文章编号：2095-2163(2021)11-0180-05 中图分类号：TP391 文献标志码：A

基于数据融合与迁移学习的学生表情识别研究

孙佳颖[1]，刘新颖[1]，姚 双[2]，沈 艳[3]，余冬华[1]
（1 绍兴文理学院 计算机科学与工程系，浙江 绍兴 312000；2 中国计量大学 经济与管理学院，杭州 310016；
3 哈尔滨工程大学 数学科学学院，哈尔滨 150000）

摘 要：学生表情逐渐成为感知学生状态的重要途径，因此准确的识别学生表情因具有重要价值而受到广泛的关注。本文针对学生表情识别这一问题，提出基于数据融合与迁移学习的识别模型，该模型融合 3 个数据集，以解决学生表情数据缺乏与多样性问题，同时引入迁移学习来提升预测精度。在数据集及实际学生表情图像上的实验结果表明，本文提出的模型可以准确识别学生表情，提升了预测精度。

关键词：学生表情识别；数据融合；迁移学习

Student expression recognition based on data fusion and transfer learning

SUN Jiaying[1], LIU Xinying[1], YAO Shuang[2], SHEN Yan[3], YU Donghua[1]
(1 Department of Computer Science and Engineering, Shaoxing University, Shaoxing 312000, China;
2 College of Economics and Management, China Jiliang University, Hangzhou 310016, China
3 College of Mathematical Sciences, Harbin Engineering University, Harbin 150001, China)

【Abstract】 Student expression has gradually become an important way to perceive the state of students. Therefore, the accurate recognition of student expression has important value and has been widely concerned. In this paper, a recognition model based on data fusion and transfer learning is proposed to solve the problem of lack and diversity of students´ facial expression data. Meanwhile, transfer learning is introduced to improve the accuracy of prediction. Experimental results on datasets and actual students´ facial expressions show that the proposed model can accurately identify students´ facial expressions and improve the prediction accuracy.

【Key words】 student expression recognition; data fusion; transfer learning

0 引 言

近年来，随着人工智能的迅速发展，表情识别技术在人机交互，安全等领域取得了丰硕的成果。面部表情可以表现出丰富的含义，通过分析学生面部表情，可以帮助老师更好的了解学生的上课状态、生活状态和心理变化等等[1-2]。因此，面部表情已逐渐成为感知学生状态的重要途径，表情数据库获取的便捷性和表情识别方法的高效性，受到教育领域的广泛关注。通过表情识别可以调整学生的学习状态和教育者的教学策略。

目前可以采用不同的机器学习算法对表情进行识别。例如：决策树、SVM、深度网络等等。Xia 等人提出基于 SVM 的人脸表情识别[3]；Su 等人提出基于 Mini_Xception_SE 的双通道的表情识别[4]；Jung 等人提出基于深度学习的表情识别等等[5]，以及在常规的方法上进行改进，何俊等人基于 CK+数据集，采用迁移学习和支持向量机，提出一种改进的深度残差网络的表情识别算法[6]；王素琴等人基于 CK+数据集及迁移学习，将长短期记忆网络与 VGGNet 组合成层 VGGNET-LSTM 模型[7]；李旻择等人提出了一种基于多尺度特征卷积神经网络的实时人脸检测表情识别方法，将在 FER-2013 人脸表情数据集上训练得到的模型再迁移到 CK+数据集上，再次训练[8]。这些方法都采用了迁移学习方法，以达到简化网络训练的目的。虽然这些方法有较好的识别效果，但是采用的数据库单一，测试都是基于训练的数据库，并没有运用到实际的校园生活

基金项目：浙江省大学生科技创新活动计划暨新苗人才计划(2021R432015)；国家自然科学基金(62002227)。

作者简介：孙佳颖(1999-)，男，本科生，主要研究方向：图像识别；刘新颖(2001-)，女，本科生，主要研究方向：图像识别；姚 双(1988-)，女，博士，讲师，硕士生导师，主要研究方向：数据挖掘、创新；沈 艳(1965-)，女，博士，教授，硕士生导师，主要研究方向：系统仿真与建模、数据挖掘；余冬华(1988-)，男，博士，讲师，硕士生导师，CCF 会员(C9219M)，主要研究方向：数据挖掘、生物信息学。

