

# 基于生成对抗网络的人脸表情识别技术综述

植炜基<sup>1</sup>, 刘春雨<sup>1</sup>, 郑婉君<sup>2</sup>, 张敏<sup>1</sup>, 詹思宇<sup>1</sup>

(1. 深圳大学 物理与光电工程学院, 广东 深圳 518000;

2. 广东科学技术职业学院, 广东 珠海 519090)

**摘要:**人脸表情识别目前在人机交互、智能控制、安全和心理等领域具有广泛的应用,而深度学习的发展更是使得人脸表情识别技术取得了巨大进步。但是人脸表情识别技术仍面临着数据不足、姿态偏转、身份差异以及面部遮挡等难题。而随着生成对抗网络近年来在图像合成等方面成功,越来越多研究者将生成对抗网络应用在人脸表情识别技术中,并取得了显著的效果。有鉴于此,首先将对生成对抗网络原理进行介绍,然后从解决数据不足、姿态偏差、身份差异以及面部遮挡四个方面,对近年来生成对抗网络在人脸表情识别技术的研究进行分析和总结,最后展望生成对抗网络在人脸表情识别技术中的前景和发展方向。

**关键词:**人脸表情识别; 人机交互; 深度学习; 生成对抗网络

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 1673-629X(2021)0001-07

## Survey of Facial Expression Recognition Technology Based on Generative Adversarial Network

ZHI Wei-ji<sup>1</sup>, LIU Chun-yu<sup>1</sup>, ZHENG Wan-jun<sup>2</sup>, ZHANG Min<sup>1</sup>, ZHAN Si-yu<sup>1</sup>

(1. School of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen 518000, China;

2. Guangdong Polytechnic of Science and Technology, Zhuhai 519090, China)

**Abstract:** Facial expression recognition has been widely used in the fields of human-computer interaction, intelligent control, security and psychology, and the development of deep learning has made great progress in facial expression recognition technology. However, the face expression recognition technology still faces the problems of insufficient data, pose-invariant, identity-invariant and occlusion affects. In recent years, with the success in image synthesis, more and more researchers have applied the generative adversarial network to facial expression recognition technology and achieved remarkable results. In view of this, firstly, we will introduce the principle of generative adversarial network, and then analyze and summarize the research of generative adversarial network in facial expression recognition technology in recent years from four aspects of solving data shortage, pose-invariant, identity-invariant and facial occlusion. Finally, we look forward to the prospect and development direction of generative adversarial network in facial expression recognition technology.

**Key words:** facial expression recognition; human-computer interaction; deep learning; generative adversarial network

## 0 引言

面部表情是最有力和通用的人类传达情感和意图的信号,在人际交流中起到非常重要的作用<sup>[1]</sup>。与此同时,人脸表情识别技术(facial expression recognition, FER)在人机交互(human-computer interaction, HCI)、智能控制、自动驾驶等领域也扮演着重要的角色。著名的人脸表情识别系统由 Ekman and Friesen 于 20 世纪 70 年代提出,根据他们的研究,人类有六种主要情感,每种情感以唯一的表情来反映人的一种独特的心理活动。这六种情感被称为基本情感,由愤怒(anger)、高兴(happiness)、悲伤(sadness)、惊讶

(surprise)、厌恶(disgust)和恐惧(fear)组成<sup>[2]</sup>。随后 Ekman 和 Friesen 更进一步地细化及完善了面部表情,并提出了基于运动单元(AUs)来描述面部表情的面部动作编码系统(FACS)<sup>[3]</sup>。人脸表情识别主要有三大步骤:预处理、特征提取以及分类。传统的人脸表情识别技术大多数用的都是人工特征(handcrafted features)<sup>[4]</sup>,如梯度方向直方图特征(histograms of oriented gradients, HOG)<sup>[5]</sup>、局部二值模式特征(local binary pattern, LBP)<sup>[6]</sup>、尺度不变特征变换(scale invariant feature transform, SIFT)<sup>[7]</sup>等。随着人脸表情识别技术逐渐从实验室数据集测试走向真实场景下的识

收稿日期:2020-08-31

基金项目:广东省自然科学基金(2016A030310305);广东省教育厅 2018 年科研项目(2018GKTSCX083)

作者简介:植炜基(1995-),男,硕士研究生,研究方向为深度学习、计算机视觉;通讯作者:张敏,教授,研究方向为计算机测控。

别以及计算机显卡性能的提高,网络结构设计优化等深度学习技术的不断发展,FER 的研究逐渐转向深度学习的方法<sup>[4]</sup>,并取得了许多显著的进展。例如,P. Liu, S. Han 等人提出的深度信念网络(deep belief network)<sup>[8]</sup>在 CK+和 JAFFK 数据库上准确率都达到了当前最高水平;T. Devries 等人提出的多任务学习网络(multi-task learning)<sup>[9]</sup>也分别在 TED 库和 FER2013 库取得了很好的效果。尽管深度学习技术使得人脸表情识别技术得到飞快发展,但在实际应用方面 FER 仍面临着许多问题,Sariyanidi<sup>[10]</sup>提出的目前 FER 所面临的几大挑战,包括人脸姿态偏差、身份差异、人脸遮挡等,以及现存数据库样本不足造成深度学习网络的过拟合现象,都极大地影响了人脸表情识别的准确率。

