基于姿态注意力的特定角度人脸正面化网络

解奕鹏[1,2],秦品乐[1,2],曾建潮[1,2],闫寒梅[3],柴 锐[1,2],赵鹏程[1,2]
(1. 中北大学 大数据学院,山西 太原 030051;
2. 山西省医学影像与数据分析工程研究中心(中北大学),山西 太原 030051;
3. 山西警察学院 刑事科学技术系,山西 太原 030401)

摘 要:人脸正面化对人脸识别有重要意义,但实际监控场景中大姿态的人脸正面化效果通常不如较小姿态,因此提出姿态引导的特定角度生成对抗网络(Pose-Specific Generative Adversarial Network,PS-GAN)。PS-GAN 由生成器和鉴别器组成,生成器由编码器、姿态注意模块、特征转换模块以及解码器四部分组成,编码器与解码器分别对输入图像进行下采样与上采样,姿态注意模块为网络引入人脸结构先验的同时约束模型关注感兴趣区域,特征转换模块对编码器得到的侧脸特征进行变换并抑制冗余通道。首先,将连续的姿态变化划分为离散的姿态集合,单个 PS-GAN 模型由某一特定角度的数据训练;然后,将多个 PS-GAN 进行组合,使其适用于任意角度的人脸输入。在本实验室自主采集的 MASFD 数据集以及 CAS-PEAL-R1 公开数据集上进行了大量的定性与定量实验,验证了网络结构的有效性以及合理性;与现有方法相比,虽然 PS-GAN 是由受限数据集训练的,但它也能在非同源数据上有良好的视觉效果。

关键词:人脸正面化;注意力机制;生成对抗网络;人脸识别;深度学习
中图分类号:TP391.41 **文献标识码**:A **文章编号**:1673-629X(2023)07-0047-08
doi:10.3969/j.issn.1673-629X.2023.07.007

Face Frontalization Network of Specific Angle Based on Pose Attention

XIE Yi-peng[1,2],QIN Pin-le[1,2],ZENG Jian-chao[1,2],YAN Han-mei[3],
CHAI Rui[1,2],ZHAO Peng-cheng[1,2]
(1. School of Big Data,North University of China,Taiyuan 030051,China;
2. Shanxi Medical Imaging and Data Analysis Engineering Research Center (North University of China),
Taiyuan 030051,China;
3. Criminal Science and Technology,Shanxi Police College,Taiyuan 030401,China)

Abstract:Face frontalization is of great significance for face recognition,but in actual monitoring scenarios,the reconstruction results of large postures is usually not as good as that of small postures. Therefore,we propose a Pose-Specific Generative Adversarial Network (PS-GAN),which consists of a generator and a discriminator. The generator consists of an encoder,a pose attention module,a feature conversion module,and a decoder. The encoder and decoder downsample and upsample the input image,respectively. The pose attention module introduces face structure prior into the network and simultaneously constrains the model to focus on the region of interest. The feature transformation module transforms the profile features obtained by the encoder and suppresses redundant channels. Firstly,the continuous pose are divided into discrete pose sets. A single PS-GAN model is trained by data from a specific Angle,and then multiple PS-GANs are combined to make it suitable for face input from any Angle. In addition,a large number of qualitative and quantitative experiments were carried out on the MASFD data set independently collected by our laboratory and the CAS-PEAL-R1 public data set,which verified the validity and rationality of the network structure. Compared with existing methods,although PS-GAN is trained on a restricted dataset,but it also has excellent visual performance on non-homologous data.

Key words:face frontalization;attention mechanism;generative adversarial network (GAN);face recognition;deep learning

0 引 言

近年来,随着深度学习的快速发展,人脸识别性能得到了很大提升,但它们局限于接近正面的人脸识别(Near-Frontal Face Recognition,NFFR)。多项研究表

收稿日期:2022-09-06 **修回日期**:2023-01-06
基金项目:山西省重点研发项目(201803D31212-1);山西省“揭榜挂帅”重大专项(202101010101018)
作者简介:解奕鹏(1998-),男,硕士研究生,研究方向为深度学习与计算机视觉;通信作者:闫寒梅(1968-),女,副教授,研究方向为刑事图像技术。

明,NFFR 算法对大姿态的人脸识别效果不佳[1],因此姿态鲁棒性人脸识别(Pose - Invariant Face Recognition,PIFR)成为近年来的研究热点。

目前解决 PIFR 问题的方法主要分两大类:一类是学习对姿态变化鲁棒的特征,另一类即人脸正面化。第一类方法[2]依赖大量姿态分布均匀的训练数据,而现有数据大多呈现长尾分布,很难学习到对姿态鲁棒的人脸特征[3]。第二类方法[4]可以在不重新训练现有人脸识别模型的基础上,通过生成对应的正面图像进行人脸识别,提高准确率。

