

中图法分类号: TP391; TP18 文献标识码: A 文章编号: 1006-8961(2026)08-2868-14

论文引用格式: Liu J X, Li S, Qian Z X, Zhang X P and Ying Q C. 2026. Face retouching restoration based on a diffusion model. Journal of Image and Graphics, 31(8): 2868-2881 (刘佳欣, 李晟, 钱振兴, 张新鹏, 应祺超. 2026. 基于扩散模型的美颜人脸图像复原. 中国图象图形学报, 31(8): 2868-2881) [DOI: 10. 11834/jig. 250242]

基于扩散模型的美颜人脸图像复原

刘佳欣, 李晟*, 钱振兴, 张新鹏, 应祺超

复旦大学计算与智能创新学院, 上海 200082

摘要: **目的** 随着人脸美颜技术的广泛应用, 社交网络中充斥了大量经过各种类型美颜操作的人脸图像。人脸美颜技术的使用方便快捷, 在为用户提升视觉体验的同时, 也给数字内容的真实性带来了严峻挑战, 并可能应用于网络诈骗、虚假宣传等领域, 带来安全隐患。当前研究主要集中于美颜检测领域, 即判断图像是否经过美颜处理, 而缺乏对如何从美颜图像中复原原始人脸的探索。本文提出了美颜人脸图像复原 (face retouching restoration, FRR) 方法, 旨在去除因美颜处理导致的色彩纹理改变和结构特征扭曲, 从而重建原始人脸图像以便于追溯原始身份特征。**方法** 本文提出了基于美颜痕迹引导去噪过程的扩散模型 (retouching-guided diffusion model for FRR, RT-FRR)。通过多尺度篡改痕迹提取模块 (multi-scale face retouching trace extraction module, FREM) 提取美颜图像中的美颜处理痕迹特征, 并构建多尺度条件引导机制控制扩散过程。RT-FRR 采用残差预测策略, 在去噪过程中预测原始人脸图像和美颜图像之间的人脸残差, 从而实现面部纹理细节和结构特征的协同恢复。**结果** 在新构建的重度美颜人脸数据集 (extremely retouched face dataset, ExRF) 和现有公开美颜检测数据集 RetouchingFFHQ 上分别进行了实验。实验结果表明, 相比现有图像复原方案, 本文方法在常用图像质量评价指标上有显著提升。对人脸特征相似度的分析也表明本文方法可以有效恢复原始人脸的身份。**结论** 本文所提的 RT-FRR 能有效预测原始人脸和美颜人脸之间的残差, 并在视觉质量和人脸识别准确率方面均表现出色, 显著优于基线模型。

关键词: 图像复原; 美颜人脸复原 (FRR); 扩散模型; 注意力机制; 残差预测

Face retouching restoration based on a diffusion model

Liu Jiaxin, Li Sheng*, Qian Zhenxing, Zhang Xinpeng, Ying Qichao

Computation and Artificial Intelligence Innovative College, Fudan University, Shanghai 200082, China

Abstract: Objective The widespread use of face-retouching techniques across social media platforms has generated significant societal concerns that extend beyond aesthetics, giving rise to psychological consequences, such as body dysmorphic disorder, undermining trust in digital imagery and reinforcing pervasive cultural norms surrounding beauty standards. Face-retouching is widely perceived as more convenient than face swapping and has been integrated into numerous social network applications. Although existing research has made substantial progress in detecting retouched faces using a range of computational approaches, the critical task of reversing these modifications to reconstruct original facial features and restore authenticity in digital images remains largely unexplored. This disparity highlights a significant gap between detection capa-

收稿日期: 2025-06-11; 修回日期: 2026-01-13; 预印本日期: 2026-01-20

* 通信作者: 李晟 lisheng@fudan.edu.cn

基金项目: 国家自然科学基金项目 (62572125, U22B2047, 62450067); 上海市自然科学基金项目 (25ZR1401019); 中国博士后科学基金项目 (2025M771574)

Supported by: National Natural Science Foundation of China (62572125, U22B2047, 62450067); Shanghai Municipal Natural Science Foundation (25ZR1401019); China Postdoctoral Science Foundation (2025M771574)

bilities and the development of practical restoration solutions. This study formally introduces face retouching restoration (FRR) as a novel computer vision task, distinguished from related problems, such as makeup restoration and general image restoration, by its explicit focus on subtle textural modifications and extreme structural alterations characteristic of retouched facial images. The primary objective is to accurately recover authentic facial characteristics from heavily manipulated images by addressing two key technical challenges: restoring fine-grained textures, such as skin smoothing, and reconstructing facial geometry, including eye enlargement and jaw reshaping. This dual focus on texture and structure advances beyond existing methods that typically address only a single aspect of facial modification. Furthermore, this study aims to establish standardized evaluation protocols and datasets for this emerging research direction, recognizing that existing benchmarks designed for retouching detection are inadequate for assessing restoration quality. The proposed approach advances digital media forensics by providing practical tools for content authentication while simultaneously contributing to the broad discussions on image manipulation in the digital age. This study aims to shift from passive detection to active restoration, with potential applications in forensic analysis, psychological counseling, and digital media literacy initiatives.

Method The proposed retouching-guided diffusion model for FRR (RT-FRR) introduces a novel dual-component architecture that synergistically integrates diffusion-based generation modeling with targeted retouching artifact analysis. The framework incorporates a face retouching estimation module (FREM) that utilizes constrained BayarConv layers with $5 \times 5 \times 3$ kernels, specifically designed to amplify manipulation artifacts while suppressing authentic facial features. This preprocessing stage extracts multi-scale retouching traces through four successive ResNet blocks, capturing local pixel-level inconsistencies and global structural alterations. These hierarchical features serve as conditional guidance for the second component—a denoising diffusion probabilistic model configured with 1 000 timesteps that learns to predict the precise residual between retouched and original facial images. The model's U-Net backbone incorporates innovative cross-attention mechanisms at four distinct network depths, allowing progressive integration of retouching traces with diffusion priors throughout the denoising process. This multi-scale attention strategy facilitates simultaneous recovery of high-frequency textural details, such as skin pores and fine wrinkles, and macroscopic structural features, including facial contours and organ geometry. Training optimization utilizes the Adam algorithm across 100 000 iterations with a fixed learning rate of 10^{-4} using a composite loss function that balances pixel-wise reconstruction accuracy with perceptual quality metrics. The authors created the extreme retouched face dataset (ExRF) comprising 1 000 high-resolution image pairs with severe yet plausible retouching operations, manually generated using professional editing software with consistent extreme parameter settings, for a comprehensive evaluation. This dataset complements the existing RetouchingFFHQ benchmark, ensuring robust validation across the full spectrum of retouching intensities from subtle enhancements to dramatic transformations. The framework operates on 512×512 resolution images, maintaining architectural consistency with state-of-the-art diffusion models while being specifically optimized for facial feature recovery tasks. **Result** Experimental evaluation demonstrates that RT-FRR achieves state-of-the-art performance in face retouching restoration, significantly outperforming existing methods across multiple quantitative and qualitative metrics. On the ExRF dataset, which features significantly retouched faces, RT-FRR achieves a PSNR of 25.23 dB, surpassing the second-best method ResDiff by 1.9 dB, along with superior scores in SSIM at 0.806, FID at 48.802, and LPIPS at 0.241. These results highlight the model's effectiveness in recovering subtle textural details and major structural modifications, such as eye enlargement, nose reshaping, and facial contour adjustments. Comparative analysis against nine baseline methods, including makeup removal techniques, such as CSD-MIT and BeautyGAN, as well as general image restoration approaches, such as Restormer and SR3, demonstrates that RT-FRR consistently delivers high-fidelity reconstructions, particularly in cases of significant retouching where prior methods frequently fail. On the RetouchingFFHQ dataset, which features mild retouching, RT-FRR maintains strong performance with a PSNR of 25.39 dB, demonstrating robustness across different retouching intensities. Qualitative assessments reinforce these findings, with feature similarity heatmaps and t-SNE visualizations confirming that RT-FRR more accurately recovers facial identity characteristics than competing approaches. The model successfully reconstructs fine-grained skin textures while correcting exaggerated structural changes, such as unnaturally thinned jawlines or enlarged eyes, without introducing artifacts common in GAN-based or transformer-based methods. Ablation studies underscore the importance of key design choices, where the face residual generation strategy alone contributes a 4 dB PSNR improvement, while the FREM module

adds an additional 2 dB by localizing retouching traces. Visual comparisons further demonstrate RT-FRR's ability to restore natural facial proportions and skin details, even in cases where retouching has heavily distorted the original appearance. These results establish RT-FRR as a comprehensive solution for face retouching restoration, effectively addressing textural and structural recovery with unprecedented accuracy. **Conclusion** In this study, we focus on FRR, a novel task focused on recovering original facial features from retouched ones, particularly those that are significantly retouched. We propose RT-FRR to recover the textural and structural face features from a retouched face, where a face retouching estimation module is designed and incorporated to extract the retouching trace to guide the FRR. The retouching traces are integrated with the denoising U-Net of the DDPM through an attention mechanism for an accurate face residual generation. We evaluate the performance of RT-FRR on an existing face retouching dataset and the newly constructed ExRF. On both datasets, our RF-FRR outperforms existing makeup removal and image restoration schemes in the task of FRR.

