

一种改进的卷积神经网络的表情识别算法

付倩倩¹, 李 昂^{2,3}

- (1. 武汉邮电科学研究院, 湖北 武汉 430074;
2. 南京邮电大学 通信学院, 江苏 南京 210003;
3. 南京理工大学紫金学院, 江苏 南京 210023)

摘 要:随着计算机视觉领域的发展,智能化人机交互技术越来越受人们的重视。作为最直接、最有效的情感识别方式,人脸表情识别现已是人机交互领域研究的一大热点和难点。由于人脸识别容易受到光照、旋转、遮挡等复杂因素的影响,传统的人脸识别方法的准确度会大大减少。为了使机器能够快速准确地感应人脸表情,提出以卷积神经网络(convolutional neural network, CNN)来构建表情识别框架,将传统的人工神经网络和深度学习(deep learning, DL)技术结合起来,利用经典的卷积神经网络模型进行分析。将表情分为愤怒、惊讶、高兴、悲伤、恐惧五大类对不同的性别进行识别与分析。结果表明,与传统的表情识别方法相比,该方法有较好的识别效率和时效性,从而可以大大提高人机交互运用的体验感。

关键词:人脸表情识别;智能化人机交互;深度学习;卷积神经网络

中图分类号: TP391

文献标识码: A

文章编号: 1673-629X(2020)11-0080-04

doi: 10.3969/j.issn.1673-629X.2020.11.015

An Improved Facial Expression Recognition Technology Based on Convolutional Neural Network

FU Qian-qian¹, LI Ang^{2,3}

- (1. Wuhan Institute of Posts and Telecommunications, Wuhan 430074, China;
2. School of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;
3. Nanjing University of Science and Technology Zijin College, Nanjing 210023, China)

Abstract: With the development of computer vision, intelligent human-computer interaction technology has been paid more and more attention. As the most direct and effective emotion recognition method, facial expression recognition has become a hot and difficult issue in the field of human-computer interaction. Because face recognition is easily affected by complex factors such as illumination, rotation and occlusion, the accuracy of traditional face recognition methods will be greatly reduced. In order to make the machine can rapidly and accurately induction facial expressions, we put forward to use convolution neural network (CNN) to build a facial expression recognition framework. Combining traditional artificial neural network and deep learning (DL), we use the classical convolution neural network model to analyze. Expressions are classified into anger, surprise, happiness, sadness and fear to identify and analyze different genders. The results show that this method has better recognition efficiency and timeliness than the traditional expression recognition method, which can greatly improve the experience of human-computer interaction.

Key words: facial expression recognition; intelligent human-computer interaction; deep learning; convolutional neural network

0 引 言

人们要想知道对方的情绪变化,通过表情识别是最直接、最有效的。随着信息时代的发展,人们愈希望“观色”可以利用在计算机视觉领域,如果计算机能够通过识别一个人的表情以观其色,即可为场景中出

现的人物提供辅助的结构化信息,这在人机交互、安防、机器人智能化等多个领域均有广泛的应用。

文献[1-5]在残差网络的基础上设计卷积网络提取不同视角下的表情特征,引入深度可分离卷积来减少网络参数。利用特征重新标定方式提高网络表示能

收稿日期:2019-10-16

修回日期:2020-03-10

基金项目:江苏省高校自然科学基金面上项目(18KJD510004);江苏省普通高校学术学位研究生科研创新计划项目(KYLX160661)

作者简介:付倩倩(1991-),女,硕士,CCF会员(E9620G),研究方向为大数据与云计算。

力,并通过加入空间金字塔池化增强网络的鲁棒性。最后进一步优化识别结果最终实现人脸表情识别。

文献[6-8]采用深度学习算法的人脸表情识别系统运用较高的 CPU 或 GPU 硬件进行模型训练。

1 相关工作

完整的人脸表情识别一般分为以下四步:定位图像、对图像进行预处理、提取图像中表情的特征并对表情进行分类。其中最重要的是提取特征,这一部分又分为生成原先图像的特征和特征降维。图 1 所示为人脸表情识别的系统框架。其中,前两个阶段,定位图像,并且对图像进行预处理,可以减少外界因素对图像的干扰,例如光照不均、动态变化、强弱不定等。这样可以使原图变得更加清晰,质量更高,从而得到图像的表情特征的准确度也更高。为了区分各个表情,需要在特征提取阶段提取同一个体的不同表情所具有的特征,进行表情分类,主要会用到分类器对之前提取到的特征进行分类。在选用分类器的时候,尽量会选择一些性能比较完善、效率较高的级联分类器。所以,综