现有的人脸正面化方法按照合成域分为基于 2D 的方法和基于 3D 的方法。基于 3D 的方法[5-6]通常对较小姿态的人脸正面化效果比较好,而对较大姿态的人脸纹理细节损失严重,同时三维拟合时严重依赖面部关键点检测的准确性,渲染时计算量大,都使得这类方法难以在实际中应用[1]。

而生成对抗网络[7]的提出极大改善了二维图像合成的视觉效果,所以越来越多的研究人员采用基于 2D 的方法解决人脸正面化问题。例如,Huang 等人提出一种双通道生成对抗网络 TP-GAN[8],融合两通道的特征生成正面人脸图像,并用实验结果说明了 GAN 网络生成的正面人脸可以提高人脸识别的精度;Yin 等人提出 DA-GAN[9],他们为解码器添加了自注意力机制,增强了纹理细节,却忽略了编码时鲁棒的特征提取也一样重要;Hu 等人提出了 CAPG-GAN[10],使用 5 个人脸关键点作为结构先验;类似的,Tu 等人提出了 MDFR[11]框架,他们使用 18 个关键点作为结构先验。然而无论是 CAPG-GAN 还是 MDFR,他们都直接将图像与姿态图直接拼接,无法保证在网络深层依旧起作用,也没有指出不同数量关键点的姿态图对网络性能的影响;更进一步,Hao 等人提出 DGPR[12]网络,他们使用人像草图作为人脸的先验知识,但使用侧面的人像草图生成正面的人像草图时,会引入一些不必要的误差;最后,Li 等人提出 Sym-GAN[3],他们关注了偏转及俯仰角同时存在的人脸正面化工作,但该方法在非同源的测试数据上视觉效果不理想。

总体来说,上述方法除 Sym-GAN 以外,均忽略了俯角对正面化工作带来的影响,而现实生活中监控摄像头拍摄的图像大多是俯视并且具有一定的偏转,导致这些方法生成效果不理想。另一方面,上述方法将所有角度的人脸数据混合训练,导致模型对特定角度人脸生成效果不突出。因此,该文聚焦于特定偏转和俯角的人脸正面化问题,结合注意力机制,引导网络生成逼真的正面图像。

主要工作如下:

(1)以类 Pix2Pix[13]网络为骨干,提出基于姿态图引导的特定角度人脸正面化网络 PS-GAN;并将多个 PS-GAN 网络进行组合,用于人脸正面化。

(2)提出姿态注意模块,引入人脸结构先验的同时约束模型关注感兴趣区域。使用特征可视化技术展示模型的感兴趣区域。

(3)在本实验室自主采集的多角度人脸监控数据集[14](Multi-Angle Surveillance Face Dataset,MASFD)以及 CAS-PEAL-R1[15]数据集上进行了充分的定性和定量实验,验证各模块结构设计的合理性;实验结果表明该方法可以有效提高人脸正面化效果,并在非同源数据上平均人脸相似度达到 67.24%。

1 相关工作

1.1 生成对抗网络

生成对抗网络[7](Generative Adversarial Networks,GAN)最初由 Goodfellow 等人提出。GAN 由生成网络和判别网络组成,两者相互博弈:生成网络用于生成与原数据集分布接近的实例,欺骗判别网络;判别网络用来鉴别输入数据是真实数据还是由生成器伪造的。

GAN 网络的提出显著提高了二维图像生成的视觉效果。如 Yin 等人提出的 DA-GAN,在解码器部分添加了两个自注意模块,同时使用多个鉴别器,除了生成图像直接与标签进行判别外,引入了三种人脸掩膜,分别关注正面人脸的不同部位。其对抗损失表示如下:

$$L_{\text{adv}} = \sum_{j\in\{f,s,k,h\}} L_j(D_j,G) \tag{1}$$

$$L_j(D_j,G) = \min_{G_j}\max_{D}\{E_{x\in I_j}[\log(D(x))] + E_{z\in I_p}[\log(1-D(G_j(z)))]\} \tag{2}$$

其中,$j\in\{f,s,k,h\}$ 表示整幅图像、面部、五官及头发部分。

1.2 注意力机制

注意力机制分为空间注意力机制与通道注意力机制。空间注意力机制计算图像中感兴趣区域并加强。例如,Jaderberg 等人提出空间转换网络 STN[16],可以对特征图在空间中进行转换并自动搜索重要区域。

通道注意力机制关注图像在通道维度上的重要性,为不同通道的特征图分配权重。Hu 等人提出 SENet[17],分为压缩和激励两部分,压缩部分对全局信息进行压缩,在通道维度学习各通道的重要性,激励部分为各通道分配权重。

最后,Woo 等人提出 CBAM[18]模块,将通道注意力与空间注意力组合,提高图像生成质量。

但是,上述注意力机制只关注特征图本身,没有足够的先验知识做引导。

2　人脸正面化网络 PS-GAN

图 1 给出 PS-GAN 网络的主体框架及各模块结构。仅由卷积、池化等操作组成的网络往往无法准确地关注感兴趣区域,因此将姿态图与空间注意力机制相结合,提出姿态注意模块(Pose Attention Module,PAM),引入人脸结构先验的同时使网络关注感兴趣区域。其次,为将编码器提取到的侧脸高维特征转换为目标特征并去除通道冗余,提出特征转换模块(Feature Transform Module,FTM)。