Key words: image restoration; face retouching restoration(FRR); diffusion model; attention mechanism; residual prediction

0 引言

近年来,人脸美颜技术在各大社交平台中得到广泛的应用,经过瘦脸、大眼、磨皮等不同美颜操作的美颜图像在社交网络中随处可见。人脸美颜技术在丰富公众生活娱乐性和趣味性的同时,也带来了一系列安全和隐私问题。一方面,人脸美颜技术的过度滥用,可以轻易地实现“改头换面”的效果,相较于技术门槛更高的换脸技术,美颜功能被广泛地集成到了抖音、快手、微信甚至腾讯会议等各类平台,方便快捷且易于实时互动,进而更容易被不法分子用于实现网络诈骗(Bharati等,2016)等恶意目的。不法分子可以通过构建虚假形象的方式吸引观众打赏或者进行网恋诈骗等违法违规活动,侵害了用户的知情权。此外,在医美行业,也存在利用美颜技术编造美容效果进行虚假宣传的现象,严重侵害了消费者的权益。另一方面,美颜技术的滥用可能导致社会公众的审美标准扭曲,造成体象障碍、审美降级等心理疾病和社会问题(Ateq等,2024)。面对美颜技术可能导致的问题,多国已采取立法、技术审查等针对性监管措施。挪威于2021年率先立法(Stortinget,2021),明确要求广告商和社交媒体网红必须披露宣传图像是否经过美颜处理,法国和英国也正在推进类似法规的制定。美颜技术带来的影响已不再局限于图像修饰层面,而是逐渐演变为一个涉及心理健康、信息真实性维护和公共安全保障的复合型社会问题。

以往一些研究工作(Rathgeb等,2020;Jain等,2020;Ying等,2023)聚焦于美颜人脸检测问题,即判断一幅人脸图像是否经过美颜操作,已经取得了不

错的效果。但如何从美颜处理后的图像中还原其原始人脸,这一兼具挑战与实用价值的关键问题仍未得到有效解决。然而,在数字身份认证和司法取证等领域中,这一问题至关重要。特别是针对重度美颜后的图像,其图像内容经过较大程度的篡改,从而难以追溯原始身份特征,潜在的危害程度更高。例如,在网络诈骗案件中,受害者仅能提供诈骗分子的重度美颜图像,仅有检测技术无法辅助警方确认嫌疑人真实身份。因此,单纯的美颜检测已无法满足实际场景对追溯真实图像的需求,亟需一种能从美颜图像中精准复原原始人脸的技术,填补从检测伪造到还原真相的技术空白。

针对此问题,本文系统地提出了美颜人脸复原(face retouching restoration,FRR)这一创新性研究任务,其核心目标是从经过美颜处理后的人脸图像中重建出原始人脸图像。当前,计算机视觉研究领域主要有两类相关任务,并已有研究人员开展了深入探索。第一类是妆容去除任务(Wang和Fu,2016),通常作为妆容迁移的子任务,其核心原理是利用风格迁移技术实现原始面部特征的还原。另一类相关工作是图像复原任务(Zamir等,2022;韦炎炎等,2025),该任务侧重于从低质量图像中恢复出高保真的纹理细节,以解决图像可能由于拍摄设备、环境等因素的影响,导致图像模糊、失真等质量下降问题。虽然面对特定任务,研究人员已提出一些有效的解决方法,能够高质量地复原图像的纹理细节等信息。但由于这些任务与美颜复原存在显著的差异,现有的妆容去除或者图像复原方法在处理美颜复原任务时存在显著局限性。如图1所示,美颜操作往往会对人脸图像进行全方面修改。美颜后的人脸图像不仅在纹理肤色层面有精细的调整,还会在

几何层面存在显著的面容形变,例如瘦脸、大眼和隆鼻等操作。这种复合型的图像修改特性要求复原方法同时满足纹理重建和结构校正的双重需求,现有妆容去除和图像复原方法往往难以应对。

针对美颜人脸图像中丢失的纹理细节与结构特征,本研究提出一种基于美颜痕迹引导的扩散模型的美颜人脸复原方法(retouching-guided diffusion model for face retouching restoration, RT-FRR)。RT-FRR利用Ho等人(2020)提出的去噪扩散概率模型(denoising diffusion probabilistic models, DDPM)作为基础框架,通过预测美颜图像与原始人脸之间的残差(简称人脸残差)引导美颜人脸图像的复原和重建,充分利用扩散模型在图像生成领域的优势以生成高质量图像。为了有效地复原人脸结构特征,本文设计了一个多尺度篡改痕迹提取模块(multi-scale face retouching trace extraction module, FREM),通过提取多尺度的人脸美颜篡改特征,并利用注意力机制将该特征作为条件信息引导去噪扩散模型,使得扩散模型更聚焦于篡改区域的恢复。最终,将预测的人脸残差与输入的美颜图像进行融合,实现原始人脸的高保真复原。为了有效地评估和训练美颜人脸复原模型,本文构建了 ExRF (extremely retouched face dataset)数据集。该数据集包含 1 000 对经过极端美颜处理的原始和美颜人脸图像,采用商业图像处理软件进行美颜操作。与现有的 RetouchingFFHQ 数据集不同,ExRF 数据集更侧重于重度美颜图像的处理,以应对极端美颜条件下的复原任务。大量实验证明,RT-FRR 模型在美颜去除任务中具有显著优势。

综上所述,本文主要贡献包括:1)首次考虑了美颜人脸复原这一新的计算机视觉任务,旨在从经过美颜处理的人脸图像中复原出原始人脸,同时构建

了重度美颜人脸数据集 ExRF,为算法训练与评估提供了可靠的基准数据支撑。2)提出一种基于扩散模型的美颜人脸复原方法 RT-FRR。该方法通过定位美颜人脸处理过程中产生的篡改痕迹,引导扩散模型进行人脸残差预测,进而实现对原始人脸的高保真重建。

1 相关工作

当前学术界在美颜人脸复原领域尚未开展针对性研究。鉴于美颜人脸复原问题的复杂性,本文对两个密切相关的研究方向——人脸妆容去除与图像复原技术进行系统性梳理,以探寻解决美颜人脸复原问题的潜在思路与方法。

1.1 妆容去除

人脸妆容去除旨在逆向去除未知化妆过程,恢复原始面部特征(Wang 和 Fu, 2016)。过去十年间,妆容去除相关研究呈现快速发展态势,其技术路线常与妆容迁移任务进行协同探索(Gu 等, 2019; Sun 等, 2024a)。在技术发展脉络上,Chen 等人(2017)提出的 Makeup-Go 构建了基于主成分分解的成分回归网络来学习妆容差异映射,开创了该领域监督学习的先河。在之后的工作中,BeautyGAN(Li 等, 2018)创新性地引入直方图匹配损失,并设计了一种双流生成对抗网络(generative adversarial network, GAN)架构,实现了妆容迁移与去除任务的联合优化,其注意力机制的引入进一步增强了生成图像的跨身份一致性。PairedCycleGAN(Chang 等, 2018)通过添加风格判别器,有效提升了局部妆容特征的相似性度量精度,在高分辨率图像处理任务中表现出色。近年来,深度表征学习方法取得显著进展。DRL(disentangled representation learning)(Li 等, 2020)首次实现身份信息与妆容风格的解耦表征,其基于对抗训练的动态权重分配策略提高了模型的鲁棒性;PSGAN++(pose and expression robust spatial-aware GAN++)(Liu 等, 2022)通过构建妆容蒸馏网络与身份提取网络,显著提升了跨身份妆容迁移效果;SSAT(symmetric semantic-aware Transformer network)(Sun 等, 2022)提出了基于对称语义感知的 Transformer 网络,通过特征融合模块和语义对齐机制实现五官位置精准匹配。CSD-MT(content-style decoupling for unsupervised makeup transfer)(Sun 等,

