

基于多任务学习和身份约束的生成对抗 网络人脸校正识别方法

黄欣研^{1,2,3,4}, 刘芳^{1,2,3,4}, 鲍骞月^{1,2,3,4}, 李任鹏^{1,2,3,4}, 刘旭^{1,2,3,4}, 李玲玲^{1,2,3,4},
陈璞花^{1,2,3,4}, 刘洋^{1,2,3,4}

(1. 西安电子科技大学人工智能学院, 陕西西安 710071; 2. 智能感知与图像理解教育部重点实验室, 陕西西安 710071;
3. 智能感知与计算国际合作联合研究中心, 陕西西安 710071; 4. 智能感知与计算国际合作联合实验室, 陕西西安 710071)

摘要: 针对 DR-GAN (Disentangled Representation learning-Generative Adversarial Network) 方法在将大偏转角度侧脸图像生成其正脸图像的整个生成过程中, 没有考虑身份类别信息, 从而导致在身份和姿态的解耦中存在真实的侧脸图像与其生成的正脸图像身份一致性较弱的问题, 本文提出了一种基于多任务学习和身份约束的生成对抗网络人脸校正识别方法. 该方法通过借鉴多任务学习机制, 在生成网络的编码器与解码器之间构建了角度姿态分类模块和身份约束识别模块. 这两个模块不但在生成过程中实现了人脸身份和姿态的解耦, 更重要的是在由侧脸生成正脸的过程中加入了人脸身份监督信息. 在训练过程中, 该方法将身份和姿态类别直接作为身份编码特征和姿态编码特征的监督信息, 并通过设计身份特征损失函数来约束侧脸的身份编码特征逼近其正脸的身份编码特征, 实现了侧脸编码特征中身份信息和姿态信息的有效解耦, 使解码器能更准确地生成与原侧脸图像保持身份一致的正脸图像. 在 M²FPA 数据集上, 对不同角度的侧脸图像使用所提方法生成的正脸图像进行识别, 达到了更高的人脸识别准确率. 实验结果表明, 即使在偏转角度较大时, 所提方法仍然能够较好地生成保持身份一致的正脸图像, 显著提升了较大偏转角下人脸识别准确率.

关键词: 多任务学习; 身份约束; 生成对抗网络; 人脸校正; 人脸识别

基金项目: 国家自然科学基金 (No.62076192); 国家自然科学基金重点项目 (No.61836009); 长江学者及大学创新研究团队计划 (No.IRT_15R53); 高等学校学科创新引智计划 (No.B07048); 教育部重点科技创新研究项目; 国家重点研发计划; CAAI 华为 MindSpore 开放基金

中图分类号: TP391.41

文献标识码: A

文章编号: 0372-2112(2023)10-2936-14

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20211352

Multi-task Learning and Identity-constrained Generative Adversarial Network for Face Frontalization and Recognition

HUANG Xin-yan^{1,2,3,4}, LIU Fang^{1,2,3,4}, BAO Qian-yue^{1,2,3,4}, LI Ren-peng^{1,2,3,4}, LIU Xu^{1,2,3,4},
LI Ling-ling^{1,2,3,4}, CHEN Pu-hua^{1,2,3,4}, LIU Yang^{1,2,3,4}

(1. School of Artificial Intelligent, Xidian University, Xi'an, Shaanxi 710071, China;

2. Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an, Shaanxi 710071, China;

3. International Research Center for Intelligent Perception and Computation, Xidian University, Xi'an, Shaanxi 710071, China;

4. Joint International Research Laboratory of Intelligent Perception and Computation, Xidian University, Xi'an, Shaanxi 710071, China)

Abstract: For the DR-GAN (Disentangled Representation learning-Generative Adversarial Network), the identity information is not considered in the whole process of generating frontal faces from non-frontal faces with large pose variations. It results in the weak identity consistency between non-frontal faces and the generated frontal faces for disentangling pose from identity. This paper proposes a multi-task learning and identity-constrained generative adversarial network for face frontalization and recognition. Based on the multi-task learning mechanism, a pose classification module and an identity constraint recognition module are constructed between the encoder and decoder of the generative network. These two

modules consider the disentangling of face identity and pose in the generating process. More importantly, face identity supervision information is added in the process of generating faces from non-frontal faces. In the process of training, identity and pose categories are directly used as the supervision information for learning identity coding features and pose coding features. The identity feature loss function is designed to constrain the identity coding features of the non-frontal faces to approximate the identity coding features of the frontal faces. The effective disentangling of identity and pose information in the non-frontal coding feature is realized. The decoder can more accurately generate a frontal face consistent with the non-frontal face. On the M²FPA dataset, the frontal faces generated from the non-frontal faces with different poses by the proposed method are used to recognize, achieving a higher face recognition accuracy. The experimental results show that even when the pose variations are large, the proposed method can still generate a frontal face with a consistent identity, significantly improving face recognition accuracy under large pose variations.

Key words: multi-task learning; identity constraint; generative adversarial network; face frontalization; face recognition

Foundation Item(s): National Natural Science Foundation of China (No.62076192); State Key Program of National Natural Science Foundation of China (No.61836009); Program for Cheung Kong Scholars and Innovative Research Team in University (No.IRT_15R53); Fund for Foreign Scholars in University Research and Teaching Programs (No.B07048); Key Scientific Technological Innovation Research Project by Ministry of Education; National Key Research and Development Program of China; CAAI Huawei MindSpore Open Fund

1 引言

人脸识别^[1-6]作为一种基于人的面部特征信息进行身份识别的生物识别技术,被广泛应用于军事、金融、公安和日常生活等领域.由于神经网络表现出了优异的图像特征提取能力^[7-11],基于神经网络的人脸识别方法在一些标准数据集上的人脸识别准确率已经超过了人类^[12].然而,人脸识别^[13-15]本质上是一种识别不合作对象(即动态人脸)的被动生物识别技术,无约束环境下获取到的人脸通常在角度姿态上存在较大的变化,当人脸偏转角超过45°时,特别是当人脸偏转角接近90°时,不同身份之间的人脸在特征空间中相距较近,而同一身份不同角度的人脸在特征空间中相距较远,导致人脸识别准确率迅速下降^[1,4-6].识别在任意姿态下拍摄的人脸图像,对机场、火车站、银行和其他公共场所的生物识别安全控制系统非常重要,因此如何提升在大角度姿态下的人脸识别准确率成为关键问题.

目前,通过人脸校正可以有效提升大角度姿态下的人脸识别准确率.人脸校正再识别的一般流程为给一张侧脸图像,生成其对应身份的正脸图像,然后再与人脸底库中的正脸图像进行验证比对.当人脸偏转角较小时,姿态对人脸识别准确率的影响较小;然而当人脸偏转角较大时,人脸识别准确率迅速下降.如何将大偏转角下的侧脸图像生成尽可能逼真同时保持身份一致的正脸图像,成为目前亟待解决的问题.

近年来,生成对抗网络展现出强大的生成能力^[16-22],人们对基于生成对抗网络的人脸校正方法^[1,23]展开研究,将侧脸图像通过生成器生成正脸图像,再对生成的正脸图像送入判别器中进行判别.如图1(a)中

所示,Tran等人^[1]提出了的解耦表示学习生成对抗网络(Disentangled Representation learning-Generative Adversarial Network, DR-GAN)用于姿态不变人脸识别(Pose-Invariant Face Recognition, PIFR),构造了带有编码器 G_{enc} 和解码器 G_{dec} 结构的人脸旋转器 G ,编码器 G_{enc} 的输入是任意姿态的人脸图像 I ,输出是身份表示 f ,将其与姿态编码 c 和随机噪声 z 一起送入解码器 G_{dec} ,输出目标姿态的合成人脸,训练 D 在区分真实图像和合成图像的同时预测人脸的 N_d 个身份和 N_p 个姿态,希望 D 旋转后的脸与输入的真实脸具有相同的身份.然而DR-GAN仍然存在以下问题:(1)在大偏转角度侧脸图像生成其正脸图像的整个生成过程中,仅对判别器使用监督信息,未对生成器直接使用身份类别信息进行监督,使真实的侧脸图像与其生成的正脸图像身份一致性较弱;(2)在生成过程中,与编码器学习得到的身份特征 f 一起送入解码器的姿态编码 c 是使用目标姿态位置为1的one-hot向量,并非通过学习获得的姿态编码特征,未学习到好的姿态不变的身份特征 f ,难以较好地将身份表示和姿态变化解耦.

Peng等人^[23]提出了基于深度神经网络(Deep Neural Networks, DNNs)的识别、姿态估计和landmark定位的联合学习框架.通过联合优化这三个任务,学习到包含身份和非身份信息的丰富特征嵌入,并在此基础上通过特征自重构和交叉重构,将身份特征和非身份特征分离,从而获得姿态不变的身份特征.然而,最终通过重构获得的姿态不变特征,并没有直接受到身份类别的约束,在大角度姿态下学到的身份特征仍未能很好地与输入图像保持身份一致性.

