基于风格迁移的面部表情识别方法

肖世明[1] 章思远[1] 毛政翔[1] 黄 伟[1,2]

[1](南昌大学信息工程学院 江西 南昌 330031)

[2](南昌大学信息化办公室 江西 南昌 330031)

摘 要 针对一个人的面部表情可以被分解为表情成分和中性成分,提出一种新颖的基于风格迁移的面部表情识别方法。该方法通过训练循环一致生成对抗网络(Cycle-GAN)得到不同的生成器;这些不同的生成器可将不同的表情迁移到中性,因此每个生成器对应一种不同的表情。在测试阶段,输入表情图像到上述训练好的生成器中。由于只有与输入的表情对应的生成器能迁移成中性表情,因此可以通过这种方式实现面部表情识别。实验结果表明:该新方法不仅在实验室条件下获得的面部表情数据集中表现突出,而且在自然条件下获得的面部表情数据集中也有非常高的识别率。

关键词 面部表情识别 风格迁移 支持向量机 图像生成

中图分类号 TP3 **文献标志码** A DOI:10.3969/j.issn.1000-386x.2023.02.027

A FACIAL EXPRESSION RECOGNITION METHOD BASED ON STYLE TEANSFER

Xiao Shiming[1] Zhang Siyuan[1] Mao Zhengxiang[1] Huang Wei[1,2]

[1](*College of Information Engineering*, *Nanchang University*, *Nanchang 330031*, *Jiangxi*, *China*)

[2](*Information Office*, *Nanchang University*, *Nanchang 330031*, *Jiangxi*, *China*)

Abstract The facial expression of one individual person can be decomposed into two components: the expression component and the neutral component. This paper proposes a facial expression recognition method based on style transfer. In this method, different generators were obtained by training Cycle-GAN, and these different generators could migrate different expressions to neutral, so each generator corresponded to a different expression. In the testing stage, facial images were input into the trained generators. Since only the generator corresponding to the input expression could be migrated to a neutral expression, facial expression recognition could be realized in this way. The experimental results show that this method not only performs well in the facial expression dataset obtained under laboratory conditions, but also has a very high recognition rate in the facial expression dataset obtained under natural conditions.

Keywords Facial expression recognition Style transfer Support vector machine Image synthesis

0 引 言

人脸面部表情是最直接、最有效的情感表达方式。Mehrabian 等[1]做过研究表明,在人类日常交流的主要方式和途径中,传递信息最多的是人脸表情,其次是声音和语言,传递信息量分别占信息总量比重是 55%、38% 和 7%。Ekman 等[2]提出了人类共有的六类基本表情:生气、害怕、厌恶、开心、悲伤、惊讶,而其他复杂的表情都是在此基础上复合而成,例如惊喜就是惊讶加上开心,这也成为研究者研究人脸表情分类的共识。人脸表情识别可以应用于诸多领域,

收稿日期:2020-07-26。国家自然科学基金项目(61862043);江西省自然科学基金重点项目(20181ACB20006)。**肖世明**,硕士生,主研领域:计算机视觉,图像处理。**章思远**,硕士生。**毛政翔**,硕士生。**黄伟**,副教授。

如人机交互实时表情识别、驾驶员疲劳检测、谎言检测等。

面部表情识别的方法一般步骤有:人脸图像信息获取、图像预处理、图像特征表示与提取、特征分类器的训练。其中表情图像特征的提取成为影响表情识别率的关键因素。面部表情特征的提取由手工提取特征[3-5,13]到浅层学习提取特征[6-7],再到如今运用广泛的深度学习提取特征[8-12]。不管是手工提取特征或者是使用深度神经网络提取特征,表情的识别率都受到身份、性别、年龄等属性的影响。Zhang 等[11]提出的 IACNN 方法考虑了身份信息等因素对面部表情识别的影响。

研究表明,对于同一个人,可以通过比较他的当前表情和他的中性表情来判断他的表情[15]。也就是说一个人的面部表情可以分解为表情成分和中性成分[16]。根据这一研究,Kim 等[17]和 Lee 等[18]利用需判断的表情图像和对应的中性表情图像的特征差异或者图像差异来识别面部表情。然而在这些工作中,识别面部表情的个体对应的中性表情都是可以直接获得的。而在现实情况下,并不是每个给定个体对应的中性表情都是可以直接获得的。为了解决这一问题,需要构建一个根据给定表情图像生成中性表情图像的生成器,并且该生成器不会改变个体的身份信息。随着生成对抗网络(Generative Adversarial Networks,GAN)[19]的提出,越来越多基于 GAN 的图像生成方法被提出。一个 GAN 模型包括生成器和判别器两部分,通过生成器和判别器之间的对抗训练生成逼真的图像。原始的 GAN 模型使用一个随机向量作为输入,缺少必要的约束,这使得生成的图像质量参差不齐。因此,在原始 GAN 模型的基础上,Zhu 等[20]提出循环一致生成对抗网络(Cycle-consistent Generative Adversarial Networks,Cycle-GAN)。Cycle-GAN 可以完成非成对图像的图像到图像的风格迁移。该网络使用源空间图像而非随机变量作为输入,通过结合循环一致损失和对抗损失共同训练模型,可以无监督地学习源空间到目标空间的映射。Cycle-GAN 被广泛地应用于图像风格迁移。

