doi:10.3969/j.issn.1001-893x.2019.10.001
引用格式:吴进,闵育,李聪,等. 一种基于 3D-CNN 的微表情识别算法[J]. 电讯技术,2019,59(10):1115-1120. [WU Jin,MIN Yu,LI Cong,et al. A micro-expression recognition algorithm based on 3D-CNN[J]. Telecommunication Engineering,2019,59(10):1115-1120.]

# 一种基于 3D-CNN 的微表情识别算法*

吴 进**,闵 育,李 聪,张伟华

(西安邮电大学 电子工程学院,西安 710121)

**摘 要**:微表情是一种持续时间很短暂的面部表情。针对其识别率低的问题,提出了一种基于三维卷积神经网络(3D Convolutionnal Neural Network,3D-CNN)的微表情识别算法。使用 Keras 作为网络框架,在 3D-VGG-Block(3Dimension Visual Geometry Group Block,3D-VGG-Block)的基础上加入批量归一化算法以及丢弃法,提升网络深度与训练速度的同时有效地防止过拟合;针对数据集稀少的问题,采取随机设置起始帧的位置,提前设定每次读取帧序列的长度,循环操作,在将所有数据均遍历的同时,达到数据增广的目的。该算法在 CASME II 数据集上的识别率最高达 68.85%,在识别率上有一定优势。

**关键词**:微表情识别;深度学习;三维卷积神经网络;批量归一化算法;丢弃法

**开放科学(资源服务)标识码(OSID)**:

微信扫描二维码
听独家语音释文
与作者在线交流

**中图分类号**:TP183;TP391.41 **文献标志码**:A **文章编号**:1001-893X(2019)10-1115-06

# A Micro-expression Recognition Algorithm Based on 3D-CNN

WU Jin,MIN Yu,LI Cong,ZHANG Weihua

(School of Electronic and Engineering,Xi´an University of Posts and Telecommunications,Xi´an 710121,China)

**Abstract**: Micro-expression is a facial expression that lasts for a short time. For the problem of low recognition rate,a micro-expression recognition algorithm based on 3D convolutional neural network(3D-CNN) is proposed. Specifically,Keras is used as a network framework,and the batch normalization algorithm and dropout are added on the basis of 3D visual geometry group block(3D-VGG-Block). It effectively prevents overfitting while improving network depth and training speed. For the problem of rare data sets,the position of the starting frame is randomly set. Meanwhile,the length of the sequence of frames which need to be read is preseted every time,and the loop operation is performed. It achieves the goal of data augmentation while traversing all data. The recognition rate of the algorithm on the CASME II dataset is up to 68.85%,which proves its advantage in recognition rate.

**Key words**: micro-expression recognition; deep learning; 3D convolutional neural network(3D-CNN); batch normalization algorithm; dropout algorithm

## 1 引 言

人们内心的各种情绪一般都会在脸上体现出来,然而在特殊的情况下,为了掩饰心中真实的情绪,其面部表情会很细微,我们通常称之为微表

* 收稿日期:2019-01-15;修回日期:2019-02-23
基金项目:国家自然科学基金资助项目(61772417,61634004,61602377);陕西省科技统筹创新工程项目(2016KTZDGY02-04-02);陕西省重点研发计划(2017GY-060);陕西省自然科学基础研究计划项目(2018JM4018)

** 通信作者:wujin1026@126.com

情[1]。1966 年,Haggard 等[2]首次提出了微表情的概念。微表情是一种很短暂的自发式表情,一般只会持续 1/25 ~ 1/5 s,表情变化幅度非常小[3]。

近年来,微表情识别因为潜在的应用价值而逐渐引起关注[4]。传统的识别算法有很多,例如局部二值模式(Local Binary Patterns,LBP)、三正交平面的局部二值模式[5](Local Binary Patterns from Three Orthogonal Planes,LBP - TOP)、光流与 LBP - TOP 特征结合[6]、方向梯度直方图(Histogram of Oriented Gradient,HOG)等特征提取的方式。这类算法对于特征描述方式比较单一,只能很好地提取描述特征点相对明显的地方;HOG 对于遮挡的位置处理就很困难,如在比较活跃的嘴角部位,它的提取会比较粗糙。对于微表情这种细微的特征,这些算法并不能很好的描述微表情原有的特征。光流法在计算机视觉领域应用广泛[7]。它的目的是在连续的两帧图像上进行特征跟踪[8],而一个完整的微表情至少包含十多帧的视频序列,它并不能提取到长的视频帧序列的信息,微表情的识别率并不理想。

