

# 基于轻量卷积网络多层特征融合的人脸表情识别

申毫<sup>1,2,3</sup>, 孟庆浩<sup>1,2,3</sup>, 刘胤伯<sup>1,2,3\*</sup>

<sup>1</sup>天津大学电气自动化与信息工程学院, 天津 300072;

<sup>2</sup>天津大学机器人与自主系统研究所, 天津 300072;

<sup>3</sup>天津市过程检测与控制重点实验室, 天津 300072

**摘要** 基于深度学习的表情识别方法存在参数量大、实时性差等问题, 针对此问题, 提出一种基于轻量卷积网络的多层特征融合的人脸表情识别方法。首先使用改进的倒置残差网络为基本单元搭建轻量卷积网络模型, 然后采用池化、 $1 \times 1$  卷积、全局平均池化法筛选卷积网络中的浅层特征, 并对这些筛选的浅层特征与深层特征进行融合用于表情识别。在两个常用的真实表情数据集 RAF-DB 和 AffectNet 上对所提方法进行测试, 识别准确率分别达 85.49% 和 57.70%, 且模型参数量仅有  $0.2 \times 10^6$ 。

**关键词** 图像处理; 表情识别; 卷积神经网络; 浅层特征; 深层特征; 多层特征融合

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP202158.0610005

## Facial Expression Recognition by Merging Multilayer Features of Lightweight Convolutional Networks

Shen Hao<sup>1,2,3</sup>, Meng Qinghao<sup>1,2,3</sup>, Liu Yinbo<sup>1,2,3\*</sup>

<sup>1</sup>School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China;

<sup>2</sup>Institute of Robotics and Autonomous Systems, Tianjin University, Tianjin 300072, China;

<sup>3</sup>Tianjin Key Laboratory of Process Detection and Control, Tianjin 300072, China

**Abstract** Current methods of expression recognition based on deep learning exhibit problems such as a large number of parameters and poor real-time performance. To resolve these problems, this paper proposes a facial expression recognition method by merging the multilayer features of a lightweight convolutional network. First, the improved inverted residual network was considered as the basic unit for building the lightweight convolutional network model. Then, the shallow features associated with the convolutional network subjected to the methods of pooling,  $1 \times 1$  convolution, and global average pooling were merged with the deep features of the network for expression recognition tasks. The proposed method was verified by employing RAF-DB and AffectNet, which are commonly used expression datasets. Results indicate that the facial expression recognition accuracies of the proposed method when considering the RAF-DB and AffectNet datasets are 85.49% and 57.70%, respectively. Furthermore, the number of model parameters is only  $0.2 \times 10^6$ .

**Key words** image processing; expression recognition; convolutional neural network; shallow feature; deep feature; multi-layer feature fusion

**OCIS codes** 100.3008; 100.4996; 100.2000

收稿日期: 2020-06-30; 修回日期: 2020-08-14; 录用日期: 2020-08-25

基金项目: 国家重点研发计划(2017YFC0306200)、国家自然科学基金(61573253)

\* E-mail: liuyinbo@tju.edu.cn

# 1 引 言

人脸表情是反映人类情感的重要信息,心理学家研究发现,在人类交流活动中人脸表情传递的情感信息占总体信息的 55%<sup>[1]</sup>。因此,通过人脸表情获得人类情绪信息的面部表情识别(FER)技术在服务机器人、驾驶员疲劳监测、医疗服务等需要人机交互的系统中具有重要应用价值,是当前的研究热点之一<sup>[2-4]</sup>。

FER 方法大致可分为两类,即传统方法和基于深度学习的方法。传统方法包括特征提取和特征分类两个步骤:提取的特征可以是局部二值模式(LBP)<sup>[5]</sup>、方向梯度直方图(HOG)<sup>[6]</sup>及小波特征<sup>[7]</sup>等;特征分类方法主要有 K 近邻法<sup>[8]</sup>、支持向量机<sup>[9]</sup>、Adaboost 分类器<sup>[10]</sup>及神经网络等。传统方法虽然在实验室的摆拍数据集上表现良好,但在真实复杂多变的场景下效果欠佳。

