Journal of Computer Applications
计算机应用,2020,40(4):990 - 995
ISSN 1001-9081
CODEN JYIIDU
2020-04-10
http://www.joca.cn

文章编号:1001-9081(2020)04-0990-06
DOI:10.11772/j.issn.1001-9081.2019081438

基于深度可分卷积神经网络的实时人脸表情和性别分类

刘尚旺*,刘承伟,张爱丽
(河南师范大学 计算机与信息工程学院,河南 新乡 453007)
(*通信作者电子邮箱shwl2012@Hotmail.com)

摘　要:针对目前普通卷积神经网络(CNN)在表情和性别识别任务中出现的训练过程复杂、耗时过长、实时性差等问题,提出一种深度可分卷积神经网络的实时人脸表情和性别识别模型。首先,利用多任务级联卷积网络(MTCNN)对不同尺度输入图像进行人脸检测,并利用核相关滤波(KCF)对检测到的人脸位置进行跟踪进而提高检测速度。然后,设置不同尺度卷积核的瓶颈层,用通道合并的特征融合方式形成核卷积单元,以具有残差块和可分卷积单元的深度可分卷积神经网络提取多样化特征,并减少参数数量,轻量化模型结构;使用实时启用的反向传播可视化来揭示权重动态的变化并评估了学习的特征。最后,将表情识别和性别识别两个网络并联融合,实现表情和性别的实时识别。实验结果表明,所提出的网络模型在FER-2013数据集上取得73.8%的识别率,在CK+数据集上的识别率达到96%,在IMDB数据集中性别分类的准确率达到96%;模型的整体处理帧率达到80 frame/s,与结合支持向量机的全连接卷积神经网络方法所得结果相比,有着1.5倍的提升。因此针对数量、分辨率、大小等差异较大的数据集,该网络模型检测快,训练时间短,特征提取简单,具有较高的识别率和实时性。

关键词:深度可分卷积神经网络;面部检测;性别分类;情感分类;特征提取
中图分类号:TP391.4　　**文献标志码:**A

Real-time facial expression and gender recognition based on depthwise separable convolutional neural network

LIU Shangwang*, LIU Chengwei, ZHANG Aili
(*College of Computer and Information Engineering, Henan Normal University, Xinxiang Henan* 453007, *China*)

Abstract: Aiming at the problem of the current common Convolutional Neural Network (CNN) in the expression and gender recognition tasks, that is training process is complicated, time-consuming, and poor in real-time performance, a real-time facial expression and gender recognition model based on depthwise separable convolutional neural network was proposed. Firstly, the Multi-Task Convolutional Neural Network (MTCNN) was used to detect faces in different scale input images, and the detected face positions were tracked by Kernelized Correlation Filter (KCF) to increase the detection speed. Then, the bottleneck layers of convolution kernels of different scales were set, the kernel convolution units were formed by the feature fusion method of channel combination, the diversified features were extracted by the depthwise separable convolutional neural network with residual blocks and separable convolution units, and the number of parameters was reduced to lightweight the model structure. Besides, real-time enabled backpropagation visualization was used to reveal the dynamic changes of the weights and characteristics of learning. Finally, the two networks of expression recognition and gender recognition were combined in parallel to realize real-time recognition of expression and gender. Experimental results show that the proposed network model has a recognition rate of 73.8% on the FER-2013 dataset, a recognition rate of 96% on the CK+ dataset, the accuracy of gender classification on the IMDB dataset reaches 96%; and this model has the overall processing speed reached 70 frames per second, which is improved by 1.5 times compared with the method of common convolutional neural network combined with support vector machine. Therefore, for datasets with large differences in quantity, resolution and size, the proposed network model has fast detection, short training time, simple feature extraction, and high recognition rate and real-time performance.

Key words: depthwise separable convolutional neural network; face detection; gender recognition; facial expression recognition; feature extraction

0　引言

随着感知技术的发展,人体特征检测和识别成为研究热点。而人的面部特征是交流的关键因素,能够表现丰富的情感信息和性别特点,利用图像处理技术和深度学习对人脸表情和性别识别在智慧教育、公共安全监控、远程医疗中有着重要的作用。而目前的实际运用中,大多数模型难以处理背景复杂、有遮挡的多角度人脸图像,如Jeon等[1]使用方向梯度直

收稿日期:2019-08-19;**修回日期:**2019-11-01;**录用日期:**2019-11-11。
基金项目:河南省科技攻关项目(192102210290);河南省高等学校重点科研项目基础研究计划(18A510014)。
作者简介:刘尚旺(1973—),男,河南新乡人,副教授,博士,CCF会员,主要研究方向:生物图像处理、计算机视觉;　刘承伟(1996—),男,河南信阳人,硕士研究生,主要研究方向:计算机视觉、深度学习;　张爱丽(1966—),女,河南滑县人,教授,主要研究方向:信号处理、通信与网络。

