

# 基于低像素人脸图像的表情识别

刘芾, 李茂军\*, 胡建文, 肖雨荷, 齐战

长沙理工大学电气与信息工程学院, 湖南 长沙 410114

**摘要** 卷积神经网络的发展极大促进了面部表情识别技术的进步,为解决实际应用中图像识别的准确率受图像像素限制的问题,从三方面对低像素人脸图像的表情识别进行研究。首先根据研究对象像素低、特征复杂的特点,提出了一种改进的卷积神经网络。其次对图像进行基本的预处理操作后,又增加了图像增强处理,作为改进卷积神经网络模型的输入。最后将模型的输出结果进行决策融合,得到最终的识别结果。实验表明,该方法在 CK+数据集上取得了良好的效果,且识别准确率较高、效果稳定、泛化能力强。

**关键词** 图像处理; 卷积神经网络; 面部表情; 低像素; 图像识别

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP57.101008

## Expression Recognition Based on Low Pixel Face Images

Liu Fu, Li Maojun\*, Hu Jianwen, Xiao Yuhe, Qi Zhan

School of Electrical and Information Engineering, Changsha University of Science and Technology,  
Changsha, Hunan 410114, China

**Abstract** The development of convolutional neural networks has greatly promoted the advancement of facial expression recognition technology. In order to solve the problem that the accuracy of image recognition in practical applications is limited by image pixels, the expression recognition of low pixel face image is studied from three aspects. First, according to the characteristics of low pixels and complex features of the research object, an improved convolutional neural network is proposed. Second, after performing basic pre-processing on the image, image enhancement processing is added as an input to improve the convolutional neural network mode. Finally, the output results of the model are subjected to decision fusion to obtain the final recognition result. Experimental results show that this method has achieved good results on the CK+ source dataset, and has high recognition accuracy, stable results, and strong generalization ability.

**Key words** image processing; convolutional neural network; facial expression; low pixel; image recognition

**OCIS codes** 100.2000; 100.5010; 150.1135

## 1 引言

面部表情识别技术使计算机能够通过人的表情得到人的情绪信息,从而对人的状态、意图进行判断,在人机交互、安全驾驶、智能广告系统等领域均有重要意义<sup>[1-3]</sup>。CK+数据集是经典的人脸表情库,包含愤怒、厌恶、恐惧、快乐、悲伤、惊喜和蔑视的表情图像,其中的表情为视频序列,即包含了同一个表情由平静到剧烈的系列图像<sup>[4]</sup>,可从中提取出中性表情图像。

受距离等因素的影响,图像会出现人脸模糊、脸部像素点少的问题,低像素人脸图像的表情识别就是识别画质低、面部特征不明显的人脸图像,通过降采样得到的图像尺寸均为 32 pixel × 32 pixel,符合低像素特征。面部表情图像的复杂性高,在面部特征不明显的情况下,难以通过提取鲜明的特征信息进行识别。针对尺寸为 32 pixel × 32 pixel 的面部表情图像,李勇等<sup>[5]</sup>提出了一种基于改进 LeNet-5 卷积神经网络(CNN)的面部表情识别方法;江大鹏等<sup>[6]</sup>提出了一种基于局部二值

收稿日期: 2019-08-21; 修回日期: 2019-10-08; 录用日期: 2019-10-18

基金项目: 国家自然科学基金(61601061)

\* E-mail: 591338413@qq.com

模式(LBP)的 CNN 面部表情识别方法。研究表明, CNN 在低像素人脸图像的表情识别方面效果较好。本文在此基础上,对 CNN 模型进行了改进,提出了一种用于低像素人脸图像的表情识别方法,并与其他几种方法进行对比,结果表明该方法识别效果较好。

## 2 低像素人脸图像的表情识别方法

### 2.1 面部表情图像预处理

一般的表情识别方法包括:图像预处理、面部特征提取以及表情识别。图像表情识别预处理阶段是进行人脸检测,得到面部区域图像<sup>[7]</sup>。对低像素人脸图像表情的识别,还需在预处理时进行图像增强或图像超分辨。图像增强是通过改变像素的分布规律增强图像的已有信息,图像超分辨是通过增加像素点恢复一些丢失的像素信息。

本方法的图像预处理包括:人脸检测和裁剪、灰度处理、降采样、数据增强及图像增强。人脸检测的目的是在图像中准确标定出人脸的位置和大小,使用 Dlib 模型进行人脸检测, Dlib 模型可以自动估计

