中国科学: 信息科学 2023 年 第 53 卷 第 4 期: 682–698
SCIENTIA SINICA Informationis
论文

多尺度稳定场 GAN 的图像修复模型

叶学义*, 曾懋胜*, 孙伟杰, 王凌宇, 赵知劲

杭州电子科技大学通信工程学院, 杭州 310018
* 通信作者. E-mail: xueyiye@hdu.edu.cn, zms_0310@163.com

收稿日期: 2022–02–14; 修回日期: 2022–05–26; 接受日期: 2022–07–21; 网络出版日期: 2023–04–11

国家自然科学基金 (批准号: U19B2016, 60802047) 资助项目

摘要 近年来生成对抗网络 (generative adversarial network, GAN) 已经展示了它在图像修复任务中修复大面积缺失区域并生成合理语义结果的潜力, 但现有方法经常忽略缺失区域的语义一致性和特征连续性, 并对不同尺度特征的感知能力不足, 因此提出一种基于多尺度稳定场 GAN 的图像修复模型. 该模型的生成单元汲取了 U-Net 的特点, 将稳定场算子嵌入到跳跃连接中以填充编码器特征图中的缺失区域, 保持了缺失区域的语义一致性和特征连续性; 然后通过多尺度融合计算逐步加强经稳定场算子填充缺失区域的特征图的传递, 使得跳跃连接传递的信息不再来自单一的特征图, 让模型能够感知高层特征的语义信息. 在人脸和自然场景等数据集上的实验结果表明, 该模型优于其他的经典图像修复方法.

关键词 图像修复, 生成对抗网络 (GAN), 稳定场, 多尺度融合, 深度学习

1 引言

利用破损图像或训练数据中的先验信息填充或修正图像中缺失的像素通常被称为图像修复 (image inpainting), 期望在满足人类视觉感知需求的前提下尽可能与原图一致. BSCB 模型 [1]、曲率扩散修复算法 (CDD 模型) [2] 和全变分修复算法 (TV 模型) [3] 都根据图像像素间的相关性和内容相似性进行修复; Barnes 等 [4] 提出的 PatchMatch 使用纹理合成的思想完成图像修复, 该算法致力于找到近似的最近邻匹配, 即从缺失区域边缘迭代搜索最适合的补丁以合成缺失部分的内容. 此类传统的图像修复算法都不能捕捉图像中的语义信息, 在待修复图像缺失区域小、结构及纹理信息简单的情况下能取得比较理想的修复效果, 但在面对缺失区域大、结构及纹理信息复杂的情况时, 由于缺乏捕捉图像语义信息的能力, 修复效果显著下降. 遥感图像的厚云去除与大缺失区域的图像修复具有一定的相似性, Li 等 [5] 提出使用非负矩阵分解和纠错方法实现去云效果; Singh 等 [6] 提出 Cloud-GAN 来学习多云图像和无云图像之间的映射实现对云层的去除. 这对大面积缺失区域的图像修复具有一定的参考价值.

引用格式: 叶学义, 曾懋胜, 孙伟杰, 等. 多尺度稳定场 GAN 的图像修复模型. 中国科学: 信息科学, 2023, 53: 682–698, doi: 10.1360/SSI-2022-0065
Ye X Y, Zeng M S, Sun W J, et al. Image inpainting based on multi-scale stable-field GAN (in Chinese). Sci Sin Inform, 2023, 53: 682–698, doi: 10.1360/SSI-2022-0065

图像修复的挑战在于如何为缺失区域合成视觉逼真且语义合理的像素, 这些像素与图像中已知像素区域共存且具有语义一致性.

Context Encoders [7] 于 2016 年首次将生成对抗网络 (generative adversarial network, GAN) [8] 应用于图像修复, 使模型在理解图像上下文的同时, 做出对图像缺失区域的合理假设, 然而在生成精细的纹理方面表现不佳. 此后有较多 GAN 模型被提出应用于图像修复, 这些方法的生成单元大多引入 U-Net [9] 架构, 使用跳跃连接 (skip connections) 将编码器与解码器对应的特征图在通道上拼接融合, 从而保留了更多纹理信息. 近年来 GAN 的快速发展使得图像修复的效果取得了明显的改善, 一定程度上弥补了传统图像修复算法无法捕捉图像语义的不足, 实现了在缺失区域大的情况下完成图像的修复并得到合理的语义结果.

此后, 研究者们在各个方面对 GAN 进行改进以提升图像修复的效果. 部分研究者对 GAN 的判别单元进行改进从而间接提升生成单元生成样本的能力, Iizuka 等 [10] 在 Context Encoders 的基础上再加入一个判别单元, 利用全局判别单元和局部判别单元共同参与对抗来进一步优化修复效果, 但是需要后处理来保证掩膜边界附近与已知区域的颜色一致性. 大多数研究者通过对 GAN 的生成单元进行改进从而直接提升其生成样本的能力, 如采取多步渐进的方式实现对图像的修复, Zhang 等 [11] 使用多个相同的 U-Net [9] 结构实现对图像的渐进式修复; Liu 等 [12] 使用两个不同的 U-Net 结构实现粗修复到精修复的双阶段模型来完成对图像的修复, 同时在精修复阶段提出连贯语义注意力, 同时考虑缺失区域和已知区域间的相似性匹配以及缺失区域内部特征像素间的相关性. 与文献 [12] 相似, 部分研究者通过引入注意力机制实现对图像的修复, Yu 等 [13] 使用上下文感知层从已知区域中搜索与缺失区域中最相似的特征块并实现粗略匹配, 但该方法忽略了图像缺失区域内部特征像素之间的相关性, 导致了修复的图像缺失区域内语义的不连贯; Zeng 等 [14] 提出金字塔上下文编码器用于学习图像语义信息; 杨等 [15] 使用扩张的多尺度通道注意力提高模型对编码器的低级特征的利用效率; Zhao 等 [16] 引入交叉语义注意力, 利用缺失区域与已知区域之间的长距离依赖关系提高修复结果的真实性. 由于带有缺失区域的图像含有无效像素, 使用标准卷积时会同等对待有效像素与无效像素, 因此部分学者对标准卷积进行改进从而提升图像修复的效果, Liu 等 [17] 通过更新每一层的掩膜并用掩膜值重新对卷积权重标准化, 提出使用部分卷积 (partial convolution, PConv) 代替标准卷积以改善标准卷积在面对不规则缺失区域时图像修复结果可能会出现的伪影、色差、模糊等问题; Xie 等 [18] 针对 PConv 中的手工特征归一化和只考虑前向传播的缺失区域更新的两个问题进行改进, 提出以端到端的方式学习特征标准化和掩膜更新, 能有效地适应不规则缺失区域的前向传播; 杨等 [19] 使用密集连接块代替 U-Net 中的标准卷积, 用于捕捉图像缺失区域的语义信息. 也有部分研究者从其他角度对 GAN 进行改进实现图像修复, Yan 等 [20] 在 U-Net 的解码器中添加了一个特殊的移位连接层, 有效地结合了基于样本块的方法与卷积神经网络的方法; Xu 等 [21] 利用图像边缘完成图像修复, 对不完整图像的边缘进行填写, 并利用该边缘实现图像修复. 在图像修复领域中, 研究者大多聚焦于如何提升图像修复结果的质量, 也有少部分在图像修复模型的推理时间和资源占用上进行改进, 廖等 [22] 提出一种结合组卷积与注意力机制的模块用于替换标准卷积, 从而在保证图像修复质量的前提下, 降低模型的推理时间与资源占用; Shin 等 [23] 提出使用并行语义修复网络, 降低粗修复到精修复的双阶段模型的计算资源.

