

doi: 10.3969/j.issn.1001-893X.2020.09.002

引用格式: 石翠萍, 谭聪, 左江, 等. 基于改进 AlexNet 卷积神经网络的人脸表情识别[J]. 电讯技术, 2020, 60(9): 1005-1012. [SHI Cuiping, TAN Cong, ZUO Jiang, et al. Expression recognition based on improved AlexNet convolutional neural network[J]. Telecommunication Engineering, 2020, 60(9): 1005-1012.]

基于改进 AlexNet 卷积神经网络的人脸表情识别^{*}

石翠萍^{**}, 谭 聪, 左 江, 赵可新

(齐齐哈尔大学 通信与电子工程学院 黑龙江 齐齐哈尔 161000)

摘 要: 为了解决传统卷积神经网络用于人脸表情识别准确率不高的问题, 提出了一种基于改进深度 AlexNet 卷积神经网络的表情识别方法。该方法基于 AlexNet 网络的基本结构, 采用单图形处理单元(Graphics Processing Unit, GPU) 进行训练, 减少了两层卷积层和一层全连接层, 在每层卷积层后加上批标准化(Batch Normalization, BN) 代替原来的局部归一化, 并在全连接层后加上 Dropout 正则化进一步防止过拟合。与 AlexNet 模型相比, 改进的网络结构更简单、复杂度低、参数量少, 可以节省大量模型训练时间进行快速预测, 且更不易过拟合, 同时加快了模型收敛速度, 提高了网络泛化能力。在 Fer2013 数据集以及 CK+ 数据集上进行实验, 结果表明, 所提方法分别得到了 68.85% 和 97.46% 的识别率, 较其他人脸表情识别方法的识别率有一定提高。

关键词: 表情识别; 深度学习; AlexNet 网络; BN 算法

开放科学(资源服务) 标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号: TN919.8; TP183 文献标志码: A 文章编号: 1001-893X(2020)09-1005-08

Expression Recognition Based on Improved AlexNet Convolutional Neural Network

SHI Cuiping, TAN Cong, ZUO Jiang, ZHAO Kexin

(College of Communication and Electronic Engineering, Qiqihar University, Qiqihar 161006, China)

Abstract: In order to solve the problem of low accuracy of traditional convolutional neural network(CNN) for facial expression recognition, an expression recognition method based on improved deep AlexNet CNN is proposed. Based on the basic structure of AlexNet network, this method uses single graphics processing unit (GPU) to train, thus reducing two layers of convolution layer and one layer of full connection layer, adding batch normalization(BN) after each layer of convolution layer to replace the original local normalization, and adding dropout regularization after the full connection layer to further prevent over fitting. Compared with AlexNet model, the improved network has simpler structure, lower complexity, and fewer parameter, and it can save a lot of model training time for fast predictions and is more difficult to overfit. Meanwhile, it can accelerate the convergence speed of the model and improve the generalization ability of the network. Experiments on Fer2013 data set and CK+ data set show that the recognition rate of this method is 68.85% and 97.46% respectively, which is higher than other facial expression recognition methods.

Key words: facial expression recognition; deep learning; AlexNet network; BN algorithm

^{*} 收稿日期: 2019-12-23; 修回日期: 2020-04-09

基金项目: 国家自然科学基金青年基金项目(41701479); 中国博士后科学基金项目(2017M621246); 黑龙江省科学基金项目(QC2018045); 黑龙江省博士后科学基金项目(LBH-Z17052); 黑龙江省省属高等学校基本科研业务费专项(135309342); 2019 年省级大学生创新创业训练计划资助项目(201910232044)

^{**} 通信作者: scp1980@126.com

0 引言

人脸表情识别作为计算机视觉领域的一个经典研究课题,在人机交互领域也有着重要的应用。传统的表情特征提取方法^[1-2]提取效率低、资源浪费严重,特征提取不完全。因此,研究者们开始利用卷积神经网络(Convolutional Neural Network, CNN)进行表情识别研究。卷积神经网络以其局部权值共享的特殊结构,在语音识别和图像处理方面有着独特的优越性。但随着网络深度的加深,其分布逐渐发生偏移,导致梯度消失进而收敛越来越慢,训练速度慢,准确度下降。

