

深度残差卷积下多视角特征融合的人脸表情识别

关小蕊¹, 高璐², 宋文博¹, 林克正¹

(1. 哈尔滨理工大学 计算机科学与技术学院, 哈尔滨 150080; 2. 哈尔滨信息工程学院, 哈尔滨 150431)

摘要:针对现实生活中多视角下人脸表情识别不够精准、计算量大等问题,提出了一种深度残差卷积下多视角特征融合的人脸表情识别模型 MVResNet-FER。首先改进 ResNet 中的残差块,并使用深度可分离网络取代常规卷积网络。其次添加了 CBAM 模块,以增强多视角下有效特征的提取和浅层特征信息的补充。然后使用 RReLU 激活函数取代原始的 ReLu,避免梯度较大时部分节点出现失活。最后使用全局平均池化层代替全连接层实现降维,并将生成的特征向量送入 Softmax 进行分类。实验表明,本文方法在 CK+ 和 RaFD 数据集上产生了较优异的结果,能有效提高人脸表情识别的准确率。

关键词:多视角;特征融合;深度可分离卷积;残差模型;表情识别

DOI:10.15938/j.jhust.2023.02.014

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1007-2683(2023)02-0117-11

Facial Expression Recognition with Multi-perspective Feature Fusion Under Deep Residual Convolution

GUAN Xiaorui¹, GAO Lu², SONG Wenbo¹, LIN Kezheng¹

(1. School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China;
2. Harbin Institute of Information Technology, 150431, China)

Abstract: Aiming at the problems of inaccurate facial expression recognition and large amount of calculation under multi-perspective in real life, a facial expression recognition model MVResNet-FER is proposed, which is based on multi-perspective feature fusion under deep residual convolution. The residual block in ResNet is first improved and the conventional convolutional network is replaced with a depthwise separable network. Second, a CBAM module is added to enhance the extraction of effective features under multi-perspective and the supplementation of shallow feature information. Then use the RReLU activation function to replace the original ReLu to avoid deactivation of some nodes when the gradient is large. Finally, the global average pooling layer is used instead of the fully connected layer to achieve dimensionality reduction, and the generated feature vector is sent to Softmax for classification. Experiments show that the proposed method produces excellent results on the CK+ and RaFD datasets, which can effectively improve the accuracy of facial expression recognition.

Keywords: multi-perspective; feature fusion; depth separable convolution; residual model; expression recognition

0 引言

面部表情是传递人类情绪状态的最直接方式之一^[1],传统人脸表情识别算法一般分为两个步骤:

特征提取和分类器判别^[2],但是极难获取深度特征,且需要人为干预特征提取过程。目前关于人脸表情识别的研究为了实现更高的面部表情识别率而设计更加复杂的网络结构,最终导致网络模型参数量过大,训练模型花费时间长,最后的结果也难以实

收稿日期: 2022-05-20

基金项目: 国家自然科学基金(62071157);黑龙江自然科学基金(F2015040)。

作者简介: 关小蕊(1996—),女,硕士研究生;

宋文博(1998—),男,硕士研究生。

通信作者: 高璐(1982—),女,硕士,副教授,E-mail:hxcigaolu@126.com。

时呈现,再加上日常生活中多视角下人脸表情容易受复杂环境的影响,具有较大的不稳定性。Dachapally 等^[3]基于 CNN 和自动编码器,提出了 Vanilla CNN 模型用于表情识别,虽然一定程度上提高了表情识别准确率,但该网络参数量巨大,最终网络难以收敛。Fan 等^[4]提出了多区域集成框架 MRE-CNN 用于表情识别,通过提取人脸的多个局部特征信息来提升 CNN 的学习能力,虽然提高了表情识别率,但大大增加了模型的参数计算量。Mollahosseini 等^[5]利用 Inception 结构设计了更深层的神经网络来识别表情,随着网络深度的增加,庞大的参数也深深的影响了整体性能。Chollet 等^[6]在 InceptionV3 网络的基础上提出了 Xception 轻量级网络框架,采用深度可分离卷积对特征图进行解耦,缩减参数量的同时有效的避免了过拟合。Google 开发团队推出了一个移动端轻量化网络 MobileNet,该模型采用深度可分离卷积取代标准卷积,并引入宽度因子和分辨率因子两个参数,在图像处理领域取得了较好的成绩^[7-8]。Jung 等^[9]提出的深度网络基于两种不同的模型,但只适用于部分小样本数据集。Kim 等^[10]使用不同学习策略训练多个深度 CNN 完成识别任务,但是网络结构和连接方式过于复杂导致训练过程繁琐耗时。周婕等^[11]利用 TensorFlow 搭建改进的 ResNet 网络,最后结合 OpenCV 人脸检测分类器进行表情识别。Santra 等人^[12]引入局部显著性启发二元模式来识别多视角人脸表情,通过似然分布以及像素强度来计算 LSiBP 特征。Allaert 等^[13]对多种光流方法在人脸数据集上的表现进行评估,让人充分理解了光流在表情识别背景下编码面部运动的充分性。Ahadit 等^[14]提出了一种人脸表情识别系统,利用核极限学习机分类器的变分模态分解和鲸鱼优化进行人脸表情的判别。李一凡等^[15]设计了一种基于全局空间注意力的表情识别方法,解决了传统机器学习方法识别结果泛化能力差的问题。史浩等^[16]利用改进后的 VGG 模型,结合多尺度特征图融合策略提取表情信息,进行面部表情识别。李斌等^[17]在 Xception 框架中利用平均池化将注意力连接方式改为并行连接,并对特征图随机抽取部分做规范化处理,提升了改进卷积神经网络在树木识别上的效果。高明宇等^[18]提出一种基于 GoogLeNet 卷积神经网络的木节缺陷识别方法,利用 GoogLeNet 网络对木节缺陷进行图像处理,具有较高的识别准确率。

随着网络模型的深度不断增加,导致参数量大

幅度加大,出现计算复杂度高、模型整体性能不高、识别准确率低等问题。本文提出了一种深度残差卷积下多视角特征融合的人脸表情识别方法,以 Inception 模块为基础,使用了深度可分离卷积(depth-wise separable convolution, DSC),替换掉传统卷积,并添加了卷积注意力机制模块(convolutional block attention module, CBAM)使网络能够在提取到多视角下人脸更多有用的表情特征,其中深度可分离卷积分步提取通道信息和空间信息。卷积注意力机制模块添加在卷积层后使网络关注有效信息,最后使用 CK+ 和 RaFD 人脸数据库进行大量实验。实验结果表明,将 MVResNet-FER 模型运用在多角度人脸表情数据集 CK+ 和 RaFD 上取得了良好的效果,验证了模型的有效性。