方图(Histogram of Oriented Gradients, HOG)特征来检测人脸以减少光照不均匀对表情识别的影响,利用SVM在FER-2013数据集上实现了70.7%的表情识别率;但该方法抗干扰能力弱,适应性差。张延良等[2]提出通过面部关键点坐标将与微表情相关的七个局部区域串联构成特征向量来进行微表情识别,但存在局部区域微表情识别率低的缺点。罗珍珍等[3]等利用条件随机森林和支持向量机(Support Vector Machine, SVM)算法来检测人脸微笑情绪特征。戴逸翔等[4]利用智能穿戴设备来获取脑电、脉搏和血压三类生物信息,利用稀疏自编码方法对多模态情绪进行分析与识别,因为需要给每一位测试者佩戴设备,这无疑存在着成本过高不能大规模使用的局限性。

目前,有效解决自然场景下的图像分类和物体检测等图像相关任务的方法主要有传统机器学习和卷积神经网络(Convolutional Neural Networks, CNN)的方法。传统机器学习的方法一般采用手工设计特征,并利用分类器算法进行表情判定。典型的表情特征提取方法有主元分析(Principal Component Analysis, PCA)法[5]、局部二值模式(Local Binary Pattern, LBP)[6]、Gabor小波变换[7]、尺度不变的特征变换(Scale Invariant Feature Transform, SIFT)[8]等,常用的分类方法主要有隐马尔可夫模型(Hidden Markov Model, HMM)[9]、*K* 最近邻(*K*-Nearest Neighbor, *K*NN)算法[10]等。

相比传统机器学习,深度神经网络能够自主学习特征,减少了人为设计特征造成的不完备性。Tang[11]提出将CNN与SVM相结合,并且放弃了全连接CNN所使用的交叉熵损失最小化方法,而使用标准的铰链损失来最小化基于边界的损失,在其测试集上实现了71.2%的识别率。MobileNet-V2[12]中采用了多尺度核卷积单元主要以深度可分离卷积为基础,分支中采用了的线性瓶颈层结构,对表情进行了分类获得了70.8%的识别率。Li等[13]提出了一种新的保持深度局域的CNN方法,旨在通过保持局部紧密度的同时最大化类间差距来增强表情类别间的辨别力。Kample等[14]通过构建级联CNN来提高表情识别的精度。徐琳琳等[15]针对网络训练时间过长等问题,提出一种基于并行卷积神经网络的表情识别方法,获得了65.6%的准确率。CNN常被用作黑盒子,它将学习到的特征隐藏,使得在分类的准确性和不必要的参数数量之间难以抉择。为此Szegedy等[16]提出利用导向梯度反向传播的实时可视化,来验证CNN学习的特征。

对FER-2013数据集上的“愤怒”“厌恶”“恐惧”“快乐”“悲伤”“惊讶”和“中性”等表情进行识别[16],是非常困难的(见图1),需要表情分析和性别识别模型具有较强的鲁棒性和较高的计算效率。

图1　FER-2013情感数据集的样本

Fig. 1　Samples in FER-2013 emotion dataset

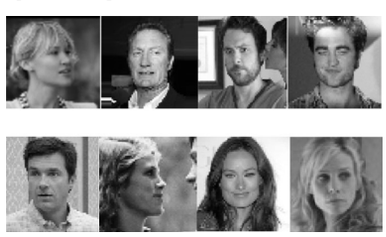


图2　IMDB数据集的样本

Fig. 2　Samples in IMDB dataset

1　本文方法

完整的实时表情和性别识别模型包括三个流程:人脸的检测与定位、特征提取和分类。针对实际应用对于人脸检测的准确度高和响应速度快的需求,使用MTCNN(Multi-Task CNN)网络对输入图像进行人脸检测,利用KCF(Kernelized Correlation Fiter)跟踪器进行人脸的定位跟踪,将人脸图像归一化输入深度可分卷积神经网络进行分类。最后,将表情识别和性别识别两个网络并联融合。图3是实时人脸表情和性别识别模型的总体框架。