深度卷积神经网络起源于 AlexNet 网络。2012 年, Alex 等人^[3]使用采用卷积神经网络的 AlexNet 在 ImageNet 竞赛图像分类任务中取得了最好成绩,使卷积神经网络在图像分类中取得了巨大成功。然而, AlexNet 模型在表情识别上表现得并不突出^[4-6]。自动表情识别(Face Expression Recognition, FER)一直是计算机视觉领域中一个具有挑战性的问题。文献[7]提出了一种深度神经网络结构,用来解决多个已知标准人脸数据集上的面部表情识别问题。文献[8]提出的方法通过在 Fer2013 上训练、在 RaFD 上测试,取得了 68% 的准确度。文献[9]提出了一种加权混合深度神经网络,用于自动提取对 FER 任务有效的特征,实现了人脸检测、旋转校正、数据增强预处理,在公开的扩展 Cohn-Kanade(CK+) 上达到了 97% 的准确度。文献[10]将面部集合特征和外观特征相结合的高维特征引入到面部表情识别中,建立了深度稀疏自编码(Deep Sparse Autoencoder, DSAE)算法,通过从数据中学习鲁棒的、有区别的特征,实现对面部表情的高精度识别,在 CK+ 数据库上识别率达到 95.79%。文献[11]提出了一种基于多级 Haar 小波的人脸特征提取方法。该方法首先使用 AdaBoost 级联目标检测器分割出信息量最大的几何成分,如眼睛、嘴、眉毛等,并提取分割后部分的 Haar 特征,最后采用 OneVsAll logistic 回归模型进行分类,在 CK+ 数据集上到达了 90.48% 的准确率。文献[12]提出了一种用于人脸表情识别的单帧图像特征提取方法,混合特征使用 SIFT 和从 CNN 模型中提取的不同层次的深度学习特征的组合,采用组合特征对表达式进行支持向量机分类。使用 CK+ 数据集测试了该方法的性能,提出的 SIFT-CNN 混合特征的 FER 准确率为 94.82%。

AlexNet 网络作为深度学习中的一种经典网络,

目前在图像分类任务中仍被大量使用。相比于大型网络(如 VGG)而言, AlexNet 深度卷积神经网络复杂度低、参数量少;相比现有的轻量级网络深度深,训练难度不会很大,表征能力强,可以节省大量模型训练时间,进行快速预测,且更不易过拟合,因此,更适合用于实时人脸识别应用。目前,大多新颖的网络都是基于 AlexNet、VGG、GoogleNet 等经典网络提出的,且结构都比 AlexNet 复杂,故在 AlexNet 的基础上进行改进,更利于进行模型的快速预测和评估。为此,本文提出一种基于 AlexNet 网络的改进模型,以提高对人脸表情识别的能力。所提方法采用单 GPU 进行训练,减少了两层卷积层和一层全连接层,在每层卷积层后加上批标准化(Batch Normalization, BN)代替原来的局部归一化(Local Response Normalization, LRN),并在全连接层后加上 Dropout 正则化进一步防止过拟合。与 AlexNet 模型相比,加快了模型收敛速度,提高了网络泛化能力。在 Fer2013 数据集^[13]和 CK+ 数据集^[14]上利用数据增强等技术进行预处理后,利用改进的 AlexNet 网络在数据集上进行测试。实验结果表明,与常用表情识别方法相比,本文提出方法能够获得更高的人脸表情识别率。

1 经典 AlexNet 网络

AlexNet 的网络结构如图 1 所示。AlexNet 网络用 ReLU 代替 Sigmoid 作为 CNN 的激活函数,并验证了 ReLU 在较深的网络里性能优于 Sigmoid,有效解决了 Sigmoid 引起的梯度弥散的问题。而且 ReLU 比 Sigmoid 的学习速度更快,节省了训练时间。采用多 GPU 训练,使用双 GPU 网络的训练时间比单 GPU 网络的训练时间少。本文提出 LRN 局部响应归一化用在 ReLU 后,不同的内核计算的神经元输出之间产生对大激活度的竞争,使得局部较大的响应值更大,而小的会变得更小,从而抑制了小的神经元,增强了模型的泛化能力。使用重叠的最大池化代替平均池化,避免了平均池化造成的模糊化效果。而且 AlexNet 中的步长比池化核的尺寸要小,池化层的输出间会有重叠,这样使特征能表现更多的内容,提高识别性能。Dropout 即以 0.5 的概率把每个隐藏的神经元的输出设置为零,以这种方式“Dropout”的神经元不参与正向传递,也不参与反向传递,所以每次提交输入时,神经网络都采样不同的体系结构,但是所有这些体系结构共享权重。这种技术减少了神经元的复杂的共同适应,因为神经元

不能依赖于特定的其他神经元的存在。由于 AlexNet 的参数数量巨大, 容易造成过拟合, 通过截取、平移、翻转以及 RGB 像素值集上进行主成分分析

(Principal Component Analysis, PCA) (对于每个训练图像, 成倍增加已有的主成分) 等方法使得数据集更丰富, 从而提高泛化能力。