为此,研究者通过不断改进网络结构和算法来解决当前所遇到的问题,如网络输入多样性<sup>[11]</sup>,将图像强度和 LBP 代码映射到三维度量空间作为 CNN 网络的输入;多网络集合<sup>[12]</sup>,将从不同网络中获取的单一特征向量集合在一起描述输入图像;网络级联<sup>[13]</sup>,首先训练深度信念网络来检测人脸以及与人脸表情相关的区域,然后用自动编码器对这些人脸的部位进行分类。而生成对抗网络(GAN)作为近年来复杂分布上无监督学习最具前景的方法之一,受到了研究者的广泛关注,几乎每周都有关于 GAN 的论文发布,在图像生成、图像风格迁移、图像域的转换等领域取得了重大的突破。因此,越来越多的研究者将 GAN 利用在人脸表情识别领域,以解决当前 FER 所遇到的各种挑战。

文中将对近年来基于生成对抗网络的人脸表情识别技术进行综述,介绍 GAN 的基本原理,总结和分析生成对抗网络如何处理和解决 FER 中数据样本不足、面部姿态、身份差异以及面部遮挡等难题,并对 GAN 在人脸表情识别技术中的发展前景和方向进行展望。

## 1 生成对抗网络 (generative adversarial network, GAN)

自人工智能概念诞生以来,人工智能的技术和理论都得到了不断发展,已经广泛应用在教育、交通、金融、医疗、娱乐等人类工作和生活的方方面面。机器学习是人工智能研究领域中的重要分支,根据学习过程中的不同经验,机器学习算法又可以分为无监督学习和监督学习。无监督学习是实现人工智能的重要一环,而生成对抗网络是现阶段无监督学习最热门的技术,近年来受到了研究者的广泛关注,在图像生成、风格转移等许多领域都有显著的成果<sup>[14]</sup>。

### 1.1 GAN 的基本原理

生成对抗网络是 Goodfellow<sup>[15]</sup>等人在 2014 年来自博弈论中的“二人零和博弈(即二人的利益之和为零,一方的所得正是另一方的所失)”中受到启发而提出的。GAN 网络主要由两个网络构成,生成器 G (Generator)和判别网络 D (Discriminator)。生成器 G 的目标是学习到输入数据的分布从而生成非常真实的图像,而判别器 D 的目标是正确辨别出真实图片和 G 生成的图片直接的差异。生成器 G 和判别器 D 都有各自的网络结构和不同的输入,其中生成器 G 的输出,即生成样本也是 D 的输入之一,而判别器 D 则会为 G 提供梯度进行权重的更新,这两个网络通过不断的改善和优化来提高自身网络的生成能力和判别能力<sup>[16]</sup>。GAN 的网络结构如图 1 所示,生成器 G 和判别器 D 的输入分别是随机噪声  $z$  和真实样本  $x$ ,  $G(z)$  则为生成器生成的尽量服从真实样本分布的人工样本,判别器的输入越接近真实样本,输出结果越接近 1;反之,越接近 0。当判别器的输入为真实样本时,输出结果为 1;当输入为  $G(z)$  时,输出结果为 0。这里 D 的目标是对输入样本做二分类判别:真(来源于真实样本  $x$  的分布)或者假(来源于生成器生成的假样本  $G(z)$ ),而 G 的目标是使自己生成的假数据  $G(z)$  在 D 上的判别结果  $D(G(z))$  和真实样本  $x$  在 D 上的判别结果  $D(x)$  一致。这两个相互对抗并迭代优化的过程使得 D 和 G 的性能不断提升,当最终双方的博弈达到纳什均衡时,生成器学到了真实样本的分布,而判别器也无法正确判别输入样本的来源了<sup>[17]</sup>。

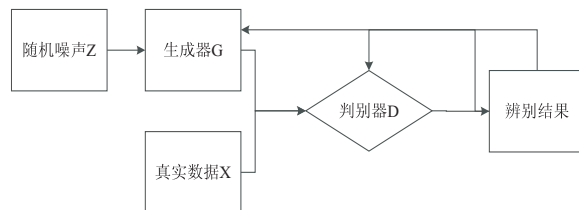


图 1 GAN 网络结构

### 1.2 GAN 的优势

生成对抗网络(GAN)的概念虽然在 2014 年才被提出来,但是由于其独特的优势和结构,在短短的几年时间里不断被研究人员所追捧,成为了深度学习领域最火的一个概念,许多学者认为 GAN 的出现大大推动了人工智能向无监督学习发展的进程。生成对抗网络及其变种已然成为了近年来机器学习领域最令人激动的想法,几乎每周都有关于 GAN 的全新论文发表,产生了种类繁多的 GAN 变形模型(如条件 GAN<sup>[18]</sup>、循环 GAN<sup>[19]</sup>等),来进一步改善模型。与其他生成模型相比,GAN 具有以下优势:

