

中图分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2019)02-0227-10

论文引用格式: Xu L L, Zhang S M, Zhao J L. Expression recognition algorithm for parallel convolutional neural networks [J]. Journal of Image and Graphics, 2019, 24(02): 0227-0236. [徐琳琳, 张树美, 赵俊莉. 构建并行卷积神经网络的表情识别算法 [J]. 中国图象图形学报, 2019, 24(02): 0227-0236.] [DOI: 10.11834/jig.180346]

构建并行卷积神经网络的表情识别算法

徐琳琳, 张树美, 赵俊莉

青岛大学数据科学与软件工程学院, 青岛 266071

摘要: **目的** 表情识别在商业、安全、医学等领域有着广泛的应用前景,能够快速准确地识别出面部表情对其研究与应用具有重要意义。传统的机器学习方法需要手工提取特征且准确率难以保证。近年来,卷积神经网络因其良好的自学习和泛化能力得到广泛应用,但还存在表情特征提取困难、网络训练时间过长等问题,针对以上问题,提出一种基于并行卷积神经网络的表情识别方法。**方法** 首先对面表情图像进行人脸定位、灰度统一以及角度调整等预处理,去除了复杂的背景、光照、角度等影响,得到了精确的人脸部分。然后针对表情图像设计一个具有两个并行卷积池化单元的卷积神经网络,可以提取细微的表情部分。该并行结构具有3个不同的通道,分别提取不同的图像特征并进行融合,最后送入SoftMax层进行分类。**结果** 实验使用提出的并行卷积神经网络在CK+、FER2013两个表情数据集上进行了10倍交叉验证,最终的结果取10次验证的平均值,在CK+及FER2013上取得了94.03%与65.6%的准确率。迭代一次的时间分别为0.185 s和0.101 s。**结论** 为卷积神经网络的设计提供了一种新思路,可以在控制深度的同时扩展广度,提取更多的表情特征。实验结果表明,针对数量、分辨率、大小等差异较大的表情数据集,该网络模型均能够获得较高的识别率并缩短训练时间。

关键词: 表情识别; 深度学习; 卷积神经网络; 并行处理; 图像分类

Expression recognition algorithm for parallel convolutional neural networks

Xu Linlin, Zhang Shumei, Zhao Junli

College of Data Science and Software Engineering, Qingdao University, Qingdao 266071, China

Abstract: **Objective** Face emotion recognition is widely applied in the fields of commercial, security, and medicine. Rapid and accurate identification of facial expressions are of great significance for their research and application. Several traditional machine learning methods, such as support vector machine (SVM), principal component analysis (PCA), and local binary pattern (LBP) are used to identify facial expressions. However, these traditional machine learning algorithms require manual feature extraction. In this process, some features are hidden or deliberately enlarged due to many human interventions, which affect accuracy. In recent years, convolutional neural networks (CNNs) have been used extensively in image recognition due to their good self-learning and generalization capabilities. However, several problems, such as difficulty in facial expression feature extraction and long training time of neural network, are still observed with neural network training. This study presents an expression recognition method based on parallel CNN to solve the aforementioned problems.

收稿日期: 2018-06-04; 修回日期: 2018-08-16; 预印本日期: 2018-08-23

基金项目: 国家自然科学基金项目(41501698); 国家自然科学基金青年科学基金项目(61702293); 虚拟现实应用教育部工程研究中心开放基金项目(ME0BNUEVRA201601)

Supported by: National Natural Science Foundation of China (41501698); Young Scientists Fund of National Natural Science Foundation of China (61702293)

Method First, a series of preprocessing operations is performed on facial expression images. For example, an original image is detected by using an AdaBoost cascade classifier to remove the complex background and obtain the face part. Then, a face image is compensated by illumination, a histogram equalization method is used to stretch the image nonlinearly, and the pixel value of the image is reallocated. Finally, affine transformation is used to achieve face alignment. The preceding preprocessing can remove complex background effects, compensate lighting, and adjust the angle to obtain more accurate face parts than that of the original image. Then, a CNN with two parallel convolution and pooling structures, which can extract subtle expressions, is designed for facial expression images. This parallel unit is the core unit of the CNN and comprises a convolutional layer, a pooling layer, and an activation function ReLu. This parallel structure has three different channels, in which each channel has different number of convolutions, pooling layers, and ReLu to extract different image features and fuse the extracted features. The second parallel processing unit can perform convolution and pooling on the extracted features by the first parallel processing unit and reduce the dimension of the image and shorten the training time of CNN. Finally, the previously merged features are sent to the SoftMax layer for expression classification. **Result** CK+ and FER2013 expression datasets that have undergone pre-processing and data enhancement are divided into 10 equal parts. Then, training and testing are performed on 10 parts, and the final accuracy is the average of the 10 results. Experimental results show that the accuracy increases and time decreases remarkably compared with traditional machine learning methods, such as SVM, PCA, and LBP or their combination and other classical CNNs, such as AlexNet and GoogLeNet. Finally, CK+ and FER2013 achieve 94.03% and 65.6% accuracy, and the iteration time reaches 0.185 s and 0.101 s, respectively. **Conclusion** This study presents a new parallel CNN structure that extracts the features of facial expressions by using three different convolutional and pooling structures. The three paths have different combinations of convolutional and pooling layers, and they can extract different image features. The different extracted features are combined and sent to the next layer for processing. This study provides a new concept for the design of CNNs, which can extend the breadth of CNN and control the depth. The proposed CNN can extract many expressions that are ignored or difficult to extract. CK+ and FER2013 expression datasets have large difference in quantity, size, and resolution. The experiments of CK+ and FER2013 show that the model can extract the precise and subtle features of facial expression images in a relatively short time under the premise of ensuring the recognition rate.