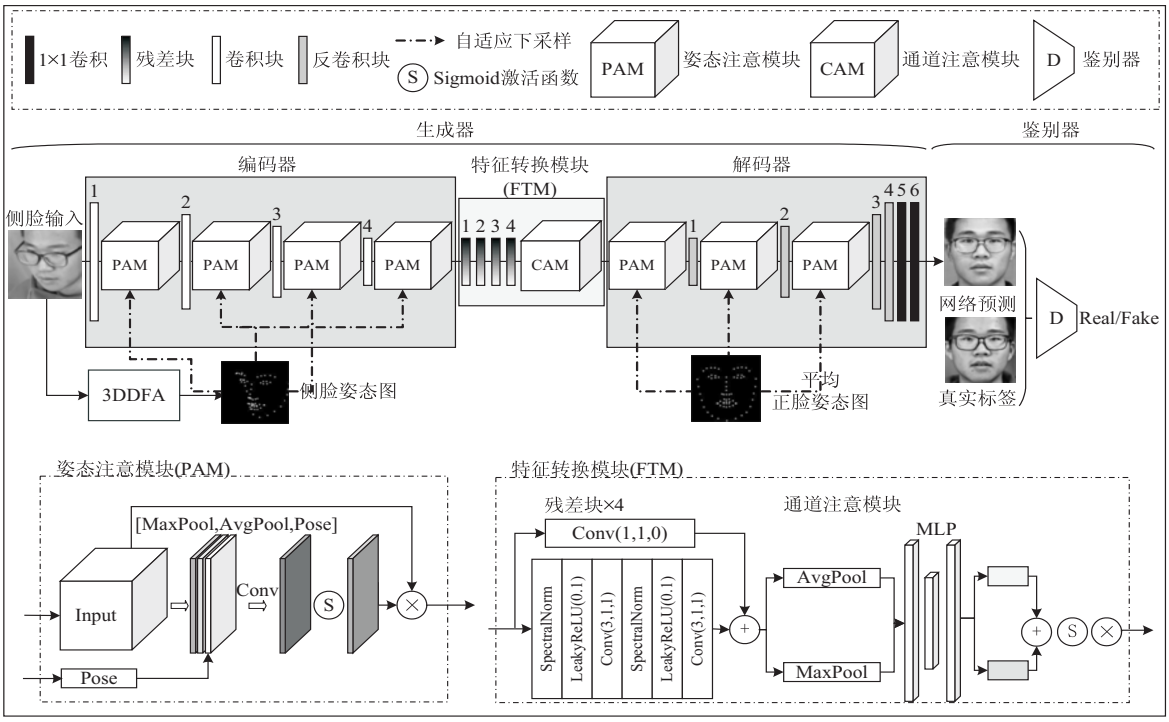

图 1　PS-GAN 网络结构及 PAM、FTM 模块结构

2.1　整体结构设计

PS-GAN 网络由生成器和鉴别器组成,生成器由编码器、姿态注意模块 PAM、特征转换模块 FTM 以及解码器四部分组成,编码器由四个卷积块组成,每个卷积块之后都添加了带有侧脸姿态的注意模块,解码器由四个反卷积块及两个卷积核为 1×1 的卷积层组成,仅在解码器的前三层添加带有平均正脸姿态的注意模块。

其次,将多个 PS-GAN 模型组合,对于任意姿态的人脸输入,首先使用人脸角度估计网络[19]计算人脸角度,再选择与该角度最接近的 PS-GAN 模型进行人脸矫正,得到最终生成结果。利用这种组合方法,解决任意角度人脸输入的问题。

2.2　姿态注意模块

2.2.1　姿态图的设计

人脸的关键点包含丰富的人脸结构信息,该文使用 3DDFA[20](3D Dense Face Alignment)获取输入图像的 68 或 8 个关键点坐标,并将其保存在灰度图中,作为姿态图。

对于侧面的人脸图像,直接使用 3DDFA 获取到的坐标信息作为侧脸姿态图;而对于正面人脸,虽然每个人的正面关键点都不相同,但五官及人脸轮廓的大致位置都有迹可循,因此使用 3DDFA 计算训练集内所有正面人脸图像平均坐标,作为平均正脸姿态图。平均正面人脸关键点的计算方法如下:

$$\mathrm{Point}_{\mathrm{avg}} = \frac{1}{N}\sum_{n=1}^{N}(\mathrm{DFA}(I_n^{gt})) \tag{3}$$

其中,N 表示训练集的人数;DFA(·)为 3DDFA 网络;I_n^{gt} 表示训练集内第 n 个人的正面人脸图像。

2.2.2　姿态图与空间注意力的结合

之前的空间注意力机制使用两种池化操作获取图像的高频细节,却无法保证这些细节的准确性。因此,将空间注意力机制[18]与姿态图相结合,提出姿态注意模块,其网络结构如图 1 所示。