1 相关基本理论

1.1 注意力机制

卷积注意力机制模块^[19]由通道注意力和空间注意力两部分组成,将两种方式融合在一起能够进行多方位卷积,提取到更重要的信息,增强了网络提取有效信息的能力。图1为CBAM原理图。

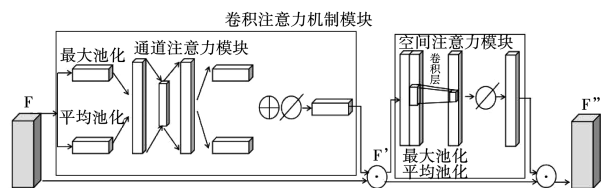


图1 CBAM 的原理图

Fig. 1 Schematic diagram of CBAM

CBAM 模块的输入是从卷积层得到的矩阵 $F \in R^{C \times H \times W}$ 。先对矩阵在通道维度上进行操作,分别是最大池化和平均池化操作,将提取到的两种特征压缩合并在一起,合并后的特征图经过单个卷积核的隐藏层生成通道权值矩阵 $F_1 \in R^{C \times H \times W}$,如式(1)所示:

$$F_1 = M_s(F) \odot F \quad (1)$$

空间注意力机制是对通道注意力机制的结果进行后续操作,是在空间维度上进行操作,再一次进行最大池化和平均池化操作,将特征图叠加压缩为一个信道描述符,经过空间压缩操作的计算得到空间权值矩阵 F_2 ,如式(2)所示。经过操作后可通过矩阵建立像素点之间的重要关系,突出重要的信息区域。

$$F_2 = M_s(F_1) \odot F_1 \quad (2)$$

其中: $M_s(F)$ 为三维的通道压缩权值矩阵; $M_s(F_1)$ 为空间压缩权值矩阵; “ $\odot$ ” 表示矩阵间乘操作。

1.2 深度可分离卷积和残差网络

传统卷积的计算过程相当于将通道和空间联合映射,而深度可分离卷积的思想是将通道和空间分开映射,可以极大地减小了网络模型的数量和计算量。具体来说,将传统卷积分为两部分,一部分是深度卷积,另一部分是一个 1×1 的卷积。具体如图2和图3所示。

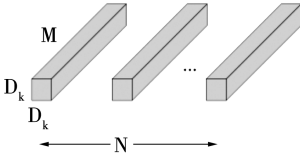


图2 传统卷积

Fig. 2 Traditional convolution graph

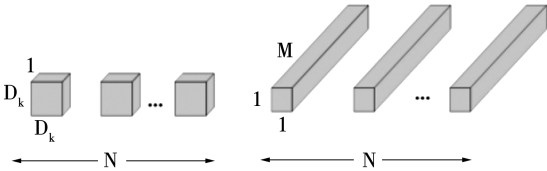


图3 深度可分离卷积

Fig. 3 Depthwise Separable Convolution

使用传统卷积进行特征提取时:若输入是 $D_e \times D_e \times M$, 卷积核为 $D_k \times D_k \times M$, 输出为 $D_f \times D_f \times N$ (N 代表个数), 则计算量如式(3)所示:

$$Conv = D_k \times D_k \times M \times D_f \times D_f \times N \quad (3)$$

使用深度可分离卷积进行特征提取时,计算分为两步。

首先计算深度卷积的计算量如式(4)所示:

$$S - conv1 = D_k^2 \times 1 \times D_f^2 \times M \quad (4)$$

其次计算 1×1 卷积的计算量如式(5)所示:

$$S - conv2 = M \times N \times D_f^2 \quad (5)$$

因此深度可分离卷积的总计算量如式(6)所示:

$$S - conv = D_k^2 \times 1 \times D_f^2 \times M + M \times N \times D_f^2 \quad (6)$$

计算量对比如式(7)所示:

$$\frac{S - conv}{Conv} = \frac{D_k^2 \times D_f^2 \times M + M \times N \times D_f^2}{D_k^2 \times M \times D_f^2 \times N} = \frac{1}{N} + \frac{1}{D_k^2} \quad (7)$$

由以上公式可看出,采用深度可分离卷积,可以大幅度的降低模型参数数量和计算量。

残差神经网络 (ResNet) 将残差的概念添加到了深度卷积神经网络中,原始的残差块由两个卷积层和两个批量归一化层组成,在卷积之后使用 shortcut 连接,来解决因网络层数过深导致的梯度爆炸和退化问题。残差块结构如图4所示。

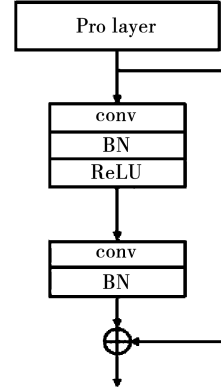


图4 残差块结构

Fig. 4 Residual block structure

残差网络由多个残差块组成,残差块需要拟合出有关恒等映射的残差映射,残差单元公式为

$$\left. \begin{aligned} y_l &= h(x_l) + F(x_l, W_l) \\ x_{l+1} &= f(y_l) \end{aligned} \right\} \quad (8)$$

其中: x_l 为第 $l-1$ 个残差单元的输出; x_{l+1} 为第 l 个残差单元的输出; $h(x_l) = x_l$ 为恒等映射; f 为 Relu 函数。

基于式(8)可以求得从深层 l 到浅层 L 的学习特征,即

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, W_i) \quad (9)$$

2 MVResNet-FER 模型

2.1 改进的 RESNET 结构

为了提取到多视角下人脸更丰富的表情信息,提高表情分类的准确率,采用改进的 ResNet18 进行人脸全局特征提取,对残差块结构进行了优化。由于 RReLU 激活函数可以增加网络的非线性表达能力,所以在残差单元的卷积层之后加入 RReLU 函数,并去除了原有结构中的两个归一化层,有效减少了其对图像空间特征信息的破坏,如图5(a)所示。为了进一步增强网络浅层特征的表达能力,在改进残差块基础上,提出一种跨层集成策略。改进后的整体残差网络结构如图5(b)所示。通过 shortcut 聚合各个残差单元输出的特征图,残差网络结构从全局残差变为局部残差。该结构充分利用了每个残差块输出的图像特征信息,不仅补充了图像浅层特征