Key words: expression recognition; deep learning; convolutional neural network (CNN); parallel processing; image classification

0 引言

人类使用不同的形式进行交流,例如语言、肢体动作以及情绪表达等。正确理解一个人的表情和表情编码是社会交流的一个重要因素。近年来,随着人工智能领域研究的深入,将人脸识别、表情识别等功能集成到机器当中,可以促进人机交流更加自然和多样化发展。如 Emotichat 软件^[1]是一个具有情感识别的聊天应用程序,它对用户的表情进行监控,每当检测到表情(例如微笑等)时,就将相应的表情插入到聊天窗口。除了人机交互之外,表情识别在医学、教育以及安全等领域都有了相应的应用^[2]。

表情识别从 20 世纪 70 年代发展至今,经过了手工提取特征到自动提取特征的发展。传统的面部表情识别方法包括特征提取和表情分类。从训练数

据中提取到的特征在最后的分类中有着至关重要的作用,因为分类器是基于提取特征的组合来做出决定的。手工提取特征的方法主要有局部二值模式(LBP)、方向梯度直方图特征(HOG)、尺度不变特征转换(SIFT)等。Shan 等人^[3]提出了一种提升 LBP 算法,并将其与支持向量机(SVM)分类器相结合。该方法在低分辨率的面部图像上有较好的鲁棒性和稳定性。Berretti 等人^[4]在深度图像的 3D 面部标志上计算 SIFT 描述符,并用 SVM 方法进行分类。Albiol 等人^[5]提出了一种 HOG-EBGM(HOG-elastic bunch graph matching)算法,该算法使用 HOG 描述符代替 Gabor 特征,对光照、旋转和小位移的变化有更好的鲁棒性。手工提取特征具有较高的准确率,在特定的情况下被广泛应用,但也有其固有的缺点。手工提取特征是由人为提取的,在这个过程中会包含对分类没有影响的特征,或忽略了对分类有较大

影响的重要特征。随着近年来深度学习和并行计算的迅速发展,将基于卷积神经网络的深度神经网络应用到表情分类问题引起了人们的关注。深度学习方法区别于传统的机器学习方法,它可以同时执行特征提取和表情分类。并且特征提取是通过反向传播和误差优化来更新迭代权重,所以可以提取出人类很难想到的关键点和预想不到的特征。Szegedy等人^[6]提出了GoogLeNet架构,是一个27层深的卷积神经网络,在ILSVRC2014分类挑战赛中,这个网络获得了6.67%的top5错误率。Burkert等人^[7]提出了一个名为DeXpression的网络,它使用了两个并行特征提取块,该架构在MMI和CK+数据集上都取得了很好的效果。杨格兰等人^[8]提出了一种基于深度时空域的卷积神经网络,采用数据驱动策略和新颖的卷积滤波响应,使得模型能够更好地抽取到视频中的动态特征和静态特征。深度学习的算法取得了较高的识别率,但还有一些缺点,如对训练数据的样本依赖性过大,且图像预处理过于耗时。存在神经网络识别率较高但训练时间过长或训练时间短但识别率不高的问题。

针对以上问题,本文对CNN结构进行优化,提出了一种基于并行结构的卷积神经网络。核心为两个并行的卷积池化结构,并行结构有3个不同的分支,可以分别对表情图像进行特征提取,再将提取出的特征连接起来送入下一层进行处理。在CK+和FER2013数据集上进行训练和测试,其正确率和速率都有明显的提高。

1 卷积神经网络

卷积神经网络是一个端到端的有监督学习过程,它包含输入层、卷积层、池化层、激活函数、全连接层、SoftMax层以及输出层等多个结构。卷积神经网络使用多个层结构组成的计算模型来学习具有多个抽象级别的数据表示,并通过反向传播算法发现大型数据集中的复杂结构。目前在处理图像、视频、语音以及文本等方面取得了较好的成果^[9]。下面具体介绍卷积神经网络的层结构。