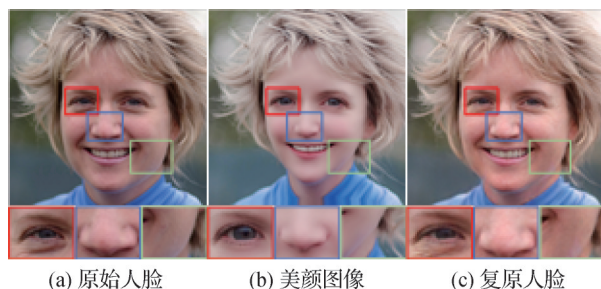


图1 人脸美颜复原

Fig. 1 Face retouching restoration ((a) original images; (b) retouched images; (c) restored images)

2024b)采用基于频域分解的无监督框架,进一步提升了模型对身份信息与妆容风格的解耦能力。

需特别指出的是,此类方法主要针对妆容迁移任务所设计,其中的妆容去除技术在应对美颜人脸复原任务时存在明显局限。美颜人脸复原与妆容去除存在本质差异:前者涉及五官比例调整等面部几何结构的改变,而后者仅处理面部纹理色彩变化。这一根本差异使得现有妆容去除方法难以直接适用于美颜复原任务。

1.2 图像复原

图像复原旨在从退化的输入图像中精确地恢复高质量的图像。早期基于卷积神经网络(convolutional neural network, CNN)的方法如BCCNN(bi-channel convolutional neural network)(Zhou等, 2015)、SRCNN(super-resolution convolutional neural network)(Dong等, 2016)和ARCNN(artifact reduction convolutional neural network)(Dong等, 2015)等通过结合原始输入与人脸特征表示,在特定场景的图像质量提升方面取得了重要突破。随着深度学习技术的发展,Transformer模型因其出色的全局建模能力(Liang等, 2021)在图像复原中得到广泛应用,其中Restormer(Zamir等, 2022)提出的高效Transformer架构在多项图像复原任务中展现出卓越性能。近年来,扩散模型凭借其精确的目标分布建模能力(Kawar等, 2022; Lin等, 2024; Wang等, 2024)在图像复原领域崭露头角。SR3(super-resolution via repeated refinements)(Saharia等, 2023)开创性地将扩散模型应用于图像超分辨率任务,DR2(diffusion-based robust degradation remover)(Wang等, 2023)通过引入生成先验的扩散过程显著提升了纹理细节的恢复质量。华夏等人(2024)提出了结合局部和非局部特征的网络方法,通过划分小图像块并采样最大值点进行自注意力运算,有效提取远距离像素依赖关系,从而提升了单图像去模糊任务的性能。OSE-Diff(one-step effective diffusion network)通过LoRA(low-rank adaptation)微调的扩散网络和变分分数蒸馏正则化,利用低质量图像单步生成高质量图像。ResDiff(residual diffusion model)(Shang等, 2024)提出的频域引导扩散策略通过引入多尺度频域信息融合机制,进一步优化了针对图像高频信息的复原效果。此外,经典的基于生成对抗网络的方法Pix2pix(image-to-image translation with conditional adver-

sarial network)(Isola等, 2017)由于其通用性,也在人脸图像复原任务中具有较强鲁棒性。肖杰等人(2025)提出了非因果选择性状态空间模型,建模图像块之间的非因果依赖关系,从而实现图像复原。

然而,值得注意的是,传统图像复原方法主要针对模糊、噪声等图像退化问题,着重于提升视觉质量指标,特别是高频纹理细节,而美颜人脸复原则面临更为复杂的挑战。美颜处理不仅会造成高频细节损失,更会引入语义层面的几何形变,导致人脸面部结构的变化,这与现有的图像复原方法针对的任务特点具有显著差异,使得现有方法在保持身份特征和校正面部结构方面表现欠佳。特别是在处理极端美颜案例时,现有的方法往往难以同时满足纹理真实性和结构准确性的双重要求。

2 数据集构建

现有研究已针对美颜人脸检测任务构建了相关数据集,例如Ying等人(2023)提出的RetouchingFFHQ构建了涵盖不同美颜程度与操作类型的大规模细粒度美颜数据集。然而该数据集中的图像大多经过精细的轻度修饰,视觉外观未产生显著变化,为考虑实际场景中可能存在极端美颜情况。本文中的美颜人脸复原任务研究更关注原始面部特征严重丢失的极端美颜案例。因此,本文首先构建了一个包含极端美颜图像及其原始版本的重度美颜人脸数据集ExRF。

本文从高质量人脸数据集FFHQ(Karras等, 2019)中筛选了1 000幅高质量人脸图像。这些图像都符合严格的质量标准,例如,图像中必须展示清晰无遮挡的正面人脸,不能有夸张表情或明显的面部妆容,同时背景需要保持简洁且不存在其他人脸。在筛选过程中充分考虑年龄、性别和种族等人口统计学特征,保证数据集分布更符合实际应用场景。随后使用商业图像处理软件WinkStudio对这些图像进行人工美颜处理,生成极端的美颜人脸图像。采用了包括上妆、大眼、瘦脸以及磨皮等多种美颜操作的混合处理,以模拟社交平台上常见的多重美颜效果,且采用较高的美颜程度,在保证美颜处理后人脸图像合理性的同时极大程度改变面部特征。为了得到更符合审美评价的极端美颜样本,聘请了10位志愿者进行美颜图像处理,在保证美颜程度较高的前

提下合理调整参数组合,并通过交叉评估的方法保留符合标准的图像。最终构建的 ExRF 数据集包含了 1 000 组经过极端美颜处理的人脸和原始人脸图像对。

图 2 展示了 ExRF 数据集中美颜人脸与原始人脸图像的对比示例。通过计算美颜图像和原始图像之间的差值并进行二值化获取对应的人脸残差图像直观展示两幅图像的差异。比较美颜前后的人脸图像以及残差图像的白色区域,可以看出极端美颜情况下的图像与原始版本存在显著的差异。特别是在面部结构上,美颜操作显著地调整了结构比例,使得在眼睛大小和下巴轮廓上有显著改变,这也符合社交网络应用中极端美颜处理操作的特点。

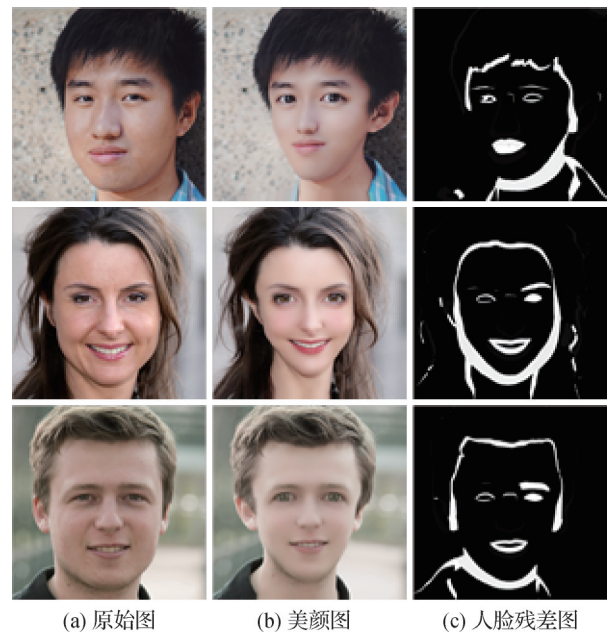


图 2 ExRF 数据集图像示例
Fig. 2 Examples of the ExRF dataset ((a) original images;
(b) retouched images; (c) face parsing residual of the
retouched and original faces)

为进一步度量二者的显著差异,本文利用 Kazemi 和 Sullivan(2014)提出的 dlib 人脸识别模型提取出全部 1 000 组图像的人脸面部特征并进行实验分析。结构表明美颜图像在特征空间与原始图像也存在显著偏移。ExRF 数据集在特征分布上的特点如图 3 所示。图 3(a)展示了人脸相似度热力图,通过对不同人脸面部特征的余弦相似度进行统计得到。可以从热力图观察到左上角热力图颜色更深,表明美颜人脸之间存在显著相似性。同时美颜和原

始图像配对的区域颜色最浅,说明美颜操作显著改变了人脸特征。图 3(b)展示了美颜人脸特征的 t-SNE(t-distributed stochastic neighbor embedding)(Van der Maaten 和 Hinton, 2008)分布图,通过将美颜人脸特征映射到二维空间进行可视化,可以看出美颜人脸和原始人脸存在显著的特征分布差异。极端美颜处理使得美颜人脸在视觉上与原始人脸形成鲜明对比,给人造成强烈的视觉冲击,从而产生欺骗性。同时往往还会影响人脸身份特征,为人脸相关场景带来安全隐患。