因此,微表情识别近几年在深度学习方面获得了更多的关注。深度学习最早由 Hinton[9]提出,经过这些年的发展研究,随后基于卷积神经网络的算法在实例分割[10]、人脸识别[11]、目标检测[12]、目标跟踪[13]、微表情识别[14]等计算机视觉领域都发展得很快,从只能处理二维空间信息的卷积神经网络(Convolutional Neural Network,CNN)扩展到了三维卷积神经网络(3Dimension Convolutionnal Neural Network,3D - CNN)。Ji 等人[15]最早设计了三维结构的 CNN,即 3D - CNN。此外,目前国内外常用的微表情数据集有 SMIC[16]、CASME[17]、CASMEII[18]等,它们是文献中最好最新的自发型数据集。由于目前可用的自发型数据集非常少,综上,微妙短暂的面部表情变化和稀少的数据集均是很大的挑战[19]。

针对上述问题,本文提出了基于改进的 3D - CNN 的网络结构,主要包括卷积层之间的堆叠方式以及特殊的数据预处理方式,同时加入最新的批量归一化算法[20](Batch Normlization,BN)以及丢弃法等,实现了将空间域的二维特征与时间域的一维特征融合的同时加快了训练速度,避免了网络过拟合,实现了数据增广,最终识别率达到 68.85%。

## 2 3D - CNN 网络结构设计

由于 CNN 只能提取静态图像的二维空间特征,应用到基于视频的微表情识别中效果不是很好,因此我们提出了基于 3D - CNN 的识别算法,它将空间域的二维特征和时间域的一维特征进行融合,如图 1 所示,两者同时作为分类依据进行识别,其网络结构以 3D - VGG - Block(3Dimension Visual Geometry Group Block,3D - VGG - Block)为基础,加入 BN 法归与丢弃法,提高识别率的同时加快训练速度,防止过拟合。

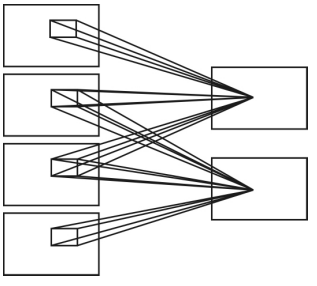

图 1 特征融合

### 2.1 网络各层的堆叠方式

众所周知,由 Simonyan 和 Zisserman[21]提出的 VGG - Block 网络结构模式在 2014 年的 Image Challenge 图像识别比赛上获得第二名。该结构模式用两个连续的卷积层连接之后使用一个池化层进行特征降维。VGG - Block 的结构如图 2 所示,它将两个卷积核为 3 × 3 大小的卷积层进行连接,3 个 3 × 3 的卷积核连接时感受野是 7 × 7,参数量为 27 × $C$,其中 $C$ 为通道数,而同样的感受野下,用 7 × 7 的卷积核时参数量为 49 × $C$。因此使用小卷积核堆叠代替大卷积核能在相同的感受野下增加网络深度,减少参数量,防止过拟合。

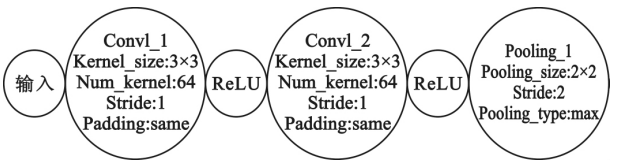


图 2 VGG - Block 结构图

### 2.2 批量归一化算法

针对深度学习网络训练困难的问题,本文加入了 BN 算法,它能够加速训练,在数据预处理中,白化预处理使特征之间的相关性降低,数据均值、标准差归一化。若数据特征维数较大,则进行 PCA 降维处理,即实现白化的第一个要求,但计算量很大。BN 算法忽略这个要求,它用式(1)进行预处理,即

