

基于三维卷积神经网络与支持向量机的微表情自动识别

何景琳,梁正友,孙宇

(广西大学计算机与电子信息学院,南宁 530004)

摘要:

不同于常规的人脸表情,微表情是一种在人们试图压抑或隐藏真实情感时产生的非常短暂的、不能自主控制的人脸表情序列。微表情识别可以应用于审犯测谎、商务谈判等场合。微表情持续时间短,动作幅度小等特点,使得人们难以用肉眼察觉和感知,目前微表情识别准确率较低。为了解决这个问题,提出采用三维卷积神经网络与支持向量机(C3DSVM)的方法来对微表情进行自动识别,首先通过C3D提取微表情在时域和空域上的特征,然后使用SVM分类器进行分类。在CASME2数据集上做实验,实验表明该方法的识别准确率比其他前沿方法高,达到88.79%。

关键词:

微表情识别; C3D; 特征提取; SVM

基金项目:

国家自然科学基金资助项目(No.61763002)

0 引言

人脸表情在我们日常生活中扮演着重要的角色。与人脸表情相比,自发式的微表情更能够反映人的真实情绪。自发性微表情是指人们尽最大努力抑制真实表情的视频片段,它的持续时间短,仅为 $1/25 \sim 1/3$ s^[1],并且动作幅度小。尽管能够通过人力进行识别,但识别的准确率并不高,经过培训的人员识别准确率不超过50%^[2]。因此,有必要使用先进的计算机视觉技术来提高识别率。

最先用于微表情识别的是手工描述特征识别方法。LBP-TOP^[3]作为手工描述特征识别的一种基准方法,在微表情识别史上首次被采用,该方法是在平面图像特征LBP的基础上,结合微表情视频在时域和空域方面提出的,通过在三个正交平面上对像素及其邻居之间的信息进行编码,从而提取微表情的特征。另外, Huang等人^[4]提出了时空全局量化模式(STCLQP),包括信号、大小和方向因素。Liu等人^[5]提出了一种基于时空空间局部纹理描述符(SLTD)的主方向平均光流(MDMO)法,更为简单有效。

近年来,深度学习技术在识别方面获得巨大成功,已经广泛应用于多个领域,如行为识别^[6]、自然语言处理^[7]、语音识别^[8]等方面,也逐步被应用到微表情自动识别当中。Kim等人^[9]提出结合时域和空域不同维度特征的提取方法,时域维度特征通过搭建CNN提取帧序列的五个不同状态信息获得,时间维度特征通过LSTM网络获得,提高了微表情识别准确率。Peng等^[10]采用两种时域维度差异的卷积神经网络(DTCNN)对不同数据集的微表情进行识别,通过一种数据增强的算法提高数据样本量,也提高了识别的准确率。

本文提出了一种三维卷积神经网络与支持向量机(3D Convolutional Neural Network and Support Vector Machine, C3DSVM)的方法,通过C3D提取微表情在时域和空域上的中间层特征,并采用SVM分类器进行分类,实现微表情的自动识别。所提出的方法在CASME2^[11]微表情数据集上做了实验,实验表明该方法在采用五折交叉验证法的情况下,提高了微表情自动识别的准确率。

1 C3D

C3D 是一种在时域和空域上进行同时进行深度学习的一种技术,可用于行为识别、场景识别、视频相似度分析等领域。Ji 等人^[12]首次提出一种能够同时作用在时间和空间上的 3D 卷积网络,在实时视频环境中识别人类行为。

C3D 的优点在于,采用三维卷积核对上一层网络中的图像感受野进行卷积操作,可以一次性提取时域特征,即可以捕捉到多个帧的动作信息。具体地,对于第 i 层网络的第 j 个图像感受野上的像素点 (x, y, z) 的特征值可以记作 a_{ij}^{xyz} ,公式为:

$$a_{ij}^{xyz} = f \left[b_{ij} + \sum_{s=0}^{S_i-1} \sum_{t=0}^{T_i-1} \sum_{r=0}^{R_i-1} w_{ij}^{str} a_{i-1}^{(x+s)(y+t)(z+r)} \right] \quad (1)$$

