计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
62(10): 2545−2564, 2025

基于深度学习的人脸去识别化综述

戴思龙　王鹏飞　王晓玲
（华东师范大学计算机科学与技术学院　上海　200062）
（sldai@stu.ecnu.edu.cn）

Review of Face De-identification Based on Deep Learning

Dai Silong, Wang Pengfei, and Wang Xiaoling
（*School of Computer Science and Technology*, *East China Normal University*, *Shanghai* 200062）

Abstract　In recent years, the rapid advancement of deep learning technology has introduced innovative solutions to the field of face de-identification. Compared with traditional image processing techniques, deep generative models have demonstrated significant advantages in this domain, including high-quality image generation and robust model performance. This article reviews and synthesizes the theoretical explorations and research outcomes of deep learning technology in addressing face de-identification challenges. It begins by outlining the network architectures and fundamental principles employed in deep learning for face de-identification. It then delves into the de-identification methods based on these technologies, covering key techniques such as face swapping and feature perturbation, and introduces the standard experimental metrics used to evaluate these methods. Furthermore, the article summarizes the main challenges currently faced by the technology, such as the stability of posture and expression, attribute disentanglement, and the adaptability to video applications, and looks forward to the pressing issues that future research needs to address. Ultimately, this article emphasizes the importance of deep learning technology in the field of face de-identification and points out the direction for future research. It aims to provide readers with in-depth insights into the field of face de-identification and inspire new ideas and directions for future studies.

Key words　face de-identification; deep learning; deep generative models; face swapping; face editing

摘　要　近年来，随着深度学习技术的迅速发展，为人脸去识别提供了全新的解决思路. 相较于传统的图像处理技术，深度生成模型在人脸去识别领域展现出了显著的优势，包括生成图像质量高、模型鲁棒性强等特点. 综述回顾并总结了近年来利用深度学习技术在人脸去识别问题上的理论探索和研究成果. 首先概述了深度学习在人脸去识别中所采用的网络架构和基本原理，接着深入讨论了基于这些技术的去识别方法，包括面部交换、特征扰动等关键技术，并介绍了评估这些技术的标准实验指标. 进一步地，总结了当前技术面临的主要挑战，如姿态与表情的稳定性、属性解耦以及视频应用的适应性等问题，并展望了未来研究中亟需攻克的难题. 最后，强调了深度学习技术在人脸去识别领域的重要性，并指出了未来研究的方向. 综述旨在为读者提供人脸去识别领域的深入见解，并激发未来研究的新思路和方向.

关键词　人脸去识别化；深度学习；深度生成模型；人脸交换；人脸编辑

中图法分类号　TP391

收稿日期：2024-03-16；**修回日期**：2025-05-21

基金项目：国家重点研发计划项目(2021YFC3340700)；国家自然科学基金项目(62136002)
This work was supported by the National Key Research and Development Program of China (2021YFC3340700) and the National Natural Science Foundation of China (62136002).

通信作者：王晓玲(xlwang@cs.ecnu.edu.cn)

DOI: 10.7544/issn1000-1239.202440191 CSTR: 32373.14.issn1000-1239.202440191

近年来, 人工智能的迅速发展对人类社会的各个方面产生了深远影响, 特别是以人脸识别技术为代表的生物识别技术在社会医疗、社交媒体和公共安全等领域的广泛应用, 为日常生活带来了巨大便利. 然而, 这种广泛应用也引发了社会对隐私泄露的担忧. 以人脸图像为例, 每月有数十亿张图像和视频被上传到社交平台, 这些图像很容易被用来推断个人敏感信息, 从而导致个人隐私泄露和滥用, 甚至危及个人财产和社会公共安全. 近年来, 国内外相继出台了《通用数据保护条例》①《中华人民共和国个人信息保护法》《隐私权法》等法律法规, 旨在保护个人数据隐私和权利. 除了出台相关法律法规外, 技术层面的人脸去识别化技术也成为解决隐私泄露问题的一种通用方法. 人脸去识别化技术通过使用技术手段处理包含人脸的图像或视频, 减少或消除其中的个人面部身份特征. Meden 等人[1]以人脸图像为例, 介绍了现有生物特征隐私增强技术的通用方法, 如特征去识别化、特征编辑、特征模糊和特征遮蔽等技术. 彭春蕾等人[2]聚焦于图像和视频 2 个方面的人脸去识别方法, 重点介绍了图像语义保持和修改去识别化方法, 并阐述和归纳了人脸隐私保护研究现状. Ravi 等人[3]则特别关注适用于主动和辅助生活(active and assisted living, AAL)领域的视觉隐私保护技术, 讨论并分类了基于视频的 AAL 监控场景的不同技术, 并根据不同技术的分类将之与隐私保护法案联系起来.

随着 Stable Diffusion[4], Sora②等图像-视频生成模型的兴起, 人脸去识别化技术在个人数据隐私保护领域面临新的挑战. 这些先进的生成模型依赖大规模的训练数据生成高质量的内容, 但其训练数据可能来自社交媒体等公共领域, 导致生成的内容可能暴露个人隐私, 例如: 通过特定提示, GPT-2③可以输出包括姓名、地址和电子邮件地址在内的个人信息[5]; 谷歌的 Imagen 可以生成具有真实身份的图像[6], 对个人隐私构成重大威胁, 甚至会给社会带来不良影响[7-10]. 针对这些性能强大的生成模型, 唐雨潇等人[11]总结了基于深度生成模型的人脸编辑研究进展, 并指出了人脸编辑技术所面临的挑战和未来的研究方向. Wang 等人[12]从信息安全的角度回顾了生成模型生成内容中的数据安全和隐私问题, 并总结了在隐私、可用性、真实性和合规性等方面所面临的挑战. 虽然生成模型可能会泄露隐私, 但得益于其强大的生成能力, 先进的生成模型也能够推动隐私保护和去识别技术在各个领域的应用. Klemp 等人[13]将 Latent Diffusion[4]应用于 Cityscapes 数据集, 该数据集是欧洲城市的驾驶记录图片集合, 是自动驾驶领域的常用数据集, 包含大量的行人面部图像, 对该数据集中的人脸图像进行去识别化能够在一定程度上避免相关企业滥用数据集中的人脸信息. Montenegro 等人[14]通过生成对抗网络(generative adversarial network, GAN)生成不具备身份特征的胸部 X 光照片, 使其保留原有医学特征的同时避免了深度模型从医学图像中提取患者的身份特征[15-16]. Kurzhals[17]则侧重于眼动追踪数据集的隐私保护, 仅通过 Stable Diffusion 中的去噪强度、初始噪声种子和提示等参数的调整来实现对数据集中出现的人脸和文本进行去识别化. 综上所述, 人脸去识别化技术不仅具有广泛的研究前景和应用领域, 而且有助于降低个人身份信息泄露的风险, 为个人数据隐私保护带来了新思路.

早期的人脸去识别技术通常使用模糊化、像素化或遮挡等图像处理方法[18], 使人脸图像足够模糊或扭曲, 以防止人脸被识别或与特定个体身份相关联. 然而, 这些处理方法会导致人脸图像失去可检测性, 无法应用于如人脸检测等下游任务中. 此外, 以往的研究已经证明传统人脸去识别化方法无法欺骗基于深度学习的人脸识别模型[19]. 另一种早期的人脸去识别方法基于 k-same 算法[20], 这种方法利用 k 个人脸的平均人脸来替换需要去识别的人脸, 从而使人脸识别的准确率低至 $1/k$. Oh 等人[21]引入一个通用的博弈论框架来避免用户的隐私泄露. Shan 等人[22]则利用肉眼难以察觉的扰动来改变图像的特征空间表示.

随着深度学习的发展, 针对人脸去识别化问题的方法可以主要分为基于面部交换的方法和基于特征扰动的方法 2 类. 基于面部交换的方法将 2 个或多个人脸图像的特征进行对齐与融合, 生成新的合成人脸图像, 使得原始人脸的身份信息无法被识别. 这

① https://gdpr-info.eu/

② https://en.wikipedia.org/wiki/Sora_(text-to-video_model)

③ https://en.wikipedia.org/wiki/GPT-2

种方法往往需要大量的人脸数据样本和精确的特征对齐算法，以确保合成的人脸具有逼真的外观和细节. 然而，这种方法的局限性在于合成人脸可能暴露其他人的身份信息，因为它引入了其他人脸的特征. 基于特征扰动的方法则直接从原始人脸图像提取身份特征，并通过添加噪声或在超平面上旋转特征等方式，生成和原始身份不同的合成人脸图像. 这种方法的合成人脸通常比基于换脸的方法更加自然，因为它不引入其他人脸的特征，使得合成的人脸图像在视觉上与原始图像更相似，却具有不同的身份. 基于特征扰动的方法关键在于平衡合成人脸的视觉质量和去识别化效果，以确保扰动后的特征能够有效地隐藏原始人脸的身份信息.

随着深度学习的不断进步，研究人员可以进一步探索和改进这 2 种方法，以提高人脸去识别化技术的效果和实际应用的可行性. 本文旨在系统总结基于深度学习的人脸去识别方法的研究现状和进展. 图 1 给出了本综述的技术框架.

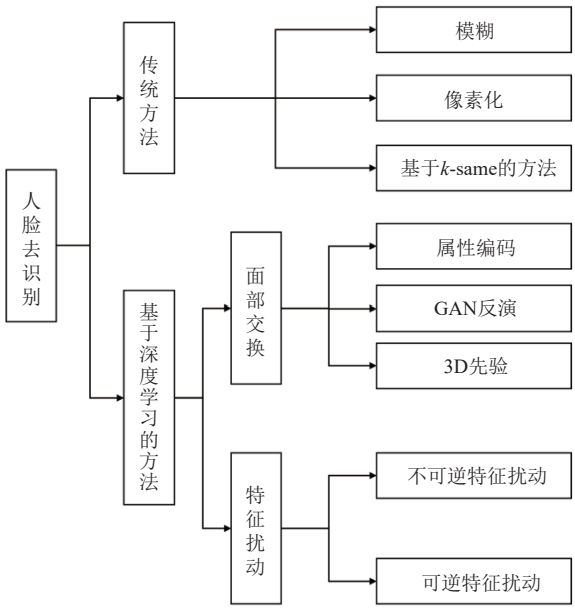


Fig. 1 Framework of face de-identification technology

图 1 人脸去识别技术框架

1 基础网络与框架

1.1 人脸去识别化

人脸去识别化(face de-identification or anonymization)被定义为一种旨在从人脸图像中移除或修改个人身份标识信息的过程，以防止通过自动化人脸识别系统或人工辨识来识别图像中的个体①. 根据去识别目标的不同，人脸去识别化技术可分为人脸身份匿名技术和可视身份隐藏技术. 前者旨在保留真实的人脸身份信息的同时，令观察人员和人脸识别模型都无法预测出身份信息；后者则令观察人员无法识别身份但能够继续让人脸识别模型提取身份信息. 这一过程通常包括应用图像处理技术、机器学习算法或深度学习模型来改变或隐藏人脸的关键特征，如眼睛、眉毛、鼻子和嘴巴等，同时尽量保持图像的其他非身份特征不变. 人脸去识别化旨在实现对个人的人脸身份信息进行保护，同时在必要时在计算机视觉、心理学和社会学等领域允许对去识别化后的图像进行进一步的分析和研究. 此外，学术界还关注于评估去识别化技术的隐私保护效果，以及在特定应用场景下对图像内容的可用性影响. 本节探讨了人脸去识别技术的核心组件，包括人脸识别模型和人脸生成网络. 人脸识别模型利用深度学习，尤其是卷积神经网络，来精确提取和识别人脸的关键特征，为去识别过程奠定了身份信息的基础. 人脸生成网络通过应用 GAN[23]、变分自编码器(variational auto-encoder, VAE)[24] 和扩散模型(diffusion model, DM)[25] 等前沿技术，生成既真实又多样化的人脸图像，同时确保不透露个体身份信息. 本节的目标是清晰地解释这些网络的结构和工作原理，以及它们如何协同工作以实现有效的人脸去识别. 这些内容为理解深度学习在人脸去识别中的应用提供了坚实的理论基础，并为后续章节的探讨做好准备.

1.2 人脸识别模型

对人脸图像进行去识别之前，首先需要使用人脸识别模型提取出原始图像的身份特征，并对身份特征进行编码. 常用的深度人脸识别模型有 DeepFace[26], FaceNet[27], ArcFace[28] 等.