传统的生成模型一般需要进行复杂的马可夫链式

的采样和推断,而 GAN 作为一种新型的生成式方法避免了反复应用马尔可夫链学习机制带来的配分函数计算,不需变分下限也不需近似推导<sup>[15]</sup>,只用到了反向传播,极大改善了生成模型的训练难度和训练效率。GAN 的设计框架非常灵活,可以针对不同的任务设计出不同类型的损失函数,由于其独特的对抗训练机制,使得在数据的概率密度不可知的情况下,依然可以使用。相比其他所有模型,GAN 可以产生更加清晰和真实的样本。GAN 可以对生成样本进行并行采样,并且可以一次性生成数据,从而大大加快了采样的速度。

## 2 基于生成对抗网络的人脸表情识别

由于生成对抗网络近年来的飞速发展及其独特的优势,其在图像生成、风格迁移等领域取得了显著的成果。因此研究人员开始将生成对抗网络用于人脸表情识别技术,试图借助 GAN 来解决目前人脸表情识别技术所遇到的挑战和瓶颈,如数据样本不足,人脸面部姿态偏差、身份差异以及人脸遮挡等。下面将对近年来基于 GAN 网络的人脸表情识别技术进行介绍和分析,总结 GAN 网络对 FER 技术的发展以及对解决 FER 当前所遇到的挑战所起到的重要作用。

### 2.1 用于数据增强

随着深度学习的发展,神经网络所需要的参数也越来越多,达到了数以百万计,而要使得这些参数可以正确地进行工作则需要大量的数据进行训练,但是实际情况中获取大量的数据是一件很困难的事情。人脸表情识别就是典型的需要大量人脸表情图像进行训练,但是大多数公众可用的 FER 数据库存在数据不足以及数据量不平衡等问题。从实验室获取到的人脸表情的图像质量和数量都无法满足训练一个很好的识别网络的要求,同时,某些表情类别,如厌恶(disgust),比起其他类别在数量上会少很多,造成了数据量不平衡。再者就是由于人脸表情分类是一个很精细的工作,某些细微的变化都会造成分类结果的不同,使得图像标注工作的难度大大增加<sup>[20]</sup>。因此,数据增强对于 FER 来说必不可少。

最常见的方法是对原始图像进行旋转、缩放、剪切等几何变化来增加数据样本<sup>[21]</sup>,利用水平翻转以及每张图片转动六个不同的角度,使得数据量扩充了 14 倍。也可以在图像中添加噪声,如高斯噪声、椒盐噪声等<sup>[22]</sup>,对原始图像进行空间变换、像素值变换,从而达到数据增强的效果。针对图像颜色空间的数据增强也是一种常见方法,包括亮度、对比度调节等<sup>[23]</sup>。但是这些数据增强的效果并不理想,而且未能解决数据量不平衡等问题。因此许多研究人员开始寻求 GAN 网络来实现 FER 的数据增强。

Xinyue Zhu 等人<sup>[24]</sup>提出了使用循环生成对抗网络(CycleGAN)来产生数据量较少的表情类别图片,达到数据增强的效果,同时使用了 LSGAN 的最小二乘法损失(least-square loss)结合其原本的对抗性损失(adversarial loss)来避免梯度消失的问题。图 2 为 CycleGAN 的结构图,参照图片和目标图片都是从原先的数据集中收集而来并输入到 CycleGAN 中分别作为 R 域和 T 域,G 和 F 是两个生成器,分别将 R 转化为 T 和将 T 转化为 R,同时对应着两个判别器 D(R)和 D(T),分别区别 R 域的真实图像与 F 生成的假的 R 域图像以及 T 域的真实图像和 G 生成的假的 T 域图像。而 G 生成的图像最后要与 T 域的图像相像,且判别器 D(T)无法辨别出,并最后作为补充数据与原始数据一起输入到分类器里进行分类,从而达到数据增强的效果。

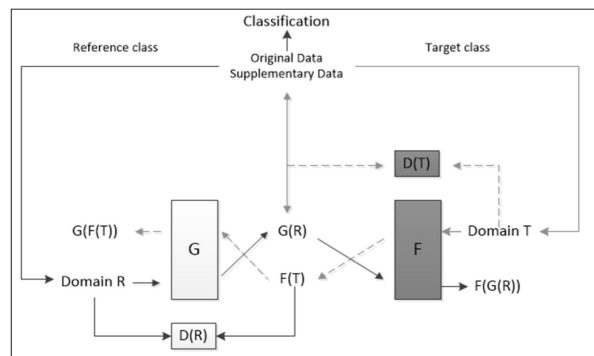


图 2 循环生成对抗网络的结构

Feifei Zhang 等人<sup>[25]</sup>提出了一个 CycleAT 模型来同时实现人脸表情图片的生成和识别,将自然环境下的 FER 问题看作是域适应的问题,文章用实验室控制下的数据作为源域,自然条件下的人脸表情数据库作为目标域,并且使用大量网上未做标注的人脸表情图片作为辅助域,通过 GAN 来产生大量标注的自然条件下的人脸表情图像,解决了 FER 在自然环境下的识别准确率的问题。