近几年,随着深度学习技术的兴起,特别是卷积神经网络(CNN)在计算机视觉领域的广泛应用,许多传统视觉领域的研究开始采用深度学习方法。FER 作为一个经典的模式识别问题,也有大量关于深度学习方面的研究出现。Yu 等<sup>[11-12]</sup>均采用集成多个 CNN 模型的方法,提升了 FER 模型的性能;Li 等<sup>[13]</sup>提出一种 Boosted-CNN 方法,该方法先训练一个 CNN 模型,在此基础上,采用随机抽样方法对每个表情进行不平衡学习,获得了较好的效果;Mollahosseini 等<sup>[14]</sup>通过向网络中加入 Inception 模块来提升 FER 模型的学习能力,在多个数据集下实现了较高的识别率;张红颖等<sup>[15]</sup>提出了一种双通道卷积神经网络,一个通道输入 LBP 图,另一个通道输入梯度图,将两种特征结合进行表情识别,在几个数据集上的测试结果表明提升了识别准确率。上述几个方法为提高 FER 模型的准确率,均采用了复杂的网络结构,识别速度有限。为了提高识别速度,Arriaga 等<sup>[16]</sup>结合残差模块和深度可分离卷积,设计了一个精简的网络结构,该网络可以实时地识别表情,但是结构过于简单,在 FER2013 数据集上只能达到基线准确率;吕海等<sup>[17]</sup>先采用人脸分割网络,得到与表情识别最相关的区域,再通过 FER 网络对分割后的图像进行识别,两个网络均采用轻量级网络,实现了不错的识别准确率和识别速度。但该方法较为复杂,首先要在表情数据集上构建一个人脸分割的数据集,然后训练人脸分割网络和表情

识别网络,步骤较为繁琐。综上所述,现有的基于深度学习的 FER 方法仍无法实现识别准确率与识别速度的平衡。

为了平衡识别准确率和识别速度,本文提出了一种基于 CNN 多层次特征融合的轻量级 FER 方法。使用改进的倒置残差网络为基本单元,搭建轻量级卷积网络模型,并在此基础上对 CNN 的深层信息和浅层信息进行融合,以提高表情识别精度;此外,针对使用全连接层直接进行深层和浅层特征融合会产生大量参数的问题,采用先对卷积网络中的几个浅层特征进行筛选,再与深层特征进行融合的方法。在两个常用表情数据集上进行实验,验证了所提方法的有效性。

## 2 基本原理

在 CNN 模型中,不同卷积层由于卷积深度不同,浅层的特征图中相对感受野较小,对局部的特征信息比较敏感,深层特征图中相对感受野较大,对整体的轮廓信息比较敏感。文献[18-19]中的研究采用全连接层对 CNN 模型深浅层的特征进行融合,虽然提升了模型性能,但也使模型参数大量增加。为减少融合所产生的参数量,首先对浅层特征进行筛选,然后再融合卷积网络深浅层特征。

### 2.1 总体网络架构设计

设计的融合多层次特征的卷积网络整体架构如图 1 所示。

该网络模型称为 ms\_model\_v1,输入图像首先经过尺寸为  $3 \times 3$  的普通卷积(Conv2D 1)进行初步的特征提取,后连接 10 个改进的倒置残差单元(Bottleneck\_M),以更轻量化的方式进一步提取特征;然后对特定几个中间层的特征图进行筛选,得到更精炼的中间层特征,并通过特征合并操作(concat)对这些特征与卷积网络主干输出的特征进行融合,达到深浅层特征融合的目的,提升模型性能;再经变形操作(reshape 1)将融合后特征的尺寸重新变为  $1 \times 1$ ,使用丢弃操作,以一定概率丢弃部分神经元,防止过拟合,使用  $1 \times 1$  卷积(Conv2D 3)将特征图个数变为与表情类别数  $k$  相同;最后经 Softmax 函数和变形操作(reshape 2)输出关于表情类别的概率向量。具体参数设置如表 1 所示,其中  $c$  为卷积核数, $s$  为步长, $t$  为 Bottleneck\_M 单元的扩张倍数。

