

# 基于 VGG16 网络的人脸情绪识别\*

蔡靖,杜佳辰,王庆,周泓任

(吉林大学 仪器科学与电气工程学院,吉林 长春 130026)

**摘要:**近年来,如何通过人工智能对人的面部表情进行识别分析成为一个研究热点,利用人工智能可以快速地分析人的面部情绪,并以此为基础进行进一步研究。在深度学习中,传统的卷积神经网络存在对面部表情特征的提取不充分以及计算机参数量较大的问题,导致分类准确率较低。因此,提出了基于 VGG16 网络的人脸表情识别算法,通过与 InceptionV3、InceptionResNetV2、ResNet50 等模型实验对比,结果表明,VGG16 神经网络在 FER2013PLUS 测试数据集上的识别准确率为 79%,准确率比传统的卷积神经网络高。

**关键词:**深度学习;卷积神经网络;情绪识别;VGG16

中图分类号: TP183;TP391.41

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.211516

中文引用格式: 蔡靖,杜佳辰,王庆,等. 基于 VGG16 网络的人脸情绪识别[J]. 电子技术应用, 2022, 48(1): 67-70, 75.

英文引用格式: Cai Jing, Du Jiachen, Wang Qing, et al. Facial emotion recognition based on VGG16 network[J]. Application of Electronic Technique, 2022, 48(1): 67-70, 75.

## Facial emotion recognition based on VGG16 network

Cai Jing, Du Jiachen, Wang Qing, Zhou Hongren

(College of Instrumentation & Electrical Engineering, Jilin University, Changchun 130026, China)

**Abstract:** In recent years, how to recognize and analyze people's facial expressions through artificial intelligence has become a research hotspot. Using artificial intelligence can quickly analyze people's facial emotions, and further research is carried out on this basis. In deep learning, the traditional convolutional neural network can not extract facial expression features sufficiently, and the amount of computer parameters is large, which leads to low classification accuracy. Therefore, a facial expression recognition algorithm based on VGG16 neural network is proposed. Compared with the model experiments of InceptionV3, InceptionResNetV2 and ResNet50, the results show that the recognition accuracy of VGG16 neural network on FER2013PLUS test data set is 79%, which is higher than that of traditional convolution neural network.

**Key words:** deep learning; convolutional neural network; emotion recognition; VGG16

### 0 引言

Hinton 等人在研究深度卷积神经网络中通过引入 ReLU 函数以及采用多 GPU 训练等方式加快了网络的训练速度并提高了网络模型的准确率<sup>[1]</sup>。LeCun 等人表明深度学习通过使用反向传播算法来发现大型数据集的复杂结构,以指示机器应如何更改用于从前一层的表示计算每一层的表示的内部参数。深度卷积网络在处理图像、视频、语音和音频方面取得了突破性进展,而递归网络则为文本和语音等顺序数据带来了光明<sup>[2]</sup>。随着科学技术,尤其是计算机技术方面的不断进步,人们对卷积神经网络的研究也不断深入,面部表情识别技术也随之得到了发展。同时,一些问题也随之而来,例如:如何提高面部表情分类的准确率,如何能够在保证准确率的前提下减少计算的参数量等。

Tang 等人<sup>[3]</sup>提出将卷积神经网络(Convolutional Neural Networks, CNN)中融入支持向量机(SVM),同时修改了全连接层中的计算损失值的方法,在表情数据集上的测试结果为 71.2%,斩获了当年 Kaggle 比赛的一等奖。崔凤焦<sup>[4]</sup>将卷积神经网络、Adaboost 和支持向量机 3 种模型算法结构进行相互比较和设计优化,在 Cohn-Kanade 表情库中实现了对人脸情绪的智能识别,平均识别率最高达到了 74.92%。谢鑫等人<sup>[5]</sup>采用了耗时加长的杜鹃算法(CuckooSearch, CS)对 ResNet50 模型进行了变更改进,提高了准确率并使算法不会在调参时陷入梯度消失或爆炸的问题,准确率达到了 74.3%。陈佳等人<sup>[6]</sup>提出一种基于卷积神经网络集成的表情识别方法,子网结构在第一阶段训练 3 个子网络,采用 Batch Normalization 方法和 Dropout 层来加速网络的收敛。接着去掉 3 个子网络的输出层,使用 SVM 进行第二阶段的训练,完成最终表情的预测,其准确率达到了 70.84%。

