西北大学学报(自然科学版)
2023 年 12 月,第 53 卷第 6 期,Dec. ,2023,Vol. 53,No. 6
Journal of Northwest University (Natural Science Edition)

·西北大学学报 110 周年刊庆·

基于深度学习的图像修复方法研究综述

彭进业[1,2],余　喆[1],屈书毅[1],胡琦瑶[1],王　珺[1,2]

(1. 西北大学 信息科学与技术学院,陕西 西安　710127;
2. 陕西省丝绸之路文化遗产数字化保护与传承协同创新中心,陕西 西安　710127)

摘要　图像修复是指通过使用计算机算法和图像处理技术还原损坏、缺失或被破坏的图像区域,其目标是使修复后的图像在视觉上具有合理的结构、纹理和连贯性,并且尽可能与原始图像的外观和信息接近。传统的图像修复技术通常基于规则和启发式方法,利用像素间的局部关系、边缘信息、纹理统计等低级特征进行图像修复,难以修复具有复杂语义的图像。近年来,深度学习技术由于其强大的特征提取能力,在图像修复任务中逐渐成为主流方法。这些方法借助大规模数据集进行训练,通过深层次的卷积神经网络或生成对抗网络自动学习图像的高级特征和复杂语义信息。然而,现有的图像修复总结研究较少,且深度学习技术更新太快,为了更好地推动深度学习技术在图像修复领域中的应用及发展,有必要对现有相关方法进行分类和总结。该文对基于深度学习的图像修复方法进行了系统回顾和全面概述,从修复策略的角度出发对图像修复方法进行系统性总结。具体分析了每类方法的优势和局限性,总结了常用的数据集、定量评价指标及代表性方法的性能对比,对图像修复领域存在的难点问题及未来研究方向进行了展望。

关键词　数字图像处理;图像修复;深度学习;计算机视觉
中图分类号:TP391.41;TP183　　**DOI**:10.16152/j.cnki.xdxbzr.2023-06-006

A survey of image inpainting methods based on deep learning

PENG Jinye[1,2], YU Zhe[1], QU Shuyi[1], HU Qiyao[1], WANG Jun[1,2]

(1. School of Information Science and Technology, Northwest University, Xi'an 710127, China;
2. Shaanxi Province Silk Road Digital Protection and Inheritance of Cultural Heritage Collaborative Innovation Center, Xi'an 710127, China)

Abstract　Image inpainting is a process that involves utilizing computer algorithms and image processing techniques to restore damaged, missing, or corrupted regions within an image. The objective of this process is to generate visually reasonable and coherent structures and textures in the repaired regions, while simultaneously being as consistent as possible with the appearance of the original image. Traditional image inpainting techniques predominantly rely on rule-based and heuristic methods, utilizing low-level features such as local pixel relationships, edge information, and texture statistics to perform inpainting tasks. However, handling images with intricate semantics through these methods has proven challenging. In recent years, the prominence of

收稿日期:2023-10-29
基金项目:国家自然学科基金(62101446);陕西省科技计划重点项目(2021ZDLGY15-06);陕西省自然科学基金(2023-JC-QN-0750)。
第一作者:彭进业,男,教授,博士生导师,从事图像处理与模式识别、多媒体信息检索、量子信息处理、文化遗产数字化技术等研究,pjy@nwu.edu.cn。

deep learning technology has grown significantly in image inpainting tasks owing to its powerful feature extraction capabilities. By leveraging large-scale datasets, these methods automatically learn high-level features and complex semantic information of images through deep convolutional neural networks or generative adversarial networks. However, there are few existing summary studies on image inpainting, while the evolution of deep learning technology is progressing rapidly. In order to facilitate the effective application and development of deep learning methods in image inpainting, a systematic categorization and summary of existing techniques is imperative. This article provides a systematic review and comprehensive overview of deep learning-based image restoration methods, offering a systematic summary of image inpainting methods from the perspective of inpainting strategies. We specifically analyzed the strengths and limitations of each method category, summarized commonly used datasets, quantitative evaluation metrics, and performance comparisons of representative approaches. Ultimately, we discussed the existing challenges in the field of image inpainting and proposed potential research avenues for future investigations.

Keywords digital image processing; image inpainting; deep learning; computer vision

图像是人类沟通交流、传递、记录与保存信息的重要手段。早期的人们以纸、墙壁、石碑为载体记录生活及艺术创作,随着时间推移,这些图像载体受到环境、气候或人为因素的影响,导致其表面出现风化、褪色、氧化及污损,不利于文化的传承。最早的图像修复技术起源于文艺复兴时期,修复师根据古老的图像和颜色痕迹来修补损坏的部分,尽可能使修复的部分与原始的绘画风格和色彩相匹配,以便保持整体的视觉一致性。这项技术很大程度上依赖于修复师的经验和对古老艺术品的理解,费时且费力。

随着计算机技术的发展,数字图像逐渐成为记录和保存信息的主要媒介。然而,数字图像在传递和存储过程中,会不可避免地出现像素丢失等质量退化问题,因此,数字图像修复技术应运而生。传统的图像修复方法的工作原理是根据图像的已知区域推断未知区域,利用纹理结构一致性、样本相似性等思想构建算法,能够修复一些破损较小的图像。当破损的区域面积较大、与已知区域无明显相关性、结构纹理较复杂时,其修复后的图像与原始图像存在明显差异,且伴有破损边缘模糊、断层等问题。

近年来,随着计算硬件的不断进步,深度学习技术在计算机视觉、自然语言处理、语音识别等多个领域取得了突破性的进展[1]。图像修复作为计算机视觉任务的基础,在特征学习和语义理解方面得到了强大的技术支撑。利用深度学习技术可以获取图像的高级语义信息,生成具有正确语义的内容,解决了传统图像修复方法的不足。其中,图像修复效果较突出的深度学习模型有 Rumelhart 等人提出的自编码器(autoencoder, AE)[2]、Goodfellow 等人提出的 GAN(generative adversarial network)[3]、Vaswani 等人提出的 Transformer[4]、Dhariwal 等人提出的 Diffusion Models[5]等。研究者们在上述模型的基础上,根据不同的数据类型、修复策略和应用场景进行改进,解决了大面积缺失的图像修复、不规则图像修复等难题。

尽管图像修复技术是许多视觉下游任务的基础,但相关的前沿综述性工作很少。因此,本文针对基于深度学习的图像修复算法的发展,从修复策略的角度出发,对图像修复算法进行系统性梳理,分类框架如图 1 所示。根据不同的修复策略,本文将基于深度学习的算法分为基于像素生成式修复、渐进式修复、基于不规则卷积修复、基于 Transformer 修复、基于扩散模型修复和基于调制修复[6-8]。为了直观地展示不同修复策略下图像修复的效果,本文介绍了不同类型图像修复方法的实验比对、常用数据集和质量评价指标。最后,重点分析了当前图像修复领域存在的难点和问题,并对未来科学热点和研究趋势进行展望。

1 基于修复策略的图像修复研究现状

修复策略从不同的角度出发,为图像修复问题提供了不同的解决方案。本节将修复策略分为 6 类:像素生成式修复,渐进式修复、基于不规则卷积修复、基于 Transformer 修复、基于扩散模型修复和基于调制修复,并对每一类方法的核心思

想和发展进程进行系统性梳理(见图 1)。

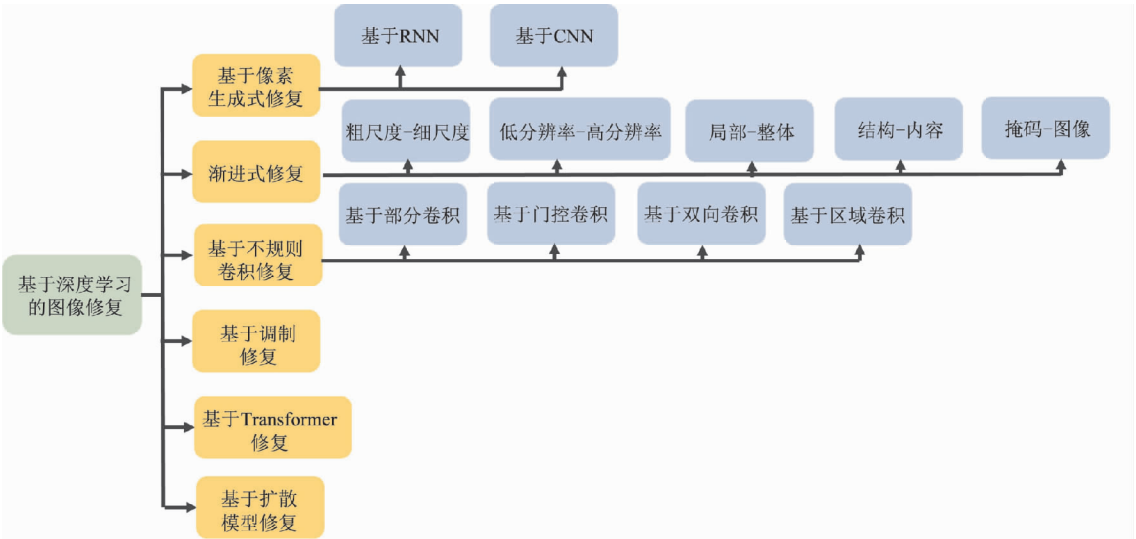


图 1　基于深度学习的图像修复方法分类框架

Fig. 1　Classification framework for deep learning-based image inpainting methods

1.1　像素生成式图像修复

基于像素生成式图像修复方法旨在通过逐个像素地生成缺失区域的像素值恢复损坏的图像。这种方法使用循环神经网络(RNN)[9]或卷积神经网络(CNN)[10],以已知区域的某个像素点为基点,通过设计算法由基点像素逐渐向周围像素进行迭代计算,生成新的像素点,从而实现未知区域的图像填充。

1.1.1　基于 RNN 的生成式图像修复

基于 RNN 的生成式图像修复算法通常将图像分解成像素序列,并使用 RNN 对整张图像的像素序列进行遍历,学习全局样本的特征分布,从而逐个生成缺失区域的像素值,其原理如图 2 所示。

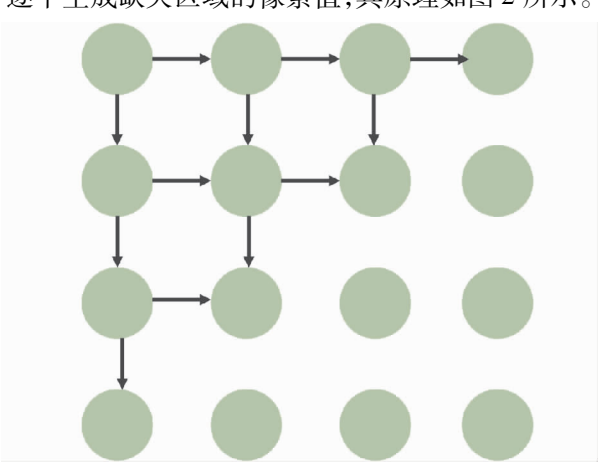


图 2　基于 RNN 的像素生成原理

Fig. 2　Principles of pixel generation based on RNN

其具体步骤分为 3 步。①数据准备。将缺失的图像分解成像素序列,并将缺失区域的像素作为输入序列。其中,每个像素通常由其坐标、周围像素的值和其他上下文信息组成。②模型构建。构建一个 RNN 模型,利用前一个时刻像素的值预测当前时刻的像素值。合理利用周围像素的值和全局特征,这些上下文信息能够很好地帮助模型理解图像的结构和纹理,从而生成更精确的像素值。③逐像素迭代生成。从图像的左上角开始,RNN 模型将根据已生成的像素和上下文信息逐个像素点地预测修复后的像素值。每一次迭代,模型都根据之前生成的像素和上下文信息进一步优化修复结果。