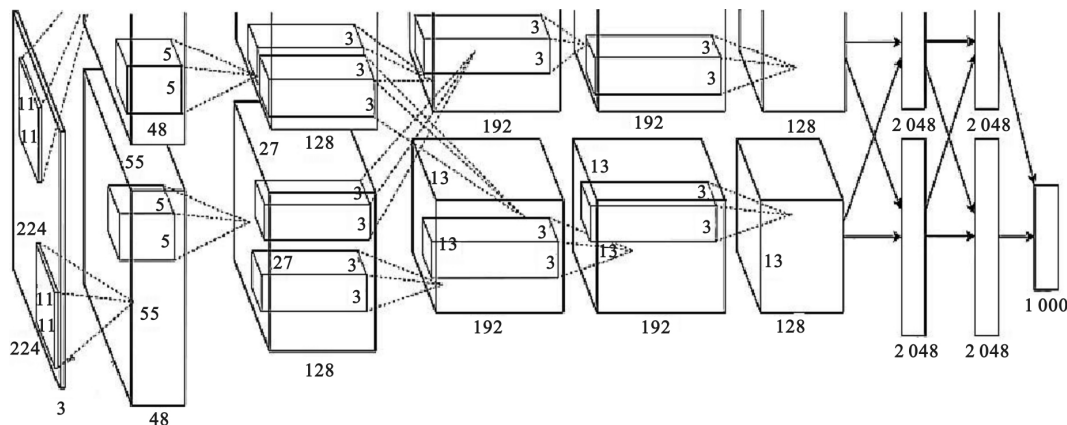


图 1 AlexNet 的网络结构

2 BN 算法

批量归一化^[15]和激活函数层、卷积层、全连接层、池化层一样, 也属于网络的一层, 简称 BN, 通常用在全连接层和卷积层之后、激活函数层之前。在网络的每一层输入 x 的时候, 又插入了一个归一化层, 也就是先做一个归一化处理(归一化至均值 0、方差为 1), 然后进入网络的下一层。它是一个可学习、有参数 (γ, β) 的网络层。设输入 x 是 d 维, 则

$$x = (x^{(1)} \ x^{(2)} \ \cdots \ x^{(d)})。 \quad (1)$$

对每一维进行归一化, 则

$$\hat{x}^{(k)} = \frac{x^{(k)} - E[x^{(k)}]}{\sqrt{\text{Var}[x^{(k)}]}}。 \quad (2)$$

如果是仅仅使用上面的归一化公式, 对网络某一层的输出数据做归一化, 然后送入下一层, 这样会影响到本层网络所学习到的特征。于是进行变换重构, 引入可学习参数 γ, β , 这是算法的关键之处, 可表示为

$$y^{(k)} = \gamma^{(k)} \hat{x}^{(k)} + \beta^{(k)}, \quad (3)$$

$$\gamma^{(k)} = \sqrt{\text{Var}[x^{(k)}]}, \quad (4)$$

$$\beta^{(k)} = E[x^{(k)}]。 \quad (5)$$

在实际测试时, 依然采用下面的公式:

$$\hat{x} = \frac{x - E[x]}{\sqrt{\text{Var}[x] + \epsilon}}。 \quad (6)$$

这里的均值和方差不再是针对某一个批量, 而是面向整个数据集。因此, 在训练过程中除了正常的前向传播和方向求导之外, 还要记录每一个批量的均值和方差, 以便训练完成后计算整体的均值和

方差, 即

$$E[x] \leftarrow E_B[\mu_B], \quad (7)$$

$$\text{Var}[x] \leftarrow \frac{m}{m-1} E_B[\sigma_B^2]。 \quad (8)$$

式中: x 表示一层 d 维的输入数据; B 表示一个尺寸为 m 的小批量, $B = \{x_1 \cdots x_m\}$; y 为 x 的线性变换; ϵ 表示一个常数; μ_B 表示小批量平均值 $\mu_B = \frac{1}{m} \sum_{i=1}^m x_i$;

σ_B^2 表示小批量差异 $\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2$ 。

3 改进方法

本文改进的网络结构中分别使用 5×5 、 3×3 的卷积核 32 个和 64 个, 保留 Relu 作为激活函数、最大池化层来缩减模型大小, 提高计算速度, 同时提高所提取特征的鲁棒性。此外, 本文采用 BN 层代替原来的 LRN 层, 放在卷积层之后。LRN 使得局部较大的响应值更大, 而小的会变得更小, 从而抑制了小的神经元, 增强了模型的泛化能力。BN 层在该基础上, 还加快了模型收敛, 打乱了训练数据, 大大减少了训练的时间。使用最大池化对图像进行下采样, 加快了模型的训练速度, 提高了提取特征的鲁棒性。接下来, 利用展平层对输入特征进行压平, 并与 2 048 个和 1 024 个神经元进行全连接, 在每层全连接后加 Dropout 正则化防止过拟合(参数设置为 0.4), 最后用 Softmax 进行分类。改进后的网络结构如图 2 所示。

