计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210100048

图像修复研究进展综述

赵露露[1]　沈　玲[2]　洪日昌[1]
1 合肥工业大学计算机与信息学院　合肥 230601
2 安徽大学互联网学院　合肥 230039
(luluzhao@mail.hfut.edu.cn)

摘　要　图像修复是计算机视觉领域中极具挑战性的研究课题。近年来,深度学习技术的发展推动了图像修复性能的显著提升,使得图像修复这一传统课题再次引起了学者们的广泛关注。文章致力于综述图像修复研究的关键技术。由于深度学习技术在解决"大面积缺失图像修复"问题时具有重要作用并带来了深远影响,文中在简要介绍传统图像修复方法的基础上,重点介绍了基于深度学习的修复模型,主要包括模型分类、优缺点对比、适用范围和在常用数据集上的性能对比等,最后对图像修复潜在的研究方向和发展动态进行了分析和展望。

关键词:图像修复;深度学习;卷积神经网络;生成对抗网络;自编码网络
中图法分类号　TP391

Survey on Image Inpainting Research Progress

ZHAO Lu-lu[1],SHEN Ling[2] and HONG Ri-chang[1]
1 School of Computer Science and Information Engineering,Hefei University of Technology,Hefei 230601,China
2 School of Internet,Anhui University,Hefei 230039,China

Abstract　Image inpainting is a challenging research topic in the field of computer vision. In recent years,the development of deep learning technology has promoted the significant improvement in the performance of image inpainting,which makes image inpainting a traditional subject attracting extensive attention from scholars once again. This paper is dedicated to review the key technologies of image inpainting research. Due to the important role and far-reaching impact of deep learning technology in solving "large-area missing image inpainting",this paper briefly introduces traditional image inpainting methods firstly,then focuses on inpainting models based on deep learning,mainly including model classification,comparison of advantages and disadvantages,scope of application and performance comparison on commonly used datasets,etc. Finally,the potential research directions and development trends of image inpainting are analyzed and prospected.

Keywords　Image inpainting,Deep learning,Convolutional neural network,Generative adversarial network,Auto encoder network

1　引言

图像修复(Image Inpainting)是一种根据图像已知内容去推测并修复出破损或缺失区域内容,使修复图像尽可能满足人类视觉感知需求的技术手段。这项工作最早起源于中世纪,修复师们凭借个人经验和丰富的想象手工复原破损的艺术品图像。随着计算机技术和图像处理技术的不断发展,数字图像修复已经成为计算机视觉(Computer Vision)和计算机图形学(Computer Graphics)领域的一项重要研究内容。该技术给人们带来了众多便利,被广泛应用于文化、生活、安防等领域,如数字文化遗产保护、图像编辑(如目标移除)、影视特效制作等,因此其一直是研究者们重点关注的问题之一。

传统的图像修复依据图像内容相似性和纹理一致性,采用基于数学和物理理论的方法,通过建立几何模型或采用纹理合成的方式来完成小区域破损图像修复。但由于计算机缺乏像人类一样的图像理解力和感知力,在大区域缺失的图像修复中,其结果往往存在内容模糊、语义缺失等问题。

近年来,以深度学习技术为代表的机器学习技术取得了质的飞跃,并在众多研究领域都取得了一系列卓越的成果。其中,卷积神经网络(Convolutional Neural Networks,CNN)作为一种前馈型的深度网络,在图像特征学习表达方面具有强大的能力,在大规模图像处理方面也具有出色表现[1-4]。另一方面,由 Goodfellow 等[5]提出的生成对抗网络(Generative Adversarial Networks,GANs),因具有巧妙的博弈对抗学习机制和拟合数据分布的巨大潜力,在计算机视觉领域也得到了广泛的应用。这些研究成果极大地弥补了图像视觉任务中传统方法在图像语义理解方面的不足,一定程度上解决了图像底层特征与高层语义之间的语义鸿沟,使得深度学习技术逐渐占领计算机视觉领域的前沿[6]。其中,基于深度学习的图像修复也掀起了一股研究热潮,并取得了令人瞩目的成果。

到稿日期:2021-01-06　返修日期:2021-01-26
基金项目:国家自然科学基金重点项目(61932009)
This work was supported by the Key Program of the National Natural Science Foundation of China(61932009).
通信作者:沈玲(sammiling315@gmail.com)

本文第 2 节简要介绍了传统图像修复方法，主要包括基于扩散的方法和基于样本的方法；第 3 节简要介绍了基于深度学习的图像修复方法中常用的深度学习的相关基础知识；第 4 节对近 5 年来具有代表性的基于深度学习的图像修复方法进行了分类和总结，包括基于自编码的图像修复方法、基于生成模型的图像修复方法和基于网络结构的图像修复方法；第 5 节介绍了深度图像修复方法中常用的图像数据集以及评价指标；第 6 节对第 4 节罗列的图像修复方法在常用数据集上的量化评测结果进行了对比分析；最后，对图像修复工作进行了总结，并对未来研究方向和发展动态进行了分析和展望。

2 传统图像修复方法

传统图像修复方法主要依据图像像素间的相关性和内容相似性来进行推测修复，依据修复最优化准则的不同，其在广义上可分为基于扩散的方法和基于样本的方法。

2.1 基于扩散的方法

基于扩散的方法主要用于修复小尺度图像破损，如照片中的小划痕。这类方法的主要思想是模仿艺术品修复师手工修补图像的过程，根据待修补区域的边缘信息，采用一种由粗到精的方法来估计等照度线的方向，并采用传播机制将已知信息传播到待修补区域来实现图像修复。这类方法主要包括基于偏微分(Partial Differential Equation，PDE)的修复技术和基于几何图像模型的变分修复技术。基于 PDE 的图像修复技术利用物理学中的热扩散方程，将待修复区域周围的信息传播到修复区域中，典型方法包括 Bertalmio 等[7]提出的用三阶偏微分方程来模拟平滑传输过程的 BSCB(Bertalmio-Sapiro-Caselles-Ballester)模型和 Chan 等[8]提出的利用曲率扩散强度的 CDD(Curvature-Driven-Diffusions)模型。基于变分法的图像修复工作通过建立图像的先验模型和数据模型，利用数学中的泛函和全变分知识，将图像修复问题转化为一个泛函求极值的变分问题。这类算法主要包括全变分(Total Variation，TV)模型[9]、Euler's elastica 模型[10]、Mumford-Shah 模型[11]、Mumford-Shah-Euler 模型[12]等。

上述偏微分方程与变分法是可以通过变分原理相互等价推导出的，因此基于扩散的方法又可称为基于变分 PDE 的图像修复算法。利用该类方法处理小尺度破损图像的效果较好，但当缺失区域较大或缺失区域及其周围纹理复杂时，修复效果趋于模糊，其原因主要有：1)该算法在变分空间对图像进行建模，且视图像为分段平滑的函数，能做到结构连续但不包含任何的纹理信息；2)该算法在本质上是由缺失区域边缘向内部扩散传播的过程，一旦待修复区域较大或纹理复杂就会失效[13]。

2.2 基于样本的方法

基于样本的方法假定图像缺失区域可以通过已知样本来表示，该方法对于大区域破损图像的修复可以取得较好的修复效果，不但可以填充任意大小的缺失区域，还可以修复破损部分的纹理细节。这类方法主要包括基于纹理合成的图像修复算法和基于数据驱动的图像修复算法。

基于纹理合成的图像修复算法的一种思路是将图像分解为结构部分和纹理部分，结构部分用基于变分 PDE 的方法来修复，纹理部分则用纹理合成的方法来填充[14-15]。另一种思路是设计一个匹配原则，全局搜索出与待修复区域相似度最高的图像块来填充缺失区域[16-22]。例如，Criminisi 等[19]提出了一种基于块的纹理合成算法，用于移除照片中的遮挡物。该算法主要通过重复以下 3 个步骤来实现：1)计算填充区域的优先权；2)传播纹理及结构信息；3)更新置信度值。相比之前的图像修复算法，Criminisi 算法[19]在视觉效果上有很大的提升，但随着填充过程的进行，置信度值迅速下降到零，优先权的计算不可靠，从而导致填充次序错误，进而影响了修复效果。为了弥补这一不足，Cheng 等[22]提出了一种更合理的优先权函数来改进 Criminisi 算法，并取得了更好的修复效果。

然而，基于纹理合成的方法大多采用一种全局搜索的方法来寻找相似块，当缺失区域为前景且纹理和结构复杂时，不但容易产生错误的匹配而且会耗费大量时间。因此，研究者们考虑利用大型的外部数据库或网络，以数据驱动的方式查询出相似的图像，用于填充缺失区域，这类方法称为基于数据驱动的图像修复方法[23-24]。其中，Hays 等[23]假定由上下文包围的区域可能具有相似的内容，然后在外部数据库中搜索与待修复图像具有足够视觉相似性的样本图像，用于填充缺失区域进而完成图像修复。但是，当外部数据库或网络中没有与待修复图像相似且合理的图像块时，这类方法会出现错误修复的情况且搜索过程的计算量较大。

传统图像修复方法在待修复图像缺失区域面积较小、结构及纹理较为简单的情况下，能取得很好的修复效果。但是，在面对较为复杂的图像修复任务时，由于缺乏对图像高层次语义的理解和感知，因此无法为缺失区域填充语义一致且合理的内容。鉴于此，越来越多的科研工作者将目光投向深度学习图像修复研究，尝试用基于深度学习的方法来获取更高质量的修复结果[25]。