1.1 卷积层

卷积层执行的是卷积操作,输入层输入的图片通过多个可训练的滤波器和可加偏执来进行卷积。在每一个卷积层中,上一层输出的表情特征图被一

个可学习的卷积核进行卷积,再经过一个激活函数,得到输出的表情特征图。卷积之后产生的表情映射特征图输出提取到的表情特征。表情特征图的计算为

$$G_i = f(G_{i-1} * W_i + b_i) \quad (1)$$

式中, G_i 为卷积层, f 为激活函数, W_i 为第*i*层的卷积核的权值向量, b_i 为偏移向量。

1.2 池化层

池化层通常跟在卷积层的后面,根据一定的规则对特征图进行下采样^[10]。作用是可以对特征图进行降维并且能够在一定程度上保持特征的尺度不变性。本文采用了两种池化方法:最大池化以及平均池化。最大池化是在输入特征图上使用最大函数来减少输入,平均池化则是使用平均函数。池化公式为

$$G_i = \text{subsampling}(G_{i-1}) \quad (2)$$

式中, G_i 为下采样层, subsampling 为相应的池化规则。

1.3 激活函数

激活函数是神经网络的一个单元,它可以将特征保留并映射出来,并加入了非线性的因素,以便更好地分类。本文采用的是Relu激活函数^[11],它比Sigmoid函数更加有效,不需要计算指数函数,减少了计算量,并且可以有效地防止梯度误差的消失。其计算为

$$R(x) = \max(0, x) \quad (3)$$

1.4 全连接层

全连接层也称为多层感知器,它将前一层所有的神经元与自身层的每个神经元相连接,用来将前面提取到的特征综合起来,基于这一特点,全连接层是神经网络中参数最多的一层。其计算公式为

$$F(x) = f(x * W + b) \quad (4)$$

式中, $F(x)$ 为全连接层, f 为激活函数, W 为全连接层卷积核的权值向量, b 为偏移向量。

1.5 SoftMax层

SoftMax回归是Logistic回归的改进,Logistic用于二分类问题,SoftMax可以用于多分类问题^[12]。本文使用SoftMax对7种表情进行分类。对于表情图像 x ,其类别*j*的概率为

$$p(y^{(i)} = j | x^{(i)}, \theta) = \frac{e^{\theta_j^T x^{(i)}}}{\sum_{i=1}^k e^{\theta_j^T x^{(i)}}} \quad (5)$$

式中, $p(y^{(i)} = j | x^{(i)}, \theta)$ 为表情图像 x 对应于每一个类别 j 的概率, $j \in \{0, 1, 2, \dots, 6\}$ 。 θ 为待拟合的参数。

2 并行卷积神经网络

2.1 图像预处理

表情图像中存在复杂背景、光照过强或过弱、角度偏差等问题。因此在送入卷积神经网络训练之前, 首先进行人脸检测、光照补偿、角度调整等预处理。本文使用了 CK+、FER2013 两个数据集。其中 CK+ 是在实验室环境下获取的正脸图像。FER2013 是在自然环境下获取的表情图像。下面具体介绍方法。

2.1.1 CK+ 数据集

Cohn-Kanade 扩展数据集(CK+) [13] 是 2010 年 Cohn-Kanade 数据集的扩展版本。它是从视频中截取的图像, 共有 123 名参与者、593 个图像序列, 包含了从平静到表情峰值再到平静的变化。该数据集是目前较为通用的人脸表情数据集。CK+ 数据集预处理方式如下:

1) 人脸检测。本文使用 Adaboost 级联分类器对原始人脸图像进行人脸检测, 获取图像中的人脸部分, 去除复杂的背景、文字等干扰信息。效果如图 1 所示。

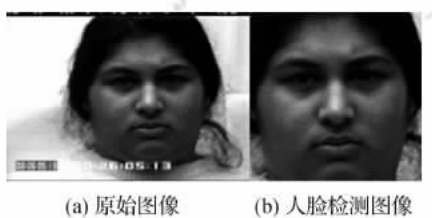


图 1 人脸检测效果图

Fig. 1 Face detection effect

((a) original image; (b) face detection effect)

2) 光照补偿。光照补偿能够在一定程度上减少光照对图像的影响。本文使用的数据集都为灰度图像, 使用了直方图均衡化对图像进行非线性拉伸, 重新分配图像的像素值。效果如图 2 所示。

3) 角度调整。角度调整是将图像中的人脸都调整到同一角度, 本文使用仿射变换实现人脸对齐。效果如图 3 所示。



图 2 光照补偿效果图

Fig. 2 Lighting compensation effect

((a) original images; (b) histogram equalization images)

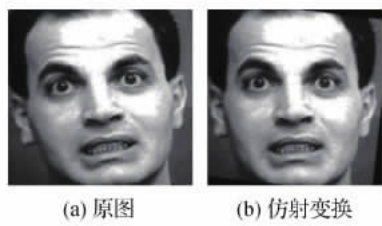


图 3 角度调整效果图

Fig. 3 Angle adjustment effect

((a) original image; (b) affine transformation)

2.1.2 FER2013 数据集

FER2013 数据集(the facial expression recognition 2013) 是 2013 年 Kaggle 面部识别挑战赛使用的一个数据集 [14]。它是由 Pierre 和 Aaron 通过谷歌关键字搜索得到的。共包含 35 887 幅人脸表情图像, 分辨率为 48×48 像素。由于该数据集是非实验室环境下得到的, 且包括遮挡、侧脸、角度等多种干扰因素的真实表情。所以目前在该数据集上的识别较为困难。对 FER2013 数据集预处理方式如下:

