

基于改进特征融合的微表情识别方法

钱泽锋¹, 钱梦莹²

(1.浙江理工大学计算机科学与技术系, 浙江 杭州 310018;

2.绍兴市环保局, 浙江 绍兴 312000)

✉591311060@qq.com; 2267110273@qq.com



摘 要: 微表情的变化是非常微小的, 这使得微表情的研究非常困难。微表情是不能伪造和压制的, 因此也成为判断人们主观情感的重要依据。本文提出了以卷积神经网络及改进长短期记忆网络特征融合为依托的微表情识别方法, 先介绍了相关的背景知识, 再介绍了实验的预处理过程、特征提取以及相应的特征融合的过程, 将所得的结果用于实验模型的预测分类。实验结果表明, 新模型具有更好的识别率。

关键词: 微表情识别; 特征融合; 微表情分类; 卷积神经网络; LSTM

中图分类号: TP3-0 **文献标识码:** A

Micro-expression Recognition Method based on Improved Feature Fusion

QIAN Zefeng¹, QIAN Mengying²

(1. Department of Computer Science and Technology, Zhejiang Sci-tech University, Hangzhou 310018, China;

2. Shaoxing Environmental Protection Bureau, Shaoxing 312000, China)

✉591311060@qq.com; 2267110273@qq.com

Abstract: Changes in micro-expressions are very small, which makes it hard to study micro-expression. Besides, micro-expression cannot be forged and suppressed, which makes it an important basis for judging people's emotions. This paper proposes a micro-expression recognition method based on convolutional neural networks and improved Long-Short-Term Memory (LSTM) network feature fusion. First, it introduces relevant background knowledge and then introduces experimental preprocessing process, feature extraction and corresponding feature fusion process. Results are used in prediction classification of experiment model. Experimental results show that the new model has a better recognition rate.

Keywords: micro-expression recognition; feature fusion; micro-expression classification; convolutional neural network; LSTM

1 引言(Introduction)

微表情的变化是非常微小的, 这使得微表情的研究非常困难。这种表达方式是不能伪造和压制的, 因此也成为判断人们主观情感的重要依据^[1-3]。

2002年, 国外学者Russell等人将首个微表情训练工具开发出来。其后, Ekman团队又研究出来一个新的微表情识别测量工具: 短暂表情识别测验。除了METT和JACBART技术之外, Ekman团队还设计了面部动作编码系统(Facial Action Coding System, FACS)^[4-5]。FACS根据人脸肌肉不同的部分, 以生物学中的解剖学为基础, 将人脸的不同区域划分成44个运动单元(Action Unit, AU), 它们彼此独立又相互联

系, 同时把不同的AU单元整合在一起, 使FACS码得以形成, 任意一种面部表情与特定的FACS码相对应。图1和图2给出了FACS编码中A₀和A₄的区别。



图1 表情是A₀

Fig.1 Facial expression A₀



图2 表情是A₄
Fig.2 Facial expression A₄

2 基本理论(Basic theory)

数据的预处理在大部分的实验中都是极其重要的, 如果微表情图像不经过预处理就会有許多噪声影响, 这会使得微表情识别率不高。在检测微表情时, 预处理图像是必需的步骤。在识别时, 通常来说, 预处理步骤是: 人脸配准(Face Alignment)、人脸切割(Face Cope)、图像归一化(Image Resize)^[6-8]; 之后则是特征提取过程, 这一步骤直接影响到后面分类的准确性; 还有对CNN和LSTM背景知识的介绍。

2.1 人脸检测及配准

人脸检测与对齐简称人脸配准, 人脸配准的好坏对识别结果有很大的影响。人脸检测的过程: 先给出一张图像, 通过计算机自动检测判断出该图像中的人脸区域有没有存在, 倘若没有包括在内, 就会进行提醒; 相反, 就会给出相应的人脸区域^[9]。2006年, Cootes等人提出局部约束模型算法(CLM算法)^[10-11]。与活动形状模型和活动表观模型对比, 这一方法不仅可以有效定位正脸, 同时可以有效解决脸部形状变化下产生的鲁棒性问题, 还解决了光照因素的鲁棒性问题。不仅如此, 这一模型的局部纹理关键点维度要低一些, 在计算速度方面要快一些。尽管局部约束模型具有以上优点, 但是其也有缺点, 就是配准结果不平稳^[12-14]。

