

基于卷积注意力的轻量级人脸表情识别方法

尹鹏博, 潘伟民*, 张海军

新疆师范大学计算机科学技术学院, 新疆 乌鲁木齐 830054

摘要 为了解决深度学习模型在人脸表情识别研究中存在的数据集需求量大、硬件配置要求高等问题,提出了一种基于卷积注意力的轻量级人脸表情识别方法。首先,用分解卷积对模型参数进行降维处理;然后,在模型中嵌入卷积注意力机制模块,以提高模型的特征提取能力;其次,针对数据集中的类别不平衡问题,采取代价敏感的损失函数对模型进行优化;最后,进行表情识别任务前将模型在人脸识别数据集上进行预训练,以提高模型提取人脸特征的能力。实验结果表明,本方法能在有效降低模型复杂度的同时保持较高水平的检测效果,且具有较强的实用性。

关键词 图像处理;注意力机制;分解卷积;轻量级模型;表情识别;代价敏感

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP202158.1210023

Lightweight Facial Expression Recognition Method Based on Convolutional Attention

Yin Pengbo, Pan Weimin*, Zhang Haijun

College of Computer Science and Technology, Xinjiang Normal University, Urumqi, Xinjiang 830054, China

Abstract In the research on facial expression recognition of deep learning models, a lightweight facial expression recognition method based on convolutional attention is proposed in this paper to solve the problems of large dataset demand and high hardware configuration requirements. First, the model parameters are decomposed and convolved for dimensionality reduction. Then, the convolutional attention mechanism module is embedded in the model to improve its feature extraction ability. For the problem of category imbalance in a dataset, the model is optimized using a cost-sensitive loss function. Finally, the model is pretrained on a face recognition dataset before performing a facial expression recognition task to improve the model's ability of extracting facial features. Experiment results show that the method effectively reduced the model complexity while maintaining a high level of detection effect along with having strong practicability.

Key words image processing; attention mechanism; decomposition convolution; lightweight model; expression recognition; cost sensitivity

OCIS codes 100.2000; 100.5010; 150.1135

1 引言

面部表情是人类传达情感状态和意图最有力、最自然、最普遍的信号之一^[1]。基本情绪包括愤怒、恐惧、厌恶、快乐、惊讶和悲伤,研究表明,人类对某些基本情绪的感知是相同的,且基本情绪可以被不

同文化背景的人感知、认同和理解。因此,人脸表情识别研究在社交机器人、课堂专注度评价以及驾驶员疲劳监测等众多领域中有着广泛的应用前景。

早期的人脸表情识别研究大多使用手工特征提取和浅层学习算法,如局部二元模式、非负矩阵分解和稀疏学习^[2]。随着深度学习技术的发展,特别是

收稿日期: 2020-08-28; 修回日期: 2020-10-09; 录用日期: 2020-11-14

基金项目: 新疆维吾尔自治区创新环境建设专项(2019XS08)

* E-mail: panweiminss@163.com

自 2013 年 FER2013 和 EmotiW 等情绪识别比赛从具有挑战性的真实场景中收集了相对充足的训练数据后,深度学习逐渐成为表情识别领域的主流技术。张红颖等^[3]提出了一种双通道卷积神经网络(CNN),该网络结合局部二值模式和梯度特征提高了模型的准确率。褚晶辉等^[4]提出了一种基于注意力机制的人脸表情识别模型,该模型通过引入残差单元提高注意力机制的特征提取能力。杨旭等^[5]在 AlexNet 中引入多尺度卷积以适应小尺寸的人脸图像,并通过融合低层次特征和高层次特征得到了效率更高的分类器。Chen 等^[6]提出了一种结合两种神经网络的模型,一种是输入为完整人脸图像的主网络,另一种是输入为含有显著表情特征区域的预处理人脸图像辅助学习网络。这些研究均通过加深或拓宽网络结构提高模型的识别效果,没有将足够多的数据用于训练大型网络。Ding 等^[7]使用生成对抗网络生成了大量的训练数据,并使用迁移学习方法弥补训练数据不足的问题^[8]。尽管基于深度学习模型的人脸表情识别研究取得了很大的进展,但深度学习模型也增加了算法的运算量,且模型的效果越好,计算量越大,模型运行所需的硬件配置也越高,如运行概率标签模型(PLD)^[9]所需的内存约为 12 GB,而实际生产环境中,设备的配置水平受成本限制,过高的配置需求会严重阻碍模型的实际应用。因此,如何降低模型对硬件设备的需求,使模型能在性能较低的小型、微型设备上正常运行也是一个重要的研究方向。

