基于扩散模型的人脸图像修复技术

郭庚辰[1] 姚剑敏[1,2] 严 群[1,2] 林智贤[1] 刘德崇[1]
GUO Gengchen YAO Jianmin YAN Qun LIN Zhixian LIU Dechong

摘 要 人脸作为人体信息最为密集的部位，人脸图像在各个研究领域都有不可替代的作用。因此，研究如何将遮挡或模糊的人脸图像恢复成真实图像是非常有意义的。针对人脸图像修复技术的研究，提出基于扩散模型的人脸图像修复技术。在现有的人脸图像修复技术基础上，解决在修复大区域破损或遮挡的情况下，修复图像出现纹理模糊及结构扭曲等问题。所提出的方法基于边缘引导的扩散模型图像修复网络，主要包括两个阶段：首先训练基于 U-Net 结构的边缘修复模型生成较为真实的缺失区域的边缘信息，然后根据已修复好的边缘信息，训练内容生成模型填充缺失部分的内容信息。实验证明对于人脸图像修复具有较好的效果。

关键词 扩散模型；人脸补全；自注意力机制；图像修复；边缘引导；混合注意力机制

doi：10.3969/j.issn.1672-9528.2024.03.048

0 引言

随着社会发展进入数字化时代，对于数字社会中广泛应用的人脸图像，解决其质量问题和应对各种挑战是保障信息安全和社会运行的重要一环。由于人脸可以标记识别个体，所以广泛地应用在比如安防系统中的人脸识别、证件办理（如身份证、护照等）、社交媒体上的照片分享等[1]。这些应用对于保障个人安全、提高工作效率和改善用户体验都具有重要作用。图像捕获时的环境条件可能因为光照、角度、遮挡等因素而不理想，导致人脸图像质量下降。这些挑战使得在处理和分析这些图像时需要采用先进的图像处理和计算机视觉技术。人脸图像修复算法在这一背景下显得尤为重要，为复杂多样的社会生活场景提供了必要的技术支持。

图像修复算法通过解析图片的上下文信息，推断出图像损失部分的语义内容以及细节。要求修复得到的图像结构纹理清晰，完整与缺失边缘部分过渡平滑，补充内容合理，尽量还原图像语义内容。而由于人脸图像在信息含义、结构、语义等方面的特殊性，使其相对于普通的图像修复任务有着更为巨大的挑战，例如人脸的某个部位被遮挡，这就无法在图像已知区域得到相似的信息进行填充，只能依赖于数据集中其他图像的人脸信息，且人脸具有更为复杂的纹理结构，这都需要图像修复技术更为精准地修复人脸图像。

1. 福州大学 福建福州 350108
2. 晋江市博感电子科技有限公司 福建晋江 362200
[基金项目] 国家重点研发计划（2022YFB3603503），福建省技术攻关重点项目（2023G007）

传统的图像修复技术主要使用的是扩散[2]和补丁[3]两种方法，但两者都有很大的局限性，比如在修复后的区域在视觉上不自然，尤其在边缘过渡区域，表现为纹理不连续或颜色不匹配。而且在大面积缺失的情况下无法提供足够的上下文信息来进行准确的修复。所以在进行人脸复杂图像修复任务中，特别是需要考虑全局上下文信息和复杂结构的情况下，基于深度学习的图像修复方法逐渐成为更为强大和灵活的替代方案。深度学习方法可以自动学习图像中的特征和结构，并提供更准确、自然的修复效果。

人脸图像修复技术本质上是条件引导的图片生成图片技术，现有深度学习图像生成技术主要分为 4 类，生成对抗网络（generative adversarial models，GAN）[4]、变分自编码器（variance auto-encoder，VAE）[5]、标准化流模型（normalization flow，NF）[6] 以及扩散模型（diffusion models，DM）[7]。扩散模型技术在条件引导图像生成技术方面具有生成质量高和多样性强的优点，基于此现状，本文采用的是以扩散模型为基本架构与混合注意力机制相融合的人脸图像修复技术在较大面积破损的人脸图像修复上有着较为优秀的表现。