通讯作者：余冬华 Email：donghuayu163@163.com

收稿日期：2021-09-16

中，不能够对学生的表情进行准确的识别。

Jain 等人提出不同年龄阶段学生网络课堂情绪识别研究，通过识别表情，身体部位和手势，可以轻松的执行在线课程[9]；Changjun 等人采用支持向量机和最近邻分类对表情进行分类，然后对学生的心理状态进行分类，对 JAFFE 表情数据库进行实验，获得了 71.35%的平均识别率[10]；Lasri 等人使用卷积神经网络对学生面部情绪进行识别，采用 Fer2013 表情数据库，但是最后得出的识别效果并不理想[11]。上述方法，在单一表情数据集上，无论采用基于迁移学习的深度网络，还是采用经典的分类模型，预测精度均有待提升。

本文针对学生表情及预测精度等问题，提出基于多源数据集及 ResNet 网络结构，并且采用迁移学习的深度学习模型。该模型迁移了 ImageNet 中的训练参数，除输出层外，冻结其余参数，以获取更优的图像识别能力，并将 CK+，JAFFE，Fer2013 这 3 种表情数据集进行融合，为表情识别提供更精准的数据源，相较于单一数据源，融合数据集可以提供更多样性的样本，以解决没有专业的学生表情数据集。

1　ResNet 网络简介

ResNet 是在 2015 年由微软实验室提出的一种网络架构，主要采用了残差结构，如图 1 所示。通过残差块的堆积组成 ResNet 网络。原来的神经网络训练时，采用的都是单通道，而 Resnet 采用的是双通道，大大提高了网络训练的正确率。。ResNet 网络结构解决了深层次神经网络正确率下降问题，其在目标检测和分类任务中都表现突出。

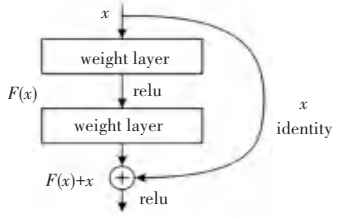

图 1　ResNet 的残差结构

Fig. 1　Residual structure of ResNet

主流的 ResNet 网络结构有 ResNet18，ResNet34，ResNet50，ResNet101. ResNet152。本次实验采用 ResNet50 来对数据库进行训练。ResNet50 的网络结构见表 1。

表 1　ResNet50 网络结构

Tab. 1　Network structure of the ResNet50

层名称	类型	核尺寸	输出大小
cov1	Conv2D	7×7×64	112×112×64
cov2_x	MaxPooling2D	3×3×64	56×56×64
	Conv2D	$\begin{bmatrix}1\times1\times64\\3\times3\times64\\1\times1\times256\end{bmatrix}\times3$	56×56×256
cov3_x	Conv2D	$\begin{bmatrix}1\times1\times128\\3\times3\times128\\1\times1\times512\end{bmatrix}\times4$	28×28×512
cov4_x	Conv2D	$\begin{bmatrix}1\times1\times256\\3\times3\times256\\1\times1\times512\end{bmatrix}\times6$	14×14×512
cov_5	Conv2D	$\begin{bmatrix}1\times1\times512\\3\times3\times512\\1\times1\times2048\end{bmatrix}\times3$	7×7×2048

2　基于数据融合与迁移学习的学生表情识别模型

本文将 CK+，Fer2013 和 JAFFE 3 个不同数据集融合在一起，以弥补学生表情数据集缺乏的不足，同时也是提供更多样性的训练样本。由于不同的数据集的图像大小并不一致，而且图像中出现了人脸以外的空白区域，会影响模型精度，所以需要对人脸区域进行识别和裁剪。