根据上述分析,本文提出一种基于风格迁移的面部表情聚类识别方法。使用 Cycle-GAN 训练不同表情的生成器,将任意给定表情图像迁移到对应中性表情图像,生成器和表情一一对应。如图 1 所示,给定含有各种随机表情的人脸图像,通过中性表情生成器将含有表情的图像迁移到中性表情图像。在这一过程中,生成的中性表情图像不会改变原图像的身份信息,同时生成器会学习到不同表情的成分,即将不同表情的表情成分"存储"在对应的生成器中。

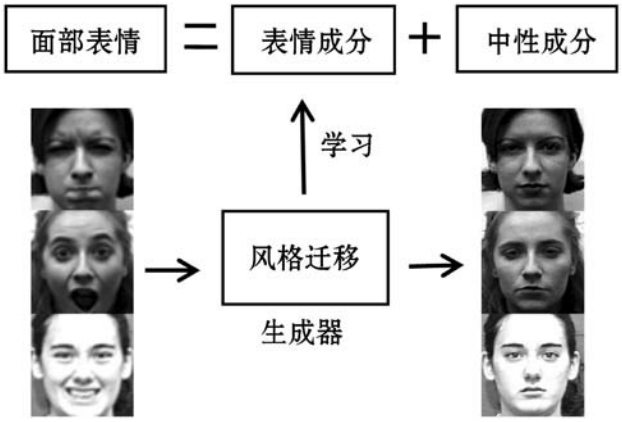

图 1　表情风格迁移示意图

输入一幅表情图像到不同个数(取决于数据集表情类别数)的生成器中生成一组图像,由于不同的生成器学习到不同的表情成分,所以这组图像中,只有学习到输入图像的表情成分的生成器可以成功将该表情图像迁移为中性表情。而其他生成的图像则保留原表情信息,因为这些生成器中不具有该表情成分,故而不能将该表情图像迁移为中性表情图像。所以只要找到迁移为中性表情对应的含有该表情成分的生成器即可识别出该图像的表情类别。在生成的图像中,表情标签只有中性和输入图像的表情标签,这样将面部表情识别的多分类问题转化为二分类问题。本文使用支持向量机(Support Vector Machine,SVM)作为分类器。在做分类任务时,使用 CNN 卷积神经网络提取面部表情特征用于分类。

1　基于风格迁移的中性表情生成

1.1　Cycle-GAN

Cycle-GAN 模型结构如图 2 所示,包含两个生成器和两个判别器,分别为 G_{AB}、G_{BA}、D_A、D_B。Cycle-GAN 可以通过学习源域(Source Domain)A 与目标域(Target Domain)B 之间的映射关系,从而完成图像到图像的风格迁移。生成器 G_{AB}学习从源域到目标域的映射 $f1:A\rightarrow B$,生成器 G_{BA}学习从目标域到源域的映射 $f2:B\rightarrow A$。判别器 D_A、D_B分别用来判断各自输入的图像是否为源域 A、目标域 B 的真实图像。