Van 等人于 2016 年提出一个新颖的 PRNN(pixel recurrent neural networks)结构[11],通过对长短期记忆 LSTM[12]层采用残差连接,构建了新颖的二维 LSTM 层:行 LSTM 和对角 BiLSTM,它们更容易扩展到更大的数据集。PRNN 可以在生成图像时利用先前生成的像素和上下文信息,具有较强的图像修复能力,并且能够生成具有细节和纹理的高质量图像。然而,由于其逐像素生成的特性,生成图像的速度较慢,同时可能面临长距离依赖问题。为了克服这些问题,后续的研究在 PRNN 的基础上进行了改进,例如 PixelCNN[13]采用了更高效的掩膜卷积结构,它允许模型在生成每个像素时只考虑其左边和上边的像素。这样的限制确保了模型生成图像的因果性,使得图像生成更加快速。在此基础上,Salimans 等人对 Pixel-

CNN 进行改进,提出了一种名为 PixelCNN + + 的改进模型[14],PixelCNN + + 使用了离散化的逻辑混合似然度,与原始 PixelCNN 使用的 256-way softmax 相比,能够更快地训练模型。这些改进使得基于 RNN 的图像生成模型在生成高分辨率图像时取得了更好的性能。

基于 RNN 的生成式图像修复算法可以利用图像中的时序和上下文信息,生成更加精确、纹理连贯的图像。由于这种方法需遍历全局像素点,因此,在处理大尺度图像时会面临计算复杂度高、耗时较长的问题。并且在遍历像素的后期阶段,像素点之间的相关性会逐渐减弱,使得该算法对于复杂的缺失图案表现不佳。

1.1.2 基于 CNN 的生成式图像修复

CNN 在处理图像数据时,由于局部连接和共享权重的结构,能够有效地捕捉图像和其他空间数据中的局部特征,有助于降低参数量,提高模型的训练效率,并且对更大范围的特征关系也能较好地处理,其生成原理见图 3。

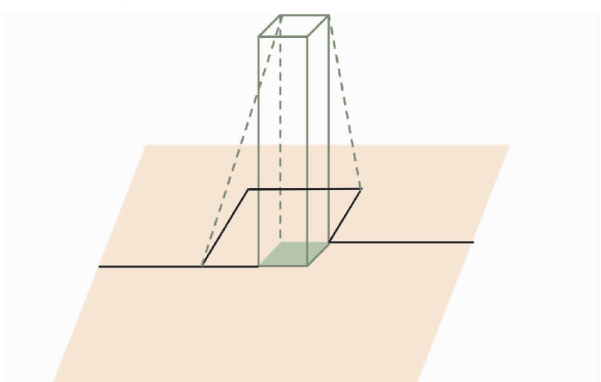


图 3 基于 CNN 的像素生成原理

Fig. 3 Principles of pixel generation based on CNN

Oliveira 等人受 CNN 的启发,提出了一种简单快速的图像修复方法,将待修复图像与加权平均内核进行卷积,计算像素邻域的加权平均值[15]。该算法的速度相比于先前的算法快 2 到 3 个数量级,从而使得修复在交互式应用中变得实用。Hadhoud 等人注意到文献[15]生成的像素是由周围的邻域像素产生的,应该将每个像素的颜色与邻近像素的一小部分颜色进行平均,并将其颜色的一小部分贡献给每个相邻像素,因此,其将中心零权重的位置修改至右下角,再进行卷积,不需要太多次的迭代卷积操作就可以修复出更高质量的图像[16]。Jain 等人发现卷积网络能够提供与小波和马尔可夫随机场方法相当的性能,在某些情况下甚至更好,因此,提出了一种更高效更快速的低级视觉的图像修复方法,结合了两个主要思想:使用卷积网络作为图像处理架构,以及从特定噪声模型合成训练样本的无监督学习过程[17]。但由于该方法仅限于加入特定的噪声类型,因此使用的局限性很大。

综上所述,像素生成式图像修复相比一些传统的图像修复方法,不需要先验信息或人工标记的辅助数据,更具有自主性和自适应性。并且可以适用于各种图像修复任务,包括缺失、遮挡、噪声和破损等,具有一定的通用性。但是该类方法存在以下缺点:①需要大量的计算资源和时间,特别是在处理高分辨率图像时;②需要大量高质量有标签的训练数据,否则修复结果可能不理想;③在修复具有复杂纹理和细节的情况下可能会产生伪影或不真实的细节。像素生成式图像修复方法在图像处理领域具有很大的研究前景,未来的研究方向将集中在改进模型的泛化能力、数据集的质量及计算效率等方面。

1.2 渐进式图像修复

渐进式图像修复策略旨在将图像由较低质量一步一步修复成高质量的图像,从子任务中获得的附加信息有助于最终结果的生成,其实现方式有多种,包括由粗尺度图像逐渐修复到细尺度图像、由局部到整体修复图像、由低分辨率逐渐修复到高分辨率、由图案结构逐渐修复出图像内容、由掩膜到图像的修复。

1.2.1 粗尺度到细尺度图像修复

Yu 等人观察到 CNN 从远程空间位置借用或复制信息方面的效率不高,导致修复大型缺失区域时产生与周围区域不一致的失真结构或模糊纹理,从而提出了一种新的基于深度生成模型的方法,不仅可以合成新的图像纹理结构,而且还可以在网络训练过程中显式地利用周围图像特征作为参考,以获得更好的预测结果[18]。该网络包括 2 个阶段:第 1 阶段是粗尺度图像修复阶段,该阶段用重建损失训练一个简单的编码器-解码器网络来得到粗尺度的图像内容;第 2 阶段是细尺度图像修复阶段,该阶段采用与第 1 阶段相同的编码器-解码器结构,并集成了上下文注意力模块,能够充分利用周围图像特征作为参考,产生合理的修复结果。虽然这种方法取得了视觉上良好的结果,但由于其特征编码分为两阶段,需要大量的计算资源。为了降低粗尺度-细尺度结构的计算资

源,Sagong 等人提出了一个新的网络结构 PEPSI,该网络由一个共享的编码网络和一个具有粗尺度路径和细尺度路径的并行解码网络组成[19]。粗尺度路径产生一个初步的修复结果,用于训练编码网络以预测上下文注意力模块的特征。同时,细尺度修复路径使用由上下文注意力模块重建的改进特征生成更高质量的修复结果。Ma 等人发现直接采用标准的卷积架构容易忽略长距离区域之间的相关性,提出了区域级卷积来局部处理不同类型的图像区域,既可以精确地重建已知区域,又能从已知区域中粗略推断出未知区域[20]。同时,引入非局部操作对不同区域之间的相关性进行全局建模,从而保证缺失和现有区域之间的视觉一致性。最后,将区域级卷积和非局部相关性集成到一个由粗到细的网络框架中,以恢复语义合理且视觉逼真的图像。

1.2.2　局部到整体图像修复

由局部到整体的图像修复将整个修复任务细分成了不同的子任务,每个子任务都从缺失区域的外层逐渐向内进行修复,最终将局部修复的结果整合在一起,完成整幅图像的修复。这样可以确保修复的结果在局部和整体上都具有合理的结构和连贯性。Zhang 等人提出一种基于局部–整体的语义图像修复方法,该方法将整个修复过程分成 4 个阶段,从缺失区域的外围逐步向中心进行修复,每个阶段旨在完成整个修复过程的一部分,并在后续阶段中进一步优化修复结果[21]。在每个阶段,网络根据之前阶段的修复结果和图像的语义信息生成新的修复结果,从而逐步填补缺失区域,得到高质量的修复结果。Li 等人设计了一种循环特征推理(RFR)网络,主要由插入式的循环特征推理模块和知识一致性注意力(KCA)模块构成[22]。类似于人类解决问题的方式,先解决较简单的部分,然后将结果作为额外信息来解决较困难的部分。RFR 模块循环地推断卷积特征图的缺失边界,然后将其作为进一步推理的线索。该模块逐步加强了对缺失区域中心的约束,使修复结果变得更加精准。Zeng 等人引入了一个深度生成模型,不仅能够输出修复结果,还输出相应的置信度图[23]。将中间过程产生的置信度图作为反馈,逐步填补缺失区域,每次迭代只保留缺失区域内置信度高的像素,并在下一次迭代中重点关注未填充的像素。由于该方法重复使用前一次迭代的部分预测结构作为已知像素,这个过程将逐渐优化最终的修复结果。

从局部到整体的的图像修复可以产生许多中间结果,合理利用这些中间结果可以减少最终修复结果的误差。但是,中间结果的生成也会消耗更多的计算时间。

1.2.3　低分辨率到高分辨率图像修复

由低分辨率到高分辨率的图像修复首先将高分辨率图像降采样为低分辨率图像,然后在低分辨率图像上进行修复,以减少计算成本。Yang 等人提出了一种混合优化方法,该方法将编码器–解码器的预测作为全局内容约束,并将缺失区域与已知区域之间的局部神经 patch 块的相似性作为纹理约束[24]。修复出来的低分辨率结果经过上采样操作将被再次精细化修复,从而生成高分辨率的修复结果。将高频残差图像添加到大的模糊图像上能够生成具有丰富细节的图像,基于此,Yi 等人提出了一种上下文残差聚合技术,通过对上下文 patch 块中的残差进行加权聚合生成缺失区域的高频残差,网络只需要在低分辨率图像上进行预测,用低分辨率图像指导高分辨率图像进行修复[25]。因此,该方法在内存和计算功率的消耗上大大减少,并且降低了对高分辨率训练数据集的需求。Kulshreshtha 等人认为增加图像尺寸会相应地减少网络在修复区域可用的局部上下文信息,因此,提出了一种新颖的由低分辨率到高分辨率的迭代优化方法,该方法通过使用低分辨率预测作为指导,在推断过程中最小化多尺度一致性损失,仅优化网络的中间特征图,在优化过程中能够生成细节丰富的高分辨率图像修复结果,同时,保持了低分辨率预测的颜色和结构[26]。Liu 等人提出了一种通过参数化坐标查询进行高分辨率图像修复的新颖框架 CoordFill,只需在低分辨率条件下对高分辨率图像进行编码,以捕捉更大的感受野[27]。该方法首先对高分辨率图像进行下采样并编码缺失区域,然后,通过基于注意力的快速傅里叶卷积参数生成网络为每个空间块产生空间自适应参数,最后,将这些参数作为一系列多层感知器的权重和偏差,输入是编码的连续坐标,输出是合成的颜色值,这种连续的位置编码有助于通过在高分辨率图像上重新采样坐标,合成逼真的高频纹理。

