

图像修复方法综述

李雪涛¹, 王耀雄², 高放^{1*}

¹广西大学电气工程学院, 广西 南宁 530004;

²中国科学院合肥智能机械研究所, 安徽 合肥 230031

摘要 图像修复技术是目前计算机视觉领域的研究热点之一, 该技术主要利用缺失区域周边或者外部辅助数据来对图片受损区域进行信息推理和修复。随着大数据时代的到来, 基于深度学习的图像修复技术以其出色的性能成为了图像处理领域内的关注点。对现有的图像修复算法进行总结, 对各个算法的模型结构、性能表现及在常用数据集上的指标进行讨论说明, 并对该领域目前所存在的问题和难点进行分析和展望。

关键词 图像修复; 深度学习; 卷积神经网络; 自编码网络; 对抗生成网络

中图分类号 O436

文献标志码 A

DOI: 10.3788/LOP212680

Review of Image Inpainting Methods

Li Xuetao¹, Wang Yaoxiong², Gao Fang^{1*}

¹School of Electrical Engineering, Guangxi University, Nanning 530004, Guangxi, China;

²Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, Anhui, China

Abstract Image inpainting is a hot topic in the field of computer vision. It is a process that enables filling in damaged regions with alternative contents by estimating the relevant information either from surrounding areas or external data. With the advent of big data, image inpainting methods based on deep learning have attracted significant attention in image processing because of their excellent performance. This paper presents a brief review of existing image inpainting approaches and discusses the network structure and performance of each algorithm, along with a comparison of widely used datasets. In view of the existing challenges in this field, this paper proposes potential research directions and developmental trends in image inpainting.

Key words image inpainting; deep learning; convolutional neural network; auto encoder network; generative adversarial network

1 引言

图像修复(image inpainting)技术历史悠久, 其主要工作原理是根据图片现存的已知内容推理出缺失区域的未知内容。该技术最早出现在文艺复兴时期, 主要用于艺术画作的修复, 然而在当时, 这项工作很大程度上依赖于修复师的经验与手法, 且十分耗时。随着计算机技术的不断发展, 数字图像修复已成为计算机视觉(computer vision)和计算机图形学(computer graphics)领域中一项重要研究内容^[1-4], 并在文物修复、图像增强和图片内容移除等诸多场景中得到广泛应用。

图像修复技术的发展经历了传统到现代的转变。传统的图片修复方法主要根据图片中已有的纹理信息和结构信息, 利用不同的算法对缺失区域的图像信息进行推断^[4]。由于当时计算机算力有限, 传统方法无法捕捉到图片中的高级语义信息, 其修复结果存在模糊、断层等现象。随着大数据时代的到来, 计算机硬件水平飞速提升, 深度学习理念也再一次进入人们的视野。卷积神经网络(CNN)^[5]的出现, 使计算机能够更好地对图片进行特征学习, 并且生成对抗网络(GAN)^[6]的出现让模型学习概率分布成为可能。基于以上几点, 图像修复技术得到跨越式的发展^[1]。

为拓展图像修复技术的研究和应用, 本文主要关

收稿日期: 2021-10-08; 修回日期: 2021-10-28; 录用日期: 2021-11-11; 网络首发日期: 2021-12-12

基金项目: 广西科技基地和人才专项(2020AC19253)、安徽省重点研究与开发计划(202104a05020041)

通信作者: *fgao@gxu.edu.cn

注近年来涌现出的图像修复新模型,包括基于自编码器(AE)^[7]的图像修复模型及基于GAN的图像修复模型,并重点阐述了各个模型之间的迭代关系和性能对比。为了让读者能够更直观地了解不同算法的创新之处,在第2、3小节最后列出了算法汇总表,并在第5小节列出了代表性算法的指标对比。最后,对图像修复领域目前所存在的问题和难点进行分析,并对学科热点和研究趋势进行展望。

2 传统图像修复方法

传统图像修复方法在广义上可以分为两类,分别是基于样本的图像修复方法和基于扩散的图像修复方法。这两类方法的主要思想都是根据像素之间的相似性,对受损区域的未知信息进行推理,再利用广播机制对生成的像素进行传播,从而完成图片修复工作。

2.1 基于样本的图像修复方法

基于样本的图像修复方法首先根据已有的图像信息生成补丁,然后运用算法模型挑选出与缺失区域匹配度良好的补丁,再对缺失区域进行填补,从而完成修复任务。根据补丁的类别,可将基于样本的图像修复方法分为基于纹理合成的修复方法和基于结构合成的修复方法^[6]。

2.1.1 基于纹理合成的方法

纹理是一种在无限二维平面上具有平稳分布的视觉图案。早期基于纹理合成的方法一般会借助统计学知识对缺失区域进行推断并填补,例如基于马尔可夫随机场(MRF)^[8]和基于高斯金字塔^[9]的模型。但这类方法忽略了相似补丁与缺失区域之间的相关性,会导致修复结果出现像素不连续的现象。针对这一点,不同学者对相似补丁与缺失区域匹配度问题进行了改进,如高效简洁的补丁缝合技术^[10]与赋予相似补丁置信度^[11]等方法。Efros等^[8]在1999年提出了基于纹理合成的方法,该方法是一种非参数式的纹理合成方法,通过查询已知像素周围相似的像素点,再利用MRF在这些相似的像素点中随机挑选出一个像素点,作为下一次迭代的输入像素点,以此往复最终完成修复任务。但该方法每次只能产生一个像素点,当图像中的像素点之间差异很大时,会产生肉眼可见的像素不连续现象。对此人们做出了很多改进,包括改进修补技术及改变信息的传播形式等。Efros等^[10]在2001年提出的方法能够较好地解决像素不连续的问题,该方法在迭代过程中生成的是与原图相似的纹理补丁,而不再是单个的像素点,该方法还通过缝合技术把生成的补丁与原图进行拼接,使修复结果的纹理更加连贯。虽然该方法在像素不连续的问题上有所改进,但是修复结果的光滑性在噪声干扰过于强烈时仍然无法达到令人满意的效果。由于不同的补丁有着不同的修复效果,Drori等^[11]提出了一种赋予相似补丁置信度的方法,该方法在迭代逼近未知区域的过

程中,根据置信度由高到低去遍历每一个搜索到的相似补丁,并选择出置信度最高的一段图片作为补丁。由于原始图片信息有限,当图片中语义结构相差较大时,搜索相似补丁的难度较大。基于这一点,人们开始引用外部数据库对算法进行辅助,即通过搜索外部数据库的图片,找出与原图背景相似的补丁。Hays等^[12]基于该思想,先利用低维场景描述符^[13]在外部数据库中快速搜索出与原图背景相似的图片,然后根据缺失区域的坐标位置挑选出与缺失区域语义符合且相似度最高的补丁,最后对图片进行无缝修补。该方法摆脱了修复过程中对二维纹理的要求,在三维结构上的修复效果也十分出色。但该模型过度依赖数据库,当目标图片在数据库中无法匹配时,会出现修复错误的现象。此外,Wei等^[9]提出了一种金字塔式的模型,该模型利用树状矢量化(TSVQ)的方法对全局变量进行搜索,提升了搜索相似纹理的运行速度。而修复效果的改善,则可以通过加强局部特征之间的匹配度来实现,具体有计算图像梯度^[14]和统计类似补丁偏移量^[15]等方式。

综上所述,基于纹理合成的方法大多利用对纹理补丁的全局搜索来得到相似补丁,其补丁来源可以是原图,也可以是外部数据库。但是由于纹理结构的复杂性、不确定性及样本数量的局限性,该类方法的修复结果很可能存在肉眼可见的像素不连续现象。

2.1.2 基于结构合成的方法

图像的结构是指由图像纹理的不同部分(全局轮廓信息)构造的视觉对象^[16],相较于纹理补丁,结构补丁能够在具有不同结构信息及不同背景的图片中较好地完成修复任务,并且能使修复结果比较自然,符合人类的视觉判断。

基于结构合成的图像修复方法一般会借助物理热流偏微分方程(PDE)来对图像进行修复^[17],但鉴于PDE无法修复大面积缺失区域的缺陷,研究人员提出了纹理合成与结构合成相结合的方法^[18],此方法既保证了结构化区域的修复效果,也能对大面积缺失区域进行有效修复。Bertalmio等^[17]在2000年提出了一种基于结构合成的图像修复方法,该方法与纹理合成方法的区别在于,结构合成方法可自动锁定需要填补的区域。此外,该方法通过PDE生成的线性图像结构,可对结构化区域(例如跨越边界的区域)进行修复,虽然对像素不连续性的改善较为明显,但对大面积缺失的受损图片的修复效果欠佳。针对这一不足,Criminisi等^[18]在2004年提出了一种全新的算法,该算法高效地结合纹理合成算法中的纹理复制过程及结构合成算法中线性结构的生成过程,通过计算各个补丁的优先级确定出最优补丁,进而向缺失区域进行图像信息传播。虽然在前人的基础上,该算法取得了很大的突破,但是在多次迭代后会出现置信度丢失的问题。2005年,Cheng等^[19]对Criminisi等^[18]所提出的优先级

函数进行了优化,解决了置信度丢失的问题,提高了系统的鲁棒性。此外,研究者们还从重新定义数据项^[20]、引入图像局部色彩方差^[21]及引入度量像素块复杂度的信息熵^[22]等角度对 Criminisi^[18]算法中的优先级计算进行了优化,从而使得修复效果更优。

