

基于 VGGNet 的人脸表情识别研究

廖清江, 刘 婷, 张星月, 董 祺, 李乐乐, 刘嘉豪

(天津商业大学信息工程学院, 天津 300134)

✉ 3264358645@qq.com; liuting@tjcu.edu.cn; 3367328354@qq.com; 3347692550@qq.com;
2406697527@qq.com; 1912842862@qq.com



摘 要:针对传统卷积神经网络(Convolutional Neural Network, CNN)在人脸表情识别中有效特征提取不够、泛化能力不强及识别准确性不高等缺点,研究选用具有更小卷积核和更深池化层的视觉几何组网络(Visual Geometry Group Network, VGGNet)进行人脸表情识别系统的设计。为了验证识别效果,在传统 CNN 和 VGGNet 框架下进行人脸表情识别系统的搭建,使用 FER2013 公开数据集进行训练和测试,并对测试结果进行对比分析。实验表明,传统 CNN 模型在人脸表情识别任务中的识别率仅为 88%,而 VGGNet 则能够取得高达 98%的识别率。此外,利用真人实际测试验证了研究搭建的 VGGNet 模型具有很好的表情识别效果。

关键词:深度学习;卷积神经网络;VGGNet;人脸表情识别;FER2013

中图分类号:TP391 **文献标志码:**A

Research on Face Expression Recognition Based on VGGNet

LIAO Qingjiang, LIU Ting, ZHANG Xingyue, DONG Qi, LI Lele, LIU Jiahao

(School of Information Engineering, Tianjin University of Commerce, Tianjin 300134, China)

✉ 3264358645@qq.com; liuting@tjcu.edu.cn; 3367328354@qq.com; 3347692550@qq.com;
2406697527@qq.com; 1912842862@qq.com

Abstract: Given the shortcomings of traditional Convolutional Neural Networks (CNN) in face expression recognition, such as insufficient effective feature extraction, weak generalization ability, and low recognition accuracy, this paper proposes to design a face expression recognition system by using Visual Geometry Group Network (VGGNet), which has smaller convolutional kernel and deeper pooling layer. To verify the recognition effect, the face expression recognition system is built under the framework of traditional CNN and VGGNet, the FER2013 public dataset is used for training and testing, and the testing results are analyzed and compared. The experiments show that the recognition rate of the traditional CNN model in the task of face expression recognition is only 88%, while the recognition rate of the VGGNet is as high as 98%. In addition, test with real persons verifies that the VGGNet model proposed in this paper has a very good expression recognition effect.

Key words: deep learning; CNN; VGGNet; face expression recognition; FER2013

0 引言(Introduction)

面部表情是除语言性交流以外的另一种传递信息的身体语言符号,是人们情感状态的重要体现之一,可以表达出个体

的心态和情感。随着人工智能技术与计算机视觉领域的迅速发展,人脸表情识别技术成为当下图像识别研究领域的热点之一。人脸表情识别作为人脸识别技术的进一步发展,现已被广

泛应用于心理学^[1]、疼痛表情识别^[2]、课堂状态检测^[3]、安全监控^[4]等领域。

早期的人脸表情识别方法基于面部的几何特征,主要依赖手工的特征提取和分类算法,容易漏掉分类识别时的关键特征,存在较大的局限性。研究者尝试将人脸表情识别建模为模板,实现对表情的识别与分类,但其泛化能力和对非刚性变形的适应性相对有限。随着机器学习技术的不断发展,研究者发现深度学习具有引入数据集规模大、能够表示复杂特征的特点,因此在人脸表情识别中应用卷积神经网络(Convolutional Neural Network, CNN)^[5]、视觉几何组 16(Visual Geometry Group 16, VGG16)^[6]、残差神经网络 50(Residual Neural Network 50, ResNet50)^[7]、三维卷积神经网络(Convolutional 3D, C3D)^[8]成为目前的主要研究方向。

为了更好地完成人脸表情识别任务,本研究选择 VGGNet 进行人脸表情识别的设计,为了对比识别效果,对传统 CNN 网络和 VGGNet 网络进行框架搭建,针对同一数据集 FER2013 分别进行测试分析,对比其损失率及识别率。为了验证 VGGNet 在真实环境中的识别效果,本研究还进行了真人实时表情测试,测试表明本文设计的方案能很好地完成人脸表情识别任务。