尽管上述大多基于 U-Net 结构的模型都取得了较好的修复效果, 但在图像修复领域, 由于输入为具有缺失区域的待修复图像, 对于缺失区域的中心, 跳跃连接的输入几乎为零, 并不能将纹理信息传递到解码器, 导致修复的结果产生纹理模糊、结构失真的内容. 从语义层面来看, 是因为这些方法未达到全局语义一致性和局部特征连续性的要求, 即忽略了缺失区域的语义一致性和特征连续性, 导致修复结果在缺失区域语义上并不连贯. 同时, 跳跃连接在 U-Net 中仅连接编码器与解码器的对应层, 这

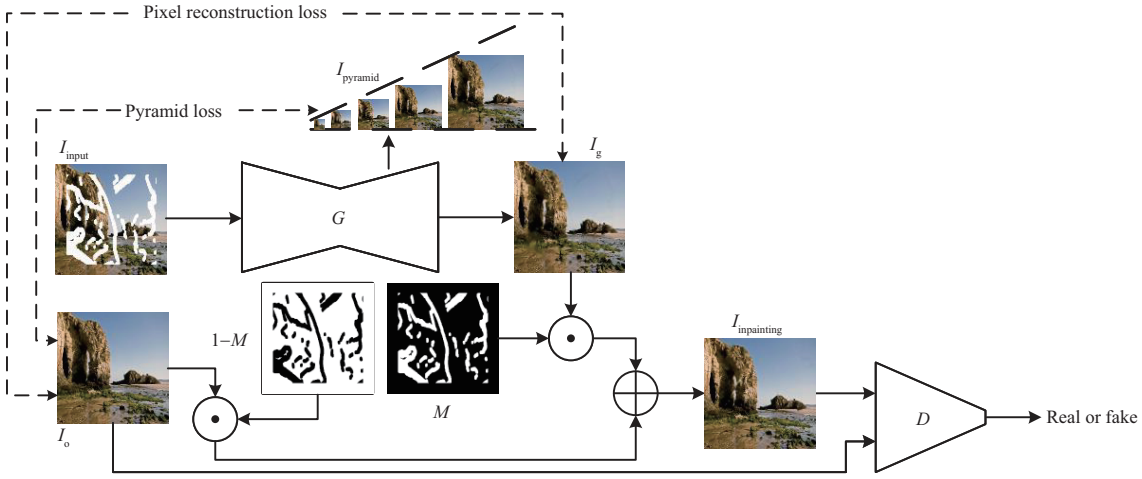

图 1　MSSF-GAN 的框图结构

Figure 1　Block diagram structure of the MSSF-GAN

种特征传递关系导致高层特征 (high-level features) 不能指导低层特征 (low-level features) 的生成, 从而保持语义连贯在解码过程中传递和加强.

因此如何保持缺失区域的语义一致性和特征连续性以及高层语义特征是目前图像修复研究的关键问题. 物理量在空间分布不随时间变化被称为稳定场, 图像纹理是物体表面结构及几何形态和成像光相互作用的稳定结果, 因此和稳定场具有相似性, 可以通过建立稳定场模型方程重建图像中缺失区域的像素点 [24]. 本文汲取 U-Net 的基本思路, 将稳定场算子引入到跳跃连接中, 在一定程度上保证了缺失区域的语义一致性和特征连续性, 再利用多尺度融合计算逐步加强经稳定场算子填充缺失区域的特征图的传递, 因为在多个尺度上融合而达到逐步加强的效果, 使得图像的高层语义特征在修复过程中得到保持. 具体来说, 先通过稳定场算子填充不同尺度的编码器特征图中的缺失区域, 使得经跳跃连接传递到解码器的特征图中的缺失区域不再是全零值, 而是可信的初值; 再利用多尺度融合将缺失区域已经被填充的高层特征以渐进的形式逐步向低层特征融合, 使得跳跃连接传递的特征信息不会只来自于单一的特征图, 让低层特征能够感知高层特征的语义信息; 同时每个尺度的解码器特征图都输出对应尺度的修复图像, 用于计算金字塔损失以逐步完善每个尺度对缺失区域的填充. 这种基于多尺度稳定场的 GAN (multi-scale stable-field generative adversarial network, MSSF-GAN) 可以适用于各种场景的图像修复任务.

2　多尺度稳定场 GAN (MSSF-GAN)

基于 GAN 的图像修复, 其生成单元接收带有缺失区域的待修复图像作为输入, 生成的输出图像相比于输入, 缺失区域和已知区域的内容都会发生改变, 但只有缺失区域需要修复, 因此最终的修复结果由输入的已知区域和输出的缺失区域组合得到.

如图 1 所示, MSSF-GAN 包括生成单元和判别单元, 图中 G 表示生成单元, D 表示判别单元, I_{o} 表示原始图像, I_{input} 表示带有缺失区域的输入图像, I_{g} 表示 I_{input} 输入 G 后输出的生成图像, I_{pyramid} 表示由不同尺度的解码器特征输出的生成图像组成的金字塔图像组, M 表示掩膜 (实验中以掩膜来拾取图像中的缺失区域, M 和 I_{g} 大小相同, M 中对应缺失区域的像素值全为 1, 其余为 0), $I_{\mathrm{inpainting}}$ 表

示最终的修复结果, 其由 I_{g} 的缺失区域修复的内容和 I_{o} 的已知区域的内容组合得到, 确保 GAN 的修复不影响图像的已知区域, 如式 (1) 所示:

$$I_{\mathrm{inpainting}} = I_{\mathrm{g}} \odot M + I_{\mathrm{o}} \odot (1 - M)\,, \tag{1}$$

其中 $\odot$ 表示矩阵逐元素相乘.

如图 1 所示, G 接收 I_{input} 作为输入, 输出 I_{g} 和 I_{pyramid} 分别和 I_{o} 计算像素重构损失和金字塔损失, 两者都是为了让修复的结果在像素层面更接近真实样本. D 接收 $I_{\mathrm{inpainting}}$ 和 I_{o} 以鉴别真假和计算对抗损失, 对抗损失帮助 D 尽可能鉴别出真实样本和修复结果, 帮助 G 尽可能生成更加逼真、合理的修复结果. G 与 D 相互对抗以优化生成结果, 整个训练过程由上述 3 个损失函数同时指导优化.

2.1 生成单元

MSSF-GAN 的生成单元 G 由编码器、解码器、嵌入到跳跃连接中的稳定场算子和多尺度融合组成, 如图 2 所示. 这种设计汲取了 U-Net 架构的特点, 使用跳跃连接将编码器中的特征信息传递到对应的解码器中, 实现在解码器特征图上的通道融合, 使得解码器的上采样计算可以借用编码器中的原始信息. 但区别于 U-Net, 对图像修复更为重要的是, 每层跳跃连接中稳定场算子的加入, 使得缺失区域在连接中传递的不再是无效信息; 不仅如此, 多尺度逐层融合的加入使编码器获得的高层语义特征得到传递和保持, 从而改善缺失区域的语义不连贯和图像修复过程中缺乏高层特征信息的问题.

如图 2 所示, 编码器含有 6 层下采样卷积层, 卷积层参数如表 1 所示, 表中 Kernel size 为卷积核大小; Stride 为步长, 即卷积核每次在特征图中水平方向或垂直方向的步进长度; Padding 为填充, 即在特征图的每一边添加的行列的数目. 图中蓝色方块示意了每层卷积计算得到的特征图, 方块上方的数字表示尺度大小, 即随着编码过程的下采样, 特征图从 128×128 直至 4×4, 下方的数字表示通道数, 随着通道数增加, 块逐渐变厚; 根据跳跃连接关系, 解码器同样是 6 层结构, 图中蓝色方块由编码器中特征图经稳定场算子和多尺度融合后传递得到, 橙色方块由前一层特征图上采样得到, 上采样过程采用插值上采样和卷积的形式代替转置卷积以避免因参数配置不当而出现输出特征图带有明显棋盘状的现象; 在各层跳跃连接中嵌入稳定场算子, 它利用该层特征图已知区域的信息初步修复缺失区域, 改善跳跃连接直接将含有缺失区域的特征图转递至解码器导致的图像修复结果语义不连贯; 最后, 经稳定场算子修复后的各层特征图再通过逐层的多尺度融合计算, 能够使各层解码器更有效地利用编码器各层特征, 同时也使得图像的高层语义特征在修复过程中得到保持.

2.1.1 基于稳定场算子的跳跃连接

U-Net 的跳跃连接中图像缺失区域被直接代入解码器, 由于此时该区域的值为零或设置为随机值, 没有考虑与已知区域的关系, 直接破坏了特征的连续性, 使得语义断裂和不连贯. 因此赋予缺失区域可信的值, 尽可能保持特征连续, 维持语义一致, 可以有效激活跳跃连接对图像修复的效能. 这里采用文献 [24] 的思路构建稳定场算子给缺失区域赋值, 该文献认为图像纹理是物体表面结构及几何形态和成像光相互作用的稳定结果, 可以通过建立稳定场模型方程来恢复图像中缺失区域的像素点, 其本质为利用缺失区域周围的已知像素点对缺失像素的影响程度不同赋予不同的权重并初始化缺失像素点, 如式 (2) 所示:

$$LG\,(r, r_i) = f, \tag{2}$$

其中 L 为线性微分算符, r 为缺失像素点, r_i 为有效像素点, $G\,(r, r_i)$ 定义为有效像素点对缺失像素点的影响函数, 为缺失区域提供能量的已知区域定义为场源 f. 其优势在于该算子可以充分保留局部特

