基于双判别器的 Wasserstein 生成对抗网络图像修复

李筱玉,张 乾,徐开丽,骆 迪,冉娅琴
贵州民族大学 数据科学与信息工程学院,贵阳 550025

摘 要:目的 针对基于生成对抗网络的图像修复模型,在修复不规则大面积缺损图像时,存在细节特征信息还原不准确、图像视觉效果欠佳及训练不稳定等缺陷问题,提出了一种基于双判别器的 Wasserstein 生成对抗网络图像修复模型。**方法** 该方法首先以编码器-解码器结构的卷积神经网络作为生成器,在生成器的编码器和解码器之间添加跳跃连接,以更好地学习图像细微特征并提升最终的修复效果;接着在全局判别器基础上引入局部判别网络,以保证局部修复结果与周围区域的一致性,并在判别器中引入 Wasserstein 距离,从而使网络的训练更加稳定以获得较为自然的图像修复效果;最后设计 VGG16 特征提取器以引入感知损失和风格损失,通过添加多个损失函数来提升图像复原效果。**结果** 在公开人脸、场景和街景数据集上进行对比分析,验证出所提出的方法定性和定量分析结果均优于对比模型,其性能评价指标值更高。**结论** 实验结果表明:所提出的方法在修复不规则大面积缺损区域的图像时,修复图像更清新,视觉效果更好。

关键词:图像修复;生成对抗网络;跳跃连接;VGG-16 特征提取模型;Wasserstein 距离
中图分类号:TP391.41 **文献标识码:**A doi:10.16055/j.issn.1672-058X.2025.0004.002

Image Inpainting Using Wasserstein Generative Adversarial Networks with Dual Discriminators

LI Xiaoyu, ZHANG Qian, XU Kaili, LUO Di, RAN Yaqin
School of Data Science and Information Engineering, Guizhou Minzu University, Guiyang 550025, China

Abstract: Objective Image inpainting models based on generative adversarial networks (GANs) suffer from deficiencies when repairing irregular and extensively damaged images, including inaccurate restoration of fine image details, unsatisfactory visual effects, and training instability. In this regard, a Wasserstein GAN-based image inpainting model with dual discriminators was proposed. **Methods** This method employed a convolutional neural network (CNN) with an encoder-decoder architecture as the generator, with skip connections added between the encoder and decoder of the generator to better learn subtle image features and enhance the final inpainting results. Additionally, a local discriminator network was introduced on the basis of the global discriminator to ensure consistency between locally restored areas and their surroundings. Wasserstein distance was incorporated into the discriminators to stabilize the training process and achieve more natural image inpainting. Furthermore, a VGG16 feature extractor was designed to introduce perceptual and style losses to improve image inpainting by incorporating multiple loss functions. **Results** Comparative analysis on public datasets of faces, scenes, and streets showed qualitative and quantitative superiority of the proposed method over baseline models, with higher performance evaluation metrics. **Conclusion** Experimental results demonstrate that the proposed method presents better visual effects and clearer restoration when repairing images with irregular and extensive damaged areas.

Keywords: image inpainting; generative adversarial networks; skip connections; VGG-16 feature extractor; Wasserstein distance

收稿日期:2023-10-24 **修回日期:**2023-12-14 **文章编号:**1672-058X(2025)04-0009-08
基金项目:贵州民族大学校级科研项目(GZMUZK[2021]YB23);黔教合(YJSKYJJ[2021]121).
作者简介:李筱玉(1995—),女,贵州威宁人,硕士研究生,从事图像修复研究.
通信作者:张乾(1984—),男,贵州贵定人,教授,博士,从事机器学习与模式识别研究. Email:331394262@qq.com.
引用格式:李筱玉,张乾,徐开丽,等. 基于双判别器的 Wasserstein 生成对抗网络图像修复[J]. 重庆工商大学学报(自然科学版),2025,42(4):9-16.
LI Xiaoyu, ZHANG Qian, XU Kaili, et al. Image inpainting using Wasserstein generative adversarial networks with dual discriminators[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(4): 9-16.

1 引 言

图像修复[1]指运用图像的已知信息恢复缺损区域的像素特征。其目的在于根据待修复图像已知区域的内容,通过合适的算法预测缺损区域的内容,得到最终的复原图像。其被广泛应用于计算机视觉领域,例如缺损文物修复、生物医学图像修复和公安办案过程中的人脸修复等。在信息化时代,图像已成为人们获取外界信息的重要途径,因此对缺损图像进行修复具有重大的意义和价值。

在深度学习之前,非学习图像修复算法大致分为基于扩散的方法[1-3]和基于样本的方法[4-6]。基于扩散的方法能将待修补区域的边缘以原有角度延伸到区域里面而修复图像。然而,由于基于扩散的方法不考虑全局图像结构,因此,只能有效填补小洞,在处理大面积缺损时效果较差。为了解决基于扩散修复方法的缺点,基于样本的方法从整个图像的已知区域中搜索有效信息,并将该信息复制或重新定位到缺损位置。尽管基于样本的算法在面对较大区域的简单图像缺损时表现良好,但在修复不规则大面积缺损图像时,因为对图像的语义信息无法理解,致使修复图像的视觉效果不好。