1 相关理论

1.1 扩散模型

扩散模型的算法理论基础是通过变分推断（variational inference）训练参数化的马尔可夫链（markov chain），它在许多任务上展现了超过 GAN 等其他生成模型的效果，例如最近非常火热的 OpenAI 的 DALL-E 2，Stability.ai 的 Stable Diffusion 等。

最早扩散模型的提出是由 Jascha Sohl-Dickstein 等人在 2015 年提出 [8]，其目的是消除对训练图像连续应用的高斯噪声，可以将其视为一系列去噪自编码器。真正让扩散模型成为图像生成的主流是在 2020 年 DDPM 的提出。扩散模型包括两个步骤。

（1）固定的（或预设的）前向扩散过程 q：该过程会逐渐将高斯噪声添加到图像中，直到最终得到纯噪声。

（2）可训练的反向去噪扩散过程：训练一个神经网络，从纯噪音开始逐渐去噪，直到得到一个真实图像。

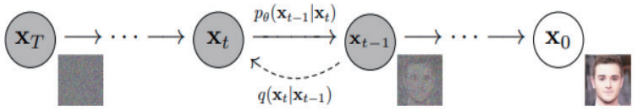


图 1 扩散模型基本结构图

前向与后向的步数由下标 t 定义，并且有预先定义好的总步数 T（DDPM 原文中为 1000）。t=0 时为从数据集中采样得到的一张真实图片，t=T 时近似为一张纯粹的噪声。

1.1.1 前向过程

$q(x_0)$ 是真实数据分布（也就是真实的大量图片），从这个分布中采样即可得到一张真实图片 $x_0 \sim q(x_0)$。定义前向扩散过程为 $q(x_t|x_{t-1})$，即每一个 step 向图片添加噪声的过程，并定义好一系列$0 < \beta_1 < \beta_2 < \cdots < \beta_t < 1$，则有：

$$q(x_t|x_{t-1}) = \boldsymbol{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t \boldsymbol{I}) \quad (1)$$

式中：N 为正态分布，均值和方差分别为$\sqrt{1-\beta_t}$和β_t，因此通过采样标准正态分布，有：

$$x_t = \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t}\epsilon \quad (2)$$

图 2 前向过程图

1.1.2 逆向过程

反向过程的核心在于如何得到 $q(x_t|x_{t-1})$ 的逆过程 $q(x_{t-1}|x_t)$，这个过程无法直接求出来，所以本文使用神经网络去拟合这一分布。本文使用一个具有参数的神经网络去计算 $q_\theta(x_{t-1}|x_t)$，假设反向的条件概率分布也是高斯分布，且高斯分布实际上只有两个参数：均值和方差，那么神经网络需要计算的实际上是：

$$q_\theta(x_{t-1}|x_t) = N(x_{t-1}, \mu_\theta(x_t, t), \sum_\theta(x_t, t)) \quad (3)$$

1.2 注意力机制

注意力机制（atention mechanism）是一种在计算机科学和人工智能领域中使用的技术，它模拟人类在处理信息时的注意力分配过程。这种机制最初是由神经网络领域引入的，特别是在自然语言处理和计算机视觉任务中取得了显著的成功。

在深度学习中，注意力机制的核心思想是模型能够在处理输入数据时选择性地关注或强调其中的特定部分，而不是简单且一视同仁地对待所有输入。更为重要的是，注意力机制能够对可变长度的数据进行全局建模，大大增强了网络捕获远距离依赖关系的能力，协助网络更好地理解图像信息。因此 Yu 等人在 2018 年提出将注意力机制引入图像修复任务中，自此之后，注意力机制便成为了图像修复领域中的基础方法，各种针对于注意力机制的改进算法层出不穷。对于图像修复任务而言，由于图像中不同区域的特征对于预测不同位置像素值的影响程度不同，因此在图像修复过程中，需要通过注意力机制来判断图像的全局特征对于当前局部特征的影响系数。