CNN 在表情识别领域有着广泛的应用,如残差网络(Resnet-Net)。已有研究表明,提高模型准确率最有效的手段是不断增加卷积的层数,但这种方式会使网络参数急剧增加。Szegedy 等^[10]对 CNN 进行分解,在同等状态下有效降低了网络的参数规模。人脸表情是由眼睛、嘴型等关键点的肌肉运动决定^[11],注意力机制可使模型具有关注特定特征的能力,因此,在神经网络模型中引入注意力机制有助于对人脸特征的提取。近年来,双向编码器表征模型(BERT)等预训练模型在多个领域中都取得了很大的发展,将模型在与人脸识别相关数据集上进行预训练后再迁移到表情识别领域可解决人脸表情数据量不足的问题。针对数据集中存在的样本类别不平衡问题,可通过生成对抗网络为样本较少的类别生成数据,以平衡数量差距,但生成模型本身的误差较大,表情识别模型在误差较大的数据集上训练会进一步放大误差,可采用代价敏感的损失函数对模型

进行优化避免该问题。

综上所述,本文提出了一种轻量级卷积模型(LCM),该模型基于迁移学习的思路,通过分解多路通道卷积减少模型参数,实现模型的轻量化;同时采用代价敏感的损失函数对模型进行优化,以解决样本类别不平衡的问题。在 LCM 的基础上嵌入卷积注意力机制,得到基于卷积注意力的轻量级模型(LCMA)。实验结果表明,相比 LCM,LCMA 在性能不变的情况下识别效果有明显提高;且在检测效果相当的情况下,LCMA 在参数规模、内存占用等性能方面具有一定优势。

2 模型结构

2.1 分解卷积

受 Google 的 Inception 模型^[10]结构的启发,本模型将一个由较大滤波器构成的卷积层分解为由多个小滤波器构成的卷积层,以降低计算复杂度。就计算复杂度而言,较大滤波器构成的卷积层计算成本往往较高。如滤波器为 5×5 卷积层的计算成本是滤波器为 3×3 卷积层的 2.8 倍($25 : 9$)。虽然 5×5 卷积层比 3×3 卷积层提取到的特征更多,但用两个串联的 3×3 卷积层对一个 5×5 卷积层进行分解,不会明显降低卷积层对特征的提取能力,同时还可以节省 28% 的计算量。分解前后网络的感受野是一样的,而两个级联滤波器的数据表征能力会更强;且分解后两个级联滤波器多使用了一个激活函数,也会增加模型的非线性表达能力。为了有效减少模型参数,在级联卷积之前再加入一个 1×1 卷积层,使最终输出的特征矩阵大小不变,同时可以降低卷积层的参数量。

2.2 改进的卷积注意力机制模块

Woo 等^[12]提出的卷积注意力机制模块(CBAM)是一种用于前馈 CNN 的注意力模块。给定一个输入特征图,CBAM 会沿着输入特征的通道维度和空间维度依次推断注意力图,然后将注意力图与输入特征图相乘,以进行自适应特征优化。输入的特征图分别经过基于 height 和 width 的平均值池化和最大值池化,平均值池化对特征图上的每一个像素点都有反馈,而最大值池化在梯度反向传播计算时只在特征图中响应最大的地方有梯度反馈。由于 CBAM 是轻量级的通用模块,因此可忽略该模块的开销,将其无缝集成到任何 CNN 架构中,同时可以与基础 CNN 一起进行端到端训练。

在表情识别任务中,不同的表情与特定的区域

相关,面部动作编码系统^[11]定义了与面部表情相关的 46 个动作单元,每个动作单元由特定面部肌肉群的运动构成,一个面部表情由若干个动作单元组成。模型更需关注与当前表情相关性最高的局部特征,因此,对输入特征图进行基于 height 和 width 的最大值池化。空间域特征可以准确描述面部表情特征,而通道域容易受色彩、光照等因素的影响产生偏差,因此,空间域特征提取的优先级高于通道域特征提取。改进的 CBAM 结构如图 1 所示,其中,avg 和 max 分别表示平均值池化和最大值池化,MLP 为多层感知器。输入特征图首先经过空间注意力模块进行特征提取,并将经过基于 height 和 width 的最大值池化输入连接起来送入卷积层中;然后将输出特征送入通道注意力模块,利用特征的通道关系生成通道注意力图,通过共享权重压缩特征图的空间维度;最后用平均值池化和最大值池化聚焦空间信息。