3 深度学习的相关基础知识

3.1 自编码器网络

自编码器(Auto Encoder，AE)的概念最早由 Rumelhart 等[26]于 1986 年提出。自编码器是一个由输入层、隐含层和输出层构成的神经网络，属于无监督学习范畴，最初主要用于特征抽取和数据降维。自编码器的模型如图 1 所示，网络中输入数据 $\boldsymbol{X}$ 转化为编码向量 $\boldsymbol{X}'$ 的过程相当于是一个编码过程，目的是将输入压缩成潜在空间表征；通过编码向量 $\boldsymbol{X}'$ 重构原始输入 $\boldsymbol{X}$ 的过程可看作一个解码过程，目的是重构来自潜在空间表征的输入。

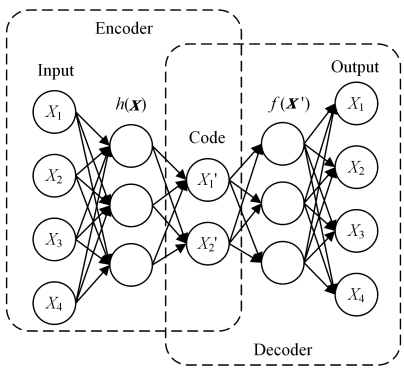


图 1 自编码器网络结构图

Fig. 1 Structure diagram of deep auto-encoder network

2006 年，Hinton 等[27]通过改进传统自编码器结构，提出

了深度自编码器(Deep Auto- Encoder,DAE)网络,通过训练多层神经网络来重建高维输入向量,将高维数据转换为低维代码,做到了有效的数据降维。Masci 等[28]于 2011 年提出的卷积自编码器(Convolutional Auto Encoder,CAE)使用卷积层和池化层替代了传统自编码器中的全连接层,利用卷积神经网络的卷积和池化操作实现无监督特征的提取。基于卷积的深度自编码网络的训练采用 Bengio 等[29]针对深度置信网络(Deep Belief Networks,DBN)提出的非监督贪心逐层训练算法。在深度网络的训练中,首先用非监督逐层贪心训练方法完成隐含层的预训练,然后使用反向传播(Back Propagation,BP)算法对整个神经网络进行系统性的参数优化调整。

3.2 变分自编码器

变分自编码器(Variational Auto-Encoder, VAE)是由 Kingma 等[30]于 2013 年提出的一种深度生成模型,其结构图如图 2 所示。

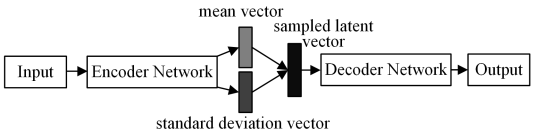


图 2 VAE 结构图

Fig. 2 Structure diagram of VAE

VAE 也是由编码器和解码器组成的,可以看作自编码器的一种变体结构,两者结构相似,但原理截然不同。与传统的自编码器通过数值的方式描述潜在空间不同,VAE 是以概率的方式描述对潜在空间的观察。

3.3 生成对抗网络

Goodfellow 等[5]于 2014 年提出的生成对抗网络(Generative Adversarial Networks,GANs)作为一种生成模型被广泛应用于计算机视觉领域,如图像生成、图像去噪、图像修复等。生成对抗网络的灵感来源于博弈论中零和博弈的思想,其网络结构如图 3 所示。

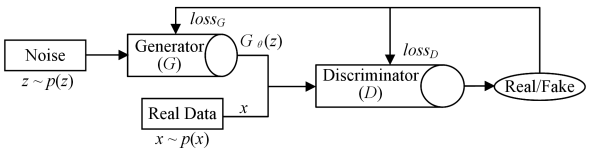


图 3 GAN 的结构图

Fig. 3 Structure diagram of GAN

该网络通常包括生成器 G(Generator)和判别器 D(Discriminator)两个部分。在训练过程中,生成器 G 的目标就是尽量生成真实的图片去欺骗判别器 D;而判别器 D 的目标就是尽量把生成器 G 生成的图片和真实的图片区分开来。

GAN 的目标函数如式(1)所示:

$$\min_G \max_D V(D,G)=E_{x\sim p_{data}(x)}[\log D(x)]+E_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{1}$$

其中,E 表示分布函数的期望值,$p_{data}(x)$代表真实样本的分布,$p_z(z)$表示定义在低维的噪声分布。将网络参数 θ_g 的 G 映射到高维的数据空间得到 $G(z,\theta_g)$。

4 基于深度学习的图像修复方法

随着深度学习技术的发展和硬件设备先进性的提高,近年来,研究者们将深度学习技术引入图像修复中,提出了众多基于深度学习的图像修复方法。根据模型结构的不同,基于深度学习的图像修复方法可以分为三大类:基于自编码的图像修复方法、基于生成模型的图像修复方法和基于网络结构的图像修复方法[31]。基于自编码的图像修复方法指以编码-解码结构为基本架构,从不同角度进行优化来提升修复效果的一类方法,这类方法也是目前深度学习图像修复领域的主流方法。按照优化角度的不同,基于自编码的图像修复方法又可细分为基于网络优化的方法、基于注意力机制的方法以及基于结构信息约束的方法。而基于生成模型的图像修复方法指基于图像生成网络,借助生成模型强大的图像生成能力来完成修复任务的一类方法。基于网络结构的图像修复方法则指利用深度生成网络自身可以拟合图像数据分布的能力来实现图像修复的方法。基于深度学习的图像修复方法的分类框架如图 4 所示。

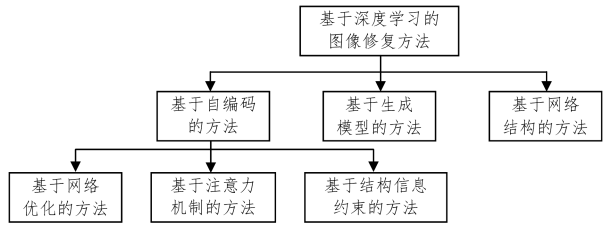


图 4 基于深度学习的图像修复方法分类框架图

Fig. 4 Classification frame diagram of image inpainting methods based on deep learning

4.1 基于自编码的图像修复方法

近年来,研究者们尝试将自编码网络中的编码-解码(Encoder-Decoder)结构应用到图像修复任务中,通过编码器捕获图像缺失区域周围的上下文信息并进行编码,提取图像的潜在特征表示,通过解码器来还原图像原始数据以生成缺失区域的内容,同时通过增加各种约束或借助 GANs 的生成对抗思想来不断优化修复结果,进而提出了大量基于自编码的图像修复方法。为了便于对这类方法进行总结,本文按照优化角度的不同又将其细分为基于网络优化的方法、基于注意力机制的方法以及基于结构信息约束的方法。其中,基于网络优化的方法指在编码-解码基本架构上,通过改进卷积算子或加入优化模块来提升图像修复效果;基于注意力机制的方法指在编码-解码架构上,通过引入注意力机制,在特征空间中寻找与未知区域最相似的特征匹配块来合成图像完整特征,从而提高图像细节修复的效果;而基于结构信息约束的方法则指先提取图像结构信息,再利用结构信息引导缺失区域的内容修复[31]。

部分文献中提出的方法可能包含多种优化方式,在进行分类时难以进行明确的区分。因此,本文依据每种修复算法重点优化的角度来对基于自编码的图像修复方法进行分类。

4.1.1 基于网络优化的方法

最初，Pathak 等[32]受自编码器的启发，将编码-解码网络结构引入深度学习图像修复工作中，提出了一种名为 Context Encoder 的网络，其网络结构如图 5 所示。具体来说，Context Encoder 结合了编码-解码网络结构和 GANs 的对抗思想，将图像重构损失和对抗损失作为约束来提高图像的修复效果。他们的实验结果表明，仅使用重构损失，即预测图像与实际图像像素值之间的欧氏距离并将其作为损失函数，修复效果较为模糊，对抗损失的加入使得生成的缺失区域的图像更加真实。但该文中增加的对抗损失约束仅仅作用于修复区域，忽略了图像全局结构的一致性和修复区域边界像素值的连续性，因此存在图像边界扭曲和局部模糊的问题。针对这一问题，Iizuka 等[33]在保留 Context Encoder 网络[32]中的判别器作为局部判别器的同时，又增加了作用于整张图像区域的全局判别器。作用于不同区域的两种判别器共同作用，使得修复后的图像既符合全局语义，又更具真实性。但是，该方法缺乏对图像纹理细节的考虑，不能对复杂的结构性纹理进行修复，同时还包含复杂的后处理过程。

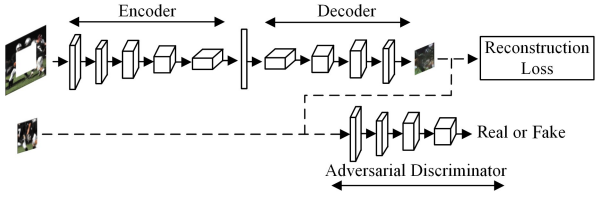


图 5 Context Encoder 的结构图[32]

Fig. 5 Structure diagram of Context Encoder[32]