本文采用多尺度卷积融合模块 [9] 与混合注意力机制 [10]，可以在保持图像整体视觉连通性的同时也可以保证完整与破损过渡区域的连贯性，以及符合语义的人脸五官部位图像。

2 本文模型

2.1 边缘生成网络

为了尽可能捕获图像纹理细节，生成大面积不规则缺失的边缘，本文采用的边缘修复网络由 12 个普通卷积和 4 个残差卷积与一个自注意力模块组成。残差卷积的引入可以有效扩大感受野大小，自注意力机制通过自适应调节目标特征与周围特征相关性权重方式加强特征之间的联系，以保证前景和背景的一致性。

边缘生成网络 G_e 的输出边缘结果为：

$$\boldsymbol{B}_{pre} = G_e(B_{canny}, M_g, I) \quad (4)$$

式中：B_{canny} 为 Canny 算子得到的掩膜边缘图像；M_g 为掩膜灰度图像；I 为掩膜图像；B_{pre} 为边缘生成网络的输出。

2.2 人脸补全网络

人脸补全网络由 12 个普通卷积与 4 个并行的空洞卷积以及 1 个混合注意力模块。其中 4 个空洞卷积核大小最初都为 3*3，扩张率一次为 1、2、4、8，经过扩展后，每个卷积核可以变成 3*3、7*7、15*15、31*31，图 3 展示了该空洞卷积模块的结构，通过 4 个空洞卷积块可以得到更为丰富的全局和局部上下文信息，从而保证输出图像的整体连通性与破损区域与完整区域的语义一致性。通过图 3 可以发现经过空洞卷积模块后，可以得到 4 个不同扩张率空洞卷积所得到的特征信息，如果直接将这 4 个特征图拼接输出，可能会导致这些通道的相关性较差。基于此引入混合注意力模块可以通过对特征图的不同通道以及不同区域有不同的关注度来解决这一问题。图 4 为基于注意力机制的空洞卷积融合模块，该