受到Tran等人^[1]提出的解耦表示学习生成对抗网

络和Peng等人^[23]提出的多任务学习机制的启发,为了获得更好的姿态不变身份特征,本文针对使用现有基于生成对抗网络的人脸校正方法,在由侧脸图像生成正脸图像的过程中,因为没有考虑身份信息约束,生成的正脸图像与真实正脸图像间的一致性变弱,大偏转角下的人脸识别准确率降低的问题,提出了基于多任务学习和身份约束的生成对抗网络人脸校正识别方法.在生成网络的编码器和解码器之间设计了身份约束识别模块、角度姿态分类模块和噪声特征编码模块(图1(b)).编码器 G_{enc} 的输入是任意姿态的人脸图像 I ,将编码器得到的初始编码特征 f 分别经过身份特征提取网络 M_{id} 、姿态特征提取网络 M_{po} 和噪声特征提取网络 M_z 获得身份编码特征 f_{id} 、姿态编码特征 f_{po} 和噪声编码特征 f_z ,使用身份识别网络 C_{id} 的 N_d 个身份类别约束身份编码特征 f_{id} 与输入图像的身份一致性,使用姿态分类网络 C_{po} 的 N_p 个偏转角类别约束姿态编码特征 f_{po} ,从而将姿态编码特征、噪声编码特征(如光照)与身份编码特征解耦,并将解耦后得到的身份编码特征 f_{id} 送入解码器 G_{dec} ,输出目标姿态的合成人脸,采用全局和局部(眼睛、鼻子、嘴)的判别器组 D_i ($i=1,2,3,4$)对全局和局部是真实图像和合成图像进行区分.由于在身份约束识别模块中引入了一个由侧脸身份编码逼近其正脸身份编码的身份约束识别函数,该模块能够拉近同一身份下不同偏转角输入图像的身份编码特征之间的距离.该校正方法实现了侧脸编码特征中身份信息和姿态信息的解耦,最终通过解码器能够较好地生成保持身份一致性的正脸图像,有效地提升大角度姿态下的人脸识别准确率.

2 相关工作

2.1 传统的人脸校正方法

传统的人脸校正方法根据基于的图像域可以分为二维和三维的方法.早期的人脸校正方法通常基于图像块形变的策略.Gao等人^[24]提出仿射形变,对原始图像和目标图像的每对三角形区域进行变换.Bookstei等人^[25]基于薄板样条线进行变换.Taigman等人^[26]借助三维人脸模型推断多姿态人脸对应的正面化姿态的关键点位置,将二维图像关键点与三维模型关键点投影位置之间的残差添加到正面化姿态的均匀形状中作为补偿.Berg等人^[27]根据每种姿态的通用关键点坐标进行三角定位,然后通过分段仿射扭曲将多姿态人脸图像转换为正面姿态的均匀形状.Ashraf等人^[28]提出使用图像块重建正面化人脸的策略,对于侧脸图像的每个图像块,采用堆栈流的方法生成正脸的相应图像块.

上述基于图像块进行正面化生成的方式无法处理

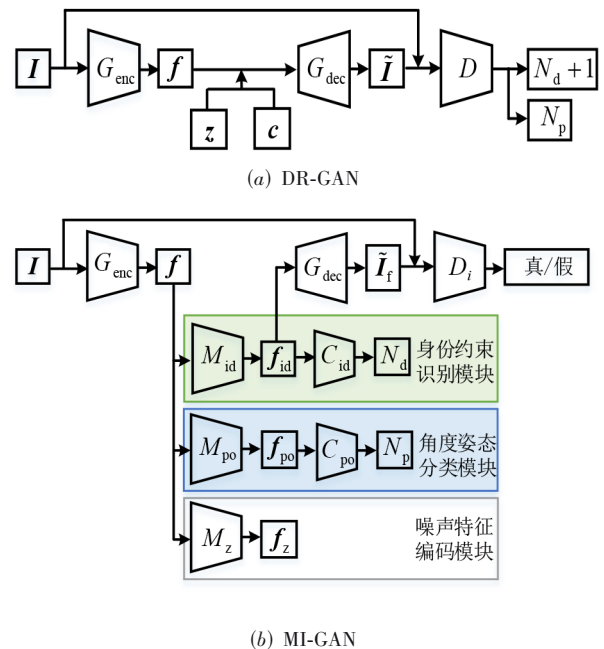


图1 DR-GAN和MI-GAN的架构对比

图像块中的非线性扭曲,Beymer等人^[29]提出基于光流对不同姿态的图像建立密集的逐像素语义对应关系.Hassner等人^[30]通过计算二维人脸图像中的关键点与三维人脸模型关键点之间的映射关系,将二维图像的像素点映射到三维模型,生成对应的正面化人脸图像.Blanz等人^[31]提出3DMM可形变模型方法,通过可形变的3D模型和计算机图形技术来模拟图像重建过程.Romdhani等人^[32]提出一种新的拟合算法,采用像素强度、边缘和镜面高光等多个特征,在给定多个特征上利用贝叶斯框架最大化形状和纹理参数的后验概率.

2.2 基于深度学习的人脸校正方法

随着深度学习的快速发展,基于卷积神经网络和生成对抗网络衍生出一系列方法.这些方法可以分为两类.一类是学习一个从侧脸图像生成对应的正脸图像的模型,再对生成的正脸图像进行识别,属于像素级的人脸校正.另一类是将侧脸图像中的姿态和身份信息互相解耦,来学习姿态不变人脸特征表示,属于特征级的人脸校正.

基于神经网络学习非线性变换的能力,相关方法^[33-35]首次尝试提取姿态鲁棒的特征,然后重建正面化人脸图像.Yim等人^[33]使用卷积神经网络首先提取姿态鲁棒的特征,然后使用全连接层重建人脸.Zhang等人^[34]采用一个隐层的自编码器生成正脸图像.Kan等人^[35]使用栈式自编码器逐步生成正脸图像,降低了每个自编码器的生成难度.

借助生成对抗网络强大的生成能力,一系列方法诞生了.TP-GAN^[36]基于多路生成器生成正脸图像,不

仅能够恢复全局信息,同时也较好地恢复局部纹理细节. PIM^[37]在生成保持身份一致性的正面化人脸图像的同时,学习了姿态不变身份特征. FNM^[38]采用无监督的学习方式,使用不同身份的侧脸和正脸训练模型. DA-GAN^[39]基于人脸解析网络提取生成正脸图像的不同区域进行分别判别. CAPG-GAN^[40]使用关键点 heatmap 来代表人脸的全局姿态信息,指导生成对应姿态的人脸图像. Cao 等人^[41]提出 DREAM-Block, 基于残差的思想学习侧脸到正脸的特征转换,通过头部姿态估计器对输入侧脸进行人脸偏转角的估计. DR-GAN^[1]在生成器中嵌入姿态信息和随机噪声,从而将身份特征与其他面部变化(如姿态)分离,使编码器提取到输入侧脸图像的姿态不变特征. Peng 等人^[23]基于多任务学习机制学习身份特征和非身份特征,通过特征重建使侧脸的身份特征逼近其正脸的身份特征,来获取姿态不变特征.

3 基于多任务学习和身份约束的人脸校正

为了在较大偏转角下也能够生成保持身份一致性

的正脸图像,本文在此基础上提出了基于多任务学习和身份约束的生成对抗网络(Multi-task learning and Identity-constrained Generative Adversarial Network, MI-GAN)、以下分别对 MI-GAN 的功能模块结构、生成网络、对抗网络和总的损失函数进行具体介绍.

3.1 MI-GAN 功能模块结构

MI-GAN 功能模块结构示意图如图 2 所示,它由两部分组成,分别是生成网络和对抗网络. 受多任务学习机制启发,在生成网络的编码器和解码器之间构建了身份约束识别模块、角度姿态分类模块和噪声特征编码模块,分别使用身份类别和姿态类别作为学习姿态编码特征和身份编码特征的监督信息,并引入身份特征损失函数约束侧脸的身份编码特征逼近其正脸的身份编码特征,实现侧脸编码特征中身份信息和姿态信息的解耦,使侧脸的编码特征通过解码器能够生成保持身份一致性的正脸图像. MI-GAN 通过对输入至解码器的编码特征引入身份约束,从而在较大偏转角下也能够生成保持身份一致性的正脸图像.

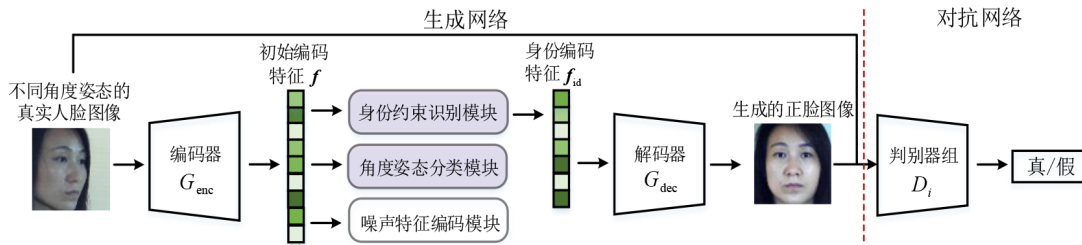


图2 MI-GAN 功能模块结构示意图

3.2 生成网络

本节主要对生成网络的结构以及在其中构建的身份约束识别模块、角度姿态分类模块进行具体介绍.