对输入特征,首先在通道维度上进行最大池化与平均池化,各得到与输入特征大小相同、通道数为 1 的特征图,再与其对应的姿态图进行拼接,得到特征块,最后经过卷积、Sigmoid 激活函数后得到空间注意力权重图,再对权重图与输入特征进行点积得到更新后的特征图。

2.3　特征转换模块

输入图像经过编码器得到侧脸的高维特征,再使用特征转换模块 FTM 对侧脸特征进行转换[21],FTM 模块结构如图 1 所示。

其中 SpectralNorm 表示光谱归一化[22]；使用斜率为 0.1 的 LeakyReLu 作为激活函数。侧脸特征首先经过四层残差块进行特征变换，再使用通道注意模块为每个通道赋予权重，去除冗余通道的同时增强感兴趣通道。在通道注意中，该特征首先在长、宽两个维度上进行平均池化以及最大池化，然后将这两个特征向量输入全连接层，再将两特征向量的对应元素相加，得到通道注意力权重图，最后将该权重图与转换后的特征在通道维度进行点积得到新的特征图。

2.4 网络结构分析

在 PS-GAN 生成器的基础上去除姿态注意模块、通道注意模块后，记为 Backbone 网络。该文使用特征可视化技术[23]分析 Backbone 网络编码器与解码器的感兴趣区域，如图 2 所示，其中深色的区域为模型关注区域。

由图 2 第一行可知，编码器在第一、二层关注了人脸区域，而从第三层开始关注头发、背景等非必要区域。因此，对于编码器，不需要添加更多的卷积层辅助提取侧脸特征，但为使编码器在各尺度都关注人脸区域，所以给编码器的每个尺度都添加带有侧脸的姿态注意模块。

图 2 Backbone 网络特征可视化

由图 2 第二行可知，Backbone 的解码器的前三层恢复正面人脸的大致轮廓，即人脸共有的属性，共性信息；第四层反卷积将特征恢复到原来大小，后两层 1×1 卷积辅助交替恢复人脸的全局与局部属性，即个性信息。而平均正脸姿态图仅包含人脸共性信息，因此仅为解码器的前三层添加带有平均正脸姿态图的姿态注意模块，辅助网络快速生成人脸共性信息的同时不影响人脸个性信息的恢复。

2.5 损失函数

本节将介绍用到的损失函数。为了使损失函数关注人脸区域，使用人脸分析方法将图像的背景区域扣除[24]：

$$L = L_{\text{adv}} + \lambda_{\text{pixel}} L_{\text{pixel}} + \lambda_{\text{lpips}} L_{\text{lpips}} + \lambda_{\text{ip}} L_{\text{ip}} + \lambda_{\text{tv}} L_{\text{tv}} \tag{4}$$

其中，L_{adv} 表示对抗损失，L_{pixel} 表示多尺度像素损失，L_{lpips} 表示感知损失，L_{ip} 表示身份保持损失，L_{tv} 表示全变分正则化项，λ_* 表示不同损失的权重。

2.5.1 对抗损失

GAN 网络由生成器与鉴别器组成，两者的对弈过程表示如下：

$$L_{\text{adv}} = \min_G \max_D \{ E_{x \in I_f} [\log(D(x))] + E_{z \in I_p} [\log(1 - D(G(z)))] \} \tag{5}$$

其中，E 表示求期望；$x \in I_f$ 表示 x 来自真实的正脸图像集；$D(x)$ 表示鉴别器；$z \in I_p$ 表示 z 来自真实的侧脸图像集；$G(z)$ 表示生成器。

2.5.2 多尺度像素损失

在正面化结果 I^G 上使用多尺度像素损失[3]来约束生成内容一致性：

$$L_{\text{pixel}} = \frac{1}{S} \sum_{s=1}^{S} \sum_{w,h,c=1}^{W_s,H_s,C} \| I^G_{s,w,h,c} - I^{gt}_{s,w,h,c} \|_1 \tag{6}$$

其中，S 表示尺度数，取 $S = 3$，I^{gt} 为正脸标签。

2.5.3 感知损失

使用感知相似性损失[25]保持图像的结构信息。具体的，使用 Conv3-64、Conv3-128、Conv3-256、Conv3-512 及最后一层全连接计算损失：

$$L_{\text{lpips}} = \frac{1}{L} \sum_{l=1}^{L} \sum_{h_l=1,w_l=1}^{H_l,W_l} \| w_l (\text{Vgg}^l(I^G) - \text{Vgg}^l(I^{gt})) \|_1 \tag{7}$$

其中，Vgg^l 表示网络提取的第 l 层的特征图，w_l 表示对第 l 层赋予的权重。

2.5.4 身份保持损失

使用 LightCNN-29V2[26]提取身份特征。具体的，使用该网络的最后一个池化层及最终的网络输出共同作为人脸高维特征[3,24]，具体公式如式(8)：

$$L_{\text{ip}} = \| \varphi^{\text{pool}}(I^G) - \varphi^{\text{pool}}(I^{gt}) \|_1 + \| \varphi^{\text{output}}(I^G) - \varphi^{\text{output}}(I^{gt}) \|_1 \tag{8}$$