2.1　数据集融合

本文采用 OpenCV 中的 LBP 特征级联检测器检测人脸。LBP（Local Binary Pattern，局部二值模式）是一种用来描述图像局部纹理特征的算子，具有旋转不变性和灰度级不变性等显著的优点。LBP 编码示意，如图 2 所示，将中心像素值作为一个阈值，与所有领域的像素值相比，当像素值大于阈值时设置为 1，否则设置为 0。从图 1 中可以看出中心的像素值为 5，将阈值与周围的像素值比较后，生成一个 0，1的矩阵，从第一开始顺时针排序生成一个二进制数 01 100 101，转换为十进制数为 101。对整幅图的像素值进行计算后得到图像的 LBP 纹理图，进而得到 LBP 直方图。

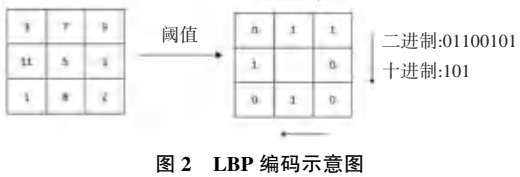

图 2　LBP 编码示意图

Fig. 2　Coding diagram of the LBP

对于检测到的人脸,需要裁剪成符合 ResNet 网络输入的、统一大小格式,裁剪的流程如图 3 所示,分为人脸区域的绘制,人脸区域的采集和图片的放缩。对 3 个数据集进行人脸检测与裁剪后,可以获得统一大小的人脸图像,即 3 个不同的数据集融合为一个数据集。

图 3 人脸图片裁剪过程

Fig. 3 Face image cropping process

融合数据集见表 2,标签为 anger 的数据数量有 155;标签为 disgust 的数据数量有 155;标签为 Fear 的数据数量有 155;标签为 Happy 的数据数量有 155;标签为 Natural 的数据数量有 155;标签为 Sadness 的数据数量有 155;标签为 Surprised 的数据数量有 155。

表 2 融合数据集

Tab. 2 Fusion datasets

数据集	Anger	Disgust	Fear	Happy	Natural	Sadness	Surprised
CK+	60	60	43	73	36	39	52
Fer	65	72	78	78	52	65	65
Jaffe	30	29	36	36	40	31	30
All	155	137	121	181	118	135	147

2.2 基于 ResNet 的迁移学习模型

采用 ResNet 的迁移学习,可以在较低成本下获得较好的学习模型。本次设计采用了 ResNet 网络,然后迁移 ImageNet 数据集上的权重参数,流程如图 4 所示。将 ImageNet 中训练好的权重参数作为初始参数,迁移到 ResNet 网络中,重新构建 ResNet 的全连接层,将原来 1 000 个分类,修改为符合表情数据集的 7 个分类。将除了 ResNet 的 fc 层的其他层 requies_grad 设置为 False,冻结除了全连接层的所有层,在实际的训练中只训练全连接层权重。

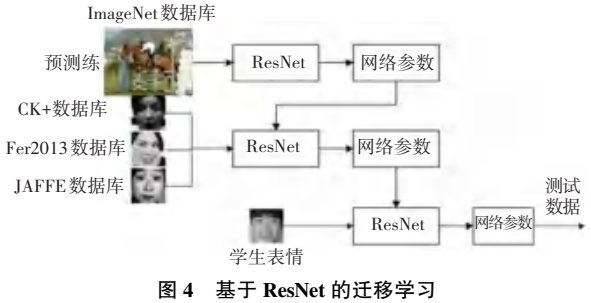

图 4 基于 ResNet 的迁移学习

Fig. 4 Transfer learning based on ResNet

3 实验结果与分析

为了比较预测准确率,分别采用 CK+,JAFFE,Fer2013 和融合数据集,训练迁移后的 ResNet50 网络,分别称为模型 1,模型 2,模型 3,模型 4,最后用这 4 个模型预测学生表情。

3.1 数据集的介绍

CK+表情数据集一共包括 8 种基本表情,分别为:生气,蔑视,高兴,悲伤,惊奇,讨厌,害怕,中性。数据集是从平静到表情表现峰值的图片,本次实验选用其中比较明显的图片。数据集中存在大量重复的照片,选取其中具有代表性的图片,其中生气 135 张,蔑视 177 张,高兴 207 张,悲伤 84 张,惊奇 249 张,讨厌 75 张,中性 108 张。