1 CNN 基本原理(CNN basic principles)

卷积神经网络是一种深度学习模型,最早被应用于图像数据处理,现不仅广泛应用于计算机视觉领域的图像和视频处理,还应用于自然语言处理和语音识别。不同于传统的机器学习, CNN 包含多个层次的处理单元,由输入层、卷积层、池化层(下采样层)、全连接层和输出层组成。其中, CNN 的核心是卷积操作,即在输入数据上进行卷积操作,使用滤波器提取图像特征,使 CNN 具有局部感知性和参数共享的特点,从而减少了需要训练的参数数量,提高了模型的效率和泛化能力。 CNN 的工作原理是通过模仿人类大脑的视觉系统达到自动从图像中提取特征,并进行分类、分割、检测操作。图 1 为卷积神经网络的结构图。

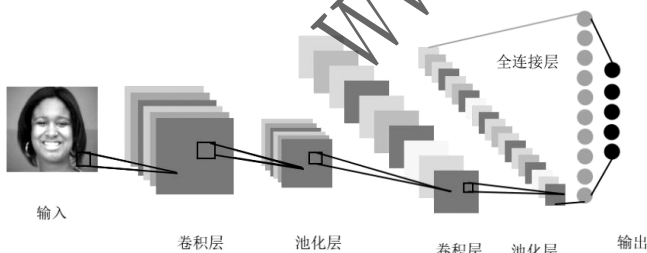


图 1 卷积神经网络的结构图

Fig. 1 Structure diagram of convolutional neural networks

2 VGGNet 原理及结构(Principle and structure of VGGNet)

VGGNet 是由牛津大学的研究团队 Karen Simonyan 和 Andrew Zisserman 等在 2014 年提出的一种深度卷积神经网络模型,该模型在 2014 年的 ImageNet 图像识别比赛中取得了分类任务第 2 名和定位任务第 1 名的好成绩,证明增加卷积神经

网络的层数可以在一定条件下影响网络最终的性能,使图片分类识别错误率大大降低,同时提升了卷积神经网络的拓展性。

VGGNet 的网络结构包含多个连续的 3×3 卷积层代替 5×5 卷积层,每个卷积层后面紧跟着一个 2×2 的最大池化层进行下采样,便于提取多尺度的图像特征。VGGNet 由多个卷积层和池化层组成,最后连接全连接层进行分类。通过深层堆叠的卷积层和池化层的设计,使得网络在不增加太多参数的情况下增加网络的深度,提高网络的深层表达能力。常用的版本是 VGG16 和 VGG19,本研究选用 VGG16 结构。

3 基于 VGGNet 的人脸表情识别的设计(Design of facial expression recognition based on VGGNet)

3.1 系统设计流程

本研究主要对人脸表情进行识别,通过人脸表情判断当前测试者所处的精神状态和情绪诉求,在输入端进行人脸检测,在输出端显示检测者包括中性(Neutral)、高兴(Happy)、生气(Angry)、惊讶(Surprise)、厌恶(Disgust)、难过(Sad)和恐惧(Fear)的情绪类别。

系统流程图如图 2 所示。

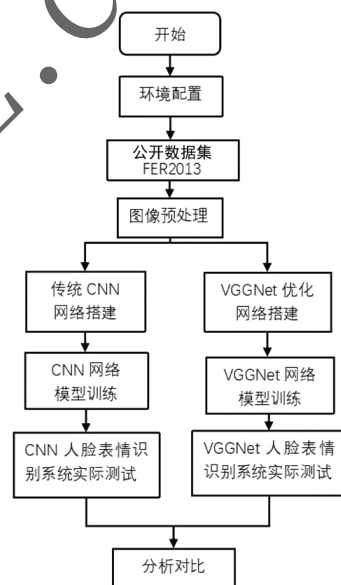


图 2 系统流程图

Fig. 2 System flowchart

3.2 数据集选择

采用 FER2013 公开数据集,该数据集最初是在 2013 年的 Kaggle 竞赛中进行面部表情分类而创建的,它包含 35 887 幅灰度图像,每幅图像的大小为 48×48 像素。FER2013 公开数据集中的每幅图像都被标记为 7 个不同的情绪类别之一,包括中性(Neutral)、高兴(Happy)、生气(Angry)、惊讶(Surprise)、厌恶(Disgust)、难过(Sad)和恐惧(Fear)。每个样本都有 1 个对应的情绪标签。FER2013 公开数据集分为 28 709 幅图像的训练集、3 589 幅图像的验证集及 3 589 幅图像的测试集;它的特点在于质量较高、图像样本更具有多样性。一些图像可能存在模糊、光照不均匀、遮挡等问题,这使得面部表情分类任务具有一定的挑战性。