其中，$\varphi^{\text{pool}}(\cdot)$ 表示网络在最后一个池化层提取的特征，$\varphi^{\text{output}}(\cdot)$ 表示网络的最终输出结果。

2.5.5 全变分正则化项

GAN 网络生成的图像往往会存在大量人工伪影，因此添加全变分正则化项以减少伪影[9]：

$$L_{tv} = \sum_{c=1}^{C} \sum_{w=1,h=1}^{W,H} \| I^G_{c,h,w+1} - I^G_{c,h,w} \|_1 + \| I^G_{c,h+1,w} - I^G_{c,h,w} \|_1 \tag{9}$$

其中，W 和 H 分别表示图像的高和宽。

3 实验与结果分析

3.1 数据集

使用 MASFD 以及 CAS-PEAL-R1 数据集进行实验。MASFD 数据集由本实验室自主采集，共包含了 4 253人的 23 种角度组合。CAS-PEAL-R1 数据集为 CAS-PEAL 数据集的共享版本，包含 1 040 位志愿者的 30 900 张人脸图像，该文使用其姿态子库 21 840 幅

图像验证人脸正面化效果。姿态子库包含三种俯仰变化(抬头、平视、低头)和每种俯仰姿态下七种水平深度旋转姿态变化。另外为了验证 PS-GAN 网络的泛化性,又拍摄了 80 张数据集外的侧脸图像,用于验证其在非同源数据上的视觉效果。

对于 MASFD 数据集,在 4 253 人中随机选择 850 人作为测试集,其余 3 403 人作为训练集,共训练 20 个单角度模型;对于 CAS-PEAL-R1 数据集,随机选择 831 人作为训练集,209 人作为测试集,共训练 13 个单角度模型。

3.2　对比实验

3.2.1　定性比较

使用 TP-GAN[8]、Sym-GAN[3]、DA-GAN[9]、CAPG-GAN[10] 四种方法与 PS-GAN 网络进行对比,在 MASFD 数据集中的定性结果如图 3 所示。

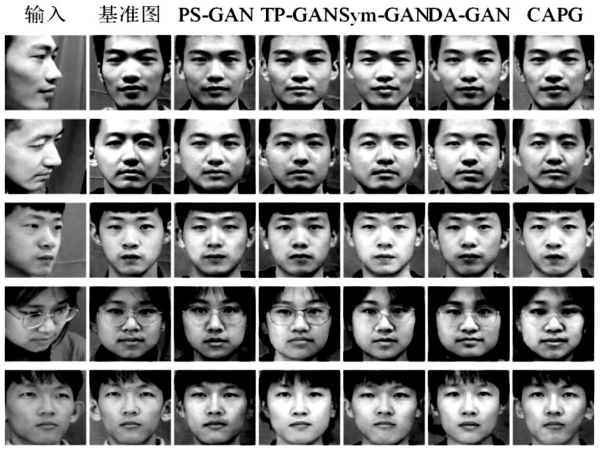

图 3　不同方法在多个角度上生成效果对比

由图 3 可知,文中方法相比其他方法有更少的人工伪影,文中方法只针对特定角度进行训练,网络关注较小范围的姿态变化,与其他方法相比,文中方法在整体结构和局部细节上均与标签更加相似。

