DOI:10.13878/j.cnki.jnuist.20221010004

左心悦[1]　郝子娴[1]　杨有[2]

多尺度语义学习的人脸图像修复

摘要

针对卷积神经网络在图像修复过程中难以兼顾修复结果的局部细节和全局语义一致性问题,以生成对抗网络为基础,提出一种多尺度语义学习的编解码人脸图像修复模型.首先,将人脸图像用门控卷积分解为具有不同大小的感受野和特征分辨率的分量,用不同尺寸的卷积核提取多尺度特征,通过提取合适的局部特征来提升修复结果的细节;其次,将提取的多尺度特征输入至语义学习模块,从通道和空间两个角度学习特征之间的语义关系,从而增强修复结果的全局一致性;最后,引入跳跃连接将编码端的特征补充到解码端中减少采样造成的细节信息损失,改善修复结果的纹理细节.在 CelebA-HQ 人脸数据集上进行实验,结果表明提出的模型在峰值信噪比、结构相似性、ℓ_1 三个性能指标上均有显著提升,修复的结果在视觉上局部细节和全局语义更合理.

关键词

图像修复;多尺度;语义学习;卷积神经网络;生成对抗网络

中图分类号 TP391.4

文献标志码 A

收稿日期 2022-10-10

资助项目 重庆市研究生联合培养基地项目(2019-45);重庆市教育委员会人文社会科学研究规划项目(21SKGH044)

作者简介

左心悦,女,硕士生,研究方向为计算机视觉.2020210516095@stu.cqnu.edu.cn

杨有(通信作者),男,博士,副教授,研究方向为计算机视觉.20130958@cqnu.edu.cn

1　重庆师范大学 计算机与信息科学学院,重庆,401331

2　重庆师范大学 重庆国家应用数学中心,重庆,401331

0　引言

图像修复(Image Inpainting)的目的是根据图像的已知内容重构缺失或损坏的区域,使修复的区域与整体内容保持一致.人脸图像修复作为其中一个重要的分支,在诸多领域有着重要的应用价值,比如面部修饰[1]和修复老照片[2]等领域.为了解决这一具有挑战性的任务,基于纹理合成的传统方法[3-4]主要是在图片的已知区域中寻找相似的纹理匹配块.但人脸图像不同于其他图像,面部五官具有固定的几何特性,传统方法输出的结果通常在语义方面存在局限.例如,人脸图像缺失的区域是鼻子,而已知区域没有与之对应的相似的纹理匹配块,因此无法产生语义上合理的结果.

随着深度学习的发展,卷积神经网络(Convolutional Neural Network,CNN)和生成对抗网络(Generative Adversarial Network,GAN)[5]在图像修复领域取得了一系列卓越的成果.CNN 具有强大的特征识别和提取能力,GAN 是生成模型的一种,核心思想源于博弈论的纳什均衡.GAN 由生成器和判别器组成,生成器学习真实数据样本的潜在分布并尽可能生成新的数据样本,判别器努力判断输入的数据是真实数据还是生成的数据样本,两者在对抗中学习.Pathak 等[6]将 GAN 的思想引入到图像修复中,并采用自编码器作为生成器生成修复结果,实验表明修复的结果既符合语义又具有真实性,该方法掀起了基于深度学习的图像修复的研究热潮.基于深度学习的修复方法最初都是针对规则的矩形破损区域进行修复,但在人脸修复的具体应用场景中,破损区域通常都是不规则的.针对此问题,Liu 等[7]首次提出用部分卷积替换 U-Net 中的普通卷积,实现了对任意形状任意大小缺失区域的图像修复;Yu 等[8]提出了门控卷积,在特征层的不同空间位置为每个通道建立了可学习的动态特征选择机制,以改善训练期间的掩码更新问题;Yang 等[9]提出了可学习结构知识融合网络,该网络分为两个阶段,第一阶段生成人脸边缘先验信息,第二阶段利用生成的边缘先验信息辅助进行图像修复.