信息,并且有效改善了过拟合,对人脸表情识别准确率的提升起到积极作用。

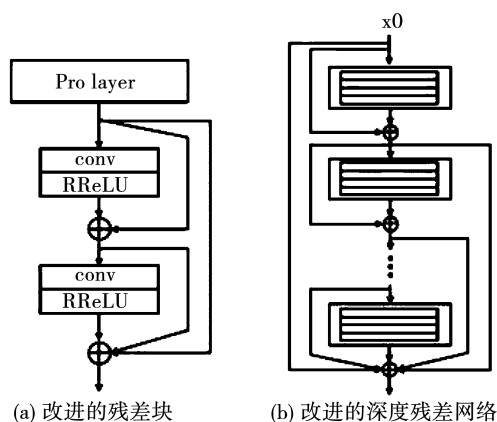


图 5 改进的残差结构

Fig. 5 Improved residual structure

改进的深度残差网络的输出为

$$X_{out} = \sum_{l=0}^N X_l = X_0 + X_1 + \cdots + X_N \quad (10)$$

其中: X_l 为第 l 个残差单元的输出; X_{out} 为浅层特征 and 所有残差单元的输出总和。

为了解决 ReLU 函数可能会在训练过程中导致部分节点失活的问题,激活函数选择 RReLU 函数,其函数为

$$\text{RReLU}(x) = \begin{cases} x, & x > 0 \\ ax, & x < 0 \end{cases} \quad (11)$$

RReLU 相较于 ReLU 的优点是: RReLU 函数中的 a 是从给定均匀分布的范围内随机抽取的值,在测试环节就会固定下来,解决了 ReLU 函数可能导致神经元死亡的问题。

2.2 MVRESNET-FER 模型结构

为了更好的加快网络收敛的速度,提高多视角下人脸表情识别的准确率,减少网络训练过程中参数过多带来的计算负担,本文借鉴了 Inception V1 网络结构的特性,因为在增加网络通道和深度的同时,减轻网络的计算量。

Inception V1 结构能实现特征图的通道变换,将 H 通道变换到 I 通道,然后将不同尺寸的特征图融合在一起形成新的特征。由于 H 和 I 的取值通常情况下很大,会导致运算量过大,为了更好的进行通道变换,在 Inception 中引入 1×1 卷积,使用先降维再升维的方法节省计算量。 1×1 卷积不仅能够对特征通道进行降维和升维,减少计算量,而且保留了图像的非线性表达能力。首先将 H 通道特征图降低

为 K 通道,再升高为 I 通道,由于 K 的值比 H 和 I 都要小的多,这样很大程度上减少了参数并减轻了模型的计算量。Inception V1 模型结构如图 6 所示。

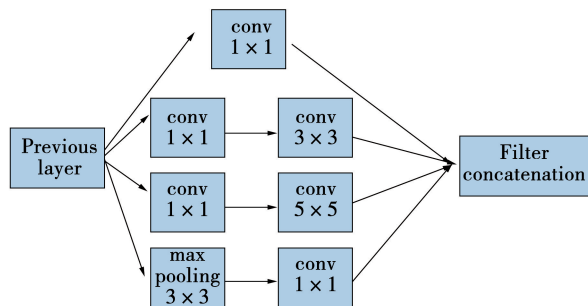


图 6 Inception V1 模型结构

Fig. 6 Inception V1 model structure

本文结合了 Inception 网络结构与残差神经网络的良好特性,并在此基础上做了改进,提出了 MVResNet-FER 模型。该模型大大减轻了网络的计算负担,并结合改进的残差学习模块补充浅层特征信息,避免梯度爆炸和消失,使网络能更好的获取图像的全局特征。

MVResNet-FER 模型整体结构如图 7 所示。模型共有 15 个卷积层:4 个 1×1 卷积,2 个常规卷积层,9 个深度可分离卷积,其中 8 个深度可分离卷积组成了四层残差块,残差块中常规卷积层使用深度可分离卷积层来取代,卷积核大小为 3×3 。同时在网络模型的第一个常规卷积层和最后一个深度可分离卷积层后各添加一个 CBAM 模块,以增强面部关键特征的提取,其中先进行通道注意力后进行空间注意力,增强多视角下人脸的有效特征并抑制无效的特征。为了防止梯度较大时部分节点失活,提升模型的线性表达,模型采用 RReLU 激活函数取代原始的 ReLU。为了更有效的缩减冗余参数,最后使用全局平均池化层(global average pooling, GAP)代替全连接层实现降维生成特征向量,减少了参数量,最后将特征向量传入 Softmax 进行类别的预测。

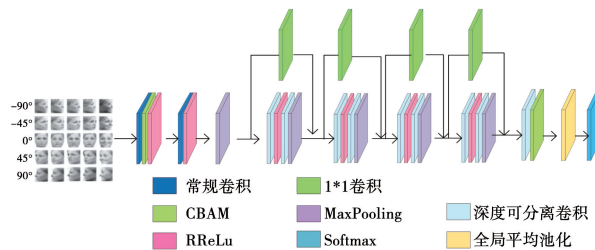


图 7 模型结构图

Fig. 7 Model structure diagram

模型结构中具体网络参数配置如表 1 所示。

表 1 MVResNet-FER 网络配置

Tab. 1 Configuration of the MVResNet-FER

主网络	输入大小	卷积核	输出大小
Conv-1	64 × 64 × 1	3 × 3 × 8	64 × 64 × 8
CBAM	64 × 64 × 8	—	64 × 64 × 8
Conv-2	64 × 64 × 8	3 × 3 × 16	64 × 64 × 16
MaxPool-1	64 × 64 × 16	—	32 × 32 × 16
Inception1	32 × 32 × 16	3 × 3 × 32	16 × 16 × 32
Inception2	16 × 16 × 32	3 × 3 × 64	8 × 8 × 64
Inception3	8 × 8 × 64	3 × 3 × 128	4 × 4 × 128
Inception4	4 × 4 × 128	3 × 3 × 256	2 × 2 × 256
S-Conv9	2 × 2 × 256	3 × 3 × 8	2 × 2 × 8
CBAM	2 × 2 × 8	—	2 × 2 × 8
Global-Pool	2 × 2 × 8	—	1 × 1 × 8

表 2 为各网络结构的参数量对比情况。结果显示本文模型参数计算量较小,大大减轻了网络的计算负担。

表 2 网络参数计算量对比

Tab. 2 Comparison of network parameter calculations

模型	计算量 × 10 ⁶
Resnet	23
AlexNet	62
VGG16	138
MVRESNET-FER	14

2.3 算法步骤描述

MVResNet-FER 网络模型的具体算法步骤描述如下:

步骤 1 对不同视角下的人脸表情图像进行数据处理。

步骤 2 将经过数据处理的图像先放入常规卷积层中进行 2 次卷积运算和 1 层最大池化运算后,再放入改进的 ResNet18 卷积和最大池化运算。

步骤 3 将经过深度可分离卷积获得面部关键特征获得的特征送入全局平均池化层进行融合降维,融合后的新特征为 F_z 。

步骤 4 将步骤 3 得到的结果与真实标签一起输入到损失函数中,通过 $L_1 = - \sum_{i=1}^N y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log (1 - \hat{y}^{(i)})$ 计算得到损失。