数据集包含逗号分隔值(Comma-Separated Values, CSV)文件,该文件将图片转化为像素数据以方便存储。首先使用 Python 对 CSV 文件提取标签实现可视化,其次创建映射表并重新加载数据集,最后进行图像预处理。

3.3 图像预处理

虽然 FER2013 公开数据集中的各个图形都已经很清晰,但是拍摄角度、光线以及摄像机所自带的噪音等都会对图像的清晰度、图像的识别率产生影响,为了避免出现因图像有噪音或角度等问题而引发的识别错误问题,需要对图片进行预处理。

3.3.1 高斯模糊

高斯模糊去噪处理的目的是降低细节层次以及图像的毛刺瑕疵;工作原理是将图像从中间划分为两半,利用高斯分布的特性进行图像模糊操作,然后将图像中细分的像素周围的适当范围内乘以一个高斯系数。

Opencv 库函数框架下的 `cv2. GaussianBlur(img, (ksizex, ksizey), 0)` 可以对图像实现高斯模糊,其中 `img` 表示输入的人脸表情原图像, `(ksizex, ksizey)` 表示高斯模糊核的大小,而 0 表示通过 `ksize` 写出的 `sigma` 决定以何种方式表示高斯核。

在图像预处理时,通常会将红绿蓝(Red Green Blue, RGB)的彩色图像转换为灰度图,其原因是 RGB 彩色是一个三通道空间,在预处理时容易产生交互影响,并且彩色图像所占空间远远大于灰度图所占空间,在人脸表情识别中,灰度图不仅容易展示人脸表情轮廓,而且节省空间资源。基于以上考虑,在人脸表情预处理时将图像转换为灰度图。在 Opencv 中,转换关系为如下:

```
gray_img=cv2. cvtColor(img, COLOR_BGR2GRAY) (1)
```

其中: `cv2. cvtColor()` 函数表示转换为灰度图, `img` 表示输入的原图像, `COLOR_BGR2GRAY` 表示将彩色图像转换为灰度图。

3.3.2 直方图均衡化

为了增强图像对比度以及减少光线对图像分析的影响,可以对图像进行直方图均衡化,同时直方图均衡化在提高对比度的同时,还能提高特征提取率。在 Opencv 库中,使用直方图代码如下:

```
equ=equalizeHist(img) (2)
```

其中: `equalizeHist()` 表示直方图均衡化, `img` 表示原图像。

3.3.3 数据类型转换

在 PyTorch 框架下, `torch` 数据类型是最基本的数据类型,输入变量的数据类型由 `numpy` 类型转化为 `torch` 类型,其代码实现过程如下所示。

```
face_tensor=torch. from_numpy(face_normalized)
# 将 python 中的 numpy 数据类型转化为 torch 数据类型
face_tensor=face_tensor. type('torch. FloatTensor')
# 指定为 torch. FloatTensor 型,避免数据类型不匹配
```

4 实验及结果分析(Experiment and result analysis)

4.1 实验环境

本研究采用的深度学习框架为 Tensorflow 2. 12. 0,语言

为 Python3. 10. 9,集成开发环境为 PyCharm2023. 1 版本,选择 CPU 版本的 PyTorch 框架,导入 Numpy 库、Matplotlib 库、tqdm 库、Opencv 库等,所有实验均在 Windows 10 操作平台下进行。

4.2 网络搭建和模型训练

4.2.1 传统 CNN

定义时继承 `torch. nn. Module` 类,从中调取卷积神经网络的结构。

首先进行前向传播,训练数据样本作为卷积神经网络的输入,经过 3 次卷积和 3 次池化,图像特征信息进入 3 层全连接层进行整合,使用 Dropout 防止过拟合现象,每层设置激活函数进行激活,并传递到全连接层,得到对应的预测值。