在基于深度学习的图像修复工作中，卷积网络中的普通卷积(Vanilla Convolutions)会将图像中的所有输入像素都当作有效值进行计算，同时卷积图像中的有效像素和缺失像素导致修复后的图像出现视觉伪影，如色差、模糊以及边界扭曲等。特别是对于不规则破损图像的修复来说，这一现象更为严重。为此，Liu 等[34]用部分卷积(Partial Convolutions)代替普通卷积，提出了一种含有自动掩膜(mask)更新机制的修复模型来实现对不规则缺损图像的修复。其中，部分卷积将已知区域内的像素视为有效像素，将缺失区域内的像素视为无效像素，且只对有效像素进行卷积。此外，该模型没有采用一般的编码-解码结构，而是引入图像分割中的 U-Net 结构[35]，这种结构通过增加跃式连接(Skip Connections)的方式将编码器的每一层特征与解码器中相应层的特征连接起来。换句话说，就是将编码过程中的下采样层的特征复制传递到后边对称的解码过程中的上采样层，以补充编码过程中丢失的相关特征。该方法实现了将破损区域内外的像素区别对待，在不对输出图像进行任何后处理的情况下，也可以有效解决视觉伪影问题。但是，这种人为设定的掩膜更新机制仍然存在不合理性。例如，在掩膜更新过程中，部分卷积将含有不同数量的有效像素的区域同等对待，且无效像素将在深层逐渐消失。

为了解决上述问题，科研工作者尝试对文献[34]进行改进，设计更灵活的掩膜更新机制来提高网络对不规则破损图像的修复能力[36-37]。其中，Yu 等[36]提出了门控卷积(gated convolutions)的概念和可学习的掩膜更新机制。该方法在特征层的不同空间位置为每个通道建立可学习的动态特征选择机制，因此可以更有效地解决修复过程中的掩膜更新问题，从而提升修复效果。但该模型基于双阶段粗-细修复架构，且细修复网络含有两个分支，这种多步多分支的模型架构的参数量较大，需要耗费大量的计算资源。文献[37]则采用可学习的双向注意力图(Attention Map)模块来自动更新掩膜，在编码过程和解码过程中都对不规则掩膜进行了动态更新，因此修复后的图像内容更加清晰、连贯且合理。

考虑到在修复网络的过程中，不同尺寸的卷积核可以获取不同大小的感受野和提取不同尺度的图像特征，更有利于图像细节的修复，因此，Wang 等[38]提出了一种名为 GMCNN 的多列卷积生成网络，在编码阶段使用多分支并行 CNN 结构，各分支使用不同尺寸的卷积核将图像分解成具有不同感受野和特征分辨率的分量，以提取图像的多尺度特征。Xiao 等[39]在 U-Net 结构中添加了 inception 模块，组成了网中网(Network in Network)架构，通过多个卷积和池化操作获取图像多尺度特征表示，增强了模型对图像的语义认知理解，提高了图像细节的修复效果。文献[40]提出了一种多尺度融合网络，其中的多尺度融合模块由多个不同扩张率的扩张卷积(dilated convolutions)层分支组成。该网络在不增加参数的情况下扩大了感受野，增强了网络的特征提取能力，显著提高了图像的修复质量。

此外，研究者们还尝试从更多的角度去优化基于自编码的图像修复网络[41-44]。例如，Zhang 等[41]引入课程学习的思想，根据缺失区域的大小将图像修复划分为多个子任务，每个子任务之间用 LSTM 框架[42]连接，实现从缺失区域边界向中心逐层修复。相比文献[32]直接对图像的整个区域进行修复，该方法的修复效果有明显提升。Shen 等[43]将多组 U-Net 变体结构进行跳跃式密集连接，提出了一种新的端到端语义图像修复框架，用于挖掘图像中更多的语义信息。Hong 等[44]在 U-Net 网络解码层中添加特征融合块，使得修复结果已知区域与未知区域边界处的结构及纹理过渡得更加自然。

以上基于深度学习的图像修复方法大多采用特征归一化(Feature Normalization)进行训练，但在空间维度上进行归一化忽略了破损区域对归一化的影响，无法解决网络中常出现的均值和方差漂移问题。针对这一缺陷，文献[45]提出了一种新颖的区域归一化(Region Normalization)方法。该方法可以根据输入掩膜将空间像素划分为损坏区域和未损坏区域，分别计算每个区域的均值和方差来进行归一化。文献[45]还使用 EdgeConnect 网络[46]的图像生成器作为主干生成器，整个生成网络由编码网络、残差块(Residual Block)和解码网络 3 个部分组成。其中，编码网络中加入了 RN-B (Basic Region Normalization)模块，残差块和解码网络中加入了 RN-L (Learnable Region Normalization)模块。最终的实验结果表明，在修复网络中使用区域归一化方法可以有效解决均值和方差漂移问题，提高了图像的修复效果。RN-B 模块和 RN-L 模块作为两种即插即用的 RN 模块，可以灵活应用于各种图像修复网络。但是，RN 模块并不适用于网络中图像损坏区域与未损坏区域难以明确区分的修复网络，如门控卷积图像

修复网络[36]。因为门控卷积会极大地平滑图像特征，在修复过程中 RN 模块难以追踪潜在的破损区域，最终导致修复后的图像较为模糊。

以上图像修复方法都是在缺失区域的形态已知的前提下进行修复，而在实际生活中待修复图像的掩膜形态大多未知。因此，能自动修复图像内容而无须为图像中缺失区域指定掩膜的盲修复(Blind Inpainting)作为一种更实用的修复方式，受到了越来越多的科研工作者的关注。盲修复任务的关键是正确识别出图像中需要修复的区域，之前的盲修复工作[47-48]都是对缺失区域内填充内容为定值或高斯噪声的破损图像进行修复，因此大多数情况下可以很容易地正确识别出缺失区域，从而高效地完成修复任务。但当破损图像缺失区域填充的内容未知时，这些方法很难确定修复区域，甚至会出现非常严重的修复错误。针对这一缺陷，文献[49]提出了一种包含掩膜预测模块和鲁棒性修复模块的两阶段视觉一致性网络。该网络中的掩膜预测模块是根据语义的不一致性而不是缺失区域的填充内容来预测掩膜区域，掩膜预测模块与鲁棒性修复模块相互关联，预测模块的输出掩膜送入修复模块，在优化过程中修复模块的误差也会反向传播到掩膜预测模块，从而提高掩膜预测的准确性。因此，该网络对掩膜预测错误具有鲁棒性，有效解决了后续修复中出现的伪影问题，显著提高了模型的泛化能力。

基于网络优化的图像修复算法的模型结构特点如表 1 所列。这类方法大多采用端到端的单步优化模型，网络结构较为简单，取得了一定的修复效果。但是，该类方法在图像纹理和结构细节的修复方面仍存在不足。

表 1 基于网络优化的图像修复算法汇总

Table 1 Summary of image inpainting algorithms based on network optimization

算法名称	年份/来源	网络基本结构	模型结构创新点
CE[32]	2016/CVPR	编码-解码结构	全连接层连接编码-解码网络
G&L[33]	2017/TOG	编码-解码结构	全局&局部判别器
PConv[34]	2018/ECCV	U-Net 结构	部分卷积层，掩膜自动更新机制，不规则图像修复
Gated Conv[36]	2019/ICCV	编码-解码结构	双阶段粗-细修复网络，可学习动态特征选择机制
LBAM[37]	2019/ICCV	U-Net 结构	可学习特征再规则化，双向掩膜更新机制
GMCNN[38]	2018/NIPS	编码-解码结构	全局&局部判别器，多列卷积编码-解码网络，ID-MRF 规则项寻找相似块
DIGN[39]	2018/Arxiv	U-Net 结构	网中网模块
DMFN[40]	2020/Arxiv	编码-解码结构	全局&局部判别器，密集多尺度融合块增大感受野
GPN[41]	2018/ACM MM	U-Net 结构	引入课程学习，采用分层方式渐进式修复图像
SSDCGN[43]	2019/ACM MM	U-Net 结构	多组 U-Net 变体结构跳跃式密连接
DFNet[44]	2019/ACM MM	U-Net 结构	解码层中含融合模块可实现细节修复
RN[45]	2020/AAAI	编码-解码结构	区域归一化方法，两种即插即用的 RN 模块
VCNet[49]	2020/ECCV	编码-解码结构	掩膜预测模块和鲁棒性修复模块

4.1.2 基于注意力机制的方法

传统方法中，基于纹理合成的图像修复方法中的块匹配方法是在像素或图像块层面进行修复，缺少对图像语义和全局结构的理解。基于此，研究者们尝试将块匹配思想引入图像特征空间，在注意力机制的引导下为缺失区域寻找最相似的特征块进行特征匹配，最终提出了一系列基于注意力机制的图像修复方法，其中具有代表性的修复算法的模型结构特点如表 2 所列。