近年来,随着生成对抗网络(Generative Adversarial Networks, GANs)[7]和卷积神经网络(Convolutional Neural Network,CNNs)[8]在计算机视觉领域取得的显著成效,基于学习的方法在图像修复领域表现出极佳的优势。Pathak 等[9]基于 Goodfellow 在 GAN[8]方面的工作,引入了采用条件 GAN 的上下文编码器(Context Encoder,CE),使用重建和对抗损失从场景周围预测场景的缺损部分,能够捕捉图像语义和全局结构,但该方法只能修复单一形状的缺损区域,修复自然背景、纹理复杂的真实图像时效果不佳;Lizuka 等[10]提出 GLCICA(Globally and Locally Consistent Image Completion)方法,在修复网络中采用堆叠扩张卷积来捕获图像更大范围的信息特征,此外,引入了一个额外的判别器来确保图像的局部一致性,该方法可以修复任意形状的缺损图像,整体修复效果较好,但在修复高分辨率图像时存在修复图像部分细节失真现象;Yu 等[11]提出一种基于深度生成模型的图像修复方法,利用上下文关注模块(Contextual Attention Module,CAM)明确地利用周围的图像特征作为参考,该方法使用两个堆叠的生成网络(粗网络和精网络)生成修复图像,并使用 CAM 的精网络对中间图像进行精处理以生成最终的修复结果,但忽略了图像缺损区域内部特征的关联性,从而被修复的图像会出现断层现象;Yan 等[12]提出一个名为 Shift Net 的新网络,该网络通过深度特征重排提供图像修复,可以修复任意形状的缺损图像且在较短的时间内获得较为精细的修复图像,但随着缺损区域的增大,其修复性能急剧下降;Liu 等[13]提出使用堆叠的部分卷积运算和掩膜更新步骤来执行图像修复,能够有效学习图像缺损区域特征之间的语义信息,更好地保留图像结构,但该方法难以学习图像缺损区域和已知区域间的对应关系,导致修复图像部分细节不清晰;Li 等[14]提出一种即插即用的递归特征推理和知识一致注意力模块网络结构,可以对掩膜面积较大且形状不规则的图片进行修复,且修复后的图像语义连续,细节清晰,但该方法由于要反复进行递归,每一次只能修复一层掩膜区域,所以会导致运算时间过长;Yu 等[15]提出区域归一化(Region Normalization, RN),根据区域掩码将特征图简单地分为两个区域,即未损坏区域和损坏区域,并分别对它们进行归一化,以避免孔中错误信息的影响,提高图像的修复效果,但在修复过程中 RN 模块难以追踪待修复图像潜在的缺损区域,这会使得修复后的图像存在模糊现象;Li 等[16]提出一种新的图像修复深度生成模型,通过信道和空间自适应反归一化(CSA-RN)模块,可以有针对性地缓解每个信道中的空间均值和方差偏移,基于选择性潜在空间映射的上下文注意力(SLSM-CA)层可以选择性地利用多尺度背景信息来提高预测质量,但该方法优化网络在训练过程中不稳定,修复复杂背景图像时视觉效果欠佳。

这些深度修复模型在图像生成问题上具有显著效果,通过在足够大的数据集上进行训练,卷积神经网络在学习丰富的模式和图像语义以及用这些学习到的知识填充目标区域方面表现优越,空间中的稀疏连通性和参数共享使其具有计算效率。然而,在修复大面积不规则缺损图像时,仍然存在一些局限性,修复后的图像部分精细特征无法被还原,存在视觉效果欠佳的问题。

针对以上问题,本文以生成对抗模型为基础架构,提出基于双判别器的 Wasserstein 生成对抗网络图像修复模型。首先,以生成对抗模型为基础架构,在传统生成器中添加跳跃连接,可以使生成图像能够更好地融合细节特征;其次,在全局判别器的基础上引入局部判别,使用训练过程中更为稳定的 Wasserstein 距离(Wasserstein GAN,WGAN)[19]能够有效解决原始 GAN 因使用 KL-divergence 和 JS-divergence 出现的训练不稳定、模式崩溃等情况,加快模型的收敛速度,以保证局部修复结果与周围区域的一致性;最后,增设 VGG16 特征提取模型,该网络已经在大规模的数据集上进行

了训练,可以提取图像的高级语义和纹理特征,从而使得生成图像与原图像,在纹理结构、视觉效果方面更符合人眼对图像质量的感受。

2　相关理论

2.1　生成对抗网络

生成对抗网络(Generative Adversarial Networks, GAN)[8]是 2014 年被 Ian Goodfellow 提出的深度学习模型。其核心思想是运用相互竞争的深度神经网络(生成器和判别器)解决无监督学习所对应的问题。其中,生成器通常接收服从某种概率分布的随机向量作为输入,而后通过一系列层进行变换,并随着训练的进行,逐渐输出接近真实数据的样本,使判别器难以区分。判别器则接收或为真实数据、或为生成器生成数据的样本,而后判断输入样本是隶属于真实数据集合或是生成器生成数据的集合,并随着训练的进行,逐渐提升其判断准确率。生成器通过生成更逼真的样本来试图欺骗判别器,使其无法准确判断数据的真实性,而判别器则通过区分真实样本和生成样本,尽可能进行准确识别,二者相互对抗,通过计算逐步优化网络参数,提升生成样本的质量。图 1 为 GAN 的训练过程。