Yuchi Huang 等人<sup>[26]</sup>提出 DyadGAN 模型,使用两个条件生成对抗网络(conditional GAN)来生成两个人交流时的人脸表情图像,与以往使用面部属性的条件不同,这个模型将重点放在一个人的表情对另外一个人表情的关系和影响上。第一步利用被采访者的人脸表情作为条件来生成表情草图,第二步利用第一步生成的草图作为条件生成采访者的人脸表情图像。

Hui Ding 等人<sup>[27]</sup>提出了一个 ExprGAN 模型,通过对人脸表情图像进行编辑生成各种不同的表情类型以及一系列连续的表情强度,可应用于数据增强。编码器 G(enc)将图像  $x$  映射为  $g(x)$  来保存图像的身份特征,表情控制模块 F 将 one-hot 的表情标签  $y$  转换为表情编码 C,编码 C 可以更准确地表达表情的强度,最后



利用  $G(\text{dec})$  生成器生成想要的表情类型。

## 2.2 用于面部姿态偏差

心理学研究表明:头部偏转能够干扰人类的情感认知<sup>[28]</sup>。同时,面部姿态的偏差,对于 FER 系统的人脸特征点的提取和分析也造成了很大的阻碍和误差。目前大多数 FER 系统训练和测试的都是实验数据,这些人脸表情数据大多都是正脸的,而在实际生活中,人脸姿态会随时发生偏转,这给 FER 的实际应用造成了极大的困扰。因此,非正面人脸表情识别技术受到了越来越多人的关注。2017 年 IEEE 国际自动人脸与手势识别会议,就将头部偏转状态下的表情识别作为第三届人脸表情识别与分析挑战赛的主题<sup>[29]</sup>。目前的方法有依靠二维或三维几何模型跟踪人脸表情图像的关键点;有获取全局或局部表情特征,以外貌信息代替关键点信息;建立不同姿态样本之间的联系,借助一种姿态信息识别出另一种姿态的表情信息<sup>[29]</sup>。以上方法虽然取得了不少成果,但是在准确率和成本上仍有很大的进步空间。下面介绍使用 GAN 来进行非正面表情识别的方法,希望能给该领域带来一些新的启发。

Ying-Hsiu Lai 等人<sup>[30]</sup>利用 GAN 把非正脸的图像生成成正脸的图像,并保留了身份信息 and 表情特征,与此同时,它能够利用学习到的情感保留表征来预测输入图像的表情标签。文章所用的 GAN 的生成器采用的是编码-解码的结构模型,包括编码器  $G(\text{enc})$  和解码器  $G(\text{dec})$ ,  $G(\text{enc})$  用做学习非正脸图像的情感保留表征,而  $G(\text{dec})$  用做从学到的特征中生成正脸图像,同时  $G(\text{enc})$  提取到的情感保留表征,还能直接用作表情识别,从而实现生成器的多任务学习。而判别器  $D$  则用做区别真实的正脸图像和生成的正脸图像,经过多次迭代训练使得生成的正脸图像的质量不断提高。

Rui Huang 等人<sup>[31]</sup>提出了一个双通道生成对抗网络 TP-GAN(two-pathway GAN),利用 Local Pathway 网络学习人脸的局部特征,网络输入为侧脸图像的四个局部块,包括嘴巴、鼻子和左右眼睛,而输出的是正面的嘴巴、鼻子和眼睛。与此同时,另外一个通道是 Global Pathway,利用其学习人脸的全局特征,如人脸轮廓等底层信息,网络输入为整张侧脸图像,输出为与输入大小相同的图像,最后将局部和全局的特征结合,得到生成的正脸图像,之后再将生成图像输入判别网络。

Luan Tran 等人<sup>[32]</sup>提出 DR-GAN(disentangled representation GAN),通过编码-解码结构的生成器能够正脸化和旋转一张任意姿态的输入图像。 $G(\text{enc})$  学习输入图像的特征表示  $f(x)$ ,  $f(x)$  与姿态编码  $C$  以及一个随机噪声  $Z$  一起输入到  $G(\text{dec})$ ,生成图像并将其

输入到判别器  $D$  中。与传统 GAN 的判别器不同的是,文章所用的  $D$  是一个多任务 CNN,包含身份分类和姿态分类。除了输入一张图像的 DR-GAN,还提出了一个扩展模型,多图像输入的 DR-GAN。多图像输入的 DR-GAN 在生成器上有所不同,除了提取特征表示  $f(x)$  外,还计算了每张图片的置信系数。