2008年,Simakov等^[23]提出了一种非纹理图像的局部修复模型,该模型通过双向相似性度量准则对修复结果进行优化。2009年,Barnes等^[24]受 Simakov等^[23]的启发,结合最近邻域(NNF)的思想,提出了一种匹配补丁(patch-match)的快速随机算法,该算法利用随机场的先验信息,在一个有可能成为补丁的位置上定义最近邻域,并在该最近邻域内通过随机采样寻找补丁。该算法可以解决图像内容大面积缺失问题,但是由于需要额外的内存对最佳距离信息进行储存,

内存开销会增加。此外,针对先验性信息的应用:Huang等^[25]以相似补丁之间的偏移量作为先验条件对结构补丁的生成过程进行约束;Kwatra等^[26]以输入样例为先验信息对纹理合成进行约束;Ružić等^[27]提出了一种全局图像修复方法,该方法会根据上下文信息将图像分割成不同大小的块,再根据缺失区域上下文的语义信息找出与之匹配的候选区域。该方法虽然在速度和修复效果上有显著的提升,但当图片中语义过于复杂时,仍会出现修复错误的情况。

综上所述,基于结构合成的方法虽然可以较好地解决跨边界区域的修复问题,但是当背景复杂,像素点差异较大时,还是会存在像素不连续的现象。另外,对于图片中弯曲的结构及深度模糊的区域,该类方法修复效果也欠佳。表1为传统图像修复方法汇总。

表1 传统图像修复方法汇总
Table 1 Summary of traditional methods on image inpainting

Category	Method	Year/Source	Contribution
Exemplar-based texture synthesis	Efros et al ^[8]	1999/ICCV	Nonparametric texture synthesis
	Wei et al ^[9]	2000/ACM	Gaussian pyramid model based on Markov Random Field
	Efros et al ^[10]	2001/ACM	Simple and efficient image quilting technology
	Ballester et al ^[14]	2001/IEEE Trans	Matching local features by calculating image gradients
	Drori et al ^[11]	2003/ACM	Computing confidence for each patch
	Hays et al ^[12]	2007/ACM	Searching patches within external databases
Exemplar-based structure synthesis	He et al ^[15]	2014/IEEE Trans	Matching local features by using the statistics of similar patches
	Bertalmio et al ^[17]	2000/ACM	Using Partial Differential Equation to generate linear structural patches
	Criminisi et al ^[18]	2004/IEEE	Texture and structure information can be transmitted simultaneously
	Chen et al ^[20]	2020/Laser & Optoelectronics Progress	Improved the priority calculation formula ^[18] with the method of refining data items
	Wang ^[21]	2020/Laser & Optoelectronics Progress	Optimized the priority calculation formula ^[18] by introducing the local color variance
	Chen et al ^[22]	2020/Laser & Optoelectronics Progress	Optimized the priority calculation formula ^[18] by introducing the information entropy of measuring the complexity of the pixel block
	Cheng et al ^[19]	2005/IEEE	Optimized the priority function in Criminisi ^[18]
	Kwatra et al ^[26]	2005/ACM TOG	Using planar exemplar guidance
	Simakov et al ^[23]	2008/IEEE	A mathematical model for local restoration of untextured images
	Barnes et al ^[24]	2009/ACM	Fast stochastic calculation based on NNF
Diffusion-based techniques	Ružić et al ^[27]	2014/IEEE	Global repair algorithm combined with Markov Random Field
	Huang et al ^[25]	2014/ACM TOG	Using planar structure guidance
	Bertalmio et al ^[17]	2000/ACM	Diffusion method based on isophote lines
	Shen et al ^[28]	2002/SIAM	Combined total variation denoising model with Partial Differential Equation
	Telea et al ^[29]	2004/JGT	Fast Marching Method

2.2 基于扩散的图像修复方法

基于扩散的图像修复方法模仿修复师手工修复图像的过程,利用不同算法根据缺失区域的边界信息推导出未知区域的图像信息,推导过程是由缺失区域四周向中心传播的过程^[1]。2000年,Bertalmio等^[17]提出了一种基于等光照线和扩散思想的修复方法,该方法

可以在任何方向对缺失区域进行填补,并在计算过程中引入了平滑估计器,确保了沿等光线方向传播的正确性。2002年,Shen等^[28]提出了一种将TV去噪模型与PDE相结合的数字化修复方法,该方法广泛应用于照片修复、文本去除、数字缩放和图像边缘信息编码等领域。Telea等^[29]在此基础上提出了快速推进方法

(FMM),该方法可更快实现基于PDE的算法,其原理是沿着像素梯度来传播图像平滑估计器,从而实现图像信息的快速传播。

综上所述,基于扩散的图像修复方法适用于划痕、直线、曲线和边缘的修补,当缺失区域变大时,修复效果将会变差。另一方面,基于数值迭代的修复过程还会产生巨大的计算量。

传统修复方法在邻域搜索方面取得了巨大的突破,但由于其根据现有图片中的已有信息生成补丁,修复结果可能存在像素不连续区域。因此,当图片中掩膜区域过大时,传统方法将无法生成与原图语义结构一致的修补区域。深度学习理念的出现很好地解决了这一问题,基于深度学习的修复方法不仅可以使修复的区域更加连贯、平滑,而且还能产生多种类的修复结果。

3 基于深度学习的图像修复方法

深度学习(DL)是机器学习(ML)的子集,两者都需要提取特征,不同的是机器学习一般需要人为定义特征,深度学习可以通过模型自主学习特征^[30]。具体而言,机器学习是一种从已知数据中学习规律再对未知数据进行推测的方法,传统机器学习的主要任务是学习一个预测模型,而这个学习过程需要将已知数据的特征作为输入才能对模型进行优化。机器学习所需的特征往往需要通过人工获取,特征处理的过程包括数据预处理、特征提取及特征转换,此过程需要耗费大量人力。传统机器学习可以看作是浅层学习,而深度学习可以从数据中自动学习到有效的特征表示,并通过多层的特征转换把原始数据变成更高层次、更抽象的表示^[30],这一过程往往需要强大的计算设备作为硬件支撑。

LeCun等^[31]指出每一张图片都有特定的标签,深度学习可以通过卷积神经网络将这些标签与图片一一建立联系,通过卷积层检测出图像中各个特征的局部连接,再通过池化层将语义上相似的特征合并,从而生成具有高级语义的修复补丁。随着计算机技术的发展,人们制造出了算力更强的CPU和GPU,使得在图像处理领域运用更大的数据集及更深层次的网络成为可能。基于深度学习的图像修复方法可以广义地划分为3类,分别是基于像素生成的图像修复方法、基于自编码器的图像修复方法和基于生成对抗网络的图像修复方法。

3.1 基于像素生成的图像修复方法

基于像素生成的图像修复方法的核心思想是利用神经网络以某个像素点为基点通过算法向四周生成新的像素点。按像素生成原理可将该类方法分为基于循环神经网络(RNN)^[32]的像素生成方法和基于CNN的像素生成方法,工作原理分别如图1、2所示。

基于RNN的像素生成方法一般利用时序网络对

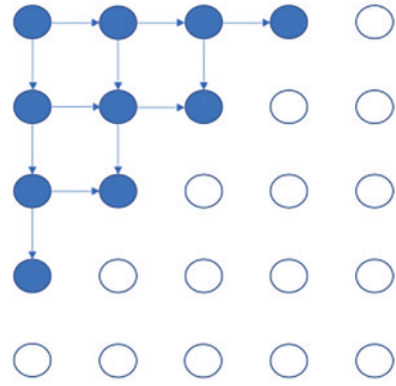


图1 基于RNN的像素生成方法

Fig. 1 RNN based pixel generation method

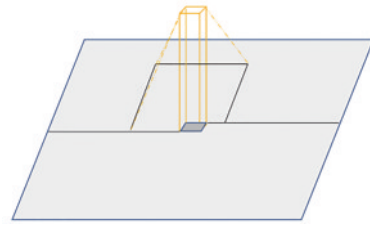


图2 基于CNN的像素生成方法

Fig. 2 CNN based pixel generation method

全局像素点进行遍历,模型再根据全局样本特征分布对缺失区域进行修复。但该方法需要遍历全局像素点,且在修复后期像素点之间的联系会逐渐减弱,导致该模型耗时长、效果差。卷积神经网络由于其局部连接和权重共享等特性可以很好地解决上述问题。目前的卷积神经网络一般由卷积层、池化层和全连接层交叉堆叠而成,并利用反向传播算法对参数进行学习优化,其中卷积层的核心工作是利用卷积核(convolution filter)对所选区域像素点进行特征学习,再依据由回归模型预测出新的像素点。van Oord等^[33]提出了一种二维RNN结构,其利用长短期记忆(LSTM)^[34]层构建像素分布,使模型便于提取全局样本特征,并利用局部和全局像素的相关性对缺失区域进行补充。Richard等^[35]受到卷积神经网络的启发,提出一种快速修复方法,该方法使用高斯核计算受损区域四周像素值的加权平均值,再利用各向同性扩散原理^[17]对受损区域进行修复,卷积核的应用使该模型的运算速度有所提升。2009年,Hadhoud等^[36]对扩散卷积核^[35]的零权重值位置进行了修改,提升了模型速度与修复质量,但该模型仍局限于修复小区域的缺失结构。Jain等^[37]提出了一种运行速度更快的降噪神经网络结构,该模型把计算任务集中在线性回归与概率估计上,并通过随机梯度下降方法随机选取某一训练批次中的少量不相关样本,把基于这些样本的平均梯度来近似作为这一批次样本的梯度,使得整个训练过程更有效率。但由于加入噪声类型的局限性,该模型只能处理前期加入的已知噪声,并且只能处理灰度图像。