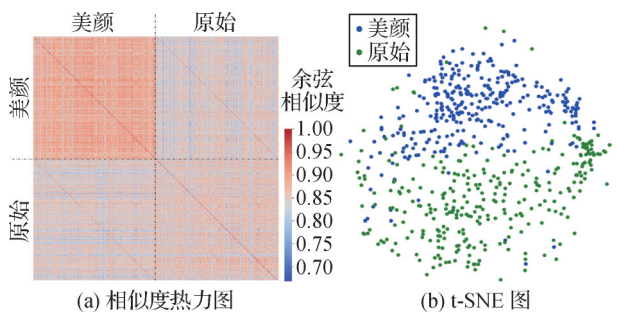


图 3 ExRF 数据集特征分布图
Fig. 3 Feature distribution of the ExRF dataset
((a) similarity heatmap; (b) t-SNE map)

3 方 法

为了精确地从美颜人脸中复原出原始人脸,本文提出了基于美颜痕迹引导的扩散模型的美颜人脸复原方法(RT-FRR)。图 4 展示了 RT-FRR 的整体架构。扩散模型由于其逐步去噪的生成机制具备高质量样本生成能力,在图像生成领域取得了卓越的效果。相比于传统的卷积神经网络、生成对抗网络和 Transformer,扩散模型能够更精确地恢复图像的高频细节,尤其是在处理复杂的图像生成任务时,其能够精确建模图像生成的目标分布,提供更加稳定的训练过程,避免了 GAN 常见的训练不稳定问题。此外,扩散模型可以在多个时间步内从图像结构到细节进行逐步优化,从而能够生成高质量精细化的图像。同时扩散模型也被证明在图像复原任务中能够有效重建高频细节和纹理特征(Wang 等, 2023)。本文采用去噪扩散概率模型(Ho 等, 2020)作为基础架构,针对性生成美颜图像与原始图像之间的人脸残差,以引导从美颜图像生成美颜前原始图像的预测。为了有效地恢复原始人脸的面部结构特征,还创新

性地设计了一个多尺度篡改痕迹提取模块 (FREM), 该模块通过多尺度特征提取分析美颜处理产生的篡改痕迹, 并将其输出结果作为条件信息引导扩散模型生成更精确的人脸残差。最终将美颜图像与生成的人脸残差相结合, 复原重建出原始人脸图像。

3.1 条件扩散模型

条件扩散模型可以在有条件引导的情况下利用马尔可夫链从高斯噪声逐步去噪, 最终生成预测图像结果。在本文的任务场景下, 给定美颜处理后的图像输入 X , 本文所提的 RT-FRR 模型将利用多尺度美颜篡改痕迹作为条件, 使扩散过程中模型聚焦于美颜操作处理过的区域, 以预测美颜人脸与原始人脸的人脸残差。

具体而言, 给定原始人脸图像, 符号定义为 Y , 人脸残差定义为原始人脸和美颜图像的差值, 即 $r_0 = Y - X$ 。扩散模型的前向扩散过程 $q(\cdot)$ 通过预设

的方差序列 $\{\beta_1, \beta_2, \dots, \beta_T\}$, 逐步地对真实人脸残差 r_0 加噪, 最终在第 T 步加噪成为高斯噪声, 这一过程可以定义为

$$q(r_t | r_{t-1}) = \mathcal{N}(r_t; \sqrt{1 - \beta_t} r_{t-1}, \beta_t I) \quad (1)$$

式中, $\mathcal{N}$ 表示高斯分布, r_t 表示第 t 步加噪后的中间结果。 I 为单位矩阵。

在扩散模型的去噪过程中, 模型在给定条件 c 的情况下, 从随机采样的高斯噪声 $\hat{r}_T \sim \mathcal{N}(0, I)$ 开始逐步生成预测的人脸残差结果, 这一过程可以表达为

$$p_\theta(\hat{r}_{0:T} | X) = p(\hat{r}_T) \prod_{t=1}^T p_\theta(\hat{r}_{t-1} | \hat{r}_t, c) \quad (2)$$

式中, $\hat{r}_t$ 表示第 t 步去噪后得到的中间残差预测。

在所提的 RT-FRR 模型中, 给定条件 c 具体包含两类引导条件, 分别是输入的美颜人脸 X 和利用多尺度篡改痕迹提取模块从美颜人脸 X 中提取出的多尺度人脸美颜篡改痕迹 f 。

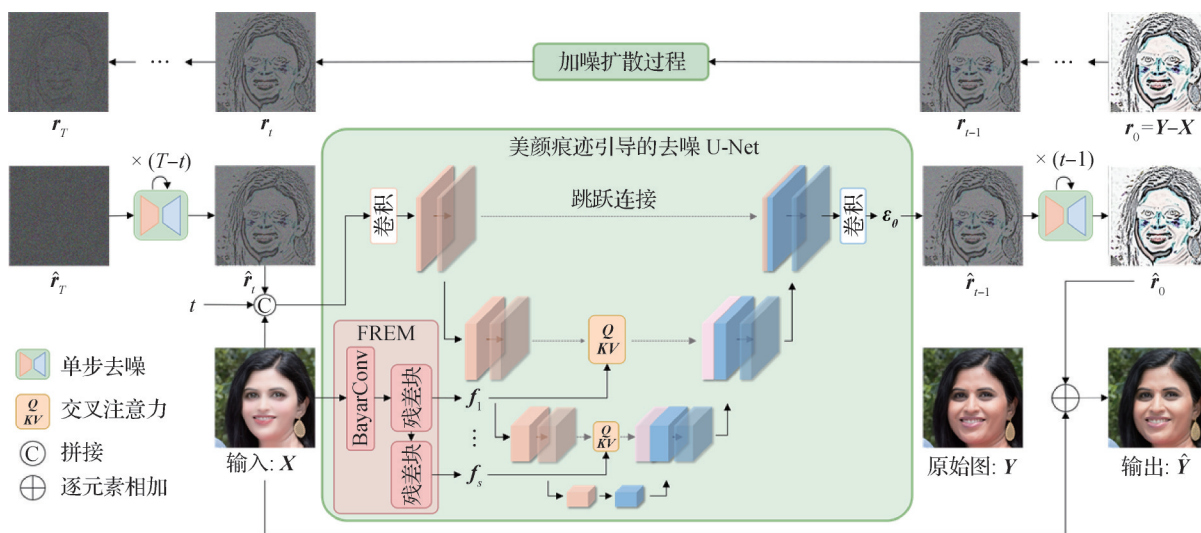


图4 RT-FRR 模型整体架构

Fig. 4 The overall pipeline of proposed RT-FRR

3.2 人脸美颜痕迹提取

本文设计了一个多尺度篡改痕迹提取模块 (FREM), 用于有效提取美颜处理产生的篡改痕迹。FREM 模块旨在提取分析美颜处理产生的篡改痕迹, 并将其输出结果作为条件信息引导扩散模型进行精准的残差预测。BayarConv (Bayar 和 Stamm, 2018) 广泛应用于图像篡改检测任务中。与传统的卷积层不同, 其通过对卷积核进行特殊约束, 能够自适应地抑制图像内容的干扰, 聚焦于图像中细微的

篡改痕迹。该模块采用了 BayarConv 作为卷积层进行篡改痕迹提取, 设置卷积核的大小为 $5 \times 5 \times 3$, 能够自适应地抑制图像内容对篡改痕迹的干扰, 从而更有效地提取篡改特征。BayarConv 需要满足以下两个条件

$$\omega(0, 0) = -1, \sum_{i,j \neq 0} \omega(i, j) = 1 \quad (3)$$

式中, ω 表示卷积核的参数, $\omega(0, 0)$ 表示卷积核中心位置元素。这种特殊设计通过约束卷积核参数, 强制卷积过程学习高频篡改痕迹, 如边缘异常、噪声异

常以及纹理失真等,同时避免图像自身语义内容带来干扰,能够有效鉴别篡改区域。相较于传统卷积方法关注图像的低级特征可能带来的过度平滑或信息丢失问题,BayarConv更能捕捉到篡改痕迹所特有的高频变化,从而更好地定位篡改区域并提供复原引导。

