

# 基于 LBP 和注意力机制的改进 VGG 网络的人脸表情识别方法

张中华<sup>1</sup>, 杨慧炯<sup>2</sup>

(1.太原师范学院计算机科学与技术学院, 山西 晋中 030619;

2.太原工业学院, 山西 太原 030008)

✉ 1940219034@qq.com; yanghj@tit.edu.cn



**摘要:**为了提高训练速度和人脸表情识别效果,提出一种基于局部二值模式(Local Binary Pattern, LBP)和注意力机制的改进视觉几何群网络(Visual Geometry Group Network, VGG 网络)的人脸表情识别方法。首先,通过 LBP 获取数据集的纹理特征。其次,利用全局平均池化层代替全连接层,并在基准模型卷积层后和全局平均池化层前引入注意力模块,创建新网络模型 NEW-VGG;通过对 NEW-VGG 做消融实验,验证模型改进的正确性。最后,融合 LBP+NEW-VGG 模型对 CK+ 和 Fer2013 两种数据集进行 10 倍交叉验证,取得了 97.98% 和 76.75% 的识别率。实验结果表明,该方法不仅能加快网络训练迭代速度,增强人脸表情识别效果,还具有较强的鲁棒性。

**关键词:**面部表情识别;局部二值模式;注意力机制

**中图分类号:**TP391 **文献标志码:**A

## Improved VGG Network for Facial Expression Recognition Method Based on LBP and Attention Mechanism

ZHANG Zhonghua<sup>1</sup>, YANG Huijiong<sup>2</sup>

(1.School of Computer Science and Technology, TaiYuan Normal University, Jinzhong 030619, China;

2.Taiyuan Institute of Technology, Taiyuan 030008, China)

✉ 1940219034@qq.com; yanghj@tit.edu.cn

**Abstract:** In order to improve training speed and facial expression recognition performance, this paper proposes an improved VGG network (Visual Geometry Group Network) based on LBP (Local Binary Pattern) and attention mechanism for facial expression recognition. Firstly, the texture features of the dataset are obtained through LBP. Secondly, the fully connected layer is replaced with the global average pooling layer and an attention module is introduced after the benchmark model convolution layer and before the global average pooling layer, so that a new network model NEW-VGG is created. Ablation experiment on NEW-VGG verifies the correctness of the model improvement. Finally, the LBP+NEW-VGG fusion model is subjected to 10-fold cross validation on the CK+ and Fer2013 datasets, achieving recognition rates of 97.98% and 76.75%. The experimental results show that the proposed method not only accelerates the iteration speed of network training and enhances the facial expression recognition effect, but also has strong robustness.

**Key words:** facial expression recognition; Local Binary Pattern; attention mechanism

## 0 引言(Introduction)

面部表情是判断人类情感和人机交互效果的重要依据,早在 20 世纪, EKMAN 等<sup>[1]</sup>系统地 将面部表情划分为生气、害

怕、厌恶、开心、悲伤、惊讶六类,这一理论的提出在表情识别领域具有跨时代的意义,同时开启了表情识别领域新世界的大门。

目前,表情识别的方法有传统算法和深度学习算法两大类。常见的传统表情识别算法有局部二值模式(LBP)<sup>[2]</sup>、Gabor小波变化法<sup>[3]</sup>和尺度不变特征变换<sup>[4-5]</sup>等。深度学习算法主要是利用神经网络模型实现自动特征提取和面部表情分类。常见的网络模型有卷积神经网络、深度置信网络和堆叠式自动编码器等<sup>[6]</sup>。

上述算法虽然在当时取得了良好的效果,但是存在一些问题,例如提取的特征单一化、模型参数过大、训练时间过长及识别精确度低等。为了避免追求高准确率而不断增加网络模型深度或不断对面部表情数据集进行扩充而将问题复杂化,本文不再对实验数据集样本进行扩充,而是对VGG网络模型进行改建,创建NEW-VGG模型,以此加快模型的训练速度,通过传统算法与深度学习算法相结合的方法,将LBP算法与NEW-VGG模型进行融合,并对两种具有代表性的数据集CK+和Fer2013进行验证,本文所提方法不仅降低了样本的训练时间,还提高了表情识别的准确率。