基于生成像素的图像修复方法相较于传统修复方法,速度和修复质量上有了较大的提升,但仅限于修复小区域的缺失区域,当图片中存在大面积缺失区域时,这类方法的修复结果仍存在不足。

3.2 基于自编码器的图像修复方法

自编码器的设计动机是为了让机器能够像人一样,从一张图片中提取出能够代表整张图片的高级信息,而不是像 PixelCNN 那样局限于关注像素级别的低维特征。传统自编码器的概念由 Rumelhart 等^[7]在 1988 年提出,其网络结构如图 3 所示。编码器部分包含输入层和隐藏层,负责压缩图片信息。解码器包含隐藏层和输出层,负责还原图片信息。受到传统自编码器的启发,Hinton 等^[38]在传统自编码器的基础上加深了神经网络,提出了深度自编码器(DAE)的概念。该结构解决了主成分分析法(PCA)中不可解释性的缺陷,通过优化初始权重来实现对输入的降维。具体而言,该结构首先通过编码器将高维信息编码为低维信息,得出隐式表示,再通过解码器还原该表示。

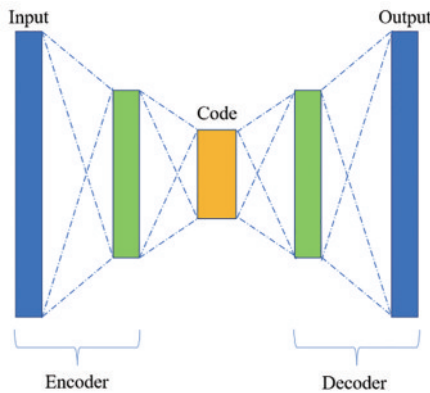


图 3 传统自编码器结构^[7]

Fig. 3 Structure of traditional auto-encoder^[7]

基于自编码器的图像修复方法以端到端的方式学习图像特征,其中编码器部分通过下采样提取图片中潜在特征表示,解码器部分再通过上采样还原图像数据。研究人员基于去噪编码器(denoising auto-encoders)^[39]、上下文编码器(CE)^[40]及全局上下文鉴别器(global discriminator)^[41]等方法改进了卷积算子和

损失函数类型,进一步提升了这类方法的修复效果。Xie 等^[39]在深度自编码器^[38]的基础上提出一种将 CNN 技术与自编码器技术融合的结构,该模型首先利用 BLS-GSM^[42]模型对数据集进行预处理,再通过由该文首次提出的堆叠稀疏去噪自编码器(SSDA)完成对受损图像的去噪、修复操作。该方法通过选择隐藏层最小的激活均值来正则化隐藏层,从而实现稀疏编码并减小模型计算量。由于 SSDA 较为依赖于监督学习,其对不同类型噪声的泛化能力较差。为克服监督学习的局限性,Pathak 等^[40]在 2016 年提出一种基于 CNN 的上下文编码器,该编码器以一种无监督学习的方式进行特征处理,并根据图片的全文信息生成缺失区域的合理语义结构,利用通道级(channel-wise)全连接层代替了传统全连接层,极大减少了参数量。此外,该方法通过结合自编码器与对抗损失函数^[4]提升了补丁的清晰度与真实度,具体网络结构如图 4 所示。实验结果表明,在重构损失函数中加入对抗损失函数可以使得生成的填补区域更加符合上下文语义环境,并且修复细节更加清晰,但是该方法所使用的对抗损失函数仅仅增强了对修复区域的约束,而忽略了修复区域的全局结构一致性,会使修补边界存在模糊区域。针对这一点,Iizuka 等^[41]提出了一种新的架构,该架构可以实现对局部语义和全局语义的理解,进而在保证修复区域全局结构一致性的同时完成对图像的补全。该模型具体由 3 个网络组成,分别是主体网络(completion network)、全局上下文鉴别器和局部上下文鉴别器(local discriminator)。其主体网络加入了空洞卷积(dilated convolutions),可在不增加参数量的同时增加输出的信息量。此外,全局鉴别器会以完整的图像作为输入来识别场景的全局一致性,而局部鉴别器则用来观察修复区域周围的一个小区域来判断已修复区域的连贯性。虽然该方法可以有效地优化修复结果的全局不一致性,但由于对纹理信息的捕捉能力较差及对快速匹配^[29]和泊松混合^[43]技术的过度依赖,其计算量过大,难以处理背景复杂的图像。

除上述方法外,人们还通过引入注意力机制对模型进行优化^[44-46]。这类方法通过上下文注意力模块获

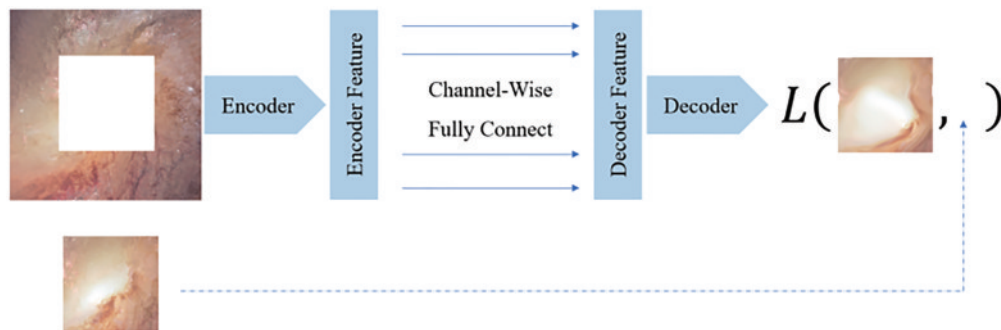


图 4 上下文编码器^[40]

Fig. 4 Context encoder^[40]

取远距离空间信息,并通过融合注意力分数来获得更真实的结果。基于 Iizuka 等^[41]提出的方案,Yu 等^[44]在模型中引入注意力机制,并提出了一种平行双编码器方案。该方案是一个由粗到细的过程,具体而言:首先通过含有空洞卷积层的自编码器产生粗糙修复图像,再将该粗糙修复结果作为输入传输给平行编码器,其中一条编码器通过全文注意力卷积层(contextual attention layer)提取背景特征及网络感兴趣的像素值,另一条编码器则通过空洞卷积对存在修复潜力的特征进行提取;两条编码通道的输出特征值进行聚合后的结果输入单通道解码器,即可解码得出精细的修复结果。该网络解决了距离相隔较远的像素点之间的融合问题,成效十分显著,但是由于注意力机制层需要额外的算力来计算注意力评分,该网络计算成本较高。针对 Yu 等^[44]所提模型中计算成本的问题,Sagong 等^[45]在 2019 年提出了 parallel extended decoder path for semantic inpainting (PEPSI)模型,不同于 Yu 等^[44]平行编码的思路,该模型提出一种并行解码网络,并在此基础上加入共享机制,极大程度减小了系统参数量。此外,该方法还对注意力机制模块进行了改进,用欧氏距离代替余弦相似度,使模型可以更好地学习已知区域与补丁之间的相似关系。Shin 等^[46]提出了 PEPSI 的升级版 Diet-PEPSI,其在 PEPSI 的基础上加入了速率自适应空洞卷积,并且根据膨胀速率产生动态特征,从而进一步减小了系统对硬件设施的要求。

由于现存算法捕获数据分布的能力有限,高分辨率缺失图片的修复问题一直是该领域的难点。基于修复结果可迭代优化的理念,研究人员采用多阶段模型逐步提升修复效果。Yang 等^[47]提出了一种将上下文编码器与纹理补丁相结合的多尺度神经网络补丁生成(MNPS)模型,该模型利用 CE 编码器的输出作为纹理补丁合成模型的输入,再通过传播机制向缺失区域传播纹理信息,从而完成修复任务。此外,该模型还借鉴了风格迁移(style transfer)^[48]的思想,以原图为基础尽可能真实地还原待修复区域的纹理内容,使修复细节更加自然逼真,在对高分辨率图片的修复效果上取得了进展,但是在处理速度上的表现仍然欠佳。针对计算量的问题,Yan 等^[49]提出了名为 Shift-Net 的快速连接网络,该网络在语义分割网络 U-Net^[50]的基础上,进一步引入一种全新的快速连接层,通过联系已知区域的编码器特征与待修复区域的解码器特征来最大程度地保留原图的全局信息,可使修复结果更加精细且完整。此外,该结构利用 Guidance 损失函数来训练参数,以此来增强相互连接的解码器特征与编码器特征之间的联系。该模型相较于 Yang 等^[47]的多尺度模型,处理速度有所提升,多尺度模型处理 256×256 的图像需要 40000 ms,而 Shift-Net 仅需 40 ms。该模型在速度和处理结果上都取得了进展,但由于其关注的是长方形缺失区域,修复结果会在长方形缺失区域的周围