3.2.1 生成网络结构

如图 3 中所示,生成网络包括编码器、解码器、身份约束识别模块、角度姿态分类模块和噪声特征编码模块.

编码器由多层卷积层和一层全连接层构成,输入的人脸图像大小为 $128 \times 128 \times 3$,与文献[36]中相似,每做一次卷积操作,就经过一个残差块. 通过残差块来加深网络的深度,提升网络的学习能力. 使用激活函数 ReLU,它可以造就网络的稀疏性,缓解过拟合问题,同时使用批归一化来加速网络的收敛. 最后一层全连接层输出 512 维的初始编码特征向量 f . 编码器的作用是提取不同角度姿态的输入图像的初始编码特征向量 f .

解码器的输入为 256 维的身份编码特征 f_{id} ,解码器由连续的反卷积层模块构成,与文献[36]中相似,每做一次反卷积操作,同样经过一个残差块,用来加深网络

的深度,提升网络的学习能力. 解码器中最后一层不使用激活函数以及批归一化,其余层都使用激活函数 ReLU 和批归一化,解码器最终生成大小为 $128 \times 128 \times 3$ 的正脸图像.

通过身份约束识别模块和角度姿态分类模块分别得到身份编码特征 f_{id} 和姿态编码特征 f_{po} ,实现侧脸编码特征中身份信息和姿态信息的解耦,身份约束识别模块和角度姿态分类模块分别在第 3.2.2 节和 3.2.3 节中详细介绍. 解码器对身份编码特征 f_{id} 进行反卷积得到正面化人脸图像.

噪声特征编码模块可对身份和姿态外的其他人脸外观变化(如光照变化)建模,将编码器得到的初始编码特征向量 f 送入一层输出节点数为 256 全连接层,得到 256 维的噪声编码特征.

3.2.2 身份约束识别模块

如图 4 所示,身份约束识别模块的输入为编码器得到的 512 维初始编码特征向量 f ,根据输入图像的身份类别 N_i 进行分类,使用身份分类损失函数 L_{cls_id} 拉近

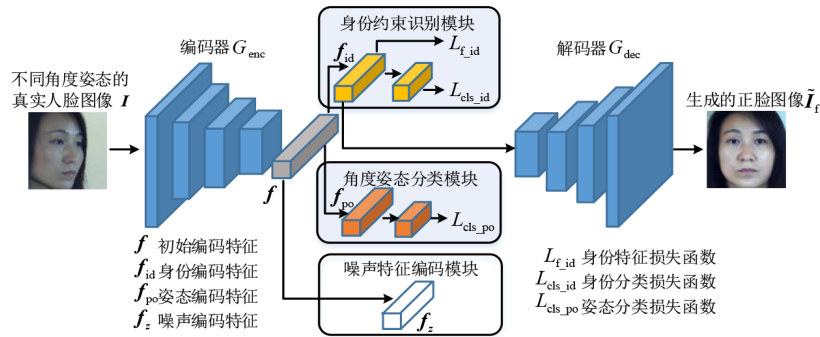


图3 生成网络结构示意图

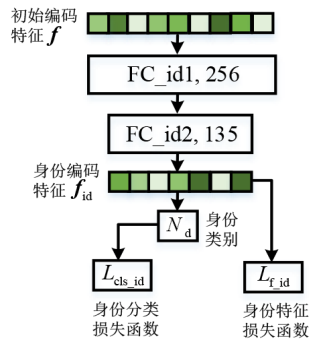


图4 身份约束识别模块结构示意图

同一身份下不同偏转角输入图像的身份编码特征之间的距离,同时引入身份特征损失函数 L_{f_id} 约束侧脸的身份编码特征逼近其正脸的身份编码特征.正脸的身份编码特征由正脸图像生成自身正脸图像所获得,因而侧脸的身份编码特征不仅包含身份信息,同时接近其真正脸的身份编码特征,通过解码器能够生成保持身份一致性的正脸图像.身份约束识别模块同样采取两层全连接层,第一层全连接层的输出节点数为256,采用激活函数ReLU,同时使用批归一化,得到256维的身份编码特征,使用身份特征损失函数约束侧脸的身份编码特征逼近其正脸的身份编码特征.由于训练时采取了135个身份的不同偏转角的人脸图像进行训练,设置第二层全连接层的输出节点数为135,对输入人脸图像的身份进行分类.为了防止过拟合,第二层全连接层采用Dropout,随机失活比率设置为0.3.

身份分类损失函数 L_{cls_id} 如式(1)所示:

$$L_{cls_id} = -\frac{1}{N} \sum_{k=1}^N \log \frac{e^{W_{id,j}^T f_{id,k} + b_{id,j}}}{\sum_{j=1}^{N_d} e^{W_{id,j}^T f_{id,k} + b_{id,j}}} \quad (1)$$

L_{cls_id} 采用softmax交叉熵损失函数, N 为训练样本数量, $f_{id,k}$ 表示第 k 个样本的身份编码特征, $W_{id,j}$ 和 $b_{id,j}$ 分别表示身份分类模块 M_{id} 最后一层全连接层权重 W_{id} 的第 j 列和偏置 b_{id} 的第 j 列, N_d 表示身份类别数,共有135种身份类别数,通过 L_{cls_id} 来约束身份约束识别模块 M_{id} 学习身份编码特征.

身份特征损失函数 L_{f_id} 如式(2)所示:

$$L_{f_id} = \|f_{id}^p - f_{id}^f\|_2^2 \quad (2)$$

L_{f_id} 采用均方差损失函数,其中 f_{id}^p 和 f_{id}^f 分别表示侧脸的身份编码特征和其正脸的身份编码特征.通过约束侧脸图像的身份编码特征尽可能接近其正脸图像的身份编码特征,使侧脸的身份编码特征通过解码器能够生成保持身份一致性的正脸图像.

3.2.3 角度姿态分类模块

如图5所示,角度姿态分类模块的输入为编码器得到的512维初始编码特征向量 f ,根据输入图像的人脸偏转角的类别 N_p 进行分类,使用姿态分类损失函数 L_{cls_po} 使在人脸偏转角相同时,即使人物身份不同,同样具有相似的全局拓扑信息.角度姿态分类模块包含了两层全连接层.第一层全连接层的输出节点数为256,得到256维的姿态编码特征,采用激活函数ReLU,同时使用批归一化.由于训练时有13种偏转角的人脸图像,设置第二层全连接层的输出节点数为13,为了防止过拟合,第二层全连接层采用Dropout,随机失活比率设置为0.3,对输入人脸图像的偏转角进行分类.

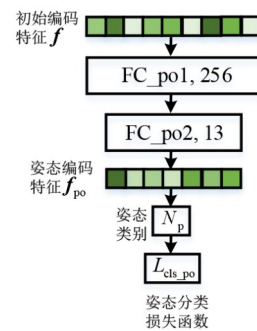


图5 角度姿态分类模块结构示意图

姿态分类损失函数 L_{cls_po} 如式(3)所示:

$$L_{cls_po} = -\frac{1}{N} \sum_{k=1}^N \log \frac{e^{W_{po,j}^T f_{po,k} + b_{po,j}}}{\sum_{q=1}^{N_p} e^{W_{po,q}^T f_{po,k} + b_{po,q}}} \quad (3)$$

L_{cls_po} 采用softmax交叉熵损失函数, $f_{po,k}$ 表示第 k 个样本

的姿态编码特征, $\mathbf{W}_{po,q}$ 和 $\mathbf{b}_{po,q}$ 分别表示姿态分类模块 M_{po} 最后一层全连接层权重 \mathbf{W}_{po} 的第 q 列和偏置 \mathbf{b}_{po} 的第 q 列, N_p 表示姿态类别数, 共有 13 种姿态类别数, 通过 L_{cls_po} 来约束角度姿态分类模块 M_{po} 学习姿态编码特征.

3.3 对抗网络

对抗网络用于判别输入图像是来自真实的正脸图像 I_f 还是生成的正脸图 \tilde{I}_f . 模型训练初期时生成网络生成能力还不够强时, 对抗网络能够较好的判别 I_f 和 \tilde{I}_f , 此时会提高生成网络的生成能力, 当生成能力提升时, 生成图像 \tilde{I}_f 的数据分布接近真实图像 I_f 数据分布时, 对抗网络此时无法较好地地区分 I_f 和 \tilde{I}_f , 相应地会提升判别能力. 生成网络和对抗网络两者相互促进, 最终生成网络生成尽可能逼真的正脸图像 \tilde{I}_f , 对抗网络也无法区分输入图像是来自真实的正脸图像 I_f 还是生成的正脸图像 \tilde{I}_f .