近似白化预处理：

$$\hat{x}^i = \frac{x^i - E[x^i]}{\sqrt{\mathrm{Var}(x^i)}}。\tag{1}$$

式中：$\hat{x}^i$ 是某一网络层中某个神经元的输入，$\hat{x}^i = Wh + b$，$W$ 为该层的权重，$h$ 为上一层输出，$b$ 为不确定常数，$E[x^{(k)}]$ 是对该神经元在随机梯度下降法中一个批次所有输入数据的均值，$\sqrt{\mathrm{Var}[x^{(i)}]}$ 是该神经元一个批次所有输入数据的标准差。

然而，若只用式(1)会使网络的表达能力下降。因此，文献[21]中提出了式(2)，引入了可学习重构参数 $\gamma^i$ 以及 $\beta^i$，它们为神经元的输入加了一个线性变换，能够调节神经元的激活值，增强了网络表达特征的能力。

$$\hat{y}^i = \gamma^i \hat{x}^i + \beta^i。\tag{2}$$

式中：当 $\gamma^{(k)} = \sqrt{\mathrm{Var}[x^{(k)}]}$、$\beta^{(k)} = E[x^{(k)}]$ 时，网络能将原网络要学习的特征分布进行学习恢复。

### 2.3　Dropout 算法

Dropout 算法[22]俗称“丢弃法”，在训练深度神经网络时，若模型参数太多而训练样本太少时，训练时网络模型很容易会出现过拟合现象，本文用的数据集样本少之又少，因此有必要加入 Dropout 防止过拟合。当进行前向传播时，Dropout 能使某个神经元的激活值以一定的概率停止工作，尽可能不依赖一些局部特征，使模型的泛化性更强，有效缓解过拟合，在某种程度上达到正则化的效果。图 3 是使用 Dropout 算法前后网络的简化模型结构。

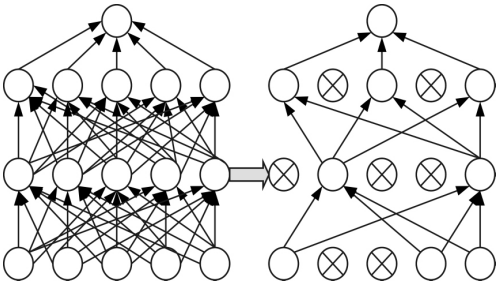

图 3　Dropout 原理图

### 2.4　网络总体结构及性能分析

本文使用 Keras 作为框架进行网络结构设计，TensorFlow 作为 Keras 的后端，其中网络结构模型是线性的图模型(Sequential)，网络总体结构以 3D－VGG－Block 为基础进行设计。图 4 是 3D－CNN 的前置网络结构，图 5 是全连接层结构。

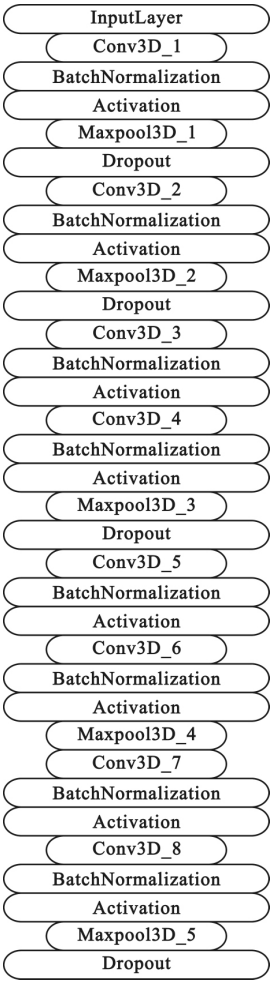


图 4　3D－CNN 的网络结构

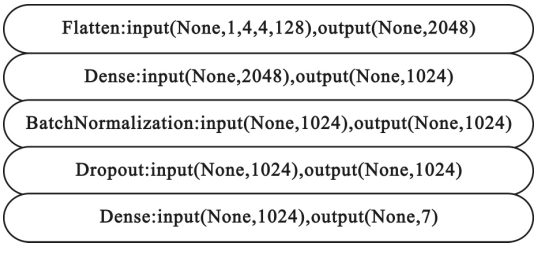


图 5　全连接层网络结构

重要网络参数的具体说明如下：

(1) 输入层：输入(None，16，128，128，3)，None 表示 batch_size 的大小，16 表示时间维度上的视频序列长度，128×128 是进行归一化处理 resize 时设置的图像分辨率，3 表示原始的图像通道数。