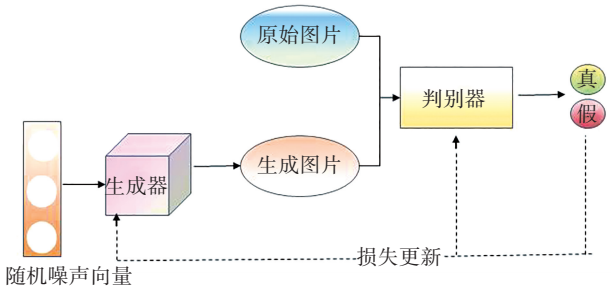

图 1　生成对抗网络模型

Fig. 1　Generative adversarial network model

2.2　WGAN 方法

在生成对抗网络中,度量原始图像和修复后图像两者分布之间的距离时,最常见的方法是 KL-divergence 和 JS-divergence。但是,这两个方法所度量的距离仍然存在以下问题:其一,对于任意两个概率分布 $P(x)$ 和 $Q(y)$,由于 $D_{\mathrm{KL}}(P \parallel Q) \neq D_{\mathrm{KL}}(Q \parallel P)$,所以 KL-divergence 度量的距离并不满足实际意义上距离的定义。为解决 KL-divergence 的不对称性问题,提出了具有对称性质的 JS-divergence 度量两个分布之间的距离,因此在众多机器学习算法中摒弃了 KL-divergence 的度量,转而采用 JS-divergence 的度量方法;其二,从距离的取值范围来看,$D_{\mathrm{JS}}(P \parallel Q) \in [0, \log 2]$,当 $P(x)=Q(y)$ 时,$D_{\mathrm{JS}}(P \parallel Q)=0$,JS-divergence 确实是度量两个分布之间距离的可行方法。但是,实际并不如此,因为在机器学习中,由于多维数据所带来的维度问题,使得分布 $P(x)$ 与分布 $Q(y)$ 在概率密度空间上并不总是有重合区域。在 $P(x)$ 和 $Q(y)$ 没有重合区域时,$D_{\mathrm{JS}}(P \parallel Q) \equiv \log 2$,这对于使用随机梯度下降法来优化目标函数是致命的,其主要原因是恒等于常数 log 2 的目标函数所提供的梯度计算恒等于 0,从而导致"梯度消失"现象,因此模型参数无法进行更新。为了度量两个分布之间的差异,同时,解决 JS- divergence 度量距离的"梯度消失"现象,基于 Wasserstein 距离估计两个分布之间距离的方法被提出[17]。

Wasserstein 距离定义:设 $\prod(P,Q)$ 为两个概率分布 $P(x)$ 和 $Q(y)$ 组合的所有可能的联合概率分布的集合,对集合中的任意联合概率分布 $\gamma(x,y)$,样本对 $(x,y) \sim \gamma$ 分布的距离 $d(x,y)$ 定义如下:

$$W(P,Q) = \inf_{\gamma \sim \prod(P,Q)} \iint \gamma(x,y)\, d(x,y) = \inf_{\gamma \sim \prod(P,Q)} E_{(x,y) \sim \gamma}[d(x,y)] \tag{1}$$

其中,$d(x,y)$ 是成本函数。对于所有有可能的联合分布 γ,能够从中采样 $(x,y) \sim \gamma$ 获得一个样本 x 和 y 且计算其成本的 $d(x,y)$,进而可以计算在联合分布 γ 下,样本对成本的期望值 $E_{(x,y) \sim \gamma}[d(x,y)]$;最终,在所有可能的联合分布中对期望值取到的下界就是 Wasserstein 距离。因此在训练的时候,WGAN 可以保证梯度的有效性,从而能够有效解决原始 GAN 因使用 KL-divergence 和 JS-divergence 出现的训练不稳定、模式崩溃等情况。

3　本文方法

3.1　网络结构

本文提出的图像修复模型整体网络结如图 2 所示。该模型由生成器模块、VGG16 特征提取模型模块和双重判别器模块构成。首先,在生成器模块中以缺损图像作为输入从而输出修复图像,编码器负责缺损图像的特征提取过程,解码器用于将卷积层输出的特征图进行还原,同时在解码与编码过程中添加跳跃连接完成多尺度信息融合,进而提升修复图像效果;其次,引入 VGG16 特征提取模型用于协助生成器,实现生成器提取图像感知特征和风格特征的能力,使得修复图像细节特征信息还原准确;最后,分别将局部与全局真实图像和修复图像输入至局部和全局判别器中,将图像压缩成合适的特征向量,从而将全局和局部判别器的输出串联成单个的 2 048 维度向量,再通过全连接层处理,进而得到输入图像的真假判别分数。由于本文使用了 WGAN 来衡量两个分布之间的差距,所以在把图像输入到判别器后,对其进行独立的梯度惩罚。