表 2 基于注意力机制的图像修复算法汇总

Table 2 Summary of image inpainting algorithms based on attention mechanism

算法名称	年份/来源	网络基本结构	模型结构创新点
HR[50]	2017/CVPR	编码-解码结构	双阶段粗-细修复网络，利用预训练好的 VGG 网络来提取特征匹配相似块
CII[51]	2018/ECCV	编码-解码结构&U-Net 结构	利用预训练好的 VGG 网络来提取特征，特征补丁交换
Shift-Net[52]	2018/ECCV	U-Net 结构	shift 连接层，编码层特征移入解码层
CA[53]	2018/CVPR	编码-解码结构	双阶段粗-细修复网络，全局&局部判别器，引入内容感知层
CSA[54]	2019/ICCV	U-Net 结构	双阶段粗-细修复网络，特征空间相似性匹配，像素间的连续性
PEPSI[55]	2019/CVPR	编码-解码结构	共享编码网络&并行解码网络，多步修复
PEPSI++[56]	2019/Arxiv	编码-解码结构	基于 PEPSI 的改进，速率自适应扩张卷积层
PEN-Net[57]	2019/CVPR	U-Net 结构	金字塔上下文编码网络&多尺度解码网络
MED[58]	2020/ECCV	U-Net 结构	结构分支&纹理分支，特征均衡化
RFR[59]	2020/CVPR	U-Net 结构	渐进式修复，循环特征推理
HRII[60]	2020/ECCV	编码-解码结构	双阶段粗-细修复网络，指导性上采样，高分辨率修复
CRA[61]	2020/CVPR	编码-解码结构	双阶段粗-细修复网络，利用上下文残差聚合机制实现超高分辨率图像修复

早期的基于注意力机制的图像修复方法主要是受到传统方法中块匹配思想的启发[50-51]。最初，Yang 等[50]提出了一种用于高分辨率图像修复的多尺度特征块合成方法，该方法首先采用 Context Encoder[32]作为全局图像内容预测网络，得到一个粗略的修复结果，再利用预训练好的 VGG-19 网络提取图像特征，在特定特征层上寻找与缺失区域最相似的特征块进行匹配，迭代完成纹理细节的修复。该方法有利于对高频纹理细节的修复，但数百次梯度下降迭代优化过程需要耗费大量的计算资源。Song 等[51]对此做出了改进，提出了一种基于学习的两阶段图像修复方法，第一阶段通过编码-解码网络得到粗略的修复结果，然后用预训练好的 VGG-19 网络提取特征，在固定特征层通过卷积运算来计算相似度以寻找最佳匹配块并进行块交换；第二阶段使用 U-Net 结构，根据特征块交换结果还原出修复图像。该方法中的块交换过程实现了将已知区域纹理信息传播到未知区域，且最终在修复时间上取得了比文献[50]更优的结果。不同于上述两种使用预训练好的 VGG 网络提取特征再进行块匹配的两阶段粗-细修复方法，Yan 等[52]直接在 U-Net 结构中引入了移位连接(shift-connection)层，提出了一种名为 Shift-Net 的单步端到端图像修复网络。该网络将图像编码层中图像已知区域的特征信息移入对应的解码层，以引导解码空间中缺失区域内部

特征的补全，进而提升修复效果。

此后，Yu 等[53]正式将注意力机制引入图像修复网络，提出了一种含有内容感知层的双阶段粗-细图像修复模型，以解决卷积神经网络无法从图像较远区域提取信息的问题。该模型的网络结构如图 6 所示。

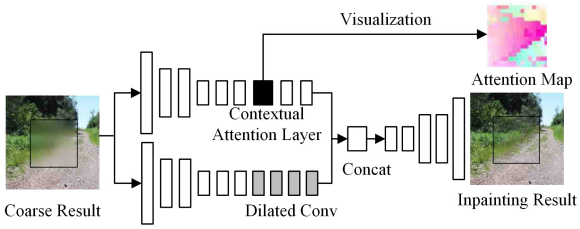


图 6 内容感知图像修复模型图[53]

Fig. 6 Model diagram of image inpainting with contextual attention[53]

该模型中基于注意力机制设计的内容感知层将提取的特征图分为前景区域(待修复区域)和背景区域(已知区域)，用卷积的方法从背景区域的特征空间中寻找与前景区域最相似的特征块，来实现远距离寻找特征匹配块的目的。但该方法忽略了图像缺失区域内部特征之间的相关性，导致修复后的图像常出现明显的断层现象。针对这一问题，Liu 等[54]提出的连贯性语义注意力机制不仅考虑了图像缺失区域和已知区域的相似性匹配，还重点关注了待修复区域深层特征之间的相关性，有效解决了文献[53]的修复结果中常出现的色彩断层及边界扭曲问题。Sagong 等[55]在文献[53]的基础上进行改进，提出了一种名为 PEPSI 的共享编码、并行解码的粗细修复网络，在注意力模块中使用欧氏距离替换文献[53]使用的余弦相似度，来计算已知区域与待修复区域特征块的相似匹配度，在保证修复效果的同时，大大缩短了修复时间。该方法的扩展版本 Diet-PEPSI[56]用速率自适应扩张卷积层代替原来网络中的一般扩张卷积层，进一步降低了资源损耗。

在深度卷积神经网络中，网络深层提取的特征主要包含高级结构语义信息，浅层提取的特征主要包含低级纹理细节。基于注意力机制，将低级纹理特征和高级结构语义特征有效结合，可以取得纹理一致、结构连续、内容连贯的修复效果[57-58]。例如，Zeng 等[57]提出了一个含有金字塔上下文编码网络和多尺度解码网络的图像修复模型。其中，金字塔上下文编码网络中的注意力转移网络，首先在高级特征图中计算缺失区域与已知区域的块相似度，然后将相似度得分转移到下一层用于指导低层特征图上的特征补全，以金字塔式的路径逐层补全特征图，直到补全到低级像素层。该模型相当于同时在图像层面以及特征层面进行修复，保证了纹理和语义的一致性。不同于大多数基于双阶段粗-细修复网络架构的修复方法，文献[58]提出了一种可以在特征层级上同时修复网络纹理和结构语义的交互编码-解码网络模型。该模型首先将编码器网络的特征分为浅层纹理特征和深层结构特征两个分支，分别通过多尺度修复模块进行孔填充，再利用基于双边注意力机制的特征均衡化将两种特征融合，最后将含有结构和纹理信息的特征解码还原成图像。模型中设计的双边注意力机制使得当前的特征点是由其周围以及全局特征点加权组成的，保证了图像局部信息和全局信息的一致性。

文献[41]提出的一种图像层面的渐进式修复方法取得了一定的修复效果，特别是对于缺失区域较大的图像。但该方法存在计算成本高、仅针对规则缺失区域进行修复等问题。因此，Li 等[59]提出了一种作用于特征层面的渐进式图像修复网络——循环特征推理网络。为了保证循环过程中不同特征图之间的一致性，该网络在特征推理过程中特别设计了一种知识一致注意力模块，该模块的当前循环过程中每个像素最终的注意力得分是由前面循环过程中的注意力得分和当前注意力得分加权获得。实验结果表明，这种作用于特征层面的渐进式修复方法可以取得更精细的修复效果。

由于计算机处理器内存和计算资源的限制，现有的深度学习图像修复方法只能处理较低分辨率的图像。考虑到低分辨率的图像可以通过上采样的方式提高分辨率，研究者们将高分辨率图像修复工作分为低分辨率修复和上采样两个过程，提出了基于注意力机制的高分辨率图像修复方法[60-61]。其中，Zeng 等[60]提出了一种包含反馈机制的迭代修复模型。该模型首先利用置信度预测机制来预测置信度图；然后将输出的置信度图作为反馈，使得模型在每一次迭代过程中仅保留缺失区域内的高置信度像素，通过这种方式不断缩小孔区域，得到一个低分辨率修复结果；最后利用基于注意力机制的指导性上采样网络计算低分辨率结果的特征块相似度，来指导高分辨率图像的重建。Yi 等[61]提出了一种上下文残差聚合机制，用于超高分辨率图像修复。该方法首先利用两阶段粗-细修复网络得到一个低分辨率修复结果；然后直接将低分辨率修复结果上采样得到一个模糊的图像；最后加入高频残差得到所需的高分辨率修复图像。该方法中，基于上下文注意力思想设计的上下文残差聚合机制利用高层特征图来计算注意力得分，然后在多个低层特征图中进行注意力转移，从而实现在多个抽象层上匹配上下文信息的目的，其中细修复过程和残差聚合过程共享注意力得分。基于注意力机制的高分辨率图像修复取得了很好的修复效果，但这些方法并不是直接对高分辨率图像进行修复，修复网络的输入、输出均是低分辨率图像。

上述基于注意力机制的修复方法表明，将注意力机制引入图像修复方法中来指导图像特征图的补全，可以有效地提升图像纹理细节和语义的修复效果。但是，大多数基于注意力机制的方法是基于双阶段粗-细修复架构的，这种网络结构的参数量较大，且第一阶段的粗修复结果对最终的修复效果的影响较大。

4.1.3 基于结构信息约束的方法

现有的图像修复方法大多由于缺少结构信息的约束，无法恢复缺失区域内精细的结构细节信息，从而导致修复后的图像出现过度平滑或模糊的区域。基于结构信息约束的方法尝试增加图像结构信息约束来引导图像内容修复，这类方法主要遵循艺术手工匠对缺损图像艺术品的修复方式，即先根据已知区域的信息对缺失区域的结构进行预测，再基于完整的结构信息和已知区域信息进行图像内容的填充。