FER2013 数据集中人脸图像的大小为 48×48 像素, 使用数据增强的方法, 将人脸图像随机剪裁为 10 个 42×42 像素的图像, 然后对剪裁后的人脸图像分别进行镜像处理, 最终将数据集扩大了 20 倍。通过这一方法, 扩大了数据集的数量并实现了空间不变性。效果如图 4 所示。

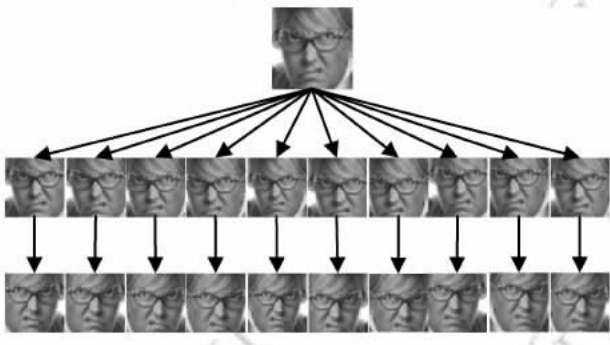


图 4 数据增强效果图

Fig. 4 Data enhancement

2.2 并行卷积神经网络架构设计

卷积神经网络通过卷积层提取特征,通过池化层降低特征图的维度并进一步提炼特征,全连接层用于将所有特征综合起来。不同的层结构提取出的

特征不同,本文提出的卷积神经网络使用了并行的卷积池化处理单元,可以提取出不同的特征并综合起来输出给下一层,还可以进一步提升速率。神经网络结构如图5所示。

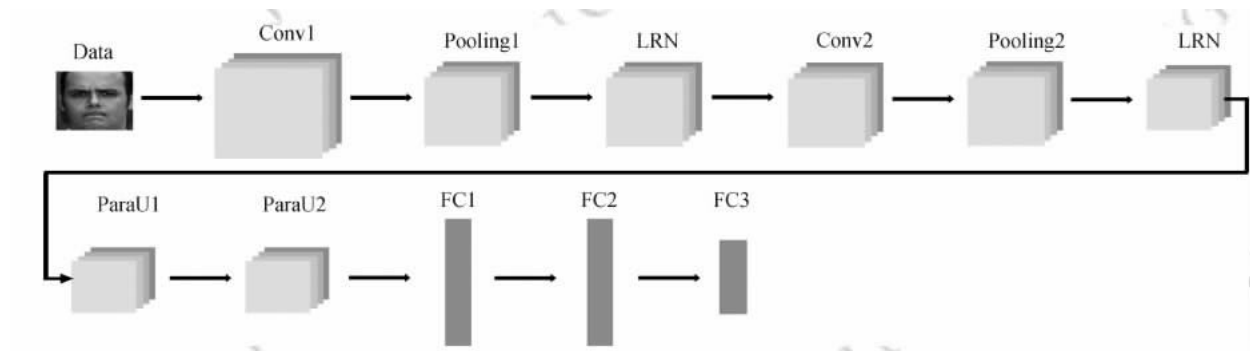


图5 并行卷积神经网络架构

Fig. 5 Parallel convolutional neural network architecture

图5中,ParaU单元是该神经网络的核心单元,它由卷积层、池化层和激活函数 Relu 组成,共分为3个分支,分别对输入的表情特征图进行特征提取,然后再连接起来输出给下一层。本文首先对 ParaU 单元的使用数量进行了对比实验,发现使用一个 ParaU 单元时,由于对特征提取不充分,识别率没有达到较高水平,后期随着使用 ParaU 单元数量的增加,识别率上升到一定水平之后没有再增加,但时间却不断增加。根据识别率与时间的最优规划,这里选择使用两个 ParaU 单元。ParaU 单元的内部构造如图6所示。

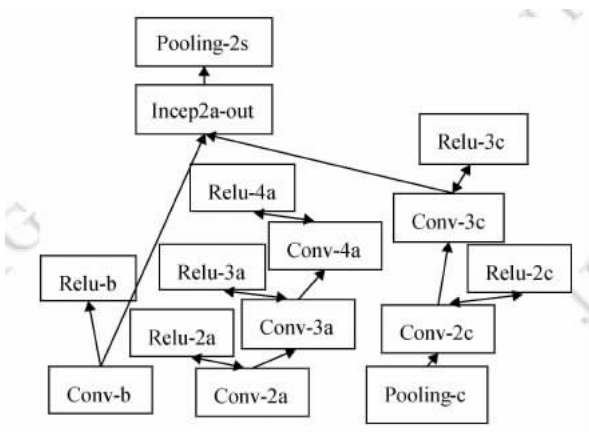


图6 ParaU 单元架构

Fig. 6 ParaU unit architecture

该模型包括16个卷积层,2个连接层和3个全连接层。前两个卷积层分别有64个卷积核,对应的卷积核大小分别为 7×7 和 3×3 。ParaU1中包括

7个卷积层,a通道使用了3个连续的卷积,分别有64个、128个、128个卷积核,对应的卷积核大小分别是 1×1 、 3×3 、 3×3 。b通道使用了两个连续的卷积,分别有64个、28个卷积核,对应的卷积核大小分别是 1×1 、 3×3 。c通道在卷积层之前加入了一个池化层,用以减少图像的尺寸,后面使用了两个连续的卷积层,均使用了128个卷积核,对应的卷积核大小是 1×1 、 3×3 。3个并行通道分别提取特征之后,提取出的特征在 Incep2a-out 层中连接起来,送入池化层,缩小尺寸并进一步提取特征之后送入下一个并行处理单元。在两个并行单元之后,连接了3个全连接层。这里为了防止过拟合的问题,在3个全连接层上都加入了 Dropout 层^[15],其值为0.4。最后输入给 SoftMax 层,用以实现表情图像的分类。参数设置如表1所示。