步骤 5 使用 $f(x) = \frac{1}{n} \sum_{i=1}^n f_i(x)$ 计算反向传播误差,同时使用 $x = x - \eta \nabla f_i(x)$ 调整参数并更新 x 。

步骤 6 保存准确率最高的模型,然后将测试图片输入到训练好的网络中得到新的特征向量 D ,最后通过 Softmax 进行分类预测,输出表情类别。

3 实验结果与分析

3.1 实验环境

本文提出的 MVResNet-FER 模型的硬件环境为:Windows 10 (64 位),Tensorflow 2.0, GPU (GTX 1080Ti 显卡 128G)。软件环境为:开发环境 (PyCharm Community Edition X64), 开发语言 (Python 3.6),深度学习框架 (Pytorch 0.4)。

3.2 评价指标

本文实验使用准确率 (Accuracy) 指标评价 MVResNet-FER 算法的优劣。准确率能够表示被正确分类的数据和总数据之间的比重关系,具体计算为

$$Accuracy = \frac{TP + TN}{P + N}$$

(12)

其中: TP 为正确的被分类为真的数据量; TN 为被正确地划分为假的数据量; $P + N$ 为总的测试数据量。

3.3 CK + 数据集实验

The extended Cohn-Kanade Database (CK +) 数据集属于小样本数据集,据集采集了 123 名实验者的 593 张正面人脸图像。包含 8 种表情:即 Happy、Angry、Sad、Fear、Surprise、Disgust、Neutral 和 Contempt,部分样本如图 8 所示。



图 8 CK + 数据集部分表情示例

Fig. 8 Examples of partial expressions in the CK + datas

由于 CK + 数据集过小,模型训练过程中网络中的参数就会过度拟合。所以为了有足够多的数据支撑 CNN 模型训练,将现有数据进行扩充,将图片经过随机缩放、平移等方法来增加训练数据。经数据增强后的数据集数量变为原来的 9 倍,大大增加了模型的训练量,减小了过拟合的风险。其中,将 CK + 数据集的 1/5 作为测试集,4/5 作为训练集。CK

+ 各类表情数据分布如图 9 所示。

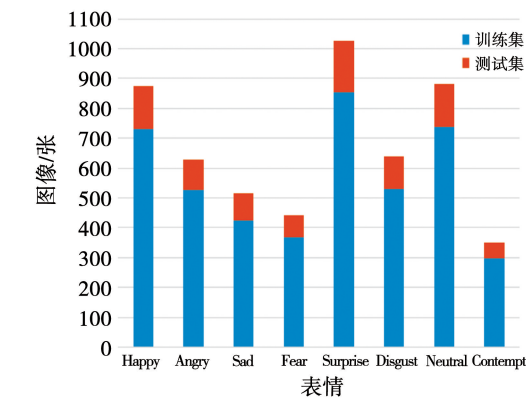


图 9 CK + 数据集各类表情分布
Fig. 9 Distribution of various expressions
in CK + dataset

为了防止出现过拟合现象,在数据增强的基础上,CK + 数据集在训练过程中采用交叉验证法,1/5 作为测试集,4/5 作为训练集。每个网络都训练 100 个 epoch。选择小批量梯度下降算法进行优化,学习率可以动态更新,随着 epoches 的增大逐渐衰减。

表 3 CK + 数据库不同视角表情识别率

Tab. 3 Expression recognition rate of different perspectives in CK + database

表情	准确率					整体准确率	平均准确率
	0°	30°	45°	60°	90°		
Happy	1. 00	0. 99	0. 97	0. 92	0. 90	0. 961	0. 956
Sad	0. 95	0. 92	0. 91	0. 90	0. 87	0. 923	0. 910
Angry	0. 99	0. 96	0. 95	0. 92	0. 89	0. 951	0. 936
Disgust	0. 96	0. 93	0. 92	0. 85	0. 82	0. 924	0. 896
Contempt	0. 94	0. 93	0. 93	0. 92	0. 88	0. 922	0. 920
Fear	0. 95	0. 94	0. 90	0. 86	0. 83	0. 919	0. 896
Neutral	1. 00	0. 97	0. 96	0. 94	0. 90	0. 950	0. 954
Surprise	1. 00	0. 98	0. 97	0. 93	0. 88	0. 962	0. 952
平均准确率	0. 973	0. 952	0. 939	0. 903	0. 871	0. 949	0. 937

表 3 为 CK + 数据库上不同视角表情的识别率。从表 3 可看出,开心的平均识别率是最高的,在 5 个角度下均达到了 90% 以上,Happy、Angry、Sad、Surprise、Neutral 和 Contempt 的识别率都达到了预期目标,而害怕和轻蔑识别率均低于 90%,害怕和轻蔑的数据样本相对较小,且 CK + 的数据样本分布不均衡容易导致误判造成准确率降低。此外,0°时人脸表情的识别率与其他 4 个角度相比是最高的,5 个

最终将验证损失值最低的模型存储下来作为最终的模型进行测试。

图 10 为在 CK + 数据集上进行训练时损失值 loss 的变化情况。随着迭代周期 (epoches) 的增加,模型在 CK + 数据集上损失 (loss) 逐渐降低。收敛效果很好,没有出现过拟合的情况。

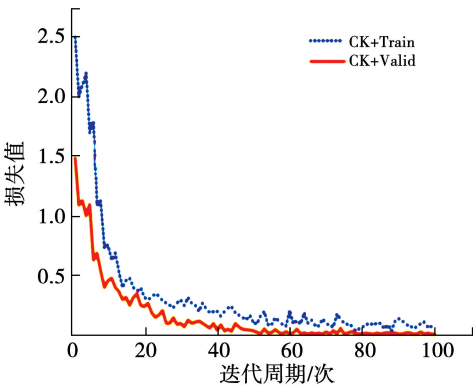


图 10 CK + 数据集上模型的损失值曲线
Fig. 10 The loss value curve of the model
on the CK + dataset

角度中 90°人脸整体识别率偏低。

图 11 为 CK + 数据集上多个角度下图像识别率的混淆矩阵,可以看出,每种面部表情都和个别表情存在轻微混淆,但整体的识别率也较高。其中,Happy 和 Neutral 表情混淆程度最小,识别率最高,Angry 和 Fear 表情更容易因为混淆出现错误,不过差距相对较小。