为了促进生成网络生成的正脸图像不仅在全局信息上有较好的恢复, 同时在局部纹理细节上与真实正脸图像保持一致性, 对抗网络采取类似于 FNM^[38] 中的人脸注意力判别器组, 包含 4 个判别器 D_i ($i=1, 2, 3, 4$), 分别对全局人脸图像、双眼、鼻子和嘴巴进行判别. 使用 RetinaFace^[42] 对真实正脸图像进行人脸关键点检测得到左眼中心、右眼中心、鼻尖以及嘴角共 5 个关键点坐标, 基于这 5 个关键点坐标对生成的正脸图像以及真实的正脸图像进行局部区域分割, 分别得到图像大小为 85×40 的双眼区域、 40×32 的鼻子区域以及 48×32 的嘴巴区域. D_i 由连续的卷积层和两层全连接层构成, 最后一层全连接层的输出节点数为 1, 用于判别输入图像的真假. 使用人脸注意力判别器组 D_i 在关注生成图像全局信息的同时对局部区域进行判别, 使生成网络生成的正脸图像在恢复全局信息的基础上同时保持局部区域与真实正脸图像的一致性. 对抗损失函数 L_{adv} 如式(4)所示:

$$L_{adv} = \sum_{i=1}^4 \log D_i(I_f^i) + \log(1 - D_i(\tilde{I}_f^i)) \quad (4)$$

其中 I_f^i 和 \tilde{I}_f^i 分别表示真实的正脸图像和生成的正脸图像对应判别器 D_i 的不同局部区域. $D_i(I)$ 的值反映了判别器 D_i 判定输入图像区域 I 为真实正脸图像对应区域的可能性. 当 $D_i(I)$ 的值越大时, 说明判别器判定输入图像区域 I_f^i 为真实正脸图像对应区域的可能性越大. 所以当训练判别器时, 最大化 L_{adv} 使判别器 D_i 能够判别出输入图像区域是来自生成的正脸图像还是真实的正脸图像, 来提升判别器的判别能力. 当训练生成器时, 最小化 L_{adv} 使判别器 D_i 无法区分输入图像区域是来自生成的正脸图像还是真实的正脸图像, 来提升生成器的生成能力. 以这样的方式交替迭代的训练判别器和生成器, 使生成器能够生成接近真实数据分布的正脸图像. 使用人脸注意力判别组 D_i 使生成的正脸图像

在恢复全局信息的同时也能够较好的恢复局部纹理细节.

3.4 总的损失函数

除了式(1)~(4)中的身份分类损失函数 L_{cls_id} 、身份特征损失函数 L_{f_id} 、姿态分类损失函数 L_{cls_po} 和对抗损失函数 L_{adv} 外, 训练 MI-GAN 还使用了其他损失函数^[36,39], 包括像素损失函数 L_{pixel} 、对称损失函数 L_{sym} 、正则项 L_{tv} 和特征约束损失函数 L_f , 这些损失函数之和为 MI-GAN 的总的损失函数.

像素损失函数 L_{pixel} 的公式如式(5)所示:

$$L_{pixel} = \frac{1}{W \times H \times C} \sum_{x=1}^W \sum_{y=1}^H \sum_{z=1}^C \left| \tilde{I}_{f_{x,y,z}} - I_{f_{x,y,z}} \right| \quad (5)$$

其中 W , H 和 C 分别代表正脸图像的宽、高和通道数. \tilde{I}_f 和 I_f 分别表示生成的正脸图像和真实的正脸图像. 通过最小化像素损失函数 L_{pixel} 使生成的正脸图像尽可能接近真实的正脸图像.

对称损失函数的 L_{sym} 如式(6)所示:

$$L_{sym} = \frac{1}{W/2 \times H \times C} \sum_{x=1}^{W/2} \sum_{y=1}^H \sum_{z=1}^C \left| \tilde{I}_{f_{x,y,z}} - \tilde{I}_{f_{W-x,y,z}} \right| \quad (6)$$

通常对称是人物正脸图像的一个固有特征, 对生成正脸图像添加对称损失这个约束将这种固有特征作为一种先验知识嵌入模型当中, 可以较好地解决大角度姿态造成自遮挡的问题, 同时能够加速模型的收敛.

正则项 L_{tv} 如式(7)所示:

$$L_{tv} = \sum_{x=1}^W \sum_{y=1}^H \sum_{z=1}^C \left| \tilde{I}_{f_{x,y,z}} - \tilde{I}_{f_{x,y,z}} \right| + \left| \left| \tilde{I}_{f_{x,y,z}} - \tilde{I}_{f_{x,y,z}} \right| \right| \quad (7)$$

通常基于生成对抗网络模型生成的图像会存在许多尖峰伪影^[43], 会在降低生成图像质量的同时降低后续的人脸识别准确率. 使用 L_{tv} 正则项^[44] 来解决这个问题.

为了保持生成的正脸图像和真实正脸图像的特征一致性, 在生成网络中加入了特征一致性约束模块, 通过约束特征提取网络 B 提取的生成的正脸特征和真实的正脸进行特征一致性约束, 使得生成的正脸图像 \tilde{I}_f 和真实的正脸图像 I_f 在特征空间中尽可能接近, 身份信息保持一致. 人脸特征提取网络 B 采用 MobileFaceNet^[45], MobileFaceNet 网络参数量较少, 同时在正脸图像上能够达到比较高的人脸识别准确率, 当人脸偏转角超过 45° 时, 准确率迅速下降, 适用于人脸校正的任务. 人脸特征提取网络 B 的输入为全局正脸图像, 最后输出为 512 维的特征向量. 特征约束损失函数 L_f 如式(8)所示:

$$L_f = \left\| B^{fc}(\tilde{I}_f) - B^{fc}(I_f) \right\|_2^2 \quad (8)$$

其中 B^{fc} 表示人脸特征提取网络 B 倒数第二层全连接层, 本文使用 MobileFaceNet 作为人脸特征提取网络. L_f

约束生成的正脸图像 \tilde{I}_f 和真实的正脸图像 I_f 在特征空间中尽可能接近,保持生成的正脸图像与真实的正脸图像身份一致.

最终的损失函数由以上介绍的损失函数加权得到:

$$L_{\text{total}} = \lambda_1 L_{\text{pixel}} + \lambda_2 L_{\text{cls_po}} + \lambda_3 L_{\text{cls_id}} + \lambda_4 L_{\tilde{I}_f \text{ id}} + \lambda_5 L_{\text{adv}} + \lambda_6 L_f + \lambda_7 L_{\text{sym}} + \lambda_8 L_{\text{tv}} \quad (9)$$

本文使用批随机梯度下降方法对生成对抗网络进行了交替训练,使用式(10)更新对抗网络,使用式(11)更新生成网络.

$$\nabla_{\theta_D} [-(\log D_i(I_f^i) + \log(1 - D_i(\tilde{I}_f^i)))] \quad (10)$$

$$\nabla_{\theta_G} [\lambda_5 \log(1 - D_i(\tilde{I}_f^i)) + \lambda_1 L_{\text{pixel}} + \lambda_2 L_{\text{cls_po}} + \lambda_3 L_{\text{cls_id}} + \lambda_4 L_{\tilde{I}_f \text{ id}} + \lambda_6 L_f + \lambda_7 L_{\text{sym}} + \lambda_8 L_{\text{tv}}] \quad (11)$$

式(9)中的8个损失项在模型的优化过程中具体训练步骤见算法1.

4 实验与分析

4.1 实验设置

4.1.1 实验数据集设置

本文是基于数据集 M²FPA^[46] 进行模型的训练和测试. M²FPA 数据集包含 229 个人物共 397544 张人脸

算法1 训练生成对抗网络

输入: 侧脸-正脸图像对, 训练批次大小 $n=32$, 迭代次数 $t=200$.

输出: 训练后的生成对抗网络.

1. 侧脸-正脸图像对中随机采样一个批次(n 个样本);
2. 使用公式(10)通过批随机梯度下降的方法更新对抗网络;
3. 使用公式(11)通过批随机梯度下降的方法更新生成网络;
4. 重复执行步骤1~3,直到达到最大迭代次数 t ;
5. 得到训练后的生成对抗网络.

RGB 图像. 对每个人物, 包含了 62 种不同的姿态、4 种不同的属性和 7 种不同的光照条件. 其中 4 种不同的属性分别为表情为开心, 戴眼镜, 表情为惊讶, 正常表情且不戴眼镜. 62 种不同的姿态包含了人脸偏转角 yaw 和人脸俯仰角 pitch 的不同组合, 由于本文只针对人脸偏转角 yaw, 所以从中挑选人脸俯仰角 Pitch=0° 而包含不同人脸偏转角 yaw 的 13 种姿态, 包含了 ±90°、±75°、±60°、±45°、±30°、±15° 和 0°. 最终采用每个人物的 13 种不同的姿态、4 种不同的属性和 7 种不同的光照条件进行实验分析. 如图 6 所示, 展示了数据集 M²FPA 不同属性下的不同偏转角的人脸图像, 从上至下人脸属性分别为表情开心、正常表情且不戴眼镜、表情为惊讶和戴眼镜, 每一行从左到右人脸偏转角由 -90° 变化至 +90°.