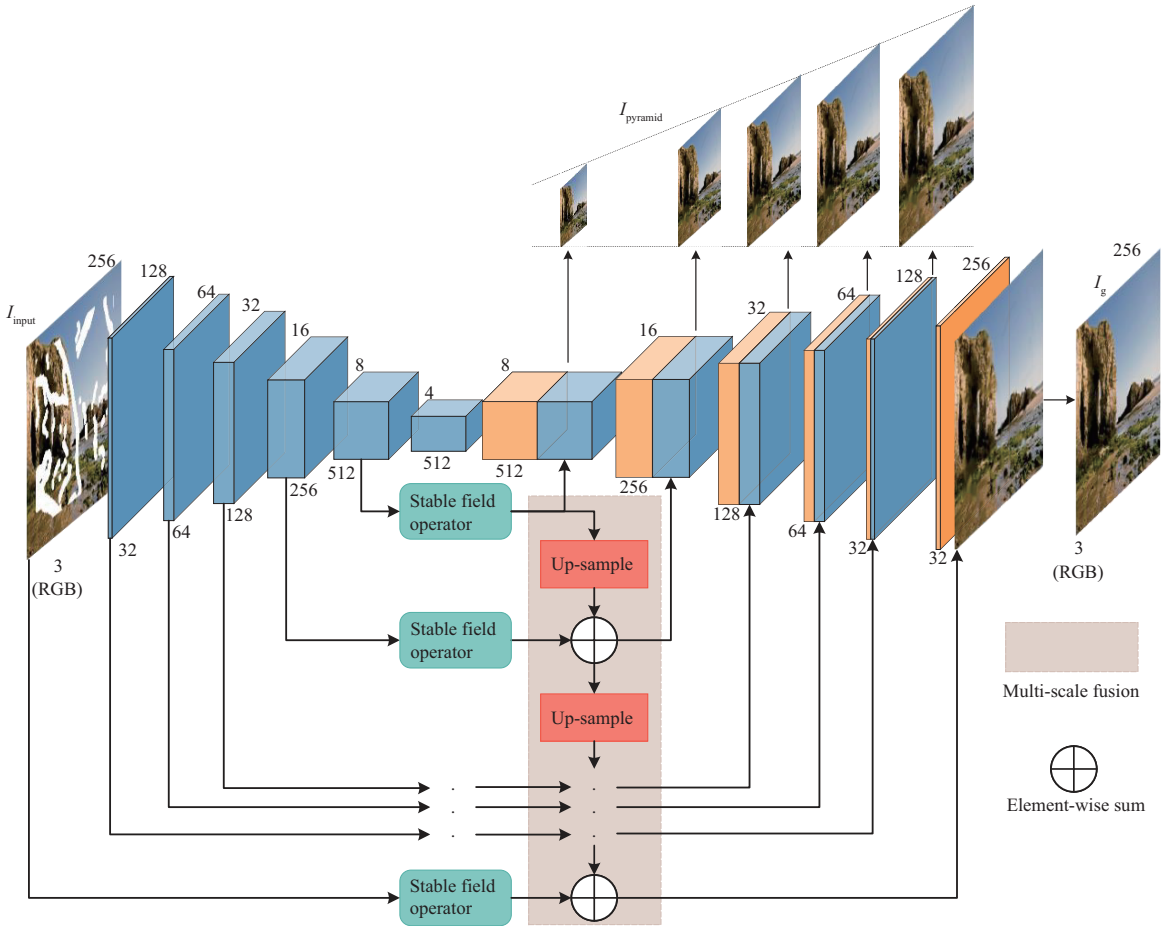

图 2　MSSF-GAN 的生成单元

Figure 2　Generator of the MSSF-GAN

表 1　生成单元卷积参数

Table 1　Parameters of the convolution in the generator

	Kernel size	Stride	Padding	Others
Down-sampling in the encoder	3	2	1	–
Up-sampling in the decoder	3	1	1	Scale factor is 2
Up-sampling in the multi-scale fusion	3	1	1	Scale factor is 2
Decoder outputs I_{pyramid}	1	1	0	–

征的连续性, 而且不需要迭代计算, 非常适合跳跃连接关系. 稳定场算子具体设计如下.

根据文献 [24] 定义, 描述图像纹理的函数 $I(r)$ 为

$$I(r)=\sum_{i=1}^{n} G(r,r_i)\, I_0(r_i), \tag{3}$$

其中 $I_0(r_i)$ 为 r_i 点的点源强度即像素值, 式 (3) 说明了缺失像素点的值可以表示为对其有影响的所有有效像素点像素值与影响函数乘积之和, 此处的有效像素点包括已知像素点和完成恢复的像素点,

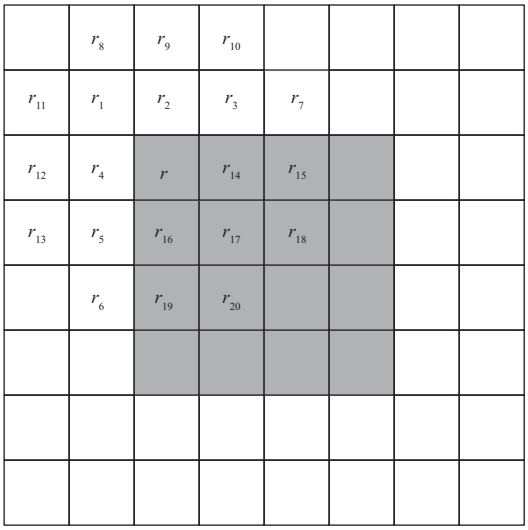

图 3 稳定场算子填充缺失区域的示意图

Figure 3 Missing regions filled by the stable field operator

距离较远的有效像素点对缺失像素点产生的影响可忽略不计. 即稳定场算子利用缺失区域周围的已知像素点对缺失像素的影响程度不同赋予不同的权重并进行填充.

图 3 以矩形缺失区域为例说明本文的 MSSF-GAN 利用稳定场算子填充编码器特征图中的缺失区域的过程, 如图中所示, 灰色背景的矩形区域为缺失区域, 其余为已知区域. 利用该特征图中的有效像素对缺失区域中的像素 r 进行填充, 考虑到当已知像素点 r_i 距离缺失像素点 r 较远时, r_i 对 r 的影响程度将很小, 因此在对 r 进行填充时只考虑其周围 20 个点对其的影响. 在此例中 $r_1 \sim r_{20}$ 是能对 r 产生有效影响的 20 个点, 但 $r_{14} \sim r_{20}$ 是缺失区域中的缺失像素, 因此只能利用 $r_1 \sim r_{13}$ 对 r 进行填充.

然后从方向导数的角度定义单个点 r_i 对 r 的影响因子 $g(r, r_i)$, 如式 (4) 所示:

$$g(r, r_i) = \sqrt{1 - \left(\frac{|I_0(r) - I_0(r_i)|}{I_0(r_i)}\right)^2} = \sqrt{1 - \left(\frac{||\nabla I_0(r)| \times \cos\theta \times d_{rr_i}|}{I_0(r_i) + \varepsilon}\right)^2}, \tag{4}$$

其中 d_{rr_i} 为 r_i 与 r 之间的距离, $\nabla I_0(r_i)$ 表示 r_i 点像素的梯度, θ 为梯度 $\nabla I_0(r_i)$ 和向量 $\boldsymbol{rr}_i$ 的夹角, ε 的作用是为了防止分母为 0. 考虑所有点的影响因子并归一化计算, 有

$$G(r, r_i) = \frac{g(r, r_i)\,\alpha(r_i)}{\sum_i g(r, r_i)\,\alpha(r_i)}, \tag{5}$$

其中 $\alpha(r_i) = \begin{cases} 0, & r_i \in \text{缺失区域}, \\ 1, & r_i \notin \text{缺失区域}, \end{cases}$ 式 (3) 得到求解.

利用式 (3) 完成对 r 的填充后再对缺失区域的其他像素进行填充, 但此时 r 已经从缺失像素点变为有效像素点, 以此类推逐步完成缺失区域全部像素点的填充.

稳定场算子在填充缺失区域像素时, 会充分考虑周围有效像素的信息传递, 因此有效延续了该像素与其周围有效像素之间的语义相关性, 该像素与其周围像素之间不会是突变的, 而是连续的; 同时, 当填充继续深入缺失区域时, 该区域已经填充的像素和已知区域的像素都作为有效像素点共同参与计算, 在尽可能维持特征连续和语义一致的同时又在一定程度上抑制误差的影响.