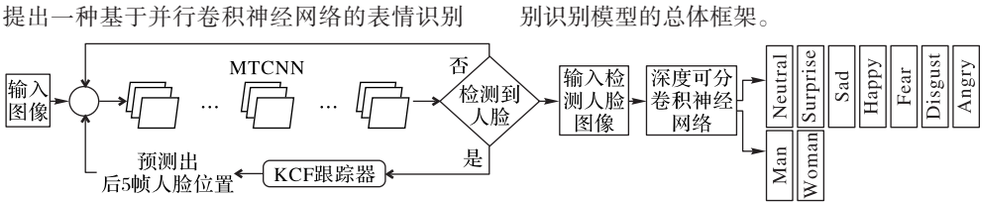


图3　人脸表情和性别识别框架

Fig. 3　Facial expression and gender recognition framework

1.1　多尺度人脸检测与跟踪

MTCNN算法使用图像金字塔,可适应不同尺度的人脸图像,网络结构如图4所示。该算法由快速生成候选窗口的P-Net(Proposal Netwaork)、进行高精度候选窗口过滤选择的R-Net(Refine Network)和生成最终边界框与人脸检测点的O-Net(Output Network)三层网络级联组成。通过人脸关键点来对齐不同角度的人脸,网络由粗到细,使用降低卷积核数量和大小、增加网络深度和候选框加分类的方式,进行快速高效的人脸检测。

加入KCF跟踪算法不仅能够解决实际运用中人脸图像角度多、有遮挡的检测问题,还能提高人脸检测速度。该算法使用目标周围区域的循环矩阵采集正负样本,利用脊回归训练目标检测器,并通过循环矩阵在傅里叶空间可对角化的性质将矩阵的运算转化为向量的Hadamad积,即元素的点乘,降低了运算量。先使用MTCNN算法对人脸进行检测,将检测的人脸坐标信息传递给跟踪算法KCF中,以此作为人脸检测基础样本框,并采用检测1帧、跟踪5帧的跟踪策略,最后更新检测人脸的帧,进行MTCNN模型更新,防止跟踪丢失。

1.2　卷积神经网络

卷积神经网络本质是一个多层感知机[17],包含众多神经元,由输入层、隐含层和输出层组成,输入层是将每个像素代表一个特征节点输入进来,隐含层的卷积层和池化层是对图像进行特征提取的核心,在图像的卷积操作中,每个神经元内部把前一层输入的图像矩阵与多个大小不同的卷积核进行卷积求和,后跟一个加性偏置。将加性偏置和乘性偏置作为激活函数的参数求解,经过线性整流函数(Rectified Linear Unit,

ReLU)激活函数后输出新值,从而构成新的特征图像。卷积层每个神经元的输出为:

$$Y_j^L = f\left(\sum_{i=1}^{N^{L-1}} Y_i^{L-1} \otimes w_{ij}^L + b_j^L\right) \tag{1}$$

其中:L和$L-1$表示为网络的层的深度;$f()$表示为激活函数;“⊗”表示卷积操作;Y_j^L表示为L层第j个输出的特征图像;Y_i^{l-1}表示$L-1$层输出的特征图像;w_{ij}^L和b_j^L分别表示L层的乘性偏置和加性偏置。

为了跟本文设计深度可分卷积神经网络作对比,构建和使用Bergstra等[18]提出的一个标准的全连接卷积神经网络,网络由9个卷积层、线性整流函数ReLU、批量标准化和最大池化层组成。该模型包含大约600 000个参数。在FER-2013数据集中验证了此模型,实现了66%的表情识别准确度。

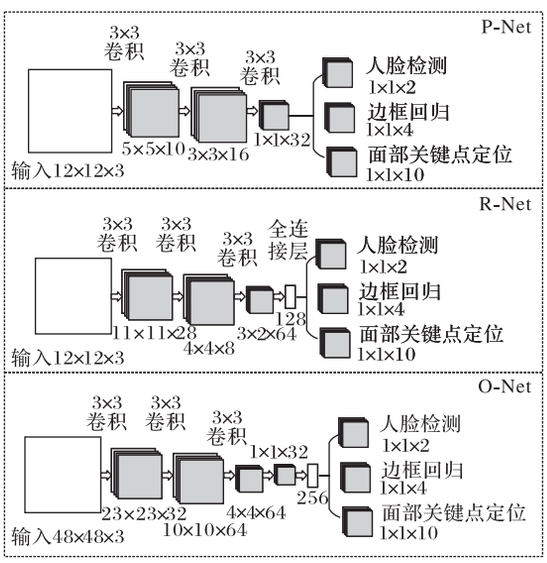


图4 MTCNN网络结构

Fig. 4 Network structure of MTCNN

1.3 深度可分卷积神经网络

由图5可知,该卷积神经网络主要由6个卷积层和3个最大池化层构成,每一个卷积层进行卷积操作后进行一个same填充,当卷积核移动步长为1时,图像尺寸不变,同时为了固定网络层中输入的均值和方差并避免梯度消失问题,将每层神经网络任意神经元的输入值的分布拉回到均值为0、方差为1的比较标准的正态分布,使用批规范化方法,在每一层加上一个批规范化(Batch Normalization, BN)操作,并用ReLU函数激活,后面连接3个全连接层和1个输出层的Softmax函数,在全连接层之后使用一个Dropout的方法,在训练中随机丢弃神经元防止过度训练。本文设计的卷积神经网络结构如图5所示,其中c为卷积核的大小,n为卷积核的数量,s为卷积步长,p为池化窗口的大小,same表示使用same的填充方式,ReLU为激活函数,Sep-Conv为深度可分卷积。