## 1 基于LBP和注意力机制的改进VGG模型 (Improved VGG model based on LBP and attention mechanism)

本文首先对原始数据集进行图片归一化处理,以消除原始数据集中与表情识别无关的干扰因素;其次使用LBP算法对已处理数据集图片的纹理特征进行提取,同时对VGG网络模型进行改建,创建了新的神经网络模型NEW-VGG,并将LBP纹理特征与NEW-VGG网络进行级联;最后通过Softmax分类器对面部表情进行分类。本文表情识别总体流程图如图1所示。

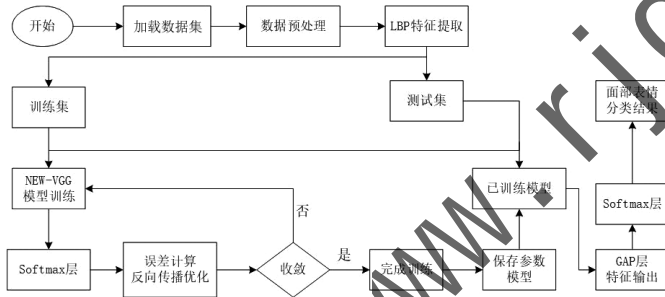


图1 表情识别总体流程图

Fig. 1 Overall flowchart of facial expression recognition

### 1.1 LBP特征提取

LBP是一种用来描述图像局部纹理特征的算子,与SIFT(Scale-Invariant Feature Transform)和HOG(Histogram of Oriented Gradient)算法的作用类似,都可以对图片的局部纹理特征进行提取,对于原始的LBP算法来说,其最大的缺点在于它只能覆盖一个固定的小区域,这并不能满足不同尺寸和不同频率纹理的需求。继原始LBP算法之后又出现了许多改进的LBP算法,例如LBP等价模式、多尺度LBP(Multiscale Block LBP)和CS-LBP(Center-Symmetric LBP)等,本文通过比较后决定采用CircularLBP算法,因为其具有旋转不变性、灰度不变性及光照变化不敏感等显著的优点。利用LBP算法对归一化的数据集进行纹理特征的提取,通过多次实验,对比LBP( $R=1, P=8$ )、LBP( $R=3, P=8$ )和LBP( $R=1, P=6$ )不同参数下表情识别的效果,最终采用半径 $R=1$ 和采样点 $P=6$ 时的最

佳结果。不同参数下的特征提取效果图如图2所示。

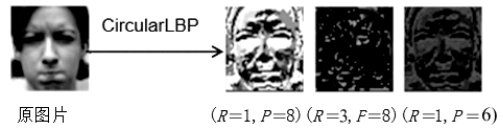


图2 不同参数下的特征提取效果图

Fig. 2 Effect image of feature extraction under different parameters

### 1.2 改进后的VGG网络

为快速、准确地对面部表情实现分类,需要对VGG-16模型进行一定的改进,改进后的模型NEW-VGG如图3所示。众所周知,在VGG-16网络模型中,第一个全连接层FC1有4096个节点,上一层POOL2有 $7 \times 7 \times 512 = 25088$ 个节点,则该传输需要 $4096 \times 25088$ 个权值,需要消耗很大的内存。LIN等<sup>[7]</sup>设计了一个维度自适应全局平均池化(Global Average Pooling, GAP)层取代传统CNN中的全连接层,用于融合学习到的深度特征。借鉴LIN等<sup>[7]</sup>提出的方法,在NEW-VGG模型中利用自适应全局平均池化层(GAP)取代传统VGG-16中的全连接层,解决全连接层容易造成的过拟合、参数超多及模型臃肿等问题,极大地降低了模型的训练速度。此外,在VGG-16模型的第一层卷积层后和全局平均池化层前加入注意力机制模块,可以使模型更多地关注面部表情中一些重要的特征。在本文模型的每个卷积层之后还添加了BN层、ReLU激活函数和 $2 \times 2$ 最大池化层。BN层和ReLU激活函数可以使卷积层线性输出的数据以非线性化的形式来表达,避免梯度爆炸和弥漫的问题。Softmax层则是进行表情图片的输出分类。