上所述,要想准确地识别人脸表情,至为重要的一步就是取决于能否精细地提取不同表情的特征,然后进行分类。而整个过程中的重点则是把能提取到的特征点高效地分类,从而提高表情识别率^[9-13]。

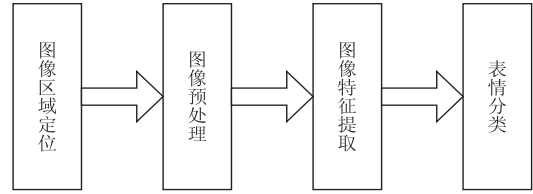


图 1 人脸表情识别系统框架

在通过对表情图像的特征提取后,各个表情就有了自己的特征,但由于表情特征提取的时候是多维提取,所以识别起来非常耗时,速度就比较慢,系统效率低。想要解决这一问题就得通过降维处理,把高维转成低维,这样就能降低难度,减少时间,提高效率,从而提高特征的有效性。

2 算法提出

卷积神经网络的结构示意图如图 2 所示。

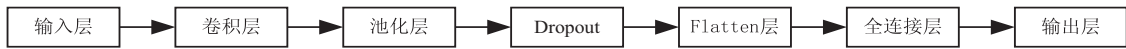


图 2 卷积神经网络结构

卷积层是通过卷积操作来提取特征的,将图像中每一个元像素乘上滤波器矩阵对应的元素所得到的。并且在卷积过程中,卷积核必须滑动到输入图像的每一个位置,图像上的每一个点都需要进行卷积运算,只有图像上的任意一个点都通过了运算,才可以完整地获得图像的特征。经过无数次的实验可知,卷积层的数量越多,特征提取的能力越强^[7-8,14-16]。

如果第一层为卷积层,则该层的卷积计算公式为:

$$X_j^l = f(\sum_{i \in M_j} X_i^{l-1} * W_{ij}^l + B^l) \quad (1)$$

其中, X_i^{l-1} 为前一层神经元 i 的输出, M_j 为神经元 j 的卷积核的大小,也就是输入层的感受野, W_{ij}^l 为该层神

经元第 j 个输入所对应的权值, B^l 为该层偏置的大小^[9-10,15-16]。

假如 l 层为池化层,则该层的池化公式为:

$$x_j^l = f[\beta^l \times D(x_i^{l-1}) + B^l] \quad (2)$$

其中, β^l 为可训练参数, B 为该层偏置^[9]。

假设 l 层为全连接层,则该层的全连接层计算公式为:

$$\sigma_j^l = f(\sum_{i=1}^n x_i^{l-1} * \omega_{ji}^{(l)} + b_j^{(l)}) \quad (3)$$

其中, n 层为上一层的神经元个数, $\omega_{ji}^{(l)}$ 为该层神经元 j 与前一层神经元 i 的连接强度, $b_j^{(l)}$ 为该层神经元的偏置, $f(x)$ 为偏置函数。