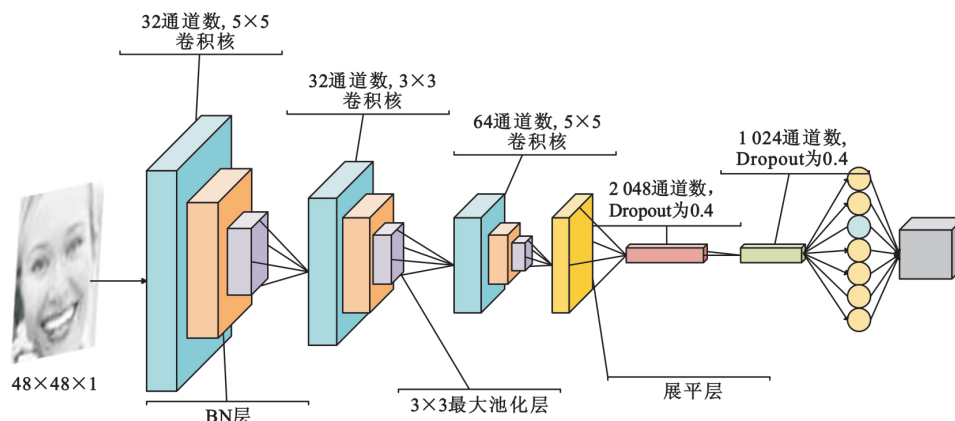


图2 改进后的网络模型

卷积神经网络主要由输入层、卷积层、Relu层、池化层和全连接层叠加起来构成。本文模型结构采用了3层卷积、3层BN归一化层、3层最大池化层和2层全连接层构成。卷积层都采用0填充,防止在卷积操作时图像角落的像素被遗漏,导致图像特征提取不完整,进而影响模型性能。采用Dropout正则化,对所有的神经元按照40%的概率进行消除处理,在很大程度上简化了神经网络结构,防止神经网络过拟合,从而提高模型的分类能力。

Dropout正则化(随机失活)可以起到防止过拟合现象的作用。根据Dropout正则化思想,对于网络的每一层,随机地丢弃一些单元,随机丢弃单元的概率可以自己根据实际情况进行设置,概率范围为0~1。本文进行了大量的实验,分别从20%、40%、60%的失活概率进行实验,结果表明,40%的随机失活概率效果最佳,故将Dropout正则化参数设为40%。本文Dropout正则化的随机失活概率设置为40%,可以起到简化神经网络结构的作用,从而起到正则化的效果。

原始图像通过卷积神经网络的输入层进入网络,以 48×48 的形状大小作为输入,利用卷积神经网络的局部感知和参数共享两个特点,通过对图像的局部像素进行感知,最后将这些局部感知的像素合并在一起,保证了学习到的卷积核对输入的局部像素具有最强的响应,学习到更多有用的特征。权值参数的共享降低了网络模型的复杂度,减少了权值的数量,并不影响网络的分类性能,有助于提高表情识别率;改进的网络在卷积层后加上BN层,BN层在防止过拟合、梯度消失的同时还提升了训练的稳定性,加快了模型的收敛速度,提高了网络泛化能力,打乱了样本的训练顺序(这样就可以避免一张图像被多次训练),进而提高了表情识别率;原始图

像输入网络后,经过3次下采样,图像的形状大小变为 5×5 ,缩减了模型的大小;原始图像经过数据增强后,训练集变大,网络学习过程中可以有效地从大量样本中学习到相应的特征,避免了复杂的特征提取过程,使算法的鲁棒性更强,从而进一步提高了表情识别率。综上所述,本文改进的网络可以有效提取图像特征,表情分类性能良好。

AlexNet网络有5层卷积和3层全连接,全连接层的参数很多,占整个网络的80%左右。本文减少一层全连接层,同时减少全连接层的通道数,进一步减少了网络的参数量,降低了网络的复杂度。本文在网络改进过程中主要改变卷积层的层数,其他参数设置相同。表1给出了三种改进方法的精度对比。由对比结果可知,保留5层卷积或减少1层卷积的分类精度,均比减少2层卷积的分类精度低。因此,本文采用减少2层卷积层和1层全连接层作为本文最终的改进模型,并用于人脸表情识别。