图6 数据集 M²FPA 中的部分人脸图像

实验中使用 135 个人物作为训练集, 共包含 135×13×4×7=49 140 张人脸 RGB 图像, 59 个人物作为测试集, 保证训练集和测试集不重叠. 对测试集中的每个人物, 选出人脸偏转角为 0° 并且表情和光照正常的人脸图像作为人脸底库, 共包含 59 张图像. 剩余的 59×12×4×7=19 824 张人脸图像作为测试集. 测试人脸识别准确率时, 对测试集中的所有图像进行人脸校正得到正面化人脸图像, 再通过人脸特征提取网络 MobileFaceNet^[45] 提取人脸特征向量. 实验中采用两个特征向量之间的余弦距离作为两张人脸图像之间相似度的评估标准, 两个特征向量之间的余弦距离越小, 说明两张人脸图像之间相似程度越高. 对于测试集中的每一张人脸, 人脸校正后与人脸底库中的每张图像进行特征

向量相似度计算, 如果匹配到的相似度最大的图像的身份与测试图像身份一致, 则说明识别正确. 对于多姿态人脸识别任务来说, 不同人脸偏转角的人脸识别准确率差别较大, 所以实验中按人脸偏转角对测试集中的人脸图像进行分类, 共分为 6 组, 分别对应于人脸偏转角为 ±15°, ±30°, ±45°, ±60°, ±75°, ±90°, 对于每一种偏转角各有 1652 张待测试图像. 实验时对同一人脸偏转角的正负方向的识别准确率取平均值. 实验中准确率的定义为匹配到相似度最大的人脸图像身份与测试图像身份一致的图像数量占总测试图像数量的百分比.

4.1.2 实验环境设置

实验所使用的计算机为 12 核 24 线程的 Intel Xeon E5-2650 CPU, 其主频为 2.20 GHz, 内存为 512 GB, GPU

是两块 GeForce RTX 2080 显卡,每块显卡的显存为 12 GB. 采用的编程语言是 Python,深度学习框架采用的是 Pytorch,程序均在 Ubuntu 18.04 系统下运行.

4.1.3 实验参数设置

原始 M²FPA 数据集的图像大小为 512×512×3,为了与人脸校正相关方法^[36,40,46]中采取的输入图像大小 128×128×3 保持一致,实验中将原始图像大小缩小至 128×128×3. 同时人脸特征提取网络使用 MobileFaceNet, MobileFaceNet 在 MS-Celeb-1M^[47]数据集上预训练,并基于 M²FPA 正脸集进行微调,在整个网络训练过程中固定参数不变. 网络初始学习率为 0.000 1,每训练一代学习率线性衰减,直至学习率为 0. 训练总代数为 200 代,训练批次大小为 32,采取 Adam^[48]优化策略,网络训练时参数更新更加稳定. 损失函数权重 $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7$ 和 λ_8 分别设置为 10, 1, 1, 10, 0.1, 100, 0.3 和 0.000 1.

表 1 为 MI-GAN 中编码器 G_{enc} 、角度姿态分类模块 M_{po} 、身份约束识别模块 M_{id} 、解码器 G_{dec} 和判别器 D 的具体网络层级结构. 对于编码器 G_{enc} 的每一个卷积模块包含了一次卷积操作和一个残差块,残差块由两层卷积层构成,卷积核大小为 3×3,步长为 1,输出通道数与输入通道数保持一致. 同样对于解码器 G_{dec} 的每一个反卷积模块包含一次反卷积操作和一个残差块. 为了验证本文方法的有效性,分别在编码器 G_{enc} 的卷积模块 Conv5 和全连接层 FC 后嵌入角度姿态分类模块 M_{po} 和身份约束识别模块 M_{id} ,进行对比实验分析.

4.2 MI-GAN 实验结果

在本节中,对所提 MI-GAN 方法在 M²FPA 数据集上进行实验验证,定量分析了 MI-GAN 方法的有效性,并给出了对部分侧脸图像使用 MI-GAN 生成正脸图像的实验结果.

4.2.1 MI-GAN 的对比实验

在数据集 M²FPA 上,将 MI-GAN 与相关方法 MobileFaceNet^[45], DR-GAN^[1], TP-GAN^[36], CAPG-GAN^[40] 和 PPGF^[49] 的人脸识别准确率进行对比,所有方法的人脸识别准确率是计算 5 次后取平均值,详细的实验结果如表 2 中所示. MobileFaceNet^[45] 是一种常见的用于人脸识别的人脸特征提取网络. DR-GAN^[1] 用学习到的姿

表 1 MI-GAN 网络参数设置

编码器 G_{enc} 的具体网络层级结构				
层数	卷积核参数	步长	归一化	激活函数
Conv1	7×7×64	1	批量归一化	ReLU
Conv2	5×5×64	2	批量归一化	ReLU
Conv3	3×3×128	2	批量归一化	ReLU
Conv4	3×3×256	2	批量归一化	ReLU
Conv5	3×3×512	2	批量归一化	ReLU
FC	节点数为 512		批量归一化	ReLU
角度姿态分类模块 M_{po} 的具体网络层级结构				
FC_po1	节点数为 256		批量归一化	ReLU
FC_po2	节点数为 13		批量归一化	ReLU
身份约束识别模块 M_{id} 的具体网络层级结构				
FC_id1	节点数为 256		批量归一化	ReLU
FC_id2	节点数为 135		批量归一化	ReLU
噪声特征编码模块 M_z 的具体网络层级结构				
FC_z1	节点数为 256		批量归一化	ReLU
解码器 G_{dec} 的具体网络层级结构				
Deconv1	8×8×64		批量归一化	ReLU
Deconv2	3×3×32	4	批量归一化	ReLU
Deconv3	3×3×16	2	批量归一化	ReLU
Deconv4	3×3×8	2	批量归一化	ReLU
Conv1	3×3×3	1	批量归一化	ReLU
判别器 D_i 的具体网络层级结构				
Conv1	3×3×64	2	批量归一化	ReLU
Conv2	3×3×128	2	批量归一化	ReLU
Conv3	3×3×256	2	批量归一化	ReLU
Conv4	3×3×512	2	批量归一化	ReLU
Conv5	3×3×512	2	批量归一化	ReLU
FC1	节点数为 1024		批量归一化	ReLU
FC2	节点数为 1			
判别器 $D_i (i = 2,3,4)$ 的具体网络层级结构				
Conv1	3×3×64	2	批量归一化	ReLU
Conv2	3×3×128	2	批量归一化	ReLU
Conv3	3×3×256	2	批量归一化	ReLU
Conv4	3×3×512	2	批量归一化	ReLU
FC1	节点数为 1024		批量归一化	ReLU
FC2	节点数为 1			

表 2 MI-GAN 与对比方法的人脸识别准确率对比

单位: %

模型	±15°	±30°	±45°	±60°	±75°	±90°
MobileFaceNet ^[45]	100.0	95.0	91.5	55.4	22.3	8.3
DR-GAN ^[1]	91.8	87.2	82.9	74.9	64.8	39.2
TP-GAN ^[36]	99.8	98.9	96.5	79.8	65.5	40.5
CAPG-GAN ^[40]	99.8	98.8	96.4	76.7	62.3	41.8
PPFG ^[49]	99.5	98.2	96.7	78.1	66.4	42.7
MI-GAN	100.0	98.9	96.8	80.4	67.1	44.3

态不变身份特征来定量评估人脸识别性能. TP-GAN^[36]基于多路生成器生成正脸图像,全局生成器用于生成正脸的全局表现,4个局部生成器分别对侧脸图像的4个局部区域(左眼、右眼、鼻子和嘴巴)进行校正,基于正脸关键点位置匹配,全局生成器融合局部校正结果生成正脸图像. CAPG-GAN^[40]使用关键点 heatmap 代表人脸的全局姿态信息,指导生成对应姿态的人脸图像. 伪侧脸生成器(Pseudo Profile Facial Generator, PPFG)^[49]在最小化与原侧脸图像像素差异的同时保持与其对应正脸的身份一致性来重建伪侧脸,通过提取伪侧脸特征进行人脸识别. 从表2中可以看出,随着偏转角不断增大,所有方法的准确率都会下降,这是由于在大偏转角下,自遮挡等因素造成侧脸图像语义丢失严重以及纹理形状变化较大,生成保持身份信息不变的正脸图像难度相应增大. 与相关方法对比,MI-GAN提升了不同偏转角的人脸识别准确率,尤其是在较大偏转角下. 当直接使用人脸特征提取网络 MobileFaceNet 对侧脸图像提取特征进行识别时,随着偏转角度的增加,在偏转角达到 60° 时,人脸识别准确率迅速下降. 当偏转角小于 60° 时,MI-GAN的人脸识别准确率优于DR-GAN,与TP-GAN, CAPG-GAN 和 PPFG 的准确率相当. 当偏转角超过 60° 时,MI-GAN的识别性能优于

所有对比方法. 这表明了MI-GAN通过对侧脸编码特征添加身份约束,使生成的正脸图像能够更好地保持身份一致性,提升了较大偏转角下的人脸识别准确率.