虽然上述方法在人脸图像修复领域取得了重大进展,但在特征表达方面仍存在一定限制导致输出的结果局部细节和全局不一致.CNN 中感受野和图像特征的大小影响模型的学习能力,有效地增加感受野、提取合适的特征图来扩大局部特征范围有利于提升修复效果,两者的大小取决于卷积核的尺寸.同时,CNN 由于其固有的特性很

难对离缺失区域距离较远的区域建模,使得修复的结果出现伪影、模糊的纹理导致整体内容不一致.针对以上问题,本文提出一种多尺度语义学习的模型来实现对不规则破损的人脸图像修复,通过加强局部特征表达、对远距离空间的多尺度特征建模来提升修复结果局部细节和全局一致性.主要贡献点如下:

1)提出一种基于 CNN 结构的多尺度提取特征的策略,用尺寸不同的卷积核提取不同尺度的人脸图像特征并获取大小不同的感受野,以增强局部特征表达,从而提升修复结果的细节.

2)设计了一种语义学习模块从通道和空间两个角度学习多尺度特征之间的语义关系来提升生成图像内容的整体一致性.

1 相关工作

1.1 基于 GAN 的图像修复

生成对抗网络在图像修复领域取得了重大的突破,尤其面对复杂的图像修复任务,克服了传统方法语义理解困难的限制. Pathak 等[6] 将 Encoder-Decoder 引入到修复任务中,结合了 GAN 的对抗性思想提出了一个名为 Context-Encoder 的网络,并使用重构损失和对抗性损失作为约束条件来提升修复的效果.Iizuka 等[10]将 Context-Encoder[6] 中的判别器保留为局部判别器,同时增加一个全局判别器.Yu 等[11]引入了上下文注意力机制通过对远距离空间特征建模修复图像.以上算法针对的是破损区域为规则的矩形图片,但在实际应用中,图像破损的区域通常是不规则的.Liu 等[7]用部分卷积代替普通卷积实现了对不规则破损图像的修复.Yu 等[8]提出了门控卷积,在特征层的不同空间位置为每个通道建立了可学习的动态特征选择机制,以改善训练期间的掩码更新问题.Wang 等[12]提出了多列卷积生成网络,该网络在编码阶段使用不同大小的卷积核来获得不同大小的感受野.Yu 等[13]提出了一种新颖的区域归一化,它可以根据输入掩码将空间像素分为损坏和未损坏的区域,并分别计算每个区域的平均值和方差.Liu 等[14]设计了一个连贯的语义注意层,对缺失区域的特征进行语义关联建模.虽然上述方法在不规则破损人脸图像修复中取得了一定成果,但生成的结果局部细节和全局一致性差,存在整体结构扭曲、局部细节纹理模糊的问题.

1.2 人脸图像修复

人脸修复是图像修复的一个重要分支,人脸图像具有特殊性,五官具有明显的几何结构特性,且人脸图像不止有正脸,还包含侧脸等角度,因此人脸图像修复是一项具有挑战性的任务.人脸修复可分为单元修复方法[15-16] 和多元修复方法[17-20],只要输出的结果自然合理,人脸图像修复也可以产生多种结果.近年来深度学习技术[21] 取得了重大进展,在分类[22]、行为识别[23]、人脸图像修复等领域都取得了许多杰出的成果,Sun 等[15] 提出了一种在社交媒体照片中进行脸部修复的方法,根据损坏的图像的上下文信息,在适当的位置生成面部位置,并根据面部位置补全缺失的部分.Banerjee 等[16] 提出了一个多尺度的 GAN,直接根据提供的人脸特征生成视觉上真实的背景像素和背景,如头发、脖子和衣服.Zheng 等[17]将 VAE 与 GAN 结合起来,并行地生成和重建网络,以实现多样性的修复.Zhao 等[18] 提出了一个无监督的跨空间生成模型用于人脸修复.Liu 等[19] 设计了一个概率多样化的 GAN,用于生成多种修复结果.Peng 等[20] 提出使用一个分层的量化变分自编码器,首先学习自回归分布,然后分割结构和纹理特征.但多样修复生成模型在训练中容易崩溃且参数量较大.

2 方法

2.1 模型整体设计

本文提出的多尺度语义学习的人脸图像修复整体采用生成对抗网络模型,由生成器和判别器组成,如图 1 所示.生成器包括三个步骤:第一步,输入破损的人脸图像,通过三个并行的编码端提取具有不同大小的感受野和特征分量的多尺度特征;第二步,提取的多尺度特征输入至多尺度语义学习模块中来学习语义关系;第三步,将编码端的特征通过跳跃连接补充到解码端进行解码,减少采样造成的信息损失,输出修复好的预测图.