Xi Yin 等人<sup>[33]</sup>提出了一个 FF-GAN 模型,结合 3D Morphable Model<sup>[34]</sup> 和 GAN 来进行人脸姿态转正。3DMM Coefficients 得到输入人脸的姿态信息,通过 3D 模型可以得到全局的信息,但是缺少细节信息,而原始的输入图像含有细节信息,所以将两者相结合输入到生成器中便可以产生含有细节的正脸图像。

## 2.3 用于身份差异

身份差异(identity variations)也是人脸表情识别技术中的一大难题,尤其是处理小规模样本数据库的时候。因为每个人在年龄、种族、性别、文化背景以及外表和面部动作风格等方面都会有所不同,所以面部表情也就因人而异了。文献[35]指出由于身份差异,同一个人做不同表情之间的差异小于不同人做同一个表情的差异,这就给表情识别甚至是表情标注都带来了很大的困扰。如图 3 所示,测试者 A 的开心表情和伤心表情的差异要小于测试者 A 与测试者 B 同样做开心表情之间的差异。最近也有许多研究去尝试解决这个问题<sup>[36-38]</sup>,有的用迁移学习的方法,有的使用更深层的网络结构和更多的训练数据<sup>[39]</sup>,有的使用特定的模块去限制身份差异对 FER 的扰动<sup>[40]</sup>。虽然这些方法在一定程度上提升了识别准确率,但是身份差异对 FER 的影响依然存在。

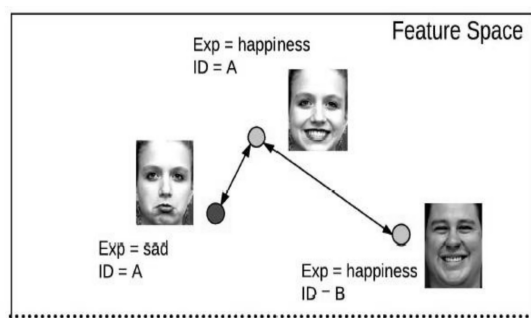


图 3 身份差异大于表情差异

Huiyuan Yang 等人<sup>[35]</sup>提出了一个 IA-gen 的模型(identity-adaptive generation),用输入的图像重新生成表情图像来解决身份差异的问题。模型分为两层,上层部分用 cGANs 来生成同一个人的不同表情,作为该主体的子空间,下层部分在不涉及其他主体的情况下,只用该主体的输入图像以及生成的子空间进行人脸表情识别,因此身份差异的扰动可以得到很好的缓解。

Jiawei Chen 等人<sup>[41]</sup>提出了 PPRL-VAGN,结合了变分编码器(VAE)<sup>[42]</sup> 和 GAN 来学习摆脱了身份信息

的图像的特征,并且可以生成保留了表情信息的新图像。给定一张带有身份标签  $y(id)$  和表情标签  $y(e)$  的人脸图像  $I$ ,通过 encoder 来学习人脸图像的特征  $f(I)$ ,  $f(I)$  与身份编码串联,通过 decoder 生成表情相同但身份信息不同的新图像,并且一共有三个判别器,分别判别输入图像  $I$  的真伪,估计输入图像的人物身份以及对输入图像的表情进行分类。

Jie Cai 等人<sup>[43]</sup>提出了一个 IF-GAN 网络去减弱身份偏差对人脸表情识别的干扰。通过输入任意身份的表情图片和一张平均脸的中立表情图片到 GAN 中进行训练,生成出平均脸的各种表情,这样训练好的网络就可以输入任意身份的表情图片,生成出平均脸的对应该表情图片,再用来进行表情识别,从而就可以过滤掉身份信息,只保留了表情信息,避免了身份偏差对 FER 的干扰。

Kamran Ali 等人<sup>[44]</sup>提出了 DE-GAN 网络,将人脸表情特征从身份信息中提取和解耦出来。DE-GAN 将人脸表情图像、噪声以及身份编码作为输入,输出人脸表情特征以及不同身份的相应表情图像,其中人脸表情特征可用作 FER,而输出图像的人脸身份由输入的身份编码决定。

Huiyuan Yang 等人<sup>[45]</sup>提出了 DeRL 模型,因为人脸表情是表情成分和中立成分的结合,所以给一张任意表情的图像作为模型的输入,其对应的中立表情可以由训练好的 GAN 生成,通过这个步骤,表情部分的信息就被过滤出来存放到生成器中。这样不管是否有身份偏差,通过模型都可以把表情成分提取出来用作 FER。

Youngsung Kim 等人<sup>[46]</sup>利用对比表征学习的方法来解决 FER 中的身份偏差。当看到一张模棱两可的表情时,人脑会将其与他以前看到过的表情进行比较,从而提取它们之间的差异。受此启发,输入真实图像  $X$  到生成网络中,生成出参考图像,然后将两者输入到对比编码网络,输出对比特征,最后进行分类。