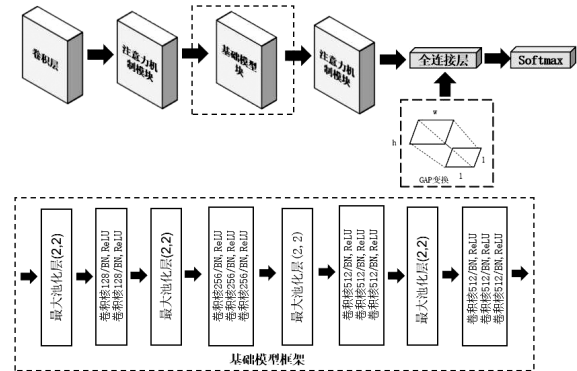


图3 改进后的NEW-VGG网络模型

Fig. 3 Improved VGG network model

### 1.3 注意力机制模块

在NEW-VGG网络中,为了达到快速准确的分类效果,采用增加不同区域特征的关注程度,加入注意力模块的方法,即引入注意力机制,这已经被证明在像素级计算机视觉任务中能取得不错的效果,可以使模型更多地关注面部表情一些重要的特征。

本文所提注意力机制模块采用混合注意力机制(CBAM)<sup>[8]</sup>,CBAM模块会沿着通道和空间两个维度依次获取注意力特征图,然后将注意力特征图与输入特征图相乘,对特征进行自适应修饰。相较于只具有通道注意力机制的

SENet<sup>[9]</sup>来说, CBAM 模块可以取得更好的效果。同时, CBAM 模块还是轻量级的, 在保证网络模型速度稳定的情况下, 还可以提高对面部表情的识别率。混合注意力机制(CBAM)网络框架图如图 4 所示。

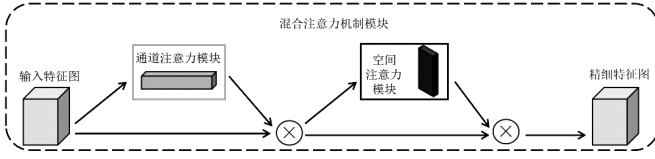


图 4 CBAM 网络框架图

Fig. 4 CBAM network framework

对于本文来说, LBP 提取的纹理特征经过一次卷积后得到在空间维度上的特征图  $F$  会分别沿着全局最大池化层(Maxpool)和全局平均池化层(Avgpool)两个方向进行压缩得到空间描述特征  $F_{cmax}$  和  $F_{cavg}$ , 从而提高特征图  $F$  经过通道注意力模块的计算效率; 空间描述特征  $F_{cmax}$  和  $F_{cavg}$  会经过一个共享网络 MLP, 将两个特征相加后经公式(1)得到通道注意力权重系数  $W_c(F)$ 。

$$W_c(F) = \sigma(MLP(F_{cmax}) + MLP(F_{cavg})) \quad (1)$$

$$F1 = W_c(F) \otimes F \quad (2)$$

其中:  $\sigma$  为 sigmoid 操作, 通过公式(2)将通道注意力权重系数  $W_c(F)$  乘以输入特征图  $F$  生成通道注意力特征图  $F1$ , 通道注意力模块网络框架如图 5 所示。