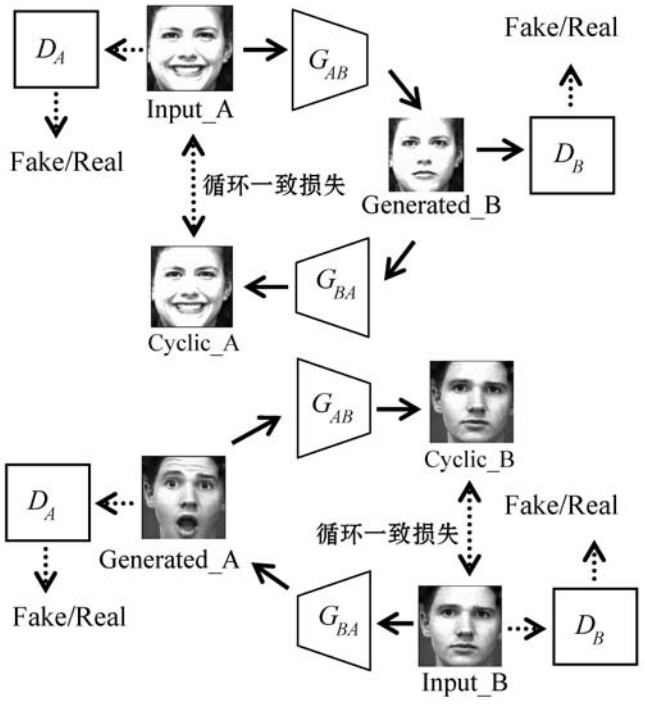

图 2 Cycle-GAN 的网络结构

Cycle-GAN 的损失函数由两部分组成：

1）生成对抗损失(Generative Adversarial Loss)：

$$L_{\mathrm{GAN}}(G_{AB},D_B,A,B)=E_{b\in P_B}[\log D_B(b)]+E_{a\in P_A}[\log(1-D_B(G_{AB}(a)))] \tag{1}$$

式(1)为 $A\rightarrow B$ 的生成对抗损失函数，a、b 分别为来自源域 A、目标域 B 的图像。生成器 G_{AB}尽可能生成与目标域 B 逼近的图像 $G_{AB}(a)$，判别器 D_B则判断输入图像是否为真实的目标域 B 图像。

$$L_{\mathrm{GAN}}(G_{BA},D_A,B,A)=E_{a\in P_A}[\log D_A(a)]+E_{b\in P_B}[\log(1-D_A(G_{BA}(b)))] \tag{2}$$

式(2)为 $B\rightarrow A$ 的生成对抗损失函数。生成器 G_{BA}尽可能生成与源域 A 逼近的图像 $G_{BA}(b)$，判别器 D_A则判断输入图像是否为真实的源域 A 图像。

2）循环一致损失(Cycle Consistency Loss)。

只有生成对抗损失是无法训练模型的，因为根据上述损失函数，生成器 G_{AB}可以将所有的源域 A 图像都映射为目标域 B 的同一幅图像。例如可以将所有高兴表情转换为中性表情，但是这些中性表情都是同一个人。同理，生成器 G_{BA}也有同样的问题。所以在 Cycle-GAN 中引入了循环一致损失函数 L_{cyc}，公式如下：

$$L_{\mathrm{cyc}}(G_{AB},G_{BA})=E_{a\in P_A}[\|G_{AB}(G_{BA}(b))-a\|_1]+E_{b\in P_B}[\|G_{AB}(G_{BA}(b))-a\|_1] \tag{3}$$

生成器 G_{AB}和 G_{BA}分别学习 f_1和 f_2两个映射的同时，要求 $G_{AB}(G_{BA}(b))\approx b$ 以及 $G_{BA}(G_{AB}(a))\approx a$。即目标域 B 的图像 b，经过 f_2映射得到图像 $G_{BA}(b)$，再经过 f_1映射得到的图像 $G_{AB}(G_{BA}(b))$，两者之间要尽可能相似。同样，对于源域 A 的图像 a，经过 f_1映射得到图像 $G_{AB}(a)$，再经过 f_2映射得到的图像 $G_{BA}(G_{AB}(a))$，两者之间要尽可能相似。这样就保证了在两个域之间的图像转换不会映射为同一幅图像。

最终网络的所有损失加起来为：

$$L(G_{AB},G_{BA},D_A,D_B)=\lambda_1 L_{\mathrm{GAN}}(G_{AB},D_B,A,B)+\lambda_2 L_{\mathrm{GAN}}(G_{BA},D_A,B,A)+\lambda_3 L_{\mathrm{cyc}}(G_{AB},G_{BA}) \tag{4}$$

式(4)中 λ_1、λ_2、λ_3分别为调节生成损失、对抗损失和循环一致损失所占权重的超参数。在所有的损失函数中，对于生成器来说需要最小化损失函数，对于判别器来说需要最大化损失函数。

1.2 中性表情生成方法

根据前文分析，面部表情的识别可以通过比较表情图像和对应的中性表情图像之间的不同完成。Cycle-GAN 可以学习源域与目标域之间的映射关系，将不同的表情作为源域 X：$\{x_i\}_{i=1}^{N}\in X$，x_i为不同的表情标签(例如开心、难过、吃惊等)，N 为数据集中的不同表情标签数，中性表情作为目标域 y，各自训练不同的生成器 G_{x_iy}。生成器 G_{x_iy}学习不同表情到中性表情的映射关系，即在 G_{x_iy}“存储”了不同的表情成分。