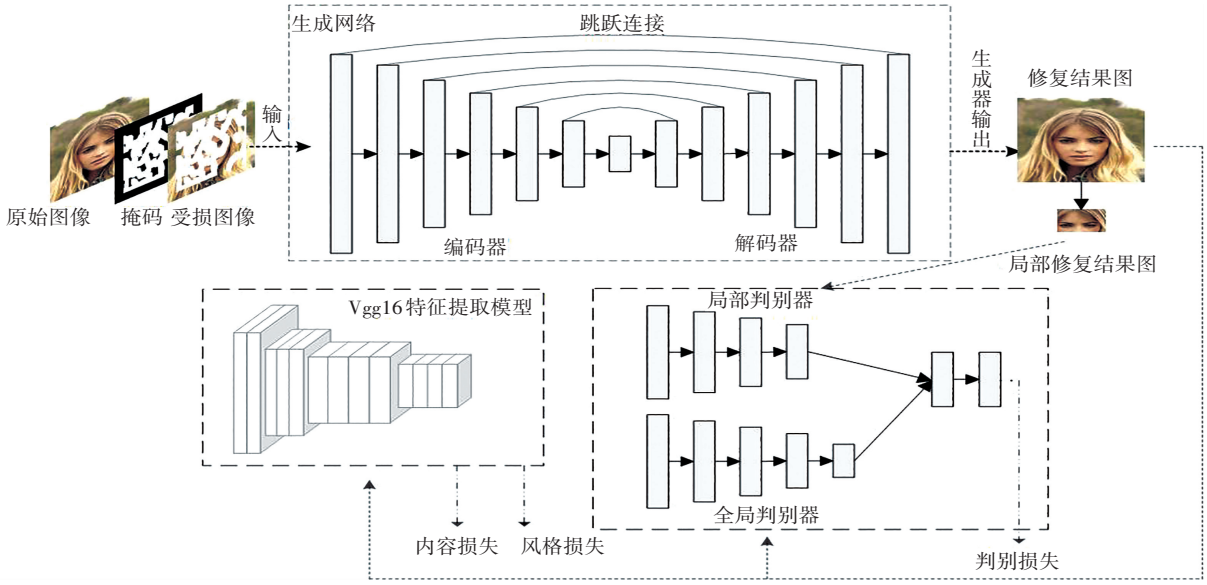

图 2 修复网络体系结构

Fig. 2 Inpainting network architecture

3.1.1 生成器

生成器是基于编码器-解码器结构的主干网络,模型结构如图 3 所示。生成器由卷积块和反卷积块构成。卷积块共 5 块设计,由卷积层、归一化层和 ReLU 激活层组成,用于对待修复图像缺损区域的特征进行提取;反卷积块共 6 块设计,由反卷积层、归一化层和 Leaky_ReLU 激活层构成,用于将卷积层输出的特征图进行还原。在编码和解码过程中使用不同的激活函数,解码时将 ReLU 激活函数更换为 Leaky_ReLU 激活函数的原因,主要是为了解决 ReLU 激活函数进入负区间后,导致的神经元不学习问题(当输入小于零时,导数为 0,导致无法更新权重)。Leaky_ReLU 激活函数由于导数总是不为零,减少了静默神经元的出现,允许基于梯度的学习,使得神经元在训练过程中能够保留一定的激活值。此外,对编码过程和解码过程进行跳跃连接,从而使得在信息传递过程中保留低层的特征,更好地学习到细微的特征以提升最终的修复效果。

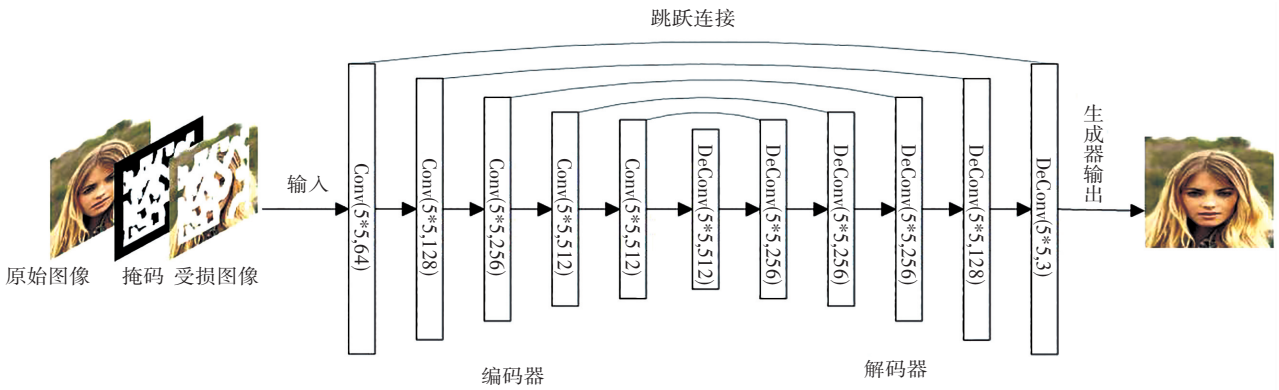

图 3 生成器网络结构图

Fig. 3 Structure of the generator network

3.1.2 判别器

在此使用双重判别器,分别从整体和局部两个方面对修复图像进行判别,保证了修复图像整体结构细节特征信息的还原。判别器结构如图 4 所示。分别将局部与全局真实图像和修复图像输入至局部和全局判别器中,将图像压缩成合适的特征向量,从而将全局和局部判别器的输出串联成单个的 2 048 维度向量。该向量由一个全连接层进行处理,输出一个连续的值。全连接层使用 Sigmoid 激活函数使输出值在[0,1]范围内,该值表示图像是真实的,而不是修复的概率。全局判别器网络将整个图片重新缩放到 256×256 作为输入,其由 5 块卷积块和全连接层构成。局部判别器遵循和全局判别器类似的配置,但输入是已修复图像中间区域的 128×128 像素块,由于初始输入分辨率为全局判别器的一半,因此不需要使用全局判别器的第一层;输出为 1 024 维向量,表示已修复区域周围信息,由 4 块卷积块和全连接层构成。卷积块由卷积核大小为 5、步长为 2 的卷积层、层归一化及其斜率为 0.2 的