该网络结构由以下部分组成:

1)经过预处理之后得到的64×64像素的学生头部图片作为输入层。

2)c1层使用64个大小为11×11的卷积核对图像进行卷积操作,即每个神经元具有一个11×11的感受野,步长为4,使用same的填充方式,激励函数为ReLU。

3)s1层采用了128个3×3大小的池化窗口对图像进行降维,池化方式为最大池化,步长为2。

4)c2层采用了192个大小为5×5的卷积核,步长为1。

5)s2层采用了192个大小为3×3的池化窗口,池化方式为最大池化,步长为2。

6)c3层使用256个3×3的卷积核,步长为1。

7)c4使用了256个大小为3×3的卷积核,步长为1。

8)c5使用256个大小为3×3的卷积核,步长为1。

9)c6使用深度可分离卷积块。

10)s3采用大小为3×3的池化窗口进行池化,池化方式为最大池化,步长为2。

11)使用4 096个神经元对256个6×6的特征图进行全连接,再进行一个dropout随机从4 096个节点中丢掉一些节点信息,得到新的4 096个神经元。

该网络包含4个剩余深度可分离卷积,其中每个卷积后面是批量归一化操作和ReLU激活函数。最后一层应用Softmax函数产生预测。图5显示了完整的最终网络架构,将其称为迷你Xception。该架构在性别分类任务中获得95%的准确度。此外,在FER-2013数据集中情感分类任务中获得了73.8%的准确度。最终模型的权重可以存储在855 KB的文件中。通过降低模型的计算成本使其具有实时性,并且能够连接两个模型并在同一图像中使用。

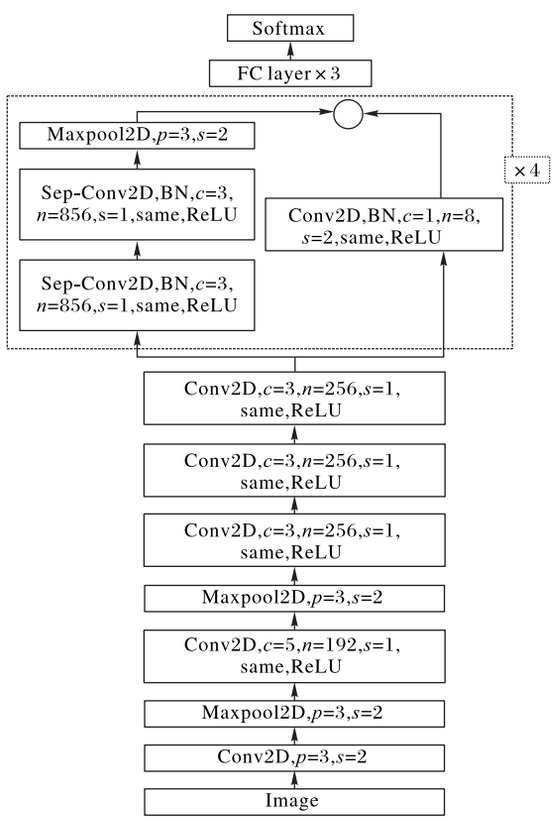


图5 深度可分卷积神经网络的结构

Fig. 5 Deepwise separable convolution neural network structure

1.4 深度可分离卷积单元

本文模型受到Xception[19]架构的启发,结合了残差模块[20]和深度可分离卷积[21]的使用。残差模块修改两个后续图层之间所需的映射,以便学习的特征成为原始特征图和所需特征的差值。通过“捷径链接”的方式,直接将输入的x传输到中间,将该中间结果作为初始结果$H(x)$,为了使网络的参数更容易学习,将网络的学习目标从完整残差块的输出$F(x)$改成新的目标值$H(x)$和x的差值。因此,后层网络训练的目标是将输出结果逼近于0,使随着网络加深,预测准确率

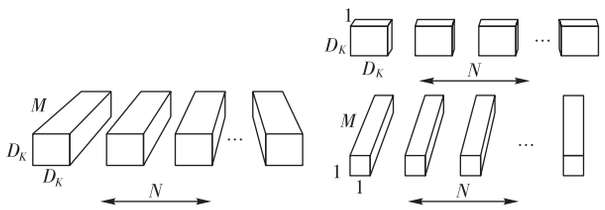
不下降,修改的期望函数$H(x)$见式(2):

$$H(x) = F(x) + x \tag{2}$$

深度可分离卷积由两个不同的层组成:深度方向卷积和点方向卷积。将传统的卷积分为两步:第一步,在每个M输入通道上应用一个$D \times D$滤波器,然后应用N个$1 \times 1 \times M$卷积滤波器将M个输入通道组合成N个输出通道;第二步,应用$1\times1\times N$卷积将特征图中的每个值结合起来。Xception结构增加了每一层网络的宽度和深度,同时也大大减少了网络的参数。深度可分卷积将标准卷积的计算量减少至$1/N + 1/D^2$。