其次,展示了 PS-GAN 与 Sym-GAN 在不同俯仰角上的正面化效果,如图 4 所示。

图 4　Sym-GAN 与 PS-GAN 在俯仰角上生成效果对比

此外,在非同源的数据上进行测试,生成结果如图 5 所示。其中,第一行为输入,第二行为输出。

图 5　文中方法在非同源数据上的生成效果

最后,展示了 PS-GAN 在两数据集在各角度下的生成结果,如图 6 所示。其中,第一列为基准图像,奇数行为网络输入,偶数行为网络输出。

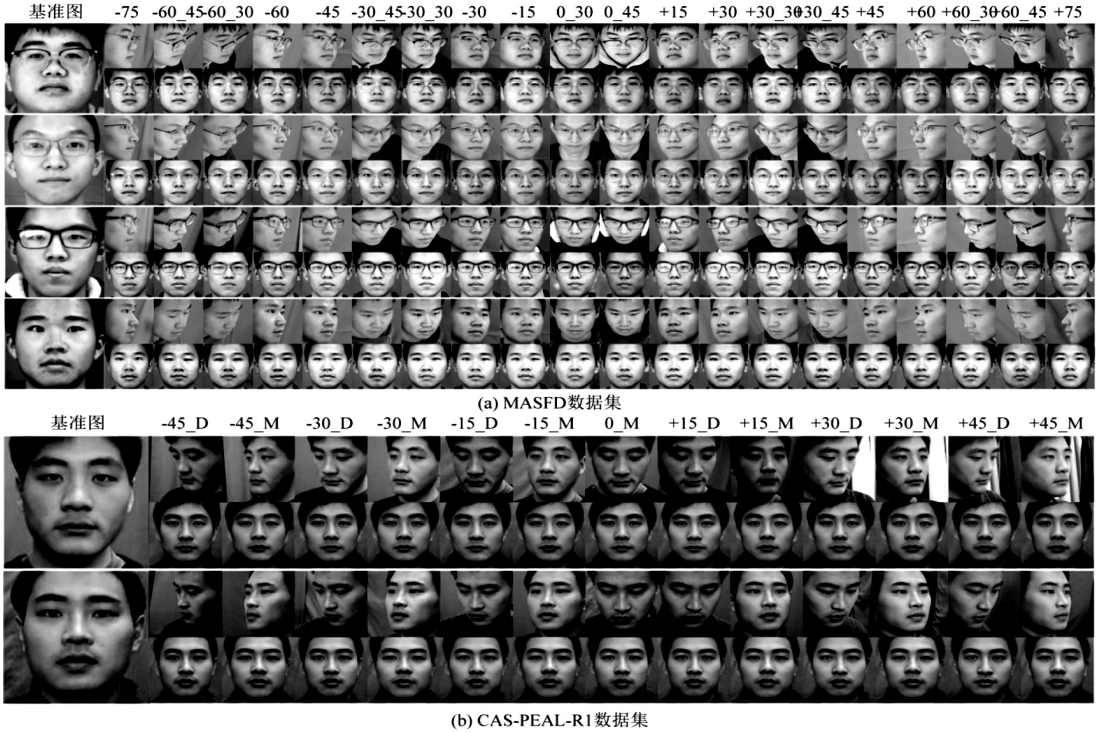

图 6　文中方法在不同数据集上各个角度的生成效果

3.2.2 定量比较

该文使用 Rank-1 指标在两数据集上对 PS-GAN 及上述方法进行定量实验,其定量结果如表 1 与表 2 所示。由于 CAS-PEAL-R1 数据集未给出具体的俯仰角度,因此,使用俯视、平视、仰视三种视角进行标注。其中第一行为俯仰角,第二行为偏转角。由表 1 和表 2 可知,文中方法在较大角度下依旧表现良好,说明用单个角度数据对模型进行训练是有效的。

表 1 不同方法在 MASFD 数据集上的 Rank-1 识别率 %

方法	俯角 0°					俯角 30°			俯角 45°		
	偏转 ±15°	偏转 ±30°	偏转 ±45°	偏转 ±60°	偏转 ±75°	偏转 0°	偏转 ±30°	偏转 ±60°	偏转 0°	偏转 ±30°	偏转 ±60°
TP-GAN	98.83	98.65	98.21	97.76	96.36	97.63	97.32	96.34	96.72	96.13	95.31
Sym-GAN	99.85	98.70	98.44	98.07	97.14	98.76	98.25	97.65	97.38	97.14	95.85
DA-GAN	98.73	98.73	98.43	98.01	97.36	98.65	97.44	96.73	97.86	96.86	96.18
CAPG-GAN	99.15	99.34	98.96	98.69	97.32	99.21	98.67	97.84	98.34	97.64	96.35
PS-GAN(Ours)	99.85	99.73	99.23	98.13	97.54	99.24	98.65	97.84	98.87	98.23	97.73

表 2 不同方法在 CAS-PEAL-R1 数据集上的 Rank-1 识别率 %

方法	仰视				平视			俯视			
	偏转 0°	偏转 ±15°	偏转 ±30°	偏转 ±45°	偏转 ±15°	偏转 ±30°	偏转 ±45°	偏转 0°	偏转 ±15°	偏转 ±30°	偏转 ±45°
TP-GAN	98.88	98.95	98.87	97.63	100.00	99.94	98.71	97.68	97.73	97.47	95.88
Sym-GAN	99.71	99.93	99.76	95.25	100.00	100.00	99.72	100.00	99.72	99.46	96.95
DA-GAN	99.75	99.78	99.62	97.94	100.00	100.00	99.71	98.97	98.97	98.86	98.13
CAPG-GAN	99.35	99.43	99.32	97.52	100.00	99.93	99.37	98.59	98.68	98.59	97.83
PS-GAN(Ours)	99.94	100.00	99.94	98.95	100.00	100.00	99.94	100.00	99.95	99.87	99.24

但是 Rank-1 指标无法体现模型生成人脸与真实人脸的相似程度,因此又使用人脸识别方法计算平均人脸相似度得分(Average Similarity Score,ASS)与方差(Variance),计算方法如下所示:

$$\mathrm{ASS} = \frac{1}{N}\sum_{i=1}^{N}(1 - \mathrm{FR}(I^G, I^{gt})) \times 100\% \tag{10}$$

$$\mathrm{Variance} = \frac{1}{N}\sum_{i=1}^{N}((1 - \mathrm{FR}(I_i^G, I_i^{gt})) \times 100\% - \mathrm{ASS})^2 \tag{11}$$