## 2.4 用于人脸遮挡

根据人类在现实生活环境中复杂的面部表情,人脸遮挡问题成为了表情识别中的新挑战和研究热点<sup>[47]</sup>。人脸遮挡问题主要来自于在真实的自然环境下收集到的图像会存在口罩、墨镜、围巾或者是手部运动等遮挡<sup>[48]</sup>,这些局部遮挡的产生具有不可预测性,对人脸定位、特征提取以及分类识别都会造成很大的影响,而人脸遮挡在现实生活中处处可见,所以要想实现 FER 技术在现实生活中的更好利用,必须要解决遮挡人脸的表情识别问题。

对于有遮挡的人脸图像,首先考虑的就是想办法将遮挡部分补全使得整张图片看上去真实连贯,这样

才能方便特征提取,人脸定位等后续的工作。文献[49]就提出了一个基于深度生成模型的人脸修复算法,利用两个判别器,分别判断整个图像区域和缺失区域的真假,最后使用一个解析网络进一步完善缺失区域的生成图像。

Wang Suqin 等人<sup>[50]</sup>提出了一种基于生成对抗网络的遮挡人脸的表情识别算法,先对遮挡人脸图像填补修复,再进行表情识别。模型分为人脸填补和表情识别两个模块,人脸填补模块由生成器和判别器组成,表情识别模块在判别器的基础上,使用了其部分卷积层和池化层,将其作为特征提取器,额外加入两层全连接层和 Softmax 层构成了表情分类器。

WANG Haiyong 等人<sup>[51]</sup>提出了一种基于改进生成对抗网络的表情识别模型,先利用由自动编码器构成的生成器和两个判别器的对抗学习对遮挡人脸图像填补修复,再在全局判别器后面添加多分类层,利用全局判别器的部分卷积层并在后面添加多分类层构成表情分类器进行表情识别。由于有局部判别器和全局判别器两个判别器,所以模型在修复图像时能在全局和细节这两个层面上更好地恢复遮挡图像。

Yao Nai-Ming 等人<sup>[52]</sup>提出了一种对人脸局部遮挡图像进行用户无关表情识别的方法,该方法包括一个基于 Wasserstein 生成对抗网络(WGAN)的人脸图像生成网络,能够为图像中的遮挡区域生成上下文一致的补全图像;以及一个表情识别网络,能够通过表情识别任务和身份识别任务之间建立对抗关系来提取用户无关的表情特征并推断表情类别。

Qingyan Duan 等人<sup>[53]</sup>提出了 BoostGAN 来解决遮挡且非正脸的人脸图像的去遮挡和正脸化。BoostGAN 是第一个利用 GAN 同时解决正脸化和去人脸遮挡的方法,与以往基于 GAN 的人脸补全不同,BoostGAN 同时利用像素级和特征级信息作为监督信号从多个尺度上保护身份信息。经 BoostGAN 补全和修正的图像可以用于 FER,对提高 FER 准确率有很大帮助。

## 3 结束语

心理学家 ARusseli 指出,在生活的日常交流中,信息的传递主要由语言和表情两者完成,其中言语所传递出的信息仅仅占到 7%,声音传递的信息为 38%,而通过表情传递的情感信息高达 55%。由此可见,表情在人们的日常交际交流中有着至关重要的地位。首先,人脸表情识别是涉及到多种邻域的一个交叉学科,其中包括生物学、心理学、机器视觉、模式识别等领域。从科研角度上说,人脸表情识别的研究将会在某种程度上促进相关邻域上的发展,而其他邻域的发展也将

推动着人脸表情识别的发展以及对该技术的需求也将提高。作为人工心理理论研究和情感计算研究的重要内容,人脸表情识别的研究将赋予计算机理解人脸表情变化的能力,其将极大改善人机交互呆滞不前的现状。其次,人脸表情识别在智能监控、智能医疗、教育学习、游戏娱乐、广告媒体、人机交互等领域将有着广泛的应用。

GAN 是生成模型发展史上的里程碑,成功地解决了其他生成模型至今存在的问题。文中对 FER 技术做了一个简要的概述,对 GAN 的原理和优缺点也进行了简单的阐述,主要介绍了 GAN 网络在人脸表情识别技术的应用,从 FER 技术目前所面临的挑战入手,分别从 FER 的数据不足、姿态偏差、身份差异以及人脸遮挡这四个方面介绍 GAN 网络对 FER 技术的重要作用,简述了其中的算法结构和 GAN 相关变形结构的原理。

虽然利用 GAN 网络对人脸表情识别系统的性能有了很大的提高,但是仍存在着许多不足,有待进一步的研究和改进;提到的基于 GAN 的人脸表情识别系统都是针对静态图像进行的而实际生活中人的情绪变化有一定的时序性,且会在短时间内发生变化,静态图像只能反映某一时刻人的表情状态,后续的工作应该考虑利用 GAN 来研究针对视频序列的 FER。