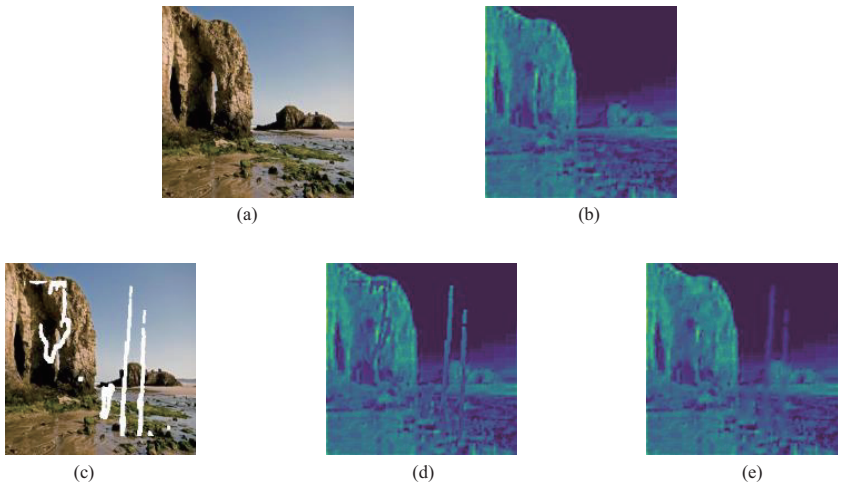

图 4　稳定场算子赋值的结果. (a) 原图; (b) 原图的特征图; (c) 含缺失区域的图像; (d) 含缺失区域图像的特征图; (e) 含缺失区域图像的特征图经稳定场算子填充的结果

Figure 4　Results of the assignment of the stable field operator. (a) Original image; (b) feature map of the original image; (c) image with missing regions; (d) feature map of the image with missing regions; (e) feature map of the image filled by the stable field operator

图 4 示例了稳定场算子对编码器特征图缺失区域的填充结果, 其中图 4(a) 为原图; 图 4(b) 为原图对应的特征图; 图 4(c) 为带有缺失区域的待修复图像; 图 4(d) 为待修复图像对应的未经稳定场算子处理的特征图, 其缺失区域全为零值, 若将该特征图通过跳跃连接传递至解码器, 在解码过程中会将缺失区域中的零值视为有效值进行计算, 导致解码得到的修复结果带有纹理模糊、结构失真的内容; 图 4(e) 为经稳定场算子处理后的特征图, 其缺失区域都被赋予了相对可信的初值, 这些特征图将输入至多尺度融合, 每个尺度的局部纹理得到强化后传递至解码器, 在一定程度上可以改善修复结果产生纹理模糊、结构失真的内容的问题, 有效保证了语义的连贯性.

从图 4(d) 和 (e) 的比较可以看出, 稳定场算子加入跳跃连接可以使传递至解码器的特征图中缺失区域内的零值变为接近于原图特征图的有效值, 避免了零值参与解码过程的卷积计算, 从而有效改善缺失区域和已知区域特征之间的语义相关性以及缺失区域内部的语义一致性不高和特征不连续的问题.

2.1.2　多尺度融合

文献 [12, 14, 18~20, 22] 的跳跃连接都只连接编码器与解码器的对应层, 仅将单层的编码器特征信息传递至对应层的解码器中, 导致低层特征无法感知高层特征的语义信息, 阻碍了修复过程中语义的传递, 导致修复区域的特征不连续和语义不连贯. MSSF-GAN 利用多尺度融合计算逐步加强经稳定场算子填充缺失区域的特征图的传递, 使各层解码器更有效地利用编码器各层特征, 特别是高层语义特征, 以改善修复结果缺乏高层特征信息的问题.

如图 2 所示, 多尺度融合将缺失区域已经被填充的高层特征以渐进的形式逐步向低层特征融合, 提高对各层次特征特别是高层特征的利用率. 定义从高层到低层的特征图分别为 $\mathrm{EF}_1, \mathrm{EF}_2, \mathrm{EF}_3, \ldots$, 经稳定场算子填充后输出特征图为 SFF_L $(L = 1, 2, 3, \ldots)$, 有

$$\mathrm{SFF}_L = \mathrm{SF}\left(\mathrm{EF}_L\right), \tag{6}$$

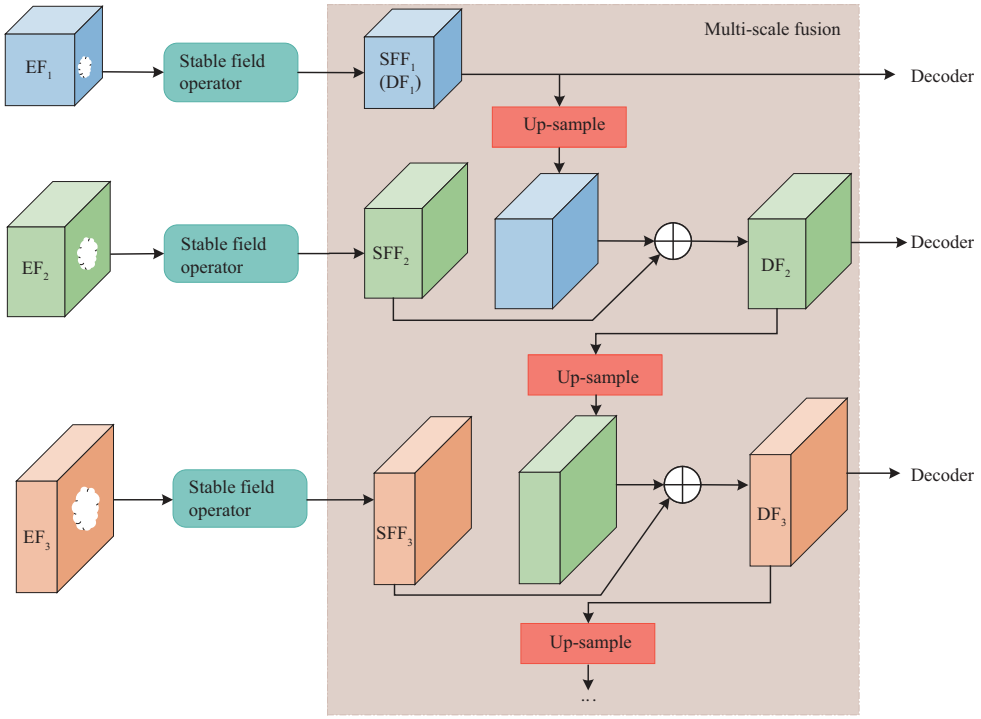

图 5 多尺度融合计算的示意图

Figure 5 Multi-scale fusion

其中 $\mathrm{SF}(\cdot)$ 为稳定场算子. 则多尺度融合计算的过程为

$$\begin{aligned} \mathrm{DF}_1 &= \mathrm{SFF}_1, \\ \mathrm{DF}_2 &= \mathrm{SFF}_2 \oplus \mathrm{Up}\left(\mathrm{DF}_1\right), \\ &\cdots, \\ \mathrm{DF}_L &= \mathrm{SFF}_{L-1} \oplus \mathrm{Up}\left(\mathrm{DF}_{L-1}\right), \end{aligned} \tag{7}$$

其中 DF_L $(L = 1, 2, 3, \ldots)$ 为传递至解码器的特征图, $\mathrm{Up}(\cdot)$ 为上采样计算 (多尺度融合计算过程中的上采样也以插值上采样和卷积的形式代替转置卷积, 参数如表 1 所示).

图 5 以 $L = 3$ 为例并结合稳定场算子说明多尺度融合计算的过程. EF_1 是编码器中的最高层特征图, 对应图 2 中编码器特征图的尺度大小为 8×8, EF_2 和 EF_3 是紧邻 EF_1 之后的两层特征图, 依次对应图 2 中的该层的紧邻左侧两层, 分别经稳定场算子填充缺失区域后得到 SFF_1, SFF_2 和 SFF_3. 此后, 将 SFF_1 传递至解码器的同时对其进行上采样, 再将 DF_1 上采样后与 SFF_2 以逐元素求和的形式融合, 得到特征图 DF_2 后传递至解码器, 同时进行上采样用于与后续特征图融合. 这样跳跃连接传递的特征信息将不会只来自于单一的特征图, 实现了让低层特征感知高层特征的语义信息, 并以逐层融合叠加的方式维持并强化语义的传递.

2.2 判别单元

对于单一的矩形缺失区域, Iizuka 等 [10] 提出使用一个额外的局部判别单元来改善修复效果, 但实际待修复图像中任意位置都可能有多个任意形状的缺失区域, 这种额外的局部判别单元并不适用; 同

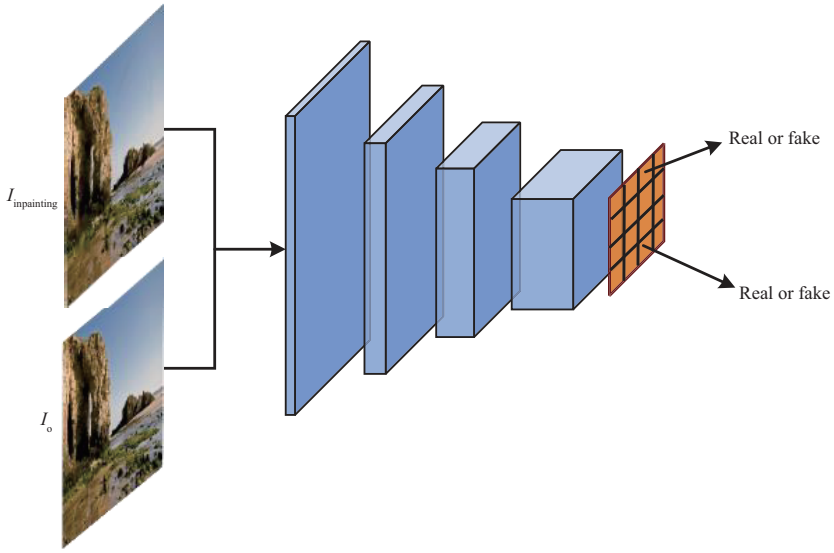

图 6　MSSF-GAN 的判别单元

Figure 6　Discriminator of the MSSF-GAN

表 2　判别单元各层卷积层参数

Table 2　Parameters of the convolution in the discriminator

Layer	Kernel size	Stride	Padding	Input size	Output size	Input channel	Output channel	Receptive field
1	5	2	1	256	127	3	64	5
2	5	2	1	127	63	64	128	13
3	5	2	1	63	31	128	256	29
4	5	1	1	31	29	256	512	61
5	5	1	1	29	27	512	1	93

时 GAN 的判别单元直接将输入样本映射成一个介于 0~1 之间的概率值, 是对整张图像的评价, 并不适合图像修复这类要求高细节保持的领域. 而 PatchGAN [25] 将输入样本映射为矩阵 $\boldsymbol{X}$, 矩阵中每个元素 $\boldsymbol{X}(i,j)$ 都代表了判别单元对输入样本中的一块区域的判别输出, 使得判别单元可以关注到图像细节.