(2) 卷积层的滑窗步长与 padding 均为 1，Conv3D_1、2 卷积核数量为 32；Conv3D_3，4 为 64，Conv3D_5，6、7、8 为 128。

(3) 第一个池化窗口为(1，2，2)，其余均为(2，2，2)。

(4) 激活层：Relu 函数。

(5) Dropout 丢弃率均为 0.25，全连接层的为 0.5。

(6) 全连接层输入向量长度为 2 048，压缩特征向

量长度提取更有用的信息，将其设为1 024，最后用 softmax 对长为1 024的输出向量分类，神经元个数为7。

其中，为了减少计算量以及避免降采样而丢失有用的信息，因此第一个池化层设置为(1，2，2)，先只在二维的空间域采样，时域的序列先完整保留，经过后面的池化层对时间域的序列特征不断采样降维，对时间域的序列特征进行提取与压缩。

本文的网络结构不仅通过特殊的网络堆叠方式减少了参数量，而且加入 Dropout 层，通过设置丢弃率，避免影响网络的泛化能力，防止过拟合。由于网络的输入数据维度是5维的，训练和测试时能同时处理多个视频序列，执行效率会好一些，并且本文通过在每个卷积层后加了 BN 层，将卷积层输出的特征图使用 BN 层进行归一化处理，再用 ReLU 函数进行非线性计算来加快训练速度。

## 3 实验与分析

### 3.1 实验环境

本文使用的实验环境：操作系统是 Ubuntu16.04；用于 GPU 加速的底层软件平台是 CUDA8.0；Keras2.0.8 使用 TensorFlow 后端；TensorFlow 1.2.0 - rc0；GPU 为 GeForce GTX1080Ti，其显存是11 GB；CPU 为 Intel® Xeon® CPU E5 - 2620；内存为 SKhynix，总共64 GB；硬盘为 Inter SSD 540 s，总共480 GB。

### 3.2 CASME II 数据集及预处理

CASME II 是中国科学院心理研究所傅小兰团队研发的第二代微表情数据集，它由26名中国人进行拍摄采集，共255个序列大约3 000多个诱发的面部微动作，使用的都是200 frame/s的高速相机，图像分辨率为280 pixel × 340 pixel，在拍摄录制时，光照非常充足且稳定，而且对面部的高光部分进行了处理，所有的微表情都是自发的、动态的。该数据集共有7类微表情，分别是高兴、厌恶、惊讶、抑郁、害怕、悲伤及其他，图6是该数据集中的一个样本实例。

图6 样本实例

由于微表情数据集中每个视频序列的图像帧数都大不相同，而且它们每个都不能把所有的图像都放在网络中进行训练，因此本文设计了一种随机训练视频帧序列中的某一段子序列，不断地循环，最终每个完整的视频帧序列中的每一张图像都可以训练到。即首先得到视频帧序列的总长度 $N$，随机生成起始帧的位置 $R$，设置一个帧序列的长度为16，$R$ 在 $0 \sim N-16$ 之内，每次训练起始帧到终止帧的图像，即 $R \sim R+16$。这样设计的好处是每次都可以生成不同的训练数据，实现了数据增广，正好克服了微表情数据集样本稀少的问题。表1是 CASME II 数据集处理前后的具体数据分析。

表1 CASME II 具体数据分析

| 表情类别 | 原序列数 | 处理后序列数 |
|---|---|---|
| 高兴 | 32 | 512 |
| 厌恶 | 63 | 1 008 |
| 惊讶 | 25 | 400 |
| 抑郁 | 27 | 432 |
| 害怕 | 2 | 32 |
| 悲伤 | 7 | 112 |
| 其他 | 99 | 1 584 |