图像中脸部特征点的坐标<sup>[8]</sup>,在 OpenCV 库中处理数据。以此为基础剪裁图像,使图像特征集中在面部。灰度处理是将彩色图像转化为灰色图像的过程。降采样是为了标准化输入 CNN 模型中的图像尺寸,使用双线性内插法,保证了重新采样后的图像面部位置与原图像相同。使用 CNN 进行图像识别,训练数据量的大小直接影响了最终识别效果,数据量越大,效果越好。常用的数据增强方法有镜像、旋转以及加入噪声等<sup>[9]</sup>,对原始数据进行镜像及不同角度、不同方向的旋转,使数据增强为原始数据的 13 倍,再对已有数据添加不同系数的噪声(椒盐噪声、高斯噪声、泊松噪声以及斑点噪声),最终数据增强为原始数据的 130 倍。对图像进行直方图均衡化并用局部二值模式获得增强后的图像,其中直方图均衡化又称直方图平坦化,实质是对图像进行非线性拉伸,重新分配图像像素值,使一定灰度范围内的像素值数量大致相等。局部二值模式是描述图像局部纹理特征的算子,具有旋转和灰度不变性等优点,可用于提取图像的局部纹理特征。具体预处理流程如图 1 所示。

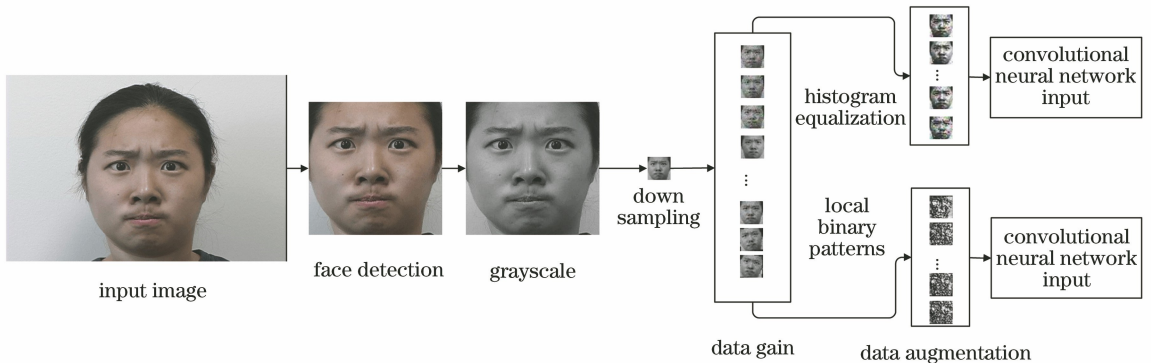


图 1 面部表情预处理流程

Fig. 1 Pipeline of the facial expression preprocessing

### 2.2 改进的卷积神经网络模型

随着计算机处理能力的发展, CNN 在图像识别方面取得了惊人的成绩<sup>[10]</sup>,新的网络模型结构不断被提出,基于 CNN 的图像识别方法效率也不断提高,逐渐取代了传统面部表情图像识别方法。2012 年的 ImageNet 面向对象识别的挑战 (ILSVRC) 中, Krizhevsky 等<sup>[11]</sup>利用 CNN 模型 AlexNet 夺得冠军。2014 年的 ILSVRC 的比赛中, CNN 模型 GoogLeNet 架构在分类方面获得第一名<sup>[12-13]</sup>。Simonyan 等<sup>[14]</sup>提出的 CNN VGGNet 取得了分类组第二名和定位项目组第一名的成绩。VGGNet 在加深网络层数的同时为了避免参数过多,所有层

都采用  $3 \times 3$  的小卷积核,卷积层步长设置为 1。多个卷积层与非线性激活层交替的结构,使其比单一卷积层结构能提取出更深层、更好的特征。在 ILSVRC 2015 的比赛中, He 等<sup>[15]</sup>提出的 ResNet 获得冠军, ResNet 中有一种叫 Shortcut Connection 的连接方式,理论上可在网络层不断加深的情况下,使网络一直保持最优状态。没有像素限制的情况下,面部表情中有足够多的特征信息用于优化模型参数,从而得到良好的识别效果。对于低像素人脸图像,需要对不明显的特征信息充分利用,在此基础上,提出了一种针对低像素人脸图像的表情识别的 CNN 模型,以便更好地提取面部特征。