因此本文的 MSSF-GAN 的判别单元以 PatchGAN 为基础设计了一个 5 层卷积层的卷积网络, 如图 6 所示. 判别单元接收真实样本 I_{o} 或图像修复结果 $I_{\text{inpainting}}$ 作为输入, 输出一个特征矩阵, 矩阵中的每个值都对应输入图像不同位置和不同语义的感受野大小的区域的不同评价. 各层卷积层具体参数及感受野大小如表 2 所示, 表中 Input size 和 Output size 分别表示各层卷积层的输入和输出特征图的尺度大小; Input channel 和 Output channel 分别表示卷积层的输入和输出的通道数; Receptive field 表示感受野大小, 即各层输出特征图上的点在输入图像上映射的区域大小. 输出特征矩阵中所有值的感受野可以覆盖整个输入图像, 因此不再需要全局判别单元.

GAN 的训练过程是生成单元和判别单元博弈的过程, 两者都在寻找损失函数比较低的值, 理想状态为两者达到纳什均衡, 但实际上每个模型在更新的过程中 (如生成单元) 成功地降低损失可能会造成博弈的另一个模型 (如判别单元) 损失的升高. 因此 GAN 的训练过程并不稳定, 且难以收敛. 甚至

两者在博弈过程中虽达到了均衡, 但两者都在不断地抵消对方的进步, 使得两者都没有达到一个满意的状态. 而 WGAN [26] 将稳定 GAN 训练的问题转换为求解最优的利普希茨连续 (Lipschitz continuity) 函数的问题, SN-GAN [27] 通过对网络参数进行谱归一化 (spectral normalization) 的方式使网络满足利普希茨连续条件, 从而达到稳定训练和加速收敛的目的. 因此 MSSF-GAN 的判别单元通过引入 SN-GAN 中提出的谱归一化并应用于前 4 层卷积层来进一步稳定模型的训练过程.

2.3 损失函数

损失意味着学习或者训练过程中计算结果与真实值 (期望结果) 的差距, 从而指导下一步的训练或者学习向正确的方向进行, 损失函数定义了损失的计算方式, 与模型的结构和最后的结果密切相关. 由于 MSSF-GAN 将稳定场算子引入跳跃连接并且借助多尺度融合的逐层强化实现特征连续和语义保持, 同时利用各层解码器的特征图输出对应尺度的彩色修复图像以实现过程监控, 构建了全新的图像修复 GAN 的结构, 因此本文细化了损失的类型, 重新定义了损失函数 $\mathcal{L}$ 来优化 MSSF-GAN 的训练过程, 如式 (8) 所示:

$$\mathcal{L} = \lambda_{\text{adv}}\mathcal{L}_{\text{adv}} + \lambda_{\text{pyramid}}\mathcal{L}_{\text{pyramid}} + \mathcal{L}_{\text{rec}}, \tag{8}$$

其中, $\mathcal{L}_{\text{adv}}$ 为对抗损失, $\mathcal{L}_{\text{pyramid}}$ 为金字塔损失, $\mathcal{L}_{\text{rec}}$ 为像素重构损失, $\lambda_{\text{adv}} > 0$ 和 $\lambda_{\text{pyramid}} > 0$ 为对应损失的权重值. 不同的权重使模型能够权衡各部分损失对整体损失的贡献, 考虑到图像修复的结果应尽可能与原图一致, 而像素重构损失 $\mathcal{L}_{\text{rec}}$ 是为了让输出在像素层面上更接近原图, 因此 $\mathcal{L}_{\text{adv}}$ 和 $\mathcal{L}_{\text{pyramid}}$ 的权重需小于 $\mathcal{L}_{\text{rec}}$, 即 $0 < \mathcal{L}_{\text{adv}}, \mathcal{L}_{\text{pyramid}} < 1$. 以下分别讨论.

2.3.1 对抗损失

对抗损失定义对抗学习过程中的损失, 对抗学习的目的是区分输入样本是真实样本还是生成样本, 因此期望损失函数能使得真实样本和生成样本的距离尽可能大. 所以利用 hinge 函数来定义 MSSF-GAN 中判别单元的对抗损失, 如式 (9) 和 (10) 所示:

$$\mathcal{L}_{\text{adv}} = -\mathbb{E}_{I_{\text{inpainting}}\sim P_{\text{data}}(I_{\text{inpainting}})}\left[D\left(I_{\text{inpainting}}\right)\right], \tag{9}$$

$$\mathcal{L}_D = \mathbb{E}_{I_{\text{o}}\sim P_{\text{data}}(I_{\text{o}})}\Big[\text{ReLU}\big(1 - D\left(I_{\text{o}}\right)\big)\Big] + \mathbb{E}_{I_{\text{inpainting}}\sim P_{\text{data}}(I_{\text{inpainting}})}\Big[\text{ReLU}\big(1 + D\left(I_{\text{inpainting}}\right)\big)\Big], \tag{10}$$

其中 D 为判别单元, $P_{\text{data}}\left(I_{\text{o}}\right)$ 为真实样本分布, $P_{\text{data}}\left(I_{\text{inpainting}}\right)$ 为修复图像样本分布, $I_{\text{o}} \sim P_{\text{data}}\left(I_{\text{o}}\right)$ 表示 I_{o} 属于 $P_{\text{data}}\left(I_{\text{o}}\right)$, $I_{\text{inpainting}} \sim P_{\text{data}}\left(I_{\text{inpainting}}\right)$ 同理, $\mathcal{L}_D$ 为判别单元对抗损失, 也为判别单元的总损失.

对于 D 而言, 只有 $D\left(I_{\text{o}}\right) < 1$ 的样本和 $D\left(I_{\text{inpainting}}\right) > -1$ 的样本 (即未被合理区分的样本) 才会对结果产生影响, 因此 hinge 损失函数可以使训练更加稳定, 同时也更快收敛.

2.3.2 像素重构损失

由于 MSSF-GAN 跳跃连接所传递的特征图中的缺失区域不再是随机值或全零值, 而是经稳定场算子计算得到的更为可信的初值; 同时考虑到生成单元输出的生成图像会同时改变输入图像 I_{input} 的缺失区域和已知区域; 因此将像素重构损失 $\mathcal{L}_{\text{rec}}$ 分解成图像缺失区域的像素重构损失 $\mathcal{L}_{\text{hole}}$ 和已知区域的像素重构损失 $\mathcal{L}_{\text{valid}}$. $\mathcal{L}_{\text{hole}}$ 目标是指导缺失区域的输出在语义上更接近原图, $\mathcal{L}_{\text{valid}}$ 目标是使生成单元输出的生成图像在已知区域不会有太大改变, 如式 (11) 所示:

$$\mathcal{L}_{\text{rec}} = \lambda_{\text{hole}}\mathcal{L}_{\text{hole}} + \lambda_{\text{valid}}\mathcal{L}_{\text{valid}}, \tag{11}$$

表 3　数据集训练样本和测试样本的分配

Table 3　Training and test datasets

Dataset	Training samples	Test samples	Total samples
CelebA-HQ [28]	29000	1000	30000
Facade [29]	506	100	606
Place2 [30]	1803460	36500	1839960

其中, $\lambda_{\text{hole}} > 0$ 和 $\lambda_{\text{valid}} > 0$ 为对应的权重值. 考虑到缺失区域的修复内容为所需的内容, 因此权重应满足 $\lambda_{\text{hole}} > \lambda_{\text{valid}}$. $\mathcal{L}_{\text{hole}}$ 和 $\mathcal{L}_{\text{valid}}$ 都采用 L_1 损失来衡量 I_{g} 和 I_{o} 在像素层面的差异, 分别定义为

$$\mathcal{L}_{\text{hole}} = \|I_{\text{g}} \odot M - I_{\text{o}} \odot M\|_1 = \|I_{\text{inpainting}} - I_{\text{o}}\|_1, \tag{12}$$

$$\mathcal{L}_{\text{valid}} = \|I_{\text{g}} \odot (1-M) - I_{\text{o}} \odot (1-M)\|_1 = \|I_{\text{g}} \odot (1-M) - I_{\text{input}}\|_1. \tag{13}$$

2.3.3　金字塔损失

MSSF-GAN 利用多尺度融合计算逐步加强了每个尺度的特征图的传递, 并将各层解码器特征图通过一层卷积层 (卷积参数如表 1 所示) 输出不同尺度的修复图像与真实样本计算 L_1 损失, 从而得到金字塔损失, 可以逐步完善每个尺度对缺失区域的填充, 如式 (14) 所示:

$$\mathcal{L}_{\text{pyramid}} = \sum_{l} \|I_{\text{o}}^{l} - I_{\text{pyramid}}^{l}\|_1, \tag{14}$$

其中, I_{pyramid}^{l} 为 I_{pyramid} 中的各个尺度的图像修复结果, I_{o}^{l} 由 I_{o} 下采样至与 I_{pyramid}^{l} 具有相同尺度大小得到. I_{pyramid} 实现了多尺度融合的过程监控, 使得解码器的过程损失更快地产生效果.