LeakyRelu 激活层构成。此外,使用 WGAN 来衡量两个分布之间的差距,因为它在解决众所周知的训练不稳定问题方面被证明是有效的。

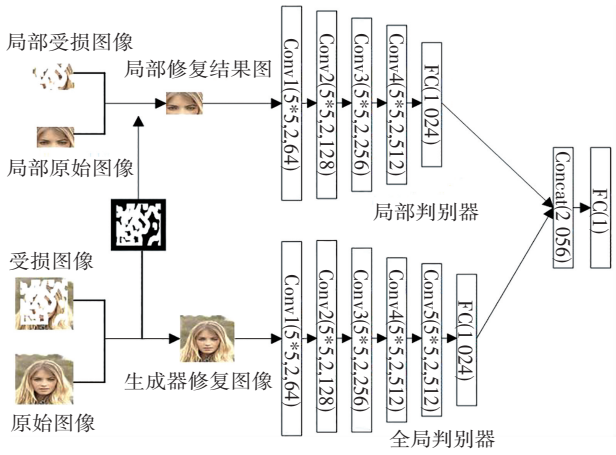

图 4　判别器结构

Fig. 4　Discriminator structure

3.1.3　VGG16 特征提取模型

VGG16 特征提取模型通过向原始生成器损失函数中添加感知和风格损失项,以协助生成器生成更优的图像修复效果,是目前计算机视觉领域中,应用较为广泛的卷积神经网络模型之一。其采用更深、更小的卷积核,最大限度地增加了模型的深度,使得模型更具有提取图像特征的能力。它包括 16 个卷积层、5 个池化层和 3 个全连接层。它将输入的 RGB 彩色图像进行卷积、池化和全连接等操作,最终输出结果。其中,卷积层使用 ReLU 激活函数,可以增强模型非线性特征提取的能力,还设置了一个 Padding = 1 的参数,该参数可以使图像尺寸不变;最大池化层通过保留最大特征点来进行降维,这可以避免过拟合情况的发生;全连接层将卷积层得到的特征信息进行分类。其模型没有使用特殊的随机初始化技术,因此性能较为稳定,结果容易复现,且模型参数较少,可以在较短时间内完成数据训练。其模型结构如图 5 所示。

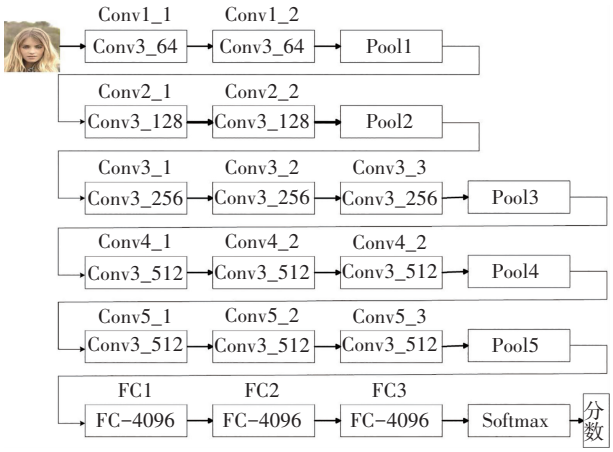

图 5　VGG16 特征提取模型

Fig. 5　VGG16 feature extraction model

3.2　损失函数

为了使修复图像有更清晰的视觉效果,损失函数分为鉴别器损失函数和修复损失函数两部分,下面将分别进行介绍。

鉴别器损失函数:

$$L_{\mathrm{d}}=D(I_{\mathrm{gt}})-D(I_{\mathrm{out}})+\gamma(\parallel\nabla_{O}(I_{s})D(I_{\mathrm{out}})\parallel_{2}-1)^{2} \tag{2}$$

式(2)中,I_{gt}表示真实图像,I_{out} 表示修复图像,γ 为梯度惩罚系数。

感知损失[18],通过预训练的神经网络来计算真实图像和生成图像之间的特征差异,可以提取图像的高层结构特征。感知损失为

$$L_{\mathrm{per}}=\frac{1}{N}\sum_{j=1}^{N}\frac{\parallel\varphi_{j}(I_{\mathrm{gt}})-\varphi_{j}(I_{\mathrm{out}})\parallel_{1}}{C_{j}\times H_{j}\times W_{j}} \tag{3}$$

式(3)中,φ_{j} 表示预训练 VGG16 网络中的卷积层提取出生成图像和真实图像的特征,$C_{j}\times H_{j}\times W_{j}$ 表示特征图的大小。通过预训练出的网络来捕捉图像的特征可以得到更好的修复结果。