2.2 人脸切割及图像归一化

在人脸对齐后要实施人脸切割, 也就是把之前图像中人脸区域进行有效划分。对于不同的图像处理技术, 如人脸检测、表情以及年纪识别等均要实施图像归一化, 利用这一操作, 可以令图像数量不发生变化, 同时消除外部环境影响^[15]。利用这一技术, 不仅能够消除无关的影响因素, 还能在后续操作中使提取准确率得以保证, 使分类识别准确率提升。图像归一化处理有两个方面, 一是尺寸归一化; 二是灰度值归一化。

2.3 微表情检测特征提取

在各种图像分类识别的模型中, 特征提取都是至关重要的一环, 是算法的核心步骤。有效的特征提取方法不仅能够降低不必要因素的干扰, 获得完整而准确的图像特征信息, 还能够降低数据维数, 减少数据运算, 使之后的图像更好地分类^[16-18]。此外, 有效的特征提取方法稳定性好, 抗干扰能力突出。

2.4 卷积神经网络

卷积神经网络(CNN)为神经网络拓展的一类形式。神经网络的关键环节是把一个或多个传入输送到一个模型之中, 并且获得一个传出, 这个环节叫作神经元, 如图3所示。

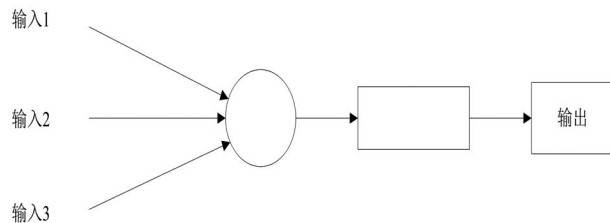


图3 神经元基本构造图

Fig.3 Basic neuron structure diagram

与传统神经网络一样, CNN也是层级的网络, 不过层的作用与模式各异。卷积运算为获取图像特点的很好方式。从卷积核之中获取特点之后, 把获得的特点传送到完整的联结层。CNN模型涵盖卷积与池化执行。每一个执行还涵盖对应的传入、激活、剪枝、传出几个板块, 如图4所示。

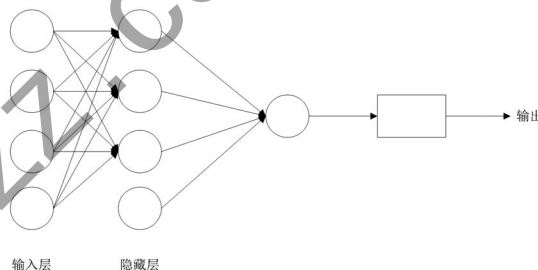


图4 卷积神经网络的构造图

Fig.4 Structure diagram of convolutional neural network

2.5 长短期记忆网络

循环神经网络(RNN)可以通过独特的记忆单元Unit将所有的时序进行一些改变。所以, RNN的输入数据、权值比重等影响着循环神经网络的输出数据。而对于面部微表情的一些序列来说, 它们彼此之间是有一定的关联的, 这也说明了不仅微表情的像素彼此之间是牵连的, 而且序列之间的元素也是有一定关系的, 因此通过RNN进行学习可以得到更好的效果。

长短期记忆(Long Short-Term Memory, LSTM)网络是RNN的一种特殊形式, 它具有记忆能力, 所以能更好地表现前后信息特征之间的相关性。RNN中存在梯度消失等问题, 但是LSTM的出现很好地解决了这个问题。

3 基于CNN和改进LSTM特征融合的表情识别 (Expression recognition based on CNN and improved LSTM feature fusion)

微表情图像序列的特征提取对之后的识别研究起着关键性的作用, 如何提取有效的特征对于微表情识别至关重要, 有效的图像特征能极大提高最终的识别率。本节通过将深度学习与微表情识别相结合, 提出一种有效的特征提取方法, 通过

CNN及改进的LSTM获取相应特征整合构成了微表情图像序列的新特点,而且把获得的特点使用在最后的预估划分中。

3.1 CNN和改进LSTM特征融合模型

因为通常特征表达符和我们需要的预估情感划分相互之间存在一定的距离,所以微表达辨别的精准性并不高。为求缩小特征表达符与微表情类别相互之间的距离,此小节给出了一种新的融合特征描述符,并将该描述符用于微表情识别,CNN与LSTM特征集中于全部图像序列之中。除此之外,我们还通过CNN-LSTM特征融合层将两者融合生成全局特征,并对LSTM结构提出改进,然后将全局特征输入分类器进行预测分类,如图5所示。

