

doi:10.3969/j.issn.1673-9833.2022.06.006

基于 ACNN 和 Bi-LSTM 的微表情识别

朱文球, 李永胜, 黄史记, 阳昊彤

(湖南工业大学 计算机学院, 湖南 株洲 412007)

摘 要: 针对微表情动作幅度小、强度低等缺点, 提出了一种基于带有注意力机制的卷积神经网络 (ACNN) 和双向长短期记忆网络 (Bi-LSTM) 相结合的神经网络结构。实验采用 CASME II 数据集, 为了减少出现过拟合的风险, 首先将预处理后的特征向量经过预训练的 VGG16 网络提取出基本特征, 接着对输出特征进行裁剪, 得到带有局部特征的 24 个微表情识别块和带有整个图片特征的全局特征向量; 然后将 24 个识别块分别经过局部识别块注意力卷积神经网络 (BR-ACNN) 提取出带有注意力信息的局部特征, 将全局特征向量经过全局注意力卷积神经网络 (GR-ACNN) 提取出带有注意力信息的全局特征; 最后, 将提取的局部和全局特征, 经过 Bi-LSTM 提取出微表情序列之间的相关性信息。实验结果显示, 5 折交叉验证平均准确率为 0.69, UF_1 为 0.638 2, UAR 为 0.675 0。CASME II 数据集上结果显示, 所提算法模型相对 OFFApexNet 模型, 其 UF_1 提高了 0.028 1, UAR 提高了 0.096 9; 相对 ATNet 模型, 其 UF_1 提高了 0.007 2, UAR 提高了 0.032 0。

关键词: 微表情识别; 长短时记忆网络; 注意力网络; 迁移学习; 识别块

中图分类号: TP391.41

文献标志码: A

文章编号: 1673-9833(2022)06-0034-08

引文格式: 朱文球, 李永胜, 黄史记, 等. 基于 ACNN 和 Bi-LSTM 的微表情识别 [J]. 湖南工业大学学报, 2022, 36(6): 34-41.

Micro-Expression Recognition Based on ACNN and Bi-LSTM

ZHU Wenqiu, LI Yongsheng, HUANG Shiji, YANG Haotong

(College of Computer Science, Hunan University of Technology, Zhuzhou Hunan 412007, China)

Abstract: In view of the flaws of micro-expression characterized with a small amplitude and low intensity, a neural network structure has thus been proposed based on the combination of convolution neural network with attention mechanism (ACNN) and bi-directional long short-term memory (Bi-LSTM). CASME II data set has been adopted in the experiment so as to reduce the risk of over-fitting, with the basic features extracted from the preprocessed feature vectors through the pre-trained VGG16 network, followed by the cropping of the output features, thus obtaining 24 micro-expression recognition blocks with local features and global feature vectors with the whole picture features. Next, based on an extraction of local features with attention information from 24 recognition blocks through local recognition block attention convolution neural network (BR-ACNN), global features with attention information are to be extracted as well from global feature vectors through global attention convolution neural network (GR-ACNN). Finally, the correlation information between the micro expression sequences can be extracted by Bi-LSTM based on the extracted local and global features. The experimental results show that the average accuracy rate of 5-fold cross validation is 0.69,

收稿日期: 2021-12-02

基金项目: 国家重点研发计划基金资助项目 (2018AAA0100400); 湖南省自然科学基金资助项目 (2021JJ50058); 湖南省教育厅开放平台创新基金资助项目 (20K046); 湖南省战略性新兴产业科技攻关与重大科技成果转化基金资助项目 (2019GK4009)

作者简介: 朱文球 (1969-), 男, 湖南株洲人, 湖南工业大学教授, 主要研究方向为智能信息处理与生物特征识别,
E-mail: ZWQ@hut.edu.cn

UF_1 is 0.638 2, and UAR is 0.675 0. The results on the CASME II data set show that the proposed algorithm model, compared with OFFApexNet model, is 0.028 1 higher in UF_1 , and 0.096 9 higher in UAR ; while compared with ATNet model, it has increased by 0.007 2 in UF_1 and by 0.032 0 in UAR .

Keywords: micro-expression recognition; long and short term memory network; attention mechanism; transfer learning; identification block

0 引言

相对于普通表情,微表情是一种持续时间短并且动作幅度小的表情变化。微表情是人们在试图掩盖内在情绪时产生的一种自发性表情,是一种既无法伪造也无法抑制的表情^[1-4]。1966年,Haggard等发现了一种快速且不易被人眼所察觉的面部表情,并首次提出了微表情的概念^[2]。起初这种微小且短暂的面部变化并没有引起其他同行研究人员的注意。直至1969年,P. Ekman等^[5]在研究抑郁症患者的一段录像时,发现处于微笑表情表现的病人会出现短暂的痛苦表情,该患者用微笑等积极性表情来掩饰内心的焦虑^[3]。和宏表情不同的是,微表情的持续时间仅为1/25~1/5 s,因此仅凭人眼对微表情进行识别并不能满足精确辨别的需要,因而使用现代人工智能手段识别微表情是非常有必要的。