风格损失[19],通过平方误差函数衡量生成图像与真实图像在风格上的差异,可以使模型能够从图像的背景中学习整体样式信息,有效解决感知损失缺乏保持风格一致性的问题。风格损失为

$$L_{\mathrm{style}}=\frac{1}{N}\sum_{j=1}^{N}\parallel\boldsymbol{M}_{\mathrm{G}}(\varphi_{j}(I_{\mathrm{gt}}))-\boldsymbol{M}_{\mathrm{GM}}(\varphi_{j}(I_{\mathrm{out}}))\parallel_{1} \tag{4}$$

式(4)中,$\boldsymbol{M}_{\mathrm{GM}}(\cdot)$表示计算伽马矩阵(GramMatrix)。

修复模型总的损失由式(5)所示:

$$L_{\mathrm{total}}=\lambda_{\mathrm{d}}L_{d}+\lambda_{s}\lambda_{\mathrm{style}}+\lambda_{p}L_{\mathrm{pre}} \tag{5}$$

式(5)中的 L_{total} 为整体损失目标函数,λ_{d}、λ_{s}、λ_{p} 为平衡不同损失的超参数。

4　仿真实验与结果分析

4.1　实验环境

本次实验在操作系统 Ubuntu20.04 下采用 Pytorch1.11.0 框架 CUDA11.0 进行训练和模型测试,实验平台配置是 GPU NVIDIA GeForce RTX 3080。

4.2　数据集

文中实验采用公开数据集 CelebA[20]、Places2[21]、Paris Street View[22] 和 Irregular masks 掩码集[13]。CelebA 数据集包含 10 177 个名人身份,202 599 张图像,在本次实验中随机选取 44 000 张图像,训练集 40 000 张图像,测试集 4 000 张图像。Places2 数据集包含超 1 000 万张图像和 400 多个不同类型的场景环境,实验随机选取 36 000 张图像,训练集 33 000 张图像,测试集 3 300 张图像。Pairs Street View 是街景数据集,该数据集包含 14 900 张训练图像和 100 张从巴黎街景中收集的测试图像,实验训练集 14 900 张图像,测

试集 100 张图像。Irregular masks 掩码集包含 12 000 个不规则掩码,每个掩码中的掩码面积占图像总大小的 0~60%。此外,根据掩码面积将不规则数据集划分为 3 个区间,即 0~20%、20%~40%和 40%~60%,其每个间隔时间都有 4 000 个不规则掩码。

4.3 评价指标

采用性能评价指标峰值信噪比(PSNR)和结构相似性指数(SSIM)定量评价修复图像的性能。PSNR 是常用的衡量信号失真的指标,其值越高修复图像的效果越好;SSIM 是量化修复图像和真实图像间结构相似性的指标,其值范围为 0~1,越大代表图像越相似。

PSNR 的公式定义如下所示:

$$R_{\mathrm{PSN}} = 10\ \ln\left(\frac{\mathrm{MAX}_I^2}{E_{\mathrm{MS}}}\right) \tag{6}$$

式(6)中,E_{MS} 为均方误差(Mean Squared Error),用来评估修复后的图像 C 与原始图像 O 之间像素级别上的差异程度;MAX_I 是像素最大值,通常为 255。

E_{MS} 公式定义如下:

$$E_{\mathrm{MS}} = \frac{1}{M \cdot N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}[O(i,j) - C(i,j)]^2 \tag{7}$$

式(7)中,$M \cdot N$ 为图像尺寸,O 表示真实图像,C 表示修复结果图,i 表示图像第 i 行的像素,j 表示图像第 j 列的像素。

SSIM 的公式定义如下所示:

$$I_{\mathrm{SSM}} = \frac{(2u_x u_y + c_1)(2\sigma_{xy} + c_2)}{(u_x^2 + u_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{8}$$

式(8)中,μ 表示均值,σ 表示标准差,c_1 和 c_2 表示常数。SSIM 的范围为 0~1,值越接近 1,说明图像越相似,修复效果越好。

4.4 定性分析

为了直观地验证本文算法的有效性,将本文提出的算法与对比算法在 CelebA、Places2、和 Paris Street View 数据集上进行比较,图 6 为 5 种算法修复的 9 幅图像视觉效果对比图。