其中,而 S_i 和 T_i 分别是三维卷积核的高度和宽度, R_i 是三维卷积核的时域维度大小, w_{ij}^{str} 为连接上层图像感受野的三维卷积核位置 (s, t, r) 的权重。 $f(x)$ 表示激活函数,常用的激活函数有 ReLU (Rectified Linear Unit)、sigmoid 等。Hinton 等人^[13]提出实验中使用 ReLU 比 sigmoid 更好,因此,实验中采用 ReLU。ReLU 可以表示为:

$$f(t) = \max(0, t) \quad (2)$$

最后一层输出层采用 Softmax 激活函数,Softmax 在输出层中表示当前神经元输出的概率,公式为:

$$f(t) = \frac{e^t}{\sum_c e^t} \quad (3)$$

其中, c 为输出层神经元个数。

2 C3DSVM 微表情自动识别

2.1 微表情数据集

目前自发的微表情数据集较少,仅有的三个数据集分别是 CASME1^[14]、CASME2 和 SMIC^[15]。本文的全部实验采用 CASME2 数据集。

CASME2 是中国科学院心理研究所收集的 CASME 数据库的升级版本。CASME2 包含由 200fps 相机记录的 26 个受试者的 255 个微表情视频序列。获得的微表情样品由 AU 编码,包括三个部分:起始、顶点和结束。微表情数据集可以分为七类:高兴、惊讶、恐惧、悲伤、厌恶、压抑、其他。

在我们的实验中,由于两个类的样本量不足以进

行特征学习的训练,因此不使用样本过少的视频序列类别(即恐惧和悲伤)来进行识别,即其余的 246 个样本用于实验,如表 1 所示。

表 1 实验中使用的微表情数据集情况

数据集	受试者	总样本数量	类别	标注(数量)
CASME2	26	246	5	高兴(32) 厌恶(63) 惊讶(25) 压抑(27) 其他(99)

2.2 C3DSVM

C3DSVM 微表情自动识别方法就是通过 C3D 提取微表情在时域和空域上的中间层特征,再将中间层特征送入 SVM 分类器,实现微表情的分类。图 1 为微表情自动识别示意图。

图 1 注释:输入视频序列尺寸为 $(16 \times 112 \times 112)$,其中 16 表示为微表情视频帧长, 112×112 分别表示帧的长和宽。

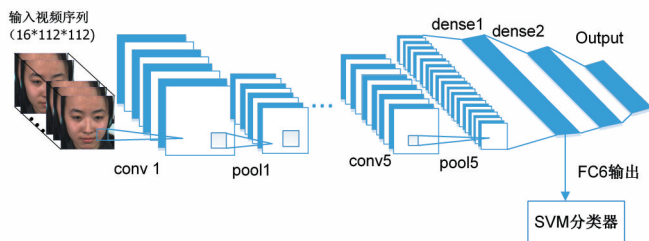


图 1 C3DSVM 微表情自动识别方法

图 2 注释:蓝色表示卷积层(conv),黄色表示最大池化层(pool),绿色表示完全连接层(dense)。所有的卷积层的卷积核大小为 $(3 \times 3 \times 3)$ 。除了 pool1 的降采样核大小是 $(1 \times 2 \times 2)$,其他 pool 层均为 $(2 \times 2 \times 2)$,并且全部采用的是最大采样,每个完全连接层有 4096 个神经元。

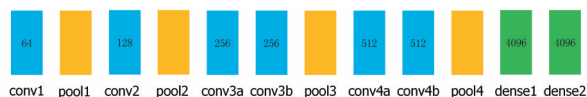


图 2 实验的网络结构

如图 1 所示,输入的微表情序列在 C3D 网络结构上经过反复的训练后,在 dense1 这个中间层输出特征,并送入 SVM 分类器,实现分类。具体的网络结构如图 2 所示。所提出的三维卷积神经网络包括 6 个卷积层、4 个最大池化层、2 个完全连接层以及 1 个输出层。除了最后一层输出层采用 Softmax,其余层的激活函数都采用 ReLU,参数训练采用随机梯度下降算法(Stochas-

tic Gradient Descent, SGD)。

后面的分类采用 SVM 分类器。SVM 是用于分析数据的一种监督学习算法,在小样本训练集上能够得到比其他算法好很多的结果^[16],也具备泛化能力。实验中的 SVM 分类器采用径向基函数(Radical Basis Function, RBF),RBF 是空间中任一点 x 到某一中心 c 之间欧氏距离的单调函数,其作用往往是局部的,即当 x 远离 c 时函数取值很小,形式为:

$$k(\|x - c\|) = \exp \left[\frac{-\|x - c\|^2}{(2t)^2} \right] \quad (4)$$

其中 c 为核函数中心, t 为函数的宽度参数,控制了函数的径向作用范围。RBF 通过将样本映射到一个更高维的空间,实现非线性可分的样本。微表情由非线性的图片序列构成,使用 RBF 核函数的 SVM 分类器能够有效实现分类。