出现像素不连续及语义断层的现象。

观察到大部分的图像修复模型^[40,44,47,49]缺失区域是规则图案,Liu 等^[51]在 2018 年提出一个包含部分卷积层(partial convolutional layer)的网络结构,该结构可以对不规则缺失区域进行修复操作。部分卷积层包含归一化卷积层^[52]和掩膜自动更新机制,具体而言,该模型首先将掩膜转化成二进制输入,然后让卷积层只对非缺失区域像素进行卷积操作,而将缺失区域视为无效像素,以此避免错误的判断迭代影响最终的输出结果。虽然在不规则缺失图片上的修复效果表现突出,但是该模型在像素区域划分时需要人为干预,当有效区域像素分配不均,例如图像中出现稀疏结构(栅栏、网状结构)时,该方法的修复效果并不是很理想。针对这一问题,Xie 等^[53]在 2019 年提出一种可学习的双向注意映射模块,其以端到端的方式学习特征并对掩码进行动态更新,能有效适应像素分布不均的情况,可使修复结果更加清晰且符合上下语义。

针对全文注意力卷积层(contextual attention layer)^[44]和快速连接层^[49]结构中的语义断层现象及部分卷积层(partial convolutional layer)^[51]中的语义错误、边界扭曲现象,Liu 等^[54]在 2019 年提出了一种基于相干语义注意力机制层(CSA layer)的网络结构。该模型适用于任意形状缺失区域的修复,并且可以构建出已知区域与缺失区域的深层次特征的内部联系。此外,为了能够使系统稳定运行并提升修复精度,该模型通过一致性损失(consistency loss)函数来减小相干语义注意力机制层与预处理模型的输出特征之间的误差,并利用特征补丁鉴别器(feature patch discriminators)和补丁鉴别器(patch discriminator)^[55]提升修复结果的分辨率。尽管该算法具有很好的性能表现,但是当整体网络深度没有设置好时,系统可能会发生信息丢失,并导致错误修复。

由于缺少先验结构信息的引导,基于自编码器的修复结果往往存在模糊区域,这种现象在结构信息丰富的区域更为明显。针对这一点,Nazeri 等^[56]提出了一种可以再现原图结构细节的图像修复方法。该模型由一个边缘生成器和一个图像补全网络组成,具体而言,边缘生成器对图像的缺失区域(规则和不规则)产生边缘图谱(edge map),图像补全网络再利用边缘图谱作为先验信息对缺失区域进行填充。该方法弥补了图像修复领域中卷积神经网络与边缘合成网络结合的空缺,并且在全局结构信息的修复上取得了显著突破,但是对精细化局部结构的处理还有所欠缺。针对这一问题,Li 等^[57]在 2019 年提出了视觉结构重建(PRVS)网络,该网络在带有部分卷积层的语义分割网络^[50]的基础上,加入了视觉结构重建层(VSR layer)。视觉结构重建层中包含结构生成器和特征生成器,首先结构生成器根据待修复区域的边缘信息更新缺失范围,特征生成器再利用更新后的缺失区域来指导模型生成新

的特征,这是一个循序渐进、由四周向中心扩散的修复过程。该方法可以较好地完成较大缺失区域的修复任务,并保证了修复结果的合理性。虽然视觉结构重建网络^[58]能够较好地解决大面积缺失的图像修复问题,但是边缘提取器在提取边缘信息的时候丢失了很多有用的图片信息,容易导致修复区域与原图存在较大的结构性差异。针对这一问题,Ren等^[58]提出了一种全新的两阶段网络,该网络由一个结构重构器和一个纹理生成器组成。为了恢复原图的结构信息,该模型在第1阶段使用原图中边缘平滑的图像来表示图像场景中的全局结构,维持边缘平滑^[59-60]的目的是去除高频纹理信息,并同时保留尖锐的边缘和低频结构信息。第2阶段,模型通过外观流^[61]技术来建立缺失区域与未缺失区域之间的联系,以此来生成更符合上下语境的纹理结构。

虽然全局编码器^[40]和平行自编码器^[44]能够较好地提取图像中的高级语义特征,但是这些方法却都忽略了高级语义与低级语义(纹理细节)间的结合过程,从而导致模型在结构连贯性和纹理一致性上表现欠佳。基于这一点,Zeng等^[62]在2019年提出了金字塔上下文编码器网络(PEN-Net),该网络以语义分割网络^[50]结构为基础,主要分为3个模块,分别为金字塔上下文编码器、多尺度译码器和基于对抗损失函数的训练模块。首先系统通过金字塔上下文编码器提取出潜在特征,并通过注意力转移网络(ATN)优化编码效果;然后通过跳跃连接融合重建特征和潜在特征,并传输给解码器进行解码;最后系统再通过L1损失函数和对抗损失函数优化参数。虽然该网络能够得到结构完善、细节精细的修复结果,但是在高分辨率图片的修复效果上表现依然欠佳。2020年,Yi等^[63]针对高分辨率图像修

复问题,提出了上下文残差聚合网络(CRA),该网络通过高频残差连接模块使修复结果最大程度地保留了丰富而详细的纹理信息,从而实现高分辨率图片的处理。且由于网络的补全模块只在低分辨率图像上运行,内存成本和计算时间显著降低。

大面积缺失图片的修复工作一直是该领域的难点,除了运用边缘生成器的方法^[56-58]外,另一种思路是利用循环渐进式的网络对图片进行修复。Li等^[64]基于这一思想,提出了循环特征推理(RFR)网络,该网络主要由循环特征推理模块和信息连贯注意力(KCA)模块构成。该网络模拟的是人类解决问题的思想,具体而言,该模型首先会解决较简单的修复区域,即通过RFR模块反复推断缺失区域的边界信息,然后再使用先前的推断信息作为额外的信息输入来辅助解决较难修复的区域。RFR模块工作的方式由于逐步加强了对缺失区域中心的约束,所以可对修复结果起到逐渐优化的效果。为了能让RFR在特征图中捕获相隔较远的信息,进一步开发了KCA模块。KCA中的注意力机制分数也是按照循环累加的模式得出的,因此可以有效改善修复区域边界不连续的现象。

综上所述,基于自编码器的图片修复方法仍然是当前图片修复领域的主流框架,虽然该方法相较于传统方法可以在修复区域和修复结果上取得较大的突破,但大部分基于自编码器的图片修复方法是端到端的学习模式,导致这类方法的解释性很差。此外,随着模型的加深和参数数量的增加,计算成本也是该方法需要优化的另一个问题。

3.3 基于生成对抗网络的方法

GAN由Goodfellow等^[6]在2014年提出,其网络结构如图5所示。

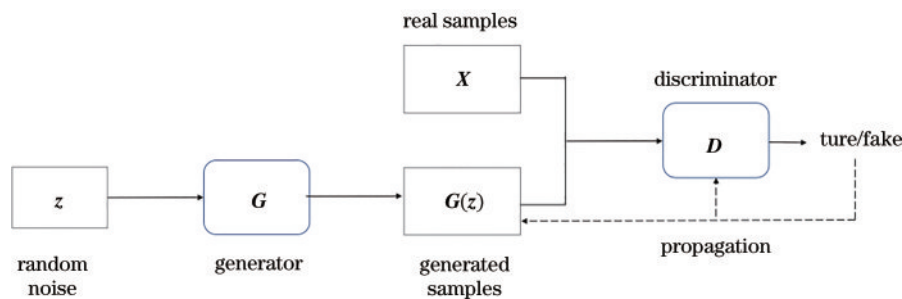


图5 GAN的结构图

Fig. 5 Structure diagram of GAN

该结构主要分为两个部分:一个为判别器(D),用于区分真实样本与生成样本;另一个为生成器(G),用于生成假样本来欺骗判别器。在训练过程中,生成器会生成与原图近似的假图片,而判别器则会将假图片辨别出来,这是一个博弈的过程,所以称之为对抗网络。假设真实样本分布为 $z \sim p_r$,生成器样本分布为 $z \sim p_g$,用于优化GAN的损失函数为

$$\min_G \max_D \left\{ E_{z \sim p_r} \log [D(z)] + E_{z \sim p_g} \log \{1 - D[G(z)]\} \right\}, \quad (1)$$

式中: E 表示分布函数的期望值。

基于生成对抗网络的图片修复技术的主要思想是通过预训练模型学习原图的先验分布,再通过鉴别器和生成器的博弈过程来优化修复结果。2015年,受到Goodfellow等^[6]的启发,Radford等^[65]提出了CNN与无

监督学习相结合的深度卷积生成对抗网络(DCGANs),该网络将学习到的特征运用在图片的生成中。在此基础上,Isola等^[55]在2017年提出一种名为PatchGAN的模型,该模型将条件对抗生成网络(CGANs)应用于图像修复领域。该网络的生成器以语义分割网络^[50]为基础,加入了跳跃式连接来保留更多的图片信息,从而使生成的补丁细节更加充沛。此外,该网络还设计了名为PatchGAN的鉴别器,该鉴别器通过关注图片的高频结构信息来判别补丁的真伪,并运用L1损失函数来优化修复结果的低频结构信息。由于该鉴别器可直接用于鉴别生成的补丁而不是鉴别整张图片,因此极大提升了模型的运行速度并节省了运算成本。