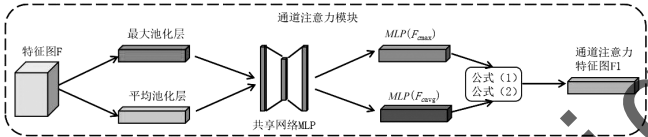


图 5 通道注意力模块网络框架

Fig. 5 Network framework of channel attention module

空间注意力模块的使用是对通道注意力模块的补充, 其模块内部操作和通道注意力模块类似。首先将通道注意力机制模块得到的特征图  $F1$  作为其输入特征图。其次输入特征图  $F1$  经过最大池化层(Maxpool)和平均池化层(Avgpool)在通道维度上聚合信息, 得到特征描述符  $F_{smax}$  和  $F_{savg}$ , 将特征描述符  $F_{smax}$  和  $F_{savg}$  横向拼接后传入一个  $7 \times 7$  的卷积核中进行卷积操作; 通过公式(3)得到空间注意力权重系数  $W_s(F)$ 。

$$W_s(F) = \sigma(f^{7 \times 7}([F_{smax}; F_{savg}])) \quad (3)$$

$$F2 = W_s(F) \otimes F1 \quad (4)$$

其中:  $\sigma$  为 sigmoid 操作, 通过公式(4)将空间注意力权重系数  $W_s(F)$  乘以输入特征图  $F1$  生成空间注意力特征图  $F2$ , 这样就可以使输入的特征图变成更为精细的特征图, 从而进一步增强 NEW-VGG 网络的鲁棒性。空间注意力模块网络框架如图 6 所示。

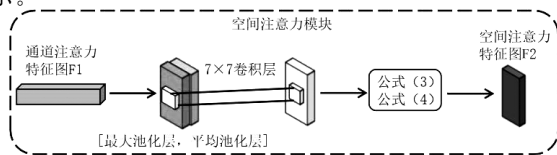


图 6 空间注意力模块网络框架

Fig. 6 Network framework of spatial attention module

## 2 实验(Experimental)

### 2.1 实验条件

本文实验的开发环境是基于 Python 语言的 PyTorch 框架, 实验软硬件平台为 64 位 Linux 操作系统, CPU 为第三代 Intel Xeon Gold 系列, 主频为 2.2 GHz, 内存为 32 GB, GPU 型号为 NVIDIA RTX A4000, 显存是 16 GB。本文选取了两种著名的人脸表情数据集 CK+ 数据集、Fer2013 数据集。两种数据集在样本数量、参与者的种族和参与者的年龄上具有较大差异, 甚至数据集中还有些图片并不是关于人脸表情的。本文选取这两种数据集可以确保网络模型的可靠性。本文实验方法采用消融实验法, 探索全局平均池化层和注意力模块对 VGG-16 网络的影响。本文验证方法采用 10 倍交叉验证法, 将预处理好的 CK+ 和 Fer2013 数据集分为 10 份进行实验以测试算法的准确性, 其中 9 份作为训练数据, 1 份作为测试数据。

由于 CK+ 和 Fer2013 两种实验数据集样本数量存在较大差异, 本文对两种数据集设置了不同的网络模型参数(表 1)。

表 1 两种数据集的网络模型参数

Tab.1 Network model parameters for two datasets

| 数据集     | 样本数量/张 | 学习率  | 测试迭代次数/次 | 批次数量/张 | 最大迭代次数/次 |
|---------|--------|------|----------|--------|----------|
| CK+     | 593    | 0.01 | 150      | 128    | 100      |
| Fer2013 | 35 887 | 0.01 | 300      | 64     | 200      |

### 2.2 实验结果和分析

为了验证本文改进方法的正确性, 对其进行消融实验, 并在 CK+ 数据集上做多次测试, 选取平均识别率作为最终的测试结果(表 2)。由表 2 中的数据可知, 融合了注意力机制且用平均池化层替代全连接层的 NEW-VGG 模型, 比基准模型的识别准确率高。同时还发现 NEW-VGG 在 CK+ 数据集上训练迭代一次的平均时间为 0.1 s, 而基准模型训练迭代一次的平均时间为 0.2 s, 对比后发现 NEW-VGG 模型训练速度更快, LBP 和 NEW-VGG 融合后的识别率进一步提高, 达到较好的识别效果。