输入 CNN 模型的图像尺寸为 32 pixel × 32 pixel, 增加 CNN 的层数以增加网络模型的非线性, 使决策函数的识别能力更强。为避免由加深网络层数导致的梯度消失和梯度爆炸问题, 网络需要有更复杂的结构。Srivastava 等<sup>[16]</sup>提出了一种网络连接结构 Highway Networks, 该结构中的一部分

特征不需要处理就可以直接通过某些网络层, 使结构更容易优化。结合该结构和 Shortcut Connection 连接方式, 避免了使用更深层网络出现的梯度消失和梯度爆炸问题。实验使用的 CNN 模型如图 2 所示, 其中 @ 前的数字为特征图的个数, @ 后的数字为特征图的大小 (pixel × pixel)。

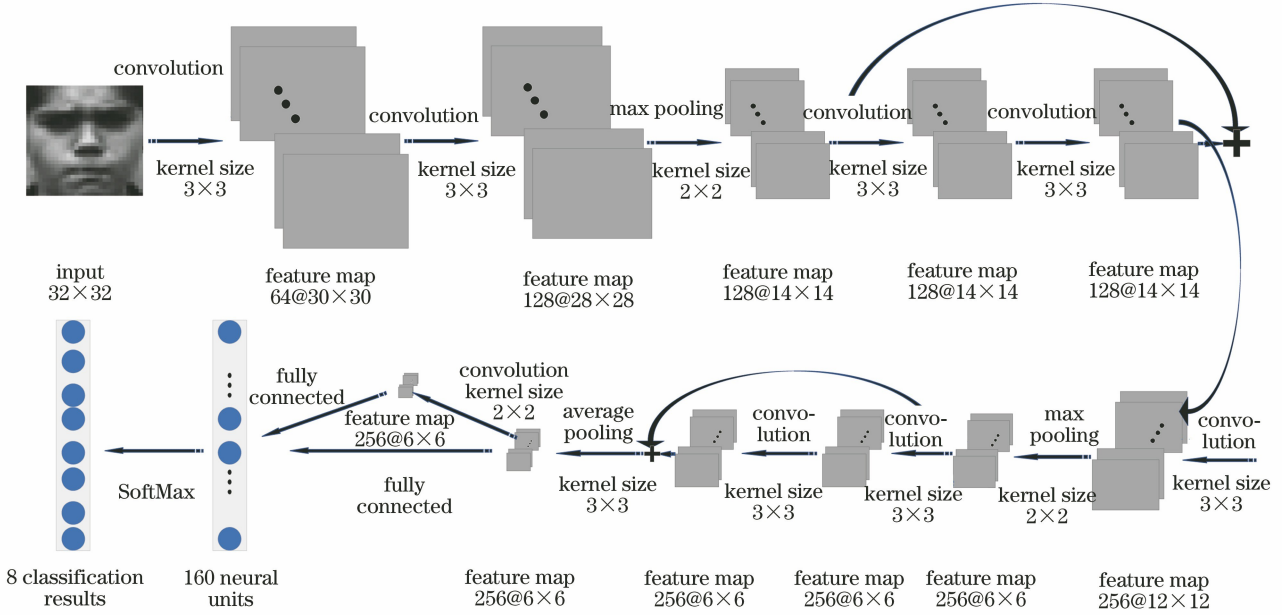


图 2 CNN 模型  
Fig. 2 CNN model

在卷积操作和池化操作中, 输入特征图的尺寸、卷积和池化操作的内核的大小、步长以及输出特征图的尺寸的数学关系式为

$$\omega_2 = (\omega_1 - f + 2p) / s + 1, \quad (1)$$

式中,  $\omega_2$  为卷积或池化操作后的特征图尺寸,  $\omega_1$  为卷积或池化操作前特征图的尺寸,  $f$  为卷积和池化操作的内核的尺寸,  $p$  为零填充的像素个数,  $s$  为步长。