针对大面积缺失图片的修复问题,Yeh等^[66]在2017年提出了深度生成模型(DGMs),该模型对含有丰富语义信息的受损图片有着良好的修复效果,且可以对任意不规则的大面积缺失区域进行修复。该模型以受损区域周围的特征为参考,在这些特征所组成的潜在空间中利用上下文损失函数和先验损失函数来搜索与受损区域最相似的编码信息,再利用DCGAN模型^[65]作为图片生成器对受损区域进行补充,其在补丁生成过程中利用先验损失函数纠正不真实的补丁。由于不存在最好的编码信息,修复结果在一定程度上出现了上下图片未对齐的现象,再加上GANs训练至最优状态较为困难,该模型在达到最优状态之前捕捉数据分布的能力较差,因此该模型的另一个缺陷是无法对密集的数据分布即高分辨率图像进行修复。

尽管DCGAN模型^[65]能够生成具有清晰边界且符合上下文语义的补丁,但它很难将生成器与判别器训练至同时收敛的状态,会导致模型失效及梯度消失问题。针对这一问题,Arjovsky等^[67]在2017年提出一种名为Wasserstein GAN(WGAN)的全新网络结构,其主要创新在于提出了基于测地线(EM)距离的损失函数。EM距离是从真实值概率分布转移至生成器概率分布的最优路径,解决了两个概率分布间差异非对称性度量问题,即当两个分布不重叠或者距离很远时,GAN都无法进行梯度训练。因此,WGAN不需要保持生成器与鉴别器的收敛平衡,就能够达到最优的修复效果。此外,WGAN最引人注目的优势在于其可以通过已经训练到最优状态的鉴别器来持续地估计EM距离,而EM距离与观察到的样本质量密切相关,所以绘制EM距离的学习曲线对于调试超参数和超参数搜索都有着关键的作用。基于WGAN的思想,Lou等^[68]在2018年以遥感图片的修复问题为导向,提出了名为Modified GAN的网络,该网络在WGAN的基础上加入了RMSProp优化算法,使得该模型在非凸问题上也能保持良好的性能。但由于空间异质性的存在,该模型无法对场景复杂的遥感图片进行修复。

局部卷积的操作似乎可以解决自编码器中普通卷积将所有像素都视为有效像素的问题,但是在局部卷

积中无效像素会逐层消失,即掩膜信息完全依赖于最后一层的特征图,而在局部卷积模型中的最后一层特征图里,自编码器却未对每个像素点的位置信息加以限定,从而导致模型无法确定该像素点是在受损区域内还是在受损区域外,进而影响修复效果。针对这一问题,Yu等^[69]在2019年提出了一种基于门控卷积的GAN网络(SN-PatchGAN),该网络是一个三阶段网络,具体而言:其中一阶段网络由加入门控卷积的自编码器构成,目的是生成粗糙的修复结果;二阶段网络是由加入全文注意力层和门控卷积的双通道自编码器组成,以提高修复结果的精细化程度;三阶段网络由其首次提出的光谱归一化马尔可夫鉴别器(SN-PatchGAN)构成。门控卷积可以提供一个动态特征的选取机制,可从输入数据学习到相应像素点的门控值和位置信息,从而有目的地对图片进行特征提取,以减少无效像素的干扰。虽然该模型在修复效果上取得了显著的突破,但是由于门控卷积不能有效地表示有效特征之间的相关性,当采样区域混入无效信息时可能会导致修复结果产生误差。针对这一问题,Wang等^[70]在2021年提出了动态选择网络(DSNet),该网络的创新之处在于提出了两个全新的动态选择模块,分别是有效性迁移卷积(VMC)模块和区域复合归一化(RCN)模块。与固定网格的卷积相比,VMC模块可以根据图像中有效信息的分布实时调整卷积形状,因此可以更有效地避免无效信息。而RCN部分相较于以往单调归一化的方式,会根据数据分布对RCN模块所包含的3种归一化方法进行动态组合,从而对数据实行最有效的归一化操作。该网络充分利用了图片中的有效信息,可使梯度下降得更加准确,修复结果更加清晰。

变分自编码器(VAE)是一种类似于传统编码器的深度生成模型,由Kingma等^[71]在2013年提出。该模型与基于数字计算的传统自编码器的不同之处在于其通过学习数据的概率分布来实现对图像的预测。但是通过真实样本的先验分布来观察潜在空间会限制其生成结果的多样性,针对这一点,Zheng等^[72]在2019年提出了一种可以产生多种修复结果的GAN(PICNET),该网络的创新之处在于将VAE与GAN进行结合,其主要思想是利用两条路径对图片进行修复:其中第1条路径通过传统的VAE根据缺失区域推导出先验分布,称为重建路径;第2条路径利用来自上一条路径的先验分布和VEA进行耦合,再进一步对图片进行修复,称为生成路径。此外,该网络加入了一种全新的自注意机制层,能够利用短期和长期的上下文信息来确保图像域的外观一致性,在某种程度上该模型的修复效果优于只使用GAN的模型。受输出多样性的启发,Zhao等^[73]在2020年提出了无监督跨空间翻译生成对抗网络(UCTGAN),该网络主要由条件编码器模块、流形投影模块和生成模块等组成。通过结合流形

投影模块和生成模块,可以将实例图像空间和条件补全图像空间投影到普通的低维流形空间中,并以一种无监督的方式学习两个空间的图像映射关系,进而大大提高了修复样本的多样性。此外,为了更好地理解全局信息,该网络还引入了一种新的交叉语义注意力机制层,利用已知区域和已修复区域之间的长期依赖关系,提高修复结果的真实感和外观一致性。

GAN以其出色的多样性修复结果成为目前热门课题,但是由于存在模型失效和梯度消失的情况,其训练过程需投入大量的人力物力来对其进行维护。此

外,由于难以训练至最优状态,当数据量过大时,GAN将无法精确捕捉数据分布的情况,这使其仅局限于处理低分辨率的图片。

综上所述,基于深度学习的图像修复方法汇总如表 2 所示,相较于传统的图片修复方法,基于深度学习的图片修复方法在图片的特征学习上有着显著的突破,并且在大面积缺失修复任务上及捕捉高级语义的任务上表现都很出色,能够有效避免结构断层和模糊的现象。但是基于深度学习的图像修复方法大都是以端到端的模式搭建的,可解释性较差。

表 2 基于深度学习的修复方法汇总

Table 2 Summary of image inpainting methods based on deep learning

Category	Method	Year/Source	Contribution
Pixel-generation-based techniques	Richard et al ^[35]	2001/VIIP	Fast image restoration method based on diffusion convolution kernel (Gaussian)
	Hadhoud et al ^[36]	2008/SIP	The position of zero weight value of diffusion convolution kernel ^[35]
	Jain et al ^[37]	2008/NIPS	A neural network structure for denoising
	Xie et al ^[39]	2012/NIPS	Stacked sparse denoising Auto-encoders
	Pathak et al ^[40]	2016/CVPR	Context encoder to capture more semantic information
	Iizuka et al ^[41]	2017/ACM	Global and local context discriminators added to the auto-encoder
	Yu et al ^[44]	2018/CVPR	A parallel encoder model based on attention mechanism
	Sagong et al ^[45]	2019/CVPR	A shared encoding network with two parallel decoding tasks
	Shin et al ^[46]	2020/NNLS	Adaptive dilated convolutional layers added to PEPSI ^[45] model
	Yang et al ^[47]	2017/CVPR	Multi-scale neural patch synthesis approach
Auto-encoder-based techniques	Yan et al ^[49]	2018/ECCV	A special shift-connection layer Shift-Net
	Liu et al ^[51]	2018/ECCV	A partial convolution structure based on U-Net structure
	Xie et al ^[53]	2019/ICCV	A learnable bidirectional attention module which can automatically update the mask
	Liu et al ^[54]	2019/ICCV	A network architecture based on coherent attention mechanism layer
	Nazeri et al ^[56]	2019/Arxiv	A two-stage adversarial model EdgeConnect
	Li et al ^[57]	2019/ICCV	A progressive reconstruction of visual structure network
	Ren et al ^[58]	2019/ICCV	A novel two-stage network which can generate texture structures consistent with context semantics
	Zeng et al ^[62]	2019/CVPR	A pyramid context encoder network combining high-level semantics and texture information
	Yi et al ^[63]	2020/CVPR	A context residual aggregation network for high resolution image inpainting
	Li et al ^[64]	2020/CVPR	A cyclic feature inference network for recovering the large missing regions of damaged images
GAN-based techniques	Radford et al ^[65]	2015/Arxiv	DCGANs combining Convolutional Neural Network (CNN) and unsupervised learning
	Isola et al ^[55]	2017/CVPR	PatchGAN based on patch identification
	Yeh et al ^[66]	2017/CVPR	DGMs to repair irregular regions and capture richer semantics
	Lou et al ^[68]	2018/PRRS	RMSProp optimization algorithm is added to WGAN to maintain good performance on non-convex problems
	Yu et al ^[69]	2019/ICCV	SN-Patch GAN network based on gated convolution to improve the details and semantic accuracy of repaired results
	Wang et al ^[70]	2021/IEEE	The validity transfer convolution and region compound normalization modules to realize the dynamic selection of valid information
	Zheng et al ^[72]	2019/CVPR	PIC NET to generate a variety of repair results
Zhao et al ^[73]	2020/CVPR	An unsupervised cross-space translation generative adversarial network	

4 图像修复的相关数据集

人工智能领域的 3 个重要组成部分是算力、算法和数据。一个模型想要获得好的性能,就必须充分学习图像与标签之间的关系,因此图像数据集已然成为深度学习领域中不可或缺的一部分。对于数据驱动的深度学习,数据集的选择是成功训练模型的关键所在,