为了适应扩散模型的结构并提供有效的信息引导,将 BayarConv 层的输出特征($f_{\text{BayarConv}}$)送入多个可训练的 ResNet 残差块($f_{\text{ResBlocks}}$)(He 等, 2016)进行进一步处理,以提取多尺度人脸美颜篡改痕迹 f ,可以表示为

$$f = \{f_1, f_2, \dots, f_S\} = f_{\text{ResBlocks}}(f_{\text{BayarConv}}(X)) \quad (4)$$

式中, $f_i(i \in \{1, 2, \dots, S\})$ 表示第 i 个残差块提取到的特征, S 为多尺度篡改痕迹提取模块设定的尺度数量超参。随着 i 的改变,网络提取到的特征 f_i 呈现出明显的特性变化。当 i 较小时,FREM提取出的 f_i 主要捕获局部纹理异常,如磨皮导致的皮肤高频细节丢失;随着 i 的增大, f_i 开始表征区域结构畸变,如大眼操作引起的眼眶几何变形;当 i 达到最大时, f_i 可以编码全局语义不一致特征,如整体面部比例失调。这种层级递进的特征提取机制使FREM能够全面建模从像素级到语义级的美颜痕迹。

3.3 美颜痕迹引导的去噪 U-Net

基于3.2节提取到的多尺度人脸美颜篡改痕迹特征 f ,本研究设计了一个美颜痕迹引导的去噪 U-Net 网络 $\varepsilon_\theta(\cdot)$,通过交叉注意力机制将美颜痕迹信息融入人脸残差的去噪过程。不同于传统 U-Net 采用的跳跃连接,本文方法采用交叉注意力机制将多尺度美颜痕迹映射到 U-Net 的对应中间层特征,具体实现如下

$$\begin{aligned} Q_i &= W_Q^{(i)} \cdot M_i \\ K_i &= W_K^{(i)} \cdot f_i \end{aligned} \quad (5)$$

$$V_i = W_V^{(i)} \cdot f_i$$

$$M'_i = f_{\text{softmax}}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i \quad (6)$$

式中, M_i, M'_i 分别表示 U-Net 第 i 层的输入和输出特征, $W_Q^{(i)}, W_K^{(i)}, W_V^{(i)}$ 为可学习的投影矩阵, d_k 表示查询矩阵 Q 的列维度。值得注意的是,本文刻意避免在 U-Net 的最浅层集成美颜痕迹特征,这种设计选择是为了确保模型优先聚焦于人脸结构特征的恢复,而避免过度关注像素级细节。

本文所提的 RT-FRR 模型的训练目标函数设计如下

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{y_0, f, \varepsilon_t \sim \mathcal{N}(0, I)} \left[\left\| \varepsilon_t - \varepsilon_\theta(r_t, X, f, t) \right\|^2 \right] \quad (7)$$

以下算法 1 和算法 2 中的伪代码分别给出了 RT-FRR 在训练和推理阶段的详细过程。

算法 1: RT-FRR 训练过程。

- 1) $f = \text{FREM}(X)$;
- 2) $r_0 = Y - X$;
- 3) while not converged do;
- 4) $t \sim \text{Uniform}\{1, \dots, T\}$;
- 5) $\varepsilon_t \sim \mathcal{N}(0, I)$;
- 6) $r_t = \sqrt{\alpha_t} r_0 + \sqrt{1 - \alpha_t} \varepsilon_t$;
- 7) $\nabla_\theta \left\| \varepsilon_t - \varepsilon_\theta(r_t, X, f, t) \right\|^2$;
- 8) end while。

算法 2: RT-FRR 推理过程。

- 1) $f = \text{FREM}(X)$;
- 2) $r_T \sim \mathcal{N}(0, I)$;
- 3) for $t=T:1$ do;
- 4) $z \sim \mathcal{N}(0, I)$ if $t > 1$, else $z = 0$;
- 5) $r_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(r_t - \frac{1 - \alpha_t}{\sqrt{1 - \alpha_t}} \varepsilon_\theta(r_t, X, f, t) \right) + \sigma_t z$;
- 6) end for;
- 7) $\hat{Y} = X + r_0$ 。

4 实验

4.1 实验设置

4.1.1 实现细节

本文中将特征尺度层级数 S 设定为4,扩散时间步 T 设置为1 000。训练过程中采用 Adam 优化器(Kingma 和 Ba, 2015)对 RT-FRR 模型进行优化,训练批量大小为16,总训练迭代次数为10万次,初始学习率固定为 10^{-4} 。模型实现采用深度学习 Pytorch 框架,实验在 NVIDIA RTX 3090 显卡上完成。在数据划分方面,ExRF 数据集按照 80% 训练样本、10% 验证样本和 10% 测试样本的标准比例进行划分。为了全面评估模型性能,本文还使用了为美颜检测任务而构建的 RetouchingFFHQ 数据集(Ying 等, 2023)进行泛化性测试,该数据集主要包含经过不同程度

美颜处理的样本,且轻度、中度美颜样本居多,数据来源包括旷视、腾讯和阿里提供的美颜接口。所有实验图像均统一缩放至 512×512 像素分辨率以确保评估一致性。

4. 1. 2 评估指标

为了评价所提方法和对比方法的效果,本文综合选取两类评价指标。第1类是基于像素相似度的峰值信噪比(peak signal to noise ratio, PSNR)和结构相似性指数(structural similarity index measure, SSIM)(Wang等,2004);第2类为感知质量评估指标,包括有参考的学习感知图像块相似度(learned perceptual image patch similarity, LPIPS)(Zhang等,2018)和无参考的评估指标FID(Fréchet inception distance)(Heusel等,2017)。为衡量复原图像在特征层面的保真度,额外采用人脸识别模型(Kazemi和Sullivan,2014)提取的人脸身份特征进行相似度分析。

4. 1. 1 对比方法

为了验证本文方法的有效性,将提出的RT-FRR模型与9种相关方法进行了系统对比。其中包括Makeup-go方法(Chen等,2017)、BeautyGAN方法(Li等,2018)、SSAT方法(Sun等,2022)以及CSD-MT方法(Sun等,2024b)等针对妆容去除任务的基线模型,以及针对图像复原任务的Pix2pix方法(Isola等,2017)、Restormer方法(Zamir等,2022)、SR3方法(Saharia等,2023)、DR2方法(Wang等,2023)、OSEDiff方法(Wu等,2024)以及ResDiff方法(Shang等,2024)。这些基线模型涵盖了基于卷积神经网络、基于生成对抗网络、基于Transformer和基于扩散模型的4类主流架构。为确保对比实验的公平性,所有基线模型均采用原始论文配置进行复现训练。

4. 2 定量分析

如表1所示,在重度美颜的ExRF数据集上,本文所提的RT-FRR在美颜人脸复原任务中展现出卓越的性能优势。从实验数据可以看出,虽然BeautyGAN等方法在妆容去除领域取得了显著效果,但这些方法难以处理面部结构变化,面对更具有挑战性的美颜人脸图像复原任务时性能显著下降。与原始图像相比,这些方案难以获得有效的复原效果。传统图像复原方法如Restormer和SR3,虽然能一定程度上恢复图像结构特征,但在面对原始面部特征严重丢失的情况下表现欠佳。相比之下,本文所提

方法通过多尺度篡改痕迹提取模块有效捕捉不同尺度的人脸美颜篡改痕迹,能够为扩散模型提供精确的复原引导,生成高质量的复原人脸图像。RT-FRR生成的复原人脸以25.23 dB的PSNR和0.806的SSIM显著优于其他基线模型,相比表现次优的ResDiff,PSNR提升了3.09 dB,SSIM提升了0.03。在感知质量评估度量方面,所提的RT-FRR生成的复原图像的FID和LPIPS也具有显著优势,分别为48.802和0.241,这表明RT-FRR生成的复原人脸与真实人脸分布更接近。

表1 ExRF数据集的定量对比
Table 1 Quantitative comparison on the ExRF dataset

方法	PSNR/dB ↑	SSIM ↑	FID ↓	LPIPS ↓
输入	17.76	0.578	87.589	0.286
Makeup-go	13.22	0.442	82.735	0.399
BeautyGAN	16.50	0.563	87.869	0.391
SSAT	16.44	0.575	101.079	0.381
CSD-MT	18.38	0.598	77.852	0.326
Pix2pix	23.33	0.667	73.672	0.323
Restormer	20.93	0.681	59.661	0.362
SR3	18.14	0.718	51.589	0.329
DR2	17.70	0.560	73.028	0.373
OSEDiff	22.03	0.748	51.294	0.289
ResDiff	22.14	0.776	50.447	0.281
RT-FRR(本文)	25.23	0.806	48.802	0.241