为了验证 MVResNet-FER 网络模型的有效性,

将 CK + 数据库上的实验结果与表情识别的一些经典算法和未嵌入 CBAM 模块的本文方法进行比较,表 4 为在本文模型与其他几种模型准确率的对比。对比实验的经典算法模型有:Attention Net^[20]、Xception^[21]、LBP + SVM^[22]。

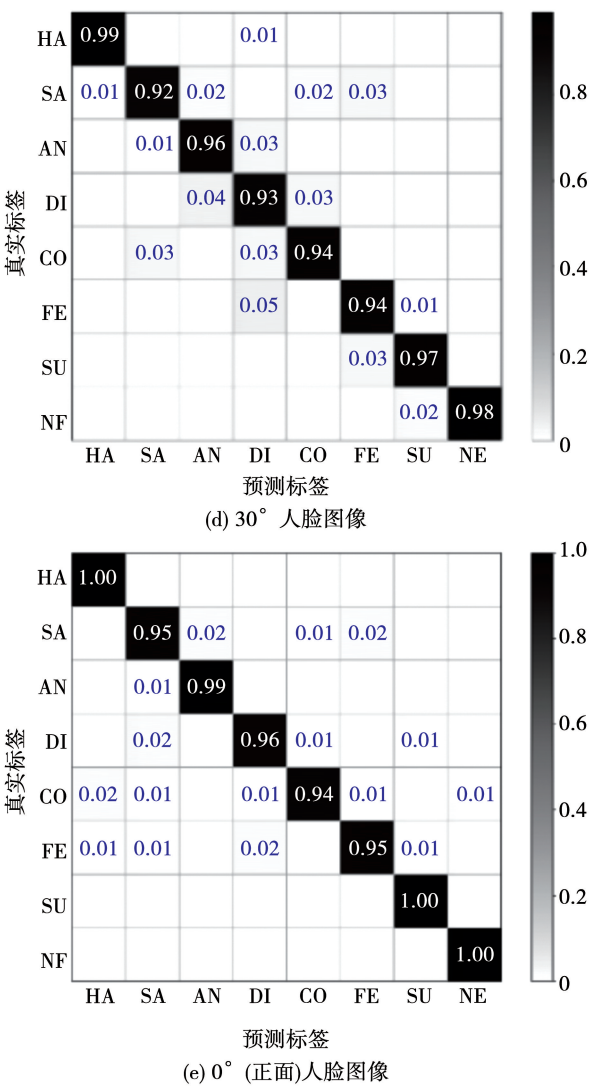
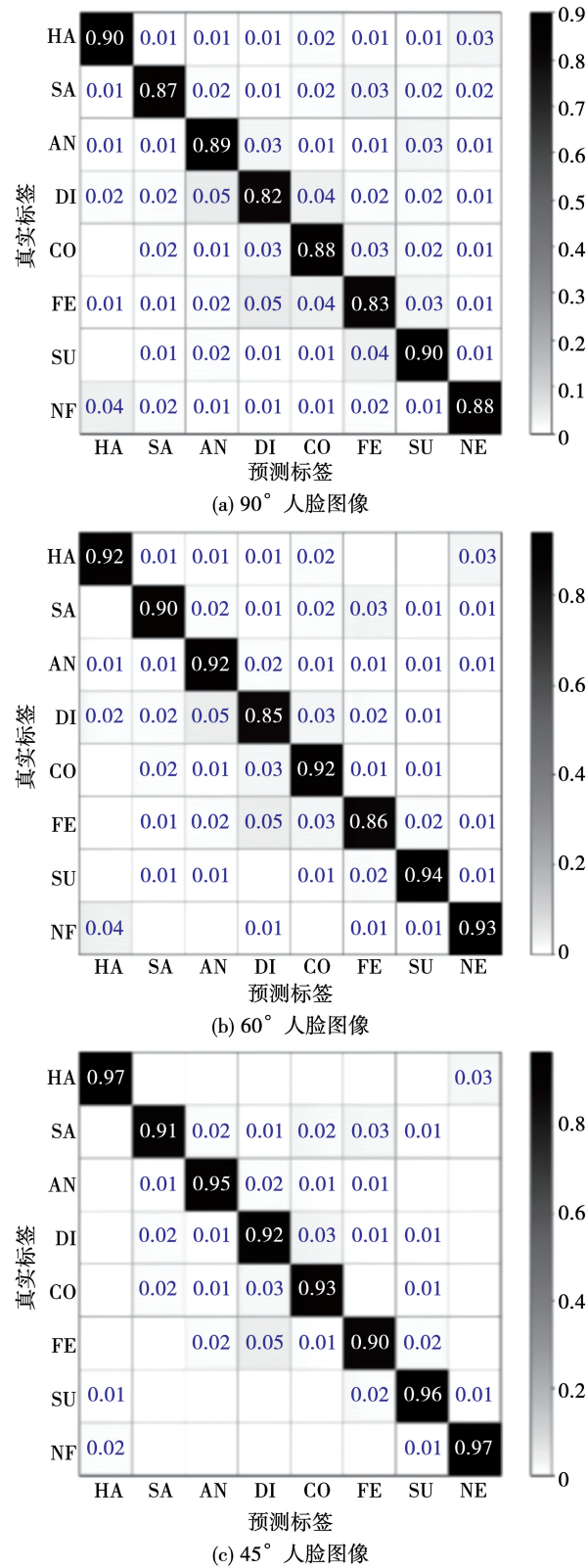


图 11 CK + 数据集中不同角度图像识别率的混淆矩阵

Fig. 11 Confusion matrix of image recognition rate at different angles in CK + dataset

表 4 不同模型在 CK + 数据集上的整体准确率

Tab. 4 Overall accuracy of different models on the CK + dataset

模型	准确率/%
Attention Net	90. 30
Xception	94. 50
LBP + SVM	90. 30
MVResNet-FER(未嵌入 CBAM)	92. 71
MVResNet-FER(本文方法)	94. 90

本文模型与其他模型进行了人脸表情识别的准确率对比,如图 12 所示。

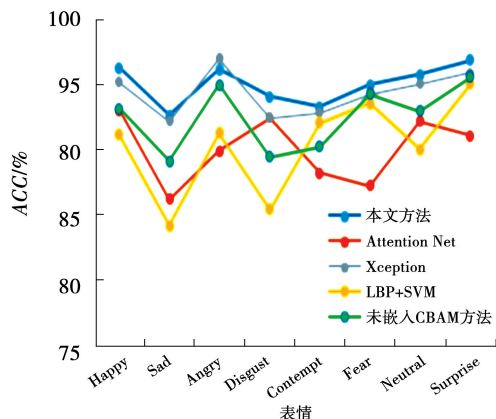


图 12 CK + 数据集下不同模型的准确率

Fig. 12 The accuracy of different models under the CK + dataset

由实验结果可知, MVResNet-FER 模型在 CK + 上得到的识别率高于 Attention Net 和 LBP + SVM, 模型准确率提高了 4.6%, 且网络参数和运行时间缩减了很多, 同时也比 Xception 方法的准确率提升了 0.4%, 且训练时间较 Xception 减少一半, 参数和计算复杂度也降低了许多, 对算法的整体性能提升更加明显。未添加注意力模块的 MVResNet-FER 模型识别率达到了 92.71%, 而嵌入 CBAM 模块之后的模型准确率提升了 2.19%, 由此可知, 使用改进的 ResNet 网络和 CBAM 模块对模型的准确率有了