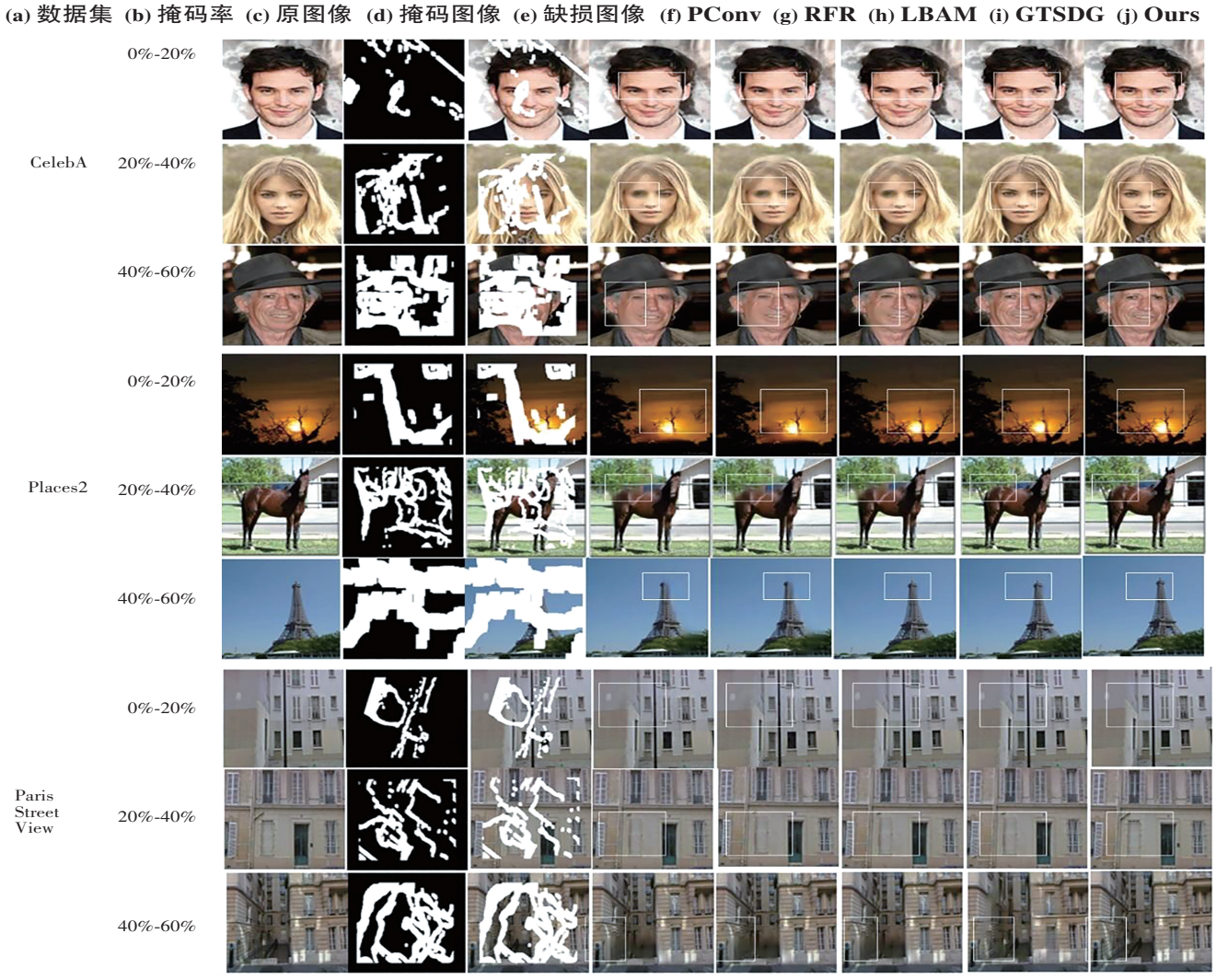

图 6 不同模型对不同数据集的图像修复结果的定性比较

Fig. 6 Qualitative comparison of image inpainting results on different datasets by different models

由图 6 可以看出:当掩码面积较小时,PConv[13]、RFR[14]、LBAM[23]和 GTSDG[24]和本文模型对受损图像的修复结果均取得了不错的效果。例如在第 1 行 CelebA 数据集上人脸图像修复、在第 4 行 Places2 场景数据集上日落图像的修复和在第 7 行 Paris Street View 数据集上墙壁的修复。但随着不规则掩码面积的增大,PConv、RFR 、LBAM 和 GTSDG 修复的图像与真实图像相比,存在部分纹理细节不够清晰的模糊现象。例如在对 CelebA 数据集上第 2 行和第 3 行人脸图像修复时,被修复的人脸图像在眼睛及皮肤纹理方面存在伪影和不精细,在对 Paris Street View 数据集第 9 行墙壁及其投射于墙壁的光影修复时,修复图像墙壁纹理结构和光影的视觉效果不够好。而本文方法的修复结果在视觉感知下,相较于对比方法,修复结果呈现的视觉效果更优。例如在第 3 行 CelebA 数据集上通过对比放大框里人脸图像皮肤纹理和眼睛的修复结果可以看出:采用本文方法呈现出更为自然的视觉效果,有效避免了对比方法进行修复出现的问题。在第 9 行的 Paris Street View 数据集上通过对比放大框里墙壁及其投射光影的修复结果可以看出:本文修复模型对投射于墙壁光影的修复效果更佳。通过不同数据集对比,证明了本文方法在修复随机不规则大面积缺损图像的优势,其修复图像的效果明显优于对比模型。

4.5　定量分析

为验证算法的有效性,在图像修复的定量分析中,采用评价指标峰值信噪比(PSNR)和结构相似性指数(SSIM)来评估修复图像的质量,通过比较修复图像和真实图像之间的相似度衡量所提出模型的性能。

表 1 所示为 5 种模型在公开街景(Paris Street View)、场景(Places2)和人脸(CelebA)数据集上,结合 3 种掩码率(0~20%、20%~40%、40%~60%)的定量分析结果。从表 1 可以看出:本文模型在掩码率较小时,与 PConv、RFR 、LBAM 和 GTSDG 的修复结果相比差距不大,主要是由于图像缺损区域的面积较小,生成图像的局部细节特征信息和区域模糊问题不明显,因此修复图像的效果视觉差异不大;但随着不规则掩码率的增大,本文所提模型性能评价指标高于对比模型,说明所提出模型相较于对比模型,修复图像视觉效果更好,这也证明了本文方法的可行性。