表1 Fer2013和CK+数据集改进过程分类精度对比表

改进方法	精度/%	
	Fer2013 数据集	CK+数据集
保留5层卷积	67.30	96.45
减少1层卷积	67.20	96.95
减少2层卷积	68.85	97.46

4 实验结果与分析

本节首先对本文所用的数据库进行介绍,并进行数据预处理;其次,利用本文提出的改进模型在表情数据集上进行训练;最后,利用训练得到的模型在数据集上测试,采用多种评价指标进行衡量,并与多种较新的表情识别方法进行比较,以证明提出方法的有效性。本文实验都是在JetBrains PyCharm2017.

1 x64和Kera2.1.4上实现的,采用GeForce 940MX

显卡。

4.1 数据库处理

4.1.1 Fer2013 人脸表情数据集

Fer2013 人脸表情数据集是人脸表情识别研究最常用的数据集, Fer2013 数据集本身就已划分了训练集、验证集和测试集, 且其他研究者都是在 Fer2013 数据集上进行测试的。选用该数据集, 有利于在相同条件下将本文方法与其他相关方法进行比较。

该数据集共包含 35 887 张人脸图片, 其中训练集 28 709 张, 验证集 3 589 张, 测试集 3 589 张。数据集中的图片均为灰度图片, 大小为 48 pixel×48 pixel。样本被分为 0=anger(生气)、1=disgust(厌恶)、2=fear(恐惧)、3=happy(开心)、4=sad(伤心)、5=surprised(惊讶)、6=normal(中性) 7 类。该数据集的测试集存在着许多标签的错误, 导致在该数据集上测试精度不是很高, 且人眼在该数据集上的识别率为 (65±5)%。图 3 给出了 Fer2103 人脸表情数据集对应的 7 种表情。



图 3 Fer2013、CK+数据集可视化

Fer2013 数据集的训练集、验证集和测试集的每个类别的数量分布如表 2 所示。

表 2 Fer2013 数据集的类别数量分布

类别	训练集	验证集	测试集
生气	3 995	467	491
厌恶	436	56	55
恐惧	4 097	496	528
高兴	7 215	895	879
伤心	4 830	653	594
惊讶	3 171	415	416
正常	4 965	607	626

4.1.2 CK+人脸表情数据集

CK+数据集是在 Cohn-Kanade 数据集的基础上扩展来的, 发布于 2010 年。该数据集的图像可视化如图 3 所示。本文将该数据集的 20% 划分为测试集用于测试模型, 80% 用于训练模型。表 3 给出了 CK+

数据集的训练集和测试集的每个类别数量分布。

表 3 CK+数据集的类别数量分布

类别	训练集	测试集
生气	108	27
厌恶	142	35
恐惧	60	15
高兴	166	41
伤心	67	17
惊讶	199	50
正常	43	11

4.2 数据增强

一般而言, 比较成功的神经网络需要大量的参数, 很多的神经网络的参数都是数以百万计, 而使这些参数正常工作需要大量的数据进行训练。实际情况中, 并没有这么多的数据可用于训练。因此, 深度学习中的数据增强就起到了很大的作用。它不仅能增强训练的数据量, 提高模型的泛化能力, 还能增加噪声数据, 提高模型的鲁棒性。本文利用了随机缩放、翻转、平移、旋转数据增强技术对数据进行增强。图 4 和图 5 分别为 Fer2013 数据集和 CK+数据集的数据增强后的数据样本分布情况。