T. Pfister等^[6]利用三维正交平面局部二值(local binary patterns from three orthogonal planes, LBP-TOP)算法,将特征提取方式由XY方向扩展到XY、XT和YT组成的3个正交平面上,即将以往的静态特征提取拓展到随时间信息变化的动态特征提取。Peng M.等^[7]提出了一种深度学习的自动抠图(ATNet)方法,即利用ResNet10从顶点帧中提取出微表情空间信息,再利用长短期记忆网络(long short-term memory, LSTM)从相邻帧中提取出微表情时间信息,并且在特征融合后进行识别。Liu J.等^[8]提出了利用原始灰度图训练空间注意力模型SAIN_GRAY,但是该模型并没有考虑图像间的序列相关性。Quang N. V.等^[9]利用胶囊网络(CapsuleNet)进行了微表情识别,其首先提取出微表情的顶点帧,接着把顶点帧输入胶囊网络中提取空间信息,最后分类输出,这种方式可以比较有效地提取出空间信息,但是忽略了微表情的时间相关性。

1 相关工作

1.1 面部表情编码系统

面部表情编码系统(facial action coding system,

FACS)是P. Ekman等^[5]在1976年提出的。微表情有两种分类标准:基于情绪的分类和基于面部动作编码系统的分类^[10]。FACS依据解剖学按区域将人脸划分为44个运动单元(action unit, AU),每个动作单元表示一个特定的面部局部动作,如动作单元1表示眉毛内侧上扬(inner brow raiser),而动作单元5表示拉动上嘴唇向上运动(upper lip raiser)等。人脸面部有42块肌肉,丰富的表情变化是由多种肌肉共同作用的结果。有些面部肌肉可以受意识控制,被称为“随意肌”;还有一些面部肌肉不受意识控制,被称为“不随意肌”。微表情的产生通常是由一个或多个运动单元(AU)共同作用的结果,如代表高兴的微表情是由AU16和AU12共同作用的结果,其中AU16表示下嘴唇向下拉动,AU12表示嘴角上扬。FACS是微表情识别的重要依据,是对眉毛、脸颊与嘴角等面部关键点特征的运动记录。

1.2 带有注意力机制的卷积神经网络

为了解决微表情识别中的持续时间短、动作强度低等缺点,本文采用基于带有注意力机制的卷积神经网络(convolution neural network with attention mechanism, ACNN)的改进算法。ACNN是在卷积神经网络的基础上添加了注意力网络,使得ACNN模型在微表情识别中不仅可以提取整个面部的特征,还可以专注于表情的局部细节变化,从而提取出更细微的面部特征。课题组基于ACNN算法,并对其进行了改进,即增加了局部注意力网络、全局注意力网络,以及双向长短期记忆网络(bi-directional long short-term memory, Bi-LSTM),以适用微表情视频序列的识别。CNN具有提取出微表情抽象特征的能力,带有局部注意力网络的ACNN用于提取微表情变化中关键局部单元的运动信息,而带有全局注意力网络的ACNN可以提取全局的变化信息。本文中将局部ACNN和全局ACNN相结合,期望改进后的ACNN模型具有既关注全局又兼顾细节的能力。

1.3 双向长短期记忆网络

普通的卷积神经网络以及全连接层都有一个共同的特点,即在处理具有连续性的序列时都不能“记

忆”时序间的相关信息。循环神经网络 (recurrent neural network, RNN) [11] 新增了一个可以保存状态信息的隐藏层, 这个隐藏层包括了序列输入的某些历史信息, 并随着输入序列的增加不断自我更新。然而普通循环神经网络最大的缺点, 是随着训练规模与层数的增大很容易产生长期依赖问题 (long-term dependencies problem), 即在对较长序列学习时容易产生梯度消失和梯度爆炸等问题。为了解决循环神经网络的上述问题, 20 世纪 90 年代初, Hochreiter 等提出了长短期记忆网络 (LSTM) [12]。LSTM 中每一个单元块包括输入门 (input gate)、遗忘门 (forget gate) 和输出门 (output gate)。其中, 输入门用来决定当前时刻的输入数据有多少可以保存到本状态单元; 遗忘门用来表示上一时刻的状态单元有多少可以保存到本状态单元; 输出门控制当前状态单元有多少可以用来输出。双向长短期记忆网络 (Bi-LSTM)

在单向 LSTM 中增加反向传播层, 使得 LSTM 模型不仅可以利用历史序列信息还可以利用未来的信息。Bi-LSTM 相对于 LSTM 可以更好地提取出微表情中的特征信息与序列信息。

2 本文方法介绍

2.1 方法概述

本文提出了一种基于带有注意力机制的卷积神经网络 (ACNN) 和双向长短期记忆网络 (Bi-LSTM) 相结合的神经网络结构。为了实现对微表情较小幅度的动作进行精准地捕捉, 本文对传统的 ACNN 框架进行了改进。改进后的 ACNN 框架会对面部多个微表情区域块提取出不同的特征信息, 同时除了微表情区域块的特征提取, 也增加了对全局信息的处理。改进后的模型架构如图 1 所示。