模块能够获取图像上丰富的远程上下文信息，通过具有不同感受野的卷积核对特征图的学习，能够从不同视角预测缺失区域的像素结果。

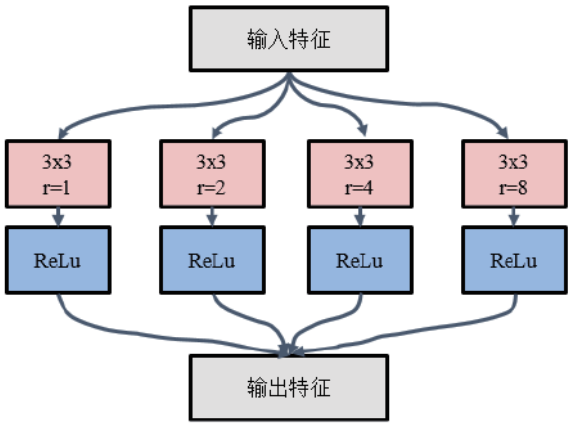


图 3 空洞卷积融合模块

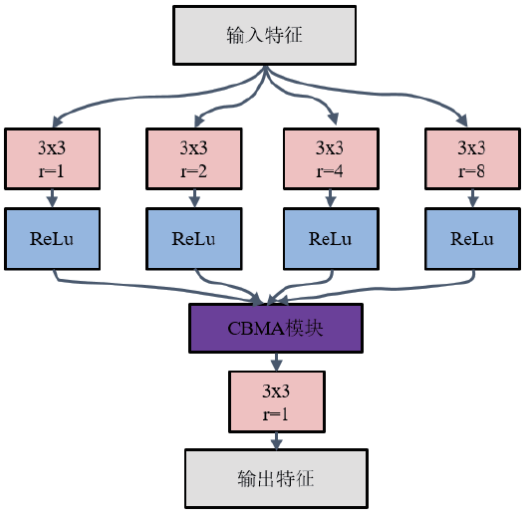


图 4 基于注意力机制的空洞卷积融合模块

3 实验结果分析

3.1 数据集准备

本次实验所采用的数据集为 Helenface[11] 和 CelebA-HQ[12]。Helen Face 数据集是一个用于人脸关键点检测的数据集，其中包含了 2330 张人脸图片，涵盖了不同的姿态、表情、光照等多种条件。每张图片都有对应的标签，包括人脸关键点位置信息。这个数据集是由 Helen 课题组在 2011 年开发的，用于训练和评估人脸关键点检测算法。在训练图像修复算法中，由于 Helen Face 数据集中的图像涵盖了不同的姿态、表情、光照等多种条件，也可以用于训练人脸修复算法。CelebA-HQ 数据集是 CelebA 数据集的升级版，是一个由高分辨率人脸图像和相关属性标签组成的数据集。CelebA-HQ 数据集中每张图片都有多种属性标签，这可以让图像修复算法在修复图像的同时考虑属性信息，如性别、年龄、发色等，从而更好地修复图像，更接近真实人脸的样子。

3.2 实验设置

本次实验所采用的硬件配置为 NVIDIA Tesla P40 GPU（16 GB），采用 adam 优化器来提高模型训练的稳定性。编译软件使用 PyCharm。所设置的迭代次数为 2000，学习率为 2e-4，每批次喂入 64 对图像。

在深度学习领域，常用的两大学习框架分别为 Tensorflow 与 PyTorch，选择合适的学习框架会使实验达到事半功倍的效果。PyTorch 使用动态计算图，这意味着在运行时可以动态构建、修改和调整计算图，使得调试和实验更加直观。而 Tensorflow 使用静态计算图，计算图在构建阶段被定义，然后在执行阶段进行调用。这提供了一些优化的机会，但也可能导致较为繁琐的调试和修改过程。相比而言 PyTorch 相对于 Tensorflow 有更好的灵活性。因此这里选择 PyTorch 作为本次实验的学习框架。

3.3 评价指标

人脸图像修复是一种特殊的图像处理技术，由于人脸图像修复的最终目标是产生质量高、外观自然的图像，而人的主观感受对于图像的质量评价具有决定性的影响。当本文评估人脸图像修复的效果时，客观数据并不总能提供准确的结果。因此，在人脸图像修复领域，本文通常会采用主观视觉评估为主、客观数据评估为辅的方法来判断图像修复的效果。客观数据评价通常采用两个指标来衡量：峰值信噪比（PSNR）和结构相似性（SSIM）。

峰值信噪比（PSNR）是一种常用于衡量图像质量的客观评价指标，特别是在图像重建和复原领域中经常使用。PSNR 的计算基于图像的原始版本和重建版本之间的均方误差（mean squared error，MSE）。PSNR 的公式为：

$$M_{\mathrm{MSE}} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{i=0}^{n-1}[I(i,j)-K(i,j)]^2 \quad (5)$$

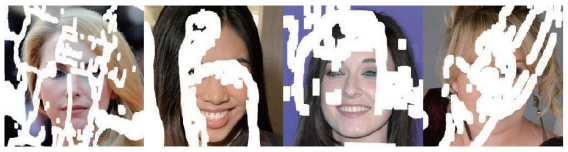
$$P_{\mathrm{PSNR}} = 10\log_{10}\left(\frac{M_{\mathrm{MAX}}{}^{I}_{2}}{M_{\mathrm{MSE}}}\right) \quad (6)$$

式中：m 和 n 分别代表图像的长和宽；I 和 K 分别代表修复后的图像和真实图像。

结构相似性（SSIM）是一种用于评价图像质量的客观评价指标。它考虑了图像的亮度、对比度和结构等方面的相似性，与人眼对图像的感知相对应。SSIM 的值范围在 [0, 1] 之间，值越接近 1 表示图像越相似。

3.4 结果分析

图 5 为本次实验得到的人脸图像修复效果对比图，表 1 为客观评价指标对比结果，通过将先前经典两种图像修复方法 PIC 和 RFR 作为对照组，可以得到本文所提到的人脸图像修复结果要优于先前两种方法。