3 实验

3.1 实验过程

预处理。微表情识别的预处理流程包括人脸识别、裁剪以及规范化处理环节。如图 3 所示,首先在微表情预处理阶段检测人脸区域,对每个视频序列的第一帧采用 Haar 人脸检测器^[17]进行人脸识别,第一帧识别得到的人脸区域矩形识别框适用于剩余帧,最后将整个视频序列对应的人脸区域裁剪下来。在规范化处理中,由于每个视频的大小尺寸不统一,我们首先通过平面线性差值的方法将每一帧的平面尺寸统一为 112×112 。其次,由于帧序列长度不统一,在时域序列上(Temporal Interpolation Method, TIM)^[18]通过 TIM 的方法统一帧数为 150。经过这样的调整,每个样本大小统一为 $112 \times 112 \times 150 \times 3$ (3 为 RGB 通道)。最后,为了获取更细致的特征,等步长截取视频帧,如图 4 所示,实验以 16 帧为步长无重叠地截取每个样本中的 16 帧,即新的样本大小为 $112 \times 112 \times 16 \times 3$,实验中由于微表情出现的高峰在视频时域上的中间段,尽量取该段的帧序列。

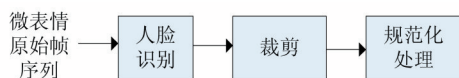


图3 微表情识别的预处理流程

验证方法和指标说明。本文的实验均采用五折交叉验证的方法,即将样本随机打乱后平均分为 5 份,4

份用作训练集,剩余 1 份用作测试集,循环下来,就能得到五个测试集的预测结果(准确率)。本文的指标采用该准确率。

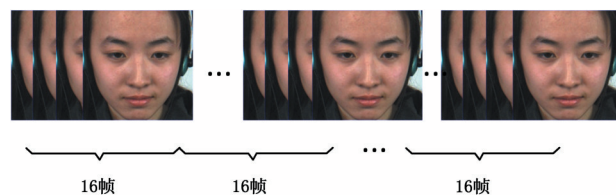


图4 以16帧为步长截取原始帧序列

训练参数及硬件说明。网络结构如 2.3 节所述,用 Keras 实现。训练时,目标函数使用均方误差(Mean Square Error, MSE);初始学习率设置为 0.01, epoch 设定为 60。以上参数通过反复试验确定。实验的主要硬件设备是两个 NVIDIA Titan X GPU。

3.2 实验结果及对比

(1)和其他方法的对比

实验结果如表 2 所示,与其他前沿方法相比,本文提出的 C3DSVM 的识别方法获得了最高的准确率,为 88.79%。表 2 前两行是手工特征描述的识别方法,后面为深度学习方法。相比之下,C3DSVM 作为一种深度学习方法,能够自动提取特征,省去了人工寻找特征的步骤,也提高了识别准确率。

相比二维卷积神经网络,C3DSVM 能够提取在时域和空域上的特征。二维卷积神经网络在提取视频特征上,是将时域上的特征图堆叠得到,容易丢失时域的信息。对于微表情本身动作幅度小的特性来说,使用 C3DSVM 能够有效提取时域上的特征,实现了很好的分类。

表2 不同微表情识别方法的准确率

方法	准确率(%)
LBP-TOP+SVM ^[3]	63.41
MMFL ^[19]	59.81
MOST ^[9]	60.98
DTSCNN ^[10]	66.67
CNN ^[20]	74.25
Bi-WOOF ^[21]	59.26
C3DSVM (本文)	88.79

(2)不同条件下方法对比

使用 SVM 分类能够提高识别准确率。图 5 表示 C3DSVM 和 C3D 的五折识别准确率,图 6 表示对应的平均识别准确率。由这两个表格可以看出,C3DSVM

的总体识别准确率高于 C3D,这是由于 SVM 适用于样本量较少的数据集,C3DSVM 不直接在 C3D 网络的最后一层分类,而将 C3D 提取出的 dense1 中间层特征放入 SVM,在小样本的数据集下 SVM 能够发挥其优势,实现微表情有效分类。