\* 基金项目:吉林省科技发展计划项目(20190303043SF);吉林省教育厅“十三五”科学技术项目(JJKH20200964KJ)

针对以上人脸表情识别准确率较低的问题,本研究提出了一种基于 VGG16 模型算法的人脸表情识别方法来对人脸进行识别,并达到较好的准确率。

## 1 卷积神经网络

卷积神经网络<sup>[7]</sup>是近些年发展起来,并引起人们广泛重视研究的一种高效的识别算法,主要用于图像目标的识别,它的结构和生物大脑的神经网络比较相似,其是由一系列功能不相同的层级所构成的。卷积神经网络的结构一般主要包括其特有的卷积层、池化层这两部分及全连接层这第三部分。在卷积神经网络中,卷积层与池化层一般情况下都会呈多次交替地出现,因此不同的模型具有不同的层级结构。

### 1.1 卷积层

卷积层是卷积神经网络的核心,主要功能就是对卷积进行运算,实现特征提取操作。卷积核(Convolutional Kernel)的功能就是对输入矩阵进行加权平均的一种权值定义函数,卷积核在一个特征图上不断地移动,与输入矩阵根据一定的规则相乘及求和便可直接得到其特征<sup>[8]</sup>,可以直接通过一个卷积核实现对一个输入矩阵进行特征提取,这种特征提取方法是由低层次向高级提取<sup>[9]</sup>,这样做可以促使提取得到的特征更真实有效。卷积层可以满足下列公式:

$$x_{ij} = f\left(\sum_{a=0}^2 \sum_{b=0}^2 y_{ab} z_{ab} - y\right) \quad (1)$$

式中, $y_{ab}$ 为卷积核中的值, $z_{ab}$ 为输入矩阵的值, $x_{ij}$ 为经过卷积计算后得到的特征矩阵的值。

通过卷积运算的到的特征值,可以判断输入区域内是否含有所需要的特征。

### 1.2 池化层

池化层主要处理卷积层输出的维度较大的特征,可以大大减小矩阵的尺寸,方便进行后续处理。池化层可以在减少数据量的同时,避免发生过拟合。池化层将特征分为几个较小的区域,取其最大值或平均值<sup>[9]</sup>,得到的特征维度更小。目前常见的池化方法有:平均池化、最大池化、随机池化、谱池化。

### 1.3 全连接层

全连接层的作用是将上一层的各个局部特征综合起来,其每个节点都与上一层的节点有连接,它通过权值矩阵组成一个全新的图<sup>[9]</sup>,变成全局特征。全连接层可以将特征矩阵转换为单个的值,因此可以在减小数据量的同时,减小特征位置对分类带来的影响。通过全连接层得到的向量,其中每个值代表样本属于每个类的可能性的大小<sup>[8]</sup>,因此可以通过最后的输出向量,判断样本的类别。

## 2 基于 VGG16 网络的情绪识别系统

### 2.1 数据集

FER2013 数据集由 Pierre Luc carrier 和 Aaron Courville 通过搜索相关的情感关键词在网络上爬取,这些图像都

是经过人工标注的。它分为三部分,其中测试集共有 28 708 张表情图片,公有验证集和私有验证集各 3 589 张,共 35 886 张表情图片。每张图片是由大小为 48×48 像素的灰度图组成,分为 7 个类别,分别为快乐、惊讶、悲伤、愤怒、厌恶、恐惧、中性。

原始的数据集并没有直接以图片的形式给出,而是以保存在 csv 格式文件中,使用 pandas 解析 csv 文件,得到 jpg 文件并储存到对应文件夹下,然后将数据集分为训练集和测试集。