将生成器输出的预测图与真实图同时输入至判别器判断真假,通过对抗学习提高模型的修复能力,同时在判别器加入了谱归一化[24] 解决生成对抗网络训练不稳定问题.

2.1.1 多尺度特征提取

为了扩大感受野的范围增强局部特征来提升修复质量,采用在编码端提取多尺度特征的方法解决.将破损的人脸图像输入至三个并行的编码器,每个编码器分别使用 3×3、5×5、7×7 的卷积核提取特征以获得不同大小的感受野,从而得到丰富的信息来

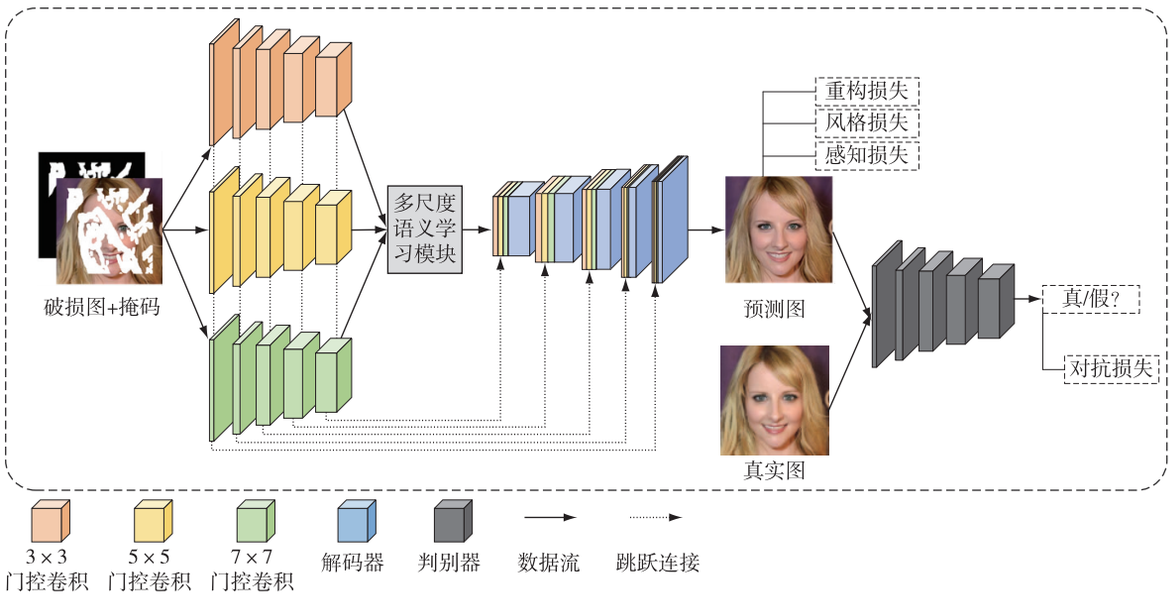

图 1　多尺度语义学习的人脸图像修复模型

Fig. 1　Overview of face image inpainting with multi-scale semantic learning

提升修复结果的细节.普通卷积将破损像素和已知像素同等对待同时输入至卷积层,导致修复的结果模糊,部分卷积[7]中人为设定的掩码更新机制不合理,比如在网络深层无效像素会消失,因此,模型采用门控卷积[8]提取特征.门控卷积有助于改善修复细节,提升整体颜色一致性,特别是修复有不规则破损区域的图像.门控卷积具有灵活的掩码更新机制,与硬门控机制不同,门控卷积能自动从数据中学习软掩码,即使在网络深层仍然能够根据掩码学习到不同通道中的特征来进行图像修复.同时,本文在每个门控卷积层加入批量归一化,以防止训练期间梯度消失.该操作可以表示为

$$\text{Gating} = \sum\sum W_g \cdot I, \tag{1}$$

$$\text{Feature} = \sum\sum W_f \cdot I, \tag{2}$$

$$I' = \text{BN}(\text{Ø}(\text{Feature}) \odot \sigma(\text{Gating})), \tag{3}$$

其中,I 表示输入的特征图,Gating 表示门控,Feature 表示卷积后的特征图,W_g 和 W_f 分别表示不同的卷积核,I' 表示门控卷积层输出的特征图,Ø 是 LeakyReLU 激活函数,$\odot$ 表示对应位置元素相乘,σ 表示 Sigmoid 激活函数,因此门控值在 0 和 1 之间取得,BN(·)代表批量归一化.