该网络模型设置了不同的层结构,不同的层结构产生不同的特征向量。图7提供了该模型的可视化。可以从可视化中看出,前面单元对于图像的抽象化程度较低,并且大多数边缘的抽象程度较低。第1个并行模块创建了3个不同比例的并行路径,这3个路径分别提取了3种不同的特征,并在 Incep2a-out 中组合在一起。第2个并行模块对上一步综合的图像特征图进一步提取特征,同时降低了维度。该可视化显示,ParaU 块的使用是创建抽象特征表示的有效方法,并且能够创建多条路径来提取不同的特征。

本文设计的 CNN 网络由于采用并行结构有效地改善了对细微表情特征提取不够、神经网络训练时间过长等缺点。其主要设计优点有:一是学习并

表 1 神经网络参数
Table 1 Neural network parameters

层	卷积核/步长	输出	Conv-a/d	Conv-2a/2d	Conv-3a/3d	Conv-b/e	Conv-2c/2f	Conv-3c/3f	参数数量/k
Conv1	$7 \times 7/2$	$21 \times 21 \times 64$							3.2
Pooling1	$3 \times 3/2$	$10 \times 10 \times 64$							0
Conv2	$3 \times 3/1$	$10 \times 10 \times 192$							110
Pooling2	$3 \times 3/2$	$5 \times 5 \times 192$							0
ParaU1			64	96	128	64	32	64	215
ParaU2			64	96	128	64	32	64	225
FC1		$1 \times 1 \times 4\,096$							1\,052
FC2		$1 \times 1 \times 1\,024$							4\,195
FC3		$1 \times 1 \times 7$							7.1

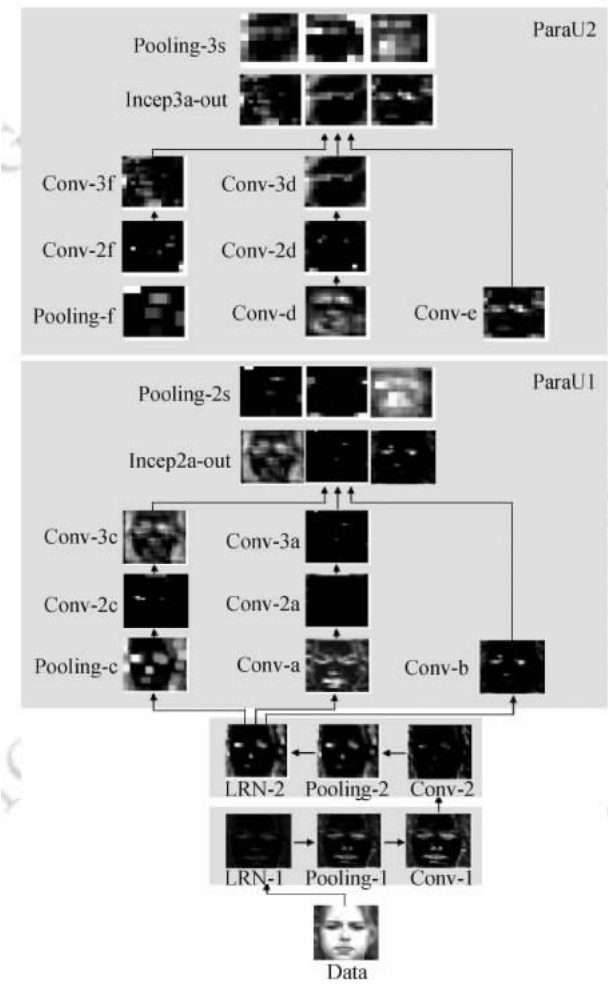


图 7 可视化图像

Fig. 7 Visual image

优化了 GoogLeNet 结构,采用了并行卷积池化处理结构,采用多个分支分别对表情图像提取特征并进行综合,使得图像表情特征的代表能力更强;二是网

络均采用了 1×1 或 3×3 的小卷积核,不仅可以加强对表情特征的提取能力并且计算简单,大大缩短了神经网络的训练时间。

3 实验与结果

3.1 实验准备

实验环境为: 64 位 Microsoft Windows 10 操作系统, CPU 为 E5-4620 v4, 主频 3.5 GHz, 显卡型号为 NVIDIA Quadro M2000, 显存为 2 GB。使用基于 Caffe 的深度学习平台。采集 CK+ 数据集 2 940 幅, FER2013 数据集训练集 572 796 幅, 测试集 71 620 幅。由于数量的差异, CK+ 处理为 227×227 像素大小进行输入, FER2013 处理为 42×42 像素大小进行输入。为了保证实验结果的有效性, 本次实验采取了 10 倍交叉验证。其中将 CK+ 数据集分别平均分为 10 组。每次选取其中的 9 组作为训练集, 另外 1 组作为测试集, 进行 10 次实验, 最后的结果取 10 次实验结果的平均值。FER2013 数据集将原始测试集的扩展图像分为 10 份, 每次取 1 份进行测试, 最后取 10 个预测结果的均值作为最终得分。这样保证了每个样本都可以作为测试集和训练集, 避免因为某些数据误差较大而影响最终的识别率。