Cycle-GAN 包含两个生成器和两个判别器，训练模型时，每次只将一类表情图像作为源域输入到生成器 G_{x_iy}中，中性表情图像作为目标域输入到生成器 G_{yx_i}中。将两个生成器生成的两幅图像输入到两个判别器中，得到两个表示图像真实度的数值。再通过模型定义的生成对抗损失和循环一致损失控制生成器生成更加真实且风格更加接近目标域的图像，最终完成图像的风格迁移。如此训练多次(次数由数据集表情类别数而定)，每次用数据集中不同的表情作为源域，目标域则同为数据集中的中性表情，得到不同表情到中性表情的生成器 G_{x_iy}。

在进行中性表情生成的时候，本文使用已经训练好的生成器 G_{x_iy}。如图 3 所示，输入一幅身份信息为 A 的表情图像到生成器 G_{x_iy}(此时 x_i为吃惊)中，生成对应的中性表情图像，且保留了原身份信息。将不同的表情图像输入到对应不同的生成器 G_{x_iy}中，便可得到不同表情图像对应的中性表情图像。

图 3 中性表情生成方法

2　基于风格迁移的表情识别方法

2.1　基于 CNN 的面部表情特征提取

卷积神经网络(Convolutional Neural Networks, CNN)可以有效提取图像特征用于训练分类器做分类任务。传统的卷积神经网络由三部分组成:卷积层、池化层和全连接层。卷积层用于提取和保留图像特征。池化层对特征进行降维和抽象,可以有效减少网络参数,避免过拟合现象。全连接层连接前面提取到的特征,根据不同的任务输出不同的结果。

本文使用的 CNN 网络结构如图 4 所示。网络包含五个卷积层和一个全连接层。每个卷积层的卷积核大小均为 3×3,步幅均为 1,经过每层卷积之后通道数为原来的两倍。

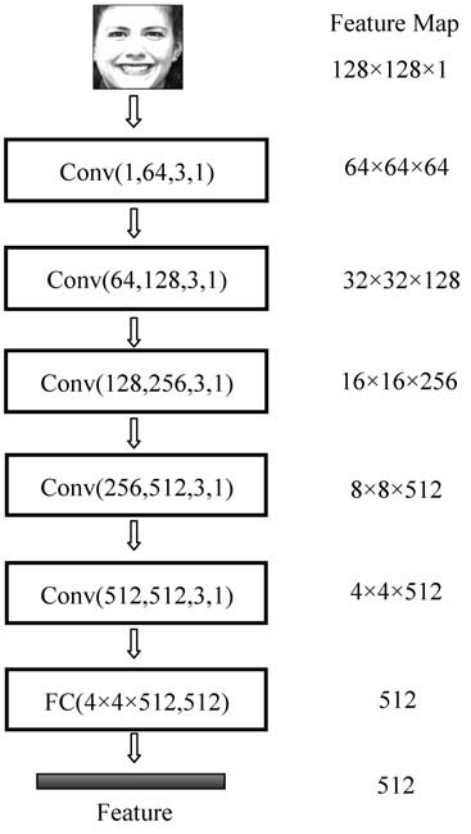

图 4　CNN 网络结构及特征图

在训练 CNN 网络时,损失函数用来计算模型输出值与真实值之间的不一致程度,损失函数值越小,模型的鲁棒性越好。本文将使用 L2 loss(均方误差)作为模型的损失函数。L2 loss 的计算公式如下:

$$L(x,y) = \frac{1}{N}\sum_{p\in P}(x(p) - y(p))^2 \qquad (5)$$

式中:N 为样本数,$x(p)$ 与 $y(p)$ 分别表示模型的输出值与目标值,P 为维数。由于 L2 loss 取平方值的缘故,会放大较大误差和较小误差之间的差值,所以通过 L2 loss 训练网络可以使得提取的不同表情特征之间的距离加大,便于更好分类。

在训练阶段,输入数据集中的单通道表情图像,大小为 128×128,经过五个卷积层之后输出的特征图为 4×4×512。将该特征图输入到全连接层之后得到 512 维的输出值。通过最小化损失函数使模型输出值逼近真实值,以得到更加鲁棒的表情特征。将表情图像输入到训练好的 CNN 网络中,得到的 512 维输出值即为提取到的表情特征。

2.2　表情识别方法

本文提出一种基于风格迁移的表情识别方法。训练把面部表情图像迁移到中性表情图像的生成器,生成器中“存储”了该表情的表情成分。输入表情图像到所有表情生成器中,得到一组表情图像。在该组图像中,只有“存储”了输入图像的表情成分的生成器可以成功迁移为中性表情,而剩下的图像则保留原表情。将表情识别任务转换为二分类任务。

