

基于双路循环生成对抗网络的多姿态人脸识别方法

徐志京, 王东*

上海海事大学信息工程学院, 上海 201306

摘要 针对非正面姿态下人脸识别率低下的问题, 提出一种基于双路循环生成对抗网络的多姿态人脸识别方法。该网络由人脸转正及人脸旋转两部分组成。人脸转正部分完成侧面人脸向正面人脸的转化, 实现多对一的姿态类别映射; 人脸旋转部分完成对正面人脸身份特征的提取及指定姿态人脸的生成, 实现一对多的姿态类别映射。训练过程中利用两条循环路径将人脸转正及人脸旋转过程结合, 一路完成人脸侧面至正面再至侧面的循环转化, 另一路则完成人脸正面至侧面再至正面的循环转化, 促使两部分内容相互利用、约束, 提高对侧面人脸的识别率。为了加快网络的收敛速度, 降低训练难度, 训练过程分为先局部后整体两个不同的阶段进行。在人脸数据集 Multi-PIE 及 CFP 上的实验结果表明, 该方法能够有效提高对侧面人脸的识别率。

关键词 图像处理; 人脸识别; 多姿态; 双路循环; 生成对抗网络

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/AOS202040.1910002

Multi-Pose Face Recognition with Two-Cycle Generative Adversarial Network

Xu Zhijing, Wang Dong*

College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China

Abstract This study proposes a multi-pose face recognition method with a two-cycle generative adversarial network to address low face recognition accuracy of non-frontal poses. The network consists of two aspects: face frontalization and face rotation. The face frontalization aspect converts profile faces to frontal faces and implements many-to-one pose category mapping. The face rotation aspect converts frontal faces to profile faces with specified poses, extracts the identity features of the frontal faces, and implements one-to-many pose category mapping. To further improve the face recognition of profile poses, two cyclic paths are used to combine the face frontalization and face rotation processes. One path is used for the cyclic conversion of profile faces to frontal faces and then to profile faces, and the other path is used for the cyclic conversion of frontal faces to profile faces and then to frontal faces. To reduce the difficulty in the training process and speed up the convergence of the network, the training process will be performed in two different stages: partial and complete training. Experiment results on Multi-PIE and CFP show that this method can effectively improve the recognition accuracy of profile poses.

Key words image processing; face recognition; multi-pose; two-cycle; generative adversarial network

OCIS codes 100.3008; 100.3010

1 引 言

近年来, 深度学习的快速发展以及大量含标注的人脸数据库的出现, 使人脸识别技术得到了显著提高^[1]。但在实际应用中, 姿态变化仍然是人脸识别中难以克服的外在因素之一^[2]。当输入人脸的姿态发生变化时, 传统的人脸识别算法的识别性能会发生明显变化。主要原因是人脸姿态在变化的同

时, 常会导致面部器官发生移位, 并且会伴随有自遮挡效应, 使得网络很难获取完整的人脸特征信息。之后出现的 3D 人脸标准化方法^[3]虽然能够在一定程度上解决以上问题, 但该类方法非常依赖 3D 建模的准确性, 而在实验过程中精确的 3D 人脸网络很难获得, 这将导致该方法难以学习到面部纹理细节。基于深度学习的方法同样能够一定程度地缓解姿态变化引起的自遮挡等问题, 但该类方法需要对

收稿日期: 2020-05-06; 修回日期: 2020-06-03; 录用日期: 2020-06-23

基金项目: 国家自然科学基金(61673259)、国家重点研发计划(2019YFB1600605)

* E-mail: 2368072842@qq.com

大量的人脸数据进行训练,如 DeepFace^[4]方法中用于训练的图片共 4.4 亿张,并且当人脸姿态变化超过 45°时,该类方法的识别性能仍会大幅度下降。究其原因,主要是姿态变化越大,网络所学得的人脸特征中混杂的姿态等非关键信息越多,导致网络无法对侧面人脸进行准确的识别。

针对传统方法存在的问题,人们开始研究更加有效的人脸识别算法,主要包含以下两种研究方向:1)由输入的侧面人脸图像生成相同身份下的正面人脸图像,再对得到的正面人脸图像进行特征提取与识别^[5-6];2)直接从侧面人脸图片中提取出具有姿态鲁棒性的人脸特征^[7-9]。基于以上两种研究思路所提出的多姿态人脸识别算法相比于传统方法均表现出了良好的识别性能。Zhang 等^[11]在孪生网络 Siamese 结构^[10]的基础上提出 DSN 网络(Deep Disentangling Siamese Network),输入一对图片,利用两张图片之间的对比信息将姿态、光照等与身份特征无关的因素剔除,提取出具有姿态鲁棒性的人脸特征。赵澜涛等^[12]提出一种基于双路卷积神经网络的多姿态人脸识别网络,将输入的人脸投影到高维特征空间并实现身份信息与姿态信息的分离,进而对侧面人脸进行识别。自 Goodfellow 等^[13]提出生成对抗网络 GAN(Generative Adversarial Network GAN)以来,GAN 网络及其衍生体已被广泛应用于图像处理领域^[14-15],并取得了良好的检测识别效果。Tran 等^[17]在条件生成对抗网络 CGAN^[16](Conditional GAN)等网络的基础上提出了 DR-GAN(Disentangled Representation Learning GAN)网络,生成器在姿态编码的约束下生成对应的侧面人脸图像,通过判别器对生成图片的姿态进行分类,最终提取出具有姿态鲁棒性的人脸特征。Ma 等^[18]在原有 DR-GAN 的基础上增加了姿态学习模块,提出了 PW-GAN 网络,充分考虑了人脸识别过程中姿态因素的重要性。Liu 等^[19]提出了一种基于编码器-鉴别器-生成器结构的多任务生成对抗网络(MTAN),去除与人脸特征无关的姿态、光照等因素,得到精确的人脸特征。