通过对 FER2013 数据集进行重新标签产生的 FER-2013PLUS 数据集在模型训练中会有更高的准确率。图 1 展示了数据集中的几个样本。



图 1 解析后的 jpg 格式数据集

### 2.2 VGG16 神经网络模型

VGG16 网络结构<sup>[10-12]</sup>如图 2 和表 1 所示,输入分辨率为 224×224 的 3 通道图像进入卷积层。在结束卷积操作后,由于数据经过非线性函数的映射之后,存在着取值区间向极限饱和和区逐渐靠拢的输入分布,通过对所有输入数据的归一化处理,可以使得数据被强制地拉回至方差为 1、平均值为 0 的标准正态分布。最后再次进入池化层,逐渐忽略局部特征信息。循环 4 轮以上操作,在卷积操作全部完成后,人脸特征信息即可进入全连接层,将包含有局部信息的特征图(包括特征图的高、宽、通道数)全部映射到 4 096 维度。VGG16 网络结构的卷积层的 kernel 大小均为 3×3,效果较好,其步长为 1,填充方式为“same”填充,与其他填充方式相比,这种填充方法可以使每一次卷积以后得到的卷积结果的尺寸大小不会发生变化。池化层使用的是 2×2 大小的池化核,填充方式与卷积层一样也为“same”填充,激励函数为 ReLU。

### 2.3 VGG16 神经网络训练过程

输入大小为 224×224×3 的图片,经过 Conv1\_1 和 Conv1\_2 两个卷积层,filter 为 3×3,卷积结束后得到 222×222×1 的矩阵,再经过矩阵填充得到 224×224×1 的矩阵。由于第一层有 64 个卷积核,原来的 224×224×1 就变成了 224×224×64。

池化层使用的卷积核是 2×2×64,步长为 2,第一层卷积结果经过池化后到达第二层,得到的矩阵维数为原来的一半,变成了 112×112×128;第二层经过池化层,矩

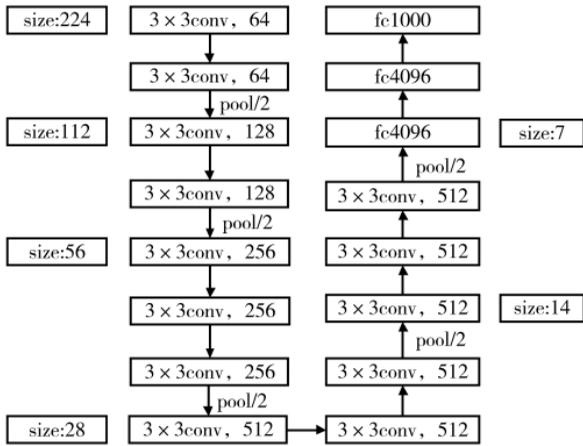


图 2 VGG16 网络

表 1 VGG16 结构及其参数量

类型	输出大小	参数量
input	(None, 224, 224, 3)	0
Block1_conv1	(None, 224, 224, 64)	1 792
Block1_conv2	(None, 224, 224, 64)	36 928
Block1_pool	(None, 112, 112, 64)	0
Block2_conv1	(None, 112, 112, 128)	73 856
Block2_conv2	(None, 112, 112, 128)	1 147 584
Block2_pool	(None, 56, 56, 128)	0
Block3_conv1	(None, 56, 56, 256)	295 168
Block3_conv2	(None, 28, 28, 256)	590 080
Block3_conv3	(None, 28, 28, 256)	590 080
Block3_pool	(None, 28, 28, 256)	0
Block4_conv1	(None, 28, 28, 512)	1 180 160
Block4_conv2	(None, 28, 28, 512)	2 359 808
Block4_conv3	(None, 28, 28, 512)	2 359 808
Block4_pool	(None, 14, 14, 512)	0
Block5_conv1	(None, 14, 14, 512)	2 359 808
Block5_conv2	(None, 14, 14, 512)	2 359 808
Block5_conv3	(None, 14, 14, 512)	2 359 808
Block5_pool	(None, 7, 7, 512)	0
flatten	(None, 25088)	0
Fc1	(None, 4096)	102 764 544
Fc2	(None, 4096)	16 781 312
Predictions	(None, 1000)	4 097 000