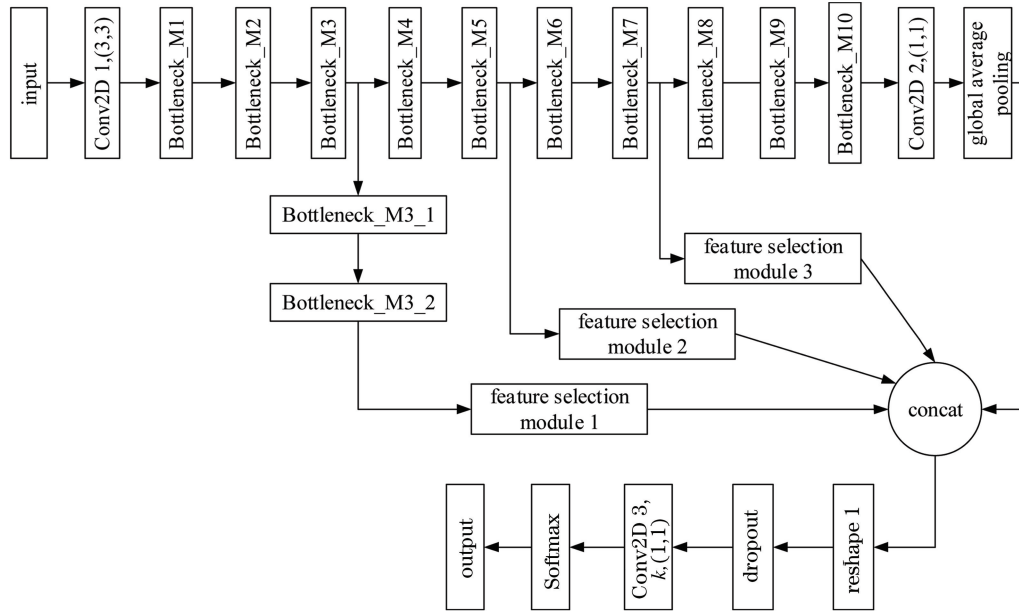


图 1 ms\_model\_v1 模型结构图  
Fig. 1 Structure of ms\_model\_v1 model

表 1 CNN 参数

Table 1 Parameters of the CNN

Layer name	<i>c</i>	<i>s</i>	<i>t</i>	Output size
Conv2D 1 (3×3)	16	2		56×56×16
Bottleneck_M1	16	1	1	56×56×16
Bottleneck_M2	24	2	5	28×28×24
Bottleneck_M3	24	1	5	28×28×24
Bottleneck_M3_1	32	1	5	28×28×32
Bottleneck_M3_2	32	1	5	28×28×32
Feature selection module 1				32
Bottleneck_M4	32	2	5	14×14×32
Bottleneck_M5	32	1	5	14×14×32
Feature selection module 2				32
Bottleneck_M6	40	1	5	14×14×40
Bottleneck_M7	40	1	5	14×14×40
Feature selection module 3				40
Bottleneck_M8	40	1	5	14×14×40
Bottleneck_M9	48	2	5	7×7×48
Bottleneck_M10	64	1	5	7×7×64
Conv2D 2 (1×1)	64	1		7×7×64
Global average pooling				64
Concat				168
Reshape 1				1×1×168
Dropout				1×1×168
Conv2D 3 (1×1)	<i>k</i>			1×1× <i>k</i>
Softmax				1×1× <i>k</i>
Reshape 2				<i>k</i>

### 2.2 改进的倒置残差单元 Bottleneck\_M

Bottleneck\_M 单元参考 MobileNetV2<sup>[20]</sup> 中的倒置残差结构。倒置残差结构的基本原理是先对特

征图进行通道扩张,在特征图通道数较多的情况下使用可分离卷积提取特征,最后再对特征图进行通道压缩。实验表明,在特征图通道数较多的情况下,

倒置残差结构有利于可分离卷积提取图像特征,且由于使用  $1 \times 1$  卷积进行特征图的通道变化,所以该结构参数量也较少。

对倒置残差结构的改进主要在于,使用 Mish 激活函数<sup>[21]</sup>替代原来结构中的 ReLU6 激活函数。相较 ReLU6, Mish 激活函数更加平滑且对负值有轻微允许,对卷积网络参数的更新更容易。

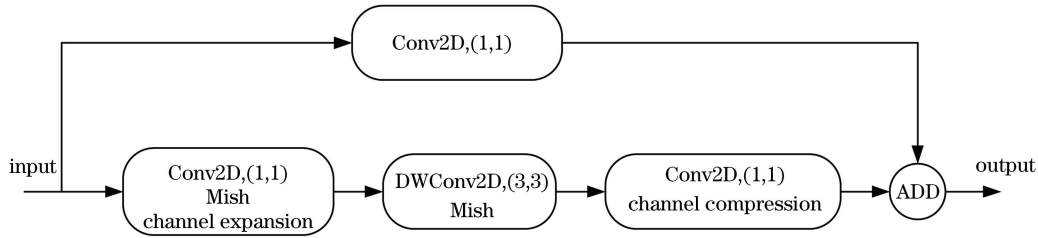


图 2 Bottleneck\_M 结构

Fig. 2 Bottleneck\_M structure

在表 1 中, Bottleneck\_M 单元的  $c$  为具有通道压缩作用的  $1 \times 1$  卷积的卷积核数,  $t$  为具有通道扩张作用的  $1 \times 1$  卷积对输入特征图的扩张倍数。