将模型第三层的输出和第五层的输出张量进行求和运算, 得到 128 个尺寸为 14 pixel × 14 pixel 的特征图, 再经过 ReLU 激活函数, 作为第六层的输入。将模型第七层和第九层的输出张量进行求和运算, 得到 256 个尺寸为 6 pixel × 6 pixel 的特征图, 再经过 ReLU 激活函数, 作为第十层的输入。第十二层是全连接层, 将第十层和第十一层的输出分别经过 ReLU 激活函数, 再将得到的张量进行串联运算作为第十二层的输入, 输出为 160 个神经元, 最后一层是 SoftMax 分类器, 输出为 8 个网络节点, 分别表示输入图像为不同表情状态的

概率。表 1 为模型的具体描述, 包括了模型每一层的类型 (Type)、对应的内核大小 (Kernel size) 和步长 (Stride) 以及每一层输出特征图的尺寸 (Output size)。

该 CNN 中包含了卷积、池化及全连接 3 种基本运算。其中卷积又分为内卷积和外卷积, 即无 0 填充情况下的卷积和有 0 填充情况下的卷积。设输入为  $M \times N$  的矩阵  $\mathbf{A}$ , 卷积核为  $m \times n$  的矩阵  $\mathbf{B}$ , 且  $M \geq m, N \geq n$ , 则其内卷积操作的输出为  $\mathbf{C} = \mathbf{A} \times \mathbf{B}$ , 对应位置的像素  $c_{ij}$  可表示为

$$c_{ij} = \sum_{s=1}^m \sum_{t=1}^n a_{i+m-s, j+n-t} \times b_{st}, \quad (2)$$

式中,  $1 \leq i \leq M - m + 1, 1 \leq j \leq N - n + 1$ , 为与矩阵  $\mathbf{A}$  的行列对应相乘, 设矩阵  $\mathbf{B}$  对应位置的像素为  $b_{st}$ , 矩阵  $\mathbf{A}$  对应位置的像素为  $a_{i+m-s, j+n-t}$ 。外卷积定义为将  $\mathbf{A}$  进行 0 填充, 填充后矩阵的行列与  $\mathbf{B}$  矩阵的行列数有关, 使其成为  $(M + 2m - 2) \times (N + 2n - 2)$  矩阵, 再与  $\mathbf{B}$  做内卷积, 可表示为

$$\mathbf{A} \times \mathbf{B} = \hat{\mathbf{A}} \times \mathbf{B}. \quad (3)$$

表1 卷积网络参数

Table 1 Convolutional network parameters

Layer	Type	Kernel size /pixel×pixel	Stride	Output number	Output size /pixel×pixel
Layer 1	Convolution	3×3	1	64	30×30
Layer 2	Convolution	3×3	1	128	28×28
Layer 3	Max pooling	2×2	2	128	14×14
Layer 4	Convolution	3×3	1	128	14×14
Layer 5	Convolution	3×3	1	128	14×14
Layer 6	Convolution	3×3	1	256	12×12
Layer 7	Max pooling	2×2	2	256	6×6
Layer 8	Convolution	3×3	1	256	6×6
Layer 9	Convolution	3×3	1	256	6×6
Layer 10	Average pooling	3×3	3	256	2×2
Layer 11	Convolution	2×2	1	512	1×1
Layer 12	Fully connected	—	—	—	160×1
Output	SoftMax	—	—	—	8×1

对矩阵  $\mathbf{A}$  进行池化,假设对其进行不重叠分块,每块大小为  $\lambda \times \tau$ ,第  $ij$  块用矩阵  $\mathbf{G}_{\lambda,\tau}^A(i,j)$  可表示为

$$\mathbf{G}_{\lambda,\tau}^A(i,j) = (\mathbf{a}_{st})_{\lambda \times \tau}. \quad (4)$$

式中,  $\mathbf{a}_{st}$  为矩阵  $\mathbf{A}$  中第  $s$  行第  $t$  列元素,  $(\mathbf{a}_{st})_{\lambda \times \tau}$  为矩阵  $\mathbf{A}$  中  $\lambda \times \tau$  个元素组成的矩阵块,  $(i-1) \times \lambda + 1 \leq s \leq i \times \lambda, (j-1) \times \tau + 1 \leq t \leq j \times \tau$ ,最大池化的定义为

$$d[\mathbf{G}_{\lambda,\tau}^A(i,j)] = \max\left(\sum_{s=(i-1)\times\lambda+1}^{i\times\lambda} \sum_{t=(j-1)\times\tau+1}^{j\times\tau} \mathbf{a}_{st}\right). \quad (5)$$