以上所提方法在取得一定效果的同时也存在着一些不足之处,如 DSN 及 PW-GAN 等方法仅研究了侧面人脸向正面人脸转化的过程,无法生成指定姿态的侧面人脸图像,且当姿态变化高于 45°时,所得的人脸特征中仍会含有较多的姿态信息。DR-GAN 及 MTAN 等方法在训练过程中将人脸转正

和人脸旋转两部分内容混合进行训练,增加了网络训练的难度,导致网络在训练过程中难以收敛,进而影响侧面人脸的识别准确率。针对以上问题,本文提出了一种基于双路循环生成对抗网络(TC-GAN)的多姿态人脸识别方法,首先将复杂的人脸识别问题细化成人脸转正和人脸旋转两部分并分别进行训练,再通过加入的双路循环结构对网络进行约束优化,最后通过多组实验验证了 TC-GAN 能够有效提高不同姿态下的人脸识别率及生成图像的质量。

2 TC-GAN

一般地,GAN 网络由生成器 G 和判别器 D 构成。生成器 G 的目的是尽可能输出与真实样本分布相似的假样本。判别器 D 的目的是计算输入的样本属于真实样本的概率值。 G 和 D 在训练过程中以对抗学习的方式交替更新各自的网络参数,即最大化判别器 D 的判别能力的同时,利用生成器 G 降低生成样本与真实样本之间数据分布的差异,最终实现纳什均衡。对应的目标函数为

$$\min_G \max_D V(D, G) = E_{x \sim P_{\text{data}}} [\ln D(x)] + E_{z \sim P_z} \{\ln \{1 - D[G(z)]\}\}, \quad (1)$$

式中: $E(\cdot)$ 代表期望值; $x \sim P_{\text{data}}(x)$ 表示 x 为来自真实样本的数据; $z \sim P_z(z)$ 表示随机噪声服从特定的数据分布(如高斯分布、均匀分布)。

在 GAN 网络的基础上,针对多姿态人脸识别问题构建了一种双路循环生成对抗网络(TC-GAN),具体结构如图 1 所示。TC-GAN 由一对 GAN 网络组成,分别完成了人脸转正及人脸旋转任务。第一组 GAN 网络由正面人脸生成器 G_N 及对应的判别器 D_N 组成,实现了侧面人脸向正面人脸的转化。相似地,第二组 GAN 网络由侧面人脸生成器 G_E 与对应的判别器 D_E 组成,实现了对输入正面人脸身份特征的提取,并与额外添加的姿态编码 c 结合,生成指定姿态下的侧面人脸图像。

2.1 人脸转正

输入侧面人脸图像后,由正面人脸生成器 G_N 生成对应的正面人脸图像。与传统 GAN 网络不同,该阶段的判别器 D_N 可视为多任务判别器,除了需要在训练过程中不断提高对真假图片的辨别能力外,还需要对输入人脸图像的身份进行准确分类。对应的输出主要由以下两部分组成: $D_N = [D_N^{(r)}, D_N^{(d)}]$,其中 $D_N^{(r)}(\cdot) \in \mathbb{R}^{(1)}$ 代表输入图片属于真实图片的概率值,概率值越大,表明生成的图片越真

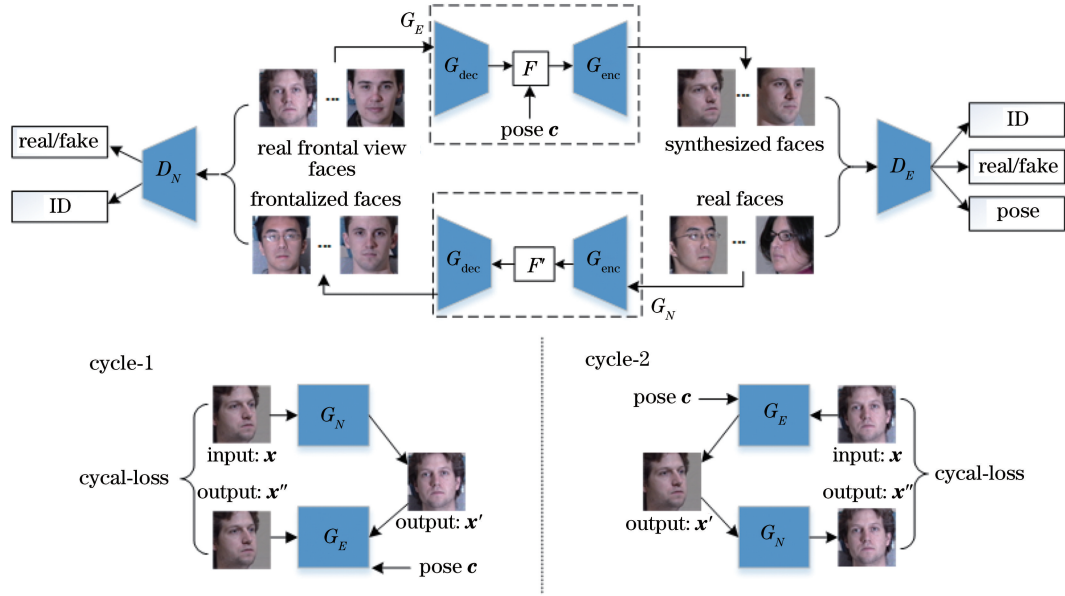


图 1 TC-GAN 网络结构

Fig. 1 Framework of TC-GAN

实, $D_N^{(d)}(\cdot) \in \mathbb{R}^{(N^d)}$ (N^d 代表训练对象的总数) 代表对输入图片的身份进行分类, 用来保证网络的输出与输入为同一个训练对象。由于生成的人脸图片仅含一种姿态, 因此无需对生成图片的姿态进行分类。

判别器 D_N 中 $D_N^{(r)}(\cdot)$ 部分对应的对抗损失函数可表示为

$$L_{D_N}^{(adv)} = E_{x_1 \sim P_f} \{-\ln[D_N^{(r)}(x_1)]\} + E_{x_2 \sim P_{data}} \{-\ln[1 - D_N^{(r)}(G_N(x_2))]\}, \quad (2)$$