当输入一个2维的数据,对于一个卷积核大小为3×3的卷积过程,正常卷积的参数量为$2 \times 3 \times 3 \times 3 = 54$,深度可分卷积的参数量为$2 \times 3 \times 3 + 2 \times 1 \times 1 \times 3 = 24$,可以看到,参数量为正常卷积的一半。加入该架构后模型大约有60 000参数,是原始CNN的1/80。

正常卷积层和深度可分离卷积之间的差异如图6所示。



(a) 正常卷积　　(b) 深度可分离卷积

图6　不同卷积之间的差异

Fig. 6　Difference between different convolutions

2　网络的训练

2.1　数据预处理

本文在训练数据集之前,先对数据集进行预处理。即,将图像数据归一化到64 × 64像素的图像;接着把归一化后的图像通过平移、翻转、灰度等方法进行数据扩充,在训练过程中以避免过拟合并提升泛化能力。另外,亦使用Dropout方法来避免过拟合。

2.2　引导反向传播可视化

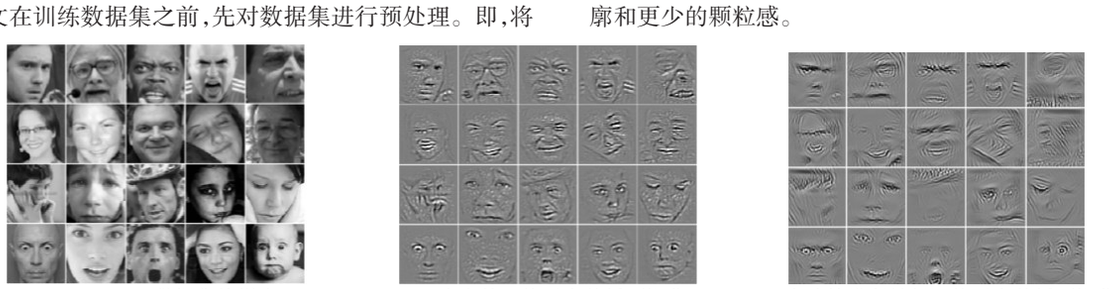
卷积神经网络模型会因为训练数据的偏向性出现偏差,在数据集FER-2013中,主要针对表情分类训练的模型偏向于西方人的面部特征。此外,佩戴眼镜也可能干扰所学习的特征,从而影响表情分类。那么当模型出现偏差时,使用实时引导的可视化技术(如引导反向传播)就变得很重要。以观察图像中的哪些像素激活更高级别特征图的元素。对于只将ReLU作为中间层的激活函数的卷积神经网络,引导反向传播是输入图像中的元素(x,y)对卷积神经网络中位于L层的特征图f^L中元素(i,j)的求导过程。当输入图像到某一层时,设置这层中想要可视化的神经元梯度为1,其他神经元的梯度设置为0,然后经过对池化层、ReLU层、卷积层的反向传播操作,得到输入空间的一张图像。因为ReLU函数的导数为:

$$f'(x) = \begin{cases} 1, & x \geqslant 0 \\ 0, & x < 0 \end{cases} \tag{3}$$

所以引导反向传播后重构的图像R滤除了所有负梯度的值。因此,选择剩余的梯度,使得它们仅增加特征图的所选元素的值。层L中的ReLU激活的CNN重建图像由式(4)给出:

$$R_{i,j}^{l} = \left(R_{i,j}^{l+1} > 0\right) * R_{i,j}^{l+1} \tag{4}$$

在FER-2013数据集中分别提取本文的网络和全连接卷积神经网络最终卷积层中的高维特征进行显示,结果如图7所示。通过对比两者高维可视化的特征显示,本文提出的具有Xception结构的网络学习到的人脸特征具有更加清晰的轮廓和更少的颗粒感。



(a) FER-2013数据集的样本　　(b) 本文的网络反向传播可视化　　(c) 全连接卷积神经网络反向传播可视化

图7　两种卷积神经网络在FER-2013数据集上的可视化效果比较

Fig. 7　Visualization comparison between two convolutional neural networks on FER-2013 dataset