平均池化的定义为

$$a[\mathbf{G}_{\lambda,\tau}^A(i,j)] = \frac{1}{\lambda \times \tau} \sum_{s=(i-1)\times\lambda+1}^{i\times\lambda} \sum_{t=(j-1)\times\tau+1}^{j\times\tau} \mathbf{a}_{st}. \quad (6)$$

用大小为  $\lambda \times \tau$  的不重叠块对最大池化后的  $\mathbf{A}_{\max}$  和平均池化后的  $\mathbf{A}_{\text{mean}}$  分别进行下采样,可表示为

$$\begin{aligned} D_{\lambda,\tau}(\mathbf{A}_{\max}) &= d[\mathbf{G}_{\lambda,\tau}^A(i,j)], \\ D_{\lambda,\tau}(\mathbf{A}_{\text{mean}}) &= a[\mathbf{G}_{\lambda,\tau}^A(i,j)]. \end{aligned} \quad (7)$$

对于全连接层的输入  $\mathbf{A}$ ,全连接层的每一个输出都可以看成前一层中第  $r$  个结点  $a_r$  乘以其权重系数  $w_r$ ,再加上一个偏置值  $b_h$  得到。如设全连接层输入有  $256 \times 2 \times 2$  个结点,即输入特征图为  $256 @ 2 \times 2$ ,输出有 80 个结点,则一共需要  $256 \times 2 \times 2 \times 80 = 81920$  个权值系数和 80 个偏置参数,则其输出向量  $\mathbf{D}$  中单个元素  $d_h$  可表示为

$$d_h = \sum_{r=1}^k w_r \times a_r + b_h, \quad (8)$$

式中,  $k$  为输入特征图的个数。

### 3 低像素人脸图像的表情识别实验

#### 3.1 数据集准备

实验使用 CK+数据集,该数据集是用于评估人脸表情识别(FER)系统的数据集,也是目前面部表情识别比较通用的数据集。包含的视频序列分别来自 123 个受试者的 593 个视频序列,持续时间为 10~60 frame 不等,显示了由平静到剧烈的系列图像。由于原始图像在不同表情上的数目分布不均衡,其中中性表情图像为该表情开始或结束时的图像,依据原始数目分布,选取每个表情序列最后 1~3 幅表情图像。用于建模的图像共 686 幅,其中 80% 用作训练集,20% 用作测试集。同一人相同表情的峰值图像不会同时出现在训练集和测试集中。

对训练集进行数据增益,将经过数据增益和未经过数据增益的测试集在同一个被训练好的模型中做测试,发现得到的识别准确率相差较小,因此对测试集不做数据增益处理,最终得到训练集图像规模为 71370 幅,测试集图像规模为 137 幅。将所有图像分别进行直方图均衡化和局部二值模式,得到包括原图在内三份规模相同的数据集。表 2 为每份测试集和训练集中 8 种表情的图像数量,8 种表情的原始图像、经直方图均衡化后的图像以及经局部二值模式的图像如图 3 所示。

#### 3.2 评价标准

面部表情识别方法主要评价标准是识别准确率和识别速度,其中识别准确率是测试集中正确识别的表情样本数和所有测试集样本数的比值。识别速度是识别模型建立好后识别每个测试集样本所花费的时间,即为测试集识别所花费时间与测试集样本

表2 人脸表情数据库  
Table 2 Facial expression dataset

Category	Angry	Disgust	Fearful	Happy	Sad	Surprised	Scorn	Neutral
Train database	14040	12220	7800	14300	8970	8580	5460	8970
Test database	27	24	15	28	15	16	12	15