4.2 MI-GAN生成图像的实验结果

本小节给出了MI-GAN对不同偏转角和不同人脸属性的部分侧脸图像的人脸校正可视化结果. 首先展示了MI-GAN在数据集M²FPA上对不同偏转角的部分侧脸图像的可视化人脸校正结果,如图7所示. 图中每个偏转角对应两列图像,左侧为输入的侧脸图像,右侧为生成的正脸图像. 由图可见,当人脸偏转角小于 75° 时,生成的正脸图像能够较好地保持身份一致性,在恢复全局信息的同时能够较好地恢复局部纹理细节. 当人脸偏转角超过 75° 时,生成的正脸图像并没有发生明显的形变,也能够较好地恢复全局信息.

图8展示了MI-GAN在数据集M²FPA上对不同人脸属性的部分侧脸图像的人脸校正可视化结果,图中每个偏转角对应3列图像,左侧为输入的侧脸图像,中间为生成的正脸图像,右侧为真实的正脸图像. 可以发现即使在较大偏转角下,生成的正脸图像在人脸的表情以及是否佩戴眼镜这些属性上,能够与真实的正脸图像保持较高的一致性,表明MI-GAN学习到的侧脸身份编码特征能够较好地保留人脸属性特征.

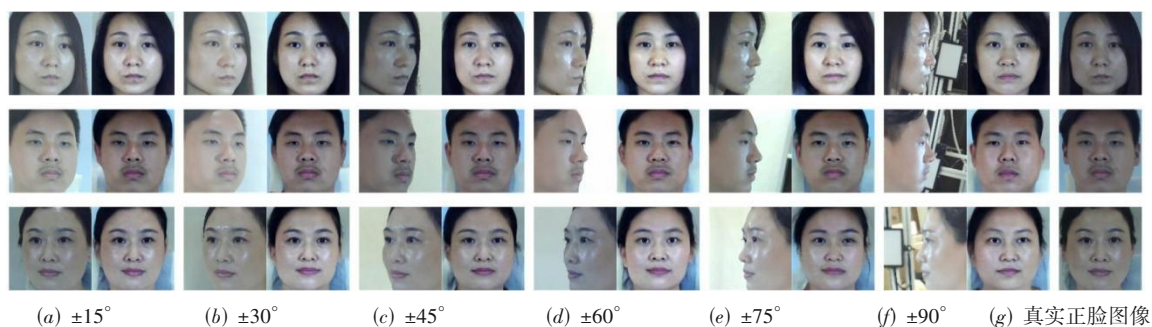


图7 MI-GAN在M²FPA数据集上对于不同偏转角的部分侧脸图像的人脸校正可视化结果

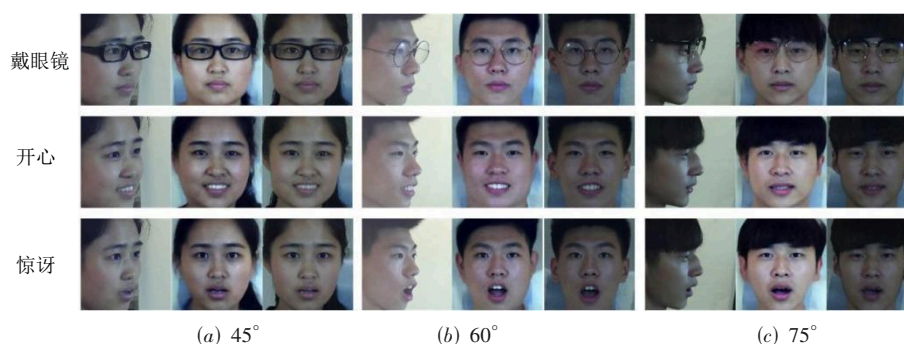


图8 MI-GAN在M²FPA数据集上对于不同人脸属性的部分侧脸图像的人脸校正可视化结果

4.3 MI-GAN的消融实验

不同于其他基于生成对抗网络进行人脸校正的模型,MI-GAN包含了身份约束识别模块 M_{id} 、角度姿态分

类模块 M_{po} 以及身份特征损失函数 L_{fid} ,为了测试不同模块的作用,在M²FPA多姿态人脸数据集上,对提出的身份约束识别模块 M_{id} 、角度姿态分类模块 M_{po} 以及身份

特征损失函数 L_{f_id} 做了消融实验, 分别为 $w/o M_{id}$ 和 L_{f_id} (删除了身份约束识别模块 M_{id} 以及身份特征损失函数 L_{f_id}), $w/o M_{po}$ (删除了角度姿态分类模块 M_{po}), $w/o L_{f_id}$ (删除了身份特征损失函数 L_{f_id}). 由于身份约束识别函数的目的是约束身份编码特征, 所以删除了身份约束识别模块的同时删除了身份特征损失函数. 表 3 展示了 MI-GAN 以及相关变种模型在不同偏转角下的人脸识别准确率, 当删除了 M_{id} 和 L_{f_id} , 只使用 M_{po} , 也就是表 3 的第一行变种模型时, 可以发现从 30° 开始, 随着偏转角度的增加, 该变种模型比其他的模型在识别正确率上明显下降. 对于表 3 的第二行变种模型 (只删除了 L_{f_id}) 和第三行变种模型 (只删除了 M_{po}), 在不同偏转角

下的人脸识别准确率也比 MI-GAN 低. 由此可以说明本文提出的角度姿态分类模块 M_{po} 、身份约束识别模块 M_{id} 和身份特征损失函数 L_{f_id} 共同促进生成尽可能逼真同时保持身份信息一致性的正脸图像. 我们在数据集 M^2FPA 上采用峰值信噪比 (Peak Signal to Noise Ratio, PSNR) 和结构相似性 (Structural Similarity Index Measure, SSIM) 评价了 MI-GAN 以及相关变种模型生成的正面人脸的图像质量, 如表 4 中所示, 可以发现 MI-GAN 在不同角度下的 PSNR 和 SSIM 都高于其他变种模型, 生成的正脸图像与真正脸图像的一致性较好. 此外, 图 9 中展示了 MI-GAN 以及相关变种模型在部分侧脸图像上的人脸校正可视化结果.

表 3 MI-GAN 与变种模型的人脸识别准确率对比

单位: %

模型	$\pm 15^\circ$	$\pm 30^\circ$	$\pm 45^\circ$	$\pm 60^\circ$	$\pm 75^\circ$	$\pm 90^\circ$
$w/o M_{id}$ 和 L_{f_id}	98.8	95.8	93.5	76.4	62.8	38.5
$w/o L_{f_id}$	98.8	96.8	95.4	78.5	65.6	42.5
$w/o M_{po}$	98.9	98.9	96.3	78.8	65.8	43.5
MI-GAN	100	98.9	96.8	80.4	67.1	44.3

表 4 MI-GAN 与变种模型在 PSNR 和 SSIM 上的实验结果对比

模型	$\pm 15^\circ$		$\pm 30^\circ$		$\pm 45^\circ$		$\pm 60^\circ$		$\pm 75^\circ$		$\pm 90^\circ$		平均值	
	PSNR/ dB	SSIM	PSNR/ dB	SSIM	PSNR/ dB	SSIM	PSNR/ dB	SSIM	PSNR/ dB	SSIM	PSNR/ dB	SSIM	PSNR/ dB	SSIM
$w/o M_{id}$ 和 L_{f_id}	15.9316	0.5934	15.4270	0.5559	15.1210	0.5520	14.5131	0.5469	14.3740	0.5376	14.0463	0.5231	14.9022	0.5515
$w/o L_{f_id}$	16.0463	0.6041	15.5050	0.5790	15.3583	0.5699	14.8049	0.5624	14.6563	0.5613	14.2214	0.5356	15.0987	0.5687
$w/o M_{po}$	16.0803	0.6069	15.7285	0.5853	15.3994	0.5749	14.8095	0.5663	14.7334	0.5690	14.3073	0.5435	15.1764	0.5743
MI-GAN	16.4599	0.6244	15.9623	0.5963	15.6697	0.5884	15.1783	0.5765	14.8538	0.5727	14.4544	0.5457	15.4297	0.5840