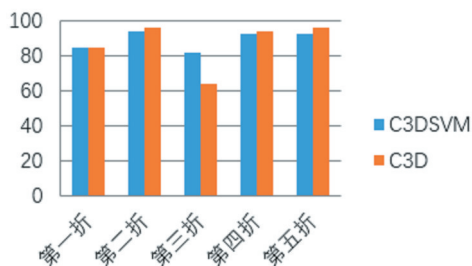


图5 使用C3DSVM和C3D的五折识别准确率(%)比较

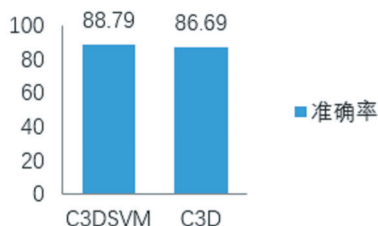


图6 使用C3DSVM和C3D的平均识别准确率(%)比较

使用预处理中 TIM 统一帧长处理的视频帧可以提高微表情识别准确率。no-TIM+C3DSVM 表示没有采用 TIM 的对比方法。图 7 表示 C3DSVM 和 no-TIM+C3DSVM 的五折识别准确率,图 8 表示对应的平均识别准确率。由这两个表格可以看出,C3DSVM 的总体识别准确率高于 no-TIM+C3DSVM,发现在采用 TIM 的情况下,C3DSVM 能够发挥出最好的效能,在使用 C3D 网络的前提下,输入层的输入大小一定,微表情帧长统一有利于网络提取特征,进而提高识别准确率。

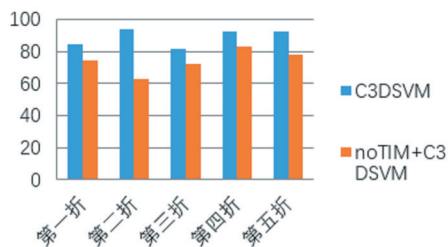


图7 使用C3DSVM和noTIM+C3DSVM的五折识别准确率(%)比较

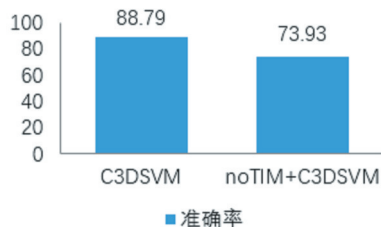


图8 使用C3DSVM和noTIM+C3DSVM的平均识别准确率(%)比较

4 结语

本文提出了 C3DSVM 的微表情自动识别方法,通过 C3D 对微表情视频序列在时域和空域上的中间层特征提取,再用 SVM 分类器分类。和传统的手工方法以及现有的深度学习方法相比,所提出的方法明显提高了识别准确率。另外,本文也探索了采用 TIM 的预处理方法对识别准确率的作用,实验结果说明采用 TIM 能够有效提高微表情识别率。

然而,深度学习极度依赖硬件设备,参与训练的参数过大,训练时间过长,训练效率不高,这也是使用深度学习技术普遍存在的弊端。另外,实验中探索网络结构设置、网络参数设置和激活函数等参数较为复杂。如何优化网络结构,优化网络参数,以进一步提高识别准确率是今后研究的重点。

参考文献:

- [1]Corneanu C, Oliu M, Cohn J F, et al. Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-related Applications[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(8):1548-1568.
- [2]Russell T A, Chu E, Phillips M L. A Pilot Study to Investigate the Effectiveness of Emotion Recognition Remediation in Schizophrenia Using the Micro-Expression Training Tool[J]. British Journal of Clinical Psychology, 2006, 45(4):579-583.