DeepFace 是引入深度学习方法解决人脸识别问题的方法. 其步骤如下：首先，利用支持向量回归器提取图像中的人脸基准点，利用 LBP 直方图进行人脸检测，并裁剪出人脸区域. 鉴于图像中的人脸可能存在非正面情况，DeepFace 引入了 3D 模型和 3D 仿射相机，将 2D 图像扭曲至 3D 人脸，并通过 Delaunay 三角化将人脸转化为具有深度的 3D 三角网，并对三角网进行偏转，实现人脸的正面化. 最后，利用深度卷积神经网络以正面化后的人脸作为输入，提取出人脸的高维表示.

① https://en.wikipedia.org/wiki/De-identification

FaceNet是一种学习人脸图像的欧氏空间特征的方法,通过卷积神经网络直接对人脸图像进行学习,并将人脸识别问题表示为特征向量积的k最近邻分类问题. 为了实现将人脸图像映射到同一可度量的欧氏空间中,FaceNet采用端到端的训练方式,直接从图像学习欧氏空间的编码方法. FaceNet的网络结构与传统深度学习模型相似,但通过L2归一化取代了分类模型中的Softmax层,通过L2归一化将人脸图像转化为特征表示,并直接将人脸图像映射到欧氏空间中,同时通过三元组损失确保映射后的向量在欧氏空间中具有可度量性.

ArcFace的训练过程通常包括预训练和微调2个阶段. 在预训练阶段,通常使用大规模的人脸数据集(如MS-Celeb-1M)和基于卷积神经网络的特征提取网络(例如ResNet[29])进行初始训练. 该网络通过卷积层、池化层和全连接层等结构,逐步提取出人脸图像的特征,并将其映射到一个较低维度的特征向量表示空间. 在微调阶段,使用ArcMargin Loss作为损失函数对网络进行微调. ArcMargin Loss通过增强特征向量的可区分性,将人脸特征向量映射到一个在角度度量上具有更好区分度的高维空间,以及通过参数化的变换(例如角度余弦),对欧氏距离和类别间的角度差异进行建模,以最小化这些差异.

1.3 人脸生成网络

使用人脸识别模型提取身份特征后,对身份特征进行去识别化,并使用人脸生成网络生成去识别化后的人脸. 现有的方法大多使用GAN、VAE和DM等深度网络进行图像生成.

GAN由生成器(generator)和鉴别器(discriminator)2个主要组件组成. 生成器是一个神经网络模型,其输入为随机噪声或其他条件. 通过多层的转换和映射操作,生成器能够生成与真实数据相似的样本以尽可能地欺骗鉴别器,使其无法区分生成的样本和真实样本. 鉴别器也是一个神经网络模型,经过多层的特征提取和分类操作后,输出一个判别分数,用于表示输入样本被判别为真实样本的概率. 在训练过程中,生成器和鉴别器通过对抗性的最优化方式进行学习. 生成器试图生成更逼真的样本,以欺骗鉴别器,而鉴别器则努力提高对真实样本和生成样本的区分能力. 这种对抗性的训练过程推动着生成器生成接近真实样本的输出,并使鉴别器变得越来越准确. GAN的训练过程可以被视作是一种博弈过程,其中生成器和鉴别器是相互竞争和学习的参与者,如图2所示. 生成器通过生成接近真实样本的输出,对抗鉴别器的判别能力,而鉴别器则努力提高对真实和生成样本的判别性能. 当生成器和鉴别器达到纳什均衡时,GAN的训练结束. 在人脸去识别问题中,鉴别器用于引导生成器生成具有足够真实和自然的视觉效果的图像. 通过提高生成的人脸图像的视觉质量,鉴别器能够更好地指导生成器的学习过程,从而提高去识别的图像合成效果.

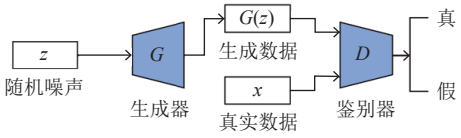


Fig. 2 Generative adversarial network (GAN)

图2 生成对抗网络

VAE是一种编码器-解码器模型,通过神经网络进行训练和优化. 如图3所示,VAE架构包括编码器和解码器2个子网络. 编码器将输入数据映射为潜在空间中的概率分布参数,例如均值和方差. 解码器则将潜在空间中的隐变量映射回原始数据空间,生成与输入数据相似的输出. VAE的目标是学习到数据的潜在表示,并从中进行采样和重构. 在训练过程中,VAE使用变分下界来约束潜在空间和原始数据之间的关系,提供了一种有效的优化方法. 变分下界允许VAE通过最大化下界来近似真实数据的潜在分布,并通过采样和解码来生成新的样本. 通过学习到的潜在表示和变分下界的约束,VAE能够生成多样化、逼真的数据样本. 它特别适用于图像生成、数据降维和特征学习等任务.

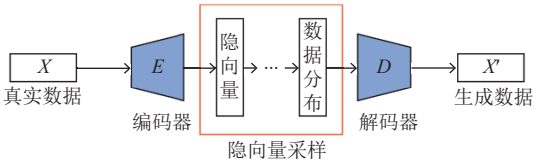


Fig. 3 VAE architecture

图3 VAE架构

DM是一种隐变量模型(latent variable model),通过马尔可夫链,在每一个时间步t中逐渐将噪声添加到数据$\boldsymbol{x}_t$中以获得后验概率$q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)$,其中$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_T$为输入数据,也是模型中的隐空间(latent space). DM具有正向的扩散过程和反向的逆扩散过程. 在正向过程中,通过T次累积对输入图像$\boldsymbol{x}_0$添加高斯噪声,由于每个时刻t只和前一时刻$t-1$有关,正向过程可以被视为马尔可夫过程:

$$q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0):=\prod_{t=1}^{T}q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}):=\prod_{t=1}^{T}N\left(\boldsymbol{x}_t;\sqrt{1-\beta_t}\boldsymbol{x}_{t-1},\beta\boldsymbol{I}\right),$$

其中$\beta_1,\beta_2,\cdots,\beta_T$为高斯分布方差的超参数. 在正向过程中, 随着$t$的增大, $\boldsymbol{x}_t$越来越接近纯噪声, 如图 4 所示, 最终收敛为标准高斯噪声$N(0,\boldsymbol{I})$. 在训练过程中, 模型学习反向的逆扩散过程的概率分布以生成新数据, 模型将从纯高斯噪声$p(\boldsymbol{x}_T):=N(\boldsymbol{x}_T;0,\boldsymbol{I})$开始, 学习联合概率分布$p_\theta(\boldsymbol{x}_{T:0})$:

$$p_\theta(\boldsymbol{x}_{T:0}):=p(\boldsymbol{x}_T)\prod_{t=1}^{T}p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t):=$$

$$p(\boldsymbol{x}_T)\prod_{t=1}^{T}N\left(\boldsymbol{x}_{t-1};\mu_\theta(\boldsymbol{x}_t,t),\sum_\theta(\boldsymbol{x}_t,t)\right).$$

根据马尔可夫规则, 逆扩散过程中的当前时刻t只取决于上一个时间步$t-1$, 故有

$$p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t):=N\left(\boldsymbol{x}_{t-1};\mu_\theta(\boldsymbol{x}_t,t),\sum_\theta(\boldsymbol{x}_t,t)\right).$$

如图 5 所示, 扩散模型的采样过程可以直观地视为去噪过程, 相比于 GAN 和 VAE 的不稳定训练, DM 的优化等同于分数匹配[30], 可以使用简单的均方误差(mean square error, *MSE*)作为损失函数.

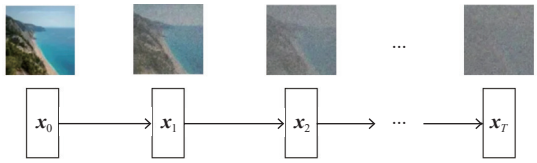


Fig. 4 Diffusion process of DM

图 4 DM 的扩散过程

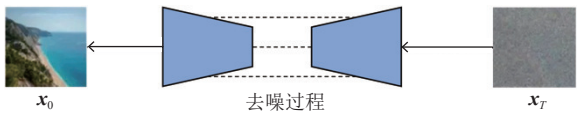


Fig. 5 The denoising process of DM

图 5 DM 的去噪过程

本节我们对人脸去识别技术中人脸识别模型和人脸生成网络的作用进行了精炼概括. 人脸识别模型通过深度学习精准地捕捉人脸特征, 为人脸去识别提供了关键的身份信息. 而人脸生成网络则通过先进的生成技术, 如 GAN、VAE 和 DM, 创造出既真实又匿名的人脸图像, 确保隐私保护的同时, 满足了去识别的需求. 这些网络的协同工作不仅展示了深度学习在人脸去识别领域的强大潜力, 也为后续的研究提供了丰富的理论支持和实践指导. 人脸识别和深度生成网络是 2 个活跃的研究课题, 但由于它们不是本文研究的重点, 建议通过综述文献来了解最新的研究进展[31-33]. 通过本节内容, 我们对人脸去识别技术的核心机制有了更清晰的认识, 为进一步探索其在保护隐私方面的应用奠定了理论基础.

2 传统的人脸去识别方法

在深度学习兴起以前, 人们主要采用传统的图像处理技术对人脸身份进行去识别化. 这些技术通常包括图像裁剪、模糊处理、像素化以及遮挡等方法, 目的是通过改变图像的特定区域来隐藏个体的身份信息. 最早的人脸去识别方法可能是对原始图像进行像素化、模糊化或使用黑色方框遮挡人脸[18]. Agrawal 等人[34]将这些方法应用于视频中, 其首先使用基于定向梯度直方图的检测器检测人物, 并利用 GrabCut 分割视频中的人与背景, 再使用基于体素的分割对生成的轮廓进行分割, 最后通过指数模糊或线积分卷积将将视频中的人脸模糊化. 这些方法不仅用于隐藏图像中的人脸, 还可以用于隐藏其他的生物特征或物体等, 被称为临时的去识别方法. 这种方法破坏了人脸图像的效用, 使得处理后的人脸图像并不能在下游任务中使用. Phillips[35]使用 PCA 投影降低人脸图像质量, 通过在人脸重建过程中减少特征脸数量以丢失人脸细节, 进而使图像更加难以识别. 相比于模糊等直接破坏图像人脸的方法, 这种方法虽然能够产生非常不自然的图像以降低人脸识别模型识别精度, 但仍然保留了一些可用于自动识别的面部特征. 随着深度学习的发展, 深度模型已经能够从被处理的图像中提取人脸的身份信息[19].

Newton 等人[20]介绍了一种基于 *k*-same 的 *k*-匿名模型. 这种方法将一组输入人脸图像划分为至少包含k个图像的子集, 并用子集平均值替换子集中的每个图像. 由于图像未对齐和图像人脸姿态差异等原因, 这种方法生成的去识别人脸会出现重影等情况. 针对上述问题, Gross 等人[36]设计一效用函数, 该函数根据图像的相似性对图像进行分组, 并利用主动外观模型[37]对齐图像. Bitouk 等人[38]将 2 个人脸图像对齐到同一公共坐标系, 并通过人脸交换来进行人脸去识别. 在 Gross 等人[36]的基础上, Samarzija 等人[39]通过面部模型对齐人脸, 以保留原始人脸的面部表情和姿态等属性.

早期的人脸去识别方法主要通过图像处理技术来实现, 这些技术包括像素化、模糊化、遮挡等. 这些方法的优点在于操作简单、易于实施, 能够直接隐藏个体的身份信息. 例如, 通过像素化处理, 可以将人脸的详细特征模糊化, 使得人脸难以被识别; 而模

糊化处理则通过降低图像的清晰度来减少人脸识别的可能性；遮挡则是直接用图形或物体覆盖人脸，从而阻止识别. 然而，这些方法也存在一些缺点. 首先，它们往往会破坏图像的视觉效果，导致图像的可用性降低. 其次，这些方法可能无法完全保护个人隐私. 随着技术的发展，一些先进的人脸识别算法能够从经过简单处理的图像中提取身份信息，这使得简单的去识别方法在面对高级人脸识别技术时可能不再有效. 新的方法，如基于 k-匿名模型、主动外观模型和人脸交换等技术，提供了更为有效和精确的去识别解决方案. 这些方法通过更复杂的图像处理技术，能够在一定程度上保留人脸图像的自然性和可用性，同时实现去识别的目的. 例如，k-匿名模型通过图像分组和平均处理来保护隐私，而主动外观模型则通过图像对齐和特征保留来减少失真. 尽管如此，这些新方法也面临挑战和局限性. 一方面，去识别过程中可能会损失图像的细节和质量，导致生成的人脸图像不自然或失真. 另一方面，这些方法的计算成本相对较高，需要更多的计算资源和时间. 此外，随着深度学习技术的发展，一些深度学习模型可能仍然能够从这些经过处理的图像中提取出身份信息，这要求去识别技术不断更新以应对新的挑战.