表 2 CK+数据集上模型的识别率

Tab.2 Model recognition rate on CK+ dataset

| 模型方法        | 识别率/% |
|-------------|-------|
| Basic-VGG   | 95.46 |
| NEW-VGG     | 96.86 |
| LBP+NEW-VGG | 97.98 |

本文比较了本文所提方法和文献[10][该文献主要探究准映射网络(AMN)在 Fer2013 数据集的性能表现]、文献[11][该文献本文首先综述了目前最流行和最先进的降维方法, 然后提出了一种新的、更有效的流形学习方法—软局部保持映射(SLPM)并探究其在 CK+ 和 Fer2013 数据集上的表现]、文献[12][该文献为了进行分类, 使用了基于样本和原型分布的库尔巴克判别法的方法]、文献[13][该文献首先使用卷积神经网络(CNN)从 VGG\_Faces 中学习面部特征, 然后将其链接到长

短期记忆网络,以利用视频帧之间的时间关系,报告了CK+面部表情数据集中的竞争结果]、文献[14][该文献提出了一种面部姿势生成对抗网络(FaPE-GAN),用于合成新的面部表情图像以增强数据集并应用于训练目的,然后学习基于LightCNN的Fa-Net模型进行表情分类]在CK+和Fer2013数据集上的准确度(表3、表4),表3和表4的结果表明,本文所提方法的识别率明显高于其他方法。

表3 CK+数据集上不同方法的比较

Tab.3 Comparison of different methods on the CK+ dataset

| 方法                                     | 识别率/% |
|----------------------------------------|-------|
| 融合对齐和非对齐人脸信息实现野外情感自动识别 <sup>[10]</sup> | 96.92 |
| 局部保持映射(SLPM) <sup>[11]</sup>           | 96.12 |
| 纹理测度与基于特征分布的分类 <sup>[12]</sup>         | 97.25 |
| 利用长短期记忆网络进行面部表情分类 <sup>[13]</sup>      | 96.32 |
| 本文方法                                   | 97.98 |

表4 Fer2013数据集上不同方法的比较

Tab.4 Comparison of different methods on the Fer2013 dataset

| 方法                                          | 识别率/% |
|---------------------------------------------|-------|
| 融合对齐和非对齐人脸信息实现野外情感自动识别 <sup>[10]</sup>      | 74.32 |
| 局部保持映射(SLPM) <sup>[11]</sup>                | 72.25 |
| 纹理测度与基于特征分布的分类 <sup>[12]</sup>              | 75.32 |
| 基于面部姿势生成对抗网络(FaPE-GAN)的表情识别 <sup>[14]</sup> | 71.21 |
| 本文方法                                        | 76.75 |

本文通过绘制混淆矩阵探究每个表情在不同数据集上的识别效果。CK+混淆矩阵中对开心、惊讶和蔑视3种表情的识别率最高,因为人们在做出这3种表情时,其面部显露特点比较明显,不易与其他表情混淆;然而,生气、憎恶、恐惧和悲伤4种表情的识别率相对较低,这是因为人在消极的情况下表露的表情的特点较为相似,这也给网络模型区分这些表情带来了一定的困难。本文所提方法在CK+数据集和Fer2013数据集上的混淆矩阵分别见表5和表6。Fer2013数据集上的混淆矩阵中只有开心和惊讶2种表情的识别率最高,其他5种表情的识别率较低。主要原因在于Fer2013数据集标签本身存在错误且受干扰因素也比较多,这也说明了原始数据集样本在实验过程中的重要性。