本文使用 SVM 作为二分类器,训练 SVM 时,将数据集中所有非中性表情图像统一标记为正类,中性表情图像标记为负类。将所有图像输入到训练好的 CNN 模型中提取表情特征,再将提取到的表情特征用于训练 SVM 做分类。

测试阶段,将通过生成器生成的一组图像输入 CNN 网络提取表情特征,提取到的表情特征用于 SVM 分类,最终得到识别结果。

方法的具体步骤如下:

Step 1　输入面部表情图像到不同的表情生成器 G_{x_iy} 得到一组图像。每幅图像用 $image\_x_i$ 表示,x_i 标记生成该图像的生成器对应的表情标签。

Step 2　将这一组图像输入到 CNN 网络,提取到一组表情特征。每幅图像对应的特征用 $Feat\_x_i$ 表示。

Step 3　将提取到的表情特征输入到 SVM 中做二分类,得到分类结果。取分类结果为负类(中性表情)的特征 $Feat\_x_i$ 对应的 x_i 为最终表情识别的结果。

3　实　验

3.1　实验设置与环境

Cycle-GAN 同时对生成器和判别器进行训练。训练使用的优化器为 Adam,学习率为 0.000 2,动量为 0.5。对于 Cycle-GAN 损失函数中的三个超参数:对抗损失超参数 λ_1,生成损失超参数 λ_2,循环一致损失超参数 λ_3,本文使用的是 Zhu 等[20]在实验中设定的值,分别设为 1、5、10。训练时 BatchSize 为 1,迭代次数

为 150。

在 CNN 网络中,在每层卷积层之后都使用了批标准化(Batch Normalization,BN)[21]处理。批标准化之后使用的激活函数为 ReLU。并且为了防止在训练时出现过拟合,在全连接层之后,使用了舍弃概率为 0.5 的 Dropout。

实验使用的深度学习框架为 PyTorch,GPU 为 NVIDIA TITAN V,显存为 12 GB。编程语言为 Python 3.5,操作系统为 Ubuntu 16.04。

3.2 数据集

本文实验使用了三个面部表情识别数据集:CK + [22]、MMI[23]、RAF-DB[24]。其中 CK + 和 MMI 为实验室环境下的表情数据集,RAF-DB 为自然环境下的表情数据集,数据集图像样本如图 5 所示。在做数据集预处理时,使用 OpenCV 人脸检测算法定位人脸区域,自动裁剪出面部表情图像。然后统一转化为大小 128 × 128 的单通道灰度图像。

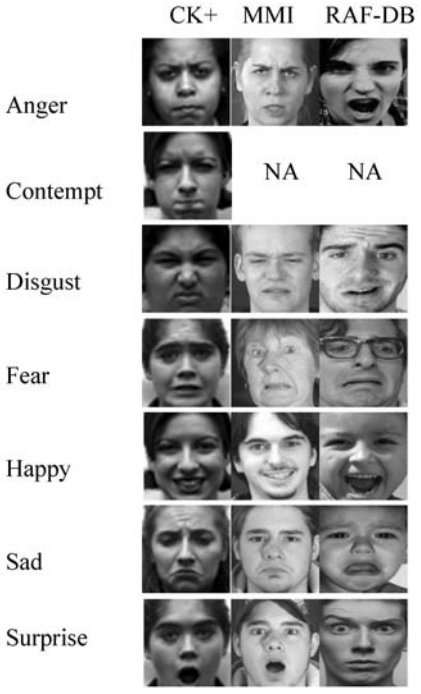

图 5 CK + 、MMI、RAF-DB 数据集的图像样本

CK + 数据集是在 Cohn-Kanada Dataset 基础上扩展得到的,被广泛地用于面部表情识别。数据集是在实验室条件下获取,包含了从 123 个人中得到的 593 个图像序列,每个表情序列都为从中性表情过渡到峰值表情。数据集包括七类表情:生气(Anger)、蔑视(Contempt)、厌恶(Disgust)、恐惧(Fear)、高兴(Happy)、悲伤(Sad)、惊讶(Surprise)。实验使用的是带有表情标签的 327 个序列,取每个序列的最后三幅图像为该表情图像,表情标签和序列标签一致。得到 981 幅表情图像,不同的表情数量分布如表 1 所示。