3 基于深度学习的人脸去识别化方法

自深度学习火热以后，人脸识别技术和深度生成网络迎来了革命性的进步，而人脸去识别方法也随之发展，能够在保护隐私的同时生成逼真的人脸图像. 现有的基于深度学习的人脸去识别方法将任务分解为特征提取、身份匿名化和匿名图像生成 3 个子任务. 在身份匿名化子任务中，可通过是否融合外部特征将人脸去识别化方法划分为基于面部交换和基于特征扰动 2 种. 基于面部交换的方法将人脸特征分为身份特征和属性特征，并替换或融合其他人脸图像的身份特征，最终生成匿名化的人脸图像. 基于面部交换的人脸去识别方法概述如表 1 所示. 基于特征扰动的方法在提取到的人脸特征中添加噪声，以实现身份匿名化，并生成与原始人脸视觉上高度相似的去识别化人脸图像. 基于特征扰动的人脸去识别方法概述如表 2 所示.

3.1 基于面部交换的人脸去识别

基于面部交换的人脸去识别化方法将参与去识别化的 2 张图像分为源图像和目标图像. 源图像为需要去识别化的人脸图像，目标图像为需要保留身份特

Table 1 Face Swapping Based Face De-Identification Method

表 1 基于面部交换的人脸去识别方法

方法	算法	编码器	生成模型	模型特点
属性编码	DeepFakes[40]	CNN	GAN	身份特征集成在模型权重内
	SimSwap[41]	ArcFace	GAN	使用身份特征注入，解耦权重和身份特征
	Nitzan 等人[42]	VGGFace2[43]，InceptionV3[44]	StyleGAN[45]	显式解耦身份和属性特征
	Faceshifter[46]	ArcFace	GAN	自适应调整身份和属性嵌入减少伪影
	InfoSwap[47]	ArcFace	GAN	通过信息瓶颈解耦身份和属性
	StyleFace[48]	VAE	StyleGAN2[49]	自适应提取人脸属性
	Smooth-Swap[50]		NCSN++[51]	光滑嵌入身份特征
	BlendFace[52]		AEI-Net[53]	在合成人脸上预训练编码器
	DiffFace[54]	ArcFace，CosFace[55]	DDPM[25]	避开不稳定对抗训练
	FIVA[56]	ArcFace	AAE[57]	噪声注入防御重建攻击
	CodeSwap[58]	VQGAN[59]	VQGAN	无需引入人脸识别网络编码身份
	FaceComposer[60]	ArcFace，PNCC[61]，Face Clip[62]	LDM[4]	对齐多个特征完成不同面部任务
	DiffusionAct[63]	DiffAE[64]	DiffAE	使用预训练生成器
GAN 反演	MegaFS[65]	ResNet50[29]，FPN[66]	StyleGAN2	无显式解耦控制身份信息
	Xu 等人[67]	pSp Encoder[68]	StyleGAN	在潜在空间解耦身份和属性
	RAFSwap[69]	pSp Encoder，BiseNet[70]	StyleGAN2	Transformer 建模全局与局部交互
	StyleSwap[71]		StyleGAN2	图像重建优化身份潜码
	E4S[72]	FaceVid2Vid[73]	StyleGAN	局部 GAN 反演解开形状和纹理的耦合
	CPP-DeID[74]	E4S，ArcFace	StyleGAN2	显式建模对去识别图像的要求

表 1（续）

方法	算法	编码器	生成模型	模型特点
3D 先验	DiscoFaceGAN[75]	3DMM VAE	StyleGAN2	多个 VAE 分别编码不同属性
	HifiFace[76]	3DMM Curricularface[77]		结合人脸 3D 形状和身份特征
	FaceController[78]	ArcFace，SEAN[79]，3DMM	SPADE[80]	集成身份和区域风格信息
	WSC-swap[81]	3DMM	GAN	避免编码器的长跳跃连接
	3dSwap[82]	pSp	EG3D	结合 3D 先验和 GAN 反演方法
	StyleIPSB[83]	3DMM	StyleGAN	对齐 3DMM 参数和生成器坐标
	DiffSwap[84]	VQGAN，ArcFace	DDIM[85]	可指定执行面部交换的区域
	Lopez[86]	Heatmap Regression Network	StarGANv2[87]	从硬件层面抑制原始身份
	StableSwap[88]	ArcFace，Deng 等人[89]	VQGAN	多阶段注入目标身份
	SelfSwapper[90]	3DMM，BiseNet	StyleGAN2	通过自监督训练模型

Table 2 Feature-Perturbation Based Face De-Identification Method

表 2 基于特征扰动的人脸去识别方法

方法	算法	扰动方式	生成模型	模型特点
不可逆特征扰动	DeepPrivacy[91]	渐进生成	GAN	渐进式生成对抗网络
	CIAGAN[92]	身份鉴别器	CGAN	高精度人脸身份识别
	Li 等人[93]	模糊	StarGAN	引入人脸认知和身份热力图
	DeepPrivacy2[94]	渐进生成	Surface-guided GAN	全身身份去识别化
	PI-Net[95]	添加噪声	GAN	潜在空间中操纵面部语义属性并添加噪声
	Ji 等人[96]	移除频域 DC 通道		引入频域信息去除身份
	PartialFace[97]	移除低频通道		低频通道隐藏视觉信息，高频通道训练身份识别模型
	PASS[98]	删除单一属性		去除人脸属性敏感信息
	A^3GAN[99]	抑制卷积去除身份特征	GAN	属性注入模块
	BenSaid[100]	修改身份潜码	StyleGAN	多层特征编码
	IDeudemon[101]	添加噪声	StyleGAN2	3D 属性解耦
	Zhang 等人[102]	动态扩散	LDM	多个扩散模型协同作用
可逆特征扰动	Zhang 等人[103]	模糊		监督升级的渐进式训练策略
	Gu 等人[104]	添加噪声	InfoGAN	用户自定义密码
	Anonym-Recognizer[105]	添加噪声	LSGAN[106]	训练额外的人脸提取器
	MfM[107]	添加噪声		二进制编码
	Cao 等人[108]	特征旋转	GAN	密码控制身份特征旋转角度

征的人脸图像. 基于面部交换的方法旨在保留源图像的身份无关特征(如姿态、表情和发型等), 并生成具有目标图像身份特征的人脸以达到去识别化的效果. 通过面部交换的方式进行人脸去识别化通常有 3 种思路: 第 1 种是基于属性编码的思想, 即使用编码器编码源图像和目标图像的人脸属性, 其中人脸属性包含了人脸的身份特征. 随后, 在潜在空间中将人脸属性解耦成身份特征和身份无关特征, 并将源图像的身份无关特征和目标图像的身份特征结合, 生成具有目标图像身份特征的去识别人脸图像; 第 2 种是基于 GAN 反演的方法, 该方法利用 GAN 的生成器和鉴别器网络, 将原始人脸图像通过生成器映射到潜在空间中, 然后再从潜在空间中反向映射回人脸图像, 目标是使得反演后的人脸图像尽可能接近原始图像, 并同时去除身份特征. 在这个过程中, 鉴别器网络用于区分生成的人脸图像和真实的人脸图像, 从而引导生成器生成更逼真的去识别人脸图像. 通过这种方式, 可以实现对人脸图像的去识别, 并且生成的去识别人脸图像在视觉上与原始图像保持一致, 避免了信息丢失和失真的问题. 第 3 种是基于 3D 先验的方法, 该方法使用 3D 人脸模型[109]提取人脸特征, 相比于人脸识别网络是为分类任务设计, 3D 人脸模型可以敏锐地捕捉到人脸的形状和局部细节, 进而为去识别人脸生成提供更多的人脸形状先验信息.

在人脸去识别化技术的研究中, 基于面部交换的方法因其能够保留原始人脸图像的身份无关特征而受到关注. 主要分为 3 种方法: 首先, 属性编码思想通过编码器将人脸属性编码, 并在潜在空间中将身份特征与身份无关特征解耦. 这种方法能够生成同时具备目标身份特征和原始人脸属性的去识别人脸图像. 然而, 它可能在图像对齐不精确时产生伪影或失真. 其次, 基于 GAN 反演的方法利用模型的生成器和鉴别器网络来获得图像潜码. 这种方法能够在保持原始图像视觉效果的同时, 生成难以被识别模型识别的去识别人脸图像. 尽管如此, 它在训练过程中需要大量的计算资源, 并且可能在面部细节处理上存在局限性. 最后, 基于 3D 先验的方法使用 3D 人脸模型[110]来提取人脸的形状和局部细节信息, 为生成去识别人脸图像提供了丰富的几何先验. 这种方法在生成逼真的人脸图像方面具有优势, 但可能需要复杂的 3D 处理流程. 综上所述, 每种方法都有其特定的应用场景和局限性. 在实际应用中, 选择合适的去识别化方法需要综合考虑隐私保护效果、图像质量和计算资源等要求.

3.1.1 基于属性编码的面部交换

基于属性编码的面部交换方法一般使用 2 个编码器 E_i 和 E_a, 分别提取目标人脸图像的身份特征和原始人脸的身份无关特征, 并利用一映射网络 M 将编码结果映射至生成器的潜在空间 W 中, 最终生成去识别的人脸图像. 如图 6 所示, 其中人脸图像取自文献 [42], 身份编码器通常为预训练的人脸识别网络. 而 DeepFakes[40] 由 1 个公共编码器 E 与 2 个特定身份解码器 D_S 和 D_T 组成. DeepFakes 将扭曲后的源人脸图像和目标人脸图像作为输入, D_S, D_T 学习将扭曲后的图像还原为原始图像. 在推理过程中, 将需要去识别化的源人脸图像作为输入, 面向目标人脸的解码器 D_T 会将输入的图像视为扭曲的图像, 进而将源人脸图像去识别化为具有目标人脸身份特征的人脸图像. DeepFakes 在训练过程中将目标人脸的身份特征集成到特定身份解码器 D_T 中, 若要更换目标人脸, 需要对模型进行重新训练. SimSwap[41] 针对 DeepFakes 只能生成单一身份特征人脸的问题, 在编码器和解码器之间添加额外的身份注入模块, 将提取出的目标人脸身份特征嵌入至原始人脸中. SimSwap 使用来自不同身份的图像进行训练, 解码器将专注于从特征中恢复图像, 其权重是和任意身份无关的, 而身份修改的任务则由身份注入模块完成, 使得训练后的 SimSwap 可以应用于任意的目标人脸而不需要重新训练模型. Nitzan 等人[42] 则分别使用基于 ResNet 的人脸识别网络和 InceptionV3[44] 对目标人脸身份特征和源人脸的身份无关属性进行编码, 明确地将人脸属性解耦为身份特征和身份无关特征, 将身份特征和属性特征连接后, 使用 4 层 MLP 和 LeakyRelu 组成的映射网将特征映射至生成器的潜在空间中, 最终生成去识别人脸的图像.

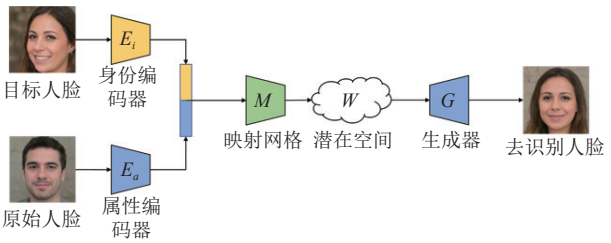


Fig. 6 Attribute encoding based face swapping method

图 6 基于属性编码的面部交换方法

上述方法直接将目标人脸的身份特征注入到源人脸的属性中, 忠实地保留了原始人脸的身份无关属性, 然而, 在人脸外观差异较大的情况下, 由于没有对源人脸的属性进行调整, 生成的人脸往往由于伪影而不够真实, 为减少生成图像的伪影, Faceshifter[46] 设计出自适应嵌入集成网络, 通过提取源人脸不同尺度的特征, 并利用自适应注意力非规范化(adaptive attentional denormalization)将身份特征和属性特征积分, 自适应地调整身份嵌入和 属性嵌入以减少生成图像的伪影. InfoSwap[47] 从信息瓶颈的角度解耦原始人脸的身份和属性, 在向前传播的过程中通过预训练的人脸识别模型量化和压缩信息流, 一次性学习身份和属性特征. InfoSwap 通过信息控制器寻找身份和属性特征的最佳解耦, 压缩后的身份和属性特征被送至自适应信息集成器以进行面部交换. Luo 等人[48] 利用 StyleGAN2 的潜在空间和风格控制机制, 引入自适应属性提取器, 提出 StyleFace 实现了百万级像素的面部交换. StyleFace 将人脸识别网络中的深层特征视为身份先验, 将身份先验映射至 StyleGAN2 的低耦合潜在空间 W 中, 并在生成器的不同层次中使用不同的损失函数以在提高保真度的同时保留源人脸的细节. 由于面部交换任务大多使用预训练的人脸识别网络作为目标人脸的身份编码器, 而人脸识别网络仅通过辨别任务进行训练, 其嵌入空间高度不光滑, 导致其难以在训练过程中为生成器指出良好的梯度方向, 进而导致生成的图像保真度较低, 使得先前的工作侧重于对生成器的网络结构进行手工设计, 并利用大量的损失函数约束模型生成高质量的