文中研究的主要是针对七种基本表情,但是现实生活中人类的表情更加复杂,而且一张人脸表情图像往往会有两个或者多个表情标签,构成混合表情,所以往后基于 GAN 的人脸表情识别系统应该考虑混合表情的识别技术。

人脸表情图像往往会存在光照强度不同、背景环境不同,这些因素都会对 FER 的准确率带来很大的干扰,往后的 FER 系统应该更多地考虑将这些干扰因素,利用 GAN 网络去解决这些困扰。

#### 参考文献:

- [1] TIAN Yingli, TAKEO K, JEFFREY F C. Recognizing action units for facial expression analysis [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2001, 23(2): 97-115.
- [2] EKMAN P, FRIESEN W V. Constants across cultures in the face and emotion [J]. Journal of Personality and Social Psychology, 1971, 17(2): 124-129.
- [3] ROLLS E T, EKMAN P, PERRETT D I, et al. Facial expressions of emotion: an old controversy and new findings [J]. Philosophical Transactions of the Royal Society of London, 1992, 335(1273): 63-69.
- [4] LI S, DENG W. Deep facial expression recognition: a survey [J]. IEEE Transactions on Affective Computing, 2018(99): 1-6.
- [5] BALTRUŠAITIS T, MAHMOUD M, ROBINSON P. Cross-dataset learning and person-specific normalisation for automatic action unit detection [C]//2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG). Ljubljana: IEEE, 2015: 1-6.
- [6] SHAN C, GONG S, MCOWAN P W. Facial expression recognition based on local binary patterns: a comprehensive study [J]. Image and Vision Computing, 2009, 27(6): 803-816.
- [7] CHU W, DE LA TORRE F, COHN J F. Selective transfer machine for personalized facial action unit detection [C]//2013 IEEE conference on computer vision and pattern recognition. Portland, OR: IEEE, 2013: 3515-3522.
- [8] LIU P, HAN S, MENG Z, et al. Facial expression recognition via a boosted deep belief network [C]//2014 IEEE conference on computer vision and pattern recognition. Columbus, OH: IEEE, 2014: 1805-1812.
- [9] DEVRIES T, BISWARANJAN K, TAYLOR G W. Multi-task learning of facial landmarks and expression [C]//2014 Canadian conference on computer and robot vision. Montreal, QC: [s. n.], 2014: 98-103.
- [10] SARIYANIDI E, GUNES H, CAVALLARO A. Automatic analysis of facial affect: a survey of registration, representation, and recognition [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(6): 1113-1133.
- [11] LEVI G, HASSNER T. Emotion recognition in the wild via convolutional neural networks and mapped binary patterns [C]//International conference on multimodal interaction. Seattle, Washington, USA: ACM, 2015: 503-510.
- [12] BARGAL S A, BARSOUM E, FERRER C C, et al. Emotion-recognition in the wild from videos using images [C]//Proceedings of the 18th ACM international conference on multimodal interaction. Tokyo, Japan: ACM, 2016: 433-436.
- [13] FATHALLAH A, ABDI L, DOUIK A. Facial expression recognition via deep learning [C]//2017 IEEE/ACS 14th international conference on computer systems and applications (AICCSA). Hammamet: IEEE, 2017: 745-750.
- [14] 伍 辉. 生成式对抗网络的研究进展综述 [J]. 工业控制计算机, 2019, 32(7): 70-71.
- [15] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]//27th international conference on neural information processing systems. Cambridge, MA, United States: [s. n.], 2014: 2672-2680.
- [16] 张倩宇, 宋彩芳. 生成式对抗网络研究综述 [J]. 电子技术与软件工程, 2019(4): 13-14.
- [17] 余胜男. 基于生成对抗网络的人脸表情识别方法研究 [D]. 合肥: 合肥工业大学, 2019.
- [18] MIRZA M, OSINDERO S. Conditional generative adversarial nets [J]. arXiv: 1411. 1784, 2014.
- [19] ZHU J, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]//