（a）缺损图像

（b）PIC 结果图像

（c）RFR 结果图像

（d）本文修复结果

图 5 实验对比结果

表 1 客观评价指标对比结果

图像修复模型	SSIM	PSNR
PIC	0.891	28.19
RFR	0.901	28.46
OURS	0.937	28.97

4 结论与展望

针对人脸图像修复问题，本文提出一种基于扩散模型的修复模型，该模型由边缘生成网络与人脸补全网络组成，结合边缘引导、多尺度空洞卷积以及混合注意力机制，在人脸修复问题上有着较为优异的表现。

本文对纹理模糊、结构不清晰、边界衔接不融洽等问题进行了改善。证明扩散模型有对于高复杂度的人脸图像修复能力。但也存在一些不足与需要改进，之后的工作可以在以下两方面考虑。

（1）对于扩散模型生成图像速度较慢的问题。由于扩散模型是迭代式生成过程，每个迭代步骤都需要进行多次的采样和计算，这会增加计算的复杂度和时间开销。因此下一步准备将重心放在提高扩散模型的迭代速度上。

（2）对于头发部位的修复一直是人脸图像修复技术中的难点，原因在于头发通常包含很多细小的细节，如毛发、光泽等，为了使修复的头发看起来自然，需要考虑这些微小的细节。光照和阴影的影响以及在头发与面部皮肤的交界处，需要保持自然的过渡。因此针对人脸图像修复中头发的修复问题可作为下一步的研究方向。

参考文献：

[1] 孙琪，翟锐，左方，等．基于部分卷积和多尺度特征融合的人脸图像修复模型 [J]. 计算机工程与科学 ,2023,45(2):304-312.

[2]BERTALMIO 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.

[3]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.

[4]IAN G, JEAN A P, MEHDI M, et al. Generative adversarial networks[EB/OL].(2014-06-10)[2023-11-01].https://arxiv.org/abs/1406.2661.

[5]KINGMA P D, WELLING M. Auto-encoding variational bayes[EB/OL]. (2013-11-20)[2023-11-18].https://arxiv.org/abs/1312.6114.

[6]DINH L, SOHL-DICKSTEIN J, BENGIO S. Density estimation using real NVP[EB/OL].(2016-05-27)[2023-11-22]. https://arxiv.org/abs/1605.08803.

[7] HO J , JAIN A , ABBEEL P .Denoising diffusion probabilistic models[EB/OL].(2020-06-19)[2023-11-26]. https://arxiv.org/abs/2006.11239.

[8]SOHL-DICKSTEIN J ,WEISS E A, MAHESWARANATHAN N, et al. Deep unsupervised learning using nonequilibrium thermodynamics[EB/OL].(2019-08-02)[2023-09-28]. https://arxiv.org/abs/1503.03585.

[9] 贺川圳．基于多尺度卷积融合的人脸图像修复方法研究 [D]. 成都：电子科技大学 ,2023.

[10] 兰治，严彩萍，李红，等．混合双注意力机制生成对抗网络的图像修复模型 [J]. 中国图象图形学报，2023, 28(11): 3440-3452.

[11]RAZAVIAN A S, AZIZPOUR H,SULLIVAN J, et al. CNN features off-the-shelf: an astounding baseline for recognition[C]//In IEEE Conference on Computer Vision and Pattern Recognition Workshops.Piscataway: IEEE, 2014: 512-519.

[12] KARRAS T, AILA T, LEHTINEN J. Progressive growing of gans for improved quality, stabilityand variation[EB/OL]. (2017-10-27)[2023-10-29]. https://arxiv.org/abs/1710.10196.

【作者简介】

郭庚辰（1998—），男，山西运城人，研究方向：深度学习、图像处理等。

姚剑敏（1978—），男，福建莆田人，博士，副研究员，研究方向：人工智能、图像处理、计算机视觉等。

（收稿日期：2024-01-23）