式中: $x_1 \sim P_f$ 表示 x_1 为真实正面人脸图像, 且取自真实分布的数据; $x_2 \sim P_{data}$ 表示 x_2 为数据集中任意姿态的人脸图像, 且取自真实分布的数据。最大化真实人脸图片对应输出值的同时最小化生成图片对应的输出值, 从而提高判别器对真假图片的判别能力。判别器中 $D_N^{(d)}(\cdot)$ 部分对输入的真实人脸图片对应身份进行分类时所产生的损失函数可表示为

$$L_{D_N}^{(id)} = E_{x_1, y_1 \sim P_f} \{-y_1 \ln[D_N^{(d)}(x_1)]\}, \quad (3)$$

式中: $x_1, y_1 \sim P_f$ 表示输入正面人脸图像 x_1 对应的身份标签为 y_1 。

与判别器 D_N 相反, 生成器 G_N 的目的是输出与真实数据分布足够相似的人脸图像, 使判别器中的 $D_N^{(r)}(\cdot)$ 无法区分输入图片的真假性。生成器 G_N 所要优化的对抗损失函数可表示为

$$L_{G_N}^{(adv)} = E_{x \sim P_{data}} \{-\ln\{D_N^{(r)}[G_N(x)]\}\}, \quad (4)$$

最大化判别器对生成图片真假性的预测值。

判别器中 $D_N^{(d)}(\cdot)$ 对生成图片的身份进行分类时产生的损失函数可表示为

$$L_{G_N}^{(id)} = E_{x, y \sim P_{data}} \{-y \ln\{D_N^{(d)}[G_N(x)]\}\}. \quad (5)$$

在图像生成等任务中, 像素级别的损失具有广泛的应用。TC-GAN 中同样引入像素级别的损失, 用来衡量输出图像与目标图像之间的差异, 提高网络的性能。人脸转正阶段的像素级别损失函数可表示为

$$L_N^{(pixel)} = \frac{1}{n} \sum_i \|x_i^{(f)} - G_N(x_i)\|_1, \quad (6)$$

式中: x_i 代表输入的侧面人脸图像; $x_i^{(f)}$ 代表输入侧面人脸 x_i 对应的正面人脸图像; $\|\cdot\|_1$ 代表 1-范数。综合上述损失函数, 对于人脸转正部分的生成器 G_N 与判别器 D_N , 需要优化的损失函数分别为

$$L_{G_N} = L_{G_N}^{(adv)} + L_{G_N}^{(id)} + L_N^{(pixel)}, \quad (7)$$

$$L_{D_N} = L_{D_N}^{(adv)} + L_{D_N}^{(id)}. \quad (8)$$

2.2 人脸旋转

在人脸旋转阶段, 输入正面人脸图像, 由生成器 G_E 结合姿态编码 c 实现指定姿态人脸图像的生成并对正面人脸图像的特征进行提取。相似地, 人脸旋转任务中的判别器 D_E 由三部分组成: $D_E = [D_E^{(r)}, D_E^{(d)}, D_E^{(p)}]$, 目的是在训练过程中不断提高对真假图片的分辨能力以及对输入对象的身份、姿态进行准确分类的能力。其中: $D_E^{(r)}(\cdot) \in \mathbb{R}^{(1)}$ 负责对输入图片的来源进行判断, 判断输入图片为真实图片的概率值; $D_E^{(d)}(\cdot) \in \mathbb{R}^{(N^d)}$ (N^d 代表训练对象的个数) 代表对输入图片身份进行分类, 保证输入与输出为同一个训练对象; $D_E^{(p)}(\cdot) \in \mathbb{R}^{(N^p)}$ (N^p 代表

训练集中姿态的种类数)代表对输入人脸姿态进行分类,保证输出人脸的姿态与指定姿态保持一致。

判别器 D_E 中 $D_E^{(r)}$ 部分所要优化的对抗损失函数和判别器 D_E 对真实图片的身份和姿态进行分类时对应的损失函数分别表示为

$$L_{D_E}^{(adv)} = E_{x_1 \sim P_{data}} \{-\ln[D_E^{(r)}(x_1)]\} + E_{x_2 \sim P_f, c \sim P_u} \{-\ln[1 - D_E^{(r)}[G_E(x_2, c)]]\}, \quad (9)$$

$$L_{D_E}^{(id)} = E_{x_1, y^{(d)} \sim P_{data}} \{-y^{(d)} \ln[D_E^{(d)}(x_1)]\}, \quad (10)$$

$$L_{D_E}^{(p)} = E_{x_1, y^{(p)} \sim P_{data}} \{-y^{(p)} \ln[D_E^{(p)}(x_1)]\}, \quad (11)$$

式中: $x_1 \sim P_{data}$ 表示 x_1 为数据集中任意姿态的人脸图像; $x_2 \sim P_f$ 表示 x_2 为真实正面人脸图像; $c \sim P_u$ 代表指定的姿态编码; $y^{(d)}$ 代表身份标签; $y^{(p)}$ 表示姿态标签。