图 2 为改进后的倒置残差网络结构图。该结构中,主干部分输入特征图首先经  $1 \times 1$  的卷积进行通道扩张,然后用  $3 \times 3$  的深度可分离卷积提取特征,最后用  $1 \times 1$  的卷积压缩通道;分支部分输入特征图经  $1 \times 1$  卷积操作后直接输出,保证与主干特征图通道数相同,在输出端对两部分输出进行叠加。

### 2.3 多层特征融合设计

对中间层输出的特征图进行筛选的特征选择模块如图 3 所示。

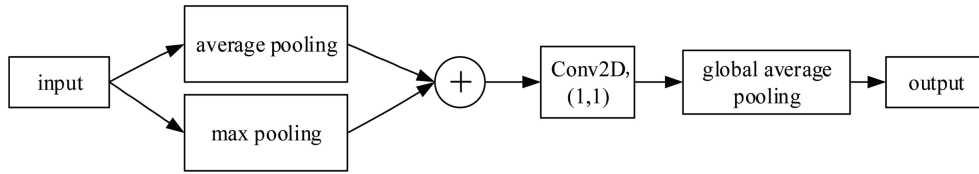


图 3 特征选择模块的流程

Fig. 3 Process of feature selection module

Bottleneck\_M3 输出的特征图尺寸为  $28 \times 28 \times 24$ 。输入图先经 2 个 Bottleneck\_M 模块,增加特征图深度,特征图尺寸变为  $28 \times 28 \times 32$ ;然后使用池化核为  $4 \times 4$ 、步长为 4 的最大池化和平均池化,得到两个  $7 \times 7 \times 32$  的特征图,对两个特征图进行元素叠加,得到一个  $7 \times 7 \times 32$  的特征图;再进行  $1 \times 1$  卷积、全局平均池化,得到 32 维的特征向量。对 Bottleneck\_M5 (特征图尺寸为  $14 \times 14 \times 32$ ) 和 Bottleneck\_M7 (特征图尺寸为  $14 \times 14 \times 40$ ) 输出的特征图,不需要增加深度,直接使用池化核为  $2 \times 2$ 、步长为 2 的最大池化和平均池化操作,然后进行元素叠加、 $1 \times 1$  卷积及全局平均池化操作,分别得到 32 维和 40 维的特征向量。经特征筛选后,尺寸为  $28 \times 28 \times 24$ 、 $14 \times 14 \times 32$  及  $14 \times 14 \times 40$  的特征图分别变为 32 维、32 维及 40 维的特征向量。

多层特征融合设计如图 1 所示,设 Bottleneck\_M3、\_M5、\_M7 及 \_M10 的输出特征图分别为  $\mathbf{X}_3$ 、 $\mathbf{X}_5$ 、 $\mathbf{X}_7$  及  $\mathbf{X}_{10}$ ,主干网络的输出为  $\mathbf{X}_{10}$  (尺寸为  $7 \times 7 \times 64$ ),经  $1 \times 1$  卷积和全局平均池化后得到的特

征向量为  $\mathbf{O}_{10}$  (64 维向量),特征筛选操作 (feature selection module) 设为  $\mathbf{G}$ ,则经筛选后,特征图变为

$$\mathbf{O}_1 = \mathbf{G}_1(\mathbf{X}_3), \quad (1)$$

$$\mathbf{O}_2 = \mathbf{G}_2(\mathbf{X}_5), \quad (2)$$

$$\mathbf{O}_3 = \mathbf{G}_3(\mathbf{X}_7), \quad (3)$$

式中:  $\mathbf{O}_1$  和  $\mathbf{O}_2$  均为 32 维向量,  $\mathbf{O}_3$  为 40 维向量。对这些筛选后的特征进行融合,则

$$\mathbf{O}_{\text{fusion}} = [\mathbf{O}_1, \mathbf{O}_2, \mathbf{O}_3, \mathbf{O}_{10}], \quad (4)$$