2.1.2　多尺度语义学习模块

为了提升修复结果全局一致性,将经过并行编码器获得的多尺度特征分别输入至多尺度语义学习模块来学习全局特征之间语义的关系,该模块由通道语义学习模块和空间语义学习模块[11]组成,从不同角度学习多尺度特征之间的语义关系来提升修复效果.

第一步,将多尺度特征输入至通道语义学习模块中,如图 2 所示.通道语义学习模块想法来源于 SENet[25],但不同的是增加了门控设计,通过对注意力扩展增加更多的非线性,来更好地拟合通道间复杂的相关性,使模型自动地学习不同通道的重要信息从而学习语义关系.首先,通过全局池化得到多尺度特征在通道层面的全局特征,将其送入两个全连接层中,使用 Sigmoid 激活函数学习每个通道之间的关系以获得不同通道的权重,预测每个通道的重要性;然后,将权重图作用到原始特征图上,将全局特征尺寸变换还原到原始特征的大小,和输入做通道级拼接;最后,将拼接后的结果并行送入全连接层,第二个分支再次使用 Sigmoid 函数激活,和第一个分支的结果逐元素相乘得到最终的输出.

第二步,将第一步的结果输入到空间语义学习模块,如图 3 所示.空间语义学习模块可从离缺失区域较远的位置学习来生成缺失部分像素,从已知区域学习语义关系来提升整体一致性.首先从输入的特征图中已知区域和缺失区域提取 3×3 像素的补丁块,计算补丁块之间的余弦相似性,计算式如下:

$$S_{i,j} = \left\langle \frac{f_i}{\|f_i\|_2}, \frac{f_j}{\|f_j\|_2} \right\rangle, \tag{4}$$

其中,f_i 和 f_j 分别表示缺失区域的第 i 个补丁块和已

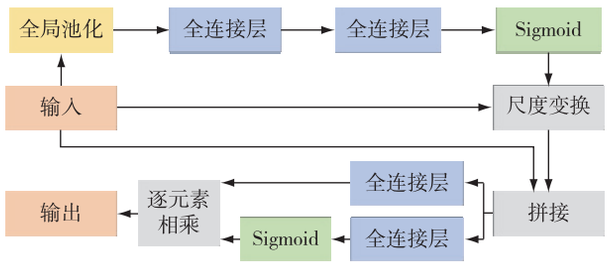

图 2 通道语义学习模块

Fig. 2 Channel semantic learning module

知区域的第 j 个补丁块.采用 Softmax 函数计算已知区域每个补丁块的注意分数,计算式如下:

$$S'_{i,j}=\left\langle \frac{\exp(S_{i,j})}{\sum_{j=1}^{N}\exp(S_{i,j})} \right\rangle . \tag{5}$$

最后基于注意力分数图重构输入的特征图生成缺失部分,计算式如下:

$$f'_i=\sum_{j=1}^{N} f_i S'_{i,j}. \tag{4}$$

第三步,将经过多尺度语义学习模块的特征从通道维度拼接,送入解码器中解码.卷积编码过程中会丢失部分信息,因此通过跳跃连接将编码器的特征补充到解码器,恢复丢失的细节信息.

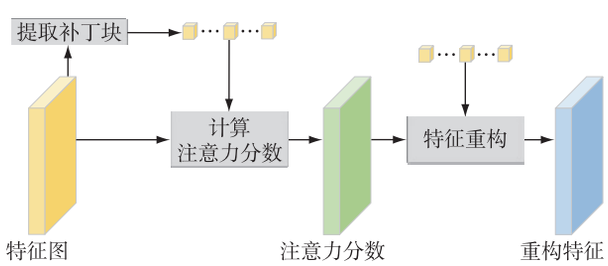

图 3 空间语义学习模块

Fig. 3 Spatial semantic learning module

2.2 损失函数

在训练过程中引入了感知损失、风格损失、重构损失和对抗损失约束生成语义合理的结果.感知损失 $\mathcal{L}_{\text{perc}}$[26]用来捕获高级语义特征,模拟人类对图像质量的视觉感知.使用 ImageNet[27]上的预训练模型 VGG-16[28]提取高级语义特征,感知损失计算式如下:

$$\mathcal{L}_{\text{perc}}=\mathbb{E}\left[\sum_i \| \boldsymbol{\phi}_{\text{pool}_i}(I_{\text{out}})-\boldsymbol{\phi}_{\text{pool}_i}(I_{\text{gt}}) \|_1\right], \tag{7}$$