其中,I^G 为生成图像,I^{gt} 为真实标签,FR(·)为计算 I^G 与 I^{gt} 之间距离的函数。不同数据集的平均人脸相似度与方差计算结果如表 3 所示。

由表 3 可知,PS-GAN 方法在同源测试集和非同源数据上的平均人脸相似度均较高,方差较小,模型稳定性较好。

表 3 不同方法及消融实验在 MASFD 数据集与非同源数据中的平均人脸相似度及其方差

方法	MASFD 测试数据集		非同源数据	
	ASS/%	Variance	ASS/%	Variance
TP-GAN	88.65	19.58	56.74	71.16
Sym-GAN	87.41	20.43	53.64	68.8
DA-GAN	89.53	16.36	54.64	57.65
CAPG-GAN	89.25	18.76	62.46	55.71
2 Resblock	87.97	17.41	64.32	74.53
3 Resblock	88.32	17.56	64.09	71.32
4 Resblock	88.68	16.56	65.80	57.22
5 Resblock	88.45	16.48	65.67	58.46
6 Resblock	88.65	16.76	65.40	55.86
Backbone	88.32	17.56	64.09	71.32
Backbone+Conv×1	88.67	17.43	65.38	65.52
Backbone+Conv×2	89.46	16.88	66.47	51.48
Backbone+Conv×3	89.65	16.97	66.32	58.37
Backbone+Conv×4	89.33	16.91	66.37	51.76
8pose	89.73	16.76	65.77	57.31
68pose	89.92	16.32	66.12	54.36
PS-GAN(Ours)	91.76	15.88	67.24	52.28

3.3 消融实验

由于 Rank-n 指标在消融实验中最终结果较接近,无法体现模型的真实性能,因此使用人脸相似度与方差作为消融实验(+30_30 角为例)指标。

实验设置一:为了验证特征转换模块的有效性及结构设计的合理性,在 Backbone 的基础上分别添加 2、3、4、5、6 个残差块进行实验。

实验设置二:由图 2 可知,网络在解码器最后交替恢复人脸局部与全局信息,为了说明解码器后卷积层的数量是否会对生成结果有影响,分别为 Backbone 添

加了 1、2、3、4 个 1×1 卷积进行实验。

实验设置三:在 Backbone 的基础上,首先将不同关键点数量的姿态图直接与输入图像在通道维度拼接,说明不同关键点姿态图对网络生成效果的影响。然后将姿态图与注意力结合,验证文中姿态引导方式的优越性。

3.3.1　不同数量残差块对生成效果的影响

本节分别为特征转换模块 PTM 添加 2、3、4、5、6 个残差块,分析不同数量残差块对生成效果的影响,实验结果如表 3 中 Resblock 所示。

由表 3 可知,在同源测试集上,添加四层残差块时平均人脸相似度最高,添加五层残差块时方差最低。在非同源数据上,随着残差块的数量增加,平均人脸相似度有所提升,且总体方差明显减小,即添加残差块可以减小生成效果的波动性。为了使网络结构精简,同时保证模型对非同源数据的泛化能力,为 PTM 模块添加四层残差块。

3.3.2　不同数量的 1×1 卷积对生成效果的影响

本节为 Backbone 添加 1、2、3、4 个 1×1 卷积,分析解码器后卷积层的数量对生成效果的影响,实验结果如表 3 中 Conv 所示。

由表可知,对同源测试集添加三层卷积层时平均人脸相似度最高,添加两层卷积层时方差最低,波动性最小;对非同源数据,在解码器之后添加两层卷积时相比 Backbone 的人脸相似度提高较明显,稳定性更好。因此,为解码器添加两层 1×1 卷积。

3.3.3　姿态注意模块的有效性

首先,在 Backbone 的基础上,比较不同姿态图对生成效果的影响。共进行了三组实验,分别是不加姿态引导(Backbone)、添加 8 关键点姿态引导(8pose)、添加 68 关键点姿态引导(68pose)的网络。每组实验均将原图与姿态图直接拼接,实验结果如表 3 所示。由表 3 可知,不论是测试集数据还是非同源数据,关键点越多的姿态图,其生成的图像平均人脸相似度越高,波动性越低。

其次,又展示了三种模型编码器与解码器的感兴趣区域,如图 7 所示。

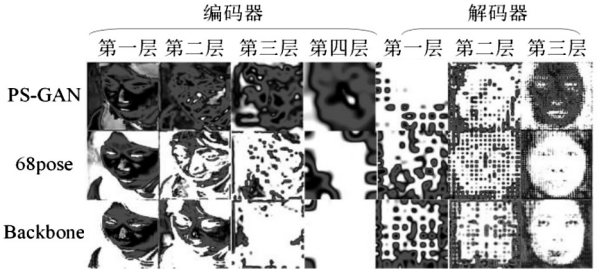

图 7　Backbone、68pose 及 PS-GAN 方法特征可视化对比

由图 7 可知,Backbone 网络在编码器第三层已经不再关注人脸区域,而 68pose 的网络在第三层时依旧关注人脸区域。这一方面反映了姿态引导的有效性,另一方面也说明这种直接与原图进行拼接的姿态引导方式很难在网络深处起作用。