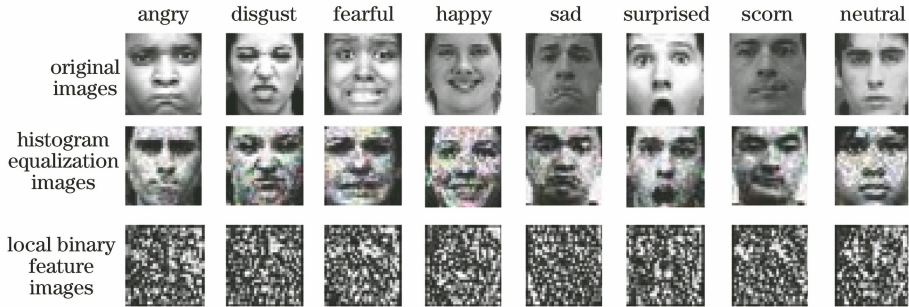


图3 8种表情示例图像

Fig. 3 8 kinds of facial expression images

数的比值。

$$A = \frac{1}{p} \sum_1^p g[f(x_b) = y_b], \quad (9)$$

$$t_d = \frac{T}{p}, \quad (10)$$

式中,  $A$  为识别准确率,  $p$  为测试集样本总数,  $g$  为指示函数,  $x_b$  为给定的样本,  $f(x_b)$  为样本通过模型后的输出,  $y_b$  为给定样本的标签,  $t_d$  为识别速度,  $T$  为所花费的总时间, 可通过最后一个测试样本识别后的时间减去第一个测试样本识别之前的时间得到。

### 3.3 实验过程

表情识别中, 由于输入 CNN 的图像像素低, 输出的识别效果会有小幅度波动, 因此引入决策融合, 做最后的图像识别。在测试阶段, 使用五个训练好的网络模型分别对测试集数据进行判断, 再使用 SoftMax 平均值投票 (SAV) 方式, 融合五个模型的判定结果, 得到最终结果, 以此提升识别效果, 测试步骤如图 4 所示。

由 2.2 小节中可知, CNN 的输出是一个 1 维向量, 向量中每个元素的值分别为图像可能是某个类别的概率大小。SoftMax 平均值投票是将五个训练好的 CNN 输出结果取平均值, 最后输出可能性最大的结果, 取三次实验的平均值。实验环境: 硬件平台: 显卡为 NVIDIA GeForce 940MX, 主频为 1122 MHz, 内存为 2.00 GB。软件平台: 操作系统为 linuxubuntu16.04, 软件为 python3.6、NVIDIA CUDA 和 cuDNN 库。为提高识别准确率, 采用的训练策略为

1) 在每个卷积层和池化层后都加入批量标准

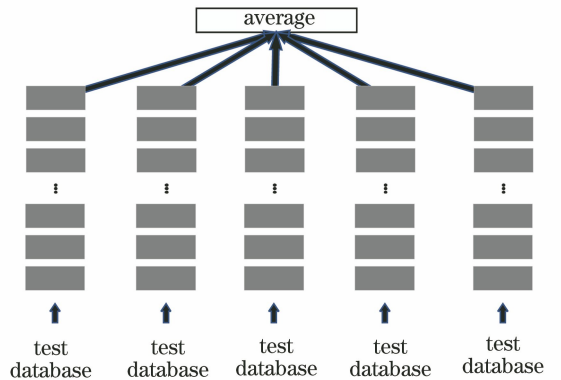


图4 SoftMax 平均值投票示意图

Fig. 4 Diagram of SoftMax average voting

化 (BN) 和 ReLU 激活函数克服梯度消失和加快训练速度。

2) 选择性加入 L2 正则化和 Droupout 缓解过拟合。

3) 采用学习率衰减的策略, 学习率初始时选择一个较大值,  $N$  轮迭代后衰减为初始学习率 1/10。优化时选用 Adam 优化算法, 使网络能够更快寻找全局最优点。

4) 在数据集分别为原始图像、直方图增强后的图像和局部二值模式特征图时训练 CNN 模型, 并对网络进行调参。

根据测试的准确率, 先确定是否需要加入 L2 正则化以及 Droupout 层, 再确定在哪层放置。根据训练时的损失确定学习率的大概范围, 根据训练损失与测试时的准确率, 通过两次二分法对学习率进行取值。

### 3.4 结果分析

为寻找人脸表情识别系统的最优情况, 在 CNN

模型中分别输入原始图像数据集、局部二值模式特征图的数据集和直方图均衡化后数据集,得到的平均识别准确率和速度如表 3 所示,实验为三次实验的平均值。

表 3 识别准确率和时间对比

Table 3 Comparison of recognition accuracy and time

Model	Original database	Local binary pattern database	Histogram equalization database
Accuracy /%	89.5	86.9	90.5
Time/s	1.54	1.27	1.56