式中:  $[\cdot]$  为特征向量合并操作;  $\mathbf{O}_{\text{fusion}}$  为 168 维向量。

如果使用全连接层进行深浅层特征融合,需要首先将特征图  $\mathbf{X}_3$  (尺寸为  $28 \times 28 \times 24$ )、 $\mathbf{X}_5$  (尺寸为  $14 \times 14 \times 32$ )、 $\mathbf{X}_7$  (尺寸为  $14 \times 14 \times 40$ ) 及  $\mathbf{X}_{10}$  (尺寸为  $7 \times 7 \times 64$ ) 展平为向量,然后拼接这四个向量,最终得到融合特征向量的维度为  $28 \times 28 \times 24 + 14 \times 14 \times 32 + 14 \times 14 \times 40 + 7 \times 7 \times 64 = 36064$ ,这是所提 ms\_model\_v1 模型得到的特征向量维数的 214 倍。

### 2.4 Softmax 分类

使用 Softmax 函数输出表情类别概率, Softmax 函数将一个多维向量输入的每一维转换到

$[0, 1]$ 之间,表达式为

$$p_i = \frac{e^{x_i}}{\sum_{i=1}^k e^{x_i}}, i = 1, 2, \dots, k, \quad (5)$$

式中: $x_i$  为 Softmax 函数的输入; $p_i$  为第  $i$  类的概率。

采用交叉熵计算卷积网络模型的损失值,计算公式为

$$\text{loss}(\mathbf{y}, \mathbf{z}) = - \sum_{i=1}^k z_i \log(y_i), \quad (6)$$

式中: $y_i$  为输入样本的预测值; $z_i$  为输入样本的真实值; $\mathbf{y}$  和  $\mathbf{z}$  也分别为样本预测值和真实值,但为  $r$  维向量。将该交叉熵损失函数作为优化卷积网络模型的目标函数。

### 3 实验结果及分析

本次实验在 Ubuntu16.04 操作系统上进行,使用 Python 编程语言和 Keras 深度学习框架搭建卷积网络模型。计算机硬件配置为 Intel E5430 CPU、32G 内存、NVIDIA 980Ti 显卡(6G 显存)。

#### 3.1 数据集与训练细节介绍

为验证所提方法的有效性,采用 RAF-DB<sup>[22]</sup> 和

表 2 随机筛选后 AffectNet 数据集的类别数量

Table 2 Number of categories in AffectNet dataset after random screening

Class	Neutral	Happy	Sad	Surprise	Fear	Disgust	Angry	Contempt	Total
Train set	5978	5979	5966	5963	6378	3803	5979	3750	43796
Test set	500	500	500	500	500	500	500	500	4000

所提卷积网络模型训练输入为  $112 \times 112$  的人脸表情图像,批量大小为 20,训练 140 轮,采用自适应矩估计(Adam)优化器,初始学习率为 0.001,当测试集的 loss 值不再下降时,学习率进行 0.5 的衰减。在训练中,使用 Keras 在线数据增强策略进行数据扩充,主要包括对表情数据进行随机旋转、水平

表 3 不同模型的性能

Table 3 Performance of different models

Model name	Number of parameters	Model size/Mbit	Acc in RAF-DB/%	Acc in AffectNet/%
base_model_R	195159	2.9	83.64	56.93
base_model_M	195159	2.9	84.23	57.30
ms_model_fully	3299303	40.2	85.45	57.37
ms_model_v1_R	193927	3.0	84.81	57.43
ms_model_v1_M	193927	3.0	85.49	57.70

改进倒置残差结构对模型性能的影响如表 3 所示。base\_model\_R 为使用原始倒置残差结构(使用 ReLU6 激活函数)搭建的基准模型,base\_model\_M 为使用改进的倒置残差结构(使用 Mish 激活函数)搭建的基准模型。可以看到,相较 base\_model\_R,

AffectNet<sup>[23]</sup> 两个人脸表情公共数据集进行测试。

1) RAF-DB (real-world affective faces database):该数据库是一个大型的真实世界表情数据库,包含从互联网下载的 29672 个高度多样化的面部图像。该数据库包含基本表情和复合表情 2 部分。本次实验主要用到基本表情部分的数据,包含 12271 个训练样本和 3068 个测试样本,被标注为惊奇、害怕、厌恶、高兴、悲伤、生气、无表情 7 类基本表情。