最初，Song 等[62]尝试将图像分割理论引入图像修复，将

经过语义分割方法得到的语义标签作为结构信息来指导图像缺失区域内容的修复。由于分割图中包含了一些有用的语义信息，这一方法取得了一定的修复效果，但对具有相似语义标签但纹理不同的区域进行修复时，语义预测错误会导致图像纹理修复出现错误。Xiong 等[63]为了解决当图像缺失区域与前景区域重叠时，缺失区域难以修复的问题，提出了一种前景感知图像修复模型，该模型利用图像已知区域信息预测得到前景轮廓，将其作为结构先验进一步指导缺失区域内容的填充。Nazeri 等[46]则采用边界生成器预测得到的完整边界信息来引导内容修复，即使对一些高度结构化的场景图像进行修复，这一方法也具有良好的表现。该方法的修复网络结构如图 7 所示。

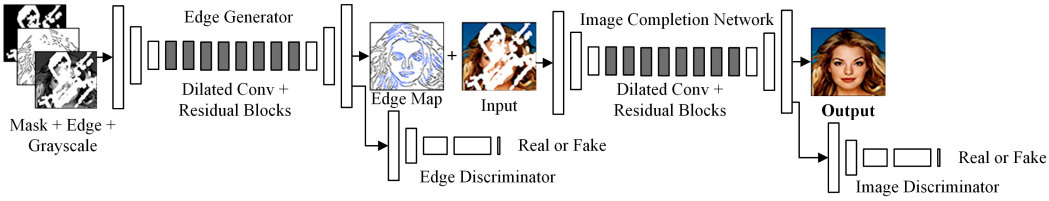


图 7 EC 网络结构图[46]

Fig. 7 Network structure diagram of EC[46]

以上方法证明了增加结构先验能有效提升图像的修复质量，但这种先结构后纹理的双阶段图像修复模型存在一定的局限性：1)网络中同时含有结构生成器和内容生成器，参数量较大；2)两类生成器采用串联结构，若前一步的结构预测不合理，后一步的内容修复工作则将受到很大影响，甚至出现大范围的修复错误。为此，Liao 等[64]在文献[62]的基础上做出了改进，提出了一种具有语义指导及评估机制的图像修复模型。该模型采用语义分割模块和图像修复模型相互作用的架构，迭代更新语义分割图和图像缺失区域的内容，在该过程中语义分割图作为结构先验不断指导图像内容修复，以一种由粗到精的方式逐步完成修复。该方法基于其对语义分割信息的可靠预测，在混合场景图像修复任务中取得了不错的效果。Li 等[65]在含有部分卷积层的 U-Net 结构中堆叠视觉结构重建层，以渐进式的方式重建缺失区域的结构和视觉特征。该网络中，每一个视觉结构重构层中更新的结构信息都用于引导特征内容的填充，逐步缩小缺失区域，最终完成修复任务。上述两种方法在提高结构先验预测的准确性的同时，确保了图像内容填充的合理性，为缺失区域生成了高质量的纹理和结构细节，而且本质上都是一种渐进式的修复方法，更有利于修复大区域的破损图像。

修复后的图像结构连续、纹理一致以及语义合理是图像修复任务的关键问题，也是其面临的巨大挑战。上述方法通过增加结构先验来引导内容修复，一定程度上提升了修复效果，但缺乏对纹理细节的考虑。针对这一缺陷，Ren 等[66]提出了一种先结构后纹理的两阶段图像修复模型。该模型在第一阶段生成边界保留的平滑图像[67-68]并将其作为全局结构先验，其中边界保留的平滑图像保留了图像的结构，但不含高频纹理，有利于全局结构的重建；第二阶段引入外表流(Appearance Flow)[69]方法为缺失区域合成高频纹理细节。文献[70]设计了一个含有结构嵌入机制的共享生成器模型，结构嵌入层用来连接图像生成分支和结构生成分支，并将结构信息合并到图像修复过程中。基于这种多任务学习架构，该模型可以同时修复图像内容和结构，且模型中的解码器可以决定是否利用结构先验，避免了错误预测对图像内容修复过程的不利影响。此外，该模型不仅将既含有边缘信息又含有纹理信息或高频细节的图像梯度图作为结构先验，还引入了注意力机制，有效提升了图像纹理细节的合成效果。

基于结构信息约束的图像修复方法的模型结构特点如表 3 所列。这类方法通过引入结构先验来指导图像内容修复，有利于修复图像中精细的结构细节，特别是对于一些高度结构化的图像，如人脸图像、建筑图像等。但与基于注意力机制的方法相似，基于结构信息约束的方法大多也采用双阶段的修复网络架构，即先预测结构信息，再进行内容修复，因此会存在模型参数量大以及因结构预测错误而对内容预测产生不利影响等问题。

表 3 基于结构信息约束的图像修复方法汇总

Table 3 Summary of image inpainting algorithms based on structural constraint

算法名称	年份/来源	网络基本结构	模型结构创新点
SPG-Net[62]	2018/BMVC	编码-解码结构	双阶段结构-内容模型，语义分割信息约束图像内容修复
FII[63]	2019/CVPR	编码-解码结构	双阶段结构-内容模型，前景轮廓约束图像内容修复
EC[46]	2019/ICCVW	编码-解码结构	双阶段结构-内容模型，边界信息约束图像内容修复
SGE-Net[64]	2020/ECCV	编码-解码结构	改进 SPG-Net，渐进式修复，更可靠语义分割信息约束图像内容修复
PRVS[65]	2019/ICCV	U-Net 结构	渐进式修复，加入视觉结构重建层，结构信息引导特征层内容修复
Stru-Flow[66]	2019/ICCV	编码-解码结构	两阶段结构-纹理模型，边界保留平滑图像引导内容修复，外表流方法做纹理补全
LISK[70]	2020/AAAI	编码-解码结构	多任务架构，共享生成器同时修复图像内容和结构

4.2 基于生成模型的图像修复方法

基于生成模型的图像修复方法指利用训练好的生成模型的强大的图像生成能力，基于破损图像已知先验分布推测未知分布的修复方法。目前比较有代表性的生成模型主要包括自回归模型(Autoregressive Model)[71]、VAE 和 GAN。表 4 列出了一些具有代表性的基于生成模型的算法。

表 4 基于生成模型的图像修复方法汇总

Table 4 Summary of image inpainting algorithms based on generative model

算法名称	年份/来源	生成模型	模型结构创新点
PixelRNN[71]	2016/ICML	自回归模型	二维 RNN 结构，逐像素点生成图像
DGM[72]	2017/CVPR	GAN 模型	利用 GAN 可以直接将噪声转换为图像的特点进行图像修复
CVAE-GAN[73]	2017/ICCV	VAE 模型，GAN 模型	结合 VAE 和 GAN，由编码器、生成器、分类器、判别器构成
PGGAN[74]	2020/CVPR	GAN 模型	双阶段训练，引入结构先验，修复更高分辨率的图像
PICNet[75]	2019/CVPR	VAE 模型，GAN 模型	结合 VAE 和 GAN，重构网络/生成网络并行实现多样性修复
UCTGAN[76]	2020/CVPR	GAN 模型	多模块双分支实现多样性修复

GooGle 团队于 2016 年提出的基于自回归模型的 PixelRNN 模型[71]，在图像修复任务中取得了不错的效果。该模型本质上是一种改进的二维循环神经网络，通过学习图像的离散概率分布来逐点预测二维图像中的像素以进行图像修复。由于是逐像素点生成图像，这种方法的计算成本高，且难以处理高分辨率图像。Yeh 等[72]将图像修复问题转变为迭代寻找与破损图像已知区域数据分布最匹配的噪声，再利用训练好的 GAN 网络得到完整图像的过程，提出了一种基于先验误差且不需要基准图片（Ground Truth）的图像修复方法。考虑到 VAE 模型生成的图像真实但模糊，而 GAN 生成的图像清晰但不那么真实，Bao 等[73]取长补短，提出了一个由编码器、生成器、分类器和判别器构成的生成网络，称之为 CVAE-GAN。他们尝试将其应用到图像修复任务中，并取得了一定的修复效果。

但是，文献[72-73]中的方法在寻找与破损图像最匹配的先验分布时，都包含复杂的迭代优化推理过程，因此对高分辨率图像的修复仍然存在局限性。为了加快推理过程，将基于生成模型的图像方法扩展到更高分辨率的图像修复任务中，Lahiri 等[74]设计了一种结合 GAN 网络和深度神经网络的双阶段训练策略：在第一阶段，训练一个将噪声分布映射到自然图像上的 GAN 网络；在第二阶段，固定前一阶段预训练好的 GAN 网络，另外训练一个用于预测输入图像的噪声先验的深度神经网络，推理出与输入图像最匹配的噪声先验，最后利用预训练好的 GAN 网络生成完整图像。该方法还引入了结构先验，在人脸图像修复任务中能很好地帮助模型保留人脸姿态信息，从而生成更真实的图像内容。

现有的图像修复方法大多可以得到视觉效果真实合理的修复结果，但均是一种一对一的修复模式，即输入一张破损图像只产生一个特定的输出。事实上，对于一张破损图像，可以有多个合理的修复结果。因此，Zheng 等[75]将 VAE 和 GAN 有效结合，提出了一种包含重构网络和生成网络且两种网络并行的多样性图像修复模型。其中，重构网络充分利用基准图像挖掘出图像数据的先验分布；生成网络预测缺失区域的潜在先验分布，同时结合重构网络生成多样化的修复结果。该模型的输入图像以及多种修复结果如图 8 所示。