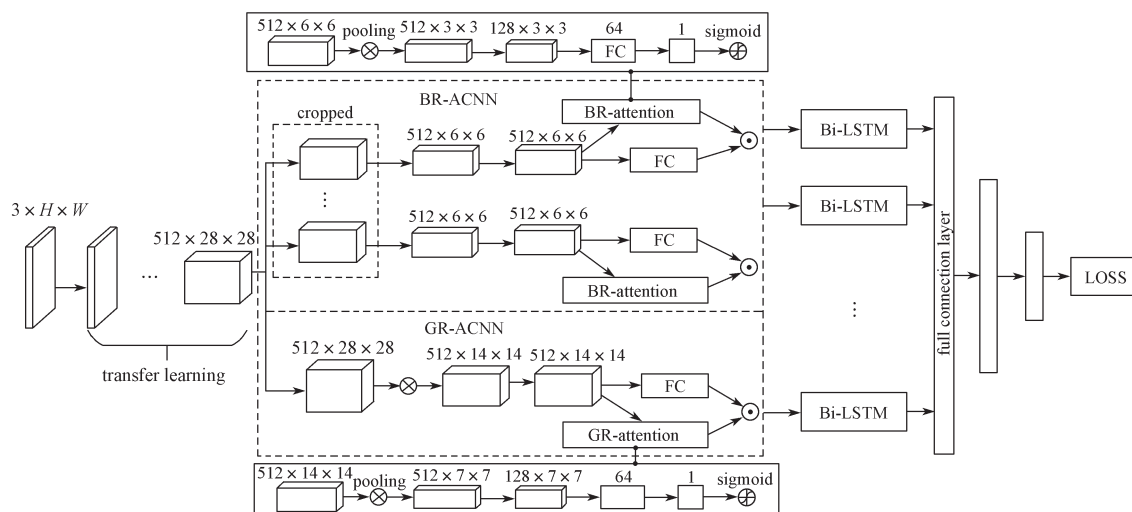


图 1 基于 ACNN 和 Bi-LSTM 的微表情识别模型

Fig. 1 Micro-expression recognition model based on ACNN and Bi-LSTM.

改进后的模型中, 网络首先将预处理后的微表情特征向量通过带有预训练权重的 VGG16 [13] 模型进行迁移学习, 并从中提取出基本面部特征, 再将每一帧提取出的面部特征分别经过局部识别块注意力卷积神经网络 (block recognition based convolution neural network with attention mechanism, BR-ACNN) 与全局注意力卷积神经网络 (global recognition based convolution neural network with attention mechanism, GR-ACNN) 提取出整体与局部信息; 接着, 把提取出的整体与局部信息进行融合, 并通过 Bi-LSTM 提取出序列的相关信息, 最后通过 3 层全连接层进行分类输出。

对于全局和局部特征的提取, 本文引入了两种不同的 ACNN, 即局部识别块注意力卷积神经网络

(BR-ACNN) 和全局注意力卷积神经网络 (GR-ACNN)。如图 1 所示, BR-ACNN 为图中虚线框内上半部分, GR-ACNN 为图中虚线框内下半部分。

2.2 局部识别块注意力卷积神经网络

由于微表情面部变化的幅度较小, 很难被有效识别, 故本实验采用面部关键区域块注意力的识别方法, 即在要识别的面部特征中增加微表情关键代表性区域的识别, 以及对应的注意力网络。详细实验方法描述如下。

1) 确定微表情识别点。首先, 使用 Dlib [14] 方法提取出 68 个面部关键点 (如图 2a); 其次在 68 个面部关键点的基础上经过转换, 成为 24 个微表情识别点, 识别点位置覆盖脸颊、嘴巴、鼻子、眼睛、眉毛等处。转换过程如下, 首先从 68 个面部关键点

中选取 16 个覆盖嘴巴、鼻子、眼睛、眉毛的识别点, 68 个面部关键点 (如图 2a) 中的提取序号为: 19、22、23、26、39、37、44、46、28、30、49、51、53、55、59、57。最终生成的微表情识别点 (如图 2b) 序号为: 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16。其次, 对于眼睛、眉毛以及脸颊部分, 由于没有直接的关键点覆盖, 本文通过关键点的中点坐标进行生成。左眼、左眉毛与左脸颊部分从 68 个面部关键点 (如图 2a) 中选取 (20, 38)、(41, 42)、(18, 59) 点对的中点坐标作为微表情识别点; 右眼、右眉毛与右脸颊部分从 68 个面部关键点中选取 (25, 45)、(47, 48)、(27, 57) 点对的中点坐标作为微表情识别点。最终生成的识别点 (如图 2b) 序号为 17、19、18、20、21、22。最后, 对于左右嘴角部分, 本文从 68 个面部关键点 (如图 2a) 中选取 49、55 两个关键点, 然后分别根据这两个点的坐标选择两个嘴角的相对偏移点, 最终得到识别点的坐标。最终生成的左右嘴角识别点 (如图 2b) 序号为: 23、24。式 (1) (2) 分别是左右嘴角识别点的计算方式。