### 3.3 数据记录

一般情况下，由于神经网络完整训练一次的时间一般都在几个小时甚至十几个小时之久，因此实验数据的记录是很必要的一步。通过分析整个训练过程中记录的数据来查找问题，然后修改网络以及微调。本文实验数据的记录使用了 Keras 的回调函数 Callbacks 的 History 模块，它是一组在训练的特定阶段被调用的函数集，通过回调函数来记录训练网络过程中网络内部的状态和统计信息。本文使用的网络模型是 Sequential，该模型类的 fit 函数会在传入到回调函数的 logs 里面包含 on_epoch_end 数据等，实验中设置 epoch 总数是200，实验记录了200个 epoch 结束时的 accuracy 值与 loss 值。

### 3.4 实验结果与分析

首先，表2给出了实验中训练 3D - CNN 网络时的网络配置参数。其中学习率设置为0.000 1，随着迭代次数的增加，为了防止损失函数发散的情况，本文设置每迭代20 000次时，学习率减少一半，就这样循环了10次，即总共迭代次数为200 000次，每个 epoch 设置迭代1 000次。实验训练网络时数据集使用预处理之后的所有数据，记录了200个 epoch 结束时的 accuracy 值与 loss 值，结果如图7和图8所示。

表 2　网络配置参数

| 参数 | 数值 |
|---|---|
| base_lr | $10^{-4}$ |
| batch_size | 16 |
| stepsize | $2\times10^{4}$ |
| max_iter | $2\times10^{5}$ |
| momentum | $9\times10^{-1}$ |
| decay | $1\times10^{-10}$ |

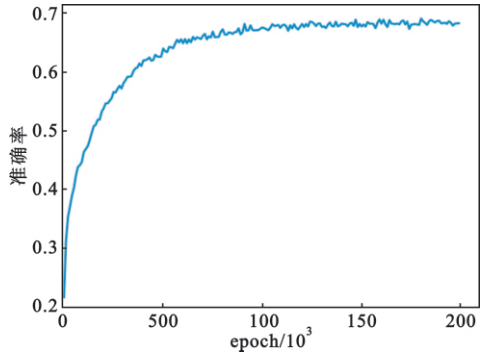


图 7　准确率曲线

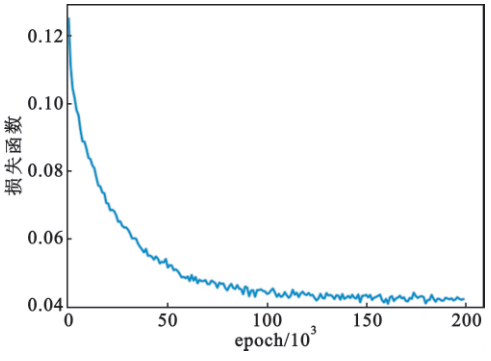


图 8　损失函数曲线

根据图 7 可以看出，当迭代到100 000次左右时，准确率的上升幅度已经比较慢; 当迭代到大约150 000次时，准确率基本已经不再上升，趋于收敛状态。测试时，使用原始的数据集，最后测试的整体准确率最高可达 68. 85%。不同表情在测试集上的分类结果如图 9 所示，表 3 是本文的算法与其他文献中算法的结果对比。其中，表情类别是高兴、惊讶的识别率较高，而害怕、悲伤的识别效果较差，其主要原因可能在于: 第一，前两者微表情的区分度相对于其他类来说比较明显; 第二，它们的样本序列也相对较多，而后者的样本序列稀缺且表情的表现幅度不易区分。

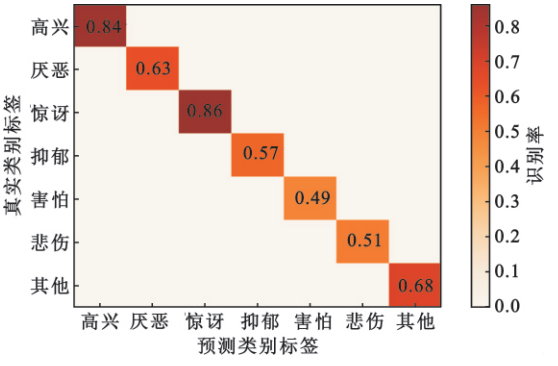


图 9　混淆矩阵

表 3　算法结果对比

| 算法 | 准确率/% |
|---|---|
| LBP - TOP[3] | 63.01 |
| LBP - TOP + OF + RF[6] | 64.46 |
| LBP - TOP + SVM