实验中根据两个数据集的数量不同, 设置了不同的网络训练参数。表 2 给出了两个数据集具体的参数设置。

3.2 实验结果

经过 10 万次的迭代训练之后, 计算 10 倍交叉

验证后的最终结果为: CK+ 数据集得到了 94.03% 的准确率, FER2013 数据集得到了 65.6% 的准确率。在 GPU 加速运算的情况下 CK+ 迭代 10 万次约需要 327 min, FER2013 迭代 10 万次约需要 175 min。图 8 和图 9 分别给出了 CK+ 和 FER2013 数据集随机挑选的一次在训练过程中正确率和损失函数的变化情况。

由图 8、图 9 可以看出迭代至 10 万次时, 训练的损失函数降到很低的位置且基本保持稳定, 说明

训练的误差已经很小。测试的识别率基本保持平稳的状态, 不再发生变化, 说明最后的结果已得到充分收敛。

为了比较本文所提出的并行卷积神经网络的性能, 与其他几种机器学习方法以及卷积神经网络, 包括 SVM、LBP、Gabor 和 AlexNet、GoogLeNet 等进行了对比实验。其中关于卷积神经网络的对比实验均使用了相同的预处理和 10 倍交叉验证的方法。识别结果如表 3 和表 4 所示。

表 2 图像网络训练参数
Table 2 Image network training parameters

数据集	学习率	测试迭代次数	权值衰减	批次数量	最大迭代次数
CK+	0.000 01	200	0.000 5	20	100 000
FER2013	0.001	100	0.000 5	100	100 000

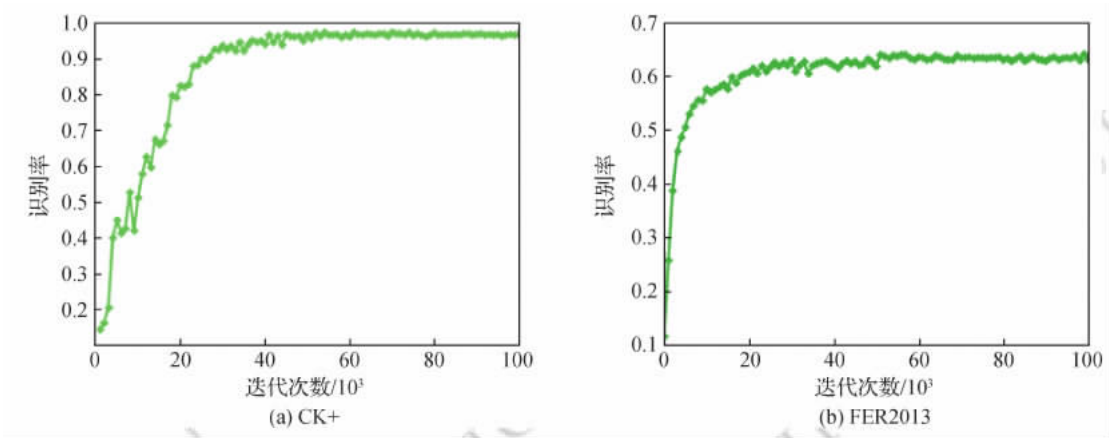


图 8 识别率变化
Fig. 8 Changes in accuracy((a) CK+ ; (b) FER2013)

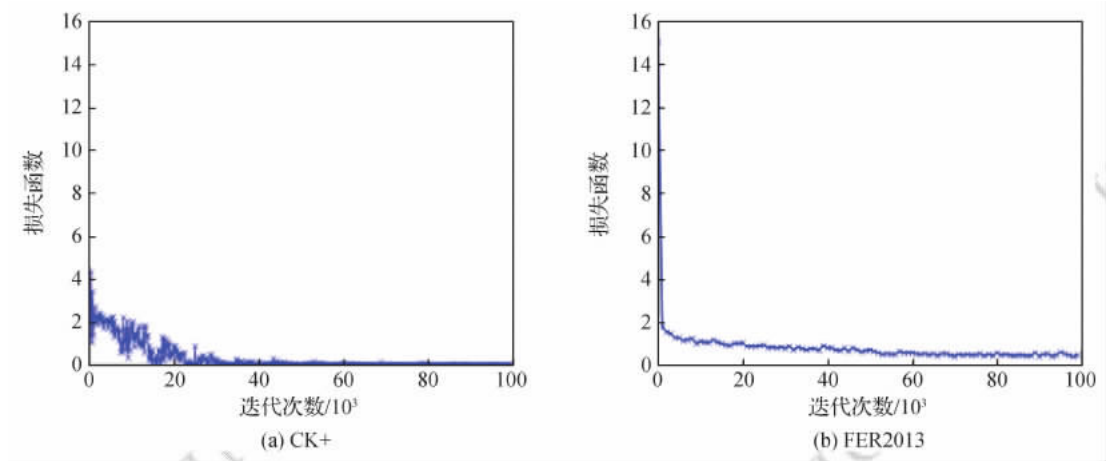


图 9 损失函数的变化
Fig. 9 Changes of loss function ((a) CK+ ; (b) FER2013)