较大的提升, 整个网络模型在使用全局平均池化层代替全连接层后减少了网络的参数量和计算量, 保证了较高的识别率。

3.4 RaFD 数据集实验

Radboud 人脸数据集 (RaFD) 是 Radboud 大学 Nijmegen 行为科学研究所整理的, 共包含 67 个模特: 20 名白种男性, 19 名白种女性, 4 个白种男孩, 6 个白种女孩, 18 名摩洛哥男性。总共 8 040 张图, 包含 8 种表情, 即愤怒、厌恶、恐惧、快乐、悲伤、惊奇、蔑视和中立。每一个表情都有多个注视方向, 且从不同的角度使用五个相机同时拍摄。RaFD 数据库中的部分样本如图 13 所示。

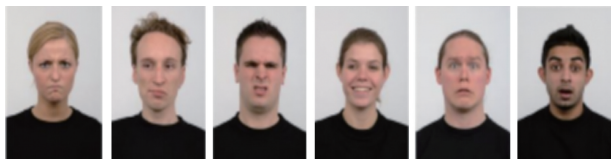


图 13 RaFD 数据集

Fig. 13 RaFD dataset

为了防止过拟合, 选择对 RaFD 数据集进行预处理操作。由于 CK + 和 RaFD 数据集的照片大小

不同, 所以将两个数据集中所有样本图片的大小统一为 64×64 的灰度图像。首先对数据集中的图像进行人脸检测, 裁剪有效面部区域。为了扩充数据集, 充分训练模型, 提高人脸图片的利用率, 本文使用小幅度转动、翻转和平移等变换对图像进行处理。如图 14 所示。



图 14 图像预处理

Fig. 14 image preprocessing

将预处理之后的图像分为两部分, 1/5 作为测试集, 4/5 作为训练集。然后送入 MVResNet-FER 网络中训练, 网络训练过程中 epoch 设为 100, 学习率设置为 0.01, 当迭代次数超过 30 后学习率每 5 轮衰减一次, 依次衰减 5%, 选取 ReLU 函数作为激活函数, 通过随机梯度下降进行网络参数的优化。最终将验证数据集损失值最低的模型作为最终的训练模型进行保存。

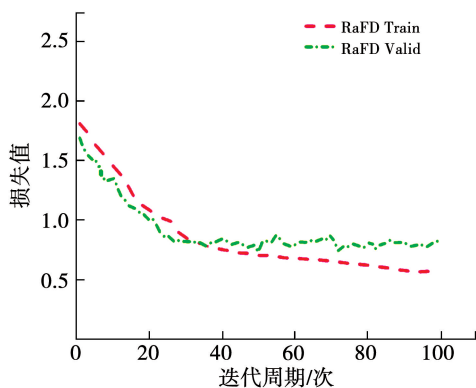


图 15 RaFD 数据集上模型的损失值曲线

Fig. 15 The loss value curve of the model on the RaFD dataset

由图 15 可知, 随着 epoches 的增加, 模型在 RaFD 数据集上的损失值 (loss) 逐渐降低, 收敛情况很好, 未出现过拟合。说明在模型训练过程中, RReLU 激活函数能够自适应地从数据中学习参数的优点可以有效避免训练过程中出现部分节点失活的问题, 此外, 由于 RReLU 函数与其他激活函数相

比,本身更具有收敛性,所以对形成损失值最低的模型有积极的影响。

表 5 为 RaFD 数据库上不同视角表情的识别率。从表 5 中可以看出,在 RaFD 数据集中,不同角度下人脸数据对模型的识别率影响不同。其中,正脸即角度为 0° 的识别率最高,达到了 98.9%,整体的平均识别率为 94.5%,也达到了预期水平。随着

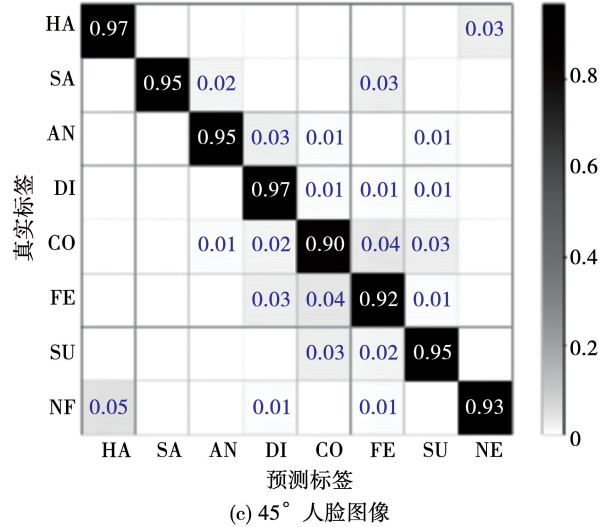
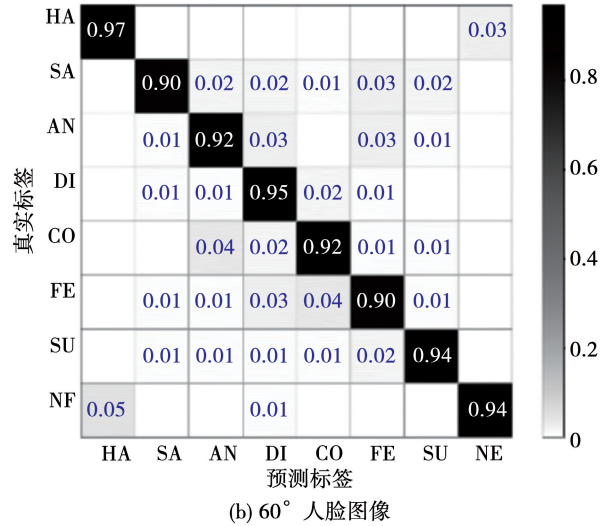
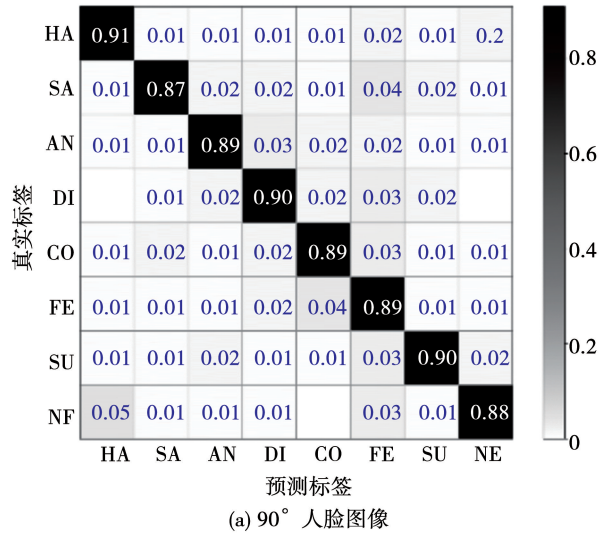
角度的变化,人脸部分结构的特征缺失,尤其是 90° 的人脸图像只能显示出半边脸,因而提取到的有效特征比较少,造成表情识别率降低,为 89.1%。通过实验结果可知,文中模型使用的 RReLU 激活函数本身所具有的随机性会带来正则化的效果,为模型性能和准确率的提升有积极的作用,添加的卷积注意力机制模块也保证了面部表情较高的识别率。