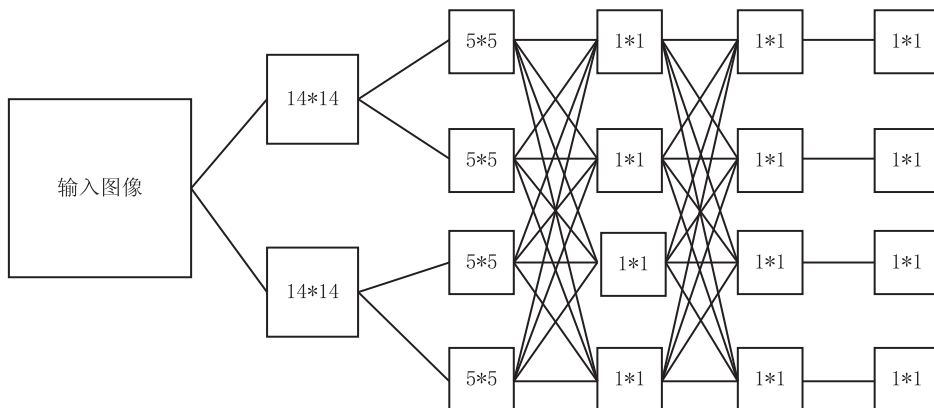


图 3 改进后的卷积神经网络

需要对上述的卷积神经网络进行改进^[11]。改进

后的卷积神经网络如图 3 所示。首先输入图像,然后

在第一层卷积层上对图像进行特征提取,这需要采用 20 个滤波器,并且每个滤波器的大小都为 5×5 ,步长都设为 2。由公式得下一层 20 个特征值为 28×28 大小的特征图。接下来再进行卷积操作,卷积核大小设置为 5×5 ,步长设为 2×2 ,经过这一层融合后就得到 45 个特征值为 5×5 大小的特征图。最后就是全连接层,这一层需根据具体情况进行调整。

将该结构单元进行 3 次叠加后,再连接 2 个全连接层 F1 和 F2 以及 1 个 Softmax 输出层 SF,就得到了改进后的模型。对该模型进行了 9 次实验,并根据实验结果分析,选取出了最优参数组合,改进要使训练避免过拟合现象。图 4 即为测试集上的混淆矩阵。采用 Dropout 技术后的测试结果要高于未采用 Dropout 技术的模型。

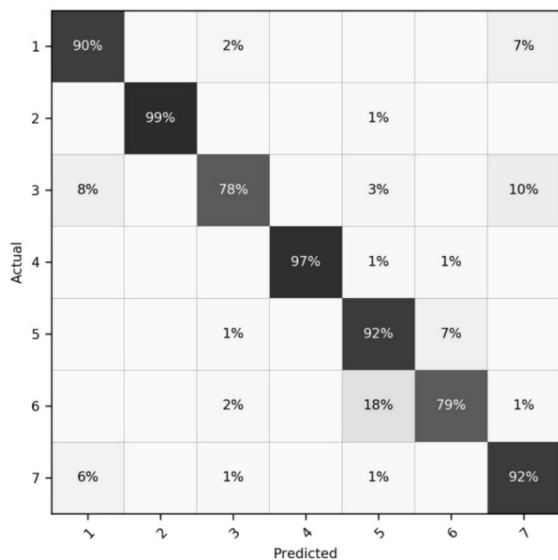


图 4 测试集上的混淆矩阵

原先网络使用的是 sigmoid 激活函数,为了提升泛化能力,故而把该函数改成 Relu 函数,因为 Relu 函数的非线性能力更强。Relu 函数的表达式为:

$$f(x) = \max(0, x) \quad (4)$$

当输入 $x \leq 0$ 时,输出为 0;当 $x \geq 0$ 时,输出为 x ,输入等于输出。

使用的 Relu 函数不会随着 x 的增加而逐渐趋于平衡,而双曲正切函数却会随之趋于平衡。因此,Relu 函数能够使神经网络的性能更加突出。

图 5 所示为改进后的卷积神经网络对人脸表情的分类图。

对比上述的三种人脸表情识别的算法不难发现,传统的表情识别方法最容易理解且逻辑清晰,步骤性很强,从采集图像到图像预处理再到特征提取最后识别,一步一步完成起来相对容易,但是效率却很低,耗时间很长。相比之下,基于卷积神经网络的算法是最简便最快速的。它把特征提取和分类合并到一块儿,

通过两个卷积层,两个池化层,一个全连接层,一个 softmax 回归层来实现。该方法比传统方法的效率高,且抗干扰性强,所接收到的图像不受外界复杂因素(例如光照等)的影响,准确度更好。由此可见,改进后的卷积神经网络的泛化能力有明显的上升,并且能够防止出现过拟合现象,它简化了层次,增加了步长,使得效率更高,准确性也更强。



图 5 改进后的卷积神经网络的人脸表情分类

3 实验

表 1 为利用传统人脸表情识别算法、基于卷积神经网络算法和改进的卷积神经网络算法得到的性别识别结果。由对比可得:改进的卷积神经网络人脸识别方法对性别的平均识别率最高为 95.1%,其次是基于卷积神经网络的人脸识别方法,为 90%,平均识别率最低的是传统的人脸识别方法。由此可知,对卷积神经网络进行改进是很有成效的,它不仅在过程上比传统的识别方法要简单得多,而且比普通卷积神经网络更加精确。