3　实验与结果分析

3.1　数据集

人脸表情分类实验在FER-2013数据库、CK+数据集上进行训练和测试,性别分类实验在IMDB数据库上进行训练和测试。

FER-2013数据集包含35 887张像素为48×48的灰度图,它已被挑战赛举办方分为了三部分:训练集28 709张、公共测试集3 589张和私有测试集3 589张。其中包含有7种表情:愤怒、厌恶、恐惧、开心、难过、惊讶和中性。CK+面部表情数据集由123个个体和593个图像序列组成,每个图像序列的最后一个图像序列都有动作单元标签,327个图像序列都有表情标签,被标记为7种表情标签:愤怒、蔑视、厌恶、恐惧、喜悦、悲伤和惊讶。IMDB性别数据集包含460 723个RGB图像,其中每个图像被标注属于“女性”或“男性”类。

3.2　参数的训练

本文利用上述数据集在深度可分卷积神经网络上进行训练,神经元一开始是随机而独特的,因此它们计算不同的更新,并将自己整合到网络的不同部分。将参数按高斯分布或者均匀分布初始化成一个绝对值较小的数[20]。绝对值过小,容易产生梯度消失问题;绝对值过大,则容易产生梯度爆炸问题。在使用正态分布初始化参数时,参数量n越大,方差越大,越可能产生训练速度慢或梯度消失问题。所以可以通过权重矩阵算法来降低初始化参数方差,进而提高训练速度,预防梯度消失[21]。

$$W = 0.01*\text{np.random.randn}(D,H) \tag{5}$$

其中:randn样本为单位标准高斯分布,均值为0。通过式(5),将每个神经元的权向量初始化为多维高斯分布中采样的随机向量,使得神经元在输入空间中指向随机方向。在训练过程中随机初始化权重和偏置,批量大小设置为120,初始学习率设置为0.01,本文使用了适应性矩估计(Adaptive moment

estimation, Adam)算法来最小化损失函数,实现学习率的自适应调整,从而保证准确率的同时加快收敛。通过对卷积神经网络权重和偏置的调整,并且使用了训练自动停止策略,当模型的在验证集和训练集上的预测能力提升,而在训练集的误差值先减小再增大,这时出现过拟合现象,训练停止。图8分别给出了FER-2013和CK+数据集训练过程中识别率的变化情况。由图8可以看出迭代至10^5次后,训练的准确率达到很高的位置且基本保持稳定,说明最后的模型已经得到充分收敛,训练停止保存模型。

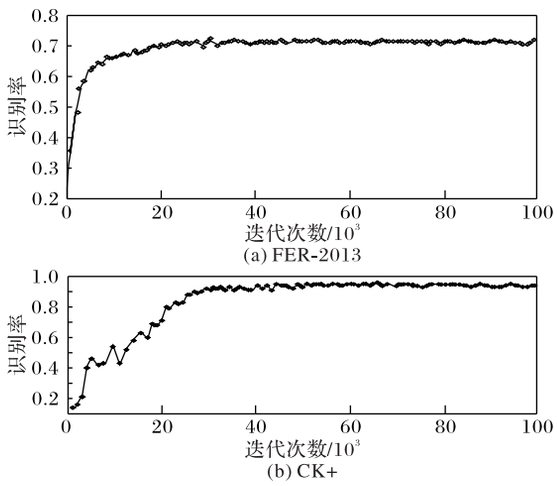


图8 两种数据集上识别率变化

Fig. 8 Change of recognition rate on two datasets

3.3 表情分类实验

人脸表情和性别识别框架中,首先加载已训练好的表情和性别分类模型以及相关配置文件,而针对待检测人脸图像,抓一帧图,找到表情和性别坐标信息,将其像素大小调整为64×64。然后,人脸图像经网络模型向前计算,与训练好的模型中的权重进行比较,得到预测的每一个情感和性别分类标签的得分值,最大值即预测结果。面部表情和性别分析视觉结果如图9所示。

图9 面部表情识别结果示例

Fig. 9 Facial expression recognition result example

实验结果为3次测验的平均值。为了比较方便,在该测试集中种表情的识别准确度结果按照混淆矩阵图表示,如表1所示。

表1 FER-2013数据集上的情感识别混淆矩阵

Tab. 1 Confusion matrix of expression recognition on FER-2013 dataset

表情类别	愤怒	厌恶	恐惧	快乐	悲伤	惊讶	中性
愤怒	**0.65**	0.03	0.20	0.03	0.11	0.02	0.06
厌恶	0.05	**0.75**	0.04	0.04	0.05	0.02	0.03
恐惧	0.02	0.01	**0.67**	0.03	0.12	0.01	0.14
快乐	0.02	0.00	0.02	**0.87**	0.02	0.02	0.05
悲伤	0.01	0.01	0.02	0.06	**0.70**	0.01	0.09
惊讶	0.03	0.00	0.10	0.05	0.02	**0.77**	0.03
中性	0.07	0.00	0.06	0.06	0.04	0.01	**0.76**