1.2.4　结构到内容图像修复

由结构到内容的图像修复,其主要目标是先恢复图像的结构信息,然后再填充细节内容。这

类方法首先通过设计算法尝试恢复图像的大致结构,包括边缘、轮廓和主要的物体形状,这一步旨在填充缺失区域,使得整体图像看起来更加完整和连贯。在结构恢复的基础上,再进一步利用周围像素的上下文信息完善修复结果。

边缘能够表现出物体的形状和轮廓,是常用的引导方式。Liao 等人提出了一种考虑场景结构和上下文的图像修复模型 E-CE[28]。之前的内容编码器使用整个图像的上下文预测缺失图像区域,E-CE 通过根据边缘结构信息恢复纹理,避免了图像不同边缘之间的上下文信息易混淆的问题。该方法首先从 mask 图像中提取边缘,并通过一个全卷积网络进行边缘修复。然后,将完成的边缘图与原始遮罩图像一起输入到修改后的上下文编码器网络中,以预测缺失区域。Nazeri 等人采用结构感知的策略,提出了一个两阶段模型 EdgeConnect,将图像修复问题分为结构预测和图像补全 2 个阶段[29]。EdgeConnect 的第 1 阶段主要用于预测缺失区域的图像结构,提取出边缘图,然后将边缘图传递给第 2 阶段,用于引导缺失区域的修复过程。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,但是对于一些精细化局部结构的处理还有欠缺。为解决以上问题,Li 等人设计了一个视觉结构重建层(VSR),解决边缘结构和特征的重建,通过共享参数使二者相互受益[30]。具体而言,VSR 采用部分卷积和瓶颈块恢复缺失区域中部分边缘信息,然后将重新构建的边缘与缺失的输入图像相结合,通过填充语义上有意义的内容逐步缩小缺失区域的范围。Ren 等人专注于细粒度纹理的修复,提出了一个两阶段模型,将图像修复任务分为结构重建和纹理生成两部分[31]。在第 1 阶段使用保留边缘的平滑图像训练一个结构重建器,修复输入图像中的缺失结构;第 2 阶段设计了一个使用外观流的纹理生成器产生图像的细节。Deng 等人采用结构引导的双分支网络用于古代壁画修复,壁画修复过程分为结构重建和内容修复[32]。在结构重建阶段,利用门控卷积和快速傅里叶卷积残差块重建受损壁画的缺失结构。在内容修复阶段,使用由结构重建阶段生成的边缘结构引导壁画的内容修复。由于图像的边缘结构通常是稀疏的,只传递图像的二进制轮廓信息,而梯度图本身不仅传递了可能的边缘信息,还包含一些纹理信息或高频细节。基于此,Yang 等人提出先预测整个梯度图,引入梯度信息嵌入方案,将学习到的结构特征明确地输入到图像修复过程中[33]。

分割技术可以预测图像中不同物体的边界和形状信息,用这些分割结果指导图像修复是非常有意义的。为解决生成模型没有利用语义分割信息约束物体的形状,从而导致边界模糊的问题,Song 等人分解了图像修复过程中类间差异和类内变化,将修复过程分解为分割预测和分割引导 2 个步骤,首先预测缺失区域的分割标签,然后生成分割引导的修复结果[34]。Yu 等人基于 Segment-Anything 模型(SAM)提出了一种名为 Inpaint Anything(IA)的新模型,该模型是一种多功能工具,结合了移除任何物体、填充任何内容和替换任何内容的功能,还能够处理多样化和高质量的输入图像[35]。

由结构到内容的图像修复可以确保修复的图像保持原有的形状和结构,同时加入了更真实的纹理和细节信息。但是仍存在以下问题:①边缘信息无法指导颜色的生成;②分割信息依赖于标签的精度,如果相同语义标签的外观差异太大,则分割信息会混淆最终的修复结果。

1.2.5 掩膜到图像修复

由掩膜预测到图像修复是盲图像修复中常用的方法。盲图像修复是指在图像中缺失或损坏的像素位置未知的情况下,通过算法自动恢复这些缺失或损坏的像素,不需要为缺失区域指定掩码,使图像看起来完整和清晰。这种技术可以广泛应用于图像去噪、修复损坏的旧照片等。

Liu 等人受到残差学习算法的启发,引入了编码器和解码器结构,并改进了 L1 损失函数处理异常值,该算法可以预测损坏区域中缺失的信息[36]。在掩膜预测的过程中会不可避免地出现预测误差,导致后续修复的图像中出现伪影。为了解决这个问题,Wang 等人提出了一个两阶段的视觉一致性网络,首先,预测语义不一致的区域,使掩码预测的可信度更高,然后,使用新的空间归一化方法修复预测的缺失区域,通过这种方式,生成了在语义上令人信服和在视觉上引人注目的内容[37]。为了跳过损坏区域的预测步骤并获得更好的结果,Phutke 等人提出了一种新的端到端架构,其中包括小波查询多头注意力变换模块和全向门控注意力模块[38]。所提出的小波查询多头注意力将经过处理的小波系数作为查询提供给多

头注意力,从而提供了编码器特征。全向门控注意力从编码器学习到的所有维度的注意力特征将被传输到相应的解码器当中。

由掩膜到图像的修复方法不需要提供手动绘制的掩膜,省时省力。但仍然面临许多挑战:①难以准确区分受损区域和有效区域,有效区域可能包含纹理、边缘和其他重要信息,而这些信息也可能在受损区域存在,使得模型难以区分;②缺乏掩膜信息,模型容易受图像中复杂结构的干扰,导致不合理的修复效果;③预测受损区域始终存在一定的误差,难以实现高质量的修复效果。

综上所述,渐进式图像修复类方法可以逐渐提高修复结果的质量,并且生成的图像具有平滑自然的过渡效果,能够避免在修复区域和原始图像之间产生明显的边缘。另一方面,研究者可以根据数据集的特性和需求,自行设计在网络训练时添加所需细节。但该类方法存在以下缺点:①多次的迭代计算生成修复结果,需要大量的计算资源;②该类方法通常需要采用两阶段网络的结构,相对于一次性修复方法更加复杂;③想要生成高质量的修复结果可能需要更多时间,难以适应一些实时或高效率需求的应用。在未来,渐进式图像修复方法的研究可能集中在以下几方面:①开发更高效的算法,减少计算成本和时间延迟;②研究能够自动调整修复速度和细节程度的方法,以适应不同的需求和场景;③研究适用于实时或互动的应用,如视频修复。

1.3　基于不规则卷积的图像修复

在传统的图像修复中,缺失的区域通常是通过周围像素的信息填充,或者通过学习深度神经网络生成缺失内容。然而,在某些情况下,修复过程可能会引入伪影或不一致性。在基于不规则卷积的图像修复中,神经网络被设计用于改进架构中的卷积操作,具有更强大的自适应性,有助于在修复缺失区域的同时更好地保留原始图像的结构和纹理。卷积核的形状和尺寸可以灵活调整,因此,不规则卷积可以适应不同形状和大小的掩膜,并对其进行有效的修复。目前,根据卷积滤波器的类型,可以将不规则卷积分为部分卷积、门控卷积、双向卷积和区域卷积。

1.3.1　基于部分卷积的图像修复

Liu 等人在 2018 年首次提出采用部分卷积进行不规则掩膜的图像修复[39]。部分卷积的操作原理如图 4 所示。传统卷积核的每个元素都被用来与图像的对应位置进行加权计算,而部分卷积中对于缺失区域内的像素,卷积核的权重被设为 0,不进行计算。这样可以避免缺失区域的信息被不正确地填充。受上述启发,Chen 等人将部分卷积应用于数字敦煌壁画的修复,使用基于部分卷积的深度神经网络作为壁画修复的基础模型,并采用滑动窗口方法进行数据增强,以解决训练过程中的样本量不足问题,并提高网络的准确性[40]。该方法在大面积不规则缺失的壁画图像上修复效果良好。Wang 等人提出了一种基于多尺度自适应部分卷积和模拟笔画形状掩膜的唐卡壁画修复方法,设计了一种基于核的多尺度自适应部分卷积,能够准确区分有效像素和无效像素,并提取多尺度对象的特征,这对提取唐卡壁画中的多尺度信息非常有效[41]。

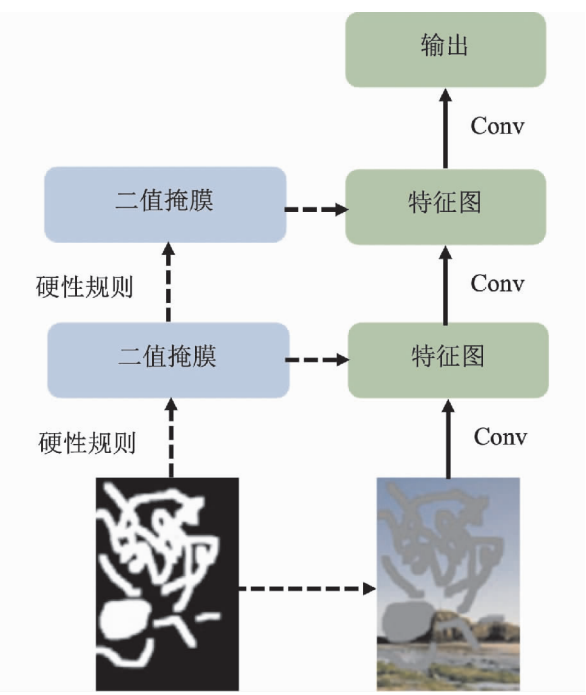


图 4　部分卷积

Fig. 4　Partial convolution

虽然部分卷积的提出大大提高了不规则图像修复的效率和精度,但是它并未精细到考虑卷积滤波器覆盖的像素数量。当滤波窗口内包含有效像素时,即使有效像素的数量非常小,当前位置的值都会变为 1。

1.3.2　基于门控卷积的图像修复

Yu 等人在 2019 年提出门控卷积[42],其基本思想是只选择部分像素参与卷积运算,而其他像素则被动态地忽略或削弱其权重。门控卷积的操作原理如图 5 所示。在门控卷积中,对于每个通

道在每个空间位置,都有一个可学习的门控机制,用于决定是否对该像素进行卷积操作。这样可以使得卷积操作对于图像中的不同区域有不同的处理方式,从而更好地应对不规则的图像修复任务。Chang 等人提出了一种自由形式掩膜视频修复模型,使用 3D 门控卷积处理自由形式掩模的不确定性[43]。Li 等人设计了一种基于门控卷积和自注意力的金字塔网络 GAP-Net,并改变了特征提取策略,该方法改善了不规则图像的修复效果并加速了网络的学习速度[44]。Xie 等人利用带有门控卷积的生成对抗网络对 CT 图像的截断区域进行图像修复,并将这些修复后的图像应用于放射治疗的剂量计算中[45]。该方法可以直接有效地对不完整的 CT 图像进行修复,并且在图像可视化和剂量学方面更接近真实标签结果。Ma 等人提出了一种新颖的密集门控卷积网络用于生成图像修复,通过修改门控卷积的网络结构,将门控卷积和密集连接的共同优点集成到一起,大大减少了网络参数,有效地改善了网络的修复效果[46]。