阵维数减半,进入到第三层,由于第三层有 256 个卷积

核,那么得到  $56 \times 56 \times 256$  的矩阵;同理,第四层有 512 个卷积核,得到  $28 \times 28 \times 512$  的矩阵;而第五层也只有 512 个卷积核,所以得到  $14 \times 14 \times 512$  的矩阵。经过第五次池化最终得到  $7 \times 7 \times 512$  的矩阵,进入到全连接层。

VGG16 的全连接层有 3 层。其每一个节点都与上一层每个节点相连接,它可以把前一层的输出特征都整体综合起来。在 VGG16 模型结构中,第一层有 4 096 个节点,第二层也有 4 096 个,第三层只有 1 000 个节点,最后再进行 softmax 成类别的数目。

如图 3 所示,图中显示了图片经过 VGG16 各层卷积层和池化层后所提取出的结果,依箭头方向分别为原始图片、第一次卷积后的图片、第一次池化后的图片、第二次卷积后的图片、第二次池化后的图片、第三次卷积后的图片、第三次池化后的图片、第四次卷积后的图片、第四次池化后的图片、第五次卷积后的图片、第五次池化后的图片。

### 3 实验结果

实验环境为 Windows 10 操作系统,采用 Python3.7 编程语言。实验过程依次是数据预处理、搭建神经网络模型、训练、测试、评估模型。

ResNet50、InceptionResNetV2、InceptionV3、VGG16 模型训练结果分别如表 2~表 5 所示。表 6 为 4 种模型的准确率对比,可以看到准确率分别为 73%、78%、78%、79%。

表 2 ResNet50 模型训练结果

	precision	recall	f1-score	support
NF	0.67	0.55	0.60	11
anger	0.66	0.62	0.64	319
contempt	0.86	0.25	0.39	24
disgust	0.89	0.47	0.62	34
fear	0.51	0.50	0.50	74
happiness	0.85	0.78	0.81	899
neutral	0.71	0.83	0.76	1 335
sadness	0.51	0.50	0.51	412
surprise	0.83	0.74	0.78	456
unknown	0.33	0.08	0.13	25
accuracy			0.73	3 589
Macro avg	0.68	0.53	0.57	3 589
Weighted avg	0.73	0.73	0.72	3 589

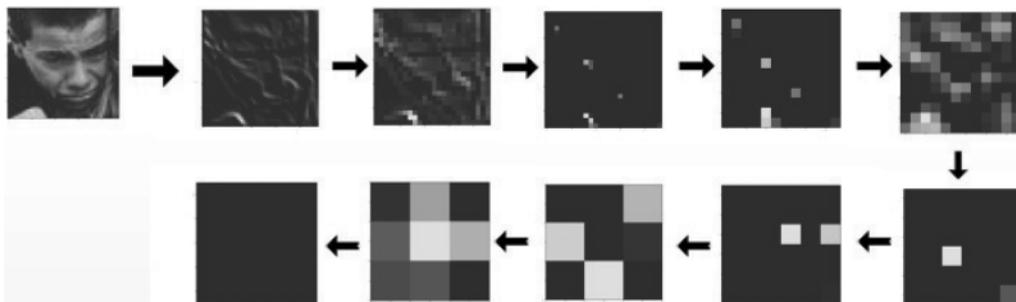


图 3 VGG16 训练图

表 3 InceptionResNetV2 模型训练结果

	precision	recall	f1-score	support
neutral	0.75	0.55	0.63	11
happiness	0.74	0.67	0.70	319
surprise	0.86	0.25	0.39	24
sadness	0.77	0.50	0.61	34
anger	0.69	0.50	0.58	74
disgust	0.85	0.89	0.87	899
neutral	0.80	0.82	0.81	1 335
fear	0.62	0.57	0.59	412
contempt	0.77	0.89	0.83	456
NF	1.00	0.08	0.15	25
accuracy			0.78	3 589
Macro avg	0.68	0.53	0.62	3 589
Weighted avg	0.73	0.73	0.78	3 589