从表1可知,本文方法对快乐表情识别率为87%,主要是因为网络在特征提取时,快乐表情的面部特征较其他表情更加明显,在Softmax函数分类的过程中产生概率也越大。惊讶和中性表情识别率分别为77%和76%;而对愤怒和恐惧的表情识别率较低分别为65%和67%,容易出现错误的识别,如图10所示。其原因是在面部特征提取和学习的过程中,两种表情的面部动作幅度都比较大,可能产生相似的面部特征,在Softmax函数分类时产生大小接近的概率值。

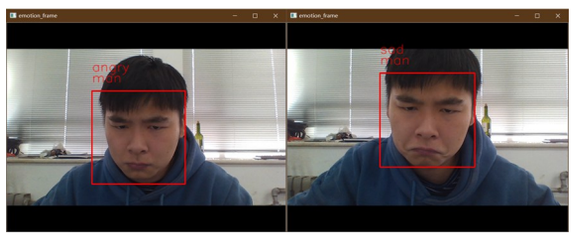


图10 易错误识别的表情对比

Fig. 10 Comparison between easily misidentified expressions

在CK+数据集上的实验采用了迁移学习的方法,将模型在FER-2013上训练得到的权重参数作为预训练结果,然后在CK+上进行微调,并采用三折交叉验证对模型性能进行评估。本文方法在CK+数据集上取得了96%的平均识别率,情感识别结果见表2。

表2 CK+数据集上的情感识别混淆矩阵

Tab. 2 Confusion matrix of emotion recognition on CK+ dataset

表情类别	愤怒	中性	厌恶	恐惧	快乐	悲伤	惊讶
愤怒	**0.99**	0.00	0.00	0.00	0.00	0.00	0.00
中性	0.00	**0.95**	0.00	0.00	0.00	0.50	0.00
厌恶	0.00	0.00	**0.98**	0.00	0.00	0.00	0.00
恐惧	0.00	0.00	0.00	**0.92**	2.00	0.00	6.00
快乐	0.00	0.00	0.00	0.00	**0.97**	0.00	0.00
悲伤	0.8	0.00	0.00	0.00	0.00	**0.92**	0.00
惊讶	0.00	0.10	0.00	0.00	0.00	0.00	**0.99**

各方法在FER-2013数据集上的识别率结果如表3所示。

表3 各方法在FER-2013数据集上的识别率对比

Tab. 3 Comparison of recognition rate among different methods on FER-2013 dataset

方法	识别率/%	方法	识别率/%
MobileNetV2[20]	69.9	Guo方法[17]	71.3
Jeon方法[1]	70.7	Kampel方法[14]	72.0
InceptionV4[16]	70.8	徐琳琳方法[15]	65.6
Tang方法[11]	71.2	本文方法	73.8

3.4 时间复杂度实验

实验环境为:64位Windows 10操作系统,CPU为Inter i5 7300HQ,主频2.5 GHz,显卡型号为NVIDIA GTX 1050ti,显存为4 GB,使用基于Tensorflow的深度学习平台。针对整体模型的实时性进行了测试。实验结果表明,通过引用深度可分卷积的轻量化网络结构,组合OpenCV人脸检测模块,表情分类模块和性别分类模型处理单帧人脸图像的时间为(0.22±0.05)ms,整体处理速度达到80 frame/s;与文献[11]所提架构的处理速度0.33 ms/frame相比,相当于1.5倍的加速,能够确保实时识别效果。

4　结语

针对卷积神经网络训练过程复杂、耗时过长、实时性差等问题,本文提出了一种基于深度可分卷积神经网络的实时表情识别和性别识别方法。利用MTCNN加上KCF的方法进行人脸的检测、跟踪。通过引入深度可分离卷积轻量化网络结构,减少模型参数数量,将参数数量同全连接CNN相比,仅占其1/80;使用反卷积方法可视化呈现了CNN模型中学习到的高级特征。最后,模型在FER-2013数据集上对人脸表情的识别达到了73.8%的高识别率,在CK+数据集上微调获得96%的准确率,在IMDB数据集上取得96%的识别率。处理单帧人脸图像的时间为(0.22±0.05)ms,整体处理速度达到80 frame/s。实验结果表明,本文模型可以堆叠用于多类分类,同时保持实时预测;可在单个集成模块中执行面部检测,进行性别分类和情感分类。后续工作将增加情感识别类型,扩充表情数据库,在真实场景下的数据集上进行训练,进一步提高识别准确率。

参考文献(References)

[1] JEON J, PARK J C, JO Y, et al. A real-time facial expression recognizer using deep neural network [C]// Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication. New York: ACM, 2016: No. 94.