日本女性面部表情数据库(The Japanese Female Facial Expression, JAFFE)提供日本人表情图像,表情区分度高,共 213 张图像,同属亚洲,与中国学生具有一定相似性,该数据集提供 7 种基本表情,分别为:中性、高兴、悲伤、惊奇、愤怒、厌恶、恐惧。数据集中每张照片大小为 256 像素×256 像素,且存在人脸以外区域,需要对数据集进行裁剪,缩放。

Fer2013 人脸表情数据集由35 886张人脸表情图片组成,其中训练集287 808张,公共测试集和私有测试集各3 589张,每一张图片大小为4 848。由于数据集比较多,并且其中有很多侧脸,遮挡和卡通图片,本次实验从中选取部分图片进行实验,并都是正脸,特征明显。

由于数据量少,在 ResNet50 中训练时,容易出现过拟合现象。使用数据增广技术来扩充数据集,如图 5 所示,对图片进行中心裁剪,调整亮度,旋转,随机裁剪等操作。

图 5 数据增广

Fig. 5 Data augmentation

3.2 不同模型性能结果分析

结合实际的情况和表情数据集的种类,将表情数据集划分为 7 种类型,分别是快乐,恐惧,悲愤,悲伤,惊讶,厌恶,中性。将 CK+,JAFFE,Fer2013 这 3 种表情数据集在 ResNet50 网络上训练。对 CK+表

情数据集进行多次的训练，即模型 1，测试准确率为 97.275%，测试集的混淆矩阵如图 6 所示；对 JAFFE 数据集进行训练，即模型 2，测试集的混淆矩阵如图 7 所示，测试集的准确率为 87.719%；对 Fer2013 数据集进行多次的训练，即模型 3，测试集的混淆矩阵如图 8 所示，测试集的准确率为 73.684%。