$$(x_{23}, y_{23}) = (x_{\text{left}} - 16, y_{\text{left}} - 16), \quad (1)$$

$$(x_{24}, y_{24}) = (x_{\text{right}} + 16, y_{\text{right}} + 16). \quad (2)$$

式中: x_{left} 、 y_{left} 为序号为 49 关键点的横、纵轴坐标;

x_{right} 、 y_{right} 为序号为 55 关键点的横、纵轴坐标;

x_{23} 、 y_{23} 分别为序号为 23 识别点的横、纵轴坐标;

x_{24} 、 y_{24} 分别为序号为 24 识别点的横、纵轴坐标。

最终得到了 24 个微表情识别点 (图 2b), 重新选取的 24 个微表情识别点在实验中将会生成以识别点为中心的 24 个 6×6 的识别块。为了提高模型的鲁棒性、增大识别块的感受野, 识别块将会作用在经过 VGG16 迁移学习处理后的 $512 \times 28 \times 28$ 特征向量上, 从而得到 24 个识别块, 每个识别块大小为 $512 \times 6 \times 6$ 。

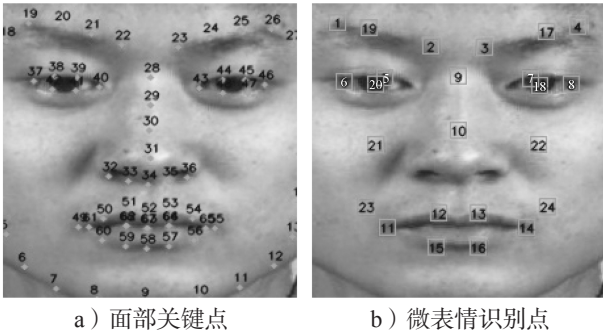


图 2 面部关键点及其处理后的微表情识别点位置图

Fig. 2 Location map of facial key points with their processed micro-expression recognition points

2) 注意力机制算法。带有注意力机制的局部识别块卷积神经网络 (BR-ACNN) 如图 1 虚线框内上半部分所示。在 BR-ACNN 中, 经过裁剪后得到的 24 个局部识别块分别经过一个全连接层和一个输出为权重标量的普通注意力网络, 最后得到 24 个带权特征向量。

在注意力网络中 (图 1 虚线框上半部分), 假设 c_i 表示第 i 个识别块的特征向量, 则可利用式 (3) 计算第 i 个识别块的注意力权重, 利用式 (4) 计算第 i 个识别块提取后且未加权的特征, 利用式 (5) 计算第 i 个识别块带有注意力权重的特征向量。这样, 经过计算后, 便可得到所有识别块的带权特征向量。

$$p_i = \varphi(c_i), \quad (3)$$

$$\tilde{c}_i = \tau(c_i), \quad (4)$$

$$a_i = p_i \cdot \tilde{c}_i, \quad (5)$$

式 (3) ~ (5) 中: p_i 为第 i 个识别块的注意力权重;

c_i 为第 i 个识别块的特征向量;

$\varphi(\cdot)$ 为注意力网络中的操作;

$\tau(\cdot)$ 为对输入特征向量的特征学习;

\tilde{c}_i 为第 i 个识别块提取后且未加权的特征;

a_i 为第 i 个识别块带有注意力权重的特征向量。

2.3 全局注意力卷积神经网络

BR-ACNN 具有学习人脸面部特征细微变化的能力, 然而完成微表情动作有时涉及面部各个部位的相互配合。在特征识别中融合全局特征, 有望提高微表情的识别效果, 所以不仅需要提取面部区块的特征, 还需要提取全局的特征。

全局注意力卷积神经网络 (GR-ACNN) 的详细结构如图 1 虚线框内下半部分所示, 其输入特征向量大小为 $512 \times 28 \times 28$ 。首先, 本文将输入的全局特征向量通过 VGG16 网络的 conv4_2 层到 conv5_2 层, 得到输出大小为 $512 \times 14 \times 14$ 的特征向量; 其次, 将大小为 $512 \times 14 \times 14$ 的特征向量, 分别经过一个全连接层和一个输出为权重标量的注意力网络; 最后, 输出带权重的全局特征向量。

2.4 双向长短期记忆网络

RC-ACNN 和 RG-ACNN 可以提取出微表情某一帧局部和整体的信息, 然而微表情视频帧之间在连续时间内是动态变化的, 因此还需要提取微表情时间上的变化信息。长短期记忆网络 (LSTM) 是一种为了克服传统循环神经网络 (RNN) 中的长期依赖问题而设计出来的新结构。而 Bi-LSTM 是在 LSTM 的基础上增加反向层, 使得新的网络结构不仅能够利用已经输入的历史信息能力, 还具有捕获未来的可用信息

