB之间是最比较合适的,且这个数值越高越好。在这个数值里面,在接近50 dB的时候,说明这个图片经过压缩后的误差是非常小的。当PSNR值大于30 dB时,人眼是察觉不出来图像压缩前后的区别的。当PSNR值在20 dB和30 dB之间时,图像间的差异可以用肉眼察觉出。当PSNR值介于10 dB和20 dB之间时,虽然也有差异,但是不会那么明显。当PSNR值低于10 dB,就会很难判断图像是否相同,也不会联想到他们是压缩前和压缩后的效果。

SSIM 可以理解为人从最直观的视觉系统上来对信息进行观察和评价,最后得出这个图片的质量的标准,图像之间的亮度、结构和对比度决定了 SSIM,可以表示为以下公式:

$$SSIM(I, \hat{I}) = \frac{(2\mu_I\mu_{\hat{I}} + C_1)(\sigma_{\hat{I}\hat{I}} + C_2)}{(\mu_I^2 + \mu_{\hat{I}}^2 + C_1)(\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2)} \quad (13)$$

其中, μ_I 和 $\mu_{\hat{I}}$ 可以用来表示 I 和 \hat{I} 的均值, σ_I 和 $\sigma_{\hat{I}}$ 表示方差, $\sigma_{\hat{I}\hat{I}}$ 表示 I 和 \hat{I} 的协方差, C_1 和 C_2 是其中的常数。SSIM 的指标数值,越大就表示这两个信号之间的相似度达到了高度契合,图像重建的质量自然也是越好。不同算法在 4 倍尺度的定量分析结果如表 1 所示。

表 1 不同 SR 算法的上采样方式和在 4 倍尺度上的定量分析 (PSNR/SSIM)

算法	上采样方式	Set5	Set14	BSD100	Urban100	Mangal09
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic	双三次插值	28.43/0.810 9	26.00/0.702 3	25.96/0.667 8	23.14/0.657 4	25.15/0.789 0
SRCNN	双三次插值	30.48/0.862 8	27.50/0.751 3	26.9/0.710 3	24.52/0.722 6	27.66/0.858 0
ESPCN	亚像素卷积	30.66/0.864 6	27.71/0.756 2	26.98/0.712 4	24.60/0.736 0	27.70/0.856 0
FSRCNN	转置卷积	30.73/0.860 1	27.71/0.748 8	26.98/0.702 9	24.62/0.727 2	27.90/0.851 7
VDSR	双三次插值	31.35/0.883 8	28.02/0.767 8	27.29/0.725 2	25.18/0.752 5	28.82/0.886 0
DRCN	双三次插值	31.53/0.885 4	28.03/0.767 3	27.24/0.723 3	25.14/0.751 1	28.97/0.886 0
LapSRN	亚像素卷积	31.54/0.886 6	28.09/0.769 4	27.32/0.726 4	25.21/0.755 3	29.09/0.890 0
DRRN	双三次插值	31.68/0.888 8	28.21/0.772 0	27.38/0.728 4	25.44/0.763 8	29.46/0.896 0
D-DBPN	双三次插值	37.57/0.958 9	33.09/0.913 2	31.83/0.895 1	30.75/0.913 3	37.65/0.974 7
EDSR	亚像素卷积	32.46/0.896 8	28.80/0.787 6	27.71/0.742 0	26.64/0.803 3	31.02/0.914 8
DRCN	亚像素卷积	31.53/0.885 4	28.03/0.767 3	27.24/0.723 3	25.14/0.751 1	28.97/0.886 0
SRFBN	转置卷积	37.71/0.959 5	33.29/0.914 2	32.02/0.897 9	31.08/0.917 9	33.42/0.912 3
RCAN	亚像素卷积	32.62/0.900 1	28.86/0.788 8	27.76/0.743 5	26.82/0.808 7	31.21/0.917 2
SRGAN	亚像素卷积	29.04/0.847 2	26.02/0.739 7	25.16/0.668 8	25.13/0.808 3	30.10/0.902 3

5 结 论

本文总结并归纳了从单图像超分辨率概念出现以来的部分成果,对其中基于学习和基于重建这两种方法进行了解释,如基于重建中的自适应滤波法、迭代反向投影法和凸集投影法,基于学习中的 KNN 算法、幻想脸和图像类推法,详细阐述了各类方法的起源、内容和优缺点。之后,在基于学习的方法中引入基于深度学习的图像 SR 方法,阐述退化模型在该方法中的重要性,从最早的 SRCNN 模型到 ESPCN、SRGAN,再到后来递归、密集连接、注意力机制的网络结构,这些基于深度学习网络的方法将图像的高频信息不断得到更多的保留,生成图像质量获得了提升^[39],且在更大的尺度(×8、×16)上也能有很好的 SR 效果^[40]。虽然图像 SR 技术得到了迅速的发展,但仍存在算法耗时长、图像退化模型未知、损失函数不够明确、算法可移植性差、评价标准不准确等问题,因此,未来可在这些方面研究。结合其他算法^[41-54]预期可以进一步提高 SR 效果。

参考文献

[1] HARRIS J L. Diffraction and resolving power[J]. JOSA, 1964, 51(7): 931-936.