图 8 输入图像及多样性修复结果[75]

Fig. 8 Input image and diverse inpainting results[75]

Zhao 等[76]将多样性图像修复任务视作一个已知边缘概率分布和联合概率分布，求解条件概率分布的问题，提出了一种无监督跨空间生成模型。该模型包含流形投影模块、编码模块和生成模块，其中由流形投影模块和生成模块组成的分支主要是将实例图像空间映射到条件完成图像空间，由编码模块组成的分支主要充当条件标签。

相比基于自编码的图像修复方法，基于生成模型的方法可以实现多样性修复。但由于生成模型存在训练不稳定、容易出现模式坍塌等问题，该类方法目前仅能处理较低分辨率的图像。

4.3 基于网络结构的图像修复方法

文献[77]认为，深度卷积网络在图像生成和修复领域所取得的优越表现主要得益于其能够从大量图像数据样本中学习先验知识的能力，然而深度生成网络在经过任何学习之前就已经拥有捕获大量低级图像统计信息的能力，即这些统计信息可以由深度卷积生成网络的结构捕获，而不是从大规模图像数据集中学习得到。为此，其提供了一种全新的修复思路，用一个随机初始化未训练的生成网络去拟合待修复图像，对单张破损图像的已知内容进行训练，将网络结构本身作为唯一的先验信息，不断迭代并反推出图像缺失区域的内容。这种修复思路十分巧妙，对于难以得到图像训练集的图像修复问题十分适用，但计算量较大，单张图像的修复需要迭代几千次。

5 相关数据集及评价指标

5.1 数据集

深度学习是数据驱动（Data-driven）的模型，以对大数据的学习为基石，用一种更加抽象的方式来表征数据分布，而学习出来的表征方式可以帮助我们更好地理解和分析数据，进而挖掘数据隐藏的结构和关系。因此，基于深度学习的图像修复方法同样需要大量的图像数据，目前最常用的图像数据集包括：ImageNet 数据集[78-79]、Paris StreetView 数据集[80]、Places2 数据集[81]、CelebA 数据集[82]和 CelebA-HQ 数据集[83]。图 9 给出了 Paris StreetView 数据集、Places2 数据集和 CelebA 数据集中的几张图像。

(a)Paris StreetView 数据集

(b)Places2 数据集

(c)CelebA 数据集

图 9 部分常用数据集

Fig. 9 Some commonly used datasets

(1)ImageNet 数据集。该数据集是一个包含约 1 400 万张图片,涵盖 2 万多个生活中常见的场景类别的大规模图像数据库。目前,ImageNet 数据集是深度学习图像处理任务中应用得最广泛的数据集。

(2)Paris StreetView 数据集。该数据集包含 14 900 张训练图像和 100 张测试图像,主要来源于现实街景,涵盖的场景内容包括建筑、树木、街道、天空等,图像分辨率为 936×537。

(3)Places2 数据集。该数据集包含约 1 000 万张图片,涵盖 400 多个不同类型的场景,广泛应用于以场景和环境为应用内容的视觉认知任务。

(4)CelebA 数据集。该数据集是香港中文大学于 2015 年发布的大型人脸属性数据集,其中包括约 20 万张名人图像,每张图片都有 40 个属性注释。

(5)CelebA-HQ 数据集。该数据集是在 CelebA 数据集上衍生出的高分辨率数据集,包含 3 万张高分辨率人脸图像,图像分辨率为 1 024×1 024。

5.2 评价指标

图像修复算法的优劣与修复后的图像质量直接相关,一般是用修复后的图像与参考图像对比来进行评估。在评估过程中,往往需要使用多种不同的评价指标进行评估才更合理。这些评价指标广义上分为主观评价指标和客观评价指标。主观评价是一种以人为观测者,对图像进行主观定性评价的方法。客观评价指根据人眼的主观视觉系统建立模型,用量化的标准去评价修复效果。图像修复方法中常用的客观评价指标包括:MSE,PSNR,SSIM 和 FID 等。下面对这几种客观评价指标进行说明。

(1)均方误差(Mean Squared Error,MSE),指两张图像像素点差平方的期望值,主要用于评价数据的变化程度。*MSE* 值越小,修复后的图像与真实图像就越相似,其计算式如式(2)所示:

$$MSE=\frac{1}{m}\sum_{i=1}^{m}(X_i-Y_i)^2 \tag{2}$$

其中,i 表示图像 X 和 Y 的像素点的变量,m 指图像像素点。

(2)峰值信噪比(Peak Signal to Noise Ratio,PSNR),是目前使用得最广泛的一种图像客观评价指标。*PSNR* 值越大,修复后的图像与真实图像之间的失真越小,图像质量就越好。其计算式如式(3)所示:

$$PSNR=10*\log_{10}\left(\frac{MAX_I^2}{MSE}\right)=20*\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \tag{3}$$

其中,MAX_I 表示图像像素值的最大值,*MSE* 表示均方误差。

(3)结构相似性(Structural Similarity Index,SSIM),是一种衡量两张图像结构相似度的指标,该值越大,图像失真程度越小,图像质量就越好。SSIM 用均值作为亮度的估计,标准差作为对比度的估计,协方差作为结构相似度的度量。其计算式如式(4)—式(7)所示:

$$L(X,Y)=\frac{2\mu_X\mu_Y+C_1}{\mu_X^2+\mu_Y^2+C_1} \tag{4}$$

$$C(X,Y)=\frac{2\sigma_X\sigma_Y+C_2}{\sigma_X^2+\sigma_Y^2+C_2} \tag{5}$$

$$S(X,Y)=\frac{\sigma_{XY}+C_3}{\sigma_X\sigma_Y+C_3} \tag{6}$$

$$SSIM(X,Y)=L(X,Y)*C(X,Y)*S(X,Y) \tag{7}$$

其中,μ_X 和 μ_Y 为图像 X 和 Y 的像素均值,σ_X 和 σ_Y 表示图像 X 和 Y 的像素的标准差,σ_{XY} 表示图像 X 和 Y 的协方差,C_1,C_2 和 C_3 是常数。

(4)FID(Fréchet Inception Distance),最初是衡量 GAN 网络优劣的一个指标,用于评价生成图片的质量(清晰度)和多样性,目前也被广泛应用于图像修复中。在计算 *FID* 时,首先使用 Inception 网络提取图像特征,然后通过均值和协方差矩阵来计算真实数据分布和生成数据分布之间的距离。*FID* 值越小,两个数据的分布越接近,生成的图片质量就越高,多样性越好。其具体计算式如式(8)所示:

$$FID(X,Y)=\|\mu_X-\mu_Y\|_2^2+T_r(\Sigma_X+\Sigma_Y-2(\Sigma_X\Sigma_Y)^{\frac{1}{2}}) \tag{8}$$

其中,μ_X 和 μ_Y 分别表示图像 X 和 Y 的均值,T_r 表示矩阵的迹,Σ 表示协方差。

6 基于深度学习的修复方法定量对比

图像修复本质上是一个病态(ill-posed)问题,修复结果的好坏更多依赖于人类的视觉感知,人为评测是一个较为主观的过程,而客观评测指标是用量化的标准去评价修复效果。在图像修复优劣评测中,一般采用以主观评测和客观评测相结合的方式。

目前,基于深度学习的图像修复算法处理的破损图像的掩膜类型主要包括规则掩膜和随机不规则掩膜。规则掩膜指位于图像中心或任意位置的固定形状的掩膜,破损区域面积占比一般为 25%;随机不规则掩膜则指位于图像任意位置且大小及形状任意的掩膜,破损区域面积的占比为 0~60%。文献[34]提出的不规则掩膜(Irregular Masks)和文献[36]提出的随机掩膜(Free-Form Masks)是最常使用的两种随机不规则掩膜类型,如图 10 所示。

(a)不规则掩膜[34]

(b)随机掩膜[36]

图 10 两种随机不规则掩膜

Fig. 10 Two random irregular masks

表 5 列出了上述各类基于深度学习的图像修复算法在常用图像数据集上的量化评测结果、修复图像分辨率、掩膜类型及缺失区域面积占比，其数值来自于各原文文献。其中，“↑”表示该项指标的值越大越好，“↓”表示该项指标的值越小越好，“—”表示相关文献中没有对该项评价指标进行数值说明，扩号中的内容指修复图像的破损面积占比。