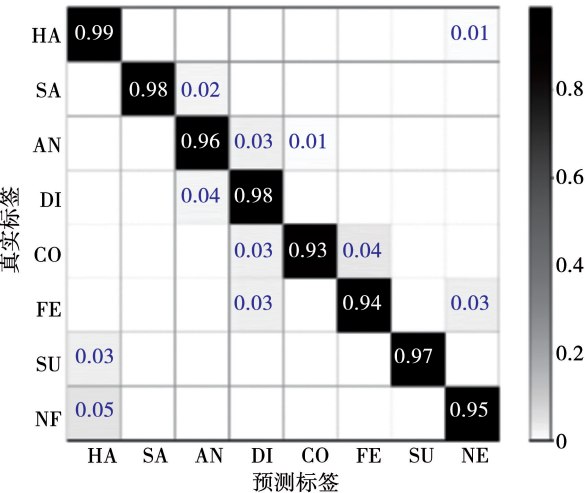
表 5 RaFD 数据库不同视角表情识别率

Tab. 5 Expression recognition rate of different perspectives in RaFD database

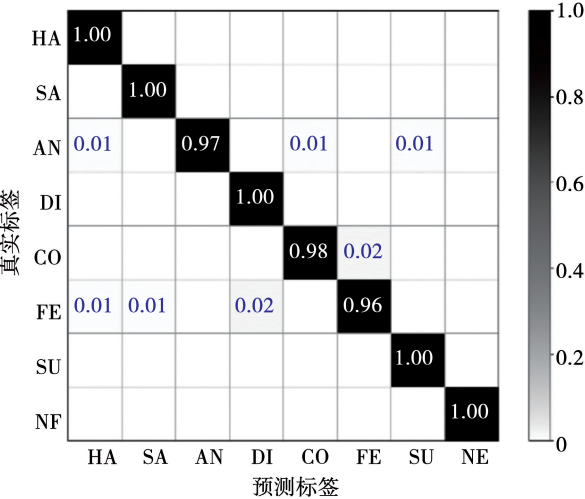
表情	准确率					整体准确率	平均准确率
	0°	30°	45°	60°	90°		
Happy	1.00	0.99	0.97	0.97	0.91	0.97	0.968
Sad	1.00	0.98	0.95	0.90	0.87	0.96	0.940
Angry	0.97	0.96	0.95	0.92	0.89	0.96	0.938
Disgust	1.00	0.98	0.97	0.95	0.90	0.96	0.960
Contempt	0.98	0.93	0.90	0.92	0.89	0.95	0.924
Fear	0.96	0.94	0.92	0.90	0.89	0.94	0.922
Neutral	1.00	0.97	0.95	0.94	0.90	0.94	0.952
Surprise	1.00	0.95	0.93	0.94	0.88	0.96	0.940
平均准确率	0.989	0.963	0.943	0.930	0.891	0.959	0.945

图 16 为 RaFD 数据集多个角度下人脸图像识别率的混淆矩阵,从混淆矩阵中可看出,虽然表情之间有轻微的混淆情况,但是每个类别的表情之间提取到的特征区分度比较高,所以整体识别率很高。其中,开心平均识别率是最高的,在各个角度下均达到了 90% 以上,且 8 种表情识别准确率均达到了 90% 以上,这与 RaFD 人脸样本分布相对均衡有关,对 MVResNet-FER 模型起到了良好的训练效果,同时也证明此模型在多视角下对人脸表情的识别率是高的。





(d) 30° 人脸图像



(e) 0° (正面)人脸图像

图 16 RaFD 数据集中不同角度图像识别率的混淆矩阵
Fig. 16 Confusion matrix of image recognition rate at different angles in RaFD dataset

本文将 MVResNet-FER 网络在 RaFD 数据集上的实验结果与其他现有网络模型进行比较,不同模型在 RaFD 数据集上的准确率如表 6 所示。

表 6 不同模型在 RaFD 上的整体准确率
Tab. 6 Overall accuracy of different models on the RaFD dataset

模型	准确率/%
Mnist-cnn	61.40
Xception	91.40
MVFE-LightNet	94.40
MVResNet-FER(未嵌入 CBAM)	92.71
MVResNet-FER	95.90

可以看出,本文网络识别准确率比 Mnist-cnn、Xception、MVFE-LightNet^[23] 和未嵌入 CBAM 的 MVResNet-FER 分别高 34.5%、4.5%、1.5% 和

3.2%。由于表情之间存在相似特征,所以添加注意力机制模块后可以帮助整个网络在多视角下具备更好的表情分类效果。同时,使用最大池化和 GAP 减少了模型训练参数和计算量,保证 MVResNet-FER 网络较高的识别率。

本文模型与其他模型进行了人脸表情识别的准确率对比,如图 17 所示。

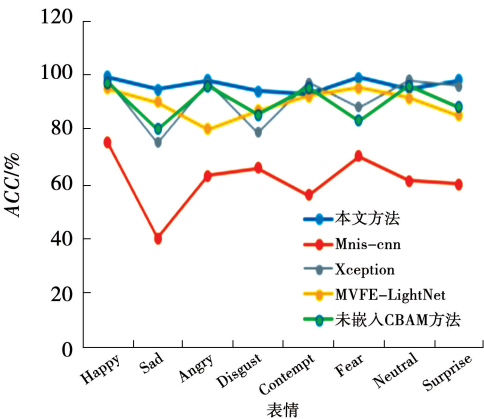


图 17 RaFD 数据集下不同模型的准确率
Fig. 17 The accuracy of different models under the RaFD dataset