[2] TSAI R. Multiframe image restoration and registration[M]. Greenwich Advance in Computer Vision and Image Processing, 1984: 317-339.

[3] DONG C, LOY C, He K, et al. Learning a deep convolutional network for image super-resolution[C]. Proceedings of the European Conference on Computer Vision, 2014: 184-199.

[4] FAN Y, YU J, LIU D, et al. Scale-wise convolution for image restoration[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 10770-10777.

[5] CHAO D, CHEN C, TANG X. Accelerating the super-resolution convolutional neural network [C]. European Conference on Computer Vision, Springer International Publishing, 2016: 391-407.

[6] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C]. IEEE Conference on Computer Vision & Pattern Recognition, 2016: 1646-1654.

[7] LAI W, HUANG J, AHUJA N, et al. Deep laplacian pyramid networks for fast and accurate super-resolution[C]. IEEE Conference on Computer Vision & Pattern

- Recognition, IEEE Computer Society, 2017:5835-5843.
- [8] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops(CVPRW), 2017:136-144.
- [9] KIM J, LEE J K, LEE K M. Deeply-recursive convolutional network for image super-resolution[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016: 1637-1645.
- [10] TAI Y Y, YANG J, LIU X. Image super-resolution via deep recursive residual network[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2017: 2790-2798.
- [11] 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: 4799-4807.
- [12] ZHANG Y, TIAN Y, KONG Y, et al. Residual dense network for image super-resolution [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2472-2481.
- [13] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4681-4690.
- [14] SHANG T, DAI Q, ZHU S, et al. Perceptual extreme super-resolution network with receptive field block[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 440-441.
- [15] DAI T, CAI J, ZHANG Y, et al. Second-order attention network for single image super-resolution [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 11065-11074.
- [16] SONG D, WANG Y, CHEN H, et al. Addersr: Towards energy efficient image super-resolution[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 15648-15657.
- [17] ZHOU Y, WU G, FU Y, et al. Cross-MPI: Cross-scale stereo for image super-resolution using multiplane images[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 14842-14851.
- [18] KONG X, ZHAO H, QIAO Y, et al. ClassSR: A general framework to accelerate super-resolution networks by data characteristic[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 12016-12025.
- [19] QIN F, HE X, WU W, et al. Image super-resolution reconstruction based on sub-pixel registration and iterative back projection[M]. Information Computing and Automation, 2008: 277-280.
- [20] STARK H, OSKOUI P. High-resolution image recovery from image-plane arrays, using convex projections[J]. JOSA A, 1989, 6(11): 1715-1726.
- [21] PATTI A J, ALTUNBASAK Y. Artifact reduction for POCS-based super resolution with edge adaptive regularization and higher-order interpolants [C]. Proceedings 1998 International Conference on Image Processing, 1998: 217-221.
- [22] ELAD M, FEUER A. Super-resolution reconstruction of continuous image sequences [C]. Proceedings of International Conference on Image Processing, 1999, 3: 459-463.
- [23] ALAM M S, BOGNAR J G, HARDIE R C, et al. Infrared image registration and high-resolution reconstruction using multiple translationally shifted aliased video frames [J]. IEEE Transactions on Instrumentation and Measurement, 2000, 49(5): 915-923.
- [24] NARATANAN B, HARDIE R C, BARNER K E, et al. A computationally efficient super-resolution algorithm for video processing using partition filters [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17(5): 621-634.
- [25] FREEMAN W T, PASZTOR E C, CARMICHAEL O T. Learning low-level vision[J]. International Journal of Computer Vision, 2000, 40(1): 25-47.
- [26] BAKER S, KANADE T. Super-resolution optical flow[M]. Carnegie Mellon University, The Robotics Institute, 1999: 36-99.
- [27] HERTZMANN A. Paint by relaxation [C]. Proceedings of Computer Graphics International. IEEE, 2001: 47-54.
- [28] 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]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1874-1883.
- [29] 刘正男, 王凤随, 付林军. 改进的联合型图像超分辨率重建算法[J]. 电子测量与仪器学报, 2020, 34(1): 111-120.
- [30] 徐永兵, 袁东, 张志良. 多注意力机制引导的双目图