3　实验结果与分析

参考图像修复研究的一般性评价, 这里同样采用定性评价和定量评价来考察本文所提的 MSSF-GAN 的性能, 此外, 还通过消融实验测试 MSSF-GAN 的不同环节对于修复过程的影响. 实验采用公开的人脸数据集 CelebA-HQ [28]、建筑外墙数据集 Facade [29] 和自然场景数据集 Place2 [30]. 数据集训练样本和测试样本分配如表 3 所示, 训练样本和测试样本分辨率大小都为 256 × 256.

实验在定量评价上将 MSSF-GAN 分别与 GL [10], CA [13], PConv [17], LBAM [18], PEN-Net [14] 和 E2I [21] 在不规则掩膜和自然场景数据集的情况下对比, 与 CA [13], CSA [12], Shift-Net [20], PEN-Net [14], UCTGAN [16] 和 PEPSI++ [23] 在居中矩形掩膜和人脸数据集的情况下对比. 正如引言中所述, 上述对比方法都是经典的图像修复方法或比较先进的方法, 如 PConv 和 LBAM 针对不规则掩膜对卷积方式进行改进; E2I 在使用不规则掩膜时具有先进性; CA 首次将注意力机制引入到图像修复中; UCTGAN 和 PEPSI++ 在使用居中矩形掩膜时具有先进性. 与上述对比方法进行对比更能体现 MSSF-GAN 的优越性. 实验中以公开的不规则掩膜数据集 [17] 和居中矩形掩膜 (宽和高都为真实样本的 1/2, 掩膜面积占比 25%) 来拾取图像中的缺失区域从而得到输入 I_{input}:

$$I_{\text{input}} = I_{\text{o}} \odot (1-M). \tag{15}$$

在训练过程中 batch size 设置为 16, 损失函数各部分权重分别为 $\lambda_{\text{adv}} = 0.1$, $\lambda_{\text{pyramid}} = 0.5$, $\lambda_{\text{hole}} = 6$, $\lambda_{\text{valid}} = 1$. 实验采用 Pytorch 框架训练和测试模型, 实验平台配置为 XEON 5112 CPU, NVIDIA GeForce RTX 2080Ti GPU.

图 7 MSSF-GAN 的部分修复结果. 图中每个子图都为一组实验结果, 每个子图中第 1 行为原图, 第 2 行为带有缺失区域的待修复图像, 第 3 行为修复的结果

Figure 7 Some inpainting results of the MSSF-GAN. Each subfigure represents a set of results, in which the first row is an original image, the second row is the image with missing regions, and the third row is the inpainting result. (a–d) Face image inpainting results for images with a centered rectangular mask; (e–h) face image inpainting results for images with an irregular mask; (i, g) scene inpainting results for images with an irregular mask; (k, l) inpainting results for images with a centered rectangular mask

3.1 定性评价

定性评价以主观观察为主, 观察图像修复的效果. 图 7 显示了 MSSF-GAN 在数据集 CelebA-HQ, Facade 和 Place2 上的部分修复结果.

图 7(a)∼(d) 四组图显示了居中矩形掩膜作为缺失区域作用于 CelebA-HQ 上的修复结果, 由于数据集中人脸整体上都处于中间的位置, 因此大部分人脸特征都因被居中矩形掩膜覆盖而丢失, 但 MSSF-GAN 也能得到全局结构和细节纹理合理的修复结果, 且图 7(a)∼(d) 的修复结果都与原图相似, 图 7(c) 的修复结果与原图相差较大, 但也满足人类视觉感知需求. 图 7(e)∼(h) 四组图显示了随机的不规则掩膜作为缺失区域作用于 CelebA-HQ 上的修复结果, 采用不规则掩膜时依然会存在部分人脸特征, 因此可以用这部分特征去指导缺失区域的修复, 得到的修复结果比居中矩形掩膜作为缺失区域时更接近于原图. 在缺失率较大时, 人脸样本修复的结果与原图都很相像, 但同时又有比较明显的差异, 在采用居中矩形掩膜时尤为明显. 原因在于, 稳定场算子利用缺失区域周围的特征信息重建缺失

表 4 居中矩形掩膜在 CelebA-HQ 上的修复性能对比

Table 4 Comparison of inpainting performance of centered rectangular masks on CelebA-HQ[a)]

	CA [13]	CSA [12]	Shift-Net [20]	PEN-Net [14]	UCTGAN [16]	PEPSI++ [23]	MSSF-GAN
PSNR (dB)	24.2377	26.1920	26.0732	25.2693	26.3833	25.5000	**26.7276**
SSIM	0.8671	0.9021	0.8671	0.8958	0.8862	0.8980	**0.9051**
L_1 (%)	2.35	1.68	1.81	1.79	**1.51**	–	1.54

a) Bold means the best.

表 5 不同缺失率的不规则掩膜在 Place2 上的修复性能对比

Table 5 Comparison of inpainting performance of irregular masks of different missing rates on Place2[a)]

	Mask	GL [10]	CA [13]	PConv [17]	LBAM [18]	PEN-Net [14]	E2I [21]	MSSF-GAN
PSNR (dB)	(0.1, 0.2]	23.83	26.27	28.32	**28.51**	26.55	28.42	28.43
	(0.2, 0.3]	20.73	24.21	25.25	**25.59**	24.05	25.16	25.34
	(0.3, 0.4]	18.61	21.95	22.89	23.31	22.47	22.93	**23.38**
	(0.4, 0.5]	17.38	20.02	21.38	21.66	21.48	21.39	**22.22**
	(0.5, 0.6]	16.37	–	19.04	–	17.83	18.36	**19.51**
SSIM	(0.1, 0.2]	0.829	0.876	0.870	0.872	0.884	0.887	**0.896**
	(0.2, 0.3]	0.721	0.763	0.779	0.785	0.799	0.806	**0.815**
	(0.3, 0.4]	0.627	0.657	0.689	0.708	0.721	0.728	**0.744**
	(0.4, 0.5]	0.533	0.572	0.595	0.602	0.604	0.614	**0.633**
	(0.5, 0.6]	0.440	–	0.484	–	0.542	0.544	**0.564**
L_1 (%)	(0.1, 0.2]	3.22	1.43	1.09	1.12	1.14	1.02	**0.86**
	(0.2, 0.3]	5.00	2.38	1.88	1.93	2.11	1.84	**1.75**
	(0.3, 0.4]	6.77	3.59	2.84	**2.55**	3.09	2.78	2.60
	(0.4, 0.5]	8.20	5.22	3.85	3.67	3.85	3.83	**3.55**
	(0.5, 0.6]	9.78	–	5.72	–	6.42	6.18	**5.22**

a) Bold means the best.

区域上的特征信息, 并结合 GAN 来完成对缺失区域的修复, 当缺失区域越大, 稳定场算子可利用的缺失区域周围的特征信息就越少, 对缺失区域中心像素的重建误差就越大, 导致了最后的修复结果在一定程度上偏离原图; 但是依靠 GAN 的训练结果依然可以生成相似的人脸图像. 图 7(i) 和 (j) 两组图显示了随机的不规则掩膜作为缺失区域作用于 Place2 上的修复结果. 图 7(k) 和 (l) 两组图显示了居中矩形掩膜作为缺失区域作用于 Facade 上的修复结果. 总之, MSSF-GAN 在上述 3 个数据集中的修复结果都能有较为逼真且合理的语义.