[2] 张延良,卢冰,洪晓鹏,等. 基于局部区域方法的微表情识别[J]. 计算机应用, 2019, 39(5): 1282-1287. (ZHANG Y L, LU B, HONG X P, et al. Micro-expression recognition based on local region method [J]. Journal of Computer Applications, 2019, 39(5): 1282-1287.)

[3] 罗珍珍,陈靓影,刘乐元,等. 基于条件随机森林的非约束环境自然笑脸检测[J]. 自动化学报, 2018, 44(4): 696-706. (LUO Z Z, CHEN J Y, LIU L Y, et al. Conditional random forests for spontaneous smile detection in unconstrained environment[J]. Acta Automatica Sinica, 2018, 44(4): 696-706.)

[4] 戴逸翔,王雪,戴鹏,等. 面向可穿戴多模生物信息传感网络的栈式自编码器优化情绪识别[J]. 计算机学报, 2017, 40(8): 1750-1763. (DAI Y X, WANG X, DAI P, et al. Stacked auto-encoder optimized emotion recognition in multimodal wearable biosensor network [J]. Chinese Journal of Computers, 2017, 40(8): 1750-1763.)

[5] LUO Y, ZHANG T, ZHANG Y. A novel fusion method of PCA and LDP for facial expression feature extraction [J]. Optik — International Journal for Light and Electron Optics, 2016, 127(2): 718-721.

[6] KUMAR P, HAPPY S L, ROUTRAY A. A real-time robust facial expression recognition system using HOG features [C]// Proceedings of the 2016 International Conference on Computing Analytics and Security Trends. Piscataway: IEEE, 2016: 289-293.

[7] 刘帅师,田彦涛,万川. 基于Gabor多方向特征融合与分块直方图的人脸表情识别方法[J]. 自动化学报, 2011, 37(12): 1455-1463. (LIU S S, TIAN Y T, WAN C. Facial expression recognition method based on GABOR multi-orientation features fusion and block histogram[J]. Acta Automatica Sinica, 2011, 37(12): 1455-1463.

[8] KUMAR V D A, KUMAR V D A, MALATHI S, et al. Facial recognition system for suspect identification using a surveillance camera [J]. Pattern Recognition and Image Analysis, 2018, 28(3): 410-420.

[9] TSENG Y T, KAWASHIMA S, KOBAYASHI S, et al. Forecasting the seasonal pollen index by using a hidden Markov model combining meteorological and biological factors [J]. Science of the Total Environment, 2020, 698: No. 134246.

[10] SUN K, KANG H, PARK H H. Tagging and classifying facial images in cloud environments based on *K*NN using MapReduce [J]. Optik — International Journal for Light and Electron Optics, 2015, 126(21): 3227-3233.

[11] TANG Y. Deep learning using linear support vector machines[EB/OL]. [2019-04-10]. http://deeplearning.net/wp-content/uploads/2013/03/dlsvm.pdf.

[12] SANDLER M, HOWARD A, ZHU M, et al. Inverted Residuals and linear bottlenecks: mobile networks for classification, detection and segmentation[EB/OL]. [2019-06-22]. https://arxiv.org/pdf/1801.04381v1.pdf.

[13] LI S, DENG W, DU J P. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2584-2593.

[14] PRAMERDORFER C, KAMPEL M. Facial expression recognition using convolutional neural networks: state of the art [EB/OL]. [2019-04-10]. https://arxiv.org/pdf/1612.02903.pdf.

[15] 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(2): 227-236. (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(2): 227-236.

[16] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, Inception-ResNet and the impact of residual connections on learning [C]// Proceedings of the 31st AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI, 2017: 4278-4284.

[17] GUO Y, TAO D, YU J, et al. Deep neural networks with relativity learning for facial expression recognition [C]// Proceedings of the 2016 IEEE International Conference on Multimedia and Expo Workshops. Piscataway: IEEE, 2016: 1-6.

[18] BERGSTRA J, COX D D. Hyperparameter optimization and boosting for classifying facial expressions: How good can a "Null" model be? [EB/OL]. [2019-04-10]. https://arxiv.xilesou.top/pdf/1306.3476.pdf.

[19] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 1800-1807.

[20] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2019-04-10]. https://arxiv.org/pdf/1704.04861.pdf.

[21] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.

This work is partially supported by the Key Science and Technology Project of Henan Province (192102210290), the Basic Research Plan of Key Scientific Research Project of Colleges and Universities of Henan Province (18A510014).

LIU Shangwang, born in 1973, Ph. D., associate professor. His research interests include biological image processing, computer vision.

LIU Chengwei, born in 1996, M. S. candidate. His research interests include computer vision, deep learning.

ZHANG Aili, born in 1966, professor. Her research interests include signal processing, communication and network.