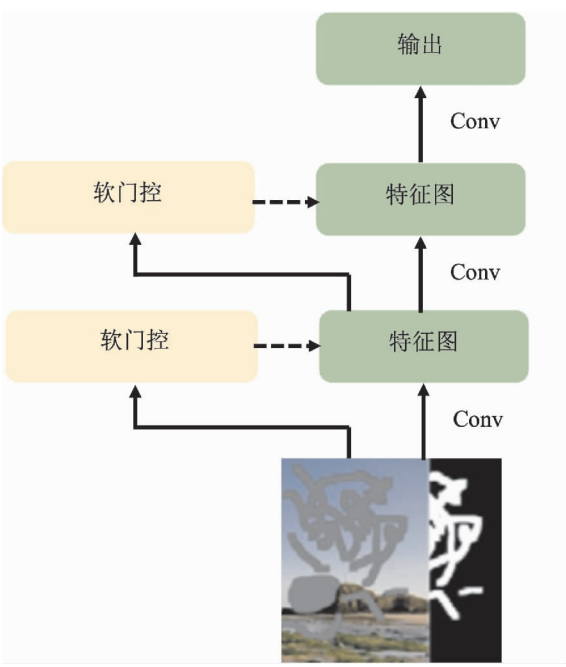


图 5 门控卷积

Fig. 5 Gate convolution

1.3.3 基于双向卷积的图像修复

传统的卷积在修复图像时只关注输入图像的局部特征,忽略了输出图像的全局特征。针对这个问题,Xie 等人在 2019 年提出利用双向卷积进行图像修复[47],其原理如图 6 所示。双向卷积不仅考虑了从输入到输出的卷积过程,还考虑了从输出到输入的卷积过程。这种正反向信息同时包含的卷积操作使得模型可以同时从输入和输出的角度来理解图像的特征和结构。同时,引入了可学习的双向注意力图,该注意力图允许模型在修复图像时同时关注缺失区域和周围的上下文信息,从而进一步提高修复结果的准确性和质量。Guo 等人提出了一种边缘引导的可学习双向注意力图 Edge-LBAM,改进不规则缺失区域的图像修复[48]。该方法引入了一个可学习的注意力图模块,用于学习特征重新归一化和掩膜更新,使其能够以端到端的方式进行训练。此外,在解码器中进一步提出了可学习的反向注意力图,用于强调填充未知像素而不是重建所有像素。该方法在生成连贯的图像结构和防止颜色不一致和模糊方面是有效的。Ma 等人将双向卷积的思想与 Transformer 技术相结合,提出了一种新颖的双向自回归 Transformer 的图像修复模型[20]。该方法利用 Transformer 学习自回归分布,还结合了掩膜语言模型,实现对丢失区域的上下文信息进行双向建模,从而对缺失的图像实现更好地修复。Guo 等人提出了一种图像修复双流网络,将结构约束纹理合成和纹理引导结构重建以耦合方式建模,这两个子任务可以交换有用的信息,从而实现相互促进[48]。此外,引入了一个双向门控特征融合模块以及上下文特征聚合模块,进一步优化结果,使得修复后的图像既具有语义合理的结构,又包含丰富的细节纹理。

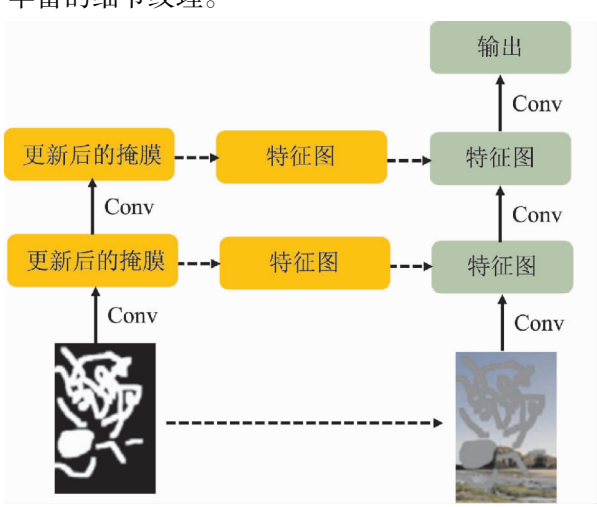


图 6 双向卷积

Fig. 6 Bi-directional convolution

1.3.4 基于区域卷积的图像修复

传统的图像修复方法通常使用全局卷积来填补图像中的缺失部分,这可能导致修复结果的细节丢失和模糊。为了解决这个问题,Ma 等人引入

了区域卷积和非局部相关的概念,在修复过程中更好地保留了图像细节和结构[20]。区域卷积将每个像素的卷积核限制在特定的区域内,从而有选择地捕捉局部细节。其中,卷积核的大小和形状可以根据具体的应用场景进行灵活调整,完全取决于需要修复的缺失区域的形状,从而更好地适应不同的图像修复任务。

综上所述,基于不规则卷积修复的方法具有更大的自由度,能够更精确地捕获和修复图像中的细节,并且适用于不同形状的掩膜,在处理复杂图像修复问题时具有优势。相对于传统卷积方法,不规则卷积可以更好地减少伪影的产生,使修复结果更自然。但该类方法相比于传统卷积,其计算复杂度更大,且超参数更难调整。未来的研究可以关注于改进不规则卷积操作的设计,以提高其性能和适应性,同时降低计算量。

1.4　基于调制修复

调制技术是一种在生成模型中使用的方法,它可以调整生成器的特征表示,从而控制生成样本的特征和风格。调制技术最初源自图像风格迁移的研究,后来被应用于生成对抗网络和其他生成模型中。近年来,调制技术在图像生成、图像修复、图像编辑等领域都有广泛的应用,可以帮助图像修复方法更好地控制图像的生成过程,从而产生更真实和合理的修复结果。在图像修复中,调制可以用于 2 个方面:特征调制和空间调制。特征调制通过学习参数调整生成器的特征表示,使其能够根据输入的条件信息生成不同风格或类别的修复结果。例如,可以通过调制参数指定修复结果的颜色、纹理或形状等特征。空间调制通过学习参数调整生成器的特征表示,使其能够根据输入的空间位置信息在不同位置上生成不同的修复内容。这样可以确保修复结果在不同位置上保持一致性和逼真性。

Zhao 等人引入了一种通用的方法,即共调制生成对抗网络 Co-Modulated GANs[49]。Co-Modulated GANs 的核心思想是将无条件调制生成器的生成能力适应到图像条件生成器中,采用了一个联合仿射变换对样式表示进行条件约束,使得生成器在生成图像的过程中灵活地控制特征的分布和统计特性。通过特征调制,Co-Modulated GANs 能够融合条件输入(例如图像的部分信息)和随机性(例如潜在向量的随机采样),从而生成多样且一致的结果。为了更好地建模图像的全局上下文信息,Zheng 等人提出了一种新颖的机制,将全局代码调制与空间代码调制级联,以便处理部分无效的特征,并更好地将全局上下文注入到空间区域中[50]。首先,从最高级别特征中提取全局风格代码 S,并对其进行 L2 规范化,然后,使用基于多层感知机的映射网络从噪声中生成一个风格代码 W,模拟图像生成的随机性,最后,将风格代码 W 与 S 组合在一起成为全局代码,用于后续的解码步骤。在解码阶段提出了一种全局-空间级联调制,通过全局调制块和空间调制块分别并行地上采样全局特征和局部特征,以实现在解码阶段连接全局上下文。该方法确保了修复后的图像在全局和局部尺度上保持一致。

综上所述,基于调制的图像修复方法可以根据需求自行调整生成器的特征表示,对于保留图像的细节和纹理非常有用。这种有选择性地修复特定频率范围内信息的方式更具有灵活性,不会对整个图像进行过度处理。但是该类方法大多数都采用生成模型,对于高质量有标签的数据有大量的需求,而配对的数据集往往难以获取。因此,基于调制的修复方法在未来可以重点关注于有效利用有限数据和弱标签,以及如何更精确地控制生成样本的属性。

1.5　基于 Transformer 修复

Transformer 是 Vaswani 等人在 2017 年提出的一种基于注意力机制的神经网络架构[51],最初用于自然语言处理任务。在传统的 RNN 或 CNN 中,序列数据会引入顺序依赖或局部性,这导致在处理长序列数据时可能面临梯度消失或梯度爆炸等问题。而 Transformer 采用完全基于自注意力机制的新型网络结构,关注序列中的所有位置,建立全局依赖,从而更好地处理长序列数据。

如图 7 所示,Transformer 由编码器和解码器组成,其中,编码器用于处理输入序列数据,而解码器用于生成输出序列数据。每个编码器和解码器由多层堆叠的自注意力机制和前馈神经网络构成。Transformer 的结构设计在应用于图像修复任务中具有一定的优势:一方面,自注意力机制允许网络在生成修复结果时对图像的各个位置进行精细的关注,从而能够更好地恢复复杂的图像结构和细节;另一方面,Transformer 可以通过调整注意力机制的尺度适应不同大小的图像修复任务,能够灵活处理小尺寸和大尺寸的图像,无需重新设计网络结构。