- [3]Pfister T, Li X, Zhao G, et al. Recognising Spontaneous Facial Micro-Expressions[C]. Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011:1449-1456.
- [4]Huang X, Zhao G, Hong X, et al. Spontaneous Facial Micro-Expression Analysis Using Spatiotemporal Completed Local Quantized Patterns[J]. Neurocomputing, 2016, 175:564-578.
- [5]Liu Y J, Zhang J K, Yan W J, et al. A Main Directional Mean Optical Flow Feature for Spontaneous Micro-Expression Recognition[J]. IEEE Transactions on Affective Computing, 2016, 7(4):299-310.
- [6]Huang W, Fan L, Harandi M, et al. Toward Efficient Action Recognition: Principal Backpropagation for Training Two-Stream Networks[J]. IEEE Transactions on Image Processing, 2019, 28(4):1773-1782.
- [7]Simpson S S, Adams N, Brugman C M, et al. Detecting Novel and Emerging Drug Terms Using Natural Language Processing: a Social Media Corpus Study[J]. JMIR Public Health and Surveillance, 2018, 4(1).
- [8]Shan C, Zhang J, Wang Y, et al. Attention-Based End-to-End Speech Recognition on Voice Search[J]. 2018:4764-4768.
- [9]Kim D H, Baddar W J, Ro Y M. Micro-Expression Recognition with Expression-State Constrained Spatio-Temporal Feature Representations[C]. Proceedings of the 2016 ACM on Multimedia Conference. ACM, 2016:382-386.
- [10]Peng M, Wang C, Chen T, et al. Dual Temporal Scale Convolutional Neural Network for Micro-Expression Recognition[J]. Frontiers in Psychology, 2017, 8:1745.
- [11]Yan W J, Li X, Wang S J, et al. CASME II: An Improved Spontaneous Micro-Expression Database and the Baseline Evaluation[J]. PLOS ONE, 2014, 9(1):1-8.
- [12]Ji S, Xu W, Yang M, et al. 3D Convolutional Neural Networks for Human Action Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1):221-231.
- [13]Hinton G E. Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair[J]. 2010.
- [14]Yan W J, Wu Q, Liu Y J, et al. CASME Database: a Dataset of Spontaneous Micro-Expressions Collected from Neutralized Faces[C]. Automatic Face and Gesture Recognition (fg), 2013 10th IEEE International Conference and Workshops on. IEEE, 2013:1-7.
- [15]Li X, Pfister T, Huang X, et al. A Spontaneous Micro-Expression Database: Inducement, Collection and Baseline[C]. Automatic Face and Gesture Recognition (fg), 2013 10th IEEE International Conference and Workshops on. IEEE, 2013:1-6.
- [16]丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述[D], 2011.
- [17]Viola P, Jones M. Rapid Object Detection Using a Boosted Cascade of Simple Features[C]. Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. IEEE, 2001, 1.
- [18]Zhou Z, Zhao G, Pietikainen M. Towards a Practical Lipreading System[C]. Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011:137-144.
- [19]He J, Hu J F, Lu X, et al. Multi-Task Mid-Level Feature Learning for Micro-Expression Recognition[J]. Pattern Recognition, 2017, 66: 44-52.
- [20]Patel D, Hong X, Zhao G. Selective Deep Features for Micro-Expression Recognition[C]. 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016:2258-2263.
- [21]Liong S T, See J, Wong K S, et al. Less is More: Micro-Expression Recognition from Video Using Apex Frame[J]. Signal Processing: Image Communication, 2018, 62:82-92.

作者简介:

何景琳(1994-),女,广东廉江人,硕士,研究方向为图像识别、深度学习

通信作者:梁正友(1968-),男,广西天等人,博士,教授,研究方向为无线传感器网络、并行分布式计算、人工智能,

E-mail: zhyliang@gxu.edu.cn

孙宇(1981-),女,广西南宁人,讲师,博士,研究方向为智能算法、图像识别、数据挖掘

收稿日期:2019-02-28 修稿日期:2019-03-05

Facial Micro-Expression Auto-Recognition Using 3D Convolutional Networks and Support Vector Machine

HE Jing-lin, LIANG Zheng-you, SUN-yu

(School of Computer and Electronics Information, Guangxi University, Nanning 530004)

Abstract:

Unlike normal face expression, micro-expression is a very brief, involuntary facial expression that produces when people try to suppress or hide real emotions. It can be used in trials such as polygraphs and business negotiations. The short duration of micro-expressions and the small intensity of motion make it difficult for people to aware and perceive with the naked eye. Currently, the micro-expression recognition accuracy is low. In order to solve this problem, proposes a three-Dimensional Convolutional Neural Network and Support Vector Machine method (C3DSVM) to recognize micro-expression. Firstly spatio-temporal features from micro-expressions are extracted, and then are classified by SVM classifier. Validates the proposed method on CASME2 dataset, and achieves 88.79% on accuracy, which is higher than other state-of-the-art methods.

Keywords:

Micro-Expression Recognition; C3D; Feature Extraction; SVM

(上接第 42 页)

作者简介:

徐丽新(1981-),女,吉林白城人,硕士,讲师,高级职业指导师,研究方向为数字图像处理、P2P

收稿日期:2019-03-05

修稿日期:2019-04-10

Research on the Constant-Degree Structured P2P System: Cycloid

XU Li-xin

(School of Information Engineering, Guangdong Engineering Polytechnic, Guangzhou 510520)

Abstract:

Structured P2P system based on DHT must maintain a restricted topology, when the network churn happens, the joining and leaving nodes will lead to high overhead. Some new P2P systems have a good character that a node maintains simply constant number of neighbors. Cycloid based on cube-connected cycle has $O(\log n)$ hops per lookup request with only seven neighbors per node.

Keywords:

Peer-to-Peer; Structured P2P; Constant-Degree P2P; Cycloid