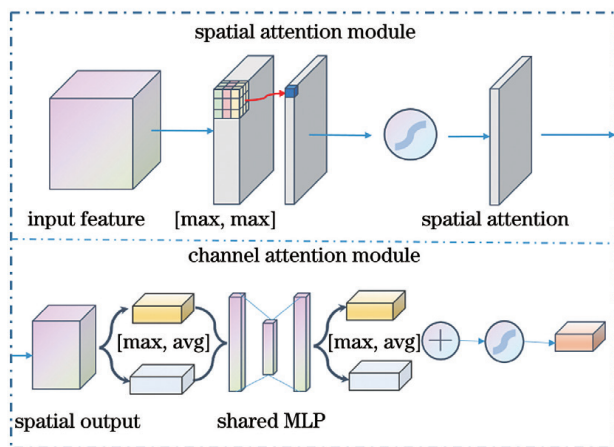


图 1 改进的 CBAM
Fig. 1 Improved CBAM

2.3 基于注意力机制的轻量级卷积模型

与大多数 CNN 模型不同,LCM 采用了流式卷积结构和多尺度编码策略增强图形特征的表达能,从而突出各种表情之间的面部形状差异。由于多种 CNN 特征具有旋转不变性,LCM 也具有很好的鲁棒性以应对姿态变化。LCM 采用基于高斯的加权函数,其均值和方差旨在模拟人类视觉机制,在 LCM 的适当位置添加 CBAM 就构成了 LCMA,如图 2 所示。其中,Conv 为卷积操作,Conv2D 为二维卷积操作,FC layer 为全连接层,Pool 为池化操作。可以发现,LCMA 共有 3 个流式卷积层,每个卷积层有四个不同比例的分解卷积链,以实现多尺度表示。输入样本数据依次通过 3 个流形卷积层后输出的特征可用于分类。两个全连接层后的

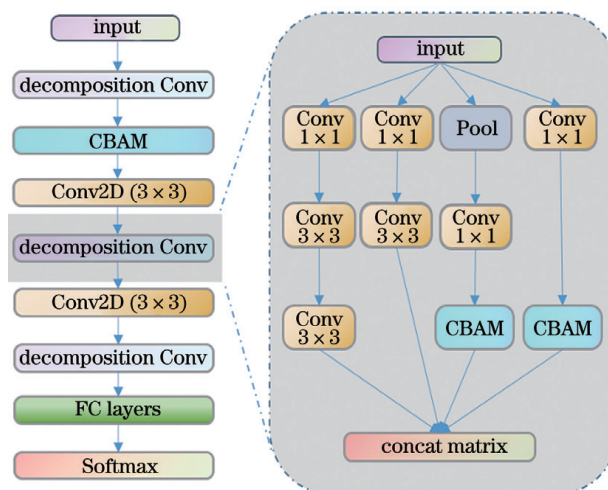


图 2 LCMA 的结构
Fig. 2 Structure of the LCMA

输出由 Softmax 函数激活。最后,通过最小化交叉熵损失训练 LCMA。

2.4 迁移学习

神经网络在图像识别领域的鲁棒性和准确率较高,但训练一个有效的网络模型往往需要大量带有标签的数据,而人脸表情识别领域现有的带标签数据集不能满足如 VGG (Visual geometry group) 等大型 CNN 的数据需求,这也是制约人脸表情识别,特别是复杂场景下表情识别技术发展的重要因素。

迁移学习指将在某个任务中训练的模型应用到另一个任务中,虽然人脸表情识别领域缺少数据集,但人脸识别领域却拥有大量的可用数据集,且相关人脸识别模型的识别准确率较高,如 VGGFace 模型。人脸识别任务在特征提取方面与表情识别有很大的相似性,因此可用迁移学习方法,将模型在人脸识别数据集上进行预训练,得到人脸特征提取器;然后将特征提取器应用在人脸表情识别相关数据集上进行二次训练,以适应人脸表情特征提取任务。结合迁移学习方法可使模型适应小数据集下的人脸表情识别任务。