4 结 论

本文提出了一种深度残差卷积下多视角特征融合的人脸表情识别方法,充分综合了 CBAM 和改进的 ResNet 网络的优势。一方面,添加 CBAM 模块,增强了多视角下有效特征的充分提取,并对提取到的特征进行特征融合。另一方面,利用改进的残差学习模块补充浅层特征信息,不仅解决了 ResNet 网络训练模型过程中效率低、复杂度高的问题,也提高了多视角下人脸表情的分类准确率和识别率。CK + 和 RaFD 数据集上的实验结果表明,本文的 MVResNet-FER 模型有明显的优势,对于多视角下特征融合后的表情识别具有较高的识别准确率。

参 考 文 献:

[1] 汤春明,赵红波,张小玉. 基于流形学习 2D-LDLP 的东亚人脸表情识别算法[J]. 计算机工程与应用, 2018,54(17):146.
TANG Chunming, ZHAO Hongbo, ZHANG Xiaoyu. East Asian Facial Expression Recognition Algorithm Based on Manifold Learning 2D-LDLP[J]. Computer Engineering and Applications, 2018,54(17):146.
[2] 汤吉训生,王荣飞. 自适应加权 LGCP 与快速稀疏表示的面部表情识别[J]. 计算机工程与应用,2017,53(1):158.

- TANG Jixunsheng, WANG Rongfei. Facial Expression Recognition Based on Adaptive Weighted LGCP and Fast Sparse Representation [J]. Computer Engineering and Applications, 2017, 53(1): 158.
- [3] DACHAPALLY P R. Facial Emotion Detection Using Convolutional Neural Networks and Representational Autoencoder Units [J]. arXiv Preprint arXiv: 1706.01509, 2017.
- [4] FAN Y, LAM J C K, LI V O K. Multi-region Ensemble Convolutional Neural Network for Facial Expression Recognition [C]//International Conference on Artificial Neural Networks. Springer, Cham, 2018: 84.
- [5] 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). IEEE, 2016: 1.
- [6] CHOLLET F. XCEPTION: Deep Learning with Depth-wise Separable Convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251.
- [7] NAN Y, JU J, HUA Q, et al. A-MobileNet: An Approach of Facial Expression Recognition [J]. Alexandria Engineering Journal, 2022, 61(6): 4435.
- [8] ZHANG F, LI Q, REN Y, et al. An Expression Recognition Method on Robots Based on Mobilenet v2-ssd [C]//2019 6th International Conference on Systems and Informatics (ICSAI). IEEE, 2019: 118.
- [9] JUNG H, LEE S, YIM J, et al. Joint Fine-tuning in Deep Neural Networks for Facial Expression Recognition [C]//IEEE International Conference on Computer Vision. IEEE, 2015: 2983.
- [10] KIM B K, LEE H, ROH J, et al. Hierarchical Committee of Deep CNNs with Exponentially-weighted Decision Fusion for Static Facial Expression Recognition [C]//7th ACM International Conference on Multimodal Interaction, 2015: 427.
- [11] 周婕, 马明栋. 基于改进的 ResNet 网络的人脸表情识别 [J]. 计算机技术与发展, 2022, 32(1): 25.
- ZHOU Jie, MA Mingdong. Facial Expression Recognition Based on Improved ResNet Network [J]. Computer Technology and Development, 2022, 32(1): 25.
- [12] SANTRA B, MUKHERJEE D P. Local Saliency-inspired Binary Patterns for Automatic Recognition of Multi-view Facial Expression [C]//2016 IEEE International Conference on Image Processing, Phoenix, AZ, 2016: 624.
- [13] ALLAERT B, WARD I R, BILASCO M, et al. A Comparative Study on Optical Flow for Facial Expression Analysis [J]. Neurocomputing, 2022, 500: 434.
- [14] AHADIT A B, JATOTH R K. A Novel Multi-feature Fusion Deep Neural Network Using HOG and VGG-Face for Facial Expression Classification [J]. Machine Vision and Applications, 2022, 33(4): 1.
- [15] 李一凡, 袁龙健, 王瑞. 基于全局空间注意力的人脸表情识别 [J]. 工业控制计算机, 2022, 35(1): 75.
- LI Yifan, YUAN Longjian, WANG Rui. Face Expression Recognition Based on Global Spatial Attention [J]. Industrial Control Computer, 2022, 35(1): 75.
- [16] 史浩, 邢瑜航, 陈炼. 基于多尺度融合注意力机制的人脸表情识别研究 [J]. 微电子学与计算机, 2022, 39(3): 34.
- SHI Hao, XING Yuhang, CHEN Lian. Research on Facial Expression Recognition Based on Multi-scale Fusion Attention Mechanism [J]. Microelectronics and Computer, 2022, 39(3): 34.
- [17] 李滨, 敬启超. 改进的卷积神经网络在树种识别中的应用 [J]. 森林工程, 2021, 37(5): 75.
- LI Bin, JING Qichao. Improved Convolutional Neural Network for Tree Species Recognition [J]. Forest Engineering, 2021, 37(5): 75.
- [18] 高明宇, 倪海明, 张博洋, 等. 一种基于 GoogLeNet 卷积神经网络的木节缺陷识别方法 [J]. 森林工程, 2021, 37(4): 66.
- GAO Mingyu, NI Haiming, ZHANG Boyang, et al. A Method for Recognizing Wood Knots Defects Based on GoogLeNet Convolutional Neural Network [J]. Forest Engineering, 2021, 37(4): 66.
- [19] 牛瑞华, 杨俊, 邢斓馨, 等. 基于卷积注意力模块和双通道网络的微表情识别算法 [J]. 计算机应用, 2021, 41(9): 2552.
- NIU Ruihua, YANG Jun, XING Lanxin, et al. Micro-expression Recognition Algorithm Based on Convolutional Attention Module and Dual-channel Network [J]. Computer Applications, 2021, 41(9): 2552.
- [20] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018: 6848.
- [21] CHOLLET F. Xception: Deep Learning with Depth Wise Separable Convolutions [C]//IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE, 2017: 1800.
- [22] ZHAO G, PIETIKAINEN M. Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 915.
- [23] 钱勇生, 邵洁, 季欣欣, 等. 基于改进卷积神经网络的多视角人脸表情识别 [J]. 计算机工程与应用, 2018, 54(24): 12.
- QIAN Yongsheng, SHAO Jie, JI Xinxin, et al. Multi-view Facial Expression Recognition Based on Improved Convolutional Neural Network [J]. Computer Engineering and Applications, 2018, 54(24): 12.