侧面人脸生成器 G_E 的结构与正面人脸生成器 G_N 相似,由编码器 G_{enc} 与解码器 G_{dec} 组成,编码器 G_{enc} 将输入的正面人脸图片转化成隐式编码 f 后,与含有姿态信息的编码 c 进行级联,最终由解码器 G_{dec} 解码输出指定姿态的人脸图像 $G_E(x, c)$ 。生成器 G_E 的目的是产生足够真实的侧面人脸图像,使判别器无法对生成图片的真假性进行准确判断。生成器 G_E 需要优化的对抗损失函数和判别器 D_E 对生成图像的身份及姿态进行分类时对应的损失函数分别表示为

$$L_{G_E}^{(adv)} = E_{x \sim P_f, c \sim P_u} \{-\ln\{D_E^{(r)}[G_E(x, c)]\}\}, \quad (12)$$

$$L_{G_E}^{(id)} = E\{-y^{(d)} \ln\{D_E^{(d)}[G_E(x, c)]\}\}, \quad (13)$$

$$L_{G_E}^{(p)} = E\{-y^{(p)} \ln\{D_E^{(p)}[G_E(x, c)]\}\}。 \quad (14)$$

相似地,人脸旋转阶段同样引入像素级别损失,对应的损失函数为

$$L_E^{(pixel)} = \frac{1}{n} \sum_i \|x_i^{(d)} - G_E(x_i, c)\|_1, \quad (15)$$

式中: x_i 代表输入的正面人脸图像; $x_i^{(d)}$ 代表 x_i 对应姿态 c 下的侧面人脸图像。

综上所述,对于人脸旋转部分的生成器 G_E 与判别器 D_E ,需要优化的损失函数分别为

$$L_{G_E} = L_{G_E}^{(adv)} + L_{G_E}^{(id)} + L_{G_E}^{(p)} + L_E^{(pixel)}, \quad (16)$$

$$L_{D_E} = L_{D_E}^{(adv)} + L_{D_E}^{(id)} + L_{D_E}^{(p)}。 \quad (17)$$

2.3 双路循环优化

为了进一步提高网络的识别性能,通过引入双路循环结构将人脸转正及人脸旋转两部分内容结合到一起,利用增加的循环损失对网络进行约束。如图1所示,TC-GAN网络中包含两条循环路径,Cycle-1用于实现侧面→正面→侧面的转换过程。

输入的侧面人脸图像 $x_i \sim P_{data}$,首先由正面人脸生成器 G_N 合成对应的正面人脸图像 $G_N(x_i)$,再由侧面人脸生成器 G_E 中的解码器得到对应的隐式编码 f 后与姿态编码 c 级联,通过生成器 G_E 中的解码器生成与输入相似的侧面人脸图像 $G_E[G_N(x_i), c]$ 。当网络训练结束后,最终生成的侧面人脸图像应与输入的侧面图像相似。Cycle-1路径对应的循环损失函数为

$$L_{cyc-1} = \frac{1}{n} \sum_i \|x_i - G_E[G_N(x_i), c]\|_1。 \quad (18)$$

相反地,Cycle-2实现的是正面→侧面→正面转换过程。Cycle-2输入的正面人脸图像 $x_i \sim P_f$,结合姿态编码 c 后由侧面人脸生成器 G_E 生成指定姿态的人脸图像 $G_E(x_i, c)$,最后通过正面人脸生成器 G_N 将输入的侧面人脸图片还原成与输入图片相似的正面人脸图像 $G_N[G_E(x_i, c)]$ 。Cycle-2路径对应的循环损失函数为

$$L_{cyc-2} = \frac{1}{n} \sum_i \|x_i - G_N[G_E(x_i, c)]\|_1。 \quad (19)$$

2.4 网络详细结构

TC-GAN中生成器与判别器的网络详细结构如表1所示。该网络结构参考CASIA-Net网络,使用深层次的网络结构,每个卷积层对应的卷积核大小均为 3×3 ,这样的设计能够减少网络参数,提高网络的非线性,从而有效地对人脸特征进行提取。为了提高TC-GAN网络的稳定性,将原有的MaxPool层和ReLU函数替换成卷积层。

生成网络中的编码器 G_{enc} 由一系列卷积层组成,首先对输入大小为 96×96 的人脸图像进行卷积,每个卷积层后面均进行批量标准化(BN),激活函数采用ELU(Exponential Linear Unit)函数,以保证TC-GAN网络的稳定性。经过一系列卷积操作后由平均池化层产生 $(1, 1, 320)$ 的三维张量,即输入人脸图像的隐式编码 $f(x) \in \mathbb{R}^{(320)}$ 。生成网络的解码结构 G_{dec} 由一系列反卷积层组成,对输入的特征张量进行反卷积操作,每次完成反卷积之后同样进行批量标准化,除最后输出层使用Tanh激活函数外,其余层均采用ELU作为激活函数,最后解码器输出为 $(96, 96, 3)$ 生成人脸图像。在进行人脸正面化的过程中,解码器 G_{dec} 直接对编码器输出的侧面人脸隐式编码 f_2 进行解码输出。在人脸旋转的过程中,由编码器 G_{enc} 得到正面人脸图像的隐式特征 f_1 后,与姿态编码 c 级联,得到大小为 $(1, 1, 320 + N^{(p)})$ 的张量。最后对融合的特征编码进行解

码,输出指定姿态的人脸图像。

判别器 D_E 与生成器的编码结构相似,(96,96,3)的输入样本经过一系列卷积核大小为3的卷积层及一个平均池化层后,由全连接层输出 $(N^{(d)} + N^{(p)} + 1)$ 维矢量,其中 $N^{(d)}$ 部分用于对输入图片对应的身份进行分类,“1”代表对输入图片属于真实图

片的概率值, $N^{(p)}$ 部分用于对输入图片的姿态进行分类。相似地,判别器 D_N 对应全连接层输出的结果为 $(N^{(d)} + 1)$ 矢量,同理, $N^{(d)}$ 用于对输入图片对应的身份进行分类,“1”用于对输入图片真假性进行判断。由于判别器的输入为正面人脸图像,仅含一种姿态,因此不需要对其人脸姿态进行判断。