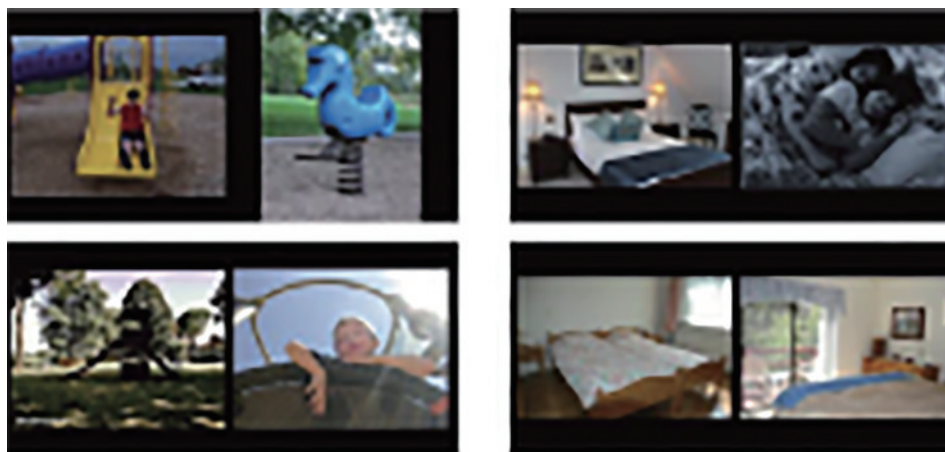


图 6 Place2 数据集^[74]

Fig. 6 Place2 dataset^[74]

2) Paris Street View 数据集^[75]包含 14900 张训练图像和 100 张测试图像,该数据集从 12 个城市中搜集了不

下面围绕一些主流数据集进行了讨论。

1) Place2 数据集^[74]总共包含 1000 多万张图像,有 400 多个独特的场景类别,拥有 5000~30000 张与真实场景一致的训练图像,该数据集主要应用于高级视觉理解任务,例如使用 CNN 学习深度场景特征的任务等,其包含的图片如图 6 所示。

同建筑的图片,以巴黎街景和巴黎郊区的图片居多,图像分辨率为 963×537 ,该数据集图片示例如图 7 所示。



图 7 Paris Street View 数据集^[75]

Fig. 7 Paris Street View dataset^[75]

3) CelebA 数据集^[76]是一个人脸数据集,该数据集收集了 202599 张名人的面部图像,其中包含 10177 个身份,5 个标注位置,每个位置都有 40 个二进制属性注释,截至 2015 年该数据集的分辨率被裁剪为 178×218 。

4) CelebA-HQ 数据集^[77]是从 CelebA 中衍生出来的 3 万张高分辨率的人脸图像,其分辨率最高可为 1024×1024 ,其图片示例如图 8 所示。

5) ImageNet 数据集^[78]包含 1400 万张图片,并且涵盖了 2 万个生活场景类别,是深度学习图像处理任务中应用最为广泛的数据集。

6) Microsoft COCO val2014 数据集^[79]是一个适用于分类和语义分割的数据集,包含 328000 张图片,这些图片总共带有 250 万个标签。

7) Nvidia Mask 数据集^[51]是一个不规则掩膜数据集,该数据集包含 55116 张训练掩膜图片和 24866 张测试掩膜图片,分辨率为 512×512 ,该数据集图片示例如图 9 所示。

8) Quick Draw Irregular Mask 数据集^[80]是一个不规则掩膜数据集,该数据集包含 50000 张训练掩膜图片和 10000 张测试掩膜图片,分辨率为 512×512 ,该数据集图片示例如图 10 所示。



图 8 CelebA-HQ 数据集^[77]
Fig. 8 CelebA-HQ dataset^[77]



图 9 Nvidia Mask 数据集^[51]
Fig. 9 Nvidia Mask dataset^[51]

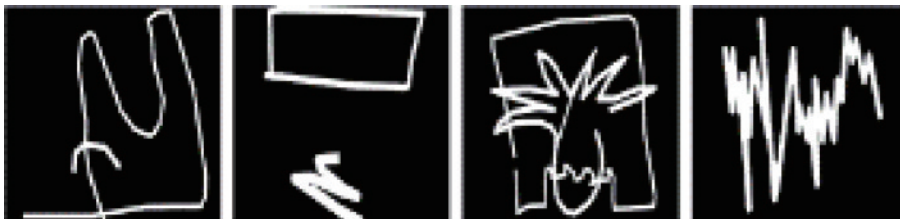


图 10 Quick Draw Irregular Mask 数据集^[80]
Fig. 10 Quick Draw Irregular Mask dataset^[80]

5 评价方法与各算法指标对比

图像修复算法性能的评判一般将修复后的图片与原图进行对比,而这个对比的过程往往需要借助不同的评判指标进行量化界定,下面对几种主流的评判指标进行说明。

1) 平均绝对误差(MAE),该指标衡量的是预测值与真实值之间的误差。MAE 越大误差越大,MAE 越小该模型预测性能较好。其计算公式为

$$E_{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \quad (2)$$

式中: \hat{y}_i 为预测值; y_i 为真实值。

2) 均方误差(MSE)。其与平均绝对误差不同的是绝对值运算换成了平方运算:

$$E_{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2. \quad (3)$$

3) 峰值信噪比(PSNR),该指标评价的是噪声影响修复结果的程度。PSNR 越大,修复结果越接近原始图片,修复效果越好。该指标是当前主流评价指标之一,其计算公式为

$$R_{PSNR} = 20 \log_{10} \frac{(V_{MAX_i})^2}{\sqrt{E_{MSE}}}, \quad (4)$$

式中: V_{MAX_i} 表示图像像素值的最大值; E_{MSE} 表示均方误差。

4) 结构相似性(SSIM),该指标衡量的是两张图片结构相似度。SSIM 值越大,修复效果越好,失真越小。SSIM 通过亮度、对比度和结构等 3 个方面来对两幅图像的相似性进行评估。亮度用均值表征,对比度用标准差表征,结构用协方差表征,其计算公式为

$$L(\mathbf{X}, \mathbf{Y}) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1}, \quad (5)$$

$$C(\mathbf{X}, \mathbf{Y}) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2}, \quad (6)$$

$$S(\mathbf{X}, \mathbf{Y}) = \frac{2\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3}, \quad (7)$$

$$SSIM(\mathbf{X}, \mathbf{Y}) = L(\mathbf{X}, \mathbf{Y}) \times C(\mathbf{X}, \mathbf{Y}) \times S(\mathbf{X}, \mathbf{Y}), \quad (8)$$

式中： μ_X 和 μ_Y 为图像 \mathbf{X} 和 \mathbf{Y} 的像素均值； σ_X 和 σ_Y 表示图像 \mathbf{X} 和 \mathbf{Y} 的像素标准差； σ_{XY} 表示图像 \mathbf{X} 和 \mathbf{Y} 的协方差； C_1 、 C_2 和 C_3 是常数。

5) Fréchet Inception Distance (FID) 是评价生成

图片质量和多样性的指标，一般用于衡量 GAN 优劣。该指标首先利用 Inception 网络提取图像特征，再利用均值和协方差矩阵两个物理量来计算真实数据分布和生成数据分布之间的距离大小。FID 越小，则两个分布越近，多样性也较优。其具体计算公式为

$$FID(\mathbf{X}, \mathbf{Y}) = \|\mu_X - \mu_Y\|_2^2 + T_r \left[\Sigma_X + \Sigma_Y - 2(\Sigma_X \Sigma_Y)^{\frac{1}{2}} \right], \quad (9)$$

式中： T_r 表示矩阵的迹； Σ 表示协方差。

结合上述评价指标与数据集，对前文所提及的部分网络进行了性能评定，具体如表 3 所示。

表 3 不同算法在常用数据集上的性能对比

Table 3 Quantitative evaluation results of algorithms on common datasets

Dataset	Method	PSNR↑	SSIM↑	FID↓	MAE↓	MSE↓	Size of image	Mask type (image-to-mask ratio)
CelebA-HQ	Sagong et al ^[45]	25.60	0.90				256×256	Square (25%)
		28.60	0.92				256×256	Irregular
	Shin et al ^[46]	25.50	0.89				256×256	Square (25%)
		28.50	0.92				256×256	Irregular
	Liu et al ^[54]	34.69	0.98		0.72	0.04	256×256	Irregular (10%–20%)
		32.58	0.98		0.94	0.07	256×256	Irregular (20%–30%)
		25.32	0.92		2.18	0.37	256×256	Irregular (30%–40%)
		24.14	0.88		2.85	0.44	256×256	Irregular (40%–50%)
		26.54	0.93		1.83	0.27	256×256	Square (25%)
		33.56	0.98		0.007		256×256	Irregular (10%–20%)
	Li et al ^[57]	27.76	0.93		0.02		256×256	Irregular (30%–40%)
		22.88	0.81		0.047		256×256	Irregular (50%–60%)
Zhao et al ^[73]	26.38	0.88		1.51		256×256	Irregular (10%–20%)	
Place2	Yu et al ^[44]	18.91			8.60	2.10	256×256	Irregular (10%–20%)
		33.75	0.94		0.49		256×256	Irregular (1%–10%)
		27.71	0.86		1.18		256×256	Irregular (10%–20%)
	Liu et al ^[51]	24.54	0.77		2.07		256×256	Irregular (20%–30%)
		22.01	0.68		3.19		256×256	Irregular (30%–40%)
		20.34	0.53		4.37		256×256	Irregular (40%–50%)
	Xie et al ^[39]	18.21	0.46		6.45		256×256	Irregular (50%–60%)
		25.59	0.78		1.93		256×256	Irregular (20%–30%)
		27.75	0.93		0.01		256×256	Square (25%)
	Nazeri et al ^[56]	21.75	0.82	8.16	3.86		256×256	Irregular (25%)
		24.92	0.86	4.91	2.59		256×256	Irregular (20%–30%)
		27.75	0.93		0.014		256×256	Irregular (10%–20%)
Li et al ^[64]	22.63	0.81		0.038		256×256	Irregular (30%–40%)	
	18.92	0.59		0.076		256×256	Irregular (50%–60%)	
Ren et al ^[58]	25.22	0.90	7.03			256×256	Irregular (20%–40%)	
Zeng et al ^[62]		0.78	15.19	9.94		256×256	Square (25%)	
Yi et al ^[63]				4.89	5.43		512×512	Irregular
				4.89	5.43		1024×1024	Irregular
				4.89	5.49		2048×2048	Irregular
				4.89	5.50		4096×4096	Irregular