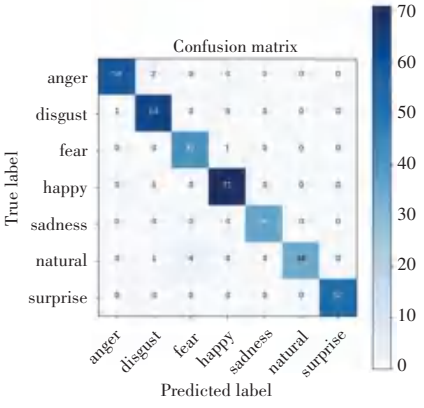

图 6　CK+测试混淆矩阵

Fig. 6　CK+ test confusion matrix

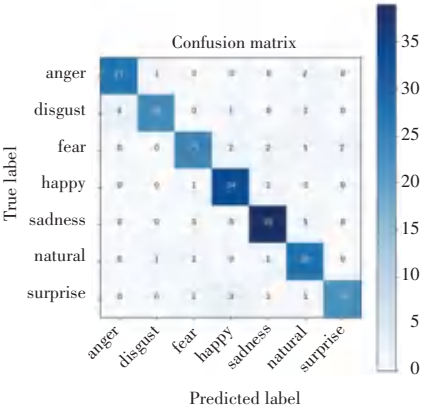

图 7　JAFFE 测试混淆矩阵

Fig. 7　JAFFE test confusion matrix

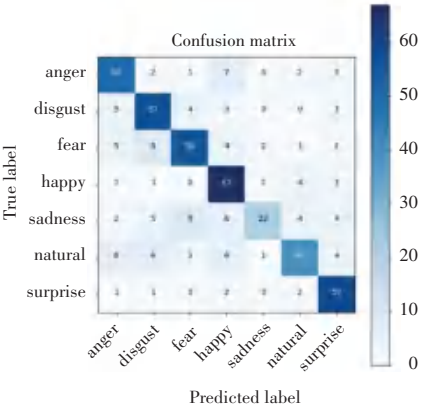

图 8　Fer2013 测试混淆矩阵

Fig. 8　Fer2013 test confusion matrix

在本文提出的融合数据集上，训练 ResNet50 网络，即模型 4，测试集的混淆矩阵如图 9 所示，测试集的准确率为 85.286%。

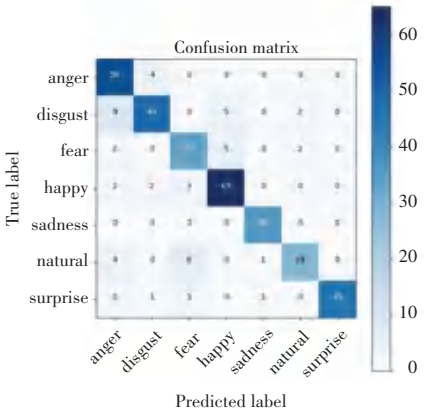

图 9　融合数据测试混淆矩阵

Fig. 9　Fusion data test confusion matrix

从测试结果来看，模型 4 以 85.28% 的准确率显著优于模型 3，略劣于模型 2，较显著劣于模型 1。但模型 1 与 2 中的数据集量较小，且图片相对单一，进而准确率较高。对于模型 3，数据量大，且图片更加多样，经过数据融合后，本文提出的模型 4 性能显著提升。在进行学生表情识别时，学生人数众多，表情图片也会更加多样，综合来看，在准确率及适应方面，模型 4 将会更加适宜学生表情识别。

3.3　学生表情测试结果分析

选取不同状态的学生表情进行实验，由于通过人的感觉给每一种表情进行分类的难度较大，所以选取了日常生活中常见到的 3 种表情，分别为 Happy，Natural，Sadness，如图 10 所示，对模型1~4进行测试。

图 10　学生表情

Fig. 10　Student expression

模型 1 的预测结果见表 3。对于图片 1，输出 Happy 的识别率为 0.688；对于图片 2，输出 Natural 的识别率为 0.463；对于图片 3，输出 Sadness 的识别率为 0.615，但是识别成 Natural 的置信度更高，为 0.94，图片 3 被识别成 Natural 类别，出现识别错误。从表 3 中可以发现，模型 1 可以识别出表情状态，但是识别出的表情置信度并不高。

表 3 模型 1 学生表情测试置信度(CK+)

Tab. 3 Student expression test confidence of Model 1 (CK+)

图片	置信度						
	Anger	Disgust	Fear	Happy	Natural	Sadness	Surprised
图片 1	0.028	0.029	0.018	0.688	0.005	0.210	0.022
图片 2	0.009	0.007	0.063	0.076	0.463	0.256	0.127
图片 3	0.210	0.001	0.040	0.034	0.94	0.615	0.006

模型 2 的测试结果见表 4。对于图片 1,输出 Happy 的置信度为 0.509;对于图片 2,输出 Natural 的置信度为 0.770;对于图片 3,输出 Sadness 的置信度为 0.031,由于 Disgust 类别的置信度更高,所以图片 3 被预测成 Disgust 类,然而,这是一个错误结果。模型 2 的学生表情识别率与实际情况存在一些误差,对图片 1 的识别置信度不高,对图片 3 出现了错误的预测。

表 4 模型 2 学生表情测试准确率(JAFFE)

Tab. 4 Student expression test accuracy rate of Model 2 (JAFFE)

图片	置信度						
	Anger	Disgust	Fear	Happy	Natural	Sadness	Surprised
图片 1	0.046	0.093	0.030	0.509	0.004	0.318	0.001
图片 2	0.003	0.000	0.012	0.138	0.770	0.000	0.075
图片 3	0.007	0.625	0.247	0.047	0.015	0.031	0.028

模型 3 的测试结果见表 5。对于图片 1,输出 Happy 的置信度为 0.626;对于图片 2,输出 Natural 的置信度为 0.578;对于图片 3,输出 Sadness 的置信度为 0.139,被错误的预测成 Happy 类。模型 3 的学生表情识别准确率与实际情况存在误差,对于图片 1 与图片 2,识别准确,但是置信度较低,而图片 3 的识别出现错误,将 Sadness 表情识别成了 Happy,且两个类别的置信度差异很大。