2.5 代价敏感的损失函数

各类机器学习任务的数据集中均存在类别不平衡问题,如经典的 CK+ 数据集包含 123 名受试者的 7 类基本表情图像 951 张:惊讶表情图像 249 张,高兴表情图像 207 张,厌恶表情图像 177 张,愤怒表情图像 135 张,悲伤表情图像 84 张,恐惧表情图像 75 张,中性表情图像 24 张。数据集中惊讶类别的图像数量约为中性表情图像数量的 10 倍,因此,解决数据集中的类别不平衡问题有助于提高模型的

类准确性。常规的神经网络损失函数可表示为

$$E(\mathbf{w}, \mathbf{b}) = \frac{1}{N} \sum_{n=1}^N L_{\text{loss}}(y_n, \hat{y}_n), \quad (1)$$

式中, \mathbf{w} 、 \mathbf{b} 均为神经网络的参数矩阵, N 为参与训练的样本数量, $L_{\text{loss}}(\cdot)$ 为交叉熵等误差函数, y_n 为第 n 个样本的真实标签, \hat{y}_n 为第 n 个样本的预测标签。神经网络训练的目的在于通过误差反向传播最小化 $E(\mathbf{w}, \mathbf{b})$, 从而得到最优的 \mathbf{w} 、 \mathbf{b} 。

考虑到类别样本不平衡现象, 样本较多的类别在损失函数中产生的影响远大于样本较少的类别, 导致分类器总是倾向样本较多的类别进行预测。分别考虑每个类别的损失函数能解决该问题, 分类别赋权的损失函数可表示为

$$E(\mathbf{w}, \mathbf{b}) = \frac{1}{N} \left\{ \sum_{m=1}^M \sum_{n=1}^{N_m} L_{\text{loss}} [y_m^n, \hat{y}_m^n(\mathbf{w}, \mathbf{b})] \right\}, \quad (2)$$

式中, $\sum_{n=1}^{N_m} L_{\text{loss}} [y_m^n, \hat{y}_m^n(\mathbf{w}, \mathbf{b})]$ 为第 m 类中所有样本的总体损失, y_m^n 和 \hat{y}_m^n 分别为第 m 类中第 n 个样本的真实标签和预测标签, N_m 为第 m 类样本的数量。分别计算每个类别样本的平均误差, 再将所有类别的平均损失值合并, 构成最终的损失函数, 以消除类别不平衡对模型产生的负面影响, 重新定义的模型损失函数可表示为

$$\bar{E}(\mathbf{w}, \mathbf{b}) = \sum_{m=1}^M E_m, \quad (3)$$

式中, E_m 为第 m 个类别的损失。

3 实验结果及分析

3.1 数据集介绍

1) VGGFace2 数据集是一个大规模人脸识别数据集, 数据集中的图像均通过搜索引擎下载得到, 包含 9131 个人的 331 万张人脸图像, 涵盖了大量姿态以及不同年龄、种族以及光照背景。该数据集在人脸识别任务上的效果较好, 利用该数据进行预训练可获得一个有效的人脸特征提取器。

2) AffectNet 数据集包含超过 40 万张从自然环境中采集的图像, 通过人工标注为这些图像打上了 7 个基本情绪类别标签。由于测试集标签不公开, 实验仅用该数据集的训练集和验证集样本。AffectNet 数据集的数据分布最佳^[13], 因此, 也被作为预训练数据集的一部分, 使模型能学习如何提取与表情识别相关的面部特征。

3) FER2013 数据集是由 Google 图像搜索接口收集的大型数据库, 所有图像均已被标记且尺寸均

被调整为 48 pixel×48 pixel。该数据集包含 28709 张训练图像, 3589 张验证图像和 3589 张测试图像, 每张图像都被标记了 7 个基本表情之一作为标签。但该数据集注释不严谨, 存在较多的错误标签。

4) FER+ 数据集是在 FER2013 数据集的基础上通过重新标注和数据补齐得到。重新标注该数据集时使用众包标记技术, 每个图像至少由 10 个不同的注释器标记, 为其赋予 7 种基本表情中的 1 种作为标签, 该数据集的标签比 FER2013 数据集更准确。

3.2 评价指标

常用的机器学习模型评价指标包括准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall)、F1 分数 (F1-score), 考虑到数据集中样本的不平衡问题, 用几何均值 (G-mean) 评价不平衡数据模型的表现指标, 可表示为

$$A_{\text{G-mean}} = \sqrt{R_{\text{Recall}} \times S_{\text{Specificity}}}, \quad (4)$$