图像. Kim 等人[50]提出 Smooth-Swap, 参考有监督的对比学习损失[111]来构造潜在空间平滑身份编码器, 并将噪声评分网络(NCSN++)[51]中的时间嵌入更改为身份嵌入, 增加输入到输出的直接跳跃连接来作为生成器. 得益于嵌入空间光滑的身份编码器, Smooth-Swap 仅使用 3 个基本的损失函数便与现有模型性能相当. Shiohara 等人[52]发现使用预训练的 ArcFace 作为身份编码器会不可避免地保留目标人脸的非身份特征(如头发等), 这是由于 ArcFace 的训练数据集偏差导致的. 故 Shiohara 等人在对目标人脸进行编码前, 通过使用交换属性的合成人脸图像预训练一个身份编码器 BlendFace 以解耦身份特征和属性特征, 并将解开的身份特征输入生成器, 作为身份损失函数引导生成器生成无属性偏差的去识别人脸. DiffFace[54]直接使用 DM 作为生成器, 避免了对抗训练带来的不稳定性, 并使用面部解析器[112]约束生成的图像遵循目标人脸的面部结构. 为避免去识别人脸被图像重建模型重识别, FIVA[56]将噪声和原始图像混合后再输入至对抗自编码器(adversarial auto-encoder, AAE)[57]进行去识别, 并通过 U-Net 架构的图像重建模型作为重建攻击模型, 通过实验表明噪声注入对防御重建攻击的有效性. CodeSwap[58]利用 VQGAN 进行大量的面部重建任务来获取面部的先验知识. 这些先验知识富含高质量的面部特征, 将作为先验码本(prior codebook)以提高生成图像的质量. CodeSwap 将图像编码至码本空间(code space)后, 通过一个基于 Transformer[113]的全局融合网络, 从码本中识别和替换原始人脸和目标人脸的特定片段以实现面部交换, 无需额外的人脸识别网络提取身份特征. 通过对齐空间信息、人脸身份和人脸描述文本[62]等多个先验, FaceComposer[60]能够作为一个统一的模型同时处理包括面部交换、人脸编辑等多种面部任务. 除常用的遮罩之外, FaceComposer 使用 PNCC[61]和 Sketch 来编码人脸的空间局部信息, 通过不同组件之间的配合来实现不同的任务, 例如, 通过引入目标人脸身份并编码原始人脸空间信息来实现面部交换. 和以往的使用现成的人脸识别模型作为身份编码器的方法不同, DiffusionAct[63]通过面部重建任务对 2 个编码器进行预训练, 这些编码器的架构和 DiffAE[64]的编码器相同, 分别编码原始人脸的身份和目标人脸的身份无关属性. 编码器完成预训练后, 将与预训练好的 DiffAE 生成器继续训练, 以进一步优化编码器权重, 通过 landmark 和 EMOCA[114]等约束人脸形状, 减少人脸图像的伪影.

3.1.2 基于 GAN 反演的面部交换

编码器对人脸的编码可能会导致人脸特征的信息损失, 进而使得生成器无法获得准确的人脸潜码, 无法生成优质的人脸图像. GAN 的反演(GAN inversion)算法有以下 3 种思路: 1)训练编码器对图片编码, 用于预测图像的潜码; 2)随机生成潜码[115], 通过优化方法来最小化预测图像和给定图像潜码之间的误差; 3)将前两者结合, 优化编码器以初始化潜在代码. 基于 GAN 反演的面部交换方法相比于属性解耦的方法往往需要更高的计算成本, 但得益于精确的人脸特征编码结果, 其图像生成能力和身份保持能力也更加强大. MegaFS[65]基于残差块和 ResNet 的主干网络设计出 HieRFE 作为人脸图像的编码器, 并提出人脸转移模块, 在没有显示特征解耦的情况下同步控制身份信息的多个属性以满足人脸交换的需求, 最终使用 StyleGAN2 生成人脸图像. 由于人脸的身份特征和属性特征在潜在空间中高度耦合, MegaFS 在无监督的情况下进行人脸的交换难以保留精细的人脸细节. Xu 等人[67]使用预训练的 pSp 编码器[68]将人脸图像反演为 18 个不同的向量, 前 7 个向量为身份特征, 其余的为属性特征, 进而对人脸的身份和属性分别进行转移和拼接, 并利用 landmark 获取源人脸的表情和姿势特征.

RAFSwap[69]利用分层编码器和 BiseNet[70]提取人脸局部属性, 并引入 Transformer 层来建模全局的身份特征和局部的人脸属性之间的交互. RAFSwap 捕捉了 GAN 反演丢失的局部人脸细节, 减少了生成人脸和目标人脸的身份误差. StyleSwap[71]则将目标图像的身份潜码反演至 StyleGAN 的潜在空间 W 中, 通过图像重建的方式迭代优化源人脸的身份潜码以提高生成人脸和目标人脸的身份相似性. Liu 等人[72]从细粒度人脸编辑的角度设计出人脸交换框架 E4S (editing for swapping), 通过局部 GAN 反演(regional GAN inversion)的方式解开人脸形状和纹理的耦合, 将人脸交换任务简化为相应的样式潜码和面部遮罩的交换, 但源人脸和目标人脸的遮罩可能无法完全对准, 进而导致交换后遮罩的一小部分区域丢失像素, 由于 E4S 提取的纹理特征包括肤色, 故模型直接使用肤色填充像素缺失的区域, 使得模型不需要训练额外的子网络来处理遮挡人脸的情况. 在 E4S 的基础上, CPP-DeID[74]通过多个损失函数平衡人脸去识别过程中对图像质量、身份保护和数据效用性的要求. 通过人为设置的权重和 E4S 框架, CPP-DeID 迭代优化去识别人脸图像在 StyleGAN 中的潜在空间分

布,使得人脸去识别过程更加可控.

3.1.3 基于3D先验的面部交换

基于GAN反演的方法能够精确地提取人脸属性潜码,但GAN的反演过程繁琐且计算成本高昂,故研究者希望使用一些现有的、容易获得的先验信息(如人脸形状和纹理等)来取代GAN反演的属性解耦表示学习过程. Deng等人[75]提出DiscoFaceGAN,将人脸图像分为身份、表情、光照条件、姿势和其他噪声等5个独立因素,并使用3DMM处理人脸图像,编码前4个独立因素,将获得的4个参数训练4个VAE,将潜码映射至更具解释性的λ空间,并将λ空间馈送到生成器以生成交换后的人脸. DisconFaceGAN利用VAE限制生成器的生成方向,可以控制生成人脸的属性来自源人脸还是目标人脸.然而,DiscoFaceGAN仅使用3DMM作为人脸特征编码器,无法模拟身份细节,导致生成的去识别人脸和目标人脸的身份相似性较弱. HifiFace[76]认为人脸的形状属于人脸身份特征,故结合人脸识别网络和3DMM提出3D形状感知身份提取器,明确地引入3D几何结构信息,并使用人脸识别网络提取的目标人脸身份特征以及3DMM提取的源人脸表情和姿势特征作为辅助监督来实现精确的人脸形状转化. FaceController[78]考虑3DMM提取的特征更加关注全局信息,故引入SEAN作为区域风格编码器,为局部面部区域提供风格的解耦表示,并将身份和区域风格信息集成到生成器中.编码器和生成器之间的长跳跃连接会将一定的原始人脸身份泄露到生成人脸中,为避免长跳跃连接, WSC-swap[81]使用2个编码器分别编码人脸图像中的身份特征和属性特征,并通过3DMM模型计算属性保留损失,进一步解开人脸图像中的身份特征和属性特征. 3dSwap[82]和StyleIPSB[83]均训练映射网络并将人脸的3D模型的参数投影至生成器的潜在空间中. 3dSwap设计了一种伪多视图优化策略,使用pSp的特征金字塔架构训练编码器将人脸图像投影至EG3D[116]的潜在空间中,并设计一种风格混合和插值组成的潜在代码操作算法以解耦人脸图像的身份特征和属性特征,最后对EG3D的生成器参数进行微调,使用一组损失函数进行强制身份转换和属性保留;相比于3dSwap将3D图像重新渲染成2D图像所需的长时间推理, StyleIPSB则提出了一个3DMM到生成器模型的映射网络,该网络将3DMM参数的语义信息转换为生成器坐标,从而实现对生成图像的精确控制,并有效减少了计算成本.与3dSwap不同的是, StyleIPSB将人脸交换任务分为3步:首先将原始人脸图像的属性转换到目标人脸图像,然后学习重演面部和原始人脸之间的残差属性,最后将面部融合到原始人脸图像的背景中. Lopez等人[86]观察到使用特殊的光学器件如低分辨率相机和散焦镜头等采集图像,能够有效滤除图像中的高频信息,故其从硬件层面出发,通过联合优化隐私保护计算相机(privacy-preserving computational camera)和热力图回归网络来获得人脸热力图,该过程在过滤掉高频信息和隐藏身份特征的同时能够保留重要特征.生成模型通过该热力图提供的原始人脸几何先验和风格提取器提取的目标人脸身份生成去识别人脸. StableSwap[88]通过感知和对抗损失优化自动编码器以将目标人脸身份潜在对齐至生成器的潜在空间中,并利用预训练好的3D人脸重建模型[89]来获取人脸图像的3D信息,最后利用多头自注意力机制来捕获面部的身份、形状和背景之间的相互作用,以实现多阶段的目标身份注入. SelfSwapper[90]通过自监督任务进行训练,通过前景网格掩膜和遮挡掩膜[70]解析原始人脸,并通过解析结果重建图像,使得模型学会如何准确地重建输入图像,同时不泄露目标人脸的身份信息.在推理过程中,模型使用从目标人脸图像中提取的身份特征,结合原始人脸的非身份特征,通过生成器网络输出重建或交换后的面部图像.

上述方法使用GAN或StyleGAN作为生成器,然而,基于GAN的生成器需要对网络架构和损失函数进行精心设计,以融合来自源人脸和目标人脸的特征. DM不以对抗的方式进行训练,可以使用简单的损失函数(如MAE等)进行训练,并能够在源人脸和目标人脸差异较大的情况下忠实地保留人脸特征. Zhao等人[84]将面部交换任务视为条件修复任务,提出基于DM的面部交换框架DiffSwap. DiffSwap使用ArcFace提取目标图像的身份特征,并使用landmark提取目标人脸的全局形状特征和源人脸的局部形状特征.多重感知机将这些特征投影至相同的维度,并将其作为条件特征输入至DM以进行训练.在推理过程中, DiffSwap可以通过构造掩码的方式指定执行面部交换的区域,并使用3D人脸重建库提取源人脸和目标人脸的面部信息,简单地将重建的源人脸形状替换为目标人脸形状,并使用交换的参数以生成去识别的人脸.

3.2 基于特征扰动的人脸去识别

基于面部交换的人脸去识别方法引入外部人脸图像,将人脸去识别任务视为面部交换任务,其人脸生成过程是有监督的,生成人脸的身份和引入的外

部人脸图像身份高度相近. 由于引入了外部的真实人脸, 生成的去识别人脸身份为外部人脸的真实身份, 仍有隐私泄露的风险. 如图 7 所示, 其中人脸图像取自文献 [93]. 基于特征扰动的方法旨在通过特征旋转、添加噪声或利用生成模型渐进生成过程等方式扰动原始人脸的身份特征, 不引入外部人脸图像以减少隐私泄露的风险. 部分研究工作考虑到去识别后的人脸图像在某些情况下需要恢复原始的图像, 故本文根据去识别后的人脸是否可以重新获取原始人脸的身份特征, 将基于特征扰动的人脸去识别方法分为不可逆的特征扰动方法与可逆的特征扰动方法.