表1 TC-GAN详细结构

Table 1 Detailed structure of TC-GAN

Encoder and generator			Decoder		
Layer	Filter/stride	Output size	Layer	Filter/stride	Output size
			FC		$6 \times 6 \times 320$
Conv11	$3 \times 3/1$	$96 \times 96 \times 32$	FConv52	$3 \times 3/1$	$6 \times 6 \times 160$
Conv12	$3 \times 3/1$	$96 \times 96 \times 64$	FConv51	$3 \times 3/1$	$6 \times 6 \times 256$
Conv21	$3 \times 3/2$	$48 \times 48 \times 64$	FConv43	$3 \times 3/2$	$12 \times 12 \times 256$
Conv22	$3 \times 3/1$	$48 \times 48 \times 64$	FConv42	$3 \times 3/1$	$12 \times 12 \times 128$
Conv23	$3 \times 3/1$	$48 \times 48 \times 128$	FConv41	$3 \times 3/1$	$12 \times 12 \times 192$
Conv31	$3 \times 3/2$	$24 \times 24 \times 128$	FConv33	$3 \times 3/2$	$24 \times 24 \times 192$
Conv32	$3 \times 3/1$	$24 \times 24 \times 96$	FConv32	$3 \times 3/1$	$24 \times 24 \times 96$
Conv33	$3 \times 3/1$	$24 \times 24 \times 192$	FConv31	$3 \times 3/1$	$24 \times 24 \times 128$
Conv41	$3 \times 3/2$	$12 \times 12 \times 192$	FConv23	$3 \times 3/2$	$48 \times 48 \times 128$
Conv42	$3 \times 3/1$	$12 \times 12 \times 128$	FConv22	$3 \times 3/1$	$48 \times 48 \times 64$
Conv43	$3 \times 3/1$	$12 \times 12 \times 256$	FConv21	$3 \times 3/1$	$48 \times 48 \times 64$
Conv51	$3 \times 3/2$	$6 \times 6 \times 256$	FConv13	$3 \times 3/2$	$96 \times 96 \times 64$
Conv52	$3 \times 3/1$	$6 \times 6 \times 160$	FConv12	$3 \times 3/1$	$96 \times 96 \times 32$
Conv53	$3 \times 3/1$	$6 \times 6 \times 320$	FConv11	$3 \times 3/1$	$96 \times 96 \times 3$
AvgPool	$6 \times 6 \times 1$	$1 \times 1 \times 320$			
FC(D_E/D_N) $N_d + N_p + 1 / (N_d + 1)$					

3 实 验

3.1 实验设备

实验环境配置如下:64 bit Windows10 操作系统,CPU 为 Intel (R) Core (TM) i7-7800X 3.50 GHz,内存 16 GB,GPU 为 NVIDIA GeForce GTX 1080 Ti,显存为 11 GB,CUDA 9.0 和 cuDNN7.0 加速包支持。软件要求:Python3.6.6, Tensorflow1.10.0 框架。

3.2 数据集介绍及预处理

Multi-PIE 是由美国卡耐基梅隆大学在多种约束条件下采集的人脸数据集,含 337 个不同身份的对象,且每个对象的人脸图片均包含 13 种姿态变化、20 种光照变化,以及多种表情变化,共计 754200 张图

片。实验过程中使用的数据集由 337 个对象在 13 种不同姿态、20 种不同光照和中性表情下的人脸图像组成。其中,前 200 个对象作为训练集对 TC-GAN 网络进行训练,剩余 137 个对象组成测试集对网络的性能进行测试。对于测试集中的每一个对象,随机选取一张正面人脸图像组成标准人脸数据库,共计 137 张图片,用于测试不同姿态下的人脸识别率。网络训练前,需要对数据集进行预处理,即提取图片中的人脸部分,降低非人脸区域对实验结果的影响。使用人脸检测器进行人脸检测,将所有图片大小对齐后,利用随机裁剪器将图片裁剪为 $96 \text{ pixel} \times 96 \text{ pixel}$ 大小的标准人脸图像,以降低训练过程中的过拟合现象。在对 Multi-PIE 数据集上进行训练和测试的过程中,本文通过加入与姿态编

码相似的光照编码,避免了训练过程中光照因素对最终实验结果的影响。

CFP是一个专门用于多姿态人脸验证的数据集,由500个对象组成,共计7000张人脸图像。每个对象包含10张正面人脸图片和4张侧面人脸图片,并且正脸与正脸间的姿态变化小于 10° ,正脸与侧脸之间的角度变化均接近 90° 。

3.3 网络训练

实验中使用Adam算法优化损失函数,学习率设置为0.0002,动量设置为0.5。每个训练批次设置为64。为了降低网络的训练难度,加快TC-GAN网络的收敛速度,本文分两个阶段对网络进行训练。

第一阶段:以对抗学习的方式分别训练 G_N 与 D_N 与 G_E 与 D_E 两项局部内容。通过优化目标

函数 L_{G_N} 及 L_{D_N} 实现对生成器 G_N 与判别器 D_N 的训练。对应生成器 G_N 与判别器 D_N 的损失函数变化趋势如图2(a)所示。相似地,生成器 G_E 与判别器 D_E 分别通过优化目标函数 L_{G_E} 与 L_{D_E} 完成训练,对应生成器 G_E 与判别器 D_E 的损失函数变化趋势如图2(b)所示。由变化趋势可看出在训练过程中判别器与生成器的损失函数均在对抗学习过程中相互竞争,总体呈现下降趋势并最终趋于平稳。

第二阶段:第一阶段训练结束后,利用循环结构将两部分内容结合,对整个网络进行训练。通过最小化两个循环结构对应的目标函数($L_{G_N} + L_{cyc-1}$)及($L_{G_E} + L_{cyc-2}$),完成对两个生成器的进一步优化。相似地,通过最小化目标函数 L_{D_N} 与 L_{D_E} 完成对两个判别器的进一步优化。