表3 CK+ 识别率对比

Table 3 CK+ accuracy comparison

方法	识别率/%	方法	识别率/%
AAM + SVM ^[16]	89.2	AlexNet	87.03
LBP/VAR + DBN ^[17]	91.4	GoogLeNet	91.29
Gabor + PCA + CRF ^[18]	91.26	本文算法	94.03

表4 CK+ 时间对比

Table 4 CK+ time comparison

方法	时间/s
AlexNet	0.146
GoogLeNet	0.3
本文算法	0.185

从表3、表4的实验结果可以看出,本文提出的算法识别率无论是相较于传统的机器学习方法还是卷积神经网络都有了明显提升。其中卷积神经网络的算法中,AlexNet 迭代一次的时间为 0.146 s,本文所提出的网络迭代一次的时间为 0.185 s,GoogLeNet 由于参数以及层数的原因,迭代一次所耗费的时间最长,为 0.3 s。由表可知,本文提出的网络在保证速率的情况下,识别率有了明显提高。表5给出了 FER2013 数据集在面部表情识别挑战赛上与其他模型识别率对比的结果。

表5 FER2013 识别率对比

Table 5 FER2013 accuracy comparison

名次	团队	识别率/%
1	RBM	71.16
2	Unsupervised	69.26
3	Maxim Milakov	68.82
4	Radu + Marius + Cristi	67.49
5	本文算法	65.6
⋮	⋮	⋮
10	sayit	62.19
⋮	⋮	⋮
50	Salty Pere	31.03

目前人类肉眼在 FER2013 数据集上的平均识别率约为 $65 \pm 5\%$ ^[19],本文算法已基本能达到肉眼识别的效果。其中本文所提出的网络迭代一次的时间约为 0.101 s,已基本可以满足实际需要。表6和

表7提供了 CK+ 和 FER2013 数据集的混淆矩阵。

表6 CK+ 混淆矩阵

Table 6 CK+ confusion matrix

	/%						
	生气	厌恶	恐惧	高兴	平静	悲伤	惊讶
生气	97.62	2.38	0	0	0	0	0
厌恶	2.38	92.86	2.38	0	2.38	0	0
恐惧	4.76	7.14	88.10	0	0	0	0
高兴	0	0	0	97.62	0	0	2.38
平静	2.38	0	0	0	97.62	0	0
悲伤	7.14	0	0	0	2.38	90.47	0
惊讶	0	0	4.76	0	0	0	95.24

表7 FER2013 混淆矩阵

Table 7 FER2013 confusion matrix

	/%						
	生气	厌恶	恐惧	高兴	平静	悲伤	惊讶
生气	45.92	1.84	11.22	5.51	18.37	12.86	4.28
厌恶	18.18	58.20	5.45	5.45	10.91	0	1.81
恐惧	9.68	1.14	42.32	4.93	14.61	14.99	12.33
高兴	2.28	0.11	1.71	83.16	6.37	3.3	3.07
平静	4.98	0.48	4.50	5.47	73.16	9.16	2.25
悲伤	9.61	1.35	11.64	4.05	27.32	43.34	2.69
惊讶	2.42	0.24	5.56	4.83	4.11	1.68	81.16

由混淆矩阵可以看出,生气、厌恶、恐惧、悲伤这几种表情识别难度较大。恐惧通常被归为悲伤,因为它们都具有拉开嘴唇和紧张前额的相同特征。悲伤与平静容易被混淆,在自然状态下,人的眼角、嘴角会微微下拉,导致分类错误。此外生气与厌恶因为具有相同的眉毛特征、狭窄以及皱起的嘴角等特征也更容易发生错误的分类。这更加证明了表情识别是一项复杂而且模糊的研究,人类的表情是非常复杂的,多种表情的混合出现,如悲伤和愤怒般的厌恶、高兴和惊讶的惊喜等增大了识别的难度。

4 结 论

本文提出了基于并行卷积神经网络的人脸表情识别算法,并在 CK+ 和 FER2013 上进行了实验,取

得了 94.03% 和 65.6% 的识别率,证明该网络结构在降低训练时长的基础上获得了较高的识别率。而这两个数据集在获取方式、图像大小、分辨率以及数量上有着较大的差异,能够获得良好的实验结果同样证明了该网络结构具有较好的鲁棒性。虽然取得了一定的成果,但还存在一些问题,例如存在一定程度的过拟合,识别率有待进一步提高等。在接下来的工作中会继续探索网络结构以及参数的设置,加强对细微表情的特征提取,细化表情感兴趣区域等。

参考文献(References)