- 像超分辨率重建算法[J]. 电子测量技术, 2021, 44(15): 103-108.
- [31] HUANG G, LIU Z, VANDERMAATEN L, et al. Densely connected convolutional networks [C]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2017: 2261-2269.
- [32] ZHANG Y L, TIAN Y P, KONG Y, et al. Residual dense network for image super-resolution [C]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018: 2472-2481.
- [33] ZHANG Y, LI K, LI K, et al. Image super-resolution using very deep residual channel attention networks [C]. Proceedings of the ECCV, 2018: 286-301.
- [34] BEVILACQUA M, ROUMY A, GUILLEMOT C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding [C]. Proceedings of the 23rd British Machine Vision Conference, 2012: 135.1-135.10.
- [35] ZEYDE R, ELAD M, PROTTER M. On single image scale-up using sparse-representations [C]. Proceedings of the 7th International Conference on Curves and Surfaces, 2010: 711-730.
- [36] MARTIN D, FOWLKES C, TAL D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]. Proceedings Eighth IEEE International Conference on Computer Vision, 2001: 416-423.
- [37] HUANG J, SINGH A, AHUJA N. Single image super-resolution from transformed self-exemplars[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5197-5206.
- [38] FUJIMOTO A, OGAWA T, YAMAMOTO K, et al. Manga109 dataset and creation of metadata [C]. Proceedings of the 1st International Workshop on coMics analysis, Processing and Understanding, 2016: 1-5.
- [39] 孙超文, 陈晓. 基于多尺度特征融合反投影网络的图像超分辨率重建[J]. 自动化学报, 2021, 47(7): 1689-1700.
- [40] CHEN X, SUN C. Multiscale recursive feedback network for image super-resolution[C]. IEEE Access, 2022, 10: 6393-6406.
- [41] 沈千里, 陈晓, 支亚京, 等. 一种新的人脸图像去噪算法[J]. 传感器与微系统, 2015, 34(11): 133-136.
- [42] 赵苏徽, 陈晓. 基于树莓派和云平台的智能灌溉系统[J]. 计算机系统应用, 2022, 31(4): 123-129.
- [43] 陈晓, 杨瑶. 基于长期递归卷积网络的无创血压测量[J]. 电子测量技术, 2022, 45(5): 51-59.
- [44] 杨瑶, 陈晓. 基于神经网络的说话人识别实验设计[J]. 实验室研究与探索, 2020, 39(9): 38-41.
- [45] 任晓晔, 陈晓, 郭妍. 基于 Fluent 和 LSTM 神经网络的超声波测风仪阴影效应补偿研究[J]. 计算机应用与软件, 2019, 36(7): 89-98.
- [46] 朱娟, 陈晓. 改进胶囊网络的有序重叠手写数字识别方法[J]. 激光杂志, 2019, 40(7): 43-46.
- [47] 郭妍, 陈晓, 任晓晔. 一种优化随机森林模型的室内定位方法[J]. 激光杂志, 2018, 39(10): 70-74.
- [48] 徐畅, 陈晓, 季仟亿. 基于稀疏编码的 Shearlet 域图像去噪[J]. 激光杂志, 2017, 38(10): 96-100.
- [49] 邹胜男, 陈晓, 陈霞. 基于改进的 UKF 算法的室内测距定位[J]. 激光杂志, 2017, 38(4): 61-65.
- [50] 陈霞, 陈晓, 邹胜男. 基于正态检验的室内定位算法[J]. 激光杂志, 2017, 38(3): 41-45.
- [51] 胡春杰, 陈晓, 陈霞. 基于改进广义 Arnold 映射的多混沌图像加密算法[J]. 包装工程, 2017, 38(3): 144-149.
- [52] 倪龙, 陈晓. 基于频散补偿和分数阶微分的多模式兰姆波分离[J]. 物理学报, 2018, 67(20): 204301-1-10.
- [53] 陈晓, 张凯. 嵌入式自动气象站故障检测系统研究[J]. 电子测量技术, 2021, 44(23): 158-164.
- [54] 胡春杰, 陈晓, 郭银. 基于多混沌映射的光学图像加密算法[J]. 激光杂志, 2017, 38(1): 110-114.

作者简介

陈晓(通信作者), 博士, 教授, 主要研究方向为现代电子系统设计、信号与信息处理、超声成像、图像处理等。

E-mail: chenxiao@nuist.edu.cn

荆茹韵, 硕士研究生, 主要从事计算机视觉方面的研究。

E-mail: jingruyun01@163.com