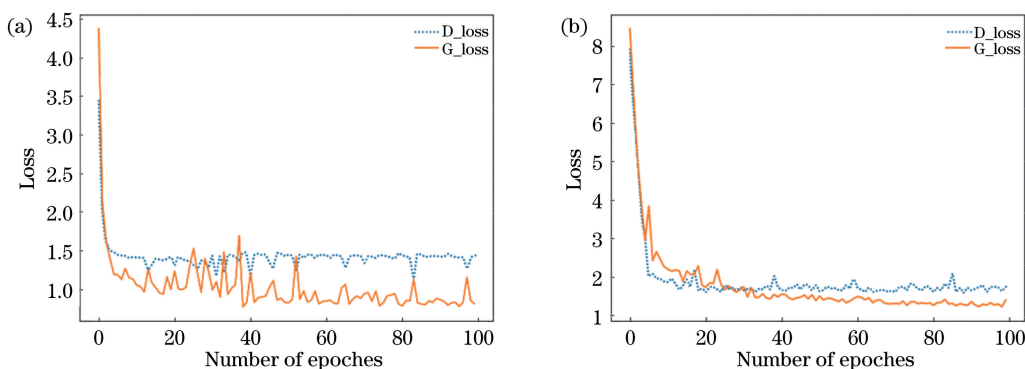


图2 损失函数变化。(a)人脸正面化阶段损失;(b)人脸旋转阶段损失

Fig. 2 Variation of loss functions. (a) Loss of face frontalization; (b) loss of face rotation

4 实验结果与分析

4.1 MULTI-PIE数据集上的评估

4.1.1 不同姿态下人脸的识别准确率

使用TC-GAN训练的结果对多姿态人脸进行识别。输入任意一张侧面人脸图片,首先由正面人脸生成器 G_N 生成相同身份下的正面人脸图像,再由侧面人脸生成器 G_E 中的解码器 G_{enc} 对生成的正面人脸图像进行特征提取,最后利用余弦距离度量生成的正面人脸图像与标准人脸数据库中人脸图像的隐式编码特征的相似度,实现对侧面人脸的识别。图3给出不同姿态下的人脸识别率。由实验结果可以看出,姿态变化越小,对应的人脸识别率越高,对于极限姿态下侧面人脸也取得了较好的识别结果。

4.1.2 不同阶段下人脸的识别准确率

TC-GAN网络的训练过程主要分为局部训练和整体训练两个阶段。为了观察不同阶段下对应网络的识别性能,本文分别在两个不同阶段对多姿态

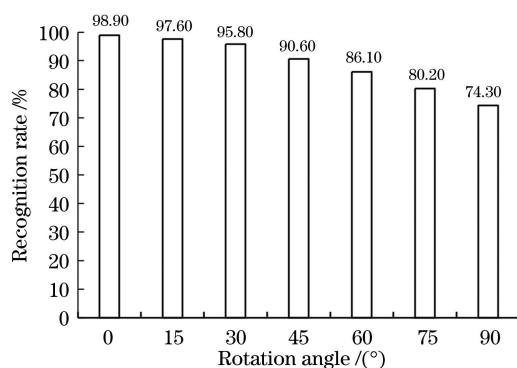


图3 不同姿态下的人脸识别率

Fig. 3 Face recognition rate for different poses

人脸进行识别,最终的实验结果如表2所示。表中stage-1代表训练过程中第一阶段训练结束后对应的多姿态人脸识别率,stage-2则代表第二阶段训练结束后对应的人脸识别率。通过对比不同阶段下的识别率发现,通过第二阶段的训练,网络的识别性能得到了较大幅度的提升。主要原因是第二阶段中加入的双路循环结构,使得网络相比于第一阶段能够

表 2 不同阶段下的人脸识别准确率

Table 2 Face recognition rates at different stages

unit: %

Stage	0°	±15°	±30°	±45°	±60°	±75°	±90°
Stage-1	98.1	95.1	92.0	84.7	80.4	75.3	70.6
Stage-2	98.9	97.6	95.8	90.6	86.1	80.2	74.3

更有效地对人脸特征进行提取及识别,证明了 TC-GAN 网络的有效性。

4.1.3 不同方法的比较

为了进一步证明 TC-GAN 网络的有效性,与现有的人脸识别算法进行比较。包括基于 GAN 网络的人脸识别算法 FF-GAN^[20]、DR-GA、MTAN 与基于非 GAN 网络的人脸识别算法 HPN、VGG-Face、TP-CNN。表 3 总结了上述方法在 Multi-PIE 数据集下的多姿态人脸识别率。从表中能够看出,TC-GAN 网络对多姿态人脸的识别效果明显优于其他方法,并且随着输入人脸姿态角度的增大,网络识别率的提高率增大。主要原因是相对于其他方法,TC-GAN 网络能够有效地对侧面人脸图片进行转

正,并能够更加准确地提取出正面人脸的身份特征,减少了人脸特征信息的丢失,进而提高了对侧面人脸的识别准确率。

4.1.4 不同因素对网络的影响

为了证明 TC-GAN 中循环损失及加入的像素级别损失对最终人脸识别结果的影响,实验过程中分别将循环损失和像素级别损失去除后,在相同的条件下再对网络进行训练。最后在相同的条件下依次对各个姿态下的人脸图像进行识别,最终测试的结果如表 4 所示。由实验结果可以看出,本文所提的循环结构以及像素级别损失对最终的实验结果有着重要的作用,并且输入的人脸姿态变化越大,提升的效果越好,对网络性能的影响越大。

表 3 不同方法下的人脸识别率

Table 3 Face recognition rates under different methods

unit: %

Method	0°	±15°	±30°	±45°	±60°	±75°	±90°	Avg1	Avg2
HPN ^[21]	95.4	84.2	78.3	72.8	61.2	47.6	29.8	0.784	0.670
VGG-Face ^[22]	98.1	97.2	96.1	92.6	84.7	62.8	34.2	0.937	0.808
TP-CNN	97.1	96.7	94.4	92.9	87.8	75.6	59.8	0.937	0.863
FF-GAN	95.7	94.6	92.5	89.7	85.2	77.2	61.2	0.915	0.852
DR-GAN ^[23]	98.1	95.0	91.3	88.0	85.8	—	—	0.916	—
MTAN	96.5	95.3	89.7	87.9	84.1	—	—	0.907	—
TC-GAN	98.9	97.6	95.8	90.6	86.1	80.2	74.3	0.938	0.891