- [1] Anderson K, McOwan P W. A real-time automated system for the recognition of human facial expressions [J]. IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics), 2006, 36(1): 96-105. [DOI: 10.1109/TSMCB.2005.854502]
- [2] Pramerdorfer C, Kampel M. Facial expression recognition using convolutional neural networks: state of the art [EB/OL]. [2018-05-20]. <https://arxiv.org/pdf/1612.02903.pdf>
- [3] Shan C F, Gong S G, McOwan P W. Facial expression recognition based on Local Binary Patterns: A comprehensive study [J]. Image and Vision Computing, 2009, 27(6): 803-816. [DOI: 10.1016/j.imavis.2008.08.005]
- [4] Berretti S, Ben Amor B, Daoudi M, et al. 3D facial expression recognition using SIFT descriptors of automatically detected key-points [J]. The Visual Computer, 2011, 27(11): 1021-1036. [DOI: 10.1007/s00371-011-0611-x]
- [5] Albiol A, Monzo D, Martin A, et al. Face recognition using HOG-EBGM [J]. Pattern Recognition Letters, 2008, 29(10): 1537-1543. [DOI: 10.1016/j.patrec.2008.03.017]
- [6] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions [C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA: IEEE, 2015: 1-9. [DOI: 10.1109/CVPR.2015.7298594]
- [7] Burkert P, Trier F, Afzal M Z, et al. DeXpression: deep convolutional neural network for expression recognition [EB/OL]. [2018-05-20]. <https://arxiv.org/pdf/1509.05371.pdf>
- [8] Yang G L, Deng X J, Liu C. Facial expression recognition model based on deep spatiotemporal convolutional neural networks [J]. Journal of Central South University: Science and Technology, 2016, 47(7): 2311-2319. [杨格兰, 邓晓军, 刘琮. 基于深度时空域卷积神经网络的表情识别模型 [J]. 中南大学学报: 自然科学版, 2016, 47(7): 2311-2319.] [DOI: 10.11817/j.issn.1672-7207.2016.07.037]
- [9] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521(7553): 436-444. [DOI: 10.1038/nature14539]
- [10] Zeiler M D, Fergus R. Stochastic pooling for regularization of deep convolutional neural networks [EB/OL]. [2018-05-20]. <https://arxiv.org/pdf/1301.3557.pdf>
- [11] Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines [C]//Proceedings of the 27th International Conference on International Conference on Machine Learning. Haifa, Israel: Omnipress, 2010: 807-814.
- [12] Wei Z. Research and implementation of face recognition based on deep learning based on Caffe platform [D]. Xi'an: Xi'an University of Electronic Science and Technology, 2015. [魏正. 基于Caffe平台深度学习的人脸识别研究与实现 [D]. 西安: 西安电子科技大学, 2015.]
- [13] 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]//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. San Francisco, CA: IEEE, 2010: 94-101. [DOI: 10.1109/CVPRW.2010.5543262]
- [14] Kotsia I, Pitas I. Facial expression recognition in image sequences using geometric deformation features and support vector machines [J]. IEEE Transactions on Image Processing, 2007, 16(1): 172-187. [DOI: 10.1109/TIP.2006.884954]
- [15] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors [EB/OL]. [2018-05-20]. <https://arxiv.org/pdf/1207.0580.pdf>
- [16] Fu Q M, Liu Q, Wang H, et al. A novel off policy Q(λ) algorithm based on linear function approximation [J]. Chinese Journal of Computers, 2014, 37(3): 677-686. [傅启明, 刘全, 王辉, 等. 一种基于线性函数逼近的离策略 Q(λ) 算法 [J]. 计算机学报, 2014, 37(3): 677-686.] [DOI: 10.3724/SP.J.1016.2013.00677]
- [17] He J, Cai J F, Fang L Z, et al. Facial expression recognition based on LBP/VAR and DBN model [J]. Application Research of Computers, 2016, 33(8): 2509-2513. [何俊, 蔡建峰, 房灵芝, 等. 基于 LBP/VAR 与 DBN 模型的人脸表情识别 [J]. 计算机应用研究, 2016, 33(8): 2509-2513.] [DOI: 10.3969/j.issn.1001-3695.2016.08.060]
- [18] Zhang B. Facial expression recognition based on Gabor and conditional random field [D]. Jinan: Shan Dong University, 2015. [张博. 基于 Gabor 和条件随机场的人脸表情识别 [D]. 济南: 山东大学, 2015.]
- [19] Goodfellow I J, Erhan D, Carrier P L, et al. Challenges in representation learning: A report on three machine learning contests [J]. Neural Networks, 2015, 64: 59-63. [DOI: 10.1016/j.neunet.2014.09.005]

作者简介



徐琳琳, 1992 年生, 女, 硕士研究生, 主要研究方向为图像识别与处理、深度学习。

E-mail: 651278169@ qq. com



张树美, 通信作者, 女, 教授, 主要研究方向为时滞非线性系统的分析与控制、图像识别与处理。

E-mail: shumeiz@ 163. com

赵俊莉, 女, 博士, 助理教授, 主要研究方向为计算机视觉、计算机图形学、虚拟现实。E-mail: zhaojl@ yeah. net

超星