表 4 InceptionV3 模型训练结果

	precision	recall	f1-score	support
neutral	0.29	0.18	0.22	11
happiness	0.77	0.62	0.69	319
surprise	0.71	0.21	0.32	24
sadness	0.72	0.38	0.50	34
anger	0.70	0.45	0.55	74
disgust	0.81	0.92	0.86	899
fear	0.78	0.85	0.81	1 335
contempt	0.65	0.52	0.58	412
unknown	0.86	0.81	0.84	456
NF	0.40	0.16	0.23	25
accuracy			0.78	3 589
Macro avg	0.67	0.51	0.56	3 589
Weighted avg	0.77	0.78	0.77	3 589

表 5 VGG16 模型训练结果

	precision	recall	f1-score	support
NF	0.23	0.64	0.34	11
anger	0.75	0.72	0.73	319
contempt	0.43	0.25	0.32	24
disgust	0.73	0.47	0.57	34
fear	0.60	0.58	0.59	74
happiness	0.84	0.91	0.88	899
neutral	0.81	0.83	0.82	1 335
sadness	0.62	0.52	0.57	412
surprise	0.85	0.86	0.85	456
unknown	0.18	0.08	0.11	25
accuracy			0.79	3 589
Macro avg	0.60	0.59	0.58	3 589
Weighted avg	0.78	0.79	0.78	3 589

表 6 模型准确率对比

模型	ResNet50	InceptionResNetV2	InceptionV3	VGG16
准确率/%	73	78	78	79

另外,利用 VGG16 模型搭建出用于识别人脸情绪的

UI 界面,如图 4 所示。



图 4 UI 结果显示

#### 4 结论

本文采用改进后的 FER2013 数据集分别对 4 种网络模型进行训练,用于寻找进行人脸表情识别的最佳模型。通过实验结果对比得知,VGG16 模型的准确率相比于其他 3 种模型准确率更高。在网络的搭建与训练过程中,发现不同的数据集和网络结构都会对模型的准确率造成影响,后续可以在数据集的预处理以及损失函数的选择等方向上进行研究,以进一步提高人脸识别模型的准确率。

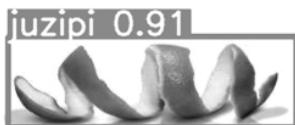
#### 参考文献

- [1] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[C]// Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012: 1097-1105.
- [2] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature: International Weekly Journal of Science, 2015, 521(7553): 436-444.
- [3] TANG Y. Deep learning using linear support vector machines[J]. arXiv preprint arXiv: 1306.0239, 2013.
- [4] 崔凤焦. 表情识别算法研究进展与性能比较[J]. 计算机技术与发展, 2018, 28(2): 145-149.
- [5] 谢鑫. 基于 ResNet50 的人脸识别模型[J]. 科技资讯, 2020, 18(35): 9-11.
- [6] 陈佳, 褚丽莉, 周影. 基于 CNN 集成学习的人脸表情识别系统的设计[J]. 电脑与信息技术, 2021, 29(1): 10-12.
- [7] 李晓阳, 谢恒义, 韩贞辉, 等. 卷积神经网络在图像识别技术中的应用研究[J]. 能源与环保, 2020, 42(6): 73-76.
- [8] 王怡文. 基于卷积神经网络及特征提取的面部表情识别算法[J]. 电脑知识与技术, 2019, 15(16): 203-205.
- [9] 李航, 厉丹, 朱晨, 等. 基于卷积神经网络的图像识别系统[J]. 电脑知识与技术, 2020, 16(10): 196-197, 200.
- [10] 戚超, 左毅, 陈哲琪, 等. 基于改进 VGG16 的大米加工精