式中, $S_{\text{Specificity}}$ 为特异度, 可表示为

$$S_{\text{Specificity}} = \frac{T_{\text{TN}}}{T_{\text{TN}} + T_{\text{FP}}}, \quad (5)$$

式中, T_{TN} 为真阴性数, T_{FP} 为假阳性数。

3.3 预训练模型

数据集中的图像特别是自然环境下采集的图像包含很多与面部表情无关的信息, 如不同的背景、光照和头部姿势。因此, 在训练模型学习有意义的特征之前, 需进行预处理, 对人脸传达的视觉语义信息进行对齐和规正。首先, 用多任务 CNN (MTCNN) 算法进行人脸检测, 去除模糊不标准的人脸样本; 然后, 用 dlib+openCV 进行人脸对齐, 并将图像转换为灰度图, 根据 openCV 内置的仿射变换, 采取 5 个关键点检测法将人脸对齐; 最后, 将人脸图像的尺寸裁剪为 48 pixel×48 pixel。

本模型适用于人脸表情识别领域, 是一个 7 分类模型, 因此用 VGGFace2 人脸识别数据集进行第一次预训练之前, 首先, 将该模型的输出层修改为与 SE-ResNet-50 (Squeeze-and-excitation ResNet-50) 模型^[14] 相同的结构, 使其适应人脸识别问题, 并具备人脸特征提取能力。然后, 将模型的输出层结构修改为表情识别的 7 分类结构。最后, 使用 AffectNet 数据集对模型进行二次预训练, 使模型具备人脸表情特征提取能力。后续在不同的数据集上进行表情识别任务时, 先将预训练模型在该数据集上进行小规模训练, 微调模型的参数以适应该数据集。

3.4 实验对比

实验的硬件环境:处理器为 Intel i5-7400,显卡为 GTX 960,内存为 16 GB。实验中每轮训练的批大小为 128,优化器选择 Adam 算法,dropout 为 0.35。

3.4.1 预训练有效性验证

为了验证使用迁移学习方法的有效性,将本模型经过 VGGFace2 和 AffectNet 数据集的不同组合进行预训练,得到该模型在 FER2013 和 FER+ 数据集上的 G-mean 指标如图 3 所示。可以发现,本模型在 VGGFace2 和 AffectNet 数据集进行预训练后的 G-mean 指标均取得了最佳实验结果。观察曲线的斜率发现,相比 AffectNet 数据集,本模型在

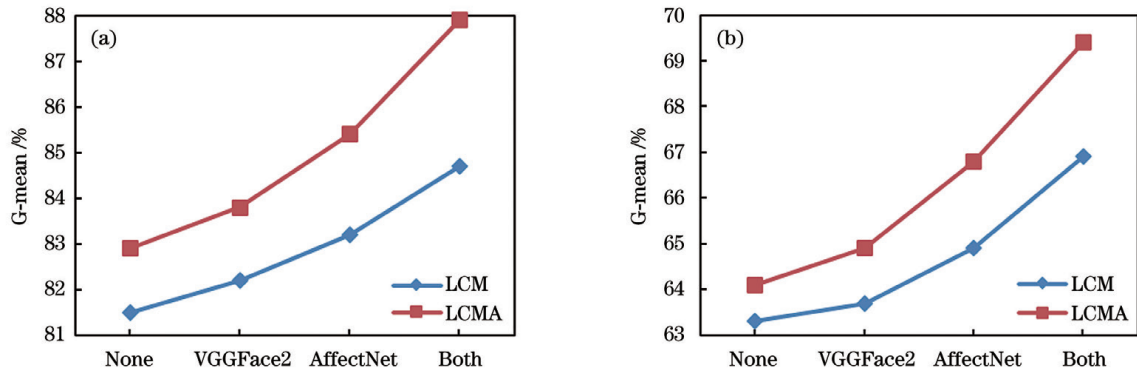


图 3 预训练前后模型的性能。(a) FER+数据集;(b) FER2013 数据集

Fig. 3 Performances of the model before and after pre-training. (a) FER+ data set; (b) FER2013 data set