表 1　在 3 种数据集上使用不同方法对评价指标进行定量分析

Table 1　Quantitative analysis of evaluation indicators using different methods on the three datasets

		Paris StreetView			Places2			CelebA		
		0~20%	20%~40%	40%~60%	0~20%	20%~40%	40%~60%	0~20%	20%~40%	40%~60%
SSIM	PConv	0.953 6	0.843 6	0.726 3	0.942 1	0.823 6	0.693 6	0.952 1	0.851 2	0.731 3
	RFR	0.966 5	0.857 8	0.751 2	0.956 4	0.857 9	0.707 4	0.971 1	0.892 1	0.746 5
	LBAM	0.956 5	0.846 1	0.732 3	0.958 5	0.828 6	0.634 7	0.962 6	0.863 7	0.726 5
	GTSGD	0.970 3	0.862 4	0.743 8	0.961 5	0.872 3	0.684 5	0.978 9	0.889 5	0.758 5
	Ours	0.973 8	0.872 1	0.774 2	0.968 2	0.897 8	0.729 9	0.989 8	0.932 8	0.782 5
PSNR	PConv	31.699 5	25.321 1	21.101 9	30.341 2	24.372 3	21.256 5	34.621 2	25.341 4	21.429 3
	RFR	32.161 7	26.687 7	22.380 5	31.102 8	25.400 9	21.913 8	35.968 1	26.246 3	23.124 1
	LBAM	32.099 5	26.321 2	22.191 9	30.441 5	25.102 3	20.956 5	35.243 8	27.251 4	22.854 2
	CTSDG	32.864 2	26.893 5	22.874 9	31.462 1	25.841 2	22.142 1	35.421 2	27.854 2	23.789 6
	Ours	33.180 2	28.123 5	24.651 3	32.513 9	26.369 2	23.221 2	36.033 5	28.983 9	25.101 2

5　结　论

针对基于生成对抗网络的图像修复模型,在修复不规则大面积缺损图像时,存在修复图像细节特征信息还原不准确、图像视觉效果差及其训练不稳定等缺陷问题,提出一种基于双判别器的生成对抗模型图像修复算法。该模型以生成对抗网络为基础架构,通过向生成器的损失函数中添加感知和风格损失项,使得生成图像与原始图像更为相似,并在其中加入跳跃连接以提高生成模型对结构信息的预测能力;而在判别模型中使用 WGAN,由于它相对 KL 散度与 JS 散度具有优越的平滑特性,理论上可以解决梯度消失问题,从而使网络的训练更加稳定,以获得更高质量的修复图像。实验结果表明:本文模型从整体上还原出缺损图像的信息,有效提升了修复准确率,修复图像的效果优于对比模型。

参考文献(References):

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

[2] 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.

[3] TSCHUMPERLE D, DERICHE R. Vector-valued image regularization with PDEs: A common framework for different applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(4): 506-517.

[4] BARNES C, SHECHTMAN E, FINKELSTEIN A, et al. PatchMatch: A randomized correspondence algorithm for structural image editing[J]. ACM Trans Graph, 2009, 28(3): 24-32.

[5] HUANG J B, KANG S B, AHUJA N, et al. Image completion using planar structure guidance[J]. ACM Transactions on Graphics (TOG), 2014, 33(4): 1-10.

[6] DING D, RAM S, RODRIGUEZ J J. Image inpainting using nonlocal texture matching and nonlinear filtering [J]. IEEE Transactions on Image Processing, 2018, 28(4): 1705-1719.

[7] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks [J]. Communications of the ACM, 2020, 63(11): 139-144.

[8] LI Z, WU J. Learning deep CNN denoiser priors for depth image inpainting[J]. Applied Sciences, 2019, 9(6): 1103-1112.

[9] 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. IEEE, 2016: 2536-2544.

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

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

[12] 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). Cham: Springer International Publishing, 2018: 3-19.

[13] 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). Cham: Springer International Publishing, 2018: 85-100.

[14] 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. IEEE, 2020: 7760-7768.

[15] YU T, GUO Z, JIN X, et al. Region normalization for image inpainting [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12733-12740.

[16] LI H, XIONG L. Region normalization for image inpainting based on dual-channel generative adversarial network [C]// Proceedings of the International Conference on Advanced Algorithms and Neural Networks (AANN 2022). SPIE, 2022, 12285: 133-137.

[17] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks[C]//International Conference on Machine Learning. PMLR, 2017: 214-223.

[18] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual losses for real-time style transfer and super-resolution [C]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 694-711.

[19] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 2414-2423.

[20] 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. IEEE, 2015: 3730-3738.

[21] 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.

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

[23] XIE C, LIU S, LI C, et al. Image inpainting with learnable bidirectional attention maps [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2019: 8857-8866.

[24] GUO X, YANG H, HUANG D. Image inpainting via conditional texture and structure dual generation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2021: 14134-14143.

责任编辑:李翠薇