3.2 定量评价

为了客观评价 MSSF-GAN 的图像修复效果, 本文选取居中矩形掩膜作用于 CelebA-HQ 上和不规则掩膜作用于 Place2 上的两个实验结果进行定量比较. 同时使用图像修复中常用的评估指标 L_1 损失、峰值信噪比 (peak signal to noise ratio, PSNR) 和结构相似度 (structural similarity, SSIM) [31] 来衡量 MSSF-GAN 修复的图像的质量. L_1 和 PSNR 比较图像之间的像素差异, SSIM 比较图像之间亮度、对比度和结构的差异. L_1 的数值越小表示图像修复质量越好, PSNR 和 SSIM 的数值越大表示图像修复质量越好. 表 4 和 5 分别列出了居中矩形掩膜作用于 CelebA-HQ 上的修复评估结果和不同缺失率

(缺失率指图像中缺失的像素个数占全部像素个数的比例) 的不规则掩膜作用于 Place2 上的修复评估结果. 其中, 表 4 中 CA, CSA, Shift-Net 和 UCTGAN 对比模型的数据来源于文献 [16], PEPSI++ 对比模型的数据来源于文献 [23]; 表 5 中使用的不规则掩膜以及 GL 和 PConv 对比模型的数据来源于文献 [17], CA 和 LBAM 对比模型的数据来源于文献 [18]. MSSF-GAN 在表 4 和 5 中的数值由与对比模型在相同数据集和相等数量的测试样本下得到, 分别为 1000 张和 500 张测试样本取均值得到.

居中矩形掩膜作用于 CelebA-HQ 时 MSSF-GAN 的 PSNR 和 SSIM 的数值均优于其他对比模型, 但在 L_1 距离的数值上比 UCTGAN 高 0.03%. 可能原因为 L_1 距离描述的是图像间对应像素点之间的像素值的差异, 并不关注图像整体的结构或语义层面的信息, 而对图像修复而言, 整体的结构和语义信息更为重要.

不规则掩膜作用于 Place2 时在缺失率 10%∼30% 的范围内 MSSF-GAN 的 PSNR 数值均低于 LBAM, 同时在缺失率 30%∼40% 的范围内 L_1 距离的数值也略高于 LBAM. 缺失率的大小对各评估指标的数值有很大影响, 标准卷积在面对不规则缺失区域时图像修复结果会出现一定的伪影、色差、模糊等问题. 由于在样本测试时随机选取掩膜, 上述实验结果数值的偏差一方面可能来源于同一缺失率区间内缺失率选取的随机性, 另一方面可能是由于 MSSF-GAN 采用了标准卷积, 在自然场景图像复杂纹理的修复上出现了一定的伪影和模糊, 而 LBAM 采用部分卷积一定程度上改善了伪影和模糊. 上述两个方面可能导致了在部分指标上 MSSF-GAN 略低于 LBAM. 但整体上 MSSF-GAN 依然优于 LBAM 和其他对比模型.

3.3 基于 CelebA-HQ 数据集的消融实验

为了证明稳定场算子、多尺度融合和各损失函数的有效性, 本文首先分别移除稳定场算子和多尺度融合并在 CelebA-HQ 上进行居中矩形掩膜和 20%∼30% 不规则掩膜的消融实验, 修复评估结果如表 6 所示; 再分别移除损失函数 $\mathcal{L}$ 中的各部分损失并在 CelebA-HQ 上进行居中矩形掩膜和 20%∼30% 不规则掩膜的消融实验, 修复评估结果如表 7 和 8 所示; 最后分别改变各损失的权重并在 CelebA-HQ 上进行居中矩形掩膜的消融实验, 具体来说, 在最优权重为 $\mathcal{L}_{\text{adv}} = 0.1$, $\mathcal{L}_{\text{pyramid}} = 0.5$, $\mathcal{L}_{\text{hole}} = 6$ 和 $\mathcal{L}_{\text{valid}} = 1$ 的基础上, 分别将对抗损失 $\mathcal{L}_{\text{adv}}$ 的权重修改为 0.05 和 1; 将金字塔损失 $\mathcal{L}_{\text{pyramid}}$ 的权重修改为 0.1 和 1; 将缺失区域的像素重构损失 $\mathcal{L}_{\text{hole}}$ 的权重修改为 3 和 8; 将已知区域的像素重构损失 $\mathcal{L}_{\text{valid}}$ 的权重修改为 0.5 和 3, 修复评估结果如表 9 所示. 实验数据均为 1000 张测试样本取均值得到.

从表 6 可以看出, 稳定场算子和多尺度融合都能对修复结果产生不同程度的积极影响, 其中多尺度融合对修复结果产生的积极影响更大, 在移除该部分后 PSNR, SSIM 和 L_1 距离 3 个指标数值改变更大; 稳定场算子对修复结果产生的积极影响虽然比多尺度融合小, 但对整体的修复效果而言也有较大的提升. 稳定场算子在填充缺失区域时会考虑缺失像素周围的有效像素, 而有效像素包括已知区域内的像素和已经被填充的缺失像素, 考虑到掩膜较大时越靠近掩膜中心填充的像素误差越大, 可能导致了稳定场算子对修复结果产生的影响并没有其他两个部分大. 从表 7 和 8 可以看出, 各部分损失都能对修复结果产生不同程度的积极影响, 其中像素重构损失 $\mathcal{L}_{\text{rec}}$ 的影响最大, 而组成像素重构损失的两个部分 $\mathcal{L}_{\text{hole}}$ 和 $\mathcal{L}_{\text{valid}}$ 中, $\mathcal{L}_{\text{hole}}$ 对修复结果的影响更大. 像素重构损失用于指导生成单元生成的结果向原图靠拢, 训练过程中若没有像素重构损失, 生成单元将会漫无目的地生成结果, 难以收敛. 对抗损失 $\mathcal{L}_{\text{adv}}$ 对修复结果的影响虽然不及其余两部分损失, 但对抗损失由生成单元与判别单元之间的对抗博弈产生, 可以进一步指导生成单元生成的结果更真实, 在移除对抗损失后, 生成单元将专注于生成结果与原图之间在像素层面的相似程度, 而不关注生成结果整体的语义连贯性, 虽然使用居中矩形掩膜时移除对抗损失的 L_1 比 MSSF-GAN 低 0.01%, 但修复结果在图像视觉层面并不及 MSSF-GAN,

表 6 稳定场算子和多尺度融合的消融实验

Table 6 Ablation study of the stable field operator and multi-scale fusion[a)]

	Centered rectangular mask			Irregular mask of a 20%∼30% missing rate		
	Without sfo	Without msf	MSSF-GAN	Without sfo	Without msf	MSSF-GAN
PSNR (dB)	25.3992	25.2589	**26.7276**	27.8512	27.1589	**28.8666**
SSIM	0.8998	0.8962	**0.9051**	0.9136	0.9073	**0.9270**
L_1 (%)	1.73	1.79	**1.54**	1.22	1.34	**1.03**

a) "sfo" means stable field operator; "msf" means multi-scale fusion; bold means the best.

表 7 损失函数的消融实验 (使用居中矩形掩膜)

Table 7 Ablation study of the loss function (for images with a centered rectangular mask)[a)]

	Without $\mathcal{L}_{\text{adv}}$	Without $\mathcal{L}_{\text{pyramid}}$	Without $\mathcal{L}_{\text{rec}}$	Without $\mathcal{L}_{\text{hole}}$	Without $\mathcal{L}_{\text{valid}}$	MSSF-GAN
PSNR (dB)	26.3919	25.3811	20.4093	22.4963	25.8945	**26.7276**
SSIM	0.8653	0.9011	0.7930	0.8057	0.8453	**0.9051**
L_1 (%)	**1.53**	1.74	3.32	2.64	1.66	1.54

a) Bold means the best.

表 8 损失函数的消融实验 (使用 20%∼30% 缺失率的不规则掩膜)

Table 8 Ablation study of the loss function (for images with an irregular mask of a 20%∼30% missing rate)[a)]

	Without $\mathcal{L}_{\text{adv}}$	Without $\mathcal{L}_{\text{pyramid}}$	Without $\mathcal{L}_{\text{rec}}$	Without $\mathcal{L}_{\text{hole}}$	Without $\mathcal{L}_{\text{valid}}$	MSSF-GAN
PSNR (dB)	27.7276	27.5728	23.1117	23.5892	27.7522	**28.8666**
SSIM	0.8734	0.9106	0.8294	0.8340	0.8686	**0.9270**
L_1 (%)	1.26	1.26	2.04	1.96	1.23	**1.03**

a) Bold means the best.

表 9 损失函数的权重的消融实验 (使用居中矩形掩膜)

Table 9 Ablation study of weights of the loss function (for images with a centered rectangular mask)[a)]

	$\mathcal{L}_{\text{adv}}$		$\mathcal{L}_{\text{pyramid}}$		$\mathcal{L}_{\text{hole}}$		$\mathcal{L}_{\text{valid}}$		Optimal weight
	0.05	1	0.1	1	3	8	0.5	3	
PSNR (dB)	26.0053	23.1636	25.7851	25.7647	25.7614	25.8856	25.8293	25.6728	**26.7276**
SSIM	0.8498	0.8131	0.8451	0.8442	0.8475	0.8466	0.8463	0.8423	**0.9051**
L_1 (%)	1.64	2.39	1.68	1.68	1.67	1.65	1.67	1.71	**1.54**

a) Bold means the best.