因此,将 68 关键点的姿态图与注意力机制结合,即文中方法。为了验证文中方法的有效性,对 68pose 和 PS-GAN 进行定性对比,结果如表 3 所示。由表 3 可知,PS-GAN 网络在两类数据上相对其他实验方法平均人脸相似度最高、波动性最低。最后,如图 7 所示,PS-GAN 模型的编码器各层在姿态注意模块的辅助下准确的关注感兴趣区域,解码器的前三层在姿态注意模块的辅助下快速捕捉人脸共性信息,证明了网络结构设计的合理性。

4　结束语

针对特定角度人脸正面化问题,通过实验验证网络的有效性与结构设计的合理性。结合注意力机制,提出 PS-GAN 网络,并使用特定角度人脸数据训练单个模型,将多个角度模型进行组合,一定程度上缓解各角度生成效果不突出的问题,但这种方法对于数据集中不存在的角度人脸生成效果一般。在后续的工作中,应当考虑如何使用数据集内有限的特定角度,在任意角度都能生成较好的结果。

参考文献:

[1] LUO H, CEN S, DING Q, et al. Frontal face reconstruction based on detail identification, variable scale self-attention and flexible skip connection[J]. Neural Computing and Applications, 2022, 34(13): 10561-10573.

[2] CAO K, RONG Y, LI C, et al. Pose-robust face recognition via deep residual equivariant mapping[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 5187-5196.

[3] 李虹霞,秦品乐,闫寒梅,等. 基于面部特征图对称的人脸正面化生成对抗网络算法[J]. 计算机应用,2021,41(3):714-720.

[4] LIU Y, CHEN J. Unsupervised face frontalization for pose-invariant face recognition[J]. Image and Vision Computing, 2021, 106: 104093.

[5] SHI L, SONG X, ZHANG T, et al. Histogram-based CRC for 3D-aided pose-invariant face recognition[J]. Sensors, 2019, 19(4): 759.

[6] ZHOUH, LIU J, LIU Z, et al. Rotate-and-render: unsupervised photorealistic face rotation from single-view images[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle: IEEE, 2020: 5911-5920.

[7] GOODFELLOW IAN J, JEAN P A, MEHDI M, et al. Generative adversarial nets[C]//Proceedings of the 27th international conference on neural information processing systems. Montreal: MIT, 2014: 2672-2680.

[8] 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]//Proceedings of the IEEE international conference on computer vision. Venice: IEEE, 2017: 2439-2448.

[9] YIN Y, JIANG S, ROBINSON J P, et al. Dual-attention GAN for large-pose face frontalization[C]//2020 15th IEEE international conference on automatic face and gesture recognition (FG 2020). Buenos Aires: IEEE, 2020: 249-256.

[10] HU Y, WU X, YU B, et al. Pose-guided photorealistic face rotation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 8398-8406.

[11] TU X, ZHAO J, LIU Q, et al. Joint face image restoration and frontalization for recognition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(3): 1285-1298.

[12] HAO J, CHEN X. Detailed feature guided generative adversarial pose reconstruction network[J]. IEEE Access, 2021, 9: 56093-56103.

[13] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu: IEEE, 2017: 1125-1134.

[14] 李红霞. 基于生成对抗网络的多角度人脸正面化研究[D]. 太原: 中北大学, 2021.

[15] GAO W, CAO B, SHAN S, et al. The CAS-PEAL large-scale Chinese face database and baseline evaluations[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2007, 38(1): 149-161.

[16] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks[J]. Advances in Neural Information Processing Systems, 2015, 28: 2017-2025.

[17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 7132-7141.

[18] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). Munich: Springer, 2018: 3-19.

[19] ZHOU Y, GREGSON J. Whenet: real-time fine-grained estimation for wide range head pose[J]. arXiv: 2005.10353, 2020.

[20] GUO J, ZHU X, YANG Y, et al. Towards fast, accurate and stable 3d dense face alignment[C]//European conference on computer vision. [s.l.]: Springer, 2020: 152-168.

[21] ZHANG P, YANG L, LAI J, et al. Exploring dual-task correlation for pose guided person image generation[J]. arXiv: 2203.02910, 2022.

[22] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks[J]. arXiv: 1802.05957, 2018.

[23] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016: 2921-2929.

[24] WEI Y, LIU M, WANG H, et al. Learning flow-based feature warping for face frontalization with illumination inconsistent supervision[C]//European conference on computer vision. Glasgow: Springer, 2020: 558-574.

[25] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 586-595.

[26] WU X, HE R, SUN Z, et al. A light CNN for deep face representation with noisy labels[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(11): 2884-2896.