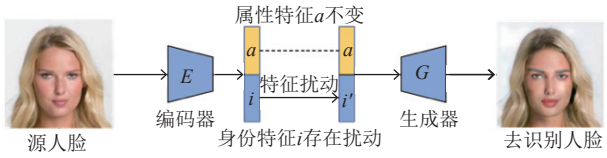


Fig. 7 Feature-perturbation based face de-identification method

图 7 基于特征扰动的人脸去识别方法

3.2.1 不可逆的特征扰动方法

DeepPrivacy[91] 使用左右眼、左右耳、左右肩和鼻子 7 个关键点来描述原始人脸的姿态信息, 并将该 7 个关键点编码为 $7\times H\times W$ 的 one-hot 编码以作为生成器的输入, 其中 $H\times W$ 为图片的分辨率. DeepPrivacy 利用 GAN 的渐进生成过程更改原始人脸的身份特征, 并没有显示地控制生成的人脸图像和原始人脸图像具有不同的身份. CIAGAN[92] 则通过额外设计的身份鉴别器来引导生成器生成具有虚假身份的去识别人脸图像, 并使用高精度视觉库(DLibrary, dlib)[117] 提取原始人脸的 landmark 来保留原始人脸的姿态和表情. 此外, 由于 landmark 保留了原始人脸的表情和姿势, 故相比于视频翻译[118] 等方法, CIAGAN 可以直接获得视频内同一身份人脸在不同帧下的姿态, 进而以较低的计算成本直接对视频中的人脸进行去识别化. Li 等人[93] 从人类认知的角度进行人脸去识别, 使得生成的去识别人脸在视觉上远离源身份而仍旧能被人脸识别网络识别, 其使用 CAM 计算身份感知激活热力图, 用于寻找并保留身份无关的人脸区域, 并利用 StarGAN[119] 生成去识别后的人脸, 使得生成的人脸图像具有模糊的视觉外观. 在原先工作的基础上, DeepPrivacy2[94] 将去识别技术推广至人体的去识别, 分别通过 DSFD[120], CSE[121], Mask R-CNN[122] 处理图像以进行人脸检测、密集姿态估计和实例分割. DeepPrivacy2 将处理结果分为 3 类: 具有密集姿势估计的个体、未通过 CSE 检测到的个体和不包含在前一类别的人脸, 并对每个类别单独进行去识别化过程, 利用 Surface-guided GAN[123] 的生成器与 StyleGAN2 的解码器进行图像重建以实现人体去识别.

差分隐私[124] 是一种具有效用保护的严格隐私概念, Chen 等人[95] 和 Ji 等人[96] 将差分隐私引入人脸去识别任务. PI-Net[95] 将人脸的不可识别性定义为 GAN 潜在空间中图像之间的相邻关系, 利用三元组损失和编码网络聚类面部语义属性, 并添加噪声至潜在空间潜码以抑制原始语义属性并更改潜码身份, 使得最终的生成图像难以被人脸识别网络识别. Ji 等人[96] 则使用离散余弦变换将人脸图像转移至频域后, 直接移除频域中的 DC 通道, 并通过人脸识别的损失对不同频率的元素进行隐私分配以实现人脸的隐私保护, 由于 DC 通道包含了大部分的视觉信息, 且直接移除 DC 通道的过程是不可逆的, 故攻击者无法通过频域信息还原原始的人脸图像. 与 Ji 等人类似, Mi 等人[97] 提出的 PartialFace 会剪枝人脸图像的低频通道以隐藏视觉信息, 并在几个随机选择的高频通道上训练身份识别模型, 使得在人脸识别时并不需要存储肉眼可识别的人脸图像.

由于差分隐私的引入, PI-Net 生成的图像具有较模糊的视觉外观, Ji 等人[96] 生成的图像更是在视觉上不具备可识别性, 其生成结果往往难以继续在下游任务中使用(如表情识别等). 为在后续任务中继续使用去识别的人脸图像, PASS[98]、A^{3}GAN[99] 和 BenSaid 等人[100] 期望通过操纵人脸属性以更改原始人脸身份. PASS 利用对抗学习减少人脸属性中的敏感信息, 在训练时只激活一个属性分类器并冻结其他分类器以帮助模型删除单一的人脸属性, 进而在匿名化原始人脸的情况下根据下游任务的要求生成具有特定属性的人脸; A^{3}GAN 将人脸去识别化分为原始身份抑制和属性注入 2 个任务, 通过抑制卷积单元来抑制原始身份信息, 并使用属性感知注入网络将属性编码对齐以匹配身份特征, 进而控制属性注入来生成客制化的去识别人脸; BenSaid 等人[100] 则利用 pSp 微调 StyleGAN 网络中不同层的特征编码以实现精确的局部细节保留与身份特征的更改, 进而生成更加自然的去识别人脸图像. 人脸去识别需要权衡身份保护效果和图像效用, 上述方法并没有对这 2 个要求进行显式地处理, Wen 等人[101] 提出 IDeudemon, 分别使用 NeRF 模型[125] 计算出的 3D 属性解耦和身份潜码进行混合以保护身份, 并通过视觉相似性辅助来调整 GAN 的损失以保留图像效用. Zhang

等人[102]通过多个DM(身份条件DM、遮罩驱动DM和文本驱动DM)的协同作用,动态决定图像生成过程中每个模型的参与程度,以动态扩散的方式控制去识别人脸图像的特征.

3.2.2 可逆的特征扰动方法

为在某些情况从匿名化后的人脸获取原始人脸的身份信息(如通过监控视频获取犯罪嫌疑人身份信息),如图8所示,其中人脸图像取自文献[108].部分工作使用密码显式控制扰动方式,即控制向原始身份特征注入特定的噪声或显式地控制身份特征的旋转角度实现了可逆的人脸去识别化.Zhang等人[103]提出了一种监督升级的渐进式训练策略,以模糊图像作为监督,使模型在模糊人脸图像的同时使生成的图像能够被人脸识别模型再识别.Gu等人[104]将人脸图像$\boldsymbol{I}\in\mathbb{R}^{H\times W\times 3}$和用户定义的密码$\boldsymbol{p}\in\{0,1\}^N$作为输入,并将$\boldsymbol{I}$和$\boldsymbol{p}$深度级联$(\boldsymbol{I},\boldsymbol{p})\in\mathbb{R}^{H\times W\times(3+N)}$,使得原始图像的每个像素都与密码的每一位高度相关,并基于InfoGAN[126]训练一个辅助网络,最大化图像域中注入的密码和身份变化之间的互信息.

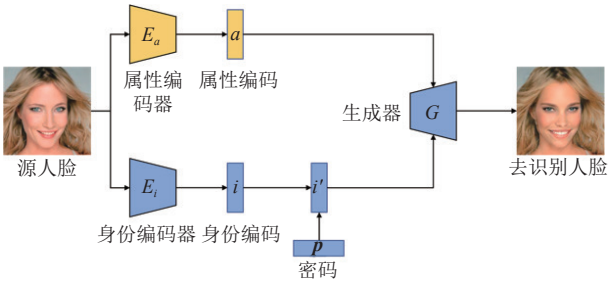


Fig. 8 Reversible feature perturbation method

图8 可逆的特征扰动方法

与Gu等人[104]不同,Peng等人[105]不从去识别人脸图像恢复原始人脸的视觉信息,而是通过额外训练一匿名人脸识别网络Anonym-Recognizer提取嵌入原始人脸图像的密码,并根据密码获得原始人脸的身份信息.由于Gu等人[104]和Peng等人[105]将人脸图像和密码深度级联,并没有将人脸图像中的身份信息和属性信息解耦,使得需要设计复杂的损失函数来约束生成的图像以远离原始身份.Pan等人[107]在Gu等人[104]的基础上,使用人脸识别网络显示地编码人脸图像的身份特征,并使用N位的二进制密码$\boldsymbol{p}\in\{0,1\}^N$和身份特征异或以形成新的身份特征编码.

上述方法直接将密码和人脸图像或身份特征级联抑制原始身份特征,这种属性抑制方法是不可控的,且会带来较高的计算成本.Cao等人[108]则受到ArcFace, CosFace等将身份特征转换到超球面空间的启发.设计一密码-身份映射函数$\boldsymbol{z}_{\text{new}}=f(\boldsymbol{z}_{\text{id}},\boldsymbol{p})$,通过密码控制身份特征的旋转角度,进而将人脸去识别.需要注意的是,这种方法在去识别人脸恢复原始身份的过程中,若密码错误,则会生成具有其他身份的人脸,如图9所示,其中人脸图像取自文献[108].

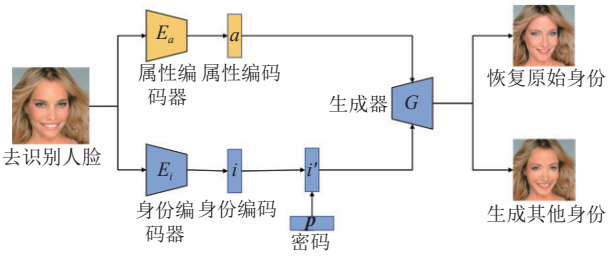


Fig. 9 Re-identification of de-identified faces

图9 去识别人脸的重识别

4 人脸去识别结果评价指标

人脸去识别也须生成新的图像,故其为图像生成任务的一个子任务,本文将过去工作中常用的评价指标分为去识别效果评估指标和图像质量评估指标.常用的去识别效果评估指标包括身份检索得分、姿态误差、表情误差和弗雷歇距离(Fréchet distance, FID)[127]等.在人脸去识别任务中,往往要求去识别后的人脸图像具有和源图像相近的姿势和表情;而身份检索指标的要求则会根据去识别方法变化,即基于面部交换的人脸去识别方法期望生成的去识别人脸具有目标人脸的身份,故其身份检索得分越高越好;基于特征扰动的人脸去识别方法由于没有引入外部身份,故越低的身份检索得分表明其对原始人脸身份的去识别化越好.由于去识别的人脸图像可能仍旧需要进行人脸相关的下游任务,故除常用的均方根误差(root mean square error, *RMSE*)、峰值信噪比(peak signal-to-noise ratio, *PSNR*)和结构相似性指数测量(structure similarity index measure, *SSIM*)[128]等,还有人脸检测得分以保证生成的人脸图像能够满足下游任务的要求.

4.1 去识别效果评估指标

对于身份检索得分,过去的工作使用不同的人脸识别网络[55]提取人脸图像的身份特征,基于面部交换的人脸去识别化方法计算目标人脸图像的身份特征x_{t}和生成的去识别人脸图像的身份特征x_{de}之间的余弦相似度,并使用该余弦相似度来评判生成的去识别人脸是否具有目标人脸的身份;基于特征扰动的人脸去识别方法则使用原始人脸图像的身份特

征x_s进行身份相似度的计算，越低的余弦相似度表明人脸去识别化的效果越好.

姿态误差和表情误差分别通常使用姿态估计器[129]和表情特征提取器[130]计算原始图像和去识别图像的人脸姿态嵌入$\boldsymbol{p}_s,\boldsymbol{p}_{de}$和表情嵌入$\boldsymbol{e}_s,\boldsymbol{e}_{de}$，并使用欧氏距离计算嵌入之间距离，并以该距离作为去识别人脸图像和源人脸图像的姿态误差和表情误差.

相比于欧氏距离，弗雷歇距离考虑2条曲线之间的最短连续路径，通过一个参数化的连续函数联系2条曲线，并在该连续函数的参数范围内找到最大的欧氏距离. 在源人脸图像和去识别人脸图像的比较中，弗雷歇距离考虑了曲线的形状和方向，对图像的全局性质更加敏感，而欧氏距离是一种点之间的距离度量，忽略了曲线的形状和方向，仅考虑点的位置. 弗雷歇距离计算公式为：

$$d_F(A,B)=\min_{\gamma}\max_{t\in[0,1]}dist(A(\gamma(t)),B(\gamma(t))),$$

其中$A(\gamma(t)),B(\gamma(t))$分别表示曲线$A$和$B$在参数$t$下的点，$\gamma$表示曲线$A$到$B$的映射. 在过去的工作中，曲线$A$和$B$表示2张人脸图像的特征曲线或轮廓线，这些曲线可以是人脸图像的边缘轮廓，也可以是2个人脸的特征点集合(如眼睛、鼻子、嘴等关键点).

4.2 图像质量评估指标

相比于去识别指标对生成人脸图像的身份特征改变、非身份特征保留的要求，图像质量评估指标更侧重于去识别人脸的图像质量，包括生成的图像有无伪影和畸变等. RMSE常用于衡量2幅图像之间的差异，过去的工作中常用其评估原始人脸图像和生成的人脸图像之间的差异，计算公式为：

$$RMSE=\sqrt{\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(\boldsymbol{I}(i,j)-\boldsymbol{I}'(i,j))^2},$$

其中m和n分别表示图像的高度和宽度，$\boldsymbol{I}(i,j)$和$\boldsymbol{I}'(i,j)$表示原始图像和去识别图像在像素点(i,j)处的像素值. *RMSE*的值越小表明2幅图像之间的差异越小，由于原始的人脸图像为真实图像，一般不会有伪影和畸变，故越小的*RMSE*值表明生成的去识别人脸图像具有更好的图像质量.