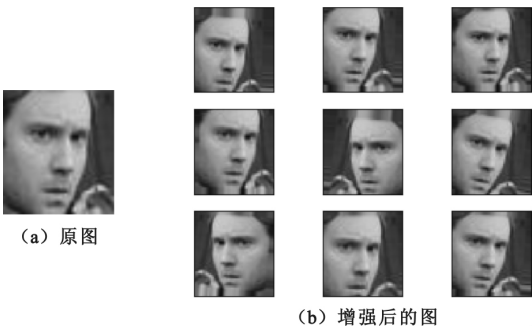


图 4 Fer2013 数据增强

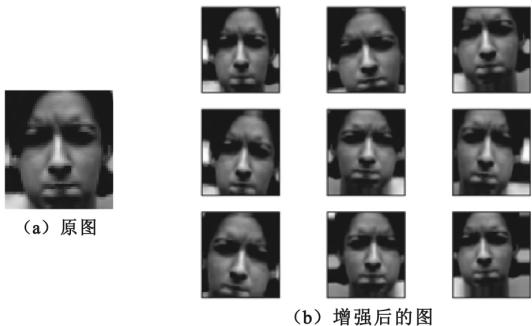


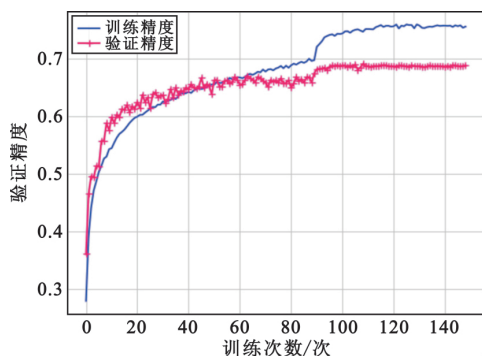
图 5 CK+数据增强

4.3 测试结果分析

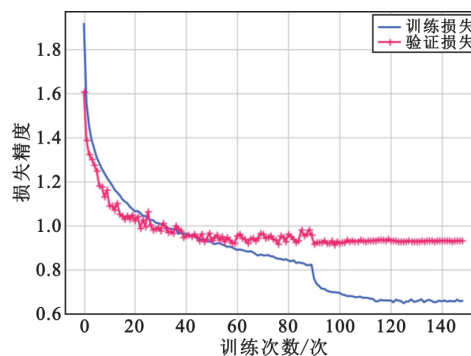
图 6 为本文利用改进的网络在 Fer2013 数据集和 CK+数据集上训练得到的训练验证精度曲线图

和损失曲线图。由图 6 可知,本文改进的网络具有良好的学习能力,随着训练次数的增加验证精度和验证损失都随着训练精度和训练损失值变化,整个

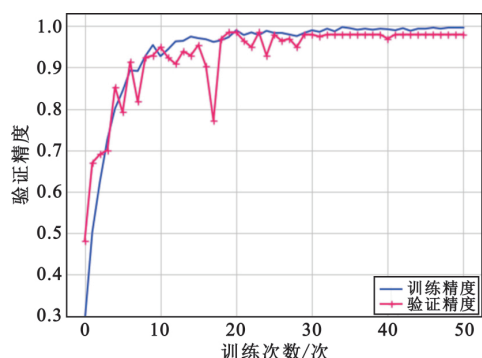
训练过程中并未出现欠拟合和过拟合现象,且在两个数据集上表现较好,表明了本文改进网络具有良好的泛化能力和学习能力。



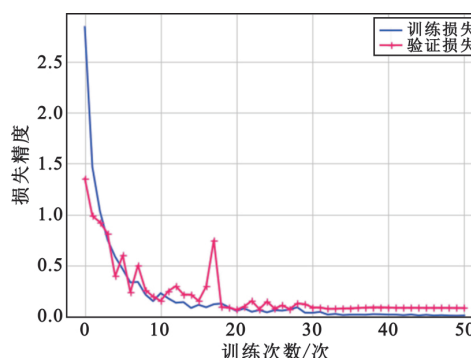
(a) Fer2013 数据集训练验证精度曲线图



(b) Fer2013 数据集训练验证损失曲线图



(c) CK+数据集训练验证精度曲线图



(d) CK+数据集训练验证损失曲线图

图 6 数据集训练过程曲线图

精确率、召回率、 F_1 值和准确率的计算公式如下:

$$P = \frac{TP}{TP+FP}, \quad (9)$$

$$r = \frac{TP}{TP+FN}, \quad (10)$$

$$F_1 = \frac{2rP}{2TP+FP+FN}, \quad (11)$$

$$accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}. \quad (12)$$

式中: TP 表示将实际正样本数预测为正样本的个数, FN 表示将实际正样本数预测为负样本的个数, FP 表示将实际负样本预测为正样本的个数, TN 表示将实际负样本数预测为负样本的个数。

在 Fer2013 和 CK+数据集上,用本文提出的模型进行实验,并进行性能分析,结果分别如表 4 和表 5 所示,表中的准确率表示每个类别预测正确的准确率;整体准确率表示七个类别的预测正确的平均

准确率;权重平均值表示各数值乘以相应的权重,然后加总求和,再除以总的单位数;数量表示每一个类别预测的数量。需要说明的是,由于 Fer2013 数据集存在着部分标签错误,导致在该数据集上进行测试,通常得到的分类精度不高。尽管如此,作为一个较大的人脸表情数据集,该数据集在表情识别领域仍然得到了广泛应用。由表 4 可见,厌恶类、恐惧类、正常类的精确率和召回率相差较大,整体准确率只有 0.671 173 75,这与 Fer2013 数据集存在标签错误和数据集里的样本错误有较大关系。如图 7 所示, Fer2013 数据集中每个类别中存在许多负样本,导入模型进行预测时会导致正负样本预测出错,致使精确率和召回率相差较大,尽管结果数值不是很高,但能达到本文的效果,说明分类性能较好。由表 5 可以看出,本文方法对数据集中每一类的分类精确率、召回率和 F_1 值都较高,整体准确率达 0.986 771 3。实验结果表明,采用本文提出模型进