这在 PSNR 和 SSIM 的对比数据中也有所体现. 从表 9 可以看出, 各损失偏大或偏小的权重都会使修复效果下降, 适当大小的权重才能使修复效果达到最优, 改变各损失的权重相当于改变模型在训练过程中各指导方向的地位, 如将对抗损失变小将会使模型更专注于在像素层面上生成与原图接近的样本, 而不关注生成结果整体的语义连贯性. 另外, 由于使用居中矩形掩膜时人脸特征基本丢失, 因此在不规则掩膜下的修复结果更接近于原图, 这在表 6∼8 中得到体现.

4 结语

本文提出了一种基于多尺度稳定场的 GAN 的图像修复模型, 通过将稳定场模型嵌入到跳跃连接中, 利用缺失区域周围的已知区域的特征信息填充编码器中特征图的缺失区域, 保证了缺失区域的语义一致性和特征连续性, 改善了图像修复的结果缺失区域语义不连贯的问题; 再通过多尺度融合计算逐步加强特征图的传递, 使得低层特征能够感知高层特征的语义信息, 提高模型对编码器中各层特征的利用率, 从而提升图像的修复效果; 最后将解码器中的各层特征图都输出对应的修复图像以计算金字塔损失, 从而逐步完善每个尺度的特征图的填充. 实验表明, 本文模型在人脸图像、自然场景图像和正面建筑图像上都能获得合理语义信息的修复结果. 但当缺失率较大时, 稳定场算子对编码器特征图的重建误差增大, 且越靠近缺失区域的中心误差越大, 导致了特征图的填充结果与原图的特征图有一定差异, 仅能得到相似的图像修复结果; 另外, 本文模型建立在 256×256 分辨率图像的基础上进行修复, 在面对高分辨率图像缺失像素大量增加的情况下, 稳定场算子对于缺失区域内部像素的预测误差可能会进一步提升, 对最后的图像修复结果也会造成相应的影响. 未来可尝试将稳定场算子进行改进以及将模型应用于高分辨率图像修复.

参考文献

1 Bertalmio M, Sapiro G, Caselles V, et al. Image inpainting. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, New Orleans, 2000. 417–424
2 Chan T F, Shen J. Nontexture inpainting by curvature-driven diffusions. J Vis Commun Image Represent, 2001, 12: 436–449
3 Shen J, Chan T F. Mathematical models for local nontexture inpaintings. SIAM J Appl Math, 2002, 62: 1019–1043
4 Barnes C, Shechtman E, Finkelstein A, et al. PatchMatch: a randomized correspondence algorithm for structural image editing. ACM Trans Graph, 2009, 28: 1–11
5 Li X H, Wang L Y, Cheng Q, et al. Cloud removal in remote sensing images using nonnegative matrix factorization and error correction. ISPRS J Photogrammetry Remote Sens, 2019, 148: 103–113
6 Singh P, Komodakis N. Cloud-GAN: cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks. In: Proceedings of IEEE International Geoscience and Remote Sensing Symposium, Valencia, 2018. 1772–1775
7 Pathak D, Krähenbühl P, Donahue J, et al. Context encoders: feature learning by inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 2536–2544
8 Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, 2014. 2672–2680
9 Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Proceedings of the Medical Image Computing and Computer-Assisted Intervention, Munich, 2015. 234–241
10 Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion. ACM Trans Graph, 2017, 36: 1–14
11 Zhang H R, Hu Z Z, Luo C Z, et al. Semantic image inpainting with progressive generative networks. In: Proceedings of the 26th ACM International Conference on Multimedia, Seoul, 2018. 1939–1947
12 Liu H Y, Jiang B, Xiao Y, et al. Coherent semantic attention for image inpainting. In: Proceedings of the IEEE International Conference on Computer Vision, Seoul, 2019. 4170–4179
13 Yu J, Lin Z, Yang J, et al. Generative image inpainting with contextual attention. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 5505–5514
14 Zeng Y H, Fu J L, Chao H Y, et al. Learning pyramid-context encoder network for high-quality image inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, 2019. 1486–1494
15 Yang H, Yu Y. Image inpainting using channel attention and hierarchical residual networks. J Comput Aided Des Comput Graph, 2021, 33: 671–681 [杨昊, 余映. 利用通道注意力与分层残差网络的图像修复. 计算机辅助设计与图形学学报, 2021, 33: 671–681]
16 Zhao L, Mo Q H, Lin S H, et al. UCTGAN: diverse image inpainting based on unsupervised cross-space translation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, 2020. 5740–5749
17 Liu G L, Reda F A, Shih K J, et al. Image inpainting for irregular holes using partial convolutions. In: Proceedings of European Conference on Computer Vision, Munich, 2018. 89–105

18 Xie C H, Liu S H, Li C, et al. Image inpainting with learnable bidirectional attention maps. In: Proceedings of the IEEE International Conference on Computer Vision, Seoul, 2019. 8857–8866
19 Yang W X, Wang M, Zhang L. Semantic face image inpainting based on U-Net with dense blocks. J Comput Appl, 2020, 40: 3651–3657 [杨文霞, 王萌, 张亮. 基于密集连接块 U-Net 的语义人脸图像修复. 计算机应用, 2020, 40: 3651–3657]
20 Yan Z, Li X, Li M, et al. Shift-Net: image inpainting via deep feature rearrangement. In: Proceedings of European Conference on Computer Vision, Munich, 2018. 3–19
21 Xu S X, Liu D, Xiong Z W. E2I: generative inpainting from edge to image. IEEE Trans Circuits Syst Video Technol, 2021, 31: 1308–1322
22 Liao N H, Zhang X J, Peng C Y, et al. LRGAN: a computational lightweight image inpainting neural network. Sci Sin Tech, 2022, 52: 447–457 [廖年鸿, 张效娟, 彭春燕, 等. LRGAN: 一种运算轻量化图像修复网络. 中国科学: 技术科学, 2022, 52: 447–457]
23 Shin Y-G, Sagong M-C, Yeo Y-J, et al. PEPSI++: fast and lightweight network for image inpainting. IEEE Trans Neural Netw Learn Syst, 2021, 32: 252–265
24 Ye X Y, Qi Z Z, He Z W, et al. Reconstructing image local regions based on directional derivative of a field. J Image Graph, 2014, 19: 998–1005 [叶学义, 齐珍珍, 何志伟, 等. 图像场方向导数的局部区域重建. 中国图像图形学报, 2014, 19: 998–1005]
25 Isola P, Zhu J Y, Zhou T H, et al. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 1125–1134
26 Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In: Proceedings of the 34th International Conference on Machine Learning, Sydney, 2017. 214–223
27 Miyato T, Kataoka T, Koyama M, et al. Spectral normalization for generative adversarial networks. In: Proceedings of International Conference on Learning Repressentations, Vancouver, 2018
28 Tero K, Timo A, Samuli L, et al. Progressive growing of GANs for improved quality, stability, and variation. In: Proceedings of International Conference on Learning Representations, Vancouver, 2018
29 Radim T, Rdaim S. Spatial pattern templates for recognition of objects with regular structure. In: Proceedings of German Conference on Pattern Recognition, Saarbrücken, 2013. 364–374
30 Zhou B, Lapedriza A, Khosla A, et al. Places: a 10 million image database for scene recognition. IEEE Trans Pattern Anal Mach Intell, 2018, 40: 1452–1464
31 Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process, 2004, 13: 600–612

Image inpainting based on multi-scale stable-field GAN

Xueyi YE*, Maosheng ZENG*, Weijie SUN, Lingyu WANG & Zhijin ZHAO

School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018*, China*
* Corresponding author. E-mail: xueyiye@hdu.edu.cn, zms_0310@163.com

Abstract Generative adversarial networks (GANs) have shown their potential to inpaint large missing regions and generate plausible semantic results in image-inpainting tasks. However, GAN-based methods often ignore the semantic consistency and feature continuity of missing regions, and lack the perceptibility of features at multiple scales. To address these issues, we propose an image-inpainting model based on GAN with a multi-scale stable field. Motivated by the U-Net architecture, its generator embeds the stable field operator into skip connections to fill the missing region in the feature map of the encoder and thus maintain the semantic consistency and feature continuity of the missing region. Further, as the benefit is gradually enhanced by multi-scale fusions, the feature information transferred by the skip connections is not obtained from only a single feature map. The model is generally enabled to perceive the semantic information of the high-level features of an image. The experimental results show that the proposed model outperforms other classical image-inpainting methods in the face and natural scene datasets.

Keywords image inpainting, generative adversarial network (GAN), stable field, multi-scale fusion, deep learning