相比于*RMSE*通过2幅图像之间每个像素值之差的平方的均值的平方根来评估图像之间的差异，PSNR更多地用于比较原始图像和经过处理后的图像之间的质量差异，其计算公式为：

$$MSE=\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(\boldsymbol{I}(i,j)-\boldsymbol{I}'(i,j))^2,$$

$$PSNR=10\lg\left(\frac{MAX^2}{MSE}\right),$$

其中*MAX*表示图像像素值的最大可能值. *PSNR*的值越高，表明2幅图像之间的相似度越高.

RMSE, *PSNR*从图像像素的角度比较2幅图像的差异，故它们并不总能准确地反映人类对图像质量的主观感知. *SSIM*则考虑了人眼对图像的感知特性，从亮度、对比度和结构比较图像. 此外，*SSIM*将图像分成多个局部窗口，并对每个窗口内的亮度、对比度和结构进行比较，为减少噪声对*SSIM*的影响，每个窗口内的像素值都会进行高斯加权，使得处于窗口中心的像素具有更高的权重. 最终，*SSIM*会将原始图像和生成的去识别图像之间的亮度相似性、对比度相似性和结构相似性综合起来，计算它们的加权乘积作为最终的*SSIM*指数. *SSIM*数值在[0,1]的范围内，“1”表示2幅图像完全相同，“0”表示2幅图像完全不同.

除*RMSE*, *PSNR*, *SSIM*等通用图像质量评价指标，过去的部分工作也会使用人脸检测器来证明相比于模糊、像素化和遮挡等传统方法，基于深度学习的人脸去识别化方法生成的人脸仍能够满足下游任务. 如使用HOG[131], SSH[132]来对生成的去识别图像进行人脸检测.

5 人脸去识别化面临的挑战

5.1 姿态与表情的变化

人脸去识别任务需要保留源人脸图片的姿态和表情. 在基于面部交换的去识别方法中，当目标人脸的姿态和表情与源人脸图片的差异较大时，生成的去识别人脸图片的姿态和表情可能会和目标人脸更加接近. 而基于特征扰动的去识别方法也会生成不够自然的人脸. 为保留原始人脸的姿态和表情，通常采用以下思路：

1)通过引入外部3D模型(如3DMM等)，显式地对源人脸的姿态和表情进行建模[75,81]，进而利用额外的损失函数有效地约束去识别人脸的生成过程.

2)在编码过程中，对源人脸的姿态和表情进行特征提取，并将编码后的特征和源人脸的其他身份无关特征进行加权.

3)针对基于面部交换的人脸去识别方法，可以对抗训练身份编码器或通过对比学习等方式[88,98]确保提取到的目标人脸身份不包含其他身份无关特征.

5.2 属性解耦的准确度

除肤色、性别等常见身份无关属性外，人脸去识

别化可能会要面对如胡须、眼镜和妆容等更复杂的、可能会遮挡或改变源人脸外观和局部结构的情况. 在这种情况下, 准确解耦人脸的身份特征和身份无关特征变得更加具有挑战性, 为解决这一问题, 有以下思路:

1)采用更细粒度的属性编码器或注意力机制对人脸图像的非身份特征进行编码[61,63], 这些编码器能够更容易获取人脸的局部信息.

2)使用 one-hot 编码显示地编码人脸的局部特征[92]. 通过这种方式, 可以将人脸的不同部位(如眼镜、胡须和妆容等)的信息独立地表示出来, 从而更好地解耦人脸身份特征和身份无关特征, 进而增加模型对人脸局部特征的关注度, 提高模型对人脸属性解耦的准确度.

3)对于多属性的解耦, 可以采用多个分类器进行迭代训练, 使得模型能够准确地解耦出单一属性[98].

5.3 视频中的人脸去识别

目前的工作主要侧重于人脸图像的去识别, 若将图像上的去识别方法应用至视频中, 需要考虑身份一致性、时序连贯性和实时性等问题. 将人脸图像的去识别方法应用至视频中有以下 3 个思路:

1)确定视频中人脸的身份, 对不同帧下的同一身份人脸, 设计额外的损失函数约束在去识别后保持同一身份.

2)采用结合 landmark、光流计算[133] 等方式, 提高人脸去识别视频的流畅度与真实性, 减少因帧之间光照、姿态的不同生成的伪影, 如 CIAGAN[92] 直接使用 landmark 编码人脸形状, 能够直接获得视频内同一身份人脸在不同帧下的姿态以减少去识别人脸伪影.

3)设计轻量级的模型或对模型进行压缩和量化[134-136], 减少模型的参数量和计算量, 在不降低准确性的前提下提高模型的推理速度, 以满足在实时监控、视频会议等场景下的实时性要求.

6 总结与展望

图 10 给出了本文总结的基于深度学习的人脸去识别技术概览. 人脸去识别技术能够有效地降低身份信息泄露的风险, 可以广泛应用于社交媒体、公共场合视频监控等领域.

和传统方式相比, 基于深度学习的人脸去识别方法有 3 个优势:

1)效用性高. 相比于传统方法直接破坏源人脸的视觉信息, 基于深度学习的人脸去识别模型生成的去识别人脸仍旧可以被检测到, 不影响图片或视频的视觉质量.

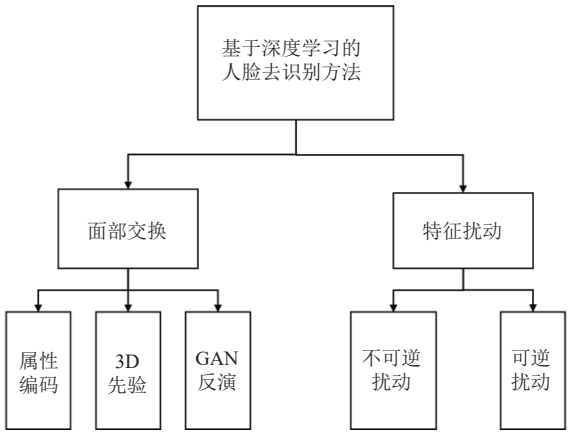


Fig. 10 Overview of face de-identification techniques based on deep learning

图 10 基于深度学习的人脸去识别技术概览

2)高质量的输出. 结合现有深度生成模型的优势, 能够生成真实、自然的去识别化人脸.

3)可逆性. 对源人脸进行去识别后, 可以通过事先保存的密钥恢复原始人脸.

但不容忽视的是, 基于深度学习的人脸去识别还面临着种种挑战. 首先, 基于面部交换的人脸去识别模型存在 3 个问题: 1)对人脸进行编码和特征解耦, 可能会导致人脸的特征信息损失, 使得生成器无法获得准确的人脸潜码, 若设计多重损失函数约束人脸特征, 则会提高模型的复杂度, 使得模型的训练成本提高. 2)基于 GAN 反演的方式虽然能够获得准确的人脸潜码, 但 GAN 反演需要较大的显存要求和时间开销, 且反演的结果很大程度上依赖于预训练的生成器模型, 若输入图像和预训练生成器所用图像并非来自同一数据集, 生成器本身的偏差会影响最终的生成结果. 3)基于 3D 先验的面部交换方法需要引入外部模型, 其生成图像的结果很大程度上依赖于 3D 先验. 其次, 基于特征扰动进行人脸去识别的思路也有 2 个限制: 1)对身份特征进行扰动可能会破坏人脸图像的视觉效用, 虽然生成的去识别图像无法被人眼辨别, 杜绝了人脸信息泄露的可能, 但生成的图片也无法在下游任务中使用. 2)对身份特征扰动需要将人脸的身份属性和非身份属性准确解耦, 若属性解耦不够准确, 生成的去识别人脸图像会不够真实、自然. 最后, 从整体而言, 该项技术面临 3 个挑战: 1)无论采用何种方案, 由于每种属性所独有的性质, 模型的通用性与针对性难以达到平衡. 2)模

型在复杂场景下的表现还不够理想，例如光线变化、遮挡和多人场景等，在这些情况下，模型往往会生成具有伪影的图像. 3)对视频中的人脸去识别还有诸多挑战，例如保持去识别身份一致性、时序连贯性和实时性等，在不同帧中，同一身份的人脸可能会具有不同的表情、姿态或遮挡. 此外，还需要保持处理后的视频具有时序连贯性，确保视频的观感和连贯性不会受到影响. 同时，人脸去识别技术还需要具备实时性，能够在视频流中实时地进行去识别操作，以满足实时监控、视频会议等应用场景的需求.

未来的研究方向可以着眼于3个方面：1)针对面部交换的思想，可以采用对抗训练的思想，确保编码器只编码人脸图像的身份特征或非身份特征，以提高属性解耦的精度. 2)针对特征扰动的思想，可以探索更加细粒度和有效的特征扰动方法，以保证在去识别的同时，尽可能地保留原始图像的视觉信息和质量. 3)现有的人脸去识别网络在视频中还具有较大的发展潜力，其中一项重要的研究方向是改进模型的鲁棒性和实时性. 在视频应用中，人脸可能会出现光照变化、姿势变化、遮挡等情况，基于图片的人脸去识别网络在处理这些情况时表现可能不稳定，导致去识别人脸质量不高.

7 结束语

本文介绍了基于深度学习的人脸去识别技术的进展，从原理和结构出发，分别介绍了传统的、基于面部交换和基于特征扰动的人脸去识别方法，并对基于深度学习的人脸去识别方法主要面临的姿态和表情变化、属性解耦准确度和视频中的人脸去识别3个方面的挑战进行阐述，最后总结了该方法的优缺点与未来的发展方向. 总体来说，相比于传统去识别方法对原始人脸图片的破坏，基于深度学习的人脸去识别方法提高了去识别人脸的图像效用，使其能够应用于下游任务，但在开发和使用人脸去识别技术时，需要考虑用户的隐私和数据安全问题，防止个人敏感信息被泄露或滥用，尤其是基于面部交换的人脸去识别方法，需要确保生成的图像不会被滥用或误用.

作者贡献声明：戴思龙和王鹏飞负责论文主要内容的撰写和校验；王晓玲对论文的结构和内容提供了重要的指导意见. 戴思龙和王鹏飞为共同第一作者.

参 考 文 献

[1] Meden B, Rot P, Terhörst P, et al. Privacy–enhancing face biometrics: A comprehensive survey[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 4147−4183

[2] Peng Chunlei, Miao Zimin, Liu Decheng, et al. Visual identity privacy protection: Research methods for face anonymization[J]. Chinese Journal of Computers, 2023, 46(11): 2431−2452 (in Chinese)
(彭春蕾, 苗紫民, 刘德成, 等. 视觉身份隐私保护：人脸匿名化研究方法[J]. 计算机学报, 2023, 46(11): 2431−2452)

[3] Ravi S, Climent-Pérez P, Florez-Revuelta F. A review on visual privacy preservation techniques for active and assisted living[J]. Multimedia Tools and Applications, 2024, 83(5): 14715−14755

[4] Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 10684−10695

[5] Carlini N, Tramer F, Wallace E, et al. Extracting training data from large language models[C]//Proc of the 30th USENIX Security Symp (USENIX Security 21). Berkeley, CA: USENIX Association, 2021: 2633−2650

[6] Carlini N, Hayes J, Nasr M, et al. Extracting training data from diffusion models[C]//Proc of the 32nd USENIX Security Symp (USENIX Security 23). Berkeley, CA: USENIX Association, 2023: 5253−5270

[7] Yang Hongyu, Li Xinghang, Hu Ze. A survey on deepfake face generation and detection technology [J/OL]. Journal of Huazhong University of Science and Technology: Natural Science Edition. [2025-04-14]. https://doi.org/10.13245/j.hust.250021 (in Chinese)
(杨宏宇, 李星航, 胡泽. 深度伪造人脸生成与检测技术综述[J/OL]. 华中科技大学学报：自然科学版. [2025-04-14]. https://doi.org/10.13245/j.hust.250021)

[8] Lai Zhimao, Zhang Yun, Li Dong. A survey of face deepfake detection technology based on Transformer[J]. Journal of Guangdong University of Technology, 2023, 40(6): 155−167 (in Chinese)
(赖志茂, 章云, 李东. 基于Transformer的人脸深度伪造检测技术综述[J]. 广东工业大学学报, 2023, 40(6): 155−167)

[9] Lin Chenhao, Shen Chao, Deng Jingyi, et al. Techniques for synthetic digital face content generation and detection[J]. Chinese Journal of Computers, 2023, 46(3): 469−498 (in Chinese)
(蔺琛皓, 沈超, 邓静怡, 等. 虚假数字人脸内容生成与检测技术[J]. 计算机学报, 2023, 46(3): 469−498)

[10] Li Zeyu, Zhang Xuhong, Pu Yuwen, et al. A survey on multimodal deepfake and detection techniques[J]. Journal of Computer Research and Development, 2023, 60(6): 1396−1416 (in Chinese)
(李泽宇, 张旭鸿, 蒲誉文, 等. 多模态深度伪造及检测技术综述[J]. 计算机研究与发展, 2023, 60(6): 1396−1416)