- 2017 IEEE international conference on computer vision (ICCV). Venice; IEEE, 2017; 2242–2251.
- [20] NG H W, NGUYEN D, VONIKAKIS V, et al. Deep learning for emotion recognition on small datasets using transfer learning [C]//International conference on multimodal interaction. Seattle, USA; ACM, 2015.
- [21] JUNG H, LEE S, YIM J, et al. Joint fine-tuning in deep neural networks for facial expression recognition [C]//2015 IEEE international conference on computer vision (ICCV). Santiago; IEEE, 2015; 2983–2991.
- [22] ZAVAREZ M V, BERRIEL R F, OLIVEIRA – SANTOS T. Cross-database facial expression recognition based on fine-tuned deep convolutional network [C]//2017 30th SIBGRAPI conference on graphics, patterns and images (SIBGRAPI). Niteroi; [s. n.], 2017; 405–412.
- [23] SANG D V, CUONG L T B, THUAN D P. Facial smile detection using convolutional neural networks [C]//2017 9th international conference on knowledge and systems engineering (KSE). Hue; [s. n.], 2017; 136–141.
- [24] ZHU X, LIU Y, LI J, et al. Emotion classification with data augmentation using generative adversarial networks [M]//Advances in knowledge discovery and data mining. Melbourne, VIC, Australia; Springer, 2018; 349–360.
- [25] ZHANG Feifei, ZHANG Tianzhu, MAO Qirong, et al. Facial expression recognition in the wild: a cycle-consistent adversarial attention transfer approach [C]//Proceedings of the 26th ACM international conference on multimedia. Seoul Republic of Korea; ACM, 2018; 126–135.
- [26] HUANG Y, KHAN S M. DyadGAN: generating facial expressions in dyadic interactions [C]//2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW). Honolulu, HI; IEEE, 2017; 2259–2266.
- [27] DING H, SRICHARAN K, CHELLAPPA R. Ex-prGAN: facial expression editing with controllable expression intensity [J]. arXiv: 1709. 03842, 2017.
- [28] NICHOLLS M E R, WOLFGANG B J, CLODE D, et al. The effect of left and right poses on the expression of facial emotion [J]. *Neuropsychologia*, 2002, 40 (10): 1662–1665.
- [29] 蒋 斌, 甘 勇, 张焕龙, 等. 非正面人脸表情识别方法综述 [J]. *计算机科学*, 2019, 46 (3): 53–62.
- [30] LAI Y, LAI S. Emotion-preserving representation learning via generative adversarial network for multi-view facial expression recognition [C]//2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018). Xi'an; IEEE, 2018; 263–270.
- [31] 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]//IEEE international conference on computer vision (ICCV). Venice, Italy; IEEE, 2017; 2458–2467.
- [32] TRAN L, YIN X, LIU X. Disentangled representation learning GAN for pose-invariant face recognition [C]//2017 IEEE conference on computer vision and pattern recognition (CVPR). Honolulu, HI; IEEE, 2017; 1283–1292.
- [33] YIN Xi, YU Xiang, SOHN K. Towards large-pose face frontalization in the wild [J]. arXiv: 1704. 06244, 2017.
- [34] BLANZ V, VETTER T, ROCKWOOD A. A morphable model for the synthesis of 3D faces [J]. *ACM SIGGRAPH*, 1999 (7): 187–194.
- [35] YANG H, ZHANG Z, YIN L. Identity-adaptive facial expression recognition through expression regeneration using conditional generative adversarial networks [C]//2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018). Xi'an; IEEE, 2018; 294–301.
- [36] ZHANG C, WANG P, CHEN K, et al. Identity-aware convolutional neural networks for facial expression recognition [J]. *Journal of Systems Engineering and Electronics*, 2017, 28 (4): 784–792.
- [37] ZHAO X, LIANG X, LIU L, et al. Peak-piloted deep network for facial expression recognition [C]//ECCV 2016. Amsterdam, The Netherlands; Springer, 2016; 425–442.
- [38] DING H, ZHOU S K, CHELLAPPA R. FaceNet2ExpNet: regularizing a deep face recognition net for expression recognition [J]. arXiv: 1609. 06591, 2016.
- [39] MOLLAHOSSEINI A, CHAN D, MAHOOR M H. Going deeper in facial expression recognition using deep neural networks [C]//2016 IEEE winter conference on applications of computer vision (WACV). Lake Placid, NY; IEEE, 2016; 1–10.
- [40] CHEN J, LIU X, TU P, et al. Learning person-specific models for facial expression and action unit recognition [J]. *Pattern Recognition Letters*, 2013, 34 (15): 1964–1970.
- [41] CHEN J, KONRAD J, ISHWAR P. VGAN-based image representation learning for privacy-preserving facial expression recognition [C]//2018 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW). Salt Lake City, UT; IEEE, 2018; 1651.
- [42] KINGMA D P, WELING M. Auto-encoding variational Bayes [J]. arXiv: 1312. 6114, 2013.
- [43] CAI J, MENG Z, KHAN A S, et al. Identity-free facial expression recognition using conditional generative adversarial network [J]. arXiv: 1903. 08051, 2019.
- [44] ALI K, HUGHES C E. Facial expression recognition using disentangled adversarial learning [J]. arXiv: 1909. 13135, 2019.
- [45] YANG Huiyuan, CIFTCI U, YIN Lijun. Facial expression recognition by de-expression residue learning [J]. *International Journal on Computer Science & Engineering*, 2018, 2 (5): 2220–2224.
- [46] KIM Y, YOO B I, KWAK Y, et al. Deep generative-contrastive networks for facial expression recognition [J]. arXiv: 1703. 07140, 2017.
- [47] CHENG Y, JIANG B, JIA K. A deep structure for facial ex-