其中,$\boldsymbol{\phi}_{\text{pool}_i}$ 是 VGG-16 的第 i 个池化层的激活图,$i\in[1,3]$,$\mathbb{E}$ 表示期望,I_{out}是模型生成的预测图,I_{gt}是真实图片,$\| \boldsymbol{x} \|$ 表示 $\boldsymbol{x}$ 的 L1 范数.风格损失 $\mathcal{L}_{\text{style}}$与感知损失 $\mathcal{L}_{\text{perc}}$计算方法类似,用来保持图像整体风格一致性.风格损失的计算式如下:

$$\boldsymbol{\phi}^{\text{style}}_{\text{pool}_i}=\boldsymbol{\phi}_{\text{pool}_i}\boldsymbol{\phi}^{\text{T}}_{\text{pool}_i}, \tag{8}$$

$$\mathcal{L}_{\text{style}}=\mathbb{E}\left[\sum_i \| \boldsymbol{\phi}^{\text{style}}_{\text{pool}_i}(I_{\text{out}})-\boldsymbol{\phi}^{\text{style}}_{\text{pool}_i}(I_{\text{gt}}) \|_1\right], \tag{9}$$

其中,$\boldsymbol{\phi}^{\text{style}}_{\text{pool}_i}$ 表示特征图对应的 Gram 矩阵.对抗损失被用来确保重建图像的一致性,它的定义如下:

$$\mathcal{L}_{\text{adv}}=\min_G \max_D \mathbb{E}_{I_{gt}}[\log D(I_{\text{gt}})]+\mathbb{E}_{I_{\text{out}}}\log[1-D(I_{\text{out}})]. \tag{10}$$

此外,计算 I_{out}和 I_{gt}之间的 ℓ_1 距离作为重构损失,计算式如下:

$$\mathcal{L}_{\text{rec}}=\mathbb{E}\| I_{\text{out}}-I_{\text{gt}} \|_1. \tag{11}$$

综上所述,总体损失函数计算式如下:

$$\mathcal{L}_{\text{total}}=\lambda_{\text{perc}}\mathcal{L}_{\text{perc}}+\lambda_{\text{sytle}}\mathcal{L}_{\text{sytle}}+\lambda_{\text{adv}}\mathcal{L}_{\text{adv}}+\lambda_{\text{rec}}\mathcal{L}_{\text{rec}}. \tag{12}$$

3 实验与结果分析

实验硬件环境采用 NVIDIA RTX3060Ti GPU,显存大小为 8 GB,CPU 为 i5-10400F,内存大小为 16 GB.网络构建由 PyTorch 深度学习框架实现,优化算法使用 Adam,训练模型时 Batchsize 为 8,具体参数为 $\lambda_{\text{perc}}=0.05$、$\lambda_{\text{sytle}}=120$、$\lambda_{\text{rec}}=1$、$\lambda_{\text{adv}}=0.1$.数据集采用 CelebA-HQ 人脸数据集和 NVIDIA 不规则掩码数据集,图像尺寸统一裁剪为 256×256 像素.为了模型的优越性,将模型与 CA[11]、GConv[8]、EC[29]、PIC[17]、RFR[30]五个经典的图像修复模型在 CelebA-HQ 人脸数据集上进行实验对比,同时进行消融实验验证模块的有效性.

3.1 定量评价

根据破损区域占整体图片的比例,在(10%,20%]、(20%,30%]、(30%,40%]、(40%,50%]四个掩码比例上做了对比实验.评价指标为峰值信噪比(Peak Signal to Noise Ratio,PSNR)、结构相似性[31](Structure SIMilarity,SSIM)、ℓ_1 距离,分别从预测图与真实图的失真程度、整体结构相似程度、平均绝对误差三个角度展开评价.定量评价如表 1 所示,本文提出的网络在不同掩码比例下性能均优于其他方法,表明提出的方法能有效地生成高质量的修复结果.

为了验证多尺度语义学习模块的重要性,表 2 为去掉多尺度语义学习模块(Multi-scale Semantic Learning,MSL)的定量结果.