注:加粗字体表示各列最优结果。“↑”代表值越高越好,“↓”代表值越低越好。

表2展示了在Ying等人(2023)提出的用于细粒度美颜检测的RetouchingFFHQ数据集上的实验效果,以进一步验证本文方法的泛化性。从表2可以看出,RetouchingFFHQ中的人脸图像所经过的美颜操作处理较为精细,主要涵盖轻度、中度的美颜处理,虽然视觉变化不明显,但仍会轻微地改变面部特征分布。从实验比较可以看出,现有方法对这种精细修改的敏感性明显不足,最佳基线模型Restormer的PSNR仅为24.59 dB。而所提的RT-FRR通过多尺度特征分析,能够有效捕捉到这些细微变化,以25.39 dB的PSNR和0.885的SSIM显著优于其他基线模型。同时,RT-FRR的FID为27.691,较次优模型提升显著,再次证明了所提方法的复原人脸更接

近于真实分布。

对比表1和表2的实验结果还可以观察到,传统方法如 Restormer 虽然在轻度美颜场景中表现尚可 (PSNR 24.59 dB),但在重度美颜场景下性能骤降 (PSNR 20.93 dB),而所提的 RT-FRR 在两个不同场景下都能取得良好的复原效果,验证了所提方法具有良好的泛化能力。此外,这些结果充分说明,通过条件扩散机制与多尺度特征提取的有机结合,RT-FRR 能够自适应地处理不同强度的美颜操作,为美颜人脸复原提供了可靠解决方案。

表2 RetouchingFFHQ数据集的定量对比
Table 2 Quantitative comparison on the RetouchingFFHQ dataset

方法	PSNR/dB ↑	SSIM ↑	FID ↓	LPIPS ↓
输入	22.21	0.776	39.637	0.181
Makeup-go	14.27	0.531	46.443	0.314
pix2pix	24.61	0.851	36.614	0.189
Restormer	24.59	0.848	28.435	0.196
SR3	23.78	0.858	35.746	0.208
DR2	21.72	0.665	39.568	0.293
OSEDiff	24.37	0.835	34.941	0.184
ResDiff	24.43	0.863	32.179	0.180
RT-FRR(本文)	25.39	0.885	27.691	0.169

注:加粗字体表示各列最优结果。“↑”代表值越高越好,“↓”代表值越低越好。

本文还对 ExRF 数据集中美颜图像、复原图像与原始图像的特征相似度进行可视化。热力图结果如图5所示。其中热力图分别通过计算原始图像与美颜图像(图5(a))、原始图像与复原图像(图5(b)—图5(e))或两组原始图像(图5(f))之间的特征余弦相似度生成。当两幅人脸图像特征相似时,热力图对应区域颜色接近深红色,每幅热力图对角线表示原始图像与其美颜版本或复原版本的对应关系。可以看出,图5(e)中对角线颜色比图5(b)—图5(d)更接近深红色,表明本文方法复原的人脸图在特征层面也能保持与原始图像更高的相似情况。

4.3 定性分析

图6展示了不同方法在 ExRF 数据集上进行复原的视觉效果对比。从图6可以看出,以 CSD-MT (Sun 等, 2024b) 为代表的妆容去除方法虽然恢复了

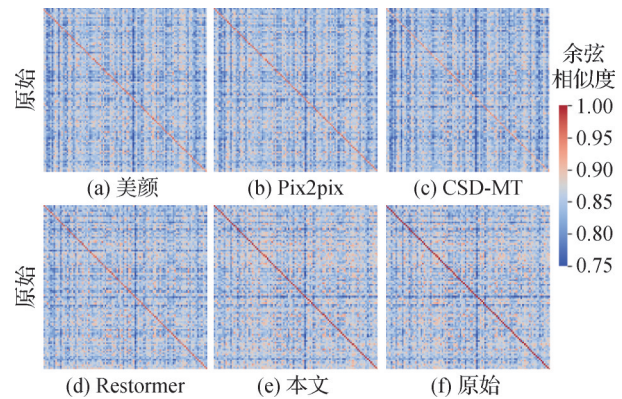


图5 特征相似度热力图

Fig. 5 Feature similarity heatmaps ((a) retouched; (b) Pix2pix; (c) CSD-MT; (d) Restormer; (e) ours; (f) original)

面部纹理风格,却忽略了结构特征的还原。Pix2pix (Isola 等, 2017) 和 Restormer (Zamir 等, 2022) 等方法生成的面部图像存在明显伪影,难以获得高质量结果。ResDiff (Shang 等, 2024) 的复原结果则出现显著色彩失真问题。相比之下,本文提出的 RT-FRR 方法在消除美颜操作的同时,既能减少伪影又能保持更好的视觉质量,其复原人脸与原始人脸最为接近。

4.4 消融实验

本节通过消融实验研究了不同尺度级别 S 的设置以及 RT-FRR 主要组件的影响。

表3展示了设置不同的特征尺度层级数 S 时的模型性能表现。实验结果表明,增加尺度级别数量通常能够提升模型性能,这为美颜感知的去噪 U-Net 提供了更丰富的引导信息。经过综合评估,本文在实验中采用 $S = 4$ 的配置方案,该设置取得了最优的图像复原效果。

为验证人脸残差生成策略与 FREM 模块的有效性,进行了3组对比实验。前两组实验分别在包含和不包含 FREM 模块的情况下采用 DDPM 直接生成复原人脸,第3组实验则在移除 FREM 的情况下生成人脸残差。表4展示了不同设置下的 FRR 性能指标。实验数据表明,采用残差生成策略时,复原图像的 PSNR 比直接生成方式提升超过 4 dB;当引入 FREM 模块后,PSNR 可进一步增加约 2 dB。这些结果充分证实了人脸残差生成策略与 FREM 模块在 RT-FRR 中的关键作用,能够显著提升美颜复原性能。

FREM 模块提取的可视化多尺度特征如图7所

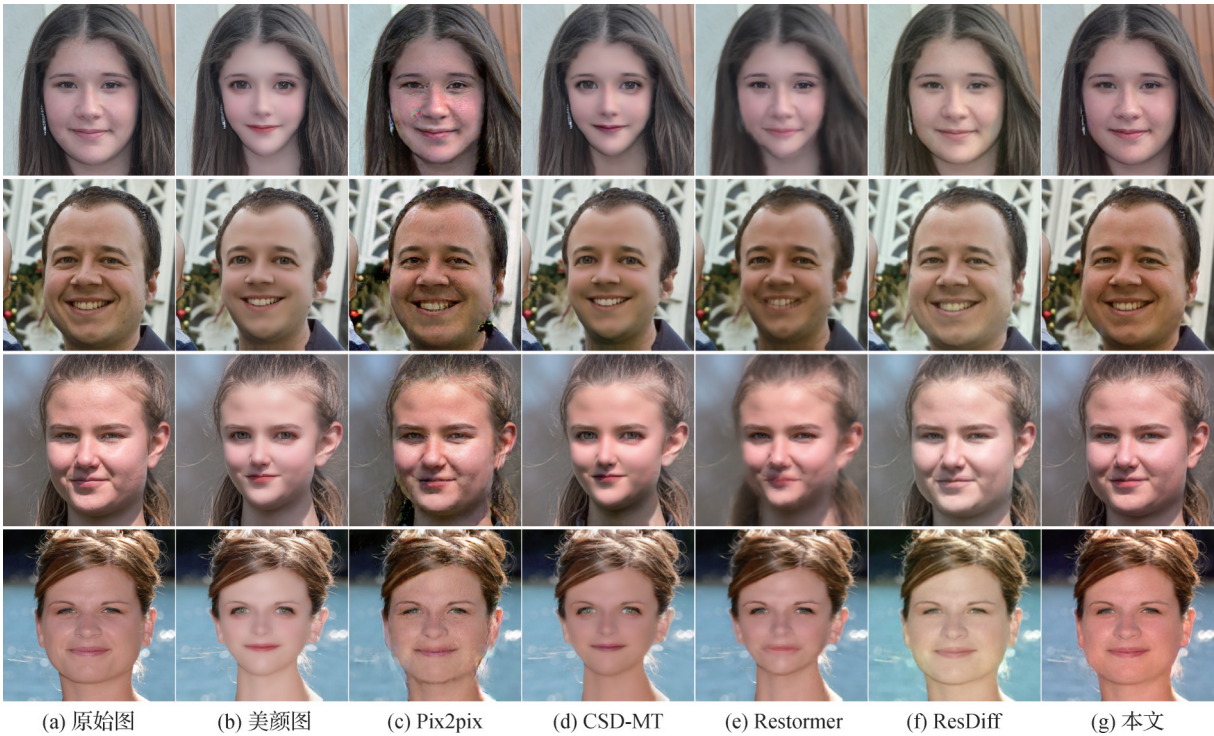


图6 不同方法复原人脸图像结果的定性对比

Fig. 6 Qualitative comparison of the recovered face images using different methods