的能力。

双向循环神经网络的结构如图3所示,图中 w 为学习权重。

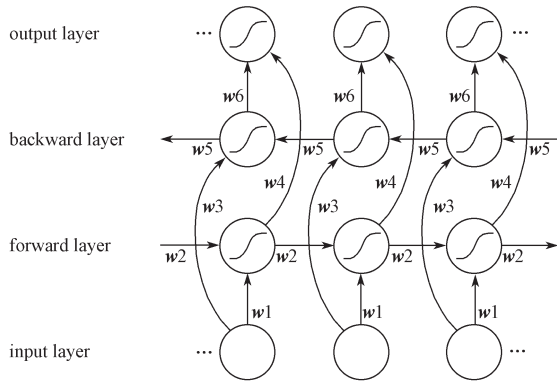


图3 双向循环神经网络结构图

Fig. 3 Structure diagram of bidirectional recurrent neural network

如图3所示,双向长短期记忆网络(Bi-LSTM)结构把双向循环神经网络每个节点换成LSTM单元。定义Bi-LSTM网络模型输入特征序列为 $X=(x_1, x_2, \dots, x_t)$,前项传播中隐藏层变量序列为 $\bar{h}=(\bar{h}_1, \bar{h}_2, \dots, \bar{h}_t)$,反向传播中隐藏层变量序列为 $\bar{h}=(\bar{h}_1, \bar{h}_2, \dots, \bar{h}_t)$,定义Bi-LSTM模型输出序列为 $y=(y_1, y_2, \dots, y_t)$,可得出以下公式:

$$\bar{h}_t = S(w_{xh}x_t + w_{hh}\bar{h}_{t-1} + b_h), \quad (6)$$

$$\bar{h}_t = S(w_{xh}x_t + w_{hh}\bar{h}_{t-1} + b_h), \quad (7)$$

$$y_t = w_{hy}\bar{h}_t + w_{hy}\bar{h}_t + b_o, \quad (8)$$

式(6)~(8)中: $S(\cdot)$ 为激活函数;

w 为Bi-LSTM的权重;

b 为偏置项。

对每个部分使用LSTM单元进行计算,LSTM单元的网络结构如图4所示。

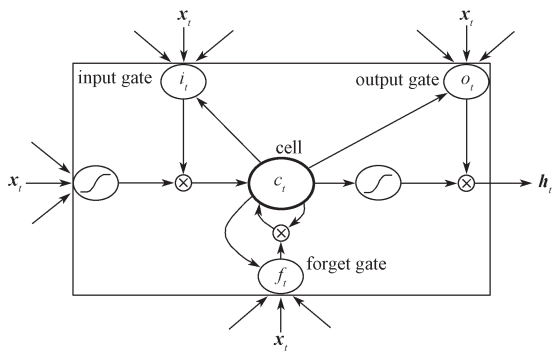


图4 LSTM单元网络结构图

Fig. 4 LSTM cell network structure diagram

LSTM的输入是经过BR-ACNN和GR-ACNN模型从帧序列中提取并进行拼接操作后的空间特征值,

本文采用单层双向的LSTM结构,其中包含一个256个节点的隐藏层。为了增加模型网络节点的鲁棒性,减少神经元之间复杂的共适应关系,在双向LSTM隐藏层和全连接层之间增加了Dropout层,以一定概率随机屏蔽神经元。

3 实验

3.1 数据集的选择与预处理

实验选用中国科学院心理研究所傅小兰团队所建立的CSAME II数据集^[15]。CSAME II微表情数据集采用200帧/s的高速摄像头,每帧大小为 280×340 像素。CSAME II微表情数据集有255个样本,数据集中参与者的平均年龄为22岁,被采集者总数为24个。该数据集包括被采集者每个样本对应的微表情标签,并分别带有起始帧、顶点帧和结束帧视频序列。标签包括压抑、厌恶、高兴、惊讶、害怕、伤心及其他分类(repression, disgust, happiness, surprise, fear, sadness, others)。本文将CSAME II数据集划分为消极、积极、惊讶3类,划分结果如表1所示。

表1 数据集划分情况

Table 1 Data set division

category	quantity	label
negative	251	repression disgust
positive	109	happiness
surprise	86	surprise

本研究首先通过时域插值模型(temporal interpolation model, TIM)^[16]将微表情数据集处理成固定10帧的输入序列,并通过双线性插值^[17]将视频帧归一化为 224×224 像素。接下来使用Dlib^[18]裁剪对齐,并提取出68个面部关键点,在对面部关键点进行实验分析后,本文选取其中的24个微表情面部运动识别点,并由24个识别点生成24个大小为 6×6 的识别块。最后,把经过预处理的视频帧和每一帧的识别块放入模型训练。

3.2 实验评估标准

由于微表情数据集的数量较少,为了保证实验的准确性,本文选用5折交叉验证,即把训练集分为5等份,每次训练选取其中1份作为验证集,其余4份作为测试集,依次类推,循环5次,直至所有数据都做过训练集和验证集。