表 1　不同算法在 CelebA-HQ 数据集的定量比较

Table 1　Quantitative comparison between different algorithms on CelebA-HQ

评价指标	模型	掩码比例 (10% 20%]	(20% 30%]	(30% 40%]	(40% 50%]
PSNR ↑	CA[11]	28.04	24.86	22.49	20.60
	GConv[8]	30.92	27.51	25.15	23.27
	EC[29]	31.23	27.94	25.51	23.49
	PIC[17]	29.13	26.02	23.74	21.79
	RFR[30]	30.54	28.35	26.44	24.70
	本文	**32.29**	**29.00**	**26.76**	**24.90**
SSIM ↑	CA[11]	0.941	0.888	0.823	0.743
	GConv[8]	0.969	0.937	0.898	0.850
	EC[29]	0.971	0.942	0.902	0.847
	PIC[17]	0.956	0.915	0.862	0.790
	RFR[30]	0.964	0.947	0.921	0.886
	本文	**0.977**	**0.953**	**0.926**	**0.888**
ℓ_1 ↓	CA[11]	0.015 0	0.025 9	0.039 0	0.054 7
	GConv[8]	0.010 6	0.018 3	0.027 2	0.037 5
	EC[29]	0.009 6	0.017 1	0.025 9	0.036 8
	PIC[17]	0.013 8	0.023 2	0.034 3	0.047 9
	RFR[30]	0.010 0	0.015 4	0.022 2	0.030 4
	本文	**0.007 6**	**0.014 0**	**0.021 1**	**0.029 6**

注:↑表示越大越好;↓表示越小越好.

表 2　多尺度语义学习模块消融实验对比

Table 2　Experimental comparison of ablation of multi-scale semantic learning modules

评价指标	模型	掩码比例 (10% 20%]	(20% 30%]	(30% 40%]	(40% 50%]
PSNR ↑	w/o MSL	32.11	28.85	26.62	24.76
	本文	**32.29**	**29.00**	**26.76**	**24.90**
SSIM ↑	w/o MSL	0.976	0.952	0.923	0.886
	本文	**0.977**	**0.953**	**0.926**	**0.888**
ℓ_1 ↓	w/o MSL	0.007 8	0.014 4	0.021 7	0.030 3
	本文	**0.007 6**	**0.014 0**	**0.021 1**	**0.029 6**

注:↑表示越大越好;↓表示越小越好.

3.2　定性评价

图 4 为所提算法和其他算法修复的效果对比.对于不规则破损的人脸图像,CA 算法无法合理地还原纹理细节如图 4c 所示,生成的结果结构扭曲纹理模糊;GN 算法生成的结果五官纹理模糊如图 4d 第四张鼻子位置;EC 算法生成的结果细节纹理模糊如图 4e 第三张眼睛位置;循环渐进式填充孔洞的 RFR 算法在掩码区域比较大时可以生成合理的结构,但掩码较小的区域无法填充所有的缺失像素如图 4f 第四张下巴位置;PIC 算法生成的结果整体结构扭曲如图 4g 第三张人脸的眼睛和鼻子.与其他算法相