2) AffectNet:该数据库包含来自互联网的超过 1000000 张图片,这些图片是通过使用情感关键词在多个搜索引擎上搜索得到的。AffectNet 中的数据被标注为无表情、高兴、悲伤、惊奇、害怕、厌恶、生气、轻视 8 类表情,包含手动标注部分和自动标注部分。手动标注部分相对而言更加可靠,所以本次实验主要使用该部分数据集,由于该数据集数据存在类别不均衡现象,厌恶、轻视两个类别数据较少,其他几个类别数据较多,所以对数据量较多的类别进行随机筛选,避免样本过度不均衡。具体训练数据和测试数据的类别数量如表 2 所示。

和垂直方向的平移、剪切变换、随机缩放、随机水平翻转等操作。

#### 3.2 实验结果和分析

本次实验测量指标为模型在测试集上的总体准确率(Acc),即分类正确的图像数量与总体数量的比值。不同模型的训练结果如表 3 所示。

base\_model\_M 在 RAF-DB 和 AffectNet 两个数据集上的准确率分别提升了 0.59 个百分点和 0.37 个百分点,证明改进的倒置残差结构的有效性。ms\_model\_v1\_R 和 ms\_model\_v1\_M 分别为在 base\_model\_R 和 base\_model\_M 基础上使用所提方法进

行深层次特征融合模型,两模型仅倒置残差单元不同,可以看到,相较 ms\_model\_v1\_R, ms\_model\_v1\_M 在 RAF-DB 和 AffectNet 两个数据集上的准确率分别提升 0.68 个百分点和 0.27 个百分点,再次证明所提改进的倒置残差结构的有效性。

深层次特征融合方法对模型性能的影响也如表 3 所示,ms\_model\_fully 为在 base\_model\_M 模型基础上使用全连接层进行深层次特征融合模型。可以看到,相较基准模型 base\_model\_M, ms\_model\_v1\_M 模型在 RAF-DB 和 AffectNet 两个数据集上的准确率分别提升了 1.26 个百分点和 0.4 个百分点,且性能与 ms\_model\_fully 模型相当,但

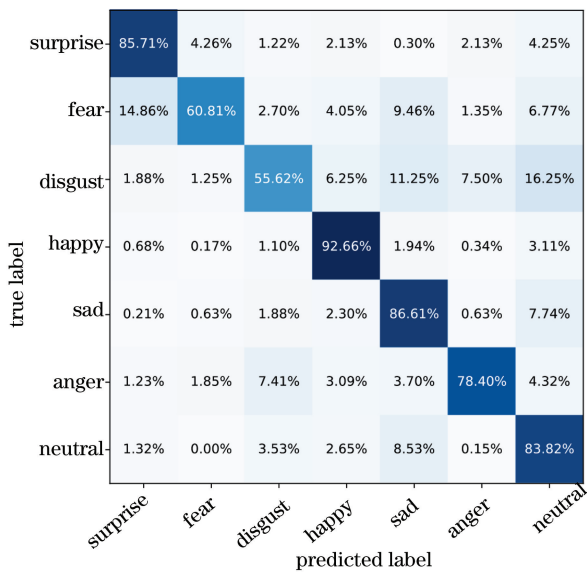


图 4 所提方法在 RAF-DB 数据集上的混淆矩阵  
Fig. 4 Confusion matrix of RAF-DB dataset using the proposed method

表 4 为各个方法在 RAF-DB 和 AffectNet 数据集上的识别准确率对比。可以看出,所提方法在两个数据集上均取得了较好的识别准确率,且由于采用轻量型的网络架构,最终模型参数量仅有  $0.2 \times 10^6$ ,较好地实现了识别准确率与识别速度的平衡。

表 4 RAF-DB 和 AffectNet 数据集上不同方法的准确率  
Table 4 Accuracy of different methods in RAF-DB and AffectNet datasets

Method	Acc in RAF-DB / %	Acc in AffectNet / %
Boosting-POOF <sup>[24]</sup>	73.19	
VGG16 <sup>[25]</sup>	80.96	51.11
DLP-CNN <sup>[22]</sup>	84.13	
pACNN <sup>[26]</sup>	83.05	55.33
gACNN <sup>[26]</sup>	85.07	58.78
E2-CapsNet <sup>[27]</sup>	85.24	
ms_model_v1_M	85.49	57.70