3.4.2 模型检测能力的对比

为了分析模型的检测能力,用经过 VGGFace2 和 AffectNet 数据集预训练的 LCM 和 LCMA 模型开展对比实验。为了对比模型在 FER2013 数据集上的表现,选取表情识别领域中引用较多的 AlexNet 模型^[15]、基于方向梯度特征的 CNN (HOG+CNN)模型^[16]、交互性端到端深度神经网络 (FaceLiveNet)模型^[17]和 Shallow CNN (SHCNN)模型^[18];同时用基于通道和空间注意力的 CNN (CSACNN)模型^[4]、基于注意力机制的 CNN (CSGE)模型^[19]作为参照,实验结果如表 1 所示。

在 FER+数据集上对比了 SHCNN 模型、CSACNN 模型、CSGE 模型、ESR-9 (Ensembles with Shared Representation based on 9 convolutional branches)模型^[20]、PLD 模型^[9]、18 层残差网络 (ResNet18)模型^[21]和 TFE (Tandem facial expression)模型^[21]的性能,实验结果如表 2 所示。可以发现,相比现有模型,本模型在检测准确率方面略有优势,而在 G-mean 指标上有明显优势。原因是本模型采用了代价敏感的损失函数优化模型,

VGGFace2 数据集上预训练后的效果提升不明显,同时在 VGGFace2 和 AffectNet 数据集上进行预训练后,本模型的性能有较大提升。原因是 VGGFace2 数据集与表情识别分属不同领域,其特征不能直接复用,而将其在 AffectNet 数据集进一步训练时模型拟合了数据的共有特征。还可以发现,相比 LCM,添加 CBAM 的 LCMA 检测效果有明显提升,原因是 CBAM 能增强模型的特征提取能力,且随着数据量的增加,模型的表现也更好。这表明用 VGGFace2 和 AffectNet 数据集对模型进行预训练能有效提升模型的表现,且卷积注意力机制对模型特征提取能力的提升也有很大的帮助。

表 1 不同模型在 FER2013 数据集上的实验结果

Table 1 Experimental results of different models on the FER2013 data set unit: %

Model	Accuracy	G-mean
SHCNN	69.1	64.7
AlexNet	61.1	55.1
HOG+CNN	61.9	56.4
FaceLiveNet	68.6	64.1
CSACNN	68.9	65.2
CSGE	65.8	59.8
LCM	67.2	65.9
LCMA	69.6	68.4

能解决类别不平衡问题;且在类别不平衡数据集上用 G-mean 作为评价指标比准确率更合理,这表明本模型的鲁棒性更强。

3.5 模型性能对比

评价深度学习模型的优劣时还需兼顾模型的迭代速度、检测速度、参数规模和内存占用等指标。

表 2 不同模型在 FER+数据集上的实验结果
Table 2 Experimental results of different models on the FER+ data set

Model	Accuracy	G-mean
SHCNN	86.5	84.9
ESR-9	87.2	85.6
PLD	84.9	83.1
ResNet18	83.1	81.1
TFE	84.3	82.5
CSACNN	86.1	84.2
CSGE	83.8	81.4
LCM	84.5	84.1
LCMA	87.4	86.9

基于 FER2013 数据集,对比了模型的迭代速度、参数规模、运行时的内存占用,结果如表 3 所示。由于对比模型中并未提及模型的性能指标,因此将这些模型按照原论文中的参数设置进行复现。对比发现,LCMA 的多个指标均保持在较低水平,原因是分解卷积的轻量级模型可显著降低模型的参数量,加快模型的运行速度,同时也保持了模型的检测效果。相比 LCM,LCMA 在添加卷积注意力机制后,各项性能指标无显著下降,但检测效果却有明显提高,这表明 CBAM 是有效且轻量的。LCMA 的检测准确率与其他模型相当,但 G-mean 指标明显更优。这表明本模型在参数规模、迭代速度和内存占用方面远低于其他模型,在实际应用中可以节省更多的成本。

表 3 不同模型在 FER2013 数据集上的性能参数
Table 3 Performance parameters of different models on the FER2013 data set

Model	Parameter size /MB	Iteration speed /min	Memory usage /GB
SHCNN	59	43	6.2
AlexNet	62	42	6.5
PLD	138	67	11.8
TFE	53	55	5.4
FaceLiveNet	73	49	7.4
ESR-9	89	71	8.6
LCM	25	31	3.7
LCMA	26	31	3.8