((a) original images; (b) retouched images; (c) Pix2pix; (d) CSD-MT; (e) Restormer; (f) ResDiff; (g) ours)

表3 超参S的消融实验

Table 3 Ablation study for the setting of S

层级数S	PSNR/dB ↑	SSIM ↑	FID ↓	LPIPS ↓
S = 1	24.41	0.799	52.108	0.248
S = 2	25.11	0.804	50.074	0.245
S = 3	25.13	0.806	49.946	0.244
S = 4	25.23	0.806	48.802	0.241

注:加粗字体表示各列最优结果。“↑”代表值越高越好,“↓”代表值越低越好。

表4 人脸残差生成策略和FREM模块的消融实验

Table 4 Ablation study for the face residual generation strategy and the FREM

残差	FREM	PSNR/dB ↑	SSIM ↑	FID ↓	LPIPS ↓
-	-	18.14	0.718	51.589	0.329
-	√	19.93	0.756	51.043	0.274
√	-	23.35	0.794	49.464	0.255
√	√	25.23	0.806	48.802	0.241

注:加粗字体表示各列最优结果。“↑”代表值越高越好,“↓”代表值越低越好。“√”表示使用对应模块,“-”表示未使用。

示,其中颜色越红的区域表示模型对该区域的关注度越高。可以看到,FREM模块中有效的引导模型

聚焦于如下颌线和眼部等关键面部区域。同时,FREM提取的多尺度特征热力图呈现出层次化的注意力分布模式,即浅层网络捕捉细粒度的像素级细节,而深层网络通过多尺度感受野可以逐步整合全局上下文信息。

消融实验的定性对比结果如图8所示。可以看到,FREM模块通过精确定位美颜痕迹,能够有效确保复原人脸的结构一致性,而使用人脸残差生成策

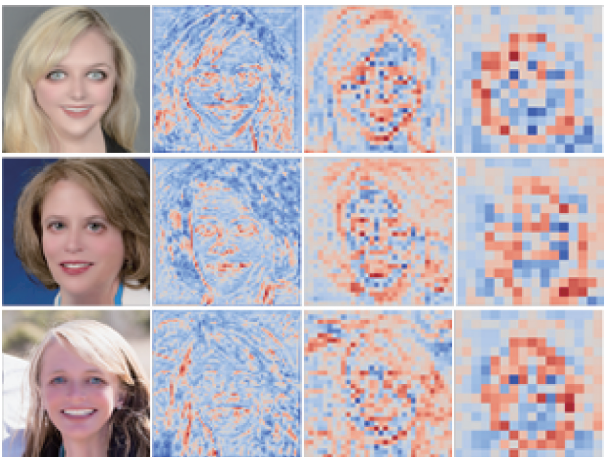


图7 FREM提取出的多尺度美颜篡改特征

Fig. 7 Multi-scale retouching features extracted by FREM
((a) original images; (b) feature 1; (c) feature 2; (d) feature 3)

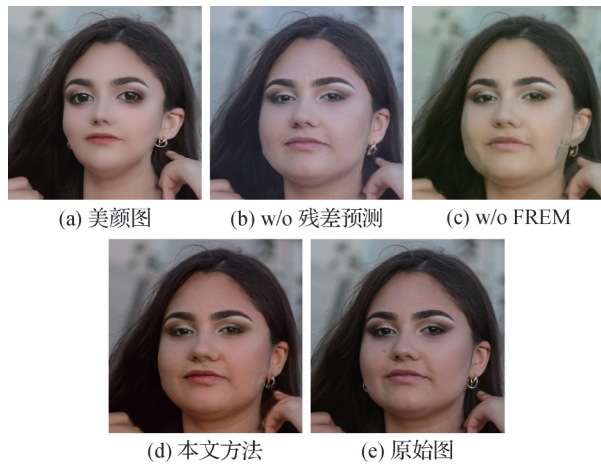


图8 消融实验的定性对比

Fig. 8 Qualitative comparison of the ablation study
((a) retouched image; (b) w/o residual prediction;
(c) w/o FREM; (d) ours; (e) original image)

略能有效保持色彩保真度。

5 结 论

本文针对美颜人脸复原(FRR)这一新颖的计算机视觉任务展开系统性研究,重点解决重度美颜人脸图像中原始特征恢复的关键技术难题,提出了一种基于美颜痕迹引导的扩散模型:RT-FRR。该方法首先设计了一个美颜痕迹估计模块(FREM),通过约束性卷积层和多尺度特征提取网络有效捕捉不同尺度的美颜篡改痕迹。在此基础上,采用交叉注意力作为特征融合机制,利用提取的美颜痕迹特征引导扩散模型的去噪 U-Net 进行人脸残差的预测,最终得到与原始人脸高度相似的美颜复原人脸,实现了面部纹理细节和结构特征的协同恢复。为了全面评估模型性能,构建了一个重度美颜数据集 ExRF,并在 ExRF 和现有公开数据集 RetouchingFFHQ 上分别进行了实验验证。实验结果表明,RT-FRR 在多个评价指标上均显著优于现有的方法,特别是在处理重度美颜图像时展现出明显的性能优势。此外,可视化评估也证实了该方法在保持面部自然度和身份特征方面的优越性。

尽管 RT-FRR 在美颜人脸复原任务中表现突出,但仍存在一定局限性。模型主要在重度美颜数据集上训练,对轻度和中度美颜的适应性有待提升;另外,扩散模型推理耗时较长,不利于实时应用。未来工作可从3个方面展开:1)构建多层次美颜与多样

人群的数据集以增强泛化性;2)研究轻量化与加速采样的扩散模型以提高效率,便于移动端设备应用;3)扩展至视频美颜复原及三维一致性建模,为动态场景下的真实还原提供支持。

参考文献(References)

- Ateq K, Alhajji M and Alhusseini N. 2024. The association between use of social media and the development of body dysmorphic disorder and attitudes toward cosmetic surgeries: a national survey. *Frontiers in Public Health*, 12: #1324092 [DOI: 10.3389/fpubh.2024.1324092]
- Bayar B and Stamm M C. 2018. Constrained convolutional neural networks: a new approach towards general purpose image manipulation detection. *IEEE Transactions on Information Forensics and Security*, 13(11): 2691-2706 [DOI: 10.1109/TIFS.2018.2825953]
- Bharati A, Singh R, Vatsa M and Bowyer K W. 2016. Detecting facial retouching using supervised deep learning. *IEEE Transactions on Information Forensics and Security*, 11(9): 1903-1913 [DOI: 10.1109/TIFS.2016.2561898]
- Chang H W, Lu J W, Yu F and Finkelstein A. 2018. PairedCycleGAN: asymmetric style transfer for applying and removing makeup//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 40-48 [DOI: 10.1109/CVPR.2018.00012]
- Chen Y C, Shen X Y and Jia J Y. 2017. Makeup-go: blind reversion of portrait edit//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 4511-4519 [DOI: 10.1109/ICCV.2017.482]
- Dong C, Deng Y B, Loy C C and Tang X O. 2015. Compression artifacts reduction by a deep convolutional network//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago, Chile: IEEE: 576-584 [DOI: 10.1109/ICCV.2015.73]
- Dong C, Loy C C, He K M and Tang X O. 2016. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2): 295-307 [DOI: 10.1109/TPAMI.2015.2439281]
- Stortinget (Norwegian Parliament). 2021. Lovvedtak 146 (2020-2021) - Endring i markedsføringskontrollloven [EB/OL]. [2025-06-11]. <https://www.stortinget.no/no/Saker-og-publikasjoner/Vedtak/Beslutninger/Lovvedtak/2020-2021/vedtak-202021-146/>.
- Gu Q, Wang G Z, Chiu M T, Tai Y W and Tang C K. 2019. LADN: local adversarial disentangling network for facial makeup and demakeup//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE: 10480-10489 [DOI: 10.1109/ICCV.2019.01058]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on*

- Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Heusel M, Ramsauer H, Unterthiner T, Nessler B and Hochreiter S. 2017. GANs trained by a two time-scale update rule converge to a local Nash equilibrium//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 6629-6640 [DOI: 10.5555/3295222.3295408]
- Ho J, Jain A and Abbeel P. 2020. Denoising diffusion probabilistic models//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #574 [DOI: 10.5555/3495724.3496298]
- Hua X, Shu T, Li M X, Shi Y and Hong H Y. 2024. Non-local feature representation embedded blurred image restoration. *Journal of Image and Graphics*, 29(10): 3033-3046 (华夏, 舒婷, 李明欣, 时愈, 洪汉玉. 2024. 融合非局部特征表示的模糊图像复原. *中国图象图形学报*, 29(10): 3033-3046) [DOI: 10.11834/jig.230735]
- Isola P, Zhu J Y, Zhou T H and Efros A A. 2017. Image-to-image translation with conditional adversarial networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5967-5976 [DOI: 10.1109/CVPR.2017.632]
- Jain A, Majumdar P, Singh R and Vatsa M. 2020. Detecting GANs and retouching based digital alterations via DAD-HCNN//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, USA: IEEE: 2870-2879 [DOI: 10.1109/CVPRW50498.2020.00344]
- Karras T, Laine S and Aila T. 2019. A style-based generator architecture for generative adversarial networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 4396-4405 [DOI: 10.1109/CVPR.2019.00453]
- Kawar B, Elad M, Ermon S and Song J M. 2022. Denoising diffusion restoration models//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: #1714 [DOI: 10.5555/3600270.3601984]
- Kazemi V and Sullivan J. 2014. One millisecond face alignment with an ensemble of regression trees//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE: 1867-1874 [DOI: 10.1109/CVPR.2014.241]
- Kingma D P and Ba J L. 2015. Adam: a method for stochastic optimization//Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA: ICLR: 1-13 [DOI: 10.48550/arXiv.1412.6980]
- Li T T, Qian R H, Dong C, Liu S, Yan Q, Zhu W W, et al. 2018. BeautyGAN: instance-level facial makeup transfer with deep generative adversarial network//Proceedings of the 26th ACM International Conference on Multimedia. Seoul, Korea (South): ACM: 645-653 [DOI: 10.1145/3240508.3240618]
- Li Y, Huang H B, Cao J, He R and Tan T N. 2020. Disentangled representation learning of makeup portraits in the wild. *International Journal of Computer Vision*, 128(8): 2166-2184 [DOI: 10.1007/s11263-019-01267-0]
- Liang J Y, Cao J Z, Sun G L, Zhang K, Van Gool L and Timofte R. 2021. SwinIR: image restoration using swin transformer//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal, Canada: IEEE: 1833-1844 [DOI: 10.1109/ICCVW54120.2021.00210]
- Lin X Q, He J W, Chen Z Y, Lyu Z Y, Dai B, Yu F H, et al. 2024. DiffBiR: toward blind image restoration with generative diffusion prior//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 430-448 [DOI: 10.1007/978-3-031-73202-7_25]
- Liu S, Jiang W T, Gao C, He R, Feng J S, Li B, et al. 2022. PSGAN++: robust detail-preserving makeup transfer and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11): 8538-8551 [DOI: 10.1109/TPAMI.2021.3083484]
- Rathgeb C, Botaljov A, Stockhardt F, Isadskiy S, Debiase L, Uhl A, et al. 2020. PRNU-based detection of facial retouching. *IET Biometrics*, 9(4): 154-164 [DOI: 10.1049/iet-bmt.2019.0196]
- Saharia C, Ho J, Chan W, Salimans T, Fleet D J and Norouzi M. 2023. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4713-4726 [DOI: 10.1109/TPAMI.2022.3204461]
- Shang S Y, Shan Z Y, Liu G X, Wang L Q, Wang X H, Zhang Z K, et al. 2024. ResDiff: combining CNN and diffusion model for image super-resolution//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI Press: 8975-8983 [DOI: 10.1609/aaai.v38i8.28746]
- Sun Z Y, Chen Y X and Xiong S W. 2022. SSAT: a symmetric semantic-aware transformer network for makeup transfer and removal//Proceedings of the 37th AAAI Conference on Artificial Intelligence. [s. l.]: AAAI Press: 2325-2334 [DOI: 10.1609/aaai.v36i2.20131]
- Sun Z Y, Xiong S W, Chen Y X, Du F, Chen W H, Wang F, et al. 2024a. SHMT: self-supervised hierarchical makeup transfer via latent diffusion models//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 510 [DOI: 10.5555/3737916.3738426]
- Sun Z Y, Xiong S W, Chen Y X and Rong Y. 2024b. Content-style decoupling for unsupervised makeup transfer without generating pseudo ground truth//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 7601-7610 [DOI: 10.1109/CVPR52733.2024.00726]
- Van der Maaten L and Hinton G E. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research*. 9: 2579-2605

- Wang J Y, Yue Z S, Zhou S C, Chan K C K and Loy C C. 2024. Exploiting diffusion prior for real-world image super-resolution. *International Journal of Computer Vision*, 132 (12): 5929-5949 [DOI: 10.1007/s11263-024-02168-7]
- Wang S Y and Fu Y. 2016. Face behind makeup//*Proceedings of the 30th AAAI Conference on Artificial Intelligence*. Phoenix, USA: AAAI Press; 58-64 [DOI: 10.1609/aaai.v30i1.10002]
- Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13 (4): 600-612 [DOI: 10.1109/TIP.2003.819861]
- Wang Z X, Zhang Z Y, Zhang X Y, Zheng H J, Zhou M Y, Zhang Y, et al. 2023. DR2: diffusion-based robust degradation remover for blind face restoration//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE: 1704-1713 [DOI: 10.1109/CVPR52729.2023.00170]
- Wei Y Y, Mao T Y, Li B A, Wang F, Li F, Zhang Z, et al. 2025. Visual and large multimodal models promote image restoration and enhancement: research progress. *Journal of Image and Graphics*, 30(5): 1197-1219 (韦炎炎, 毛天一, 李柏昂, 王飞, 李锋, 张召, 等. 2025. 视觉模型及多模态大模型推进图像复原增强研究进展. *中国图象图形学报*, 30(5): 1197-1219) [DOI: 10.11834/jig.240436]
- Wu R Y, Sun L C, Ma Z Y and Zhang L. 2024. One-step effective diffusion network for real-world image super-resolution//*Proceedings of the 38th International Conference on Neural Information Processing Systems*. Vancouver, Canada: Curran Associates Inc.: #2938 [DOI: 10.5555/3737916.3740854]
- Xiao J, Fan Z H, Li D, Fu X Y and Zha Z J. 2025. Non-causal selective state space model for image restoration. *Journal of Image and Graphics*, 30(10): 3173-3186 (肖杰, 范子豪, 李东, 傅雪阳, 查正军. 2025. 面向图像复原的非因果选择性状态空间模型. *中国图象图形学报*, 30(10): 3173-3186) [DOI: 10.11834/jig.240517]
- Ying Q C, Liu J X, Li S, Xu H S, Qian Z X and Zhang X P. 2023. RetouchingFFHQ: a large-scale dataset for fine-grained face retouching detection//*Proceedings of the 31st ACM International Conference on Multimedia*. Ottawa, Canada: ACM: 737-746 [DOI: 10.1145/3581783.3611843]
- Zamir S W, Arora A, Khan S, Hayat M, Khan F S and Yang M H. 2022. Restormer: efficient transformer for high-resolution image restoration//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 5718-5729 [DOI: 10.1109/CVPR52688.2022.00564]
- Zhang R, Isola P, Efros A A, Shechtman E and Wang O. 2018. The unreasonable effectiveness of deep features as a perceptual metric//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 586-595 [DOI: 10.1109/CVPR.2018.00068]
- Zhou E J, Fan H Q, Cao Z M, Jiang Y N and Yin Q. 2015. Learning face hallucination in the wild//*Proceedings of the 29th AAAI Conference on Artificial Intelligence*. Austin, USA: AAAI Press: 3871-3877 [DOI: 10.1609/aaai.v29i1.9795]

作者简介

刘佳欣,女,硕士研究生,主要研究方向为美颜人脸图像复原。E-mail:23210240085@m.fudan.edu.cn

李晟,通信作者,男,副教授,主要研究方向为生物数据隐私保护、多媒体内容安全和图像识别。

E-mail:lisheng@fudan.edu.cn

钱振兴,男,教授,主要研究方向为多媒体信息安全、隐写、数字水印、图像取证和密文域信号处理。

E-mail:zxqian@fudan.edu.cn

张新鹏,男,教授,主要研究方向为多媒体信息安全、AI 安全和图像处理。E-mail:zhangxinpeng@fudan.edu.cn

应祺超,男,博士后,主要研究方向为虚假新闻检测、医学图像和美颜检测。E-mail:20110240050@fudan.edu.cn