续表

Dataset	Method	PSNR \uparrow	SSIM \uparrow	FID \downarrow	MAE \downarrow	MSE \downarrow	Size of image	Mask type (image-to-mask ratio)
Paris Street View	Pathak et al ^[40]	17.59			0.10	0.23	128 \times 128	Square (25%)
	Yang et al ^[47]	18.00				0.23	128 \times 128	Square (25%)
	Yan et al ^[49]	26.51	0.90			0.02	256 \times 256	Irregular (10%-20%)
		31.71	0.95		0.011		256 \times 256	Irregular (10%-20%)
	Li et al ^[57]	26.44	0.86		0.027		256 \times 256	Irregular (30%-40%)
		22.40	0.68		0.054		256 \times 256	Irregular (50%-60%)

6 结 论

虽然图像修复这个话题已经存在了几个世纪,但是随着大数据时代的到来,图像修复依旧是当今学术界的热门话题。从算法的角度回顾了一系列的模型结构,并从性能评估及局限性方面对不同算法进行了分析。从传统技术到深度学习的跨越,图像修复技术取得了长足的进步。相较于传统技术,基于深度学习的图像修复技术能够更好地捕捉图像的特征,并产生更高级的语义信息,可使修复结果更加连贯自然。此外,基于深度学习的方法还能够通过拟合概率分布完成大面积缺失图片的修复任务,并可以产生多样性的修复结果。虽然目前基于深度学习的图像修复技术取得了巨大的成功,但依旧存在可解释性较差及计算成本较高等一系列问题。在对现有图像修复方法进行总结的基础上,针对该领域所存在的问题和难点,做出了以下展望:

1)数据集。为了能够跟踪领域内最先进的算法进展,建立一个标准化的训练集和测试集是关键所在,只有制定统一的标准,才能更好地对不同算法进行定量评估。

2)算法。图像修复技术是面向实际应用的技术,因此算法是否简洁、轻便,是算法能否成功落地的关键。因此,网络结构简化和算法蒸馏等将是未来研究的热点。

3)模型可解释性。目前大部分的算法结构是以一种端到端的模式呈现的,深度学习的黑盒模式使得其可解释性较差。因此,将模型可视化并探究模型内部工作原理依然是当前研究的热点之一。

4)安全性问题。由于图像的特征图是一种面向机器的数据结构,当模型受到人眼无法识别的对抗样本的攻击时,模型将产生错误的判断,进而影响最后的输出结果。因此,建立一个安全稳定的模型将是未来研究的趋势之一。

5)分辨率的提高。随着硬件设备的不断迭代更新,图片的分辨率也越来越高,因此算法模型不应只局限于处理低分辨率的图片,实现高分辨率的细节修复将是未来研究的另一个热门方向。

参 考 文 献

- [1] 赵露露,沈玲,洪日昌. 图像修复研究进展综述[J]. 计算机科学, 2021, 48(3): 14-26.
Zhao L L, Shen L, Hong R C. Survey on image inpainting research progress[J]. Computer Science, 2021, 48(3): 14-26.
- [2] Elharrouss O, Almaadeed N, Al-Maadeed S, et al. Image inpainting: a review[J]. Neural Processing Letters, 2020, 51(2): 2007-2028.
- [3] Jam J, Kendrick C, Walker K, et al. A comprehensive review of past and present image inpainting methods[J]. Computer Vision and Image Understanding, 2021, 203: 103147.
- [4] Rojas D J B, Fernandes B J T, Fernandes S M M. A review on image inpainting techniques and datasets[C]// 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), November 7-10, 2020, Porto de Galinhas, Brazil. New York: IEEE Press, 2020: 240-247.
- [5] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4): 541-551.
- [6] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]// Advances in Neural Information Processing Systems 27, December 8-13, 2014, Montreal, Quebec, Canada. [S.l.: s.n.], 2014: 2672-2680.
- [7] Rumelhart D E, Hinton G E, Williams R J. Learning internal representations by error propagation[M]// Collins A, Smith E E. Readings in cognitive science. Amsterdam: Elsevier, 1988: 399-421.
- [8] Efros A A, Leung T K. Texture synthesis by non-parametric sampling[C]// Proceedings of the Seventh IEEE International Conference on Computer Vision, September 20-27, 1999, Kerkyra, Greece. New York: IEEE Press, 1999: 1033-1038.
- [9] Wei L Y, Levoy M. Fast texture synthesis using tree-structured vector quantization[C]// Proceedings of the 27th annual conference on Computer graphics and interactive techniques-SIGGRAPH '00, July 23-28, 2000, New Orleans, LA, USA. New York: ACM Press, 2000: 479-488.
- [10] Efros A A, Freeman W T. Image quilting for texture synthesis and transfer[C]// Proceedings of the 28th annual conference on Computer graphics and interactive

- techniques-SIGGRAPH '01, August 12-17, 2001, Los Angeles, California, USA. New York: ACM Press, 2001: 341-346.
- [11] Drori I, Cohen-Or D, Yeshurun H. Fragment-based image completion[J]. *ACM Transactions on Graphics*, 2003, 22(3): 303-312.
- [12] Hays J, Efros A A. Scene completion using millions of photographs[J]. *Communications of the ACM*, 2008, 51(10): 87-94.
- [13] Oliva A, Torralba A. Building the gist of a scene: the role of global image features in recognition[J]. *Progress in Brain Research*, 2006, 155: 23-36.
- [14] Ballester C, Bertalmio M, Caselles V, et al. Filling-in by joint interpolation of vector fields and gray levels[J]. *IEEE Transactions on Image Processing*, 2001, 10(8): 1200-1211.
- [15] He K M, Sun J. Image completion approaches using the statistics of similar patches[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(12): 2423-2435.
- [16] Bertalmio M, Vese L, Sapiro G, et al. Simultaneous structure and texture image inpainting[J]. *IEEE Transactions on Image Processing*, 2003, 12(8): 882-889.
- [17] Bertalmio M, Sapiro G, Caselles V, et al. Image inpainting[C]//*Proceedings of the 27th annual conference on Computer graphics and interactive techniques-SIGGRAPH '00*, July 23-28, 2000, New Orleans, California, USA. New York: ACM Press, 2000: 417-424.
- [18] Criminisi A, Perez P, Toyama K. Region filling and object removal by exemplar-based image inpainting[J]. *IEEE Transactions on Image Processing*, 2004, 13(9): 1200-1212.
- [19] Cheng W H, Hsieh C W, Lin S K, et al. Robust algorithm for exemplar-based image inpainting[C]//*Proceedings of International Conference on Computer Graphics, Imaging and Visualization*, July 26-29, 2005, Beijing, China. New York: IEEE Press, 2005: 64-69.
- [20] 陈永, 陈锦, 艾亚鹏, 等. 基于序贯相似性和布谷鸟寻优的敦煌壁画修复算法[J]. *激光与光电子学进展*, 2020, 57(20): 201020.
Chen Y, Chen J, Ai Y P, et al. Dunhuang mural inpainting algorithm based on sequential similarity detection and cuckoo optimization[J]. *Laser & Optoelectronics Progress*, 2020, 57(20): 201020.
- [21] 王永飞. 一种图像缺损修复算法分析[J]. *激光与光电子学进展*, 2020, 57(12): 121005.
Wang Y F. Analysis of an algorithm for repairing image defects[J]. *Laser & Optoelectronics Progress*, 2020, 57(12): 121005.
- [22] 陈永, 艾亚鹏, 陈锦. 基于信息熵和结构特性的敦煌壁画修复算法[J]. *激光与光电子学进展*, 2020, 57(12): 121020.
Chen Y, Ai Y P, Chen J. Dunhuang mural inpainting algorithm based on information entropy and structural characteristics[J]. *Laser & Optoelectronics Progress*, 2020, 57(12): 121020.
- [23] Simakov D, Caspi Y, Shechtman E, et al. Summarizing visual data using bidirectional similarity[C]//*2008 IEEE Conference on Computer Vision and Pattern Recognition*, June 23-28, 2008, Anchorage, AK, USA. New York: IEEE Press, 2008.
- [24] Barnes C, Shechtman E, Finkelstein A, et al. PatchMatch: a randomized correspondence algorithm for structural image editing[J]. *ACM Transactions on Graphics*, 2009, 28(3): 24.
- [25] Huang J B, Kang S B, Ahuja N, et al. Image completion using planar structure guidance[J]. *ACM Transactions on Graphics*, 2014, 33(4): 1-10.
- [26] Kwatra V, Essa I, Bobick A, et al. Texture optimization for example-based synthesis[C]//*ACM SIGGRAPH 2005 Papers on-SIGGRAPH '05*, July 31-August 4, 2005, Los Angeles, California. New York: ACM Press, 2005: 795-802.
- [27] Ružić T, Pizurica A. Context-aware patch-based image inpainting using Markov random field modeling[J]. *IEEE Transactions on Image Processing*, 2015, 24(1): 444-456.
- [28] Shen J H, Chan T F. Mathematical models for local nontexture inpaintings[J]. *SIAM Journal on Applied Mathematics*, 2002, 62(3): 1019-1043.
- [29] Telea A. An image inpainting technique based on the fast marching method[J]. *Journal of Graphics Tools*, 2004, 9(1): 23-34.
- [30] 邱锡鹏. 神经网络与深度学习[J]. *中文信息学报*, 2020, 34(7): 4.
Qiu X P. Neural networks and deep learning[J]. *Journal of Chinese Information Processing*, 2020, 34(7): 4.
- [31] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.
- [32] Koutnik J, Greff K, Gomez F, et al. A clockwork RNN [C]//*International Conference on Machine Learning*, July 6-9, 2014, Glasgow, Scotland, Britain. [S. l.: s. n.], 2014: 1863-1871.
- [33] van Oord A, Kalchbrenner N, Kavukcuoglu K. Pixel recurrent neural networks[C]//*International Conference on Machine Learning*, June 19-24, 2016, New York, NY, United States. [S. l.: s. n.], 2016: 1747-1756.
- [34] Hochreiter S, Schmidhuber J. Long short-term memory [J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [35] Richard M, Chang M Y-S. Fast digital image inpainting [C]//*Appeared in the Proceedings of the International Conference on Visualization, Imaging and Image Processing (VIIP 2001)*, September 3-5, 2001, Marbella, Spain. Anaheim: Acta Press, 2001: 106-107.
- [36] Hadhoud M M, Moustafa K A, Shenoda S Z. Digital images inpainting using modified convolution based method[J]. *Proceedings of SPIE*, 2009, 7340: 73400S.
- [37] Jain V, Seung S. Natural image denoising with convolutional networks[C]//*Advances in Neural Information Processing Systems*, December 8-11, 2008, Vancouver, British Columbia, Canada. Spain: Curran Associates, Inc., 2008: 769-776.
- [38] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313(5786): 504-507.