本文采用了未加权 F_1 值 UF_1 (unweighted F_1 -score),未加权平均召回率 UAR (unweighted average recall),以及准确率 Acc (accuracy)的评

价指标。未加权 F_1 值的计算如式 (9) 所示, 未加权平均召回率的计算如式 (10) 所示, 准确率的计算如式 (11) 所示。

$$UF_1 = \frac{1}{C} \sum_{i=1}^C \frac{2TP_i}{2TP_i + FP_i + FN_i}, \quad (9)$$

$$UAR = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{N_i}, \quad (10)$$

$$Acc = \frac{TP}{TP + FP}. \quad (11)$$

式 (9) ~ (11) 中: TP_i 、 FP_i 、 FN_i 分别为第 i 个类别中真正例、假正例、假负例的数量;

C 为类别数;

N_i 为第 i 个样本的数量;

TP 为所有种类中真正例的数量;

FP 为所有种类中假正例的数量。

3.3 实验结果

训练使用 Adam 优化器, 学习率为 0.000 16, 迭代次数 epoch 设置为 100, 训练 batch_size 设置为 16。因为微表情是小数据集, 容易过拟合, 为了提高模型的鲁棒性与泛化能力, 本文对模型参数取正则化的 L_2 范数, 并在损失函数中加上 λ 倍的 L_2 参数范数, 经过多次实验, 结果表明 λ 设置为 0.01 时模型效果最好。图 5 为 5 折交叉验证的平均损失曲线, 图 6 为 5 折交叉验证的平均准确率曲线, 横坐标为迭代的 epoch 次数, 纵坐标为准确率。

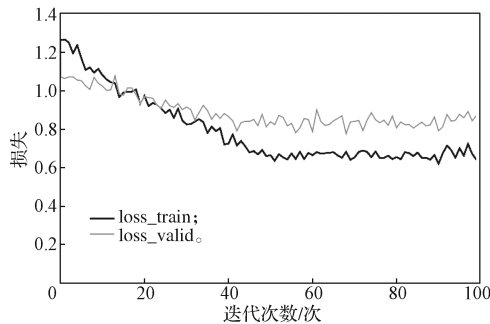


图 5 5 折交叉验证的损失曲线

Fig. 5 5-fold cross-validation loss curves

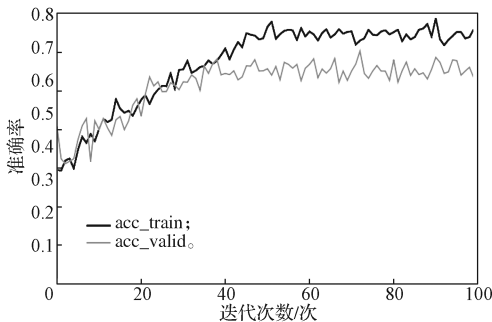


图 6 5 折交叉验证的准确率曲线

Fig. 6 5-fold cross-validation accuracy curves

5 折交叉验证结果的具体准确率数据如表 2 所示。

表 2 验证训练的准确率数据

Table 2 Training accuracy results				%
test 1	test 2	test 3	test 4	test 5
79.3	69.0	62.1	62.1	72.7

表 2 所示实验结果显示, 5 折交叉验证的平均准确率为 69.04%, UF_1 为 0.638 2, UAR 为 0.675 0。最终的混淆矩阵如图 7 所示, 预测结果在“消极”附近分布比较多, 并且准确率较高, 这主要是由数据集分布的不平衡性导致的。因为 CASME II 数据集在采集中“积极”的微表情比较难触发, 所以数据集标签为“消极”的数目远大于标签为“积极”的数目, 这才导致数据集分布不平衡, 从而影响训练精度。

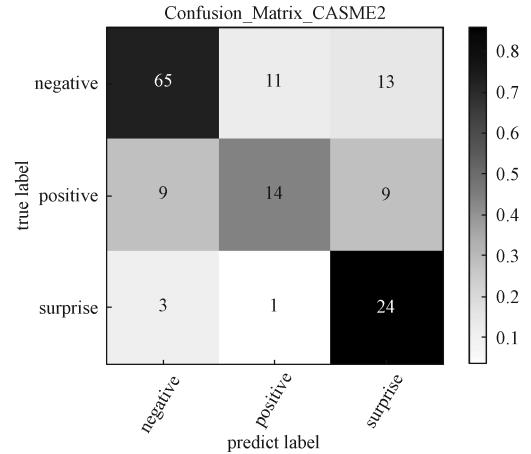


图 7 混淆矩阵图

Fig. 7 Confusion matrix diagram

3.4 数据分析

不同算法的识别效果对比如表 3 所示。

表 3 不同算法识别效果对比

Table 3 Comparison of recognition effects of different algorithms

methods	UF_1	UAR
LBP-TOP ^[19]	0.360 2	0.383 9
VGG-16 ^[13]	0.469 2	0.465 5
AlexNet ^[20]	0.508 3	0.490 1
VGG11 ^[21]	0.531 5	0.538 1
ResNet18 ^[22]	0.536 7	0.544 1
SAIN_GRAY ^[23]	0.545 0	0.527 0
GoogleNet ^[8]	0.607 0	0.574 0
OFFApexNet ^[24]	0.610 1	0.578 1
ATNet ^[7]	0.631 0	0.643 0
our model	0.638 2	0.675 0