表 4 不同因素下的人脸识别率

Table 4 Face recognition rates under different factors

unit: %

Method	0°	±15°	±30°	±45°	±60°	±75°	±90°
After removing pixel-wise loss	99.7	95.3	90.2	85.1	84.4	77.2	72.9
After removing cycle loss	99.6	95.7	92.9	86.1	83.7	76.9	71.3
TC-GAN	98.9	97.6	95.8	90.6	86.4	80.2	74.3

4.2 CFP 数据集上的评估

为了进一步验证 TC-GAN 的有效性,利用 TC-GAN 网络在 CFP 数据集上进行人脸验证实验。测试过程中的测试集由成对的人脸图片构成,有的图片对为同一个人,有的为不同的人。实验过程中,从 CFP 数据集中创建 200 个同身份的测试对及

200 个不同身份的测试对进行测试。重复上述过程 10 次,最后取 10 次结果的平均值作为最终的人脸验证结果,并与其他方法进行比较。表 5 中总结了不同方法下的人脸验证准确率,通过与其他方法的对比,发现基于 TC-GAN 的方法能够得到更高的人脸验证准确率,进一步验证了 TC-GAN 网络的有效性。

表 5 CFP 数据集上人脸验证准确率

Table 5 Recognition rate comparison on CFP database

Method	Frontal-Frontal	Frontal-Profile
VGG-Face	97.59	89.12
TP-CNN	97.35	92.41
DR-GAN	97.08	91.02
Human	96.24	94.57
TC-GAN	97.41	93.13

4.3 可视化 TC-GAN 网络的生成结果

利用 TC-GAN 中的人脸转正部分能够实现将输入的侧面人脸图像转化为相同身份下的正面人脸图像。其中部分正面人脸生成结果如图 4 所示。图中第一行代表输入网络的侧面人脸图像,姿态从

-90°依次变化到+90°,共 13 种不同的姿态变化。在第一阶段(stage-1)的任务结束后,利用该阶段训练得到的结果对输入的侧面人脸图像进行转正,生成的部分正面人脸图片如图 4 中第二行所示。在第二阶段任务结束后,同样对输入的侧面人脸图像进行转正操作,得到的结果如图 4 中第三行所示。第四行代表输入图像对应的真实正面人脸图像。通过对比不同阶段下生成图片与真实图片之间的差异,发现第二阶段下生成的正面人脸图像与真实图片更为相似。与第一阶段下生成的图片相比,第二阶段中的循环结构能够促使网络生成更加清晰的正面人脸图像,并且保留更多的细节信息。图 4 中的结果有效地证明 TC-GAN 能够更好地将侧面人脸图片还原成对应的正面人脸图像。

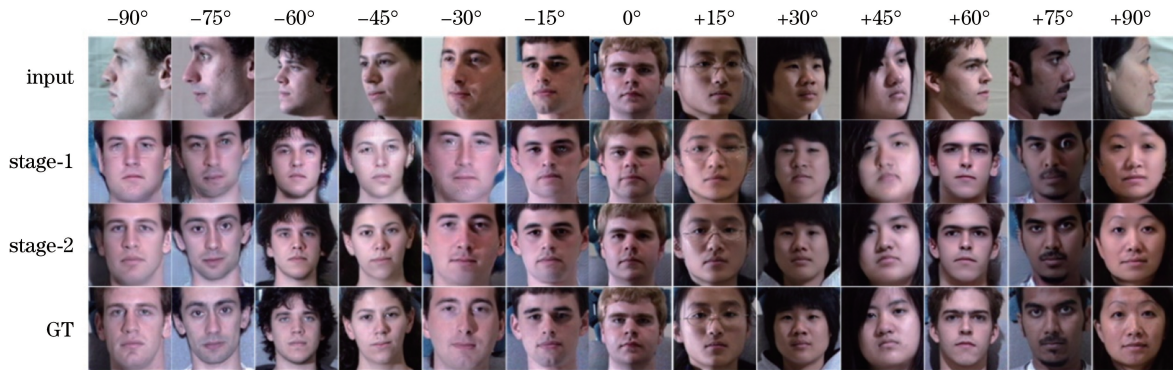


图 4 Multi-PIE 上人脸转正结果

Fig. 4 Face frontalization results on Multi-PIE

相反地,也能利用 TC-GAN 中的人脸旋转部分将输入的正面人脸图片旋转至指定姿态的侧面人脸图像。其中部分结果如图 5 所示。第一列表示输入为真实的正面人脸图像,第二列到第 14 列代表不同的旋转角度,分别对应 13 种不同的姿态。在第一阶段任务训练结束后,利用该阶段中的人脸旋转部分将输入的正面人脸图像旋转至指定的姿态,图 5 中

stage-1 行代表该阶段生成的侧面人脸图像。相似地,stage-2 行代表第二阶段训练结束后对正面人脸进行旋转所得的结果。通过对比实验结果后发现,第二阶段下得到的侧面人脸图像相比于第一阶段细节更加真实、清晰。图 5 中的结果证明了 TC-GAN 能够更加有效地对输入图片进行旋转,得到指定姿态的侧面人脸图像。