行分类,能得到较好的分类性能。

表 4 Fer2013 数据集测试指标

类型	精确率	召回率	F_1 值	准确率	数量
生气	0.60	0.62	0.61	0.623 217 94	491
厌恶	0.90	0.65	0.76	0.654 545 40	55
恐惧	0.59	0.43	0.50	0.433 712 12	528
高兴	0.87	0.87	0.87	0.873 720 10	879
伤心	0.55	0.52	0.53	0.518 518 50	594
惊讶	0.81	0.83	0.82	0.829 326 90	416
正常	0.61	0.77	0.68	0.765 175 70	626

表 5 CK+数据集测试指标

类型	精确率	召回率	F_1 值	准确率	数量
生气	0.96	1.00	0.98	1.000 000 00	25
厌恶	0.94	1.00	0.97	1.000 000 00	44
恐惧	1.00	1.00	1.00	1.000 000 00	10
高兴	1.00	0.97	0.99	0.972 973 00	37
伤心	1.00	1.00	1.00	1.000 000 00	11
惊讶	1.00	0.93	0.97	0.934 426 25	61
正常	0.90	1.00	0.95	1.000 000 00	9

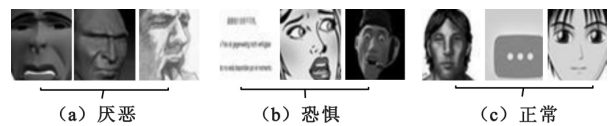


图 7 Fer2013 数据集中的负样本

在 Fer2013 和 CK+数据集上,采用多个其他表情识别方法进行测试,并与本文方法进行比较,得到的结果分别如图 8 和图 9 所示。

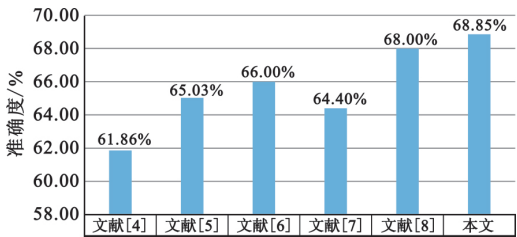


图 8 基于 Fer2013 数据集的识别率比较

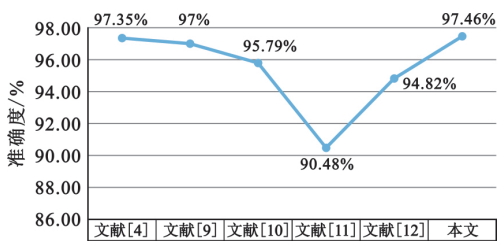


图 9 基于 CK+数据集的识别率比较

由图 8 和图 9 可知,在 Fer2013 数据集和 CK+数据集上进行测试,相比于其他表情识别方法,采用本文模型能得到更好的表情识别结果。文献[4]将手工特征信息嵌入到网络中进行训练,在不损失特征的情况下效果才最佳,但传统的手工特征提取不完全,效率低,使得分类的效果不明显;文献[5]从单个人脸图像出发进行表情识别,特征提取不够全面;文献[7]采用单一的组件构成的网络进行表情识别研究,过于单调;文献[6,8]采用通用的卷积神经网络进行表情识别,得到的分类效果并不明显;文献[11]采用传统的表情识别方法——多级 Haar 小波的人脸特征提取方法,特征提取不完全,资源浪费严重;相比之下,本文采用的深度卷积神经网络具有局部权值共享、自动提取特征和较强的自学习能力等优点,提取特征完全,分类效率高。文献[9]利用 VGG16 进行参数初始化,然后用浅层卷积神经网络提取 LBP 人脸图像特征,最后将两通道的特征进行融合;文献[10,12]采用混合特征进行表情识别,采用 VGG16 进行初始化,复杂度高、计算量大,和传统方法进行特征融合,会把提取的不完整的特征融合到一起,降低识别率;本文采用的卷积神经网络逐层对人脸图像特征进行提取,并在每层卷积层后加上 BN 层,对提取的特征进行全局归一化,相比于上述文献,不但能够防止梯度消失和过拟合,还能够提高分类效率,加快迭代速度。同时,模型的复杂度低、参数量少、结构简单,可以快速用于模型训练和预测,节省了大量训练时间。与上述文献的对比证明了本文方法的有效性。