表 3 中的数据表明, 本文改进的算法模型相对于以往的识别算法, 其精度有了相对的提高。其中相对 OFFApexNet 模型, 改进算法的 UF_1 提高了 0.028 1, UAR 提高了 0.096 9; 相对 ATNet 模型, 改进算法的 UF_1 提高了 0.007 2, UAR 提高了 0.032 0。这主要是

因为: OFFApexNet 模型依靠起始帧和顶点帧的光流变化进行微表情识别,但忽略了其他连续帧中有用的微表情信息。ATNet 模型利用 ResNet 网络从顶点帧中提取出微表情的空间信息,利用 LSTM 从相邻帧中提取出微表情的时间信息,并在特征融合后进行识别。ATNet 模型虽然考虑到连续帧中空间与时间信息,但是在空间特征提取中并没有兼顾微表情面部不同部位关注度的差异性。

4 结 语

微表情具有持续时间短、动作幅度小等特点,因此微表情数据集较小,容易产生过拟合,针对这一问题,本文在 ACNN 框架的基础上进行了改进,提出了局部注意力网络和双向长短期记忆网络结合的识别方式。首先,将预处理后的微表情视频帧通过 VGG16 提取出特征集合;然后,将提取出的特征集合分别通过局部识别块注意力卷积神经网络(BR-ACNN)和全局注意力卷积神经网络(GR-ACNN)提取出局部与全局特征;接下来,通过双向长短期记忆网络(Bi-LSTM)提取出每一帧的序列信息;最后,通过 3 层全连接层进行分类。实验结果表明,基于识别块注意力网络与双向长短期记忆网络的识别方式,可以从动作幅度小的视频帧里有效地提取出有用信息,并在实验中表现出较高的准确率。但是微表情数据集数量较少,持续时间普遍较短、强度较低,这仍然是导致实验识别率较低的主要原因,这一点在混淆矩阵中尤为明显。

在以后的研究中,对于数据集数量较少的问题,还需要进一步完善数据集,以促进微表情识别的发展;对于微表情强度低的问题,在数据预处理部分下一步,可以通过对视频起始和峰值帧之间的视频帧做线性时域插值以强化峰值帧的作用。此外,微表情识别点的选取可以依据以后的实验再优化,识别点的选取应该尽量具有代表性。

参考文献:

- [1] LI Q Y, YU J, KURIHARA T, et al. Deep Convolutional Neural Network with Optical Flow for Facial Micro-Expression Recognition[J]. Journal of Circuits, Systems and Computers, 2020, 29(1): 2050006.
- [2] LI X B, HONG X P, MOILANEN A, et al. Towards Reading Hidden Emotions: a Comparative Study of Spontaneous Micro-Expression Spotting and Recognition Methods[J]. IEEE Transactions on Affective Computing, 2018, 9(4): 563-577.
- [3] LI Y T, HUANG X H, ZHAO G Y. Joint Local and Global Information Learning with Single Apex Frame Detection for Micro-Expression Recognition[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2021, 30: 249-263.
- [4] LIONG S T, GAN Y S, ZHENG D N, et al. Evaluation of the Spatio-Temporal Features and GAN for Micro-Expression Recognition System[J]. Journal of Signal Processing Systems, 2020, 92(7): 705-725.
- [5] EKMAN P, ROSENBERG E L. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)[M]. 2nd Ed. USA: Oxford University Press, 2005: 3-35.
- [6] PFISTER T, LI X B, ZHAO G Y, et al. Recognising Spontaneous Facial Micro-Expressions[C]//Proceedings of 2011 International Conference on Computer Vision. Barcelona Spain: IEEE, 2011: 1449-1456.
- [7] PENG M, WANG C Y, BI T, et al. A Novel Apex-Time Network for Cross-Dataset Micro-Expression Recognition[C]//2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). Los Alamitos: IEEE Computer Society Press, 2019: 1-6. DOI: 10.1109/ACII.2019.8925525
- [8] LIU J T, ZHENG W M, ZONG Y. SMA-STN: Segmented Movement-Attending Spatiotemporal Network For Micro-Expression Recognition[EB/OL]. (2020-10-19). <https://arxiv.org/abs/2010.09342>.
- [9] QUANG N V, CHUN J, TOKUYAMA T. CapsuleNet for Micro-Expression Recognition[C]//2019 14th IEEE International Conference on Automatic Face & Gesture Recognition. Lille, France: IEEE, 2019: 1-7.
- [10] 程 村. 微表情识别综述 [J]. 计算机时代, 2020(9): 17-19, 23.
CHENG Cun. A Survey of Micro-Expression Recognition[J]. Computer Era, 2020(9): 17-19, 23.
- [11] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent Neural Network Regularization[J]. CoRR, 2014, abs/1409: 2329.
- [12] AOUAYEB M, HAMIDOUCE W, SOLADIE C, et al. Micro-Expression Recognition from Local Facial Regions[J]. Signal Processing: Image Communication, 2021, 99: 116457.
- [13] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[EB/OL]. (2015-04-10). <https://arxiv.org/abs/1409.1556>.
- [14] XIA Z Q, HONG X P, GAO X Y, et al. Spatiotemporal Recurrent Convolutional Networks for Recognizing Spontaneous Micro-Expressions[J]. IEEE Transactions on