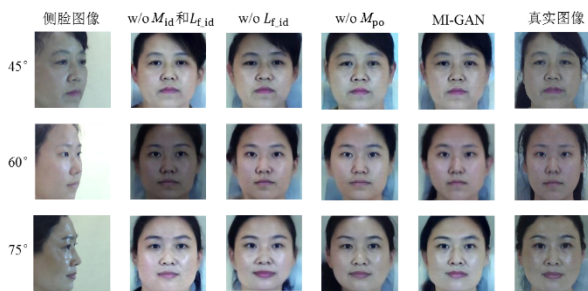


图 9 MI-GAN 与变种模型在 M^2FPA 数据集上的部分侧脸图像的人脸校正

4.4 损失函数权重的参数分析实验

考虑到不同的损失函数权重取值可能影响算法性能, 本节对 8 个损失函数权重 $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7$ 和 λ_8 的设置进行了讨论. 图 10 展示了 8 个损失函数权重的参数分析实验结果, 横坐标表示参数取值范围, 纵坐标为人脸识别准确率, 横坐标中的每个值都对应 6 个不同偏转角下的人脸识别准确率. 图中 λ_1 和 λ_4 的取值范围是

$\{5, 7.5, 10, 12.5, 15\}$, λ_2 和 λ_3 的取值范围是 $\{0.5, 0.75, 1, 1.25, 1.5\}$, λ_5 的取值范围是 $\{0.05, 0.075, 0.1, 0.125, 0.15\}$, λ_6 的取值范围是 $\{50, 75, 100, 125, 150\}$, λ_7 的取值范围是 $\{0.15, 0.225, 0.3, 0.375, 0.45\}$, λ_8 的取值范围是 $\{0.000\ 05, 0.000\ 075, 0.000\ 1, 0.000\ 125, 0.000\ 15\}$. 由于参数组合爆炸, 实验量大, 我们每次调整 1 个权重的值并固定其他 7 个权重, 得到该权重在不同参数取值下对不同偏转角人脸识别准确率的影响. 从图 10 中可以看出, 当人脸偏转角不超过 30° , 也就是偏转角度较小时, 人脸的识别准确率较高, 8 个损失函数权重参数在不同的取值下敏感性较低. 随着偏转角度的增加, 当人脸偏转角大于 30° 时, 在损失函数权重参数的取值范围内, 随着参数值的增大, 人脸识别准确率总体上表现出先升后降的趋势, 在取值范围的中间值附近, 大部分都能达到较高的人脸识别准确率. 因此, 本文损失函数权重 $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7$ 和 λ_8 分别设置为 10, 1, 1, 10, 0.1, 100, 0.3 和 0.000 1.

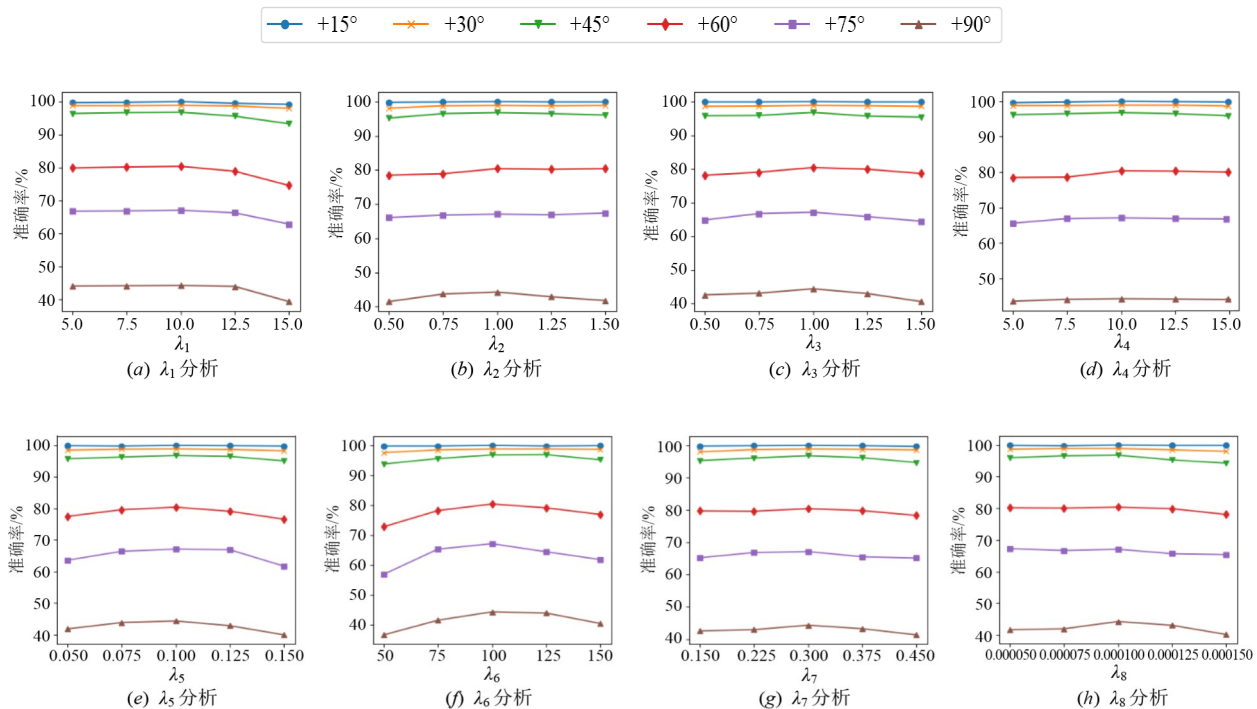


图10 损失函数权重的参数分析

5 结束语

本文提出了一种基于多任务学习和身份约束的生成对抗网络人脸校正识别方法. 该方法通过借鉴多任务学习机制,在生成网络的编码器与解码器之间构建了角度姿态分类模块和身份约束识别模块,用身份和姿态类别直接作为学习身份编码特征和姿态编码特征的监督信息,使由侧脸生成正脸的过程中加入了人脸身份监督信息,并通过设计身份特征损失函数来约束侧脸的身份编码特征逼近其正脸的身份编码特征,实现了侧脸编码特征中身份信息和姿态信息的有效解耦,通过解码器能更准确地生成与原侧脸图像保持身份一致的正脸图像,提升了大偏转角下人脸识别的准确率.

参考文献

- [1] TRAN L, YIN X, LIU X. Disentangled representation learning gan for pose-invariant face recognition[C]//Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 1415-1424.
- [2] 李倩玉, 蒋建国, 齐美彬. 基于改进深层网络的人脸识别基于改进深层网络的人脸识别算法[J]. 电子学报, 2017, 45(3): 619-625.
LI Q Y, JIANG J G, QI M B. Face recognition algorithm based on improved deep networks[J]. Acta Electronica Sinica, 2017, 45(3): 619-625. (in Chinese)
- [3] 徐先峰, 张丽, 郎彬, 等. 引入感知模型的改进孪生卷积神经网络实现人脸识别算法研究[J]. 电子学报, 2020, 48(4): 643-647.
XU X F, ZHANG L, LANG B, et al. Research on inception module incorporated siamese convolutional neural networks to realize face recognition[J]. Acta Electronica Sinica, 2020, 48(4): 643-647. (in Chinese)
- [4] SENGUPTA S, CHEN J C, CASTILLO C, et al. Frontal to profile face verification in the wild[C]//2016 IEEE Winter Conference on Applications of Computer Vision. Lake Placid: IEEE, 2016: 1-9.
- [5] KAN M, SHAN S, CHEN X. Multi-view deep network for cross-view classification[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 4847-4855.
- [6] JAMPOUR M, MAUTHNER T, BISCHOF H. Pairwise linear regression: an efficient and fast multi-view facial expression recognition[C]//11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Ljubljana: IEEE, 2015: 1-8.
- [7] JIAO L, SHANG R, LIU F, et al. Brain and Nature-Inspired Learning, Computation and Recognition[M]. Amsterdam: Elsevier Press, 2020.