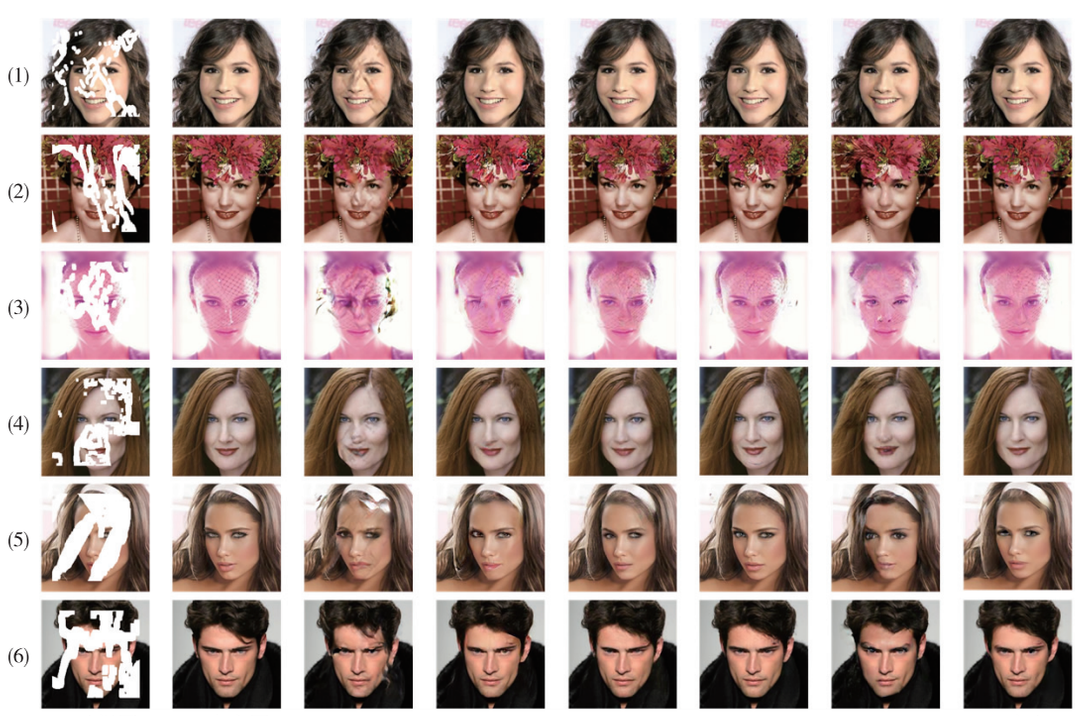

图 4　CelebA-HQ 数据集定性评价

Fig. 4　Qualitative comparison on CelebA-HQ

比,本文算法在整体和细节上均优于其他算法,局部细节逼真、全局一致性更加连贯,具有更加合理的视觉效果.

4 结论

本文设计了一种基于多尺度语义学习的人脸图像修复方法,通过不同大小的卷积核提取多尺度特征获得不同大小的感受野来增强局部特征表达,从通道和空间两个维度学习多尺度特征的语义关系,从而提升修复结果的全局一致性,弥补卷积神经网络固有特性在人脸图像修复方面的不足.同时,引入跳跃连接将编码端的特征补充到解码来恢复采样过程中丢失的细节信息,提高模型修复图像的能力.在CelebA-HQ人脸数据集上进行实验,结果表明提出的方法可以有效提升修复结果的细节,在质量和性能方面优于其他先进的方法.后续将在多尺度特征提取、语义学习模块的设计上做进一步研究,以不断提升人脸图像修复的结果.

参考文献

References

[1] Elharrouss O, Almaadeed N, Al-Maadeed S, et al. Image inpainting: a review[J]. Neural Processing Letters, 2020, 51(2): 2007-2028

[2] Wan Z Y, Zhang B, Chen D D, et al. Old photo restoration via deep latent space translation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022. DOI: 10.1109/TPAMI.2022.3163183

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

[4] Drori I, Cohen-Or D, Yeshurun H. Fragment-based image completion[C]//SIGGRAPH'03: ACM SIGGRAPH 2003 Papers. July 27-31, 2003, San Diego, California. New York: ACM, 2003: 303-312

[5] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144

[6] Pathak D, Krähenbühl P, Donahue J, et al. Context encoders: feature learning by inpainting[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 2536-2544

[7] Liu G L, Reda F A, Shih K J, et al. Image inpainting for irregular holes using partial convolutions[C]//15th European Conference on Computer Vision. September 8-14, 2018, Munich, Germany. New York: ACM, 2018: 89-105

[8] Yu J H, Lin Z, Yang J M, et al. Free-form image inpainting with gated convolution[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). October 27-November 2, 2019, Seoul, Korea (South). IEEE, 2020: 4470-4479

[9] Yang Y, Liu S, Xing B, et al. Face inpainting via learnable structure knowledge of fusion network[J]. KSII Transactions on Internet and Information Systems, 2022, 16(3): 877-893

[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 H, Lin Z, Yang J M, et al. Generative image inpainting with contextual attention[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 5505-5514

[12] Wang Y, Tao X, Qi X J, et al. Image inpainting via generative multi-column convolutional neural networks[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. December 3-8, 2018, Montréal, Canada. New York: ACM, 2018: 329-338

[13] Yu T, Guo Z Y, Jin X, et al. Region normalization for image inpainting[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12733-12740

[14] Liu H Y, Jiang B, Xiao Y, et al. Coherent semantic attention for image inpainting[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). October 27-November 2, 2019, Seoul, Korea (South). IEEE, 2020: 4169-4178