[11] Tang Yuxiao, Wang Binjun. Research progress on face editing based on deep generative models[J]. Computer Science, 2022, 49(2): 51−61 (in Chinese)

(唐雨潇, 王斌君. 基于深度生成模型的人脸编辑研究进展[J]. 计算机科学, 2022, 49(2): 51−61)

[12] Wang Tao, Zhang Yushu, Qi Shuren, et al. Security and privacy on generative data in AIGC: A survey[J]. arXiv preprint, arXiv: 2309.09435, 2023

[13] Klemp M, Rösch K, Wagner R, et al. LDFA: Latent diffusion face anonymization for self-driving applications[C]//Proc of the 2023 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 3198−3204

[14] Montenegro H, Cardoso J S. Anonymizing medical case-based explanations through disentanglement[J]. arXiv preprint, arXiv: 2311.04833, 2023

[15] Esmeral L C M, Uhl A. Low-effort re-identification techniques based on medical imagery threaten patient privacy[C]//Proc of the Annual Conf on Medical Image Understanding and Analysis. Berlin: Springer, 2022: 719−733

[16] Packhäuser K, Gündel S, Münster N, et al. Deep learning-based patient re-identification is able to exploit the biometric nature of medical chest X-ray data[J/OL]. Scientific Reports, 2022 [2023-12-23]. https://www.nature.com/articles/s41598-022-19045-3.pdf

[17] Kurzhals K. Anonymizing eye-tracking stimuli with stable diffusion[J/OL]. Computers & Graphics, 2024 [2023-12-03]. https://www.sciencedirect.com/science/article/pii/S0097849324000256

[18] Neustaedter C, Greenberg S, Boyle M. Blur filtration fails to preserve privacy for home-based video conferencing[J]. ACM Transactions on Computer-Human Interaction, 2006, 13(1): 1−36

[19] Acquisti A, Gross R. Imagined communities: Awareness, information sharing, and privacy on the Facebook[C]//Proc of the 2006 Int Workshop on Privacy Enhancing Technologies. Berlin: Springer, 2006: 36−58

[20] Newton E M, Sweeney L, Malin B. Preserving privacy by de-identifying face images[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(2): 232−243

[21] Oh S J, Fritz M, Schiele B. Adversarial image perturbation for privacy protection a game theory perspective[C]//Proc of the 2017 IEEE Int Conf on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2017: 1491−1500

[22] Shan S, Wenger E, Zhang Jiayun, et al. Fawkes: Protecting privacy against unauthorized deep learning models[C]// Proc of the 29th USENIX Security Symp (USENIX Security 20). Berkeley, CA: USENIX Association, 2020: 1589−1604

[23] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J/OL]. Advances in Neural Information Processing Systems, 2014 [2023-12-22]. https://proceedings.neurips.cc/paper_files/paper/2014/file/f033ed80deb0234979a61f95710dbe25-Paper.pdf

[24] Kingma D P, Welling M. Auto-encoding variational Bayes[J]. arXiv preprint, arXiv: 1312.6114, 2013

[25] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[J]. Advances in Neural Information Processing Systems, 2020, 33: 6840−6851

[26] Taigman Y, Yang M, Ranzato M A, et al. DeepFace: Closing the gap to human-level performance in face verification[C]//Proc of the 2014 IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 1701−1708

[27] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering[C]//Proc of the 2015 IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 815−823

[28] Deng Jiankang, Guo Jia, Xue Nianan, et al. ArcFace: Additive angular margin loss for deep face recognition[C]//Proc of the 2019 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 4690−4699

[29] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[C]//Proc of the 2016 IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770−778

[30] Batzolis G, Stanczuk J, Schönlieb C B, et al. Conditional image generation with score-based diffusion models[J]. arXiv preprint, arXiv: 2111.13606, 2021

[31] Adjabi I, Ouahabi A, Benzaoui A, et al. Past, present, and future of face recognition: A review[J]. Electronics, 2020, 9(8): 1188

[32] Gui Jie, Sun Zhenan, Wen Yonggang, et al. A review on generative adversarial networks: Algorithms, theory, and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 35(4): 3313−3332

[33] Chen Fuji, Zhu Feng, Wu Qingxiao, et al. A survey on generative adversarial networks and their applications in image generation[J]. Chinese Journal of Computers, 2021, 44(2): 347−369 (in Chinese)
(陈佛计, 朱枫, 吴清潇, 等. 生成对抗网络及其在图像生成中的应用研究综述[J]. 计算机学报, 2021, 44(2): 347−369)

[34] Agrawal P, Narayanan P J. Person de-identification in videos[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2011, 21(3): 299−310

[35] Phillips P J. Privacy operating characteristic for privacy protection in surveillance applications[C]//Proc of the 2005 Int Conf on Audio- and Video-Based Biometric Person Authentication. Berlin: Springer, 2005: 869−878

[36] Gross R, Airoldi E, Malin B, et al. Integrating utility into face de-identification[C]//Proc of the 5th Int Workshop on Privacy Enhancing Technologies (PET 2005). Berlin: Springer, 2006: 227−242

[37] Gross R, Sweeney L, De la Torre F, et al. Model-based face de-identification[C]//Proc of the 2006 Conf on Computer Vision and Pattern Recognition Workshop (CVPRW'06). Piscataway, NJ: IEEE, 2006: 161−161

[38] Bitouk D, Kumar N, Dhillon S, et al. Face swapping: Automatically replacing faces in photographs[C]//Proc of the SIGGRAPH 2008. New York: ACM, 2008: 1−8

[39] Samarzija B, Ribaric S. An approach to the de-identification of faces in different poses[C]//Proc of the 37th Int Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). Piscataway, NJ: IEEE, 2014: 1246−1251

[40] Faceswap. DeepFakes[CP/OL]. 2020[2025-03-03]. https://github.com/deepfakes/ faceswap

[41] Chen Renwang, Chen Xuanhong, Ni Bingbing, et al. SimSwap: An

efficient framework for high fidelity face swapping[C]//Proc of the 28th ACM Int Conf on Multimedia. New York: ACM, 2020: 2003−2011

[42] Nitzan Y, Bermano A, Li Yangyan, et al. Face identity disentanglement via latent space mapping[J]. arXiv preprint, arXiv: 2005.07728, 2020

[43] Cao Qiong, Shen Li, Xie Weidi, et al. Vggface2: A dataset for recognising faces across pose and age[C]//Proc of the 13th IEEE Int Conf on Automatic Face & Gesture Recognition (FG 2018). Piscataway, NJ: IEEE, 2018: 67−74

[44] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proc of the 2016 IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 2818−2826

[45] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks[C]//Proc of the 2019 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 4401−4410

[46] Li Lingzhi, Bao Jianming, Yang Hao, et al. Advancing high fidelity identity swapping for forgery detection[C]//Proc of the 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 5074−5083

[47] Gao Gege, Huang Huaibo, Fu Chaoyou, et al. Information bottleneck disentanglement for identity swapping[C]//Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 3404−3413

[48] Luo Yuchen, Zhu Junwei, He Keke, et al. StyleFace: Towards identity-disentangled face generation on megapixels[C]//Proc of the 2022 European Conf on Computer Vision. Berlin: Springer, 2022: 297−312

[49] Karras T, Laine S, Aittala M, et al. Analyzing and improving the image quality of styleGAN[C]//Proc of the 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 8110−8119

[50] Kim J, Lee J, Zhang B T. Smooth-Swap: A simple enhancement for face-swapping with smoothness[C]//Proc of the 2022 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 10779−10788

[51] Song Yang, Sohl-Dickstein J, Kingma D P, et al. Score-based generative modeling through stochastic differential equations[J]. arXiv preprint, arXiv: 2011.13456, 2020

[52] Shiohara K, Yang Xingchao, Taketomi T. BlendFace: Re-designing identity encoders for face-swapping[C]//Proc of the 2023 IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2023: 7634−7644

[53] Li Lingzhi, Bao Jianmin, Yang Hao, et al. Advancing high fidelity identity swapping for forgery detection [C]//Proc of the 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 5074−5083

[54] Kim K, Kim Y, Cho S, et al. DiffFace: Diffusion-based face swapping with facial guidance[J]. arXiv preprint, arXiv: 2212.13344, 2022

[55] Wang Hao, Wang Yitong, Zhou Zheng, et al. CosFace: Large margin cosine loss for deep face recognition[C]//Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 5265−5274

[56] Rosberg F, Aksoy E E, Englund C, et al. FIVA: Facial image and video anonymization and anonymization defense[C]//Proc of the 2023 IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2023: 362−371

[57] Makhzani A, Shlens J, Jaitly N, et al. Adversarial autoencoders[J]. arXiv preprint, arXiv: 1511.05644, 2015

[58] Luo Xiangyang, Zhang Xin, Xie Yifan, et al. CodeSwap: Symmetrically face swapping based on prior codebook[C]// Proc of the 32nd ACM Int Conf on Multimedia. New York: ACM, 2024: 6910−6919

[59] Esser P, Rombach R, Ommer B. Taming transformers for high-resolution image synthesis[C]//Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 12873−12883

[60] Wang Jiayu, Zhao Kang, Ma Yifeng, et al. FaceComposer: A unified model for versatile facial content creation[J/OL]. Advances in Neural Information Processing Systems, 2024 [2023-12-23]. https://proceedings.neurips.cc/paper_files/paper/2023/file/2b4caf39e645680f826ae0a9e7ae9402-Paper-Conference.pdf

[61] Zhu Xiangyu, Lei Zhen, Liu Xiaoming, et al. Face alignment across large poses: A 3D solution[C]//Proc of the 2016 IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 146−155

[62] Zheng Yinglin, Yang Hao, Zhang Ting, et al. General facial representation learning in a visual-linguistic manner[C]//Proc of the 2022 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 18697−18709

[63] Bounareli S, Tzelepis C, Argyriou V, et al. DiffusionAct: Controllable diffusion autoencoder for one-shot face reenactment[J]. arXiv preprint, arXiv: 2403.17217, 2024

[64] Preechakul K, Chatthee N, Wizadwongsa S, et al. Diffusion autoencoders: Toward a meaningful and decodable representation[C]//Proc of the 2022 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 10619−10629

[65] Zhu Yuhao, Li Qi, Wang Jian, et al. One shot face swapping on megapixels[C]//Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 4834−4844

[66] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proc of the 2017 IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 2117−2125

[67] Xu Yangyang, Deng Bailin, Wang Junle, et al. High-resolution face swapping via latent semantics disentanglement[C]//Proc of the 2022 IEEE/CVF Conf on Computer Vision and Pattern recognition. Piscataway, NJ: IEEE, 2022: 7642−7651

[68] Richardson E, Alaluf Y, Patashnik O, et al. Encoding in style: A styleGAN encoder for image-to-image translation[C]//Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 2287−2296

[69] Xu Chao, Zhang Jiangning, Hua Miao, et al. Region-aware face swapping[C]//Proc of the 2022 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 7632−7641

[70] Yu Changqian, Wang Jingbo, Peng Chao, et al. BiseNet: Bilateral segmentation network for real-time semantic segmentation[C]//Proc of the European Conf on Computer Vision (ECCV). Berlin: Springer, 2018: 325−341

[71] Xu Zhiliang, Zhou Hang, Hong Zhibin, et al. StyleSwap: Style-based generator empowers robust face swapping[C]//Proc of the 2022 European Conf on Computer Vision. Berlin: Springer, 2022: 661−677

[72] Liu Zhian, Li Maomao, Zhang Yong, et al. Fine-grained face swapping via regional GAN inversion[C]//Proc of the 2023 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 8578−8587

[73] Wang Tingchun, Mallya A, Liu Mingyu. One-shot free-view neural talking-head synthesis for video conferencing[C]//Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 10039−10049

[74] Meden B, Gonzalez-Hernandez M, Peer P, et al. Face deidentification with controllable privacy protection[J/OL]. Image and Vision Computing, 2023[2023-12-22]. https://www.sciencedirect.com/science/article/pii/S0262885623000525

[75] Deng Yu, Yang Jiaolong, Chen Dong, et al. Disentangled and controllable face image generation via 3D imitative-contrastive learning[C]//Proc of the 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 5154−5163

[76] Wang Yuhan, Chen Xu, Zhu Junwei, et al. HifiFace: 3D shape and semantic prior guided high fidelity face swapping[J]. arXiv preprint, arXiv: 2106.09965, 2021

[77] Huang Yuge, Wang Yuhan, Tai Ying, et al. Curricularface: Adaptive curriculum learning loss for deep face recognition[C]//Proc of the 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 5901−5910

[78] Xu Zhiliang, Yu Xiyu, Hong Zhibin, et al. FaceController: Controllable attribute editing for face in the wild[C]//Proc of the 2021 AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI 2021, 35(4): 3083−3091