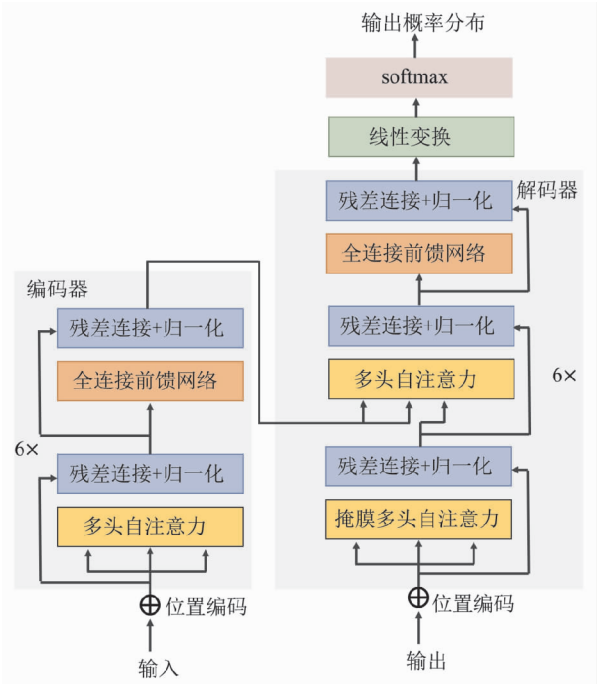


图 7 Transformer 模型架构

Fig. 7 The architecture of Transformer

Wan 等人将 Transformer 应用于图像修复领域,将 Transformer 的外观先验重构与 CNN 的纹理补充结合在一起,利用 Transformer 恢复了多样化的连贯结构以及一些粗糙的纹理,利用 CNN 在高分辨率掩膜图像的引导下增强了粗略先验的局部纹理细节[52]。Zhou 等人将 Transformer 应用于修复复杂场景的图像,提出了一种多个单应性变换融合方法 TransFill[53]。模型首先估计两个图像之间的匹配特征点,根据它们在目标图像中估计的深度将内点进行聚类,并为每个聚类估计一个单应性以进行初始图像配准,得到粗尺度的修复。然后,使用深度双边颜色转换解决颜色匹配问题,并通过像素级空间变换解决视差问题,得到进一步修复的结果。最后,通过学习一组融合掩码合并之前产生的修复结果,得到最终修复的图像。Wang 等人提出一种频率引导 Transformer 和自顶向下细化网络 FT-TDR,用于修复人脸盲图像[54]。FT-TDR 使用基于 Transformer 的网络通过建模不同块之间的关系检测要修复的受损区域,生成掩码。然后,采用了一个自顶向下的细化网络,以分层的方式恢复不同层次的特征,并生成与未遮挡的人脸区域在语义上一致的内容。由于 ViT[55] 在图像视觉领域具有巨大的应用前景,Cao 等人将 ViT 作为掩膜自动编码器,并使用来自 MAE 的注意力先验,使修复模型学习到遮挡和未遮挡区域之间更多的远距离依赖关系[56]。与先前依赖于先验直接指导的方法不同,Yu 等人在 Transformer 基础上开发了一个端到端的多模态引导图像修复网络,包含 1 个修复分支和 2 个用于语义分割和边缘纹理的辅助分支[57]。在每个 Transformer 块内,通过辅助去归一化,并提出多尺度空间感知注意模块,用来高效地学习多模态结构特征。与当前仅在像素级别上使用解码器的 Transformer 图像修复不同,Liu 等人提出了同时包含编码器和解码器的 Transformer 网络模型[58]。其中,编码器通过自注意模块捕获图像中所有 patch 块的纹理语义相关性,解码器中建立了一个动态的 patch 词汇表,用于在掩膜区域上填充 patch。在此基础上,通过概率扩散过程,提出了一个以已知区域为锚点的结构-纹理匹配注意模块,将这两者的优势结合起来进行渐进式修复。为了构建一个适合个人使用的小型计算模型,Chen 等人提出了一种结合特殊 Transformer 和传统卷积神经网络的轻量级模型,并提出了一种新的损失函数加强颜色细节[59]。Naderi 等人将 Swin Transformer 块引入人脸修复任务中,实现更大的感受野,并平衡全局和局部特征[60]。通过给每个面部部位使用单独的鉴别器,增加了修复模型的泛化能力,提高了其对语义面部部位的理解。Liao 等人提出了一种基于多尺度 Transformer 架构的参考引导的修复框架 TransRef,核心思想是在编码器的每个尺度中,引导信息逐步嵌入到缺失图像中[61]。具体而言,提出了一个参考 patch 对齐模块,用于粗略地对齐参考图像和掩膜图像。为了对粗略对齐的特征进行优化,提出了一个参考 patch 模块,首先,通过多头参考注意机制在小 patch 级别上对融合特征进行优化,然后,与掩膜图像的主要特征进行融合,最后,将来自所有尺度的融合特征进行级联,解码为完整的图像。

综上所述,基于 Transformer 修复的方法在图像的全局上下文理解方面更有优势,能够更好地理解图像内容和语义信息,修复完成的图像更具有合理性。但是,Transformer 模型通常需要大量的计算资源,应用于实时修复时可能导致较长的修复时间,并且当模型尺寸较大时,很难部署在受限的设备中。因此,基于 Transformer 的图像修复方法在未来可能集中于研究小规模的模型,尽可能在减小计算复杂性的同时,保证图像修复的性能,让其部署在移动设备或嵌入式系统中成为可能。

1.6　基于扩散模型修复

扩散模型是一类基于概率分布的生成模型,用于生成图像或其他类型的数据样本。它们通常利用随机扩散过程模拟样本生成的过程,通过逐渐去除信号中的噪声生成高质量的样本。在最近的研究中,扩散模型已经被证明可以生成高质量的图像,并且具有一些理想的属性[62-63],如分布覆盖范围、固定训练目标和易于扩展等。相比于经典的 CNN、GAN 模型,扩散模型具有更好的泛化能力,不易出现模式崩溃的问题,且不需要特定掩膜的训练就可以产生高保真的输出。

Lugmayr 等人提出了一种基于去噪扩散概率模型的图像修复方法 RePaint,该方法适用于极端掩模[64]。如图 8 所示,RePaint 由两阶段组成。第 1 阶段使用深度神经网络生成缺失像素的粗略估计。该网络在大量类似图像的数据集上进行训练,并使用卷积和反卷积层学习周围像素与缺失像素之间的关系。第 2 阶段采用去噪扩散概率模型处理来自第 1 阶段输出的粗略估计。该模型使用马尔科夫随机场建模图像像素与周围上下文之间的依赖关系,然后,基于周围像素预测缺失像素的最可能值,并根据上一次迭代的结果更新其预测,直到预测收敛为最终解决方案。RePaint 的一个关键优势是能够处理带有缺失像素、裂缝、孔洞和其他类型损伤的图像,还可以处理纹理表面,例如头发和草地。Li 等人提出了一种修复大面积缺失区域同时保留图像纹理和结构的空间扩散模型 SDM[65]。SDM 使用深度神经网络生成缺失像素的粗略估计,然后使用空间扩散模型对其进行改进。空间扩散模型的一个关键优势是能够填充大的缺失区域,同时保持原始图像的纹理和结构。Horita 等人引入结构引导解决大面积缺失的图像修复问题,提出了一种结构引导扩散模型 SGDM[66]。SGDM 由结构生成器和纹理生成器组成,都属于扩散概率模型。结构生成器用于生成边缘图像,该图像指导纹理生成器进行更具语义效率的修复。由于依赖结构生成器的输出可能会导致错误的修复,因此采用了一种联合训练方法,应用了贝叶斯去噪和动量框架[67],从数据增强中随机擦除区域,防止数据损坏并提高泛化性能。虽然 SGDM 生成的图像在结构、纹理和颜色梯度方面具有更好的泛化能力,但仍存在 2 个缺点,一是无法生成具有足够封闭边缘的图像,二是采用两个扩散模型使得计算成本较高。为了改善计算成本高和所需时间长的问题,Rombach 等人提出利用压缩模型和 UNet 架构在低维潜在空间上进行高分辨率的图像合成,该空间具有较低的复杂性[68]。压缩模型基于自编码器,它通过感知损失和基于 patch 的对抗目标训练,有助于实现自然的重建效果。这种尝试首次达到了在复杂度降低和细节保留之间的平衡点,极大提高了修复后图像的视觉保真度。

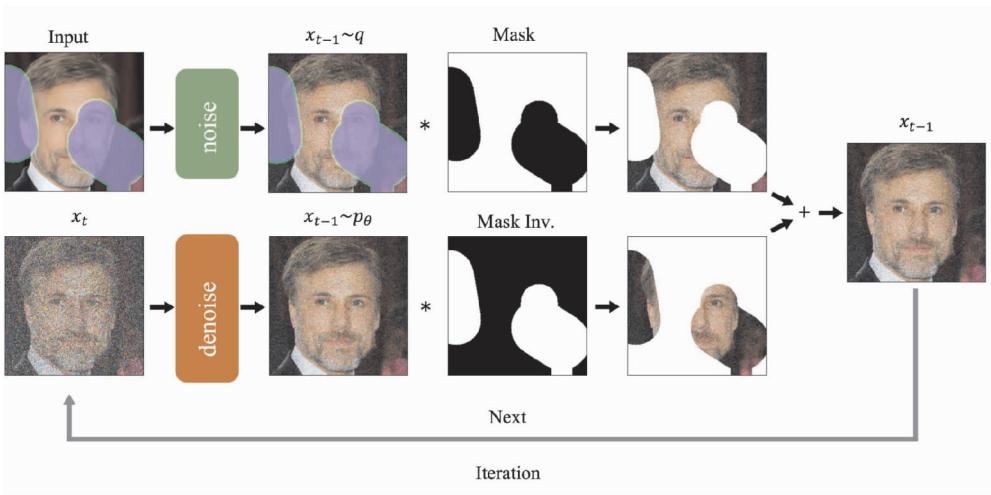


图 8　RePaint 模型架构

Fig. 8　The architecture of RePaint

综上所述,基于扩散模型的修复方法利用扩散过程填补缺失区域,通常能够生成高质量的修复结果,在修复大面积缺失的图像上具有优势,还可以生成多样性的修复结果,在创造性修复方面为研究者提供更多的思路。但是该类方法存在训练时间长、计算复杂度高的问题,扩散模型通常依

赖于一些参数的选择和调整,并且需要研究者在参数设置方面具有一定的经验。

2 图像修复相关数据集

数据集在整个机器学习流程中起着至关重要的作用,是用于训练、验证和测试机器学习模型的基础。实际上,收集大量配对的缺失图像和完整图像是相当困难的,因此,研究者大多数是利用大规模的公共数据集,通过在这些公共数据集上设计掩膜,生成缺失图像。在图像修复任务中,数据集被分为图像数据集和掩膜图像数据集。图像数据集的类别包括物体、场景、人脸等,掩膜图像数据集分为规则掩膜和不规则掩膜,规则掩膜一般由研究者直接在图像任意位置添加矩形掩膜获得。本节将介绍每个类别中的一些代表性数据集。

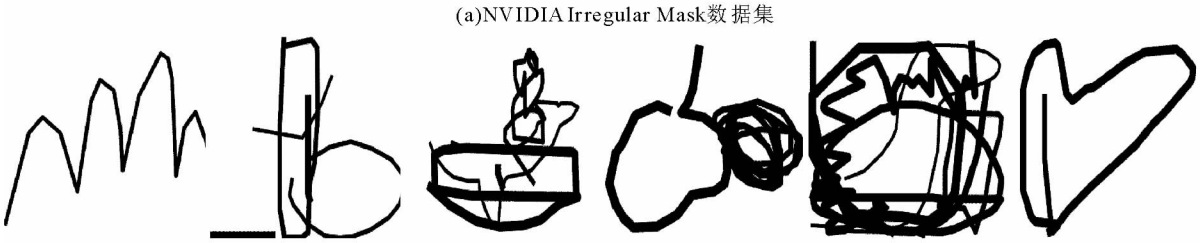
2.1 掩膜图像数据集

NVIDIA Irregular Mask 数据集[39]收集了大量不规则掩膜,其中包含 55 116 张用于训练的掩膜图像,以及 12 000 个用于测试的掩膜图像。该数据集中图像的分辨率为 512 ×512,图 9(a)展示了该数据集的样例。

Quick Draw Irregular Mask 数据集[69]是一个不规则掩膜数据集,它是基于 Quick Draw 数据集制作的。通过从 Quick Draw 数据集中随机采样笔画、随机选择笔画数量以及随机采样放大倍数,得到分辨率为 512 ×512 的 50 000 张训练掩膜图像和 10 000 张测试掩膜图像,图 9(b)展示了该数据集的样例。