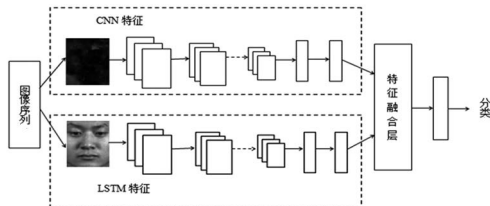


图5 CNN和LSTM融合模型

Fig.5 CNN and LSTM fusion model

从光流图中提取完整有效的的时间特征,将两帧之间的光流位移场图当作传入值来做时间LSTM构建。我们使用AlexNet构建时间LSTM与空间CNN之间的关系。在我们的CNN之中,有5个卷积层、3个池化层与4个FC层。为了防止过拟合现象的发生,我们的池化层使用最大池化策略。而CNN的传入大小为 $96 \times 96 \times 3$,其中,3表示三通道图像,如图6所示。

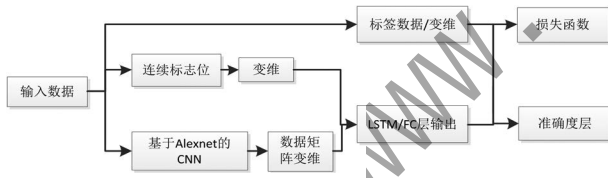


图6 新的LSTM模型结构图

Fig.6 New LSTM model structure diagram

3.2 实验过程

我们使用最常用的SMIC和CASME II微表情数据库进行实验验证,以此评估提出方法的识别准确性。我们使用具有16 GB显存的Tesla GPU进行训练学习。实验的系数设立如下:对于CNN与LSTM特征的训练环节,批处理大小的数值为50,最大轮次数设置为6,000,其中学习率取值为0.0005。

首先是预处理过程:为求达成适应的CNN,我们按照先前阐述的方式实施了人脸测试与人脸对齐,而且把图像的序列调节到 96×96 的大小。每两帧相互之间的光流位移场图当作CNN的传入。

因为图像序列的帧数并不完全一样,图像的信息比较少,为求达成拓宽它包含的信息,我们对每一个所得的序列

实施了划分,把图像切换为大量帧与帧之间的片段,而且应用片段当作传入的信息。

为了使微表情识别的效果最好,应用以下方式来拓展训练信息的起点。我们把每一个片段的帧数定义为 X ,区间为 $[2, 30]$,通过实验证明出现 X 的取值为20时最好。其中,当图像序列中的帧数高于20时,前、后 $(X-20)/2$ 帧被去除;反之,如若帧数低于20,则获取最前和最后的 $(X-20)/2$ 帧。

对于时间的LSTM,我们直接将包含的20帧图像作为起始输入数据。每20帧图像可以得到19幅时间图像,相邻的两幅图像可以取得一幅时间图像。我们把时间的数据定义为一组运动矢量 dt , dt 由第 t 帧与第 $t+1$ 帧计算而得。时间图像涵盖两个分量: dt_x 与 dt_y 。两者依次代表的是第 t 帧到第 $t+1$ 帧之间 X 与 Y 方向相应的位置改变。考虑到网络的传入大小,我们利用 dt_x 与 dt_y 的平方根计算第三个分量 dt_z ,以此来表示输入时间图像。

为了防止过拟合,我们先将图像序列中隐含的特征提取出来,接下来把该特点当作传入值传入网络实施学习与预估划分,具体模型按照之前所给出的。我们将20个处理过的 $96 \times 96 \times 3$ 大小的图像片段传入模型中获取相应的特征,最后进行特征整合的过程,在给出的CNN-LSTM特征融合模型中,核心为训练好的整合层。

在实验探讨环节中,我们研究了各异的时空特征整合方式,如出现空间数据、时态数据特征整合为一个单一结果的情况。在微表情识别实验中使用该全局特征识别效果最好。除了这些之外,我们还尝试在各异的层之间放置改进的CNN-LSTM整合层,而且探讨了可以训练的策略和与之相反的策略之间的差异。最终,我们还对某些比较经典的微表情识别方式进行了对比,例如LBP-TOP+SVM、FDM及MDMO等,以此证实我们所给出的新方式的适用性。

4 实验结果及分析(Experimental results and analysis)

根据以上的分析,我们可以得出在两种模式之下对CNN-LSTM特征融合的微表情识别会有不同的结果,非训练模式的模型表现要比训练模式的模型效果差。在SMIC数据库数据识别率过程中,训练模式要比非训练模式高出5.4%;在CASME II数据库的数据识别过程中,训练模式要比非训练模式高出5%,如表1所示。