可以发现,输入为直方图均衡化后的数据集准确率优于原始图像的数据集,而输入为局部二值模式特征图的数据集准确率最差。三者速度上并无

表 4 改进 CNN 模型的识别准确率

Table 4 Improved recognition accuracy of CNN model

unit: %

Category	Accuracy	Angry	Disgust	Fearful	Happy	Sad	Surprised	Scorn	Neutral	Total
First experiment		96.3	91.7	60.0	100.0	86.7	100.0	83.3	93.3	90.5
Second experiment		88.9	91.7	86.7	100.0	80.0	100.0	83.3	93.3	91.2
Third experiment		96.3	95.8	60.0	100.0	86.7	93.8	75.0	93.3	89.8
Total		93.8	93.1	68.9	100.0	84.5	97.9	80.5	93.3	90.5

从表 4 可以看出,Happy 与 Surprised 两种表情的识别率较高,Fearful 的识别率低且有较大波动。一方面可能是前两种特征较为鲜明,而 Fearful 和 Sad 在某种程度上有相似特征,容易混淆。另一方面是在 CK+数据集中前两种表情的原始数据量丰富,而 Fearful 数据量少,Fearful 测试集中总共 15 幅图像,及原始数据中仅有 5 幅不相似的表情图像,导致训练次数不平等。为了验证建立好的识别模型的效果,将 JAFFE 数据集作为测试集输入,识别除蔑视表情以外的其余表情,三次结果的平均识别准确率为 82.4%。

为证明所提出方法和决策融合方法对低像素人脸图像的表情识别具有针对性,分别在两种情况下进行实验,一种是用经典的浅层卷积神经网络 LeNet-5 替代改进的 CNN 模型,第二种是在不使用决策融合情况下进行实验。表 5 为以上两种情况下,三次实验的平均识别准确率和时间对比。

表 5 两种模型识别准确率和时间对比

Table 5 Comparison of recognition accuracy and time between two models

Model	LeNet-5	Without decision level
Accuracy /%	74.6	87.9
Time /s	0.59	0.31

由表 3 可知,改进后的 CNN 模型比 LeNet-5

明显差别,输入为局部二值模式特征图的数据集识别速度稍快,但 0.29 s 的速度与 3.6% 的准确率相比优势较小。分析图 3 中不同数据集的差别,认为一定的数据增强会强化图像的信息,提高识别准确率,但由于数据原始像素过低,直接提取其局部二值模式的特征图,在增强纹理信息的同时,丢失了更多信息,因此准确率最低,但速度最快。实验最终选择直方图均衡化增强的数据集输入 CNN 模型,其余训练策略的选择:在第十二层后加入 L2 正则化和 Dropout,Dropout 参数为 0.7,学习率初始值为 0.0001,在 1000 次迭代后衰减为初始学习率的 1/10。表 4 给出了改进 CNN 模型在 CK+数据集上,对 8 种表情识别的三次实验结果。

网络时识别准确率增加了 15.9 个百分点,这证明了本方法更适用于低像素人脸图像的表情识别。决策融合后的识别准确率比未经过决策融合的网络增加了 2.6 个百分点,主要原因是三次实验中的实验效果不稳定,其中两次实验的识别准确率均在 90.0% 左右,其中一次实验结果为 83.9%。而本方法每次实验结果都是由五个训练好的网络模型取平均得出,整体稳定性较高,这证明了本方法在实际中是有效可行的。

近年来,对尺寸为 32 pixel×32 pixel 的人脸图像进行表情识别的方法相继被提出。在 CK+数据集上,李勇等<sup>[5]</sup>提出了一种跨连 LeNet-5 的 CNN,对不包括中性表情在内的图像进行七分类,识别准确率为 83.74%。李宽<sup>[17]</sup>提出一种浅层 CNN,取得 97.38% 的七分类识别准确率,比本方法的识别准确率高,但跨数据集识别时,用训练好的模型在 JAFFE 数据集进行六分类时,识别准确率仅有 40.98%,泛化能力差。

## 4 结 论

针对低像素人脸图像的表情识别,提出了一种基于改进的 CNN 表情识别方法。参考 Shortcut Connection 连接方式和 Highway Networks 连接结构,通过增加卷积层的方式增加网络模型非线性,减

少了梯度消失和梯度爆炸的问题。能从更多层次中提取图像特征,反映图像信息,获得了更好的识别性能。但该方法的识别准确率相对于高像素人脸图像识别准确率仍有差距,之后希望运用反卷积等方法实现 CNN 各层特征的可视化,更好地理解各层特征,从而找到准确率更高、通用性更好的网络模型。