表 5 模型 3 学生表情测试置信度(Fer2013)

Tab. 5 Student expression test confidence of Model 3(Fer2013)

图片	置信度						
	Anger	Disgust	Fear	Happy	Natural	Sadness	Surprised
图片 1	0.000	0.013	0.016	0.626	0.012	0.004	0.329
图片 2	0.014	0.011	0.125	0.044	0.578	0.001 7	0.212
图片 3	0.052	0.077	0.017	0.680	0.018	0.139	0.016

模型 4 的测试结果见表 6。对于图片 1,输出 Happy 的置信度为 0.826;对于图片 2,输出 Natural 的置信度为 0.778;对于图片 3,输出 Sadness 的置信度为 0.780。可以看出,模型 4 对与学生表情的识别率比较好,3 个表情均以高置信度识别出来,这也是 4 个模型中,针对图片 3 唯一识别准确的模型。

表 6 模型 4 学生表情测试置信度(融合数据集)

Tab. 6 Student expression test confidence of Model 4(fusion data set)

图片	置信度						
	Anger	Disgust	Fear	Happy	Natural	Sadness	Surprised
图片 1	0.000	0.013	0.016	0.826	0.012	0.004	0.129
图片 2	0.014	0.011	0.025	0.044	0.778	0.017	0.112
图片 3	0.052	0.077	0.017	0.039	0.018	0.780	0.016

4 结束语

本文通过 CK+,JAFFE,Fer2013 这 3 个数据集,及本文的融合数据集,在这 4 个表情数据集上,迁移 ResNet50 网络参数,重置输出层,重新训练。使用平时常见的 3 种学生表情来进行测试,本文提出的基于数据融合的迁移学习模型的识别准确率和置信度均获得提升。因此,该模型可适用于基于学生表情分析的各个领域。

参考文献

[1] 郦泽坤,苏航,陈美月,等. 支持 MOOC 课程的动态表情识别算法[J]. 小型微型计算机系统,2017,38(9):2096-2100.

[2] 孙波,刘永娜,陈玖冰,等. 智慧学习环境中基于面部表情的情感分析[J]. 现代远程教育研究,2015(2):96-103.

[3] XIA L. Facial expression recognition based on SVM[C]//2014 7th International Conference on Intelligent Computation Technology and Automation. IEEE, 2014: 256-259.

[4] SU C, WEI J. Expression Recognition of Dual Channels Model System Based on Mini _ Xception _ SE [C]//2020 IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th International Conference on Data Science and Systems (HPCC/SmartCity/DSS). IEEE, 2020: 1338-1343.

[5] JUNG H, LEE S, PARK S, et al. Development of deep learning-based facial expression recognition system[C]//2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV). IEEE, 2015: 1-4.

[6] 何俊,刘跃,李倡洪,等. 基于改进的深度残差网络的表情识别研究[J]. 计算机应用研究,2020,37(5):1578-1581.

[7] 王素琴,张峰,高宇豆,等. 基于图像序列的学习表情识别[J]. 系统仿真学报,2020,32(7):1322-1330.

[8] 李旻择,李小霞,王学渊,等. 基于多尺度核特征卷积神经网络的实时人脸表情识别[J]. 计算机应用,2019,39(9):2568-2574.

[9] JAIN A, SAH H R, KOTHARI A. Study for Emotion Recognition of Different Age Groups Students during Online Class[C]//2021 8th International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, 2021: 621-625.

[10] CHANGJUN Z, SHEN P, CHEN X. Research on algorithm of state recognition of students based on facial expression [C]// Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology. IEEE, 2011, 2: 626-630.

[11] LASRI I, SOLH A R, EL BELKACEMI M. Facial emotion recognition of students using convolutional neural network[C]// 2019 third international conference on intelligent computing in data sciences (ICDS). IEEE, 2019: 1-6.