- [39] Xie J, Xu L, Chen E. Image denoising and inpainting with deep neural networks[C]//Advances in Neural Information Processing Systems, December 5-10, Lake Tahoe, Nevada, USA. Massachusetts: MIT Press, 2012: 341-349.
- [40] Pathak D, Krähenbühl P, Donahue J, et al. Context encoders: feature learning by inpainting[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2536-2544.
- [41] Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion[J]. ACM Transactions on Graphics, 2017, 36(4): 1-14.
- [42] Portilla J, Strela V, Wainwright M J, et al. Image denoising using scale mixtures of Gaussians in the wavelet domain[J]. IEEE Transactions on Image Processing, 2003, 12(11): 1338-1351.
- [43] Pérez P, Gangnet M, Blake A. Poisson image editing [C]//ACM SIGGRAPH 2003 Papers on-SIGGRAPH '03, July 27-31, 2003, San Diego, California. New York: ACM Press, 2003: 313-318.
- [44] Yu J H, Lin Z, Yang J M, et al. Generative image inpainting with contextual attention[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 5505-5514.
- [45] Sagong M C, Shin Y G, Kim S W, et al. PEPSI: fast image inpainting with parallel decoding network[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 11352-11360.
- [46] Shin Y G, Sagong M C, Yeo Y J, et al. PEPSI++ : fast and lightweight network for image inpainting[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 252-265.
- [47] Yang C, Lu X, Lin Z, et al. High-resolution image inpainting using multi-scale neural patch synthesis[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 4076-4084.
- [48] Gatys L, Ecker A, Bethge M. A neural algorithm of artistic style[J]. Journal of Vision, 2016, 16(12): 326.
- [49] Yan Z, Li X, Li M, et al. Shift-net: image inpainting via deep feature rearrangement[EB/OL]. (2018-01-29)[2021-05-04]. <https://arxiv.org/abs/1801.09392v1>.
- [50] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [51] Liu G L, Reda F A, Shih K J, et al. Image inpainting for irregular holes using partial convolutions[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11215: 89-105.
- [52] Harley A W, Derpanis K G, Kokkinos I. Segmentation-aware convolutional networks using local attention masks [C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 5048-5057.
- [53] Xie C H, Liu S H, Li C, et al. Image inpainting with learnable bidirectional attention maps[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 8857-8866.
- [54] Liu H Y, Jiang B, Xiao Y, et al. Coherent semantic attention for image inpainting[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 4169-4178.
- [55] Isola P, Zhu J Y, Zhou T H, et al. Image-to-image translation with conditional adversarial networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5967-5976.
- [56] Nazeri K, Ng E, Joseph T, et al. EdgeConnect: generative image inpainting with adversarial edge learning [EB/OL]. (2019-01-01)[2021-04-05]. <https://arxiv.org/abs/1901.00212>.
- [57] Li J Y, He F X, Zhang L F, et al. Progressive reconstruction of visual structure for image inpainting [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 5961-5970.
- [58] Ren Y R, Yu X M, Zhang R N, et al. StructureFlow: image inpainting via structure-aware appearance flow[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 181-190.
- [59] Xu L, Lu C, Xu Y, et al. Image smoothing via L 0 gradient minimization[C]//Proceedings of the 2011 SIGGRAPH Asia Conference, December 12-15, 2011, Hong Kong, China. New York: ACM Press, 2011: 1-12.
- [60] Xu L, Yan Q, Xia Y, et al. Structure extraction from texture via relative total variation[J]. ACM Transactions on Graphics (TOG), 2012, 31(6): 1-10.
- [61] Zhou T, Tulsiani S, Sun W, et al. View synthesis by appearance flow[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9908: 286-301.
- [62] Zeng Y H, Fu J L, Chao H Y, et al. Learning pyramid-context encoder network for high-quality image inpainting [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 1486-1494.
- [63] Yi Z L, Tang Q, Azizi S, et al. Contextual residual aggregation for ultra high-resolution image inpainting[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020:

- 7505-7514.
- [64] Li J Y, Wang N, Zhang L F, et al. Recurrent feature reasoning for image inpainting[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 7757-7765.
- [65] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. (2015-11-19) [2021-04-05]. <https://arxiv.org/abs/1511.06434>.
- [66] Yeh R A, Chen C, Lim T Y, et al. Semantic image inpainting with deep generative models[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6882-6890.
- [67] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks[C]//Proceedings of the 34th International Conference on Machine Learning, August 6-11, 2017, Sydney, NSW, Australia. New York: ACM Press, 2017: 214-223.
- [68] Lou S L, Fan Q C, Chen F, et al. Preliminary investigation on single remote sensing image inpainting through a modified GAN[C]//2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS), August 19-20, 2018, Beijing, China. New York: IEEE Press, 2018.
- [69] Yu J H, Lin Z, Yang J M, et al. Free-form image inpainting with gated convolution[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 4470-4479.
- [70] Wang N, Zhang Y P, Zhang L F. Dynamic selection network for image inpainting[J]. IEEE Transactions on Image Processing, 2021, 30: 1784-1798.
- [71] Kingma D P, Welling M. Auto-encoding variational bayes[EB/OL]. (2013-12-20)[2021-05-06]. <https://arxiv.org/abs/1312.6114>.
- [72] Zheng C X, Cham T J, Cai J F. Pluralistic image completion[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 1438-1447.
- [73] Zhao L, Mo Q H, Lin S H, et al. UCTGAN: diverse image inpainting based on unsupervised cross-space translation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 5740-5749.
- [74] Zhou B L, Lapedriza A, Khosla A, et al. Places: a 10 million image database for scene recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(6): 1452-1464.
- [75] Doersch C, Singh S, Gupta A, et al. What makes Paris look like Paris? [J]. ACM Transactions on Graphics, 2012, 31(4): 1-9.
- [76] Liu Z W, Luo P, Wang X G, et al. Deep learning face attributes in the wild[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 3730-3738.
- [77] Karras T, Aila T M, Laine S, et al. Progressive growing of GANs for improved quality, stability, and variation[EB/OL]. (2017-10-27) [2021-05-06]. <https://arxiv.org/abs/1710.10196>.
- [78] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [79] Lin T-Y, Maire M, Belongie S, et al. Microsoft coco: common objects in context[M]//Fleet D, Pajdla T, Schiele B, et al. Computer Vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8693: 740-755.
- [80] Isakov K. Semi-parametric image inpainting[EB/OL]. (2018-07-08)[2021-04-05]. <https://arxiv.org/abs/1807.02855>.