- Multimedia, 2020, 22(3): 626–640.
- [15] YAN W J, LI X B, WANG S J, et al. CASME II: an Improved Spontaneous Micro-Expression Database and the Baseline Evaluation[J]. PLoS One, 2014, 9(1): e86041.
- [16] ZHOU Z H, ZHAO G Y, PIETIKÄINEN M. Towards a Practical Lipreading System[C]//The 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011). Colorado Springs: IEEE, 2011: 137–144.
- [17] HUANG H, GE P. Depth Extraction in Computational Integral Imaging Based on Bilinear Interpolation[J]. Optica Applicata, 2020, 50(4). DOI: 10.37190/oa200401.
- [18] KING D E. Dlib-ml: A Machine Learning Toolkit[J]. Journal of Machine Learning Research, 2009, 10(3): 1755–1758.
- [19] 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.
- [20] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. Communications of the ACM, 2017, 60(6): 84–90.
- [21] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014. DOI: 10.48550/arXiv.1409.1556.
- [22] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S. l.]: IEEE, 2016. DOI: 10.1109/CVPR.2016.90.
- [23] 刘德志, 梁正友, 孙 宇. 结合空间注意力机制与光流特征的微表情识别方法 [J]. 计算机辅助设计与图形学学报, 2021, 33(10): 1541–1552.
- LIU Dezhi, LIANG Zhengyou, SUN Yu. Micro-Expression Recognition Method Based on Spatial Attention Mechanism and Optical Flow Features[J]. Journal of Computer-Aided Design & Computer Graphics, 2021, 33(10): 1541–1552.
- [24] GAN Y S, LIONG S T, YAU W C, et al. OFF-ApexNet on Micro-Expression Recognition System[J]. Signal Processing: Image Communication, 2019, 74: 129–139.
- (责任编辑: 廖友媛)

.....

(上接第 33 页)

- [8] JAVEED K, WANG X J. FPGA Based High Speed SPA Resistant Elliptic Curve Scalar Multiplier Architecture[J]. International Journal of Reconfigurable Computing, 2016, 2016: 6371403.
- [9] JAVEED K, WANG X J, SCOTT M. High Performance Hardware Support for Elliptic Curve Cryptography over General Prime Field[J]. Microprocessors and Microsystems, 2017, 51: 331–342.
- [10] JAVEED K, WANG X J. Low Latency Flexible FPGA Implementation of Point Multiplication on Elliptic Curves over GF(p)[J]. International Journal of Circuit Theory and Applications, 2017, 45(2): 214–228.
- [11] 尤文珠, 葛海波. 利用多基数系统的高效椭圆曲线多标量乘法 [J]. 计算机工程, 2021, 47(2): 182–187.
- YOU Wenzhu, GE Haibo. Efficient Algorithm for Multi-Scalar Multiplication of Elliptic Curves Using Multi-Base Number System[J]. Computer Engineering, 2021, 47(2): 182–187.
- [12] 韩炼冰, 段俊红, 王 松, 等. 基于 FPGA 的 Edwards 曲线标量乘法实现方法 [J]. 通信技术, 2015, 48(10): 1179–1182.
- HAN Lianbing, DUAN Junhong, WANG Song, et al. FPGA-Based Implementation of Scalar Multiplication on Edwards Curve[J]. Communications Technology, 2015, 48(10): 1179–1182.
- [13] 韩炼冰. 椭圆曲线密码算法的 FPGA 设计与实现 [D]. 成都: 电子科技大学, 2018.
- HAN Lianbing. The Design and Implementation of FPGA over Elliptic Curves Cryptography[D]. Chengdu: University of Electronic Science and Technology of China, 2018.
- [14] 刘 艳, 郎显赫, 裴少婧. 基于 ECC 与同态加密的加密算法 [J]. 计算机工程与设计, 2020, 41(5): 1243–1247.
- LIU Yan, LANG Xianhe, PEI Shaojing. Encryption Algorithm Based on ECC and Homomorphic Encryption[J]. Computer Engineering and Design, 2020, 41(5): 1243–1247.
- [15] 尤文珠, 葛海波. 二进制域上椭圆曲线密码 ECC 的高性能 FPGA 实现 [J]. 计算机科学, 2020, 47(8): 127–131.
- YOU Wenzhu, GE Haibo. High-Performance FPGA Implementation of Elliptic Curve ECC on Binary Domain[J]. Computer Science, 2020, 47(8): 127–131.
- (责任编辑: 申 剑)