表 5 算法在常用数据集上的量化评测结果

Table 5 Quantitative evaluation results of algorithms on common datasets

数据集	算法名称	*PSNR*↑	*SSIM*↑	*FID*↓	*L*1↓/%	图像分辨率	掩膜类型
ImageNet	GMCNN[38]	**22.43**	**0.8939**	—	—	256×256	规则掩膜(小于或等于 25%)
	VCNet[49]	19.58	0.6339	—	—	256×256	随机掩膜[36]
	CII[51]	—	0.5600	—	15.61	512×512	中心区域规则掩膜(20%)
	PICNet[75]	20.10	—	—	**12.91**	256×256	中心区域规则掩膜(25%)
Paris StreetView	CE[32]	17.59	—	—	10.33	128×128	中心区域规则掩膜(25%)
	GMCNN[38]	24.65	0.8650	—	—	256×256	规则掩膜(小于或等于 25%)
	DMFN[40]	25.00	0.8563	—	—	256×256	中心区域规则掩膜(25%)
	GPN[41]	19.72	—	—	8.11	128×128	中心区域规则掩膜(25%)
	SSDCGN[43]	20.57	—	—	7.33	128×128	中心区域规则掩膜(25%)
	HR[50]	18.00	—	—	10.01	128×128	中心区域规则掩膜(25%)
	Shift-Net[52]	**26.51**	**0.9000**	—	—	256×256	中心区域规则掩膜(25%)
	RFR[59]	26.44	0.8620	—	**2.75**	256×256	不规则掩膜[34](30%～40%)
Places2	PConv[34]	25.25	0.7790	—	1.88	512×512	不规则掩膜[34](20%～30%)
	LBAM[37]	25.59	0.7850	—	1.93	256×256	不规则掩膜[34](20%～30%)
	GMCNN[38]	20.16	0.8617	—	—	256×256	规则掩膜(小于或等于 25%)
	DMFN[40]	21.53	0.8079	—	—	256×256	中心区域规则掩膜(25%)
	RN[45]	25.06	0.8680	—	1.96	256×256	不规则掩膜[34](20%～30%)
	VCNet[49]	20.54	0.6988	—	—	256×256	随机掩膜[36]
	CA[53]	18.91	—	—	8.60	256×256	规则掩膜(小于或等于 25%)
	PEN-Net[57]	—	0.7809	15.190	9.94	256×256	中心区域规则掩膜(25%)
	MED[58]	**28.87**	**0.9230**	8.060	—	256×256	不规则掩膜[34](20%～30%)
	RFR[59]	22.63	0.8190	—	3.81	256×256	不规则掩膜[34](30%～40%)
	HRII[60]	19.69	0.8063	—	3.84	256×256	中心区域规则掩膜(25%)
		24.70	0.8744	—	2.20		随机掩膜[36]
	CRA[61]	—	0.8840	4.898	5.44	512×512	任意不规则掩膜
	EC[46]	21.75	0.8230	8.160	3.86	256×256	规则掩膜(25%)
		24.92	0.8610	4.910	2.59		不规则掩膜[34](20%～30%)
	PRVS[65]	25.66	0.9140	—	—	256×256	不规则掩膜[34](20%～30%)
	Stru-Flow[66]	25.22	0.9026	7.035	—	256×256	不规则掩膜[34](20%～40%)
	LISK[70]	22.46	0.8130	7.423	3.52	256×256	规则掩膜(25%)
		27.07	0.8870	**4.883**	**0.69**		不规则掩膜[34]
CelebA	CSA[54]	26.54	0.9310	—	1.83	256×256	中心区域规则掩膜(25%)
		32.58	**0.9820**	—	**0.94**		不规则掩膜[34](20%～30%)
	MED[58]	26.32	0.9100	25.510	—	256×256	中心区域规则掩膜(25%)
	RFR[59]	27.76	0.9340	—	2.12	256×256	不规则掩膜[34](30%～40%)
	LISK[70]	26.82	0.9270	1.654	2.08	256×256	规则掩膜(25%)
		33.19	0.9600	**1.227**	1.47		不规则掩膜[34]
	DGM[72]	19.40	—	—	—	64×64	中心区域规则掩膜(25%)
	PGGAN[74]	21.70	—	6.000	—	64×64	规则掩膜(40%)
CelebA-HQ	GMCNN[38]	25.70	**0.9546**	—	—	256×256	规则掩膜(小于或等于 25%)
	DMFN[40]	26.50	0.8932	—	—	256×256	中心区域规则掩膜(25%)
	PEPSI[55]	25.60	0.9010	—	—	256×256	规则掩膜(25%)
		28.60	0.9290	—	—		随机掩膜[34]
	PEPSI++[56]	25.50	0.8980	—	—	256×256	规则掩膜(25%)
		28.50	0.9280	—	—		随机掩膜[34]
	UCTGAN[76]	26.38	0.8862	—	**1.51**	256×256	中心区域规则掩膜(25%)

结束语 随着大数据时代的到来、硬件计算力的提高，以及迫切的应用需求，“大面积缺失图像”的修复任务受到了国内外学术界和工业界的广泛关注，成为了计算机视觉领域一个重要且具有挑战性的研究课题。基于深度学习的图像修复

方法主要借助图像结构连贯性和内容相似性，依靠卷积网络对图像特征强大的表达学习能力以及生成对抗网络对数据概率分布的拟合能力等来进行图像修复，该方法已成为目前图像修复领域的主流方法。本文在对现有的图像修复方法进行分类总结的基础上，针对目前研究任务中依然存在的问题和难点，对其未来的研究方向和发展趋势做出如下展望：

(1)图像修复的本质是通过挖掘已知信息来对缺失区域进行内容推测补全的一项计算机视觉任务。实现对已知信息的有效提取，并建立与未知内容的信息关联，是提高修复质量的必然需求。因此在后续研究中，提高修复模型对图像的特征表达学习能力依然是值得深入探索的问题之一。

(2)多步修复方式可以实现由粗到精的递进式修复，但该模式存在着不可忽略的缺陷，如计算资源的增加、各修复阶段之间的不良影响及训练难度的增加等。在未来的研究中，建立端到端的单步修复模型，采用模型内部优化模式，以实现纹理一致、结构连贯、语义明确的高质量图像修复，依旧是研究热点。

(3)图像修复是面向实际应用需求的研究任务，在实际应用中，破损区域通常处于未知状态。因此，盲修复将是未来研究的一个趋势。

(4)随着硬件设备技术的提高，图像分辨率得到很大提高，但目前的修复工作依然针对的是低分辨率图像。基于现有的修复技术和优化算法，实现高分辨率图像的细粒度修复将是未来研究的一个热点。

(5)生成对抗网络在图像生成中起到了关键作用，同样也被大多数图像修复方法采用，但目前生成对抗网络依然存在着自身缺陷，如模式坍塌、训练不稳定等，如何解决这些问题，也将成为图像修复研究的一项挑战。

参 考 文 献

[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[2] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409.1556, 2014.

[3] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.

[4] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[5] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]// Advances in Neural Information Processing Systems. 2014: 2672-2680.

[6] QIANG Z P, HE L B, CHEN X, et al. Survey on deep learning image inpainting methods[J]. Jurnal of Image and Graphics, 2019, 24(3): 447-463.

[7] BERTALMIO M, SAPIRO G, CASELLES V, et al. Image inpainting[C]// Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. 2000: 417-424.

[8] CHAN T F, SHEN J. Nontexture inpainting by curvature-driven diffusions[J]. Journal of Visual Communication and Image Representation, 2001, 12(4): 436-449.

[9] SHEN J, CHAN T F. Mathematical models for local nontexture inpaintings[J]. SIAM Journal on Applied Mathematics, 2002, 62(3): 1019-1043.

[10] SHEN J, KANG S H, CHAN T F. Euler's elastica and curvature-based inpainting[J]. SIAM Journal on Applied Mathematics, 2003, 63(2): 564-592.

[11] TSAI A, YEZZI A, WILLSKY A S. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification[J]. IEEE Transactions on Image Processing, 2001, 10(8): 1169-1186.

[12] ESEDOGLU S. Digital Inpainting Based on The Mumford-Shah-Euler Image Model[J]. European Journal of Applied Mathematics, 2003, 13(4): 353-370.

[13] ZHANG H Y, PENG Q C. A survey on digital image inpainting[J]. Journal of Image and Graphics, 2007, 12(1): 1-10.

[14] GROSSAUER H. A combined PDE and texture synthesis approach to inpainting[C]// European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2004: 214-224.

[15] RANE S D, SAPIRO G, BERTALMIO M. Structure and texture filling-in of missing image blocks in wireless transmission and compression applications[J]. IEEE Transactions on Image Processing, 2003, 12(3): 296-303.

[16] YAMAUCHI H, HABER J, SEIDEL H P. Image restoration using multiresolution texture synthesis and image inpainting[C]// Proceedings Computer Graphics International 2003. IEEE, 2003: 120-125.

[17] DRORI I, COHEN-OR D, YESHURUN H. Fragment-based image completion[J]. ACM Transactions on Graphics, 2003, 22(3): 303-312.

[18] ZHANG Y, XIAO J, SHAH M. Region completion in a single image[OL]. http://diglib.eg.org/handle/10.2312/egs20041002.

[19] CRIMINISI A, PÉREZ P, TOYAMA K. Region filling and object removal by exemplar-based image inpainting[J]. IEEE Transactions on Image Processing, 2004, 13(9): 1200-1212.

[20] TANG F, YING Y, WANG J, et al. A novel texture synthesis based algorithm for object removal in photographs[C]// Annual Asian Computing Science Conference. Springer, Berlin, Heidelberg, 2004: 248-258.

[21] CRIMINISI A, PEREZ P, TOYAMA K. Object removal by exemplar-based inpainting[C]// 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2003, 2: II-II.

[22] 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. 2005: 64-69.

[23] HAYS J, EFROS A A. Scene completion using millions of photographs[J]. Communications of the ACM, 2008, 51(10): 87-94.

[24] WHYTE O, SIVIC J, ZISSERMAN A. Get Out of my Picture! Internet-based Inpainting[C]// BMVC. 2009, 2(4): 5.