4 结 论

提出了一种基于卷积注意力的轻量级人脸表情识别方法,通过分解卷积有效降低了模型的参数规模,级联卷积的设计使模型经过分解后依然能保持较高的检测准确率;针对数据集中类别不平衡的问题,采取了代价敏感的损失函数。对比实验结果表明,本模型对于硬件设备的配置要求较低,可运行于如网络摄像头等小型设备上。下一步工作还将聚焦提升模型检测效果和减少模型的参数,一是通过改进迁移学习方法提高模型的特征提取能力;二是对注意力机制模块进行优化,降低模块的复杂度。

参 考 文 献

- [1] Tian Y I, Kanade T, Cohn J F. Recognizing action units for facial expression analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(2): 97-115.
- [2] Won C. Recognition of facial emotion using multi-scale LBP[J]. Journal of Korea Multimedia Society, 2014, 17(12): 1383-1392.
- [3] 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.
- [4] Chu J H, Tang W H, Zhang S, et al. An attention model-based facial expression recognition algorithm [J]. Laser & Optoelectronics Progress, 2020, 57(12): 121015.
褚晶辉, 汤文豪, 张姍, 等. 一种基于注意力模型的面部表情识别算法[J]. 激光与光电子学进展, 2020, 57(12): 121015.
- [5] Yang X, Shang Z H. Facial expression recognition based on improved AlexNet[J]. Laser & Optoelectronics Progress, 2020, 57(14): 141026.
杨旭, 尚振宏. 基于改进 AlexNet 的人脸表情识别 [J]. 激光与光电子学进展, 2020, 57(14): 141026.
- [6] Chen W D, Hu H F. Joint prominent expression feature regions in auxiliary task learning network for facial expression recognition[J]. Electronics Letters, 2019, 55(1): 22-24.
- [7] Ding H, Sricharan K, Chellappa R. Exprgan: facial expression editing with controllable expression intensity [EB/OL]. (2017-09-12) [2020-08-25]. <https://arxiv.org/abs/1709.03842v1>.

- [8] Ding H, Zhou S K, Chellappa R. FaceNet2ExpNet: regularizing a deep face recognition net for expression recognition[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: 118-126.
- [9] Barsoum E, Zhang C, Ferrer C C, et al. Training deep networks for facial expression recognition with crowd-sourced label distribution[EB/OL]. (2016-09-24) [2020-08-23]. <https://arxiv.org/1608.01041>.
- [10] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2818-2826.
- [11] Ekman P, Friesen W V. A technique for the measurement of facial action [J]. Facial Action Coding System (FACS), 1978, 13(9): 102-108.
- [12] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [13] Barros P, Churamani N, Sciutti A. The FaceChannel: a light-weight deep neural network for facial expression recognition [C]//2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), November 16-20, 2020, Buenos Aires, Argentina. New York: IEEE Press, 2020: 652-656.
- [14] Cao Q, Shen L, Xie W D, et al. VGGFace2: a dataset for recognising faces across pose and age[C]//2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), May 15-19, 2018, Xi'an, China. New York: IEEE Press, 2018: 67-74.
- [15] 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.
- [16] Zeng G H, Zhou J C, Jia X, et al. Hand-crafted feature guided deep learning for facial expression recognition[C]//2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), May 15-19, 2018, Xi'an, China. New York: IEEE Press, 2018: 423-430.
- [17] Ming Z H, Chazalon J, Luqman M M, et al. FaceLiveNet: end-to-end networks combining face verification with interactive facial expression-based liveness detection[C]//2018 24th International Conference on Pattern Recognition (ICPR), August 20-24, 2018, Beijing, China. New York: IEEE Press, 2018: 3507-3512.
- [18] Miao S, Xu H Y, Han Z Q, et al. Recognizing facial expressions using a shallow convolutional neural network[J]. IEEE Access, 2019, 7: 78000-78011.
- [19] Kang J, Li S Y. Facial expression recognition based on attention mechanism of convolutional neural network [J]. Journal of Shaanxi University of Science & Technology, 2020, 38(4): 159-165, 171. 亢洁, 李思禹. 基于注意力机制的卷积神经网络人脸表情识别 [J]. 陕西科技大学学报, 2020, 38(4): 159-165, 171.
- [20] Siqueira H, Magg S, Wermter S. Efficient facial feature learning with wide ensemble-based convolutional neural networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 5800-5809.
- [21] Li M, Xu H, Huang X C, et al. Facial expression recognition with identity and emotion joint learning [J]. IEEE Transactions on Affective Computing, 2018, 14: 1-9.