- [8] 焦李成. 神经网络系统理论[M]. 西安: 西安电子科技大学出版社, 1990.
JIAO L C. Neural Network System Theory[M]. Xi'an: Xidian University Press, 1990. (in Chinese)
- [9] 焦李成. 神经网络计算[M]. 西安: 西安电子科技大学出版社, 1993.
JIAO L C. Neural Network Computing[M]. Xi'an: Xidian University Press, 1993. (in Chinese)
- [10] 焦李成. 神经网络的应用与实现[M]. 西安: 西安电子科技大学出版社, 1993.
JIAO L C. Application and Realization of Neural Network [M]. Xi'an: Xidian University Press, 1993. (in Chinese)
- [11] 焦李成, 赵进, 杨淑媛, 等. 深度神经网络学习、优化与识别[M]. 北京: 清华大学出版社, 2017.
JIAO L C, ZHAO J, YANG S Y. Deep Learning, Optimization and Recognition[M]. Beijing: Tsinghua University Press, 2017. (in Chinese)
- [12] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 815-823.
- [13] WEI W, TIAN C, ZHANG Y. Robust face pose classification method based on geometry-preserving visual phrase[C]//2014 IEEE International Conference on Image Processing. Paris: IEEE, 2015: 815-823.
- [14] ZHANG H, ZHANG Y, HUANG T S. Pose-robust face recognition via sparse representation[J]. Pattern Recognition, 2013, 46(5): 1511-1521.
- [15] LIN L, WANG K, MENG D, et al. Active self-paced learning for cost-effective and progressive face identification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(1): 7-19.
- [16] CHEN X, ZHU Y, ZHANG Y. Adversarial learning with collaborative attention for facial makeup removal[J]. Neurocomputing, 2021, 434: 249-260.
- [17] SHAO M, Li L, WANG H, et al. Selective generative adversarial network for raindrop removal from a single image[J]. Neurocomputing, 2021, 426: 265-273.
- [18] DUAN Y, HAN C, TAO X, et al. Panoramic image generation: from 2-d sketch to spherical image[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(1): 194-208.
- [19] HAN C, DUAN Y, TAO X, et al. Toward variable-rate generative compression by reducing the channel redundancy[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(7): 1789-1802.
- [20] 方晨, 郭渊博, 王娜, 等. 基于生成对抗网络的差分隐私数据发布方法[J]. 电子学报, 2020, 48(10): 1983-1992.
FANG C, GUO Y B, WANG N, et al. Differential private data publishing method based on generative adversarial network[J]. Acta Electronica Sinica, 2020, 48(10): 1983-1992. (in Chinese)
- [21] 江泽涛, 覃露露. 一种基于U-Net生成对抗网络的低照度图像增强方法[J]. 电子学报, 2020, 48(2): 258-264.
JIANG Z T, QIN L L. Low-light image enhancement method based on u-net generative adversarial network[J]. Acta Electronica Sinica, 2020, 48(2): 258-264. (in Chinese)
- [22] 王格格, 郭涛, 余游, 等. 基于生成对抗网络的无监督域适应分类模型[J]. 电子学报, 2020, 48(6): 1190-1197.
WANG G G, GUO T, YU Y, et al. D Unsupervised domain adaptation classification model based on generative adversarial network[J]. Acta Electronica Sinica, 2020, 48(6): 1190-1197. (in Chinese)
- [23] PENG X, YU X, SOHN K, et al. Reconstruction-based disentanglement for pose-invariant face recognition[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 1632-1641.
- [24] GAO H, EKENEL H K, STIEFELHAGEN R. Pose normalization for local appearance-based face recognition [C]//Advances in Biometrics, Third International Conference. Alghero: Springer, 2009: 32-41.
- [25] BOOKSTEIN F L. Principal warps: thin-plate splines and the decomposition of deformations[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, 11(6): 567-585.
- [26] TAIGMAN Y, YANG M, RANZATO M A, et al. Deep-face: closing the gap to human-level performance in face verification[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 1701-1708.
- [27] BERG T, BELHUMEUR P N. Tom-vs-Pete classifiers and identity-preserving alignment for face verification [C]//British Machine Vision Conference. Surrey: BMVA Press, 2012: 1-11.
- [28] ASHRAF A B, LUCEY S, CHEN T. Learning patch correspondences for improved viewpoint invariant face recognition[C]//2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Salt Lake

- City: IEEE, 2008: 1-8.
- [29] BEYMER D, POGGIO T. Face recognition from one example view[C]//Proceedings of the Fifth International Conference on Computer Vision. Cambridge: IEEE, 1995: 500-507.
- [30] HASSNER T, HAREL S, PAZ E, et al. Effective face frontalization in unconstrained images[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 4295-4304.
- [31] BLANZ V, VETTER T. A morphable model for the synthesis of 3D faces[C]//Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. Los Angeles: ACM, 1999: 187-194.
- [32] ROMDHANI S, VETTER T. Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego: IEEE, 2005: 986-993.
- [33] YIM J, JUNG H, YOO B I, et al. Rotating your face using multi-task deep neural network[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 676-684.
- [34] ZHANG Y, SHAO M, WONG E K, et al. Random faces guided sparse many-to-one encoder for pose-invariant face recognition[C]//2013 IEEE International Conference on Computer Vision. Sydney: IEEE, 2013: 2416-2423.
- [35] KAN M, SHAN S, CHANG H, et al. Stacked progressive auto-encoders (SPAЕ) for face recognition across poses [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 1883-1890.
- [36] HUANG R, ZHANG S, LI T, et al. Beyond face rotation: global and local perception GAN for photorealistic and identity preserving frontal view synthesis[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2458-2467.
- [37] ZHAO J, CHENG Y, XU Y, et al. Towards Pose Invariant Face Recognition in the Wild[C]//2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 2207-2216.
- [38] QIAN Y, DENG W, HU J. Unsupervised face normalization with extreme pose and expression in the wild[C]//2019 IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 9851-9858.
- [39] YIN Y, JIANG S, ROBINSON J P, et al. Dual-attention GAN for large-pose face frontalization[C]//15th IEEE International Conference on Automatic Face and Gesture Recognition. Buenos Aires: IEEE, 2020: 249-256.
- [40] HU Y, WU X, YU B, et al. Pose-guided photorealistic face rotation[C]//2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8398-8406.
- [41] CAO K, RONG Y, LI C, et al. Pose-robust face recognition via deep residual equivariant mapping[C]//2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 5187-5196.
- [42] DENG J, GUO J, ZHOU Y, et al. Retinaface: Single-stage dense face localisation in the wild[EB/OL]. (2019) [2021]. <https://arxiv.org/abs/1905.00641>.
- [43] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[C]//4th International Conference on Learning Representations. Piscataway: IEEE, 2016: 31-38.
- [44] JOHNSON J, ALAHI A, LI F F. Perceptual losses for real-time style transfer and super-resolution[C]//European Conference on Computer Vision. Amsterdam: Springer, 2016: 694-711.
- [45] CHEN S, LIU Y, GAO X, et al. Mobilefacenets: efficient cnns for accurate real-time face verification on mobile devices[C]//Chinese Conference on Biometric Recognition. Urumqi: Springer, 2018: 428-438.
- [46] LI P, WU X, HU Y, et al. M²fpa: a multi-yaw multi-pitch high-quality dataset and benchmark for facial pose analysis[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 10042-10050.
- [47] GUOY, ZHANG L, HU Y, et al. Ms-celeb-1m: a dataset and benchmark for large-scale face recognition[C]//European Conference on Computer Vision. Amsterdam: Springer, 2016: 87-102.
- [48] KINGMA D P, BA J. Adam: a method for stochastic optimization[C]//3rd International Conference on Learning Representations. San Diego: IEEE, 2015: 1-8.
- [49] WANG G, MA J, ZHANG Q, et al. Pseudo facial generation with extreme poses for face recognition[C]//2021 IEEE Conference on Computer Vision and Pattern Recognition. Virtual: IEEE, 2021: 1994-2003.

作者简介



黄欣研 女,1992年7月出生,陕西宝鸡人.现为西安电子科技大学人工智能学院博士研究生.主要研究方向为深度学习、图像处理和计算机视觉.

E-mail: xinyanh@stu.xidian.edu.cn



刘芳(通讯作者) 女,1963年2月出生,湖南华容人.1995年毕业于西安电子科技大学.现为西安电子科技大学人工智能学院教授,博士生导师.主要研究方向为人工智能和模式识别、机器学习、图像感知和场景理解、进化计算和数据挖掘.

E-mail: f63liu@163.com



鲍骛月 男,1998年8月出生,山西朔州人.现为西安电子科技大学人工智能学院博士研究生.主要研究方向为机器学习、图像处理和模式识别.

E-mail: baoqianyue@163.com



李任鹏 男,1997年9月出生,湖南娄底人.现为西安电子科技大学人工智能学院硕士研究生.主要研究方向为深度学习和图像处理.

E-mail: lrpp121314@163.com



李玲玲 女,1988年4月出生,陕西白水人.2017年毕业于西安电子科技大学.现为西安电子科技大学人工智能学院副教授.主要研究方向为量子进化优化学习、深度学习方法与应用、复杂遥感影像理解与解译.

E-mail: llli@xidian.edu.cn



刘旭 男,1989年5月出生,陕西咸阳人.2019年毕业于西安电子科技大学.现为西安电子科技大学人工智能学院副教授.主要研究方向为机器学习/深度学习理论、图像/视频处理方法.中国电子学会会员编号:E190027251M.

E-mail: xuliu@xidian.edu.cn



陈璞花 女,1986年7月出生,陕西汉中 人.2016年毕业于西安电子科技大学.现为西安电子科技大学人工智能学院副教授.主要研究方向为机器学习、模式识别和遥感图像解译.

E-mail: phchen@xidian.edu.cn



刘洋 女,1998年6月出生,山西朔州人.现为西安电子科技大学人工智能学院博士研究生.主要研究方向为机器学习、图像处理和模式识别.

E-mail: yyyliu98@163.com