表 1 CK + 数据集表情数量分布

表情	生气	蔑视	厌恶	恐惧	高兴	悲伤	惊讶
数量/幅	135	54	177	75	207	84	249

MMI 数据集包含 236 个图像序列,从 31 个人中获得。每个序列数据集包括六类基本表情:生气(Anger)、厌恶(Disgust)、恐惧(Fear)、高兴(Happy)、悲伤(Sad)、惊讶(Surprise),相比 CK + 数据集,少了蔑视表情。每个图像序列从中性表情过渡到峰值表情,再从峰值表情过渡回到中性表情,峰值表情处在序列中段。实验选取 208 个正脸角度的序列,取每个序列中段的 3 幅图像作为表情图像,标签为该序列标签。得到 624 幅表情图像,不同的表情数量分布如表 2 所示。

表 2 MMI 数据集表情数量分布

表情	生气	厌恶	恐惧	高兴	悲伤	惊讶
数量/幅	101	103	86	125	97	112

RAF-DB 是从网页上抓取的大量人脸图像,采用众包的方式标记表情标签的自然条件下的表情数据集。RAF-DB 包含 15 339 幅标记了七类基本表情(愤怒(Anger)、厌恶(Disgust)、恐惧(Fear)、高兴(Happy)、悲伤(Sad)、惊讶(Surprise))的图片,和 3 954 幅标记了 12 类复合表情的图片,实验只用标记为基本表情的图片。由于 RAF-DB 数据集存在严重的样本比例失衡问题,例如高兴的表情图片有 5 957 幅,而恐惧的表情图片只有 355 幅。所以每个表情图片选取数量以最少的恐惧表情为标准,都为 355 幅。

3.3 实验结果与分析

本文实验在三个数据集上均采用 10 折交叉验证,实验结果取 10 次交叉验证结果的平均值。为了评价本文提出的方法的有效性,每个数据集的实验结果中均设有 Baseline。Baseline 的结果为本文提取表情特征的 CNN 网络直接用于表情分类的结果。在使用 CNN 网络进行表情分类时,在 CNN 网络最后加一层输出为 6 维或 7 维(由数据集表情类别数而定)数值的全连接层,使用交叉熵(Cross Entropy Loss)作为损失函数进行表情分类。

3.3.1 CK + 数据集

根据 2.2 节表情识别方法可知,不同表情输入到生成器生成的中性表情的图像质量对最终的识别结果有很大影响。图 6 为在 CK + 数据集上的迁移效果,单数行是输入的表情图像,偶数行是对应表情生成的中性表情图像。

图 6　CK + 数据集上的表情迁移效果

使用本文提出的方法,在 CK + 数据集上的表情识别准确率达到 98.47%,如表 3 所示。为了评估本文提出的方法在 CK + 数据集上的表现,将实验结果与近几年提出的六种表情识别方法(LBVCNN[9]、DTAGN[10]、IACNN[11]、RN + LAF + ADA[25]、ppfSVM[26]、DCMA-CNNs[12])进行对比。LBVCNN、DTAGN 和 ppfSVM 三种方法输入的为表情序列,训练数据的规模远大于本文方法使用的数据规模。可以看出,本文方法较比其他几种方法在准确率上都有提高。

表 3　不同方法在 CK + 数据集上的准确率对比

方法	准确率/%	输入
LBVCNN	97.38	图像序列
DTAGN	97.25	图像序列
IACNN	95.37	图像
RN + LAF + ADA	96.40	图像
ppfSVM	96.87	图像序列
DCMA-CNNs	93.46	图像
CNN(baseline)	87.53	图像
本文方法	98.47	图像

表 4 为本文方法在 CK + 数据集上各类表情识别的混淆矩阵,表格斜对角线的值对应各表情识别准确率,其他数值为表情识别错误率。通过混淆矩阵看出,蔑视表情的识别率最低,为 96.39%。原因是蔑视表情在数据集中的数量较少,将该表情迁移至中性表情的生成器学习到的特征较少,生成的图像质量偏差,从而影响准确率。

表 4　CK + 数据集上的混淆矩阵

表情	生气	蔑视	厌恶	恐惧	高兴	悲伤	惊讶
生气	98.11	0.51	0.32	0.25	0	0.53	0.28
蔑视	0.21	96.39	0	0.52	0.65	0.23	0
厌恶	0	0	99.47	0.19	0.34	0	0
恐惧	0.91	0.57	0	98.25	0	0	0
高兴	0	0	0	0	100	0	0
悲伤	0.77	0.71	0.21	0	0	98.3	0
惊讶	0	0	0	0	0.88	0	99.12

3.3.2　MMI 数据集

使用本文提出的方法,在 MMI 数据集上的表情识别准确率为 85.27%,如表 5 所示。同样,为了评估本文方法在 MMI 数据集上的表现,将实验结果同近几年提出的表情识别方法(IACNN[11]、STM-Explet[14]、DTAGN[10]、HOG-3D[13])进行了比较。其中 STM-Explet、DTAGN-Joint、HOG-3D 均使用图像序列作为输入,以此得到表情变化的时序信息,本文方法相比于这三个方法,准确率都有显著的提高。而对于同样使用图像输入的 IACNN 方法,本文方法更是提升了近 14 百分点的准确率。