参数量仅为 ms\_model\_fully 的 5.9%,证明了所提方法进行深层次特征融合的有效性。

图 4、5 分别为 ms\_model\_v1\_M 模型在 RAF-DB 和 AffectNet 数据集上的混淆矩阵,对角线位置对应每个表情类别的识别准确率。可以看出:在 RAF-DB 数据集上,所提方法的整体表情识别率较高,但在“害怕”和“厌恶”两种表情上识别率较低(60%左右),而“惊讶”、“高兴”、“悲伤”及“面无表情”4 种表情上识别率较高(80%以上);在 AffectNet 数据集上,各个表情的识别率均不高,只有“高兴”这一个表情达 80%,大部分表情识别率在 50%~60%,其中“厌恶”和“轻视”在 50%左右。

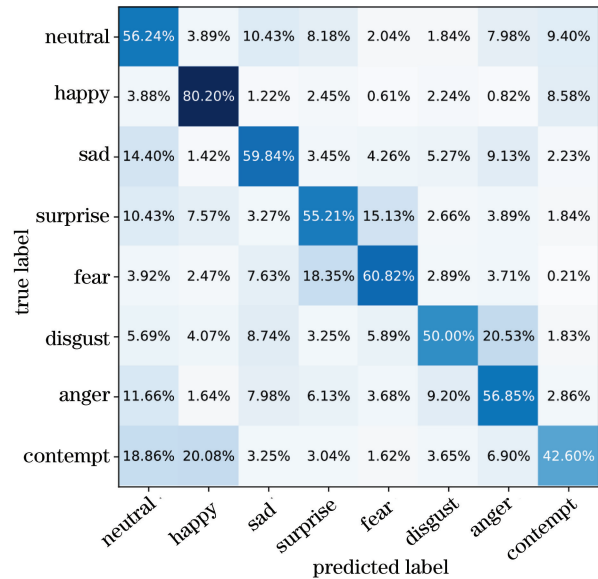


图 5 所提方法在 AffectNet 数据集上的混淆矩阵  
Fig. 5 Confusion matrix of AffectNet dataset using the proposed method

## 4 结 论

提出了一种基于多层次特征融合的轻量型卷积网络表情识别方法。采用改进的倒置残差结构设计网络模型,使模型在兼具轻量化的同时具有较好的特征提取能力,在此基础上融合卷积网络深层特征和浅层特征,进一步提升了模型准确率。在 RAF-DB 和 AffectNet 两个公开数据集上的实验结果表明,所提方法与全连接层直接融合的方法相比,性能相当,但参数量仅为后者的 6%,较好地实现了识别准确率与识别速度的平衡。下一步将对模型剪枝和权值量化进行研究。