[25] LIU Y，SHE J C，GONG Y C，et al. Survey of facial completion techniques based on deep learning[J]. Application Research of Computers，2021，38(1)：9-14.

[26] RUMELHART D E，HINTON G E，WILLIAMS R J. Learning internal representations by error propagation[R]. California Univ San Diego La Jolla Inst for Cognitive Science，1985.

[27] HINTON G E，OSINDERO S，TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural computation，2006，18(7)：1527-1554.

[28] MASCI J，MEIER U，CIREŞAN D，et al. Stacked convolutional auto-encoders for hierarchical feature extraction[C]// International Conference on Artificial Neural Networks. Springer，Berlin，Heidelberg，2011：52-59.

[29] BENGIO Y，LAMBLIN P，POPOVICI D，et al. Greedy layer-wise training of deep networks[J]. Advances in Neural Information Processing Systems，2006，19：153-160.

[30] KINGMA D P，WELLING M. Auto-encoding variational bayes [J]. arXiv：1312. 6114，2013.

[31] SHEN L. Research on image inpainting methods based on semantic perception deep model [D]. Hefei：Hefei University of Technology，2020.

[32] PATHAK D，KRAHENBUHL P，DONAHUE J，et al. Context encoders：Feature learning by inpainting[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016：2536-2544.

[33] IIZUKA S，SIMO-SERRA E，ISHIKAWA H. Globally and locally consistent image completion[J]. ACM Transactions on Graphics (ToG)，2017，36(4)：1-14.

[34] LIU G，REDA F A，SHIH K J，et al. Image inpainting for irregular holes using partial convolutions[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018：85-100.

[35] RONNEBERGER O，FISCHER P，BROX T. U-net：Convolutional networks for biomedical image segmentation[C]// International Conference on Medical Image Computing and Computer-assisted Intervention. Springer，Cham，2015：234-241.

[36] YU J，LIN Z，YANG J，et al. Free-form image inpainting with gated convolution[C]// Proceedings of the IEEE International Conference on Compu-ter Vision. 2019：4471-4480.

[37] XIE C，LIU S，LI C，et al. Image inpainting with learnable bidirectional attention maps[C]// Proceedings of the IEEE International Conference on Computer Vision. 2019：8858-8867.

[38] WANG Y，TAO X，QI X，et al. Image inpainting via generative multi-column convolutional neural networks[C]// Advances in Neural Information Processing Systems. 2018：331-340.

[39] XIAO Q，LI G，CHEN Q. Deep inception generative network for cognitive image inpainting[J]. arXiv：1812. 01458，2018.

[40] HUI Z，LI J，WANG X，et al. Image fine-grained inpainting[J]. arXiv：2002. 02609，2020.

[41] ZHANG H，HU Z，LUO C，et al. Semantic image inpainting with progressive generative networks[C]// Proceedings of the 26th ACM International Conference on Multimedia. 2018：1939-1947.

[42] HOCHREITER S，SCHMIDHUBER J. Long short-term memory[J]. Neural Computation，1997，9(8)：1735-1780.

[43] SHEN L，HONG R，ZHANG H，et al. Single-shot Semantic Image Inpainting with Densely Connected Generative Networks [C]// Proceedings of the 27th ACM International Conference on Multimedia. 2019：1861-1869.

[44] HONG X，XIONG P，JI R，et al. Deep fusion network for image completion[C]// Proceedings of the 27th ACM International Conference on Multimedia. 2019：2033-2042.

[45] YU T，GUO Z，JIN X，et al. Region Normalization for Image Inpainting[C]// AAAI. 2020：12733-12740.

[46] NAZERI K，NG E，JOSEPH T，et al. Edgeconnect：Generative image inpainting with adversarial edge learning[J]. arXiv：1901. 00212，2019.

[47] CAI N，SU Z，LIN Z，et al. Blind inpainting using the fully convolutional neural network [J]. The Visual Computer，2017，33(2)：249-261.

[48] LIU Y，PAN J，SU Z. Deep blind image inpainting[C]// International Conference on Intelligent Science and Big Data Engineering. Springer，Cham，2019：128-141.

[49] WANG Y，CHEN Y C，TAO X，et al. VCNet：A Robust Approach to Blind Image Inpainting[J]. arXiv：2003. 06816，2020.

[50] YANG C，LU X，LIN Z，et al. High-resolution image inpainting using multi-scale neural patch synthesis[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017：6721-6729.

[51] SONG Y，YANG C，LIN Z，et al. Contextual-based image inpainting：Infer，match，and translate[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018：3-19.

[52] YAN Z，LI X，LI M，et al. Shift-net：Image inpainting via deep feature rearrangement[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018：1-17.

[53] YU J，LIN Z，YANG J，et al. Generative image inpainting with contextual attention[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018：5505-5514.

[54] LIU H，JIANG B，XIAO Y，et al. Coherent semantic attention for image inpainting[C]// Proceedings of the IEEE International Conference on Computer Vision. 2019：4170-4179.

[55] SAGONG M，SHIN Y，KIM S，et al. Pepsi：Fast image inpainting with parallel decoding network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019：11360-11368.

[56] 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，2020，32(1)：252-265.

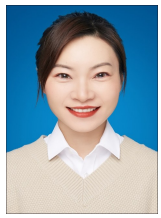
[57] ZENG Y，FU J，CHAO H，et al. Learning pyramid-context encoder network for high-quality image inpainting[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019：1486-1494.

[58] LIU H，JIANG B，SONG Y，et al. Rethinking Image Inpainting via a Mutual Encoder-Decoder with Feature Equalizations[J]. arXiv：2007. 06929，2020.

[59] LI J，WANG N，ZHANG L，et al. Recurrent Feature Reasoning for Image Inpainting[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020：7760-7768.

[60] ZENG Y, LIN Z, YANG J, et al. High-Resolution Image Inpainting with Iterative Confidence Feedback and Guided Upsampling[J]. arXiv: 2005.11742, 2020.
[61] YI Z, TANG Q, AZIZI S, et al. Contextual Residual Aggregation for Ultra High-Resolution Image Inpainting[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 7508-7517.
[62] SONG Y, YANG C, SHEN Y, et al. Spg-net: Segmentation prediction and guidance network for image inpainting[J]. arXiv: 1805.03356, 2018.
[63] XIONG W, YU J, LIN Z, et al. Foreground-aware image inpainting[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 5840-5848.
[64] LIAO L, XIAO J, WANG Z, et al. Guidance and Evaluation: Semantic-Aware Image Inpainting for Mixed Scenes[J]. arXiv: 2003.06877, 2020.
[65] LI J, HE F, ZHANG L, et al. Progressive reconstruction of visual structure for image inpainting[C]// Proceedings of the IEEE International Conference on Computer Vision. 2019: 5962-5971.
[66] REN Y, YU X, ZHANG R, et al. Structureflow: Image inpainting via structure-aware appearance flow[C]// Proceedings of the IEEE International Conference on Computer Vision. 2019: 181-190.
[67] XU L, LU C, XU Y, et al. Image smoothing via L 0 gradient minimization[C]// Proceedings of the 2011 SIGGRAPH Asia Conference. 2011: 1-12.
[68] 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.
[69] ZHOU T, TULSIANI S, SUN W, et al. View synthesis by appearance flow[C]// European Conference on Computer Vision. Springer, Cham, 2016: 286-301.
[70] YANG J, QI Z, SHI Y. Learning to Incorporate Structure Knowledge for Image Inpainting[C]// AAAI. 2020: 12605-12612.
[71] OORD A, KALCHBRENNER N, KAVUKCUOGLU K. Pixel recurrent neural networks[J]. arXiv: 1601.06759, 2016.
[72] YEH R A, CHEN C, YIAN LIM T, et al. Semantic image inpainting with deep generative models[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 5485-5493.
[73] BAO J, CHEN D, WEN F, et al. CVAE-GAN: fine-grained image generation through asymmetric training[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 2745-2754.
[74] LAHIRI A, JAIN A K, AGRAWAL S, et al. Prior Guided GAN Based Semantic Inpainting[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 13696-13705.
[75] ZHENG C, CHAM T J, CAI J. Pluralistic image completion[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 1438-1447.
[76] ZHAO L, MO Q, LIN S, et al. UCTGAN: Diverse Image Inpainting Based on Unsupervised Cross-Space Translation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 5741-5750.
[77] ULYANOV D, VEDALDI A, LEMPITSKY V. Deep image prior[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 9446-9454.
[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] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 248-255.
[80] DOERSCH C, SINGH S, GUPTA A, et al. What Makes Paris Look Like Paris? [J]. Acm Transactions on Graphics, 2012, 31(4): 1-9.
[81] ZHOU B, LAPEDRIZA A, KHOSLA A, et al. Places: A 10 million image database for scene recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(6): 1452-1464.
[82] LIU Z, LUO P, WANG X, et al. Deep learning face attributes in the wild[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015: 3730-3738.
[83] KARRAS T, AILA T, LAINE S, et al. Progressive growing of gans for improved quality, stability, and variation[J]. arXiv: 1710.10196, 2017.

ZHAO Lu-lu, born in 1997, postgraduate. Her main research interests include deep learning and image inpainting.



SHEN Ling, born in 1983, Ph.D, lecturer. Her main research interests include computer vision and machine learning.