其次运用损失函数计算损失值衡量模型预测与真实标签之间的具体差异。这里将计算的误差记为 p ,具体计算过程如下:

$$p = |z - y| \quad (3)$$

其中: z 为真实值, y 为预测值。

最后反向传播沿着网络的层级关系逆向传播梯度,将梯度信息传递回网络中的每一层。这样根据梯度更新网络中的权重和偏置,减小损失函数的值,从而训练出用于人脸表情识别的网络模型。CNN 网络训练过程如图 3 所示。

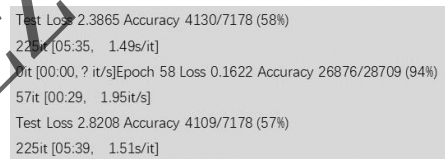


图 3 CNN 训练过程图

Fig. 3 Diagram of CNN training process

4.2.2 VGGNet

执行迁移学习,使用 PyTorch 引入经过训练的 VGG 模型,继承 `nn. Module` 的 VGGNet,并修改网络中标准交叉熵损失、学习率等参数。训练得到人脸表情识别系统模型,并引入 transform 进行图形的裁剪、旋转、缩放和亮度调节,提高模型的泛化性和抗拒性。

引入预训练模型过程如图 4 所示。

```
import torchvision. models as models
device = torch. device("cuda" if torch. cuda. is_available() else "cpu")#根据配置选择 CPU 或 GPU
model_ft = models. vgg16(pretrained=True)
```

图 4 预训练模型引入图

Fig. 4 Introduction diagram of the pre-trained model

4.3 结果对比分析

训练过程中将学习率设置为 0. 01。根据数据集样本数量特点,先选取 60 次作为迭代参数在 FER2013 公开数据集上进行实验。经过实验验证,在迭代 60 次以后,传统 CNN 的识别率为 88%,VGGNet 的识别率为 98%,二者的训练集与验证集的识别率对比如图 5 所示。

迭代 60 次以后,传统 CNN 的损失率为 18. 81%,而 VGGNet 的损失率则减小至 7. 84%,两个网络的训练集与验证集损失率对比如表 1 所示。

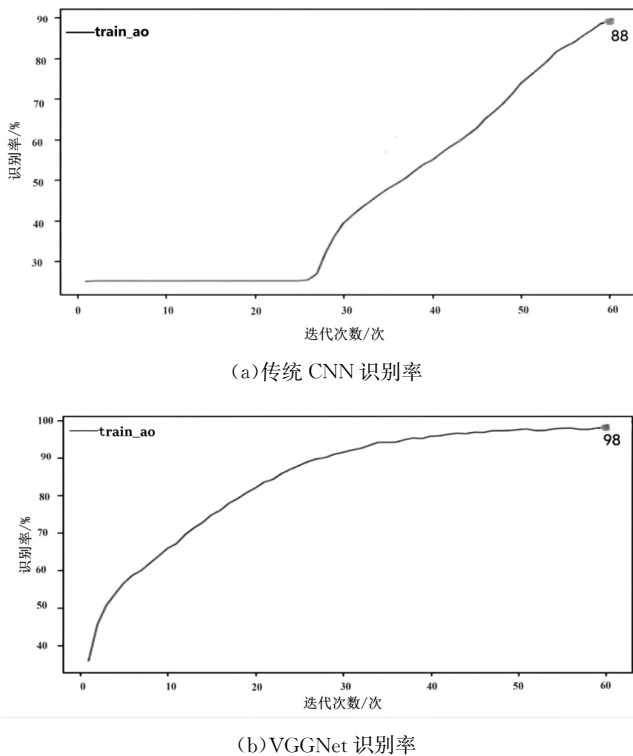


图 5 传统 CNN 与 VGGNet 识别率对比图

Fig. 5 Comparison of recognition rates between traditional CNN and VGGNet

表 1 传统 CNN 与 VGGNet 损失率对比表

Tab.1 Comparison table of loss rates between traditional CNN and VGGNet

网络类型	损失率/%
传统 CNN	18.81
VGGNet	7.84

4.4 现实应用能力评估

为测试模型在现实场景中的应用能力,保存性 FER2013 公开数据集上训练的网络模型,并运用该模型对研究团队中的成员进行面部表情识别,识别结果如图 6 所示。从识别结果来看,本文所提模型能准确识别现实生活中的人物表情。从整体效果来看,模型泛化能力较好。