[15] Sun Q R, Ma L Q, Oh S J, et al. Natural and effective obfuscation by head inpainting[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 5050-5059

[16] Banerjee S, Scheirer W J, Bowyer K W, et al. On hallucinating context and background pixels from a face mask using multi-scale GANs[C]//2020 IEEE Winter Conference on Applications of Computer Vision (WACV). March 1-5, 2020, Snowmass, CO, USA. IEEE, 2020: 289-298

[17] Zheng C X, Cham T J, Cai J F. Pluralistic image completion[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 15-20, 2019, Long Beach, CA, USA. IEEE, 2020: 1438-1447

[18] Zhao L, Mo Q H, Lin S H, et al. UCTGAN: diverse image inpainting based on unsupervised cross-space translation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 13-19, 2020, Seattle, WA, USA. IEEE, 2020: 5740-5749

[19] Liu H Y, Wan Z Y, Huang W, et al. PD-GAN: probabilistic diverse GAN for image inpainting[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 9367-9376

[20] Peng J L, Liu D, Xu S C, et al. Generating diverse structure for image inpainting with hierarchical VQ-VAE[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 10770-10779

[21] 郑征,谭磊,周楠,等.基于多头注意力卷积网络的电力负荷预测[J].南京信息工程大学学报(自然科学版),2022,14(5):535-542
ZHENG Zheng, TAN Lei, ZHOU Nan, et al. Power load prediction based on multi-headed attentional convolutional network[J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition),2022,14(5):535-542

[22] 庄志豪,王敏,王康,等.基于深度学习的地基云分类技术研究进展[J].南京信息工程大学学报(自然科学版),2022,14(5):566-578
ZHUANG Zhihao, WANG Min, WANG Kang, et al. Research progress of deep learning-based cloud classification[J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition),2022,14(5):566-578

[23] 胡凯,郑翡,卢飞宇,等.基于深度学习的行为识别算法综述[J].南京信息工程大学学报(自然科学版),2021,13(6):730-743.
HU Kai,ZHENG Fei,LU Feiyu,et al.A survey of action recognition algorithms based on deep learning[J].Journal of Nanjing University of Information Science & Technology (Natural Science Edition),2021,13(6):730-743

[24] Miyato T,Kataoka T,Koyama M,et al.Spectral normalization for generative adversarial networks [J]. arXiv e-print,2018,arXiv:1802.05957

[25] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018:7132-7141

[26] Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution [M]//Computer Vision—ECCV 2016. Cham: Springer International Publishing, 2016:694-711

[27] 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

[28] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv e-print, 2014, arXiv:1409.1556

[29] Nazeri K, Ng E, Joseph T, et al. Edgeconnect: generative image inpainting with adversarial edge learning[J]. arXiv e-print, 2019, arXiv:1901.00212

[30] Li J Y, Wang N, Zhang L F, et al. Recurrent feature reasoning for image inpainting[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 13-19, 2020, Seattle, WA, USA. IEEE, 2020:7757-7765

[31] 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: a Publication of the IEEE Signal Processing Society, 2004, 13(4):600-612

Face image inpainting with multi-scale sematic learning

ZUO Xinyue[1] HAO Zixian[1] YANG You[2]

1 College of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China

2 National Center for Applied Mathematics in Chongqing, Chongqing Normal University, Chongqing 401331, China

Abstract To address the issue that convolutional neural networks can hardly balance the local details and global semantic consistency of results in the process of image inpainting, a multi-scale semantic learning model for face image inpainting based on generative adversarial networks is proposed. First, the face image is decomposed into components with different perceptual fields and feature resolutions using gated convolution, and multi-scale features are extracted using convolution kernels of different sizes to enhance the detail of the restoration results by extracting appropriate local features. Second, the extracted multi-scale features are fed into the semantic learning module to learn the semantic relationships between features from both the channel and spatial perspectives, thus enhance the global consistency of the restoration results. Finally, skip connections are introduced to complement the features on the encoding side to the decoding side to reduce the loss of detail information caused by sampling and improve the texture details of the restoration results. Experiments on the CelebA-HQ face dataset show that the proposed model has significant improvements in three performance metrics: peak signal to noise ratio, structure similarity and ℓ_1, and the inpainting results are visually more reasonable in terms of local details and global semantics.

Key words image inpainting; multi-scale; sematic learning; convolutional neural network (CNN); generative adversarial network (GAN)