[79] Zhu Peihao, Abdal R, Qin Yipeng, et al. SEAN: Image synthesis with semantic region-adaptive normalization[C]//Proc of the 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 5104−5113

[80] Park T, Liu Mingyu, Wang Tingchun, et al. Semantic image synthesis with spatially-adaptive normalization[C]//Proc of the 2019 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 2337−2346

[81] Ren Xiaohang, Chen Xingyu, Yao Pengfei, et al. Reinforced disentanglement for face swapping without skip connection[C]//Proc of the 2023 IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2023: 20665−20675

[82] Li Yixuan, Ma Chao, Yan Yichao, et al. 3D-aware face swapping[C]//Proc of the 2023 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 12705−12714

[83] Jiang Diqiong, Song Dan, Tong Ruofeng, et al. StyleIPSB: Identity-preserving semantic basis of StyleGAN for high fidelity face swapping[C]//Proc of the 2023 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 352−361

[84] Zhao Wenliang, Rao Yongming, Shi Weikang, et al. DiffSwap: High-fidelity and controllable face swapping via 3D-aware masked diffusion[C]//Proc of the 2023 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 8568−8577

[85] Song Jiaming, Meng Chenlin, Ermon S. Denoising diffusion implicit models[J]. arXiv preprint, arXiv: 2010.02502, 2020

[86] Lopez J, Hinojosa C, Arguello H, et al. Privacy-preserving optics for enhancing protection in face de-identification[C]//Proc of the 2024 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2024: 12120−12129

[87] Choi Y, Uh Y, Yoo J, et al. StarGANv2: Diverse image synthesis for multiple domains[C]//Proc of the 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 8188−8197

[88] Zhu Yixuan, Zhao Wenliang, Tang Yansong, et al. StableSwap: Stable face swapping in a shared and controllable latent space[J/OL]. IEEE Transactions on Multimedia, 2024[2024-01-27]. https://ieeexplore.ieee.org/abstract/document/10444967/

[89] Deng Yu, Yang Jiaolong, Xu Sicheng, et al. Accurate 3D face reconstruction with weakly-supervised learning: From single image to image set[C/OL]//Proc of the 2019 IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE, 2019[2023-12-23]. https://openaccess.thecvf.com/content_CVPRW_2019/papers/AMFG/Deng_Accurate_3D_Face_Reconstruction_With_Weakly-Supervised_Learning_From_Single_Image_CVPRW_2019_paper.pdf

[90] Lee J, Hyung J, Jeong S, et al. SelfSwapper: Self-supervised face swapping via shape agnostic masked autoencoder[J]. arXiv preprint, arXiv: 2402.07370, 2024

[91] Hukkelås H, Mester R, Lindseth F. DeepPrivacy: A generative adversarial network for face anonymization[C]//Proc of the 2019 Int Symp on Visual Computing. Berlin: Springer, 2019: 565−578

[92] Maximov M, Elezi I, Leal-Taixé L. CIAGAN: Conditional identity anonymization generative adversarial networks[C]//Proc of the 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 5447−5456

[93] Li Jingzhi, Han Lutong, Chen Ruoyu, et al. Identity-preserving face anonymization via adaptively facial attributes obfuscation[C]//Proc of the 29th ACM Int Conf on Multimedia. New York: ACM, 2021: 3891−3899

[94] Hukkelås H, Lindseth F. DeepPrivacy2: Towards realistic full-body anonymization[C]//Proc of the 2023 IEEE/CVF Winter Conf on Applications of Computer Vision. Piscataway, NJ: IEEE, 2023: 1329−1338

[95] Chen Jiawei, Chen Liju, Yu C M, et al. Perceptual indistinguishability-net (PI-Net): Facial image obfuscation with

manipulable semantics[C]//Proc of the 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 6478−6487

[96] Ji Jiazhen, Wang Huan, Huang Yuge, et al. Privacy-preserving face recognition with learnable privacy budgets in frequency domain[C]//Proc of the 2022 European Conf on Computer Vision. Berlin: Springer, 2022: 475−491

[97] Mi Yuxi, Huang Yuge, Ji Jiazhen, et al. Privacy-preserving face recognition using random frequency components[C]//Proc of the 2023 IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2023: 19673−19684

[98] Dhar P, Gleason J, Roy A, et al. PASS: Protected attribute suppression system for mitigating bias in face recognition[C]//Proc of the 2021 IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2021: 15087−15096

[99] Zhai Liming, Guo Qing, Xie Xiaofei, et al. A^3GAN: Attribute-aware anonymization networks for face de-identification[C]//Proc of the 30th ACM Int Conf on Multimedia. New York: ACM, 2022: 5303−5313

[100] BenSaid E, Neji M, Alimi A M. Natural face anonymization via latent space layers swapping[C]//Proc of the 2023 IEEE Symp on Computers and Communications (ISCC). Piscataway, NJ: IEEE, 2023: 562−567

[101] Wen Yunqian, Liu Bo, Cao Jingyi, et al. Divide and conquer: A two-step method for high quality face de-identification with model explainability[C]//Proc of the 2023 IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2023: 5148−5157

[102] Zhang Yushu, Ji Junhao, Wen Wenying, et al. Understanding visual privacy protection: A generalized framework with an instance on facial privacy[J/OL]. IEEE Transactions on Information Forensics and Security, 2024[2024-2-10]. https://ieeexplore.ieee.org/abstract/document/10500471

[103] Zhang Junwu, Ye Mang, Yang Yao. Learnable privacy-preserving anonymization for pedestrian images[C]//Proc of the 30th ACM Int Conf on Multimedia. New York: ACM 2022: 7300−7308

[104] Gu Xiuye, Luo Weixin, Ryoo M S, et al. Password-conditioned anonymization and deanonymization with face identity transformers[C]//Proc of the 2020 European Conf on Computer Vision. Berlin: Springer, 2020: 727−743

[105] Peng Chunlei, Wan Shuang, Miao Zimin, et al. Anonym-Recognizer: Relationship-preserving face anonymization and recognition [C]//Proc of the 3rd Int Workshop on Human-Centric Multimedia Analysis. New York: ACM, 2022: 1−6

[106] Mao Xudong, Li Qing, Xie Haoran, et al. Least squares generative adversarial networks[C]//Proc of the 2017 IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 2794−2802

[107] Pan Yilun, Chen Juncheng, Wu Jaling. A multi-factor combinations enhanced reversible privacy protection system for facial images[C]//Proc of the 2021 IEEE Int Conf on Multimedia and Expo (ICME). Piscataway, NJ: IEEE, 2021: 1−6

[108] Cao Jingyi, Liu Bo, Wen Yunqian, et al. Personalized and invertible face de-identification by disentangled identity information manipulation[C]//Proc of the 2021 IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2021: 3334−3342

[109] Blanz V, Vetter T. A morphable model for the synthesis of 3D faces[M/OL]//Seminal Graphics Papers: Pushing the Boundaries, Volume 2. 2023[2024-01-23]. https://dl.acm.org/doi/pdf/10.1145/311535.311556

[110] Paysan P, Knothe R, Amberg B, et al. A 3D face model for pose and illumination invariant face recognition[C]//Proc of the 6th IEEE Int Conf on Advanced Video and Signal Based Surveillance. Piscataway, NJ: IEEE, 2009: 296−301

[111] Khosla P, Teterwak P, Wang Chen, et al. Supervised contrastive learning[J/OL]. Advances in Neural Information Processing Systems, 2020[2023-12-22]. https://proceedings.neurips.cc/paper_files/paper/2020/file/d89a66c7c80a29b1bdbab0f2a1a94af8-Paper.pdf

[112] Zagoruyko S, Komodakis N. Wide residual networks[J]. arXiv preprint, arXiv: 1605.07146, 2016

[113] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J/OL]. Advances in Neural Information Processing Systems, 2017[2023-12-23]. https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[114] Daněček R, Black M J, Bolkart T. EMOCA: Emotion driven monocular face capture and animation[C]//Proc of the 2022 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 20311−20322

[115] Abdal R, Qin Yipeng, Wonka P. Image2styleGAN: How to embed images into the styleGAN latent space? [C]//Proc of the 2019 IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 4432−4441

[116] Chan E R, Lin C Z, Chan M A, et al. Efficient geometry-aware 3D generative adversarial networks[C]//Proc of the 2022 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 16123−16133

[117] King D E. Dlib-ML: A machine learning toolkit[J]. The Journal of Machine Learning Research, 2009, 10: 1755−1758

[118] Wang Tingchun, Liu Mingyu, Zhu Junyan, et al. Video-to-video synthesis[J]. arXiv preprint, arXiv: 1808.06601, 2018

[119] Choi Y, Choi M, Kim M, et al. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation [C]//Proc of the 2010 IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 8789−8797

[120] Li Jian, Wang Yabiao, Wang Changan, et al. DSFD: Dual shot face detector[C]//Proc of the 2019 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 5060−5069

[121] Neverova N, Novotny D, Szafraniec M, et al. Continuous surface embeddings[J]. Advances in Neural Information Processing Systems, 2020, 33: 17258−17270

[122] He Kaiming, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//Proc of the 2017 IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 2961−2969

[123] Hukkelås H, Smebye M, Mester R, et al. Realistic full-body anonymization with surface-guided GANs[C]//Proc of the 2023 IEEE/CVF Winter Conf on Applications of Computer Vision. Piscataway, NJ: IEEE, 2023: 1430−1440

[124] Dwork C, Roth A. The algorithmic foundations of differential privacy[J]. Foundations and Trends® in Theoretical Computer Science, 2014, 9(4): 211−407

[125] Hong Yang, Peng Bo, Xiao Haiyao, et al. HeadNeRf: A real-time nerf-based parametric head model[C]//Proc of the 2022 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 20374−20384

[126] Chen Xi, Duan Yan, Houthooft R, et al. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets[J/OL]. Advances in Neural Information Processing Systems, 2016[2023-12-22]. https://proceedings.neurips.cc/paper_files/paper/2016/file/7c9d0b1f96aebd7b5eca8c3edaa19ebb-Paper.pdf

[127] Heusel M, Ramsauer H, Unterthiner T, et al. Gans trained by a two time-scale update rule converge to a local Nash equilibrium[J/OL]. Advances in Neural Information Processing Systems, 2017[2023-12-23]. https://proceedings.neurips.cc/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf

[128] Wang Zhou, Bovik A C, Sheikh H R, et al. Image quality assessment: From error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600−612

[129] Ruiz N, Chong E, Rehg J M. Fine-grained head pose estimation without keypoints[C]//Proc of the 2018 IEEE Conf on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE, 2018: 2074−2083

[130] Vemulapalli R, Agarwala A. A compact embedding for facial expression similarity[C]//Proc of the 2019 IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 5683−5692

[131] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]// Proc of the 2005 IEEE Computer Society Conf on Computer Vision and Pattern Recognition (CVPR'05). Piscataway, NJ: IEEE, 2005: 886−893

[132] Najibi M, Samangouei P, Chellappa R, et al. SSH: Single stage headless face detector[C]//Proc of the 2017 IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 4875−4884

[133] Dosovitskiy A, Fischer P, Ilg E, et al. FlowNet: Learning optical flow with convolutional networks[C]//Proc of the 2015 IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2015: 2758−2766

[134] Gou Jianping, Yu Baosheng, Maybank S J, et al. Knowledge distillation: A survey[J]. International Journal of Computer Vision, 2021, 129(6): 1789−1819

[135] Liang Tailin, Glossner J, Wang Lei, et al. Pruning and quantization for deep neural network acceleration: A survey[J]. Neurocomputing, 2021, 461: 370−403

[136] Gray R M, Neuhoff D L. Quantization[J]. IEEE Transactions on Information Theory, 1998, 44(6): 2325−2383



Dai Silong, born in 2001. Master candidate. Student member of CCF. His main research interests include multi-modal learning, time-series analysis, and healthcare applications.

戴思龙, 2001 年生. 硕士研究生. CCF 学生会员. 主要研究方向为多模态学习、时间序列分析、医疗健康应用.



Wang Pengfei, born in 1990. PhD candidate. Student member of CCF. His main research interests include multi-modal learning, time-series analysis, and healthcare applications. (pfwang@stu.ecnu.edu.cn)

王鹏飞, 1990 年生. 博士研究生. CCF 学生会员. 主要研究方向为多模态学习、时间序列分析、医疗健康应用.



Wang Xiaoling, born in 1975. PhD, professor. PhD supervisor. Senior member of CCF. Her main research interests include knowledge graph and intelligent data analysis.

王晓玲, 1975 年生. 博士，教授，博士生导师. CCF 高级会员. 主要研究方向为知识图谱、智能数据分析.