Foreground-aware 数据集[70]是一个不规则掩膜数据集,包含了标注的前景和背景信息,可以帮助算法更好地理解图像中不同部分的语义和结构。该数据集包含 10 万个用于训练的掩膜和 1 万个用于测试的掩膜,每个掩模都是大小为 256 ×256 的二值图像。

(a)NVIDIA Irregular Mask数据集



(b)Quick Draw Irregular Mask数据集

图 9 掩膜图像数据集样例

Fig. 9 An example of masked image dataset

2.2 图像数据集

常用的街景图像数据集有谷歌街景数字图像数据集 SVHN[71]、巴黎街景图像数据集 Paris StreetView[72]和城市街景数据集 Cityscapes[73]。SVNH 数据集是从 Google StreetView 中的门牌号获得的,由 99 289 张经过小裁剪的数字图像组成,涵盖了从 1 位数字到 3 位数字不等的各种房号,图像中的数字可能出现在不同位置、不同尺度或不同背景下。Paris StreetView 数据集收集了巴黎城市中不同地区的街景图像,这些图像捕捉了城市的各个角落和风貌,它包含 14 900 个训练图像和 100 个测试图像。Cityscapes 数据集侧重于城市街道场景的语义理解,收集了德国和其他城市的城市街景图像。每张图像都有详细的像素级别标注,用于表示每个像素属于不同的类别,如道路、建筑物、车辆等。该数据集总共包含 5 000 张带有精细标注的图像和 20 000 张带有粗略标注的图像。Cityscapes 数据集样例如图 10(a)所示。

常用的场景数据集包含日常场景图像数据集 MS COCO[74]、大规模多场景图像数据集 ImageNet[75]和自然场景图像数据集 Places2[76]。MS COCO 数据集收集了来自各种场景和环境的图像,涵盖了超

过 80 个不同的物体和场景类别,如人、动物、物体、交通工具、建筑物等。每张图像都有详细的标注,包括物体的边界框、物体类别、图像分割掩膜等信息。ImageNet 是根据 WordNet 层次结构组织的图像数据集,每个子集都代表一个有意义的概念,当前版本的数据集包含 21 841 个非空子集和 14 197 122 张图像。ImageNet 数据集样例如图 10(b)所示。Places2 数据集包含来自 400 个场景类别的超过 1 000 万张图像,具有不同种类的场景、不同季节、天气和光照条件。

常用的人脸图像数据集包含人脸标志数据集 Helen Face[77]、大型人脸属性数据集 CelebA[78]、高质量图像数据集 CelebA-HQ[79] 和多样化的高质量人脸数据集 FFHQ[80]。Helen Face 数据集中的人脸图像包括了不同种族、性别、年龄和表情的人脸,以及不同姿势、光照条件和背景。每张人脸图像都有对应的人工标注,标注了人脸上的关键点位置,如眼睛、鼻子、嘴巴。该数据集包含 2 000 张用于训练的图像和 330 张用于测试的图像。CelebA 数据集是一个大规模的人脸属性数据集,收集了来自互联网的名人人脸图像。该数据集多样性大、数量大、注释丰富,包括 10 177 个身份、202 599 张人脸图像、5 个地标位置以及每张图像 40 个属性注释。CelebA-HQ 数据集由 GAN 模型开发,构建了 CelebA 的高质量版本,该数据集包含 30 000 张尺寸为 1 024 × 1 024 的图像。FFHQ 数据集收集了来自 Flickr 图片分享平台的人脸图像,其中包括各种不同类型的人脸照片,包含 70 000张尺寸为 1 024 × 1 024 的图像。

常用的物体图像数据集包括建筑物数据集 Façade[81]、纹理数据集 DTD[82]、斯坦福汽车数据集 Stanford Cars[83]。Facade 数据集是来自不同城市、具有不同建筑风格的立面图像的数据集,从现代建筑到传统建筑等,包括来自不同来源的 606 张经过校正的立面图像。DTD 数据集是一个纹理数据库,收集野外的纹理图像,如植物、动物、纹理材质等。该数据集由 5 640 张图像组成,按照 47 个类别进行组织。DTD 数据集样例如图 10(c)所示。Stanford Cars 数据集收集了不同品牌、型号、颜色和角度的汽车图像,每张汽车图像都有对应的标注,标注了汽车的型号、品牌等信息。该数据集包含来自 196 类汽车的 16 185 张图像。

(a)Cityscapes数据集

(b)ImageNet数据集

(c)DTD数据集

图 10　图像数据集样例

Fig. 10　An example of image dataset

3 质量评价指标及代表性方法性能对比

在完成图像修复工作后,一般需要通过质量评价指标衡量模型的性能。质量评价方法分为主观评价和客观评价,主观评价方法需要多名观察者对修复后的图像与原图进行对比并打分,客观评价方法采用不同的属性对修复后的图像与原图进行计算。主观评价依赖于观察者的主观感受,不仅费时费力,而且缺乏公平性。客观评价借助不同的评判指标对图像进行量化界定,能够区分人眼感知不到的细微差别,从不同角度和属性出发,对图像进行更全面的评判。例如,均方误差、峰值信噪比、结构相似性指数[84]和学习感知图像块相似性通常被用来衡量重构图像的质量。初始分数[85]、Fréchet 初始距离[86]在生成对抗网络中通常被用来衡量生成图像样本的质量。本节简要介绍常用的客观评价指标的工作原理。

1)平均绝对偏差(mean absolute error, MAE,式中简记 E_{MAE})[87]衡量了修复后的图像 I_{out} 与原始图像 I_{image} 像素之间的绝对差异的平均程度,MAE 值越小说明修复后的图像越接近原始图像。其计算公式为

$$E_{\mathrm{MAE}} = \frac{1}{n}\sum_{i=1}^{n} | I_{\mathrm{out}} - I_{\mathrm{image}} | \tag{1}$$

2) 均方误差(mean square error, MSE,式中简记 E_{MSE})[88] 衡量了修复后的图像 I_{out} 与原始图像 I_{image} 之间差异的平方的平均值,MSE 值越小,修复后的图像质量越好。然而,由于平方项的存在,MSE 在处理异常值时可能会受到较大的影响,较大的误差会被放大。其计算公式为

$$E_{\mathrm{MSE}} = \frac{1}{n}\sum_{i=1}^{n}(I_{\mathrm{out}} - I_{\mathrm{image}})^2 \tag{2}$$

3) 峰值信噪比(peak signal to noise ratio, PSNR,式中简记 R_{PSNR})是一种衡量噪声影响修复结果程度的评价指标,比较修复后的图像 I_{out} 与原始图像 I_{image} 之间的相似性。较高的 PSNR 值表示重建图像与原始图像之间的差异较小,即图像的质量较高。其计算公式为

$$R_{\mathrm{PSNR}} = 20\lg\frac{P_{\mathrm{MAX}}}{\sqrt{E_{\mathrm{MSE}}}} \tag{3}$$

式中:P_{MAX} 是图像的最大可能像素值。

4) 结构相似性指数(structure similarity index measure,SSIM,式中简记 M_{SSIM})是一种用于测量两个图像之间的相似度。SSIM 能够感知结构信息的变化,它基于修复后的图像 I_{out} 和原始图像 I_{image} 之间的 3 个属性进行比较测量:亮度、对比度、结构。SSIM 的计算公式是 3 个属性的加权组合,

$$M_{\mathrm{SSIM}} = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \tag{4}$$

式中:$l(x,y)$、(x,y)、$s(x,y)$ 表示修复后的图像 I_{out} 和原始图像 I_{image} 的亮度、对比度、结构的估计值。

5) 学习感知图像块相似性(learned perceptual image patch similarity,LPIPS,式中简记 S_{LPIPS})[89] 是一种使用卷积神经网络衡量图像之间相似性的评价指标。LPIPS 首先将修复后的图像 I_{out} 和原始图像 I_{image} 分成小的图像块,然后,使用预训练的 CNN 模型提取这些图像块的特征表示,最后,通过比较这些特征表示的差异计算图像之间的相似性分数。其计算公式为

$$S_{\mathrm{LPIPS}} = \sum_{l}\frac{1}{H_i \times W_i}\| W_l \odot (\boldsymbol{\Phi}^{l}_{\mathrm{image}} - \boldsymbol{\Phi}^{l}_{\mathrm{out}}) \|_2^2 \tag{5}$$

式中:$\boldsymbol{\Phi}^{l}_{\mathrm{image}}$ 和 $\boldsymbol{\Phi}^{l}_{\mathrm{out}}$ 是预训练网络第 l 层的特征图。

6) 初始分数(inception score,IS,式中简记 S_{IS})是一种用于评价生成对抗网络生成的图像质量和多样性的指标,能够衡量生成图像的逼真度和多样性。IS 首先使用预训练的 Inception V3 模型提取修复后的图像 I_{out} 与原始图像 I_{image} 的特征向量,然后,计算每张图像的预测类别分布及分布的多样性,最终,由预测类别分布的分散程度和均衡程度综合评估修复后图像 I_{out} 的质量和多样性。其计算公式为

$$S_{\mathrm{IS}} = \exp\left(\frac{1}{N}\sum_{i=1}^{N} D_{\mathrm{KL}}(p(\boldsymbol{y} | x^i) \| p(\boldsymbol{y}))\right) \tag{6}$$

式中:x 表示生成的图像;$\boldsymbol{y}$ 表示 Inception V3 模型提取的向量;N 表示生成图像的数量;i 表示生成图像数量变量。生成具有有意义对象的图像会导致条件标签分布 $p(y | x^i)$ 的熵低,生成具有不同对象的图像会导致边缘标签分布 $p(y)$ 的熵高,因此,根据 KL 散度,IS 越高图像质量越好。

7)Fréchet 初始距离(Fréchet Inception distance,FID,式中简记 d_{FID})的计算方法基于深度特征的统计特性,通常使用预训练的 Inception

$V3$ 模型提取图像的特征,计算修复后的图像和原始图像特征的多维高斯分布,并测量这两个分布之间的相似程度。FID 越小,则表示这两个图像在特征空间中越接近,即修复后图像的质量越高。其计算公式为

$$d_{\mathrm{FID}} = \| \mu_x - \mu_y \|^2 + tr(\Sigma_x + \Sigma_y - 2(\Sigma_x \Sigma_y)^{\frac{1}{2}}) \tag{7}$$

式中:μ_x 表示原始图像的特征均值;μ_y 表示修复后图像的特征均值;$tr(\cdot)$ 表示矩阵的迹。与 Inception Score 相比,FID 更加关注图像在特征空间中的分布,因此,可以更准确地捕捉图像的质量和多样性。然而,FID 需要复杂的计算量,因为涉及到图像的特征统计分析。

结合上述质量评价指标,统计了一些代表性方法的部分实验结果,如表 1 所示,展示了不同图像修复方法在常用数据集上的性能对比结果。其中,“↑”表示该评价指标值越大图像质量越好,“↓”表示该评价指标值越小图像质量越好,“–”表示文献中没有该评价指标的数值结果。

表 1　不同图像修复方法在常用数据集上的性能对比

Tab. 1　Performance comparison of various image inpainting methods on common datasets