## 参 考 文 献

- [1] Jiang M X, Hu M, Wang X H, et al. Dual-modal emotion recognition based on facial expression and body posture in video sequences [J]. *Laser & Optoelectronics Progress*, 2018, 55(7): 071004.  
姜明星, 胡敏, 王晓华, 等. 视频序列中表情和姿态的双模态情感识别[J]. *激光与光电子学进展*, 2018, 55(7): 071004.
- [2] He Z C, Zhao L Z, Chen C. Convolution neural network with multi-resolution feature fusion for facial expression recognition [J]. *Laser & Optoelectronics Progress*, 2018, 55(7): 071503.  
何志超, 赵龙章, 陈闯. 用于人脸表情识别的多分辨率特征融合卷积神经网络[J]. *激光与光电子学进展*, 2018, 55(7): 071503.
- [3] Wang X Y, Fan J Z, Liu H M, et al. Multi-expression sequence fusion recognition based on probabilistic cooperative representation [J]. *Laser & Optoelectronics Progress*, 2019, 56(13): 131004.  
王秀友, 范建中, 刘华明, 等. 基于概率协作表示的多表情序列融合识别 [J]. *激光与光电子学进展*, 2019, 56(13): 131004.
- [4] Lucey P, Cohn J F, Kanade T, et al. The extended cohn-kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression [C] // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE, 2010: 94-101.
- [5] Li Y, Lin X Z, Jiang M Y. Facial expression recognition with cross-connect LeNet-5 network [J]. *Acta Automatica Sinica*, 2018, 44(1): 176-182.  
李勇, 林小竹, 蒋梦莹. 基于跨连接 LeNet-5 网络的面部表情识别 [J]. *自动化学报*, 2018, 44(1): 176-182.
- [6] Jiang D P, Yang B, Zou L. Facial expression recognition based on local binary mode convolution neural network [J]. *Computer Engineering and Design*, 2018, 39(7): 1971-1977.  
江大鹏, 杨彪, 邹凌. 基于 LBP 卷积神经网络的面部表情识别 [J]. *计算机工程与设计*, 2018, 39(7): 1971-1977.
- [7] Jain S, Hu C B, Aggarwal J K. Facial expression recognition with temporal modeling of shapes [C] // 2011 IEEE International Conference on Computer Vision Workshops, November 6-13, 2011, Barcelona, Spain. New York: IEEE, 2011: 1642-1649.
- [8] King D E. Dlib-ml: a machine learning toolkit [J]. *Journal of Machine Learning Research*, 2009, 10(3): 1755-1758.
- [9] Lopes A T, de Aguiar E, de Souza A F, et al. Facial expression recognition with convolutional neural networks: coping with few data and the training sample order [J]. *Pattern Recognition*, 2017, 61: 610-628.
- [10] Zhuo D, Jing J F, Zhang H H, et al. Classification of chopped strand mat defects based on convolutional neural network [J]. *Laser & Optoelectronics Progress*, 2019, 56(10): 101009.  
卓东, 景军锋, 张缓缓, 等. 基于卷积神经网络的短切毡缺陷分类 [J]. *激光与光电子学进展*, 2019, 56(10): 101009.
- [11] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C] // Proceedings of the 25th International Conference on Neural Information Processing Systems, December 3-6, 2012, Lake Tahoe, Nevada, USA. 2012: 1097-1105.
- [12] Szegedy C, Ioffe S, Vanhoucke V. Inception-v4, inception-ResNet and the impact of residual connections on learning [C] // Thirty First AAAI Conference on Artificial Intelligence, February 4-10, 2017, Hilton, San Francisco Union Square. Palo Alto, California: AAAI, 2017: 4278-4284.
- [13] 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, March 7-10, 2016, Lake Placid, NY, USA. New York: IEEE, 2016: 1-10.
- [14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2019-06-15]. <https://arxiv.org/abs/1409.1556>.
- [15] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, Nevada,

- USA. New York: IEEE, 2016: 770-778.
- [16] Srivastava R K, Greff K, Schmidhuber J. Training very deep networks [EB/OL]. [2019-06-15]. <https://arxiv.org/abs/1507.06228>.
- [17] Li K. Facial expression recognition based on shallow convolutional neural network[D]. Hefei: University of Science and Technology of China, 2019: 42-47.
- 李宽. 基于浅层卷积神经网络的人脸表情识别方法研究[D]. 合肥: 中国科学技术大学, 2019: 42-47.