表 5　不同方法在 MMI 数据集上的准确率对比

方法	准确率/%	输入
IACNN	71.55	图像
STM-Explet	75.12	图像序列
HOG-3D	60.89	图像序列
DTAGN	70.24	图像序列
CNN(baseline)	61.13	图像
本文方法	85.27	图像

MMI 数据集上各类表情识别的混淆矩阵如表 6 所示。可以看出,恐惧表情的识别率比较低,容易与厌恶表情和惊讶表情混淆。导致该现象的原因为,MMI 数据集中这三种表情图像在特征上比较相似,使得生成器学习到的表情特征也较为相似,从而影响准确率。另一方面,高兴表情极易识别,准确率达到 96.13%。

表 6　MMI 数据集上的混淆矩阵

表情	生气	厌恶	恐惧	高兴	悲伤	惊讶
生气	87.32	0	5.54	0	3.21	3.93
厌恶	2.84	88.63	0	3.83	1.93	2.37
恐惧	0	14.14	67.31	0	1.15	17.40
高兴	2.11	0	0	96.13	1.24	0.52
悲伤	4.21	6.49	2.11	3.27	80.28	3.12
惊讶	0	3.94	3.33	0	5.96	88.67

3.3.3 RAF-DB 数据集

RAF-DB 数据集的迁移效果如图 7 所示,单数行是输入的表情图像,偶数行是迁移至中性的表情图像。使用本文提出的方法,在 RAF-DB 数据集上的表情识别准确率为 78.13%。同样为了评估本文方法在 RAF-DB 数据集上的表现,将实验结果与相关方法进行比较,实验结果如表 7 所示。其中 VGG、AlexNet、baseCNN、DLP-CNN[24]是 RAF-DB 数据集作者给出的基准方法。FsNet + TcNet[27]、Boosting-POOF[28]为当前较为先进的方法。通过表 7 可知,本文方法相比于数据集给出的基准方法中准确率最高的 DLP-CNN 方法,准确率有 4% 左右的提升。并且实验结果的准确率接近于当前最好方法,证明了本文方法在自然条件下的面部表情识别效果的可靠性。

图 7 RAF-DB 数据集上的表情迁移效果

表 7 不同方法在 RAF-DB 数据集上的准确率对比

方法	准确率/%	输入
VGG	58.22	图像
AlexNet	55.60	图像
baseDCNN	72.42	图像
DLP-CNN	74.20	图像
Boosting-POOF	73.19	图像序列
FsNet + TcNet	82.02	图像序列
CNN(baseline)	71.27	图像
本文方法	78.13	图像

表 8 为 RAF-DB 数据集上各类表情识别的混淆矩阵。可以发现,“惊讶”“高兴”和“生气”这三类表情具有更高的识别率,而“厌恶”“恐惧”和“悲伤”这三类表情的识别率较低,容易产生混淆。

表 8 RAF-DB 数据集上的混淆矩阵

表情	生气	厌恶	恐惧	高兴	悲伤	惊讶
生气	77.32	1.94	2.41	7.93	2.14	8.26
厌恶	6.17	31.68	25.24	8.36	20.39	8.16
恐惧	1.09	22.62	48.16	0.98	25.19	1.96
高兴	3.33	0.91	1.28	89.27	1.08	4.13
悲伤	4.08	11.25	9.47	2.16	69.62	3.42
惊讶	6.73	1.16	1.71	6.53	1.16	82.71

4 结 语

本文提出了一种基于风格迁移的面部表情识别方法。通过训练 Cycle-GAN 得到将不同表情图像迁移到中性表情图像的生成器,即生成器学习了不同表情的表情成分。将表情图像输入到训练好的生成器中得到一组图像,该组图像只具备中性表情和输入图像的表情。通过 CNN 网络提取该组图像特征用于 SVM 做二分类任务,得到中性表情图像,生成该图像对应的表情生成器即为表情识别结果。实验结果表明该方法不仅在实验室条件下的获得的数据集中表现良好,在自然条件下获得的数据集中也有较高识别率。在 CK + 数据、MMI 数据集和 RAF-DB 数据集上的表情识别率分别为 98.47%、85.27% 和 78.13%。在今后的任务中,将进一步提升表情迁移后的图像质量和迁移效果,以便提高在更加复杂环境下的面部表情识别率。

参 考 文 献

[1] Mehrabian A, Russell J A. An approach to environmental Psychology[M]. Cambridge: The MIT Press, 1980.