数据集	方法	PSNR ↑	SSIM ↑	FID ↓	MAE ↓	LPIPS ↓	图像尺寸	掩膜类型
CelebA-HQ	PEPSI[19]	28.60	0.92	–	–	–	256 ×256	Irregular
	PRVS[30]	27.76	0.93	–	0.02	–	256 ×256	Irregular
	RePaint[65]	–	–	6.98	–	0.059	256 ×256	Irregular
	SDM[66]	–	–	4.05	–	0.052	512 ×512	Irregular
Place2	Yu, et al[18]	18.91	–	–	8.60	–	256 ×256	Irregular
	RFR[22]	22.62	0.81	–	0.038	–	256 ×256	Irregular
	HiFIll[25]	–	–	4.89	5.43	–	512 ×512	Irregular
	Ren, et al[31]	25.22	0.90	7.03	–	–	256 ×256	Irregular
Paris StreetView	Yang, et al[24]	17.59	–	–	–	–	128 ×128	Square(25%)
	PRVS[30]	26.44	0.86	–	0.027	–	256 ×256	Irregular
FFHQ	BAT-Fill[39]	22.59	0.759 2	–	0.036	0.143	256 ×256	Random
	ICT[87]	25.31	0.898 5	–	0.032	0.078	256 ×256	Random
	DGTS[59]	24.31	0.819 8	–	0.032	0.117	256 ×256	Random

4　结论与展望

图像修复作为底层任务,旨在从受损、缺失或受噪声干扰的图像中恢复出原始图像的信息,对于许多计算机高级视觉任务的成功具有关键意义。通过图像修复,可以有效地还原受损图像的内容和细节,使其恢复到更接近原始状态,从而为后续的视觉分析和应用提供更准确、可靠的数据基础。随着深度学习技术的飞速进步,涌现出大量新技术和新模型,从而推动基于深度学习的图像修复方法迈上蓬勃发展的阶段。这些方法通过应用新的模型架构、优化模型结构、采用先进的修复策略、先验信息等方面,取得了更加卓越的修复效果。然而,由于技术不断地迭代,这类方法在新技术方面的应用总结并未得到及时更新。因此,本文从修复策略的角度出发,尽可能全面地对基于深度学习的图像修复任务进行分类总结,概述了常用的图像修复数据集和质量评价指标,总结

了一些代表性方法的性能对比。在此基础上,针对该领域目前存在的难题和未来研究趋势做出以下展望。

1)如何根据不同的图像区域或损坏类型自适应地采用不同的修复策略,是一个亟需解决的难题。基于像素生成式的修复方法通常可以有效地处理噪声,通过预测像素值降低噪声的影响,但是每个像素独立地预测和生成,会导致修复结果在复杂纹理区域出现失真和模糊。渐进式图像修复策略以分阶段处理的方式在解决复杂损坏、提高修复质量以及充分利用信息传递方面具有明显的优势。然而,不同阶段之间的信息传递可能不够充分,导致前后阶段之间的一致性和连贯性下降,并且每个阶段的修复都可能引入一定的误差,这些误差可能会在后续阶段中积累,从而影响修复结果的质量。基于卷积感知的修复策略可以适应不同类型的损坏,但对于大范围的复杂损坏可能失效,并且在处理复杂纹理和结构时容易引入失真和模糊。基于调制的修复方法注重保留图像的结构和纹理特征等高频信息,修复结果通常更加自然和真实,不容易引入失真,但常常忽略低频信息,而在一些情况下,图像的低频信息也很重要。因此,现有的方法大都针对于不同类型的损坏选择不同的修复策略,难以兼顾所有损坏模式。如何实现根据图像的特点灵活地采用修复策略是一个值得深入研究的方向。

2)计算效率高、成本低的高分辨率图像修复模型是一个有待研究的热点问题。随着科技的发展和商业应用的扩展,高分辨率图像在医疗、安防、卫星图像等诸多领域的需求逐渐增加,并且随着数据的可获取性和多样性不断提升,为训练更高质量的图像修复模型创造了良好的环境。尽管一些高级的图像修复方法(如 Transformer 类)在高分辨率图像上的修复结果优异,但其计算成本和硬件要求难以在实际应用中推广。虽然 U-Net 类和 GAN 类修复方法在高分辨率图像修复中具有一定的潜力,但它们通常采用增加卷积层扩大感受野的方式使模型学习高层次的图像信息,不仅会导致模型的参数量和计算量增加,还会增加模型的复杂性,从而使模型更容易过拟合训练数据。因此,研究低计算成本的高分辨率图像修复的方法对更好地利用丰富的大规模数据具有重要意义。

3)研究一种能在多类型数据集上进行综合训练并合理修复各种类型图像的模型具有重要意义。目前基于网络架构改进的图像修复方法大多数都是针对某一类数据集进行训练,不仅适用范围受限,而且无法统一衡量方法的好坏。如果能实现利用单一模型在多类型数据集上进行综合训练,让不同类型的任务共享底层特征表示,将有助于模型更好地捕捉数据之间的共性和联系,避免了为每种数据类型设计和训练独立模型的重复工作。这不仅可以将共享的底层特征表示迁移到其他相关任务上,还简化了系统的架构和部署。

4)研究一种能够自动检测图像受损区域并根据图像类型进行合理修复的网络架构具有实际意义。现有的大多数图像修复算法都需要给网络中输入缺失区域的掩膜图像,以指导模型更精准地修复受损区域,而在实际场景中获取受损区域的掩膜图像是不现实的。虽然基于掩膜到图像的盲图像修复方法能够让模型尽可能地从输入图像中推断缺损的位置和特征,但是在自主识别缺损区域及修复内容合理性方面仍面临很大的挑战。

5)随着对自动化及自适应图像修复方法的需求增加,设计一种无参考的质量评价指标迫在眉睫。当前使用的客观质量评价指标 PSNR、SSIM、FID 等都属于全参考质量评价指标,必须采用未破损的原图作为参考图像,对修复后的图像进行对比计算。如果在未来实现了自动化及自适应的图像修复方法,就需要在符合人眼视觉判断的无参考质量评价指标方面进行深入研究。

参考文献

[1] JAM J, KENDRICK C, WALKER K, et al. A comprehensivereview of past and present image inpainting methods[J]. Computer Vision and Image Understanding, 2021, 203: 103147.

[2] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning internal representations by error propagation [M] // Readings in Cognitive Science. Amsterdam: Elsevier, 1988: 399-421.

[3] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C] // Proceedings of the 27th International Conference on Neural Information Processing Systems. New York: ACM, 2014: 2672-2680.

[4] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all You need[C] // Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.

[5] DHARIWAL P, NICHOL A. Diffusion models beat GANs on image synthesis[EB/OL]. (2021-06-01)[2023-05-25]. https://arxiv.org/abs/2105.05233.

[6] KÖHLER R, SCHULER C, SCHÖLKOPF B, et al. Mask-specific inpainting with deep neural networks[C]// German Conference on Pattern Recognition. Cham:Springer, 2014: 523-534.

[7] REN J S, XU L, YAN Q, et al. Shepard convolutional neural networks[M]// Advances in Neural Information Processing Systems (NIPS). San Francisco: Morgan Kaufmann, 2015.

[8] DAPOGNY A, CORD M, PÉREZ P. The missing data encoder: Cross-channel image completion with hide-and-seek adversarial network[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7):10688-10695.

[9] KOUTNÍK J, GREFF K, GOMEZ F, et al. A clockwork RNN[EB/OL]. (2014-02-14)[2023-05-25]. https://arxiv.org/abs/1402.3511.

[10] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4): 541-551.

[11] VAN OORD A, KALCHBRENNER N, KAVUKCUOGLU K. Pixel recurrent neural networks[C]// International Conference on Machine Learning. New York: PMLR, 2016: 1747-1756.

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

[13] VAN DEN OORD A, KALCHBRENNER N, VINYALS O, et al. Conditional image generation with PixelCNN decoders[C]// Proceedings of the 30th International Conference on Neural Information Processing Systems. New York:ACM, 2016:4797-4805.

[14] SALIMANS T, KARPATHY A, CHEN X, et al. PixelCNN++: Improving the PixelCNN with discretized logistic mixture likelihood and other modifications[EB/OL]. (2017-01-19)[2023-05-25]. https://arxiv.org/abs/1701.05517.

[15] OLIVEIRA M M, BOWEN B, MCKENNA R, et al. Fast digital image inpainting[C]// Proceedings of the International Conference on Visualization, Imaging and Image Processing (VIIP 2001), Marbella: [s.n.], 2001:106-107.

[16] HADHOUD M M, MOUSTAFA K A, SHENODA S Z. Digital images inpainting using modified convolution based method[C]// Proceedings SPIE 7340, Optical Pattern Recognition XX. Orlando: SPIE, 2009, 7340: 234-240.

[17] JAIN V, SEUNG S. Natural image denoising with convolutional networks[C]// Advances in Neural Information Processing Systems. Spain: Curran Associates, Inc, 2008: 769-776.

[18] 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 (CVPR). Salt Lake City:IEEE,2018: 5505-5514.

[19] SAGONG M C, SHIN Y G, KIM S W, et al. PEPSI: Fast image inpainting with parallel decoding network[C]//2919 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2020:11352-11360.

[20] MA Y Q, LIU X L, BAI S H, et al. Coarse-to-fine image inpainting via region-wise convolutions and non-local correlation[C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao:IJCAI,2019: 3123-3129.

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

[22] 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). Seattle:IEEE,2020: 7757-7765.

[23] ZENG Y, LIN Z, YANG J M, et al. High-resolution image inpainting with iterative confidence feedback and guided upsampling[C]// European Conference on Computer Vision (ECCV). Cham:Springer,2020:1-17.

[24] 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 (CVPR). Honolulu: IEEE, 2017: 4076-4084.

[25] YI Z L, TANG Q, AZIZI S, et al. Contextual residual aggregation for ultrahigh-resolution image inpainting[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle:IEEE, 2020:7505-7514.

[26] KULSHRESHTHA P, PUGH B, JIDDI S. Feature refinement to improve high resolution image inpainting[EB/OL]. (2002-06-29)[2023-05-25]. https://arxiv.org/abs/2206.13644.

[27] LIU W H, CUN X D, PUN C M, et al. CoordFill: Ef-

ficient high-resolution image inpainting via parameterized coordinate querying [EB/OL]. (2023-03-15) [2023-05-25]. https://arxiv.org/abs/2303.08524.

[28] LIAO L, HU R M, XIAO J, et al. Edge-aware context encoder for image inpainting[C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary: IEEE, 2018: 3156-3160.

[29] NAZERI K, NG E, JOSEPH T, et al. EdgeConnect: Structure guided image inpainting using edge prediction [C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul: IEEE, 2019:3265-3274.

[30] 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). New York: IEEE, 2020: 5961-5970.

[31] 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). New York: IEEE, 2020: 181-190.

[32] DENG X C, YU Y. Ancient mural inpainting via structure information guided two-branch model[J]. Heritage Science, 2023, 11(1): 1-17.