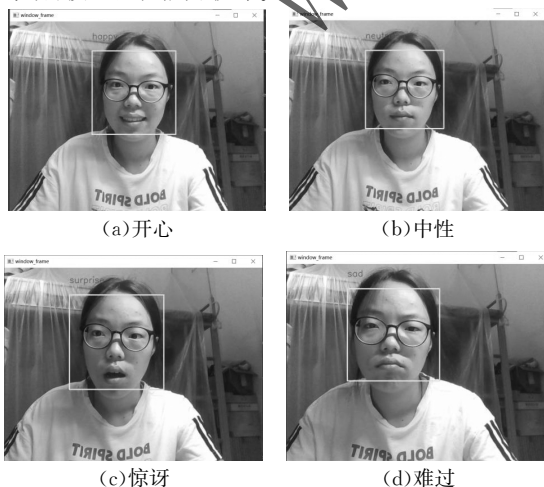


图 6 现实场景中的人脸表情识别图

Fig. 6 Facial expression recognition images in real-life scenarios

5 结论(Conclusion)

在 Tensorflow 2.12.0 深度学习框架下,利用 Python 3.10.9 语言,PyCharm 2023.1 集成开发环境,以准确识别人脸表情为目的,在 FER2013 公开数据集上搭建基于 VGGNet 的人脸表情识别系统,并对比了基于传统 CNN 和 VGGNet 两种不同网络模型在人脸表情识别中的性能。同时,在现实场景中验证了所提模型的有效性。

文中提出的人脸表情识别方案优势明显,但在某些方面仍有不足:

(1)训练时间长:本文所提系统进行人脸表情训练时需要耗费大量时间,在训练期间,对特征信息提取能力和计算能力要求较高,要求使用有强算力的 GPU、TPU 等图像处理器。

(2)数据集依然有待提高:虽然 FER2013 公开数据集拥有 30 000 余幅人脸表情图像,但是运用于实际时,依然会出现不能精确识别等问题,应采取更多其他人脸表情识别数据集训练网络,以达到提高识别精确度及不同人群的适用性的目的。

参考文献(References)

- [1] 金辉,高文. 人脸面部混合表情识别系统[J]. 计算机学报, 2000,23(6):602-608.
- [2] 卢官明,宋统帅,楼亦墨,等. 基于图的半监督深度学习及其在新生儿疼痛表情识别中的应用[J]. 南京邮电大学学报(自然科学版),2022,42(1):53-61.
- [3] 李燕苹,李一凡. 基于表情识别的课堂状态检测分析[J]. 工业控制计算机,2023,36(1):116-118.
- [4] 吴家宝,张运喜,刘强,等. 戴口罩多人脸识别系统研究[J]. 天津职业技术师范大学学报,2023,33(2):7-13.
- [5] 孙歌,王剑雄,欧琪,等. 基于 CNN 和面部关键点的表情分析研究[J]. 河北建筑工程学院学报,2022,40(4):148-153.
- [6] 方霖,童强. 基于 VGG16 网络的参数优化方法研究[J]. 湖北师范大学学报(自然科学版),2023,43(1):44-50.
- [7] 张洁,穆静,钱智哲. 改进的 ResNeXt50 神经网络面部表情识别方法[J]. 西安工业大学学报,2022,42(6):610-619.
- [8] RAJAGOPAL S D, RAMACHANDRAN B. 3D face expression recognition with ensemble deep learning exploring congruent features among expressions[J]. Computational Intelligence,2022,38(2):345-365.

作者简介:

廖清江(2002-),女,本科生。研究领域:智能信号与信息处理。
刘 婷(1981-),女,博士,副教授。研究领域:智能信号与信息处理。

张星月(2004-),女,本科生。研究领域:智能信号与信息处理。
董 祺(2004-),女,本科生。研究领域:智能信号与信息处理。
李乐乐(2002-),女,本科生。研究领域:智能信号与信息处理。
刘嘉豪(2002-),男,本科生。研究领域:智能信号与信息处理。