[2] Ekman P, Friesen W. Facial action coding system: A technique for the measurement of facial movement[M]. Palo Alto: Consulting Psychologists Press, 1978.

[3] Zhang Z, Mu X, Gao L. Recognizing facial expressions based on Gabor filter selection[C]//International Congress on Image and Signal Processing. IEEE, 2011: 1544 – 1548.

[4] He J, Cai J F, Fang L Z, et al. A method of facial expression recognition based on LBP fusion of key expressions areas [C]//Control and Decision Conference. IEEE, 2015: 4200 –

4204.

[5] Li T, Du C, Naren T, et al. Using feature points and angles between them to recognize facial expressions by a neural network approach[J]. IET Image Processing, 2018, 12(11): 1951 – 1955.

[6] Zhi R, Flierl M, Ruan Q, et al. Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2011, 41(1): 38 – 52.

[7] Lin Z, Qing L, Peng Y, et al. Learning active facial patches for expression analysis[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2012: 2562 – 2569.

[8] Sun W, Zhao H, Jin Z. A visual attention based ROI detection method for facial expression recognition[J]. Neurocomputing, 2018, 296: 12 – 22.

[9] Kumawat S, Verma M, Raman S. LBVCNN: Local binary volume convolutional neural network for facial expression recognition from image sequences[EB]. arXiv: 1904. 07647, 2019.

[10] Jung H, Lee S, Yim J, et al. Joint fine-tuning in deep neural networks for facial expression recognition[C]//Proceeding of the 2015 IEEE International Conference on Computer Vision(ICCV2015). IEEE, 2015: 2983 – 2991.

[11] 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.

[12] Xie S, Hu H. Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks[J]. IEEE Transactions on Multimedia, 2019, 21(1): 211 – 220.

[13] Klaser A, Marszałek M, Schmid. A spatio-temporal descriptor based on 3D-gradients[EB/OL]. [2023 – 01 – 04]. https://lear. inrialpes. fr/pubs/2008/KMS08/KlaserMarszalekSchmid-BMVC08-3DGradientDescriptor. pdf.

[14] Liu M, Shan S, Wang R, et al. Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition[C]//In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[15] Calder A, Young A. Understanding the recognition of facial identity and facial expression[J]. Nature Reviews Neuroscience, 2005, 6(8): 641 – 651.

[16] Wang H, Ahuja N. Facial expression decomposition[C]// Proceedings Ninth IEEE International Conference on Computer Vision. IEEE, 2003.

[17] Kim Y, Yoo B, Kwak Y, et al. Deep generative-contrastive networks for facial expression recognition[EB]. arXiv: 1703. 07140, 2017.

[18] Lee S H, Plataniotis K N K, Yong M R. Intraclass variation reduction using training expression images for sparse representation based facial expression recognition[J]. IEEE Transactions on Affective Computing, 2014, 5(3): 340 – 351.

[19] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. ACM, 2014.

[20] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[EB/OL]. [2023 – 01 – 04]. https://arxiv. org/pdf/1703. 10593. pdf.

[21] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[EB]. arXiv: 1502. 03167, 2015.

[22] Lucey L, Cohn J F, Kanade T, et al. The extended cohn-kanade dataset(ck +): A complete dataset for action unit and emotion-specified expression[C]//Computer Vision and Pattern Recognition Workshops. IEEE, 2010.

[23] Pantic M, Valstar M, Rademaker R, et al. Web-based database for facial expression analysis[C]//2005 IEEE International Conference on Multimedia and Expo. IEEE, 2005.

[24] Li S, Deng W, Du J P. Reliable crowdsourcing and deep locality- preserving learning for expression recognition in the wild[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2852 – 2861.

[25] Hu G S, Liu L, Yuan Y, et al. Deep multi-task learning to recognise subtle facial expressions of mental states[C]//Proceeding of the 15th European Conference on Computer Vision. Springer, 2018: 106 – 123.

[26] Kacem A, Daoudi M, Amor B B, et al. A novel space-time representation on the positive semidefinite cone for facial expression recognition[C]//2017 IEEE International Conference on Computer Vision(ICCV2017). IEEE, 2017: 3199 – 3208.

[27] 吕诲,童倩倩,袁志勇. 基于人脸分割的复杂环境下表情识别实时框架[J]. 计算机工程与应用,2020,56(12):134 – 140.

[28] Liu Z, Li S, Deng W. Boosting-poof: Boosting part based one vs one feature for facial expression recognition in the wild[C]//Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition. IEEE, 2017: 967 – 972.