对比三种方法可得:改进的卷积神经网络人脸识别方法对表情的平均识别率也是最高,高达 99.0%,其次是基于卷积神经网络的人脸识别方法 90%,平均识别率最低的依然是传统的人脸识别方法。由于表情的分类较多,虽然对表情的识别率没有对性别的识别

率高,但是依然能够看出,卷积神经网络表情识别系统的正确率很高,它的泛化能力也很强。但是对比之下,

改进后的卷积神经网络随着步长的增加,因此它的准确率比改进前的卷积神经网络更高,泛化能力更强。

表 1 传统的人脸表情识别方法

传统的人脸表情识别方法				基于卷积神经网络的人脸表情识别方法				改进的卷积神经网络人脸识别方法			
样本数	正确数	错误数	识别率	样本数	正确数	错误数	识别率	样本数	正确数	错误数	识别率
500 (男)	402	78	84.4%	500 (男)	455	45	91.0%	500 (男)	498	2	99.1%
500 (女)	410	90	82.0%	500 (女)	445	55	89.0%	500 (女)	497	3	99.0%
平均识别率:83.2%				平均识别率:90%				平均识别率:99.0%			

4 结束语

与传统方法和普通卷积神经网络相比,改进的卷积神经网络在处理表情图像时,在选取出了最优参数组合的过程中,降低了训练过拟合现象的几率,同时提高了泛化能力,具有准确率高、效率高的优点,。

该课题利用卷积神经网络来实现人脸表情识别,通过实验表明该算法有较高的识别率和准确度,但仍有待改善和提高。

参考文献:

- [1] 张 杨,张仁杰. 基于改进 PCA 算法的人脸识别[J]. 软件导刊,2018,17(1):32-34.
- [2] REHMAN B,ONG W H,TAN A C H,et al. Face detection and tracking using hybrid margin-based ROI techniques[J]. The Visual Computer,2020,36(3):633-647.
- [3] FU J,ALVAR S R,BAJIC I V,et al. Fddb-360:face detection in 360-degree fisheye images[C]//2019 IEEE conference on multimedia information processing and retrieval (MIPR). San Jose,CA:IEEE,2019:15-30.
- [4] 郑泽宇,顾思宇. TensorFlow 实战 Google 深度学习框架[M]. 北京:电子工业出版社,2017.
- [5] LECUN Y,BENGIO Y,HINTON G. Deep learning[J]. Nature,2015,521(7553):436-444.
- [6] REN Chao,HE Xiaohai,TENG Qizhi,et al. Single image super-resolution using local geometric duality and non-local similarity[J]. IEEE Transactions on Image Processing,2016,25(5):2168-2183.
- [7] LIU C,SUN D. On Bayesian adaptive video super resolution[J]. IEEE Transactions on Pattern Analysis and Machine In-

telligence,2014,36(2):346-360.

- [8] SCHULTER S,LEISTNER C,BISCHOF H. Fast and accurate image upscaling with super-resolution forest[C]//Proceedings of IEEE conference on computer vision and pattern recognitions. Los Angeles:IEEE,2019:200-212.
- [9] YANG C Y,MA C,YANG M H. Single-image super-resolution;a benchmark[C]//European conference on computer vision. Zürich:Springer,2014:372-386.
- [10] DONG C,LOY C C,TANG X. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2019,38(2):295-307.
- [11] 李武军,王崇骏,张 炜,等. 人脸识别研究综述[J]. 模式识别与人工智能,2006,19(1):58-66.
- [12] 李春葆,李石军,李筱驰. 数据仓库与数据挖掘实践[M]. 北京:电子工业出版社,2014.
- [13] ZHU Xiangyu,LIU Hao,LEI Zhen,et al. Large-scale bisample learning on ID versus spot face recognition[J]. International Journal of Computer Vision,2019,127(6-7):684-700.
- [14] MICHAELI T,IRANI M. Nonparametric blind super-resolution[C]//Proceeding of IEEE international conference on computer vision (ICCV). Warsaw:IEEE,2018:350-360.
- [15] SINDAGI V A,PATEL V M. DAFE-FD: density aware feature enrichment for face detection[C]//2019 IEEE winter conference on applications of computer vision (WACV). Waikoloa Village,HI:IEEE,2019:212-216.
- [16] CHANG H,YEUNG D Y,XIONG Y. Super-resolution through neighbor embedding [C]//Proceedings of IEEE computer society conference on computer vision and pattern recognitions. Los Angeles:IEEE,2019:230-236.