## 参 考 文 献

[1] Mehrabian A, Russell J A. An approach to

- environmental psychology [M]. Cambridge: MIT Press, 1974: 56-63.
- [2] Huang Y X, Chen F, Lü S, et al. Facial expression recognition: a survey [J]. *Symmetry*, 2019, 11(10): 1189.
- [3] Li S, Deng W H. Deep facial expression recognition: a survey [J]. *IEEE Transactions on Affective Computing*, 2020.
- [4] Ye J H, Zhu J T, Jiang A W, et al. Facial expression recognition: a survey [J]. *Journal of Data Acquisition and Processing*, 2020, 35(1): 21-34.  
叶继华, 祝锦泰, 江爱文, 等. 人脸表情识别综述 [J]. *数据采集与处理*, 2020, 35(1): 21-34.
- [5] Shan C F, Gong S G, McOwan P W. Facial expression recognition based on local binary patterns: a comprehensive study [J]. *Image and Vision Computing*, 2009, 27(6): 803-816.
- [6] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C] // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), June 20-25, 2005, San Diego, CA, USA. New York: IEEE Press, 2005: 886-893.
- [7] Bashyal S, Venayagamoorthy G K. Recognition of facial expressions using Gabor wavelets and learning vector quantization [J]. *Engineering Applications of Artificial Intelligence*, 2008, 21(7): 1056-1064.
- [8] Dino H I, Abdulrazzaq M B. Facial expression classification based on SVM, KNN and MLP classifiers [C] // 2019 International Conference on Advanced Science and Engineering (ICOASE), April 2-4, 2019, Zakho-Duhok, Iraq. New York: IEEE Press, 2019: 70-75.
- [9] Zhong W, Huang Y L. A facial expression recognition algorithm based on feature fusion and hierarchical decision tree technology [J]. *Computer Engineering & Science*, 2017, 39(2): 393-398.  
钟伟, 黄元亮. 基于特征融合与决策树技术的表情识别方法 [J]. *计算机工程与科学*, 2017, 39(2): 393-398.
- [10] Wang Y B, Ai H Z, Wu B, et al. Real time facial expression recognition with AdaBoost [C] // Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, August 26-26, 2004, Cambridge, UK. New York: IEEE Press, 2004: 926-929.
- [11] Yu Z D, Zhang C. Image based static facial expression recognition with multiple deep network learning [C] // ICMI'15: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, November 9-13, 2015, Seattle, WA, USA. Seattle: ACM, 2015: 435-442.
- [12] Zhou T, Lü X Q, Ren G Y, et al. Facial expression classification based on ensemble convolutional neural network [J]. *Laser & Optoelectronics Progress*, 2020, 57(14): 141501.  
周涛, 吕晓琪, 任国印, 等. 基于集成卷积神经网络的面部表情分类 [J]. *激光与光电子学进展*, 2020, 57(14): 141501.
- [13] Li H, Li G M. Research on facial expression recognition based on LBP and DeepLearning [C] // 2019 International Conference on Robots & Intelligent System (ICRIS), June 15-16, 2019, Haikou, China. New York: IEEE Press, 2019: 94-97.
- [14] Mollahosseini A, Chan D, Mahoor M H. Going deeper in facial expression recognition using deep neural networks [C] // 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), March 7-10, 2016, Lake Placid, NY, USA. New York: IEEE Press, 2016: 1-10.
- [15] Zhang H Y, Wang H S. Double-channel facial expression recognition based on local binary pattern and gradient features [J]. *Laser & Optoelectronics Progress*, 2020, 57(14): 141005.  
张红颖, 王汇三. 结合局部二值模式和梯度特征的双通道表情识别 [J]. *激光与光电子学进展*, 2020, 57(14): 141005.
- [16] Arriaga O, Valdenegro-Toro M, Plöger P. Real-time convolutional neural networks for emotion and gender classification [EB/OL]. (2017-10-20) [2020-06-29]. <https://arxiv.org/abs/1710.07557>.
- [17] Lü H, Tong Q Q, Yuan Z Y. Realtime architecture for facial expression recognition in complex scenes based on face region segmentation [J]. *Computer Engineering and Applications*, 2020, 56(12): 134-140.  
吕海, 童倩倩, 袁志勇. 基于人脸分割的复杂环境下表情识别实时框架 [J]. *计算机工程与应用*, 2020, 56(12): 134-140.
- [18] Li Y, Lin X Z, Jiang M Y. Facial expression recognition with cross-connect LeNet-5 network [J]. *Acta Automatica Sinica*, 2018, 44(1): 176-182.  
李勇, 林小竹, 蒋梦莹. 基于跨连接 LeNet-5 网络的面部表情识别 [J]. *自动化学报*, 2018, 44(1): 176-182.
- [19] He Z C, Zhao L Z, Chen C. Convolution neural

- network with multi-resolution feature fusion for facial expression recognition [J]. *Laser & Optoelectronics Progress*, 2018, 55(7): 071503.
- 何志超, 赵龙章, 陈闯. 用于人脸表情识别的多分辨率特征融合卷积神经网络 [J]. *激光与光电子学进展*, 2018, 55(7): 071503.
- [20] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.
- [21] Mishra D. Mish: a self regularized non-monotonic neural activation function [EB/OL]. (2019-08-13) [2020-06-29]. <https://arxiv.org/abs/1908.08681?context=cs.LG>.
- [22] Li S, Deng W H, Du J P. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2584-2593.
- [23] Mollahosseini A, Hasani B, Mahoor M H. AffectNet: a database for facial expression, valence, and arousal computing in the wild [J]. *IEEE Transactions on Affective Computing*, 2019, 10(1): 18-31.
- [24] Liu Z W, Li S, Deng W H. Boosting-POOF: boosting part based one vs one feature for facial expression recognition in the wild [C] // 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), May 30 - June 3, 2017, Washington, DC, USA. New York: IEEE Press, 2017: 967-972.
- [25] Simonyan K, Zisserman A. Verydeep convolutional networks for large-scale image recognition [EB/OL]. (2015-04-10) [2020-06-29]. <https://arxiv.org/abs/1409.1556>.
- [26] Li Y, Zeng J B, Shan S G, et al. Occlusion aware facial expression recognition using CNN with attention mechanism [J]. *IEEE Transactions on Image Processing*, 2019, 28(5): 2439-2450.
- [27] Cao S, Yao Y Q, An G Y. E2-capsule neural networks for facial expression recognition using AU-aware attention [J]. *IET Image Processing*, 2020, 14(11): 2417-2424.