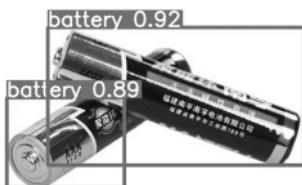
(下转第 75 页)



(a) Paper ball



(b) juzipi



(c) battery

图7 测试结果图

## 参考文献

- [1] Hu Jing, Zhang Bo. Application research of automatic garbage sorting based on TensorFlow and OpenCV[J]. Journal of Physics: Conference Series, 2021, 1883(1): 012169.
- [2] Wang Yilin, Ji Baofeng, Zhao Yunfei, et al. Intelligent garbage sorting device based on machine vision neural network[J]. Scientific Journal of Intelligent Systems Research, 2021, 3(4): 96-102.
- [3] 方海峰, 许佩雯, 万信飞, 等. 基于并联机器人可回收垃圾分拣系统的设计[J]. 组合机床与自动化加工技术, 2020(11): 24-27, 32.
- [4] 吴蓬勃, 姚美菱, 王拓, 等. 基于 TensorFlow 的垃圾分拣机器人设计[J]. 实验室研究与探索, 2020, 39(6): 117-122.
- [5] 韩伟娜, 张伟聪, 韩旭. 垃圾分拣机械手的设计及性能仿真研究[J]. 现代电子技术, 2019, 42(12): 86-89.

(上接第 70 页)

- 度分级方法研究[J]. 农业机械学报, 2021, 52(5): 301-307.
- [11] 柯溢, 卢亚玲. 基于 VGG16 的宠物犬识别方法[J]. 电子制作, 2020(21): 42-45.
- [12] 吉鑫, 陈剑云, 完颜幸幸. 用于接触网可视化接地识别的改进 VGG-16 模型[J]. 传感器与微系统, 2021, 40(2): 58-60, 72.

- [6] 王众磊, 陈海波. TensorFlow 移动端机器学习实战[M]. 北京: 中国工信出版集团, 2019.
- [7] 孙磊, 肖金球, 夏禹, 等. 改进的基于嵌入式 SoC 卷积神经网络识别模型[J]. 计算机应用与软件, 2020, 37(3): 257-260.
- [8] Sen Yu, He Yongze, Shao Mingyang. Real-time target detection based on Jetson Nano and YOLOv3-Tiny[J]. International Core Journal of Engineering, 2021, 7(2): 0049.
- [9] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: optimal speed and accuracy of object detection[J]. ArXiv: 2004.10934[cs.CV], 2020.
- [10] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2018: 8759-8768.
- [11] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2019: 658-666.
- [12] ZHENG Z, WANG P, LIU W, et al. Distance-IoU Loss: faster and better learning for bounding box regression[C]// AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [13] YUN S D, HAN D Y, SEONG J O, et al. CutMix: regularization strategy to train strong classifiers with localizable features[C]//2019 IEEE/CVF International Conference on Computer Vision(ICCV), 2019: 6022-6031.
- [14] 李季, 史晨发, 邵磊, 等. 基于改进 RRT 算法的 6-DOF 机器人路径规划[J]. 计算机应用与软件, 2020, 37(9): 221-226.
- [15] 沙勇. 基于遗传算法机器人路径规划研究[J]. 齐齐哈尔大学学报(自然科学版), 2021, 37(5): 55-56, 66.

(收稿日期: 2021-05-25)

## 作者简介:

王慧(1996-), 女, 硕士, 主要研究方向: 嵌入式系统与控制工程。

蒋朝根(1960-), 男, 硕士, 教授, 主要研究方向: 嵌入式系统及其应用与无线传感器网络。



扫码下载电子文档

(收稿日期: 2021-03-16)

## 作者简介:

蔡靖(1979-), 男, 硕士研究生, 高级工程师, 主要研究方向: 医疗器械。

杜佳辰(1998-), 男, 本科, 主要研究方向: 检测技术。

王庆(1998-), 男, 本科, 主要研究方向: 检测技术。



扫码下载电子文档

## 版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所