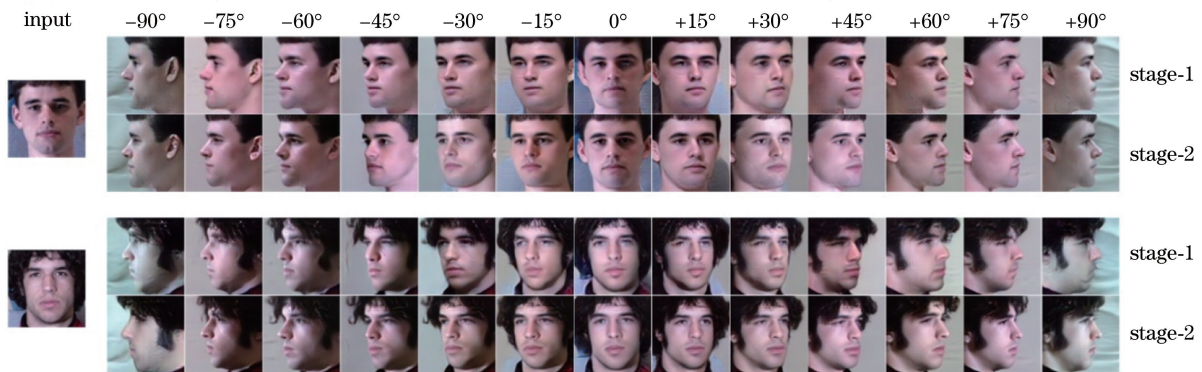


图 5 Multi-PIE 上人脸旋转结果

Fig. 5 Face rotation results on Multi-PIE

5 结 论

针对姿态变化导致的人脸识别网络性能下降问题,在生成对抗网络的基础上提出了一种能够有效提高多姿态人脸识别率的网络结构(TC-GAN)。首先将多姿态人脸识别任务进行细化并分别进行训练,再通过引入的双路循环结构实现对网络的进一步优化。多项对比实验的结果证明了所提 TC-GAN 方法的有效性,TC-GAN 网络不仅能够较好地生成指定姿态的侧面人脸图像,同时能够对侧面人脸进行有效地识别,并且姿态变化越大,提升的效果越明显。下一步将针对如何进一步减少网络训练时间及如何进一步提高侧脸识别率等问题进行深入研究。

参 考 文 献

- [1] Hu G S, Yang Y X, Yi D, et al. When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition [C]//2015 IEEE International Conference on Computer Vision Workshop (ICCVW), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 384-392.
- [2] Song X N, Feng Z H, Hu G S, et al. Dictionary integration using 3D morphable face models for pose-invariant collaborative-representation-based classification [J]. IEEE Transactions on Information Forensics and Security, 2018, 13(11): 2734-2745.
- [3] Blanz V, Vetter T. Face recognition based on fitting a 3D morphable model [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(9): 1063-1074.
- [4] Taigman Y, Yang M, Ranzato M, et al. DeepFace: closing the gap to human-level performance in face verification[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 1701-1708.
- [5] Hassner T, Harel S, Paz E, et al. Effective face frontalization in unconstrained images [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 4295-4304.
- [6] Yim J, Jung H, Yoo B, et al. Rotating your face using multi-task deep neural network[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 676-684.
- [7] Chan J S, Hsu G S J, Shie H C, et al. Face recognition by facial attribute assisted network[C]// 2017 IEEE International Conference on Image Processing (ICIP), September 17-20, 2017, Beijing, China. New York: IEEE Press, 2017: 2381-8549.
- [8] Tian Y, Peng X, Zhao L, et al. CR-GAN: learning complete representations for multi-view generation [C] // Proceedings of the 27th International Joint Conference on Artificial Intelligence, July 13-19, 2018, Stockholm, Sweden. San Francisco: Morgan Kaufmann, 2018: 942-948.
- [9] Shen Y J, Luo P, Luo P, et al. FaceID-GAN: learning a symmetry three-player GAN for identity-preserving face synthesis [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 821-830.
- [10] Gross R, Matthews I, Cohn J, et al. Multi-PIE[J]. Image and Vision Computing, 2010, 28(5): 807-813.
- [11] Zhang T, Wang H, Dong Q L. Deep disentangling Siamese network for frontal face synthesis under neutral illumination [J]. IEEE Signal Processing Letters, 2018, 25(9): 1344-1348.
- [12] Zhao L T, Lin J J. Pose-invariant face recognition with two-pathway convolutional neural network[J]. Journal of East China University of Science and Technology, 2019, 45(3): 466-470.
赵澜涛,林家骏.基于双路CNN的多姿态人脸识别方法[J].华东理工大学学报(自然科学版),2019,45(3):466-470.
- [13] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [C] // Proceedings of the 27th International Conference on Neural Information Processing Systems, December 8-13, 2014, Montreal, Canada. Cambridge: MIT Press, 2014, 2: 2672-2680.
- [14] Hu L M, Zhang Y. Facial image translation in short-wavelength infrared and visible light based on generative adversarial network [J]. Acta Optica Sinica, 2020, 40(5): 0510001.
胡麟苗,张湧.基于生成对抗网络的短波红外-可见光人脸图像翻译[J].光学学报,2020,40(5):0510001.
- [15] Liu K, Wang D, Rong M X. X-ray image classification algorithm based on semi-supervised generative adversarial networks [J]. Acta Optica Sinica, 2019, 39(8): 0810003.
刘坤,王典,荣梦学.基于半监督生成对抗网络X光图像分类算法[J].光学学报,2019,39(8):

- 0810003.
- [16] Mirza M, Osindero S. Conditional generative adversarial nets [EB/OL]. (2014-11-06) [2018-07-28]. <http://arxiv.org/abs/1411.1784>.
- [17] Tran L, Yin X, Liu X M. Disentangled representation learning GAN for pose-invariant face recognition[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1283-1292.
- [18] Ma J X, Zhou F. Multi-poses Face Frontalization based on Pose Weighted GAN [C]//2019 IEEE 3rd Information Technology, March 15-17, 2019, Chengdu, China. New York: IEEE Press, 2019: 1271-1276.
- [19] Liu Y, Wang Z W, Jin H L, et al. Multi-task adversarial network for disentangled feature learning [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 3743-3751.
- [20] Yin X, Yu X, Sohn K, et al. Towards large-pose face frontalization in the wild [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 4010-4019.
- [21] Ding C X, Tao D C. Pose-invariant face recognition with homography-based normalization [J]. Pattern Recognition, 2017, 66: 144-152.
- [22] Parkhi O M, Vedaldi A, Zisserman A. Deep face recognition[C]//Proceedings of 26th British Machine Vision Conference, September 7-10, Swansea, UK. Swansea: BMVC Press, 2015: 1-12.
- [23] Tran L, Yin X, Liu X M. Representation learning by rotating your faces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(12): 3007-3021.