表5 本文所提方法在CK+数据集上的混淆矩阵

Tab.5 The confusion matrix of the proposed method on the CK+ dataset

| 表情 | 生气   | 憎恶   | 恐惧   | 开心   | 悲伤   | 惊讶   | 蔑视   |
|----|------|------|------|------|------|------|------|
| 生气 | 0.93 | 0.00 | 0.01 | 0.00 | 0.04 | 0.02 | 0.00 |
| 憎恶 | 0.03 | 0.95 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 |
| 恐惧 | 0.00 | 0.03 | 0.90 | 0.04 | 0.00 | 0.02 | 0.01 |
| 开心 | 0.00 | 0.00 | 0.00 | 0.99 | 0.00 | 0.01 | 0.00 |
| 悲伤 | 0.06 | 0.00 | 0.02 | 0.00 | 0.92 | 0.00 | 0.00 |
| 惊讶 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.99 | 0.00 |
| 蔑视 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |

表6 本文所提方法在Fer2013数据集上的混淆矩阵

Tab.6 The confusion matrix of the proposed method on the Fer2013 dataset

| 表情 | 生气   | 憎恶   | 恐惧   | 开心   | 悲伤   | 惊讶   | 中性   |
|----|------|------|------|------|------|------|------|
| 生气 | 0.67 | 0.01 | 0.09 | 0.02 | 0.13 | 0.01 | 0.07 |
| 憎恶 | 0.18 | 0.73 | 0.04 | 0.02 | 0.00 | 0.02 | 0.01 |
| 恐惧 | 0.10 | 0.00 | 0.67 | 0.03 | 0.12 | 0.02 | 0.06 |
| 开心 | 0.01 | 0.01 | 0.00 | 0.91 | 0.03 | 0.02 | 0.02 |
| 悲伤 | 0.07 | 0.00 | 0.10 | 0.04 | 0.64 | 0.01 | 0.14 |
| 惊讶 | 0.01 | 0.00 | 0.10 | 0.03 | 0.03 | 0.81 | 0.02 |
| 中性 | 0.04 | 0.00 | 0.05 | 0.06 | 0.14 | 0.01 | 0.70 |

3 结论(Conclusion)

本文提出了一种融合LBP特征和注意力机制的面部表情识别方法,通过对VGG-16网络进行改进,创建出了NEW-VGG模型,并通过消融实验验证了全局平均池化层和注意力机制对VGG-16网络模型在速度、精度提升上的正确性。将LBP算法和NEW-VGG模型进行融合,并对CK+和Fer2013两种著名人脸表情数据集进行实验,实验结果表明,该模型在CK+和Fer2013数据集上分别取得了97.98%和76.75%的准确率,经与其他文献识别率对比可以发现,本文方法明显高于文献[10]至文献[14]的表情识别方法。但是,从实验中绘制的数据集混淆矩阵可以发现,本文所提方法针对生气、恐惧和悲伤表情的识别效果不尽如人意。在接下来的工作中,应考虑进一步探索用更好的深度学习方法增强网络,通过增强数据特征的提取,以便更加快速、准确地区分易混淆的面部表情。

参考文献(References)

[1] EKMAN P, FRIESEN W V. Facial action coding system: a technique for the measurement of facial movement[J]. Palo alto, 1978, 3(2): 5.

[2] 蒋小平, 刘俊威, 王乐乐, 等. 基于多尺度卷积神经网络和LBP算法的浮选工况识别[J]. 矿业科学学报, 2023, 8(2): 202-212.

[3] 刘胜昔, 程春玲. 改进的Gabor小波变换特征提取算法[J]. 计算机应用研究, 2020, 37(2): 606-610.

[4] 邹富标, 李艳. 基于SIFT特征提取的受电弓滑板磨损检测[J]. 北京工业职业技术学院学报, 2023, 22(1): 11-14.

[5] 黄奕玮. 基于深度学习的面部表情识别研究与实现[D]. 北京: 北京邮电大学, 2019.

[6] 刘锦峰, 黄江峰. 基于深度学习的面部表情识别研究[J]. 现代信息科技, 2022, 6(1): 86-88, 92.

[7] LIN M, CHEN Q, YAN S C. Network in network[DB/OL]. (2014-03-04)[2023-05-20]. <https://arxiv.org/abs/1312.4400>.

[8] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Springer. Proceedings of the European Conference on Computer Vision. Cham: Springer, 2018: 3-19.