表1 训练和非训练模式下的平均识别率

Tab.1 Average recognition rate in training and non-training modes

模式	SMIC	CASMEII
经过训练	53.4%	56.6%
未经过训练	48.0%	51.6%

由此我们可以推断，在训练模式的条件下，CNN-LSTM时空特征融合层可以更加准确地获取到图像序列；而在非训练模式的条件下CNN-LSTM特征融合层仅仅是将过去的神经网络各个层中提取到的时间和空间特征进行整合，因为没有对其进行提前的训练，所以获取特征信息的能力也就相对比较薄弱。

相较于FC6之后插入时空特征融合层得到的性能表现，FC7合格率更高。在SMIC数据库中提取数据时，FC7比FC6识别率高了6.2%；在CASME II数据库中提取数据时，FC7比FC6识别率高了2.5%，如表2所示。

表2 不同位置之后插入融合层的平均识别率

Tab.2 The average recognition rate of inserting the fusion layer after different positions

融合层插入位置	SMIC	CASMEII
FC7	56.3%	60.2%
FC6	50.1%	57.7%

在以上的各个实验中，为了确保对比实验输入数据的数据量相同，使用每个片段19帧的空间图像和每个片段19帧的时间图像进行微表情分类实验。

表3的结果表明，为了更好地识别微表情，可以使用特征融合，而且相较于时间/空间特征提取出来的微表情，融合特征提取出来的微表情得到的识别效果更好。出现这种情况主要是因为时空特征融合提取出了微表情图像中更为关键有用的内容，因此识别效果更好。

表3 不同特征的平均识别率

Tab.3 Average recognition rate of different features

特征	SMIC	CASMEII
融合的特征	56.3%	60.2%
时间特征	41.1%	43.1%
空间特征	41.8%	42.9%

新的特征融合模型在识别方面的功能要远远高于过去以往的方法。出现这种情况源于微表情图像序列中的时间和空间特征信息在我们提出的模型中得到了充分利用，与此同时在特征融合的作用下产生了功能更加齐全完善的时空特性。另外，在以上的种种实验过程中我们还得到一些其他的发现，利用这些方法进行识别的时候，因为CASME II的数据质量比SMIC要高，所以如果识别的对象是CASME II，那么它的成功率远远比识别数据库SMIC要高。从这一方面我们可以看出，如果数据库的样本是高质量的，那么对于微表情的研究则有一定的推动作用，如表4所示。

表4 不同识别方法的平均识别率

Tab.4 Average recognition rate of different recognition methods

方法	SMIC	CASMEII
新的融合模型	55.3%	58.7%
LBP-TOP+SVM	53.1%	56.5%
FDM	42.3%	43.2%
MDMO	52.2%	53.4%

5 结论(Conclusion)

本文对微表情识别的三个阶段，即预处理、特征提取和微表情分类进行了分析。重点介绍了三个阶段中所采用的主要方法和关键技术并对其进行比较，指出了所采用算法的基本原理和优缺点。

经过实验，利用各种层次的方式对特征进行提取，然后将新提取的特征进行特征融合，再送到分类器对微表情进行分门别类。实验最终的数据显示，利用训练的模式比未经训练的模式表现更好。根据其他的实验结果，该方法相较于经典主流特征提取算法表现出了更好的识别分类效果，验证了该方法对提取基于图像序列的微表情特征有着更好的效果。

参考文献(References)

- [1] EKMAN P, FRIESEN W V. The repertoire of nonverbal behavior: Categories, origins, usage, and coding[J]. SEMIOTICA, 1969, 1(4):49-98.
- [2] EKMAN P, ROSENBERG E L. What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system(FACS)[M]. New York: Oxford University Press, 1997, 12(2):7-10.
- [3] EKMAN P, FRIESEN W V, HAGAR J C. Facial action coding system investigator's guide[R]. Salt Lake City: A Human Face, 2002, 22(6):10-16.
- [4] SARAGIH J M, LUCEY S, COHN J F. Deformable model fitting by regularized landmark mean-shift[J]. International Journal of Computer Vision, 2011, 91(2):200-215.
- [5] 刘宇灏.微表情识别的理论和方法研究[D].南京:东南大学,2016,34(2):56-58.
- [6] 伍凯,朱恒亮,郝阳阳,等.级联回归的多姿态人脸配准[J].中国图像图形学报,2017, 22(2):257-264.
- [7] 张余敬,常丹华,刘宇,等.基于Gabor小波变换的人脸表情识别技术研究[J].计算机测量与控制,2010,18(4):906-908.
- [8] 王建超.微表情数据库的建立和微表情检测技术研究[D].青岛:山东大学,2017,45(2):15-18.
- [9] 吴雪.基于单演二值模式的微表情识别研究[D].天津:河北工