[33] YANG J E, QI Z Q, SHI Y. Learning to incorporate structure knowledge for image inpainting[J]. Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2020,34(7):12605-12612.

[34] SONG Y H, YANG C, SHEN Y J, et al. SPG-net: Segmentation prediction and guidance network for image inpainting [EB/OL]. (2018-08-06) [2023-05-25]. https://arxiv.org/abs/1805.03356.

[35] YU T, FENG R S, FENG R Y, et al. Inpaint anything: Segment anything meets image inpainting[EB/OL]. (2023-04-13) [2023-05-25]. https://arxiv.org/abs/2304.06790.

[36] LIU Y, PAN J S, SU Z X. Deep blind image inpainting[M]//CUI Z, PAN J S, ZHANG S S, et al., Eds. Intelligence Science and Big Data Engineering. Visual Data Engineering. Cham: Springer International Publishing, 2019:128-141.

[37] WANG Y, CHEN Y C, TAO X, et al. VCNET: A robust approach to blind image inpainting[C]//European Conference on Computer Vision. Cham: Springer, 2020: 752-768.

[38] PHUTKE S S, KULKARNI A, VIPPARTHI S K, et al. Blind image inpainting via omni-dimensional gated attention and wavelet queries[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Vancouver: IEEE, 2023: 1251-1260.

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

[40] CHEN M, ZHAO X D, XU D Q. Image inpainting for digital Dunhuang murals using partial convolutions and sliding window method[J]. Journal of Physics: Conference Series, 2019, 1302(3): 032040.

[41] WANG N Y, WANG W L, HU W J, et al. Thanka mural inpainting based on multi-scale adaptive partial convolution and stroke-like mask[J]. IEEE Transactions on Image Processing, 2021, 30: 3720-3733.

[42] 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). Seoul:IEEE,2019:4470-4479.

[43] CHANG Y L, LIU Z Y, LEE K Y, et al. Free-form video inpainting with 3D gated convolution and temporal PatchGAN[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Sesoul: IEEE,2020: 9065-9074.

[44] LI H A, WANG G Y, GAO K, et al. A gated convolution and self-attention-based pyramid image inpainting network[J]. Journal of Circuits, Systems and Computers, 2022, 31(12): 2250208.

[45] XIE K, GAO L G, ZHANG H, et al. Inpainting truncated areas of CT images based on generative adversarial networks with gated convolution for radiotherapy [J]. Medical & Biological Engineering & Computing, 2023,61(7): 1757-1772.

[46] MA X X, DENG Y B, ZHANG L, et al. A novel generative image inpainting model with dense gated convolutional network[J]. International Journal of Computers Communications & Control, 2023, 18(2):1-18.

[47] 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). Seoul:IEEE,2020: 8857-8866.

[48] GUO X F, YANG H Y, HUANG D. Image inpainting via conditional texture and structure dual generation [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2022: 14114-14123.

[49] ZHAO S Y, CUI J, SHENG Y L, et al. Large scale image completion via co-modulated generative adversarial networks[EB/OL]. (2021-03-18)[2023-05-25]. https://arxiv.org/abs/2103.10428.

[50] ZHENG H T, LIN Z, LU J W, et al. Image inpainting with cascaded modulation GAN and object-aware training[C]// AVIDAN S, BROSTOW G, CISSÉ M, et al. European Conference on Computer Vision. Cham: Springer, 2022: 277-296.

[51] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all You need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.

[52] WAN Z Y, ZHANG J B, CHEN D D, et al. High-fidelity pluralistic image completion with transformers[C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2022: 4672-4681.

[53] ZHOU Y Q, BARNES C, SHECHTMAN E, et al. TransFill: Reference-guided image inpainting by merging multiple color and spatial transformations[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Nashville:IEEE,2021: 2266-2267.

[54] WANG J K, CHEN S X, WU Z X, et al. FT-TDR: Frequency-guided transformer and top-down refinement network for blind face inpainting[J]. IEEE Transactions on Multimedia, 2023,25:2382-2392.

[55] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[EB/OL]. (2021-06-03)[2023-05-25]. https://arxiv.org/abs/2010.11929.

[56] CAO C J, DONG Q L, FU Y M. Learning prior feature and attention enhanced image inpainting[C]// AVIDAN S, BROSTOW G, CISSÉ M, et al. European Conference on Computer Vision. Cham: Springer, 2022: 306-322.

[57] YU Y S, DU D W, ZHANG L B, et al. Unbiased multi-modality guidance for image inpainting[C]// AVIDAN S, BROSTOW G, CISSÉ M, et al. European Conference on Computer Vision. Cham: Springer, 2022: 668-684.

[58] LIU H P, WANG Y, WANG M, et al. Delving globally into texture and structure for image inpainting[C]// Proceedings of the 30th ACM International Conference on Multimedia. New York: ACM,2022: 1270-1278.

[59] CHEN B L, LIU T J, LIU K H. Lightweight image inpainting by stripe window transformer with joint attention to CNN[EB/OL]. (2023-01-02)[2023-05-25]. https://arxiv.org/abs/2301.00553.

[60] NADERI M, GIVKASHI M H, KARIMI N, et al. SFI-swin: Symmetric face inpainting with swin transformer by distinctly learning face components distributions[EB/OL]. (2023-01-09)[2023-05-25]. https://arxiv.org/abs/2301.03130.

[61] LIAO L, LIU T R, CHEN D L, et al. TransRef: Multi-scale reference embedding transformer for reference-guided image inpainting[EB/OL]. (2023-06-20)[2023-05-25]. https://arxiv.org/abs/2306.11528.

[62] DHARIWAL P, NICHOL A. Diffusion models beat GANs on image synthesis[EB/OL]. (2021-06-01)[2023-05-25]. https://arxiv.org/abs/2105.05233.pdf.

[63] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[M]// Advances in neural Information Processing Systems. San Francisco: Margan Kaufmann, 2020.

[64] LUGMAYR A, DANELLJAN M, ROMERO A, et al. RePaint: Inpainting using denoising diffusion probabilistic models[EB/OL]. (2022-08-31)[2023-07-25]. https://arxiv.org/abs/2201.09865.

[65] LI W B, YU X, ZHOU K, et al. SDM: Spatial diffusion model for large hole image inpainting[EB/OL]. (2023-03-08)[2023-07-25]. https://arxiv.org/abs/2212.02963.

[66] HORITA D, YANG J L, CHEN D, et al. A structure-guided diffusion model for large-hole diverse image completion[EB/OL]. (2022-11-18)[2023-07-25]. https://arxiv.org/abs/2211.10437.

[67] GRILL J B, STRUB F, ALTCHÉ F, et al. Bootstrap your own latent: A new approach to self-supervised Learning[EB/OL]. (2020-09-10)[2023-05-25]. https://arxiv.org/abs/2006.07733.

[68] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022: 10674-10685.

[69] ISKAKOV K. Semi-parametric image inpainting[EB/OL]. (2018-11-13)[2023-07-25]. https://arxiv.org/abs/1807.02855.

[70] XIONG W, YU J H, LIN Z, et al. Foreground-aware image inpainting[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Long Beach:IEEE,2020: 5833-5841.

[71] YUVAL N. Reading digits in natural images with unsupervised feature learning[C]// Proceedings of the NIPS Workshop on Deep Learning and Unsupervised Feature Learning. Granada:NIPS Foundation, 2011.

[72] DOERSCH C, SINGH S, GUPTA A, et al. What makes Paris look like Paris? [J]. ACM Transactions on Graphics, 2012, 31(4):1-9.

[73] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas:IEEE,2016: 3213-3223.

[74] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context[C]// European Conference on Computer Vision. Cham: Springer, 2014: 740-755.

[75] 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.

[76] 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.

[77] LE V, BRANDT J, LIN Z, et al. Interactive facial feature localization[C]// European Conference on Computer Vision. Berlin, Heidelberg: Springer, 2012: 679-692.

[78] 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(ICCV). Santiago:IEEE,2016: 3730-3738.

[79] KARRAS T, AILA T M, LAINE S, et al. Progressive growing of GANs for improved quality, stability, and variation[EB/OL]. (2018-02-26)[2023-05-25]. https://arxiv.org/abs/1710.10196.

[80] KARRAS T, LAINE S, AILA T M. A style-based generator architecture for generative adversarial networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach: IEEE,2020:4396-4405.

[81] TYLECEK R, ŠÁRA R. Spatial pattern templates for recognition of objects with regular structure[C]// German Conference 35th on Pattern Recognition. Berlin, Heidelberg: Springer , 2013: 364-374.

[82] CIMPOI M, MAJI S, KOKKINOS I, et al. Describing textures in the wild[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus:IEEE, 2014:3606-3613.

[83] KRAUSE J, STARK M, JIA D, et al. 3D object representations for fine-grained categorization[C]//2013 IEEE International Conference on Computer Vision Workshops (ICCVW). Sydney:IEEE,2014:554-561.

[84] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: From error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[85] SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved techniques for training GANs[C]// Proceedings of the 30th International Conference on Neural Information Processing Systems. New York: ACM, 2016:2234-2242.

[86] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. GANs trained by a two time-scale update rule converge to a local Nash equilibrium[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. New York:ACM,2017:6629-6640.

[87] LOSSON O, MACAIRE L, YANG Y. Comparison of color demosaicing methods[M]// Advances in Imaging and Electron Physics. Amsterdam: Elsevier, 2010, 162: 173-265.

[88] HACCIUS C, HERFET T. Computer vision performance and image quality metrics: A reciprocal relation [C]// Computer Science & Information Technology (CS & IT). Florence: Academy & Industry Research Collaboration Center (AIRCC), 2017: 27-37.

[89] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 586-595.

(编 辑 李 静)

作者简介:

彭进业,男,1964 年生,博士,二级教授、博士生导师,文化遗产数字化保护与传播教育部创新团队负责人(带头人),文化遗产数字化保护陕西省三秦学者创新团队负责人(带头人),教育部新世纪优秀人才,陕西省三秦学者,陕西省教学名师,全国高校黄大年式教师团队核心成员(排名第二)。现任西北大学信息科学与技术学院院长、软件学院院长,新型网络智能信息服务国家地方联合工程研究中心常务副主任,文化遗产数字化国家地方联合工程研究中心副主任,陕西省面向领域应用的人工智能技术学科创新引智基地主任,陕西省丝绸之路文化遗产数字化保护与传承协同创新中心主任。兼任中国计算机自动测量与控制技术协会理事、陕西省图象图形学学会副理事长等职。先后担任 IoTaas 2020、CBD 2021、ICIPMC 2022/2023 等多个国际学术会议主席。从事智能信息处理、文物数字化保护技术等方面的研究与教学工作。主持国家重点研发课题、国家自然科学基金等科研项目 20 多项。曾在日本国立 Toyohashi 科技大学作访问学者。在 IEEE TIP、TMM、TKDE、TITB、*Journal of Cultural Heritage*、《中国科学》等国内外刊物及 CVPR、WWW、IJCAI 等重要国际学术会议上发表一批学术论文,获授权发明专利 20 多项,其中转化应用 7 项,获国家教学成果二等奖、陕西省科学技术二等奖等教学科研奖励。