本文改进方法的实用效果如图 10 所示,可以接受。本文利用改进的网络训练得到的模型,通过加载模型进行实时人脸表情检测,分类出表情类别并将其输出。由实时检测分类效果可知,本文改进网络的分类效果良好。



图 10 改进方法实用效果图

5 结束语

针对表情识别分类,本文结合深度学习和经典分类算法,在经典 AlexNet 卷积神经网络的结构上进行改进:采用单 GPU 进行训练,减少两层卷积层和一层全连接层,在每层卷积层后加上批标准化代替原来的局部归一化;保留原有的 Relu 激活函数,并在全连接层后加上 Dropout 正则化进一步防止过拟合;保留原有的最大池化层以减少特征维度,增强特征的鲁棒性;通过数据增强,克服了训练数据不足、算法的准确率低等问题。本文改进的网络结构简单、复杂度低、参数量少,可以快速、有效地进行模型的训练和预测。实验结果表明,与现有的多种表情识别方法相比,本文提出的改进模型具有更好的性能。

下一步工作将继续提高模型的识别准确率,进一步实现表情识别智能化和实时化。

参考文献:

- [1] 尚晓,吴进.采用 Gabor-Hough 变换的自适应滤波人眼定位[J].电讯技术,2016,56(3):324-330.
- [2] GIRSHICK R, DONAHUE J, DRRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Seattle: IEEE Computer Society, 2014: 580-587.
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E, et al. ImageNet classification with deep convolutional neural networks [J]. Communications of The ACM, 2017, 60(6): 84-90.
- [4] ZENG G, ZHOU J, JIA X, et al. Hand-Crafted feature guided deep learning for facial expression recognition [C]//Proceedings of 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). Xi'an: IEEE, 2018: 423-430.
- [5] LIU K, ZHANG M, PAN Z. Facial expression recognition with CNN ensemble [C]//Proceedings of 2016 International Conference on Cyberworlds (CW). Chongqing: IEEE, 2016: 163-166.
- [6] ARRIAGA O, VALDENEGRO-TORO M, PLÖGER P. Real-time convolutional neural networks for emotion and gender classification [J/OL]. Computer Science, 2017: 1-5 [2019-11-20]. <http://arxiv-org-s.vpn.uestc.edu.cn:8118/pdf/1710.07557.pdf>.
- [7] MOLLAHOSEINI A, CHAN D, MAHOOR M H. Going deeper in facial expression recognition using deep neural networks [C]//Proceedings of 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Placid: IEEE, 2016: 1-10.
- [8] SALUNKE V V, PATIL C G. A new approach for automatic face emotion recognition and classification based on deep networks [C]//Proceedings of 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA). Pune: IEEE, 2017: 1-5.
- [9] YANG B, CAO J, NI R, et al. Facial expression recognition using weighted mixture deep neural network based on double-channel facial images [J]. IEEE Access, 2018, 6: 4630-4640.
- [10] ZENG N Y, ZHANG H, SONG B Y, et al. Facial expression recognition via learning deep sparse autoencoders [J]. Neurocomputing, 2018, 273: 643-649.
- [11] GOYANI M M, PATEL N M. Multi-level HAAR wavelet based facial expression recognition using logistic regression [J]. Indian Journal of Science and Technology, 2017, 10(9): 1-9.
- [12] SUN X, LV M. Facial expression recognition based on a hybrid model combining deep and shallow features [J]. Cognitive Computation, 2019, 11(4): 587-597.
- [13] GOODFELLOW I J, ERHAN D, CARRIER P L, et al. Challenges in representation learning: a report on three machine learning contests [C]//Proceedings of International Conference on Neural Information Processing. Heidelberg: Springer, 2013: 117-124.
- [14] LUCEY P, COHN J F, KANADE T J, 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. San Francisco: IEEE, 2010: 94-101.
- [15] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]//Proceedings of International Conference on International Conference on Machine Learning. Lille: IEEE, 2015: 448-456.

作者简介:



石翠萍 女,1980 年生于山东菏泽,2016 年于哈尔滨工业大学获博士学位,现为齐齐哈尔大学副教授、硕士生导师,主要研究方向为数字图像处理、模式识别与人工智能。

谭 聪 男,1998 年生于重庆忠县,本科生,主要研究方向为数字图像处理、模式识别与人工智能。

左 江 男,1998 年生于四川彭州,本科生,主要研究方向为数字图像处理、模式识别与人工智能。

赵可新 女,1999 年生于黑龙江明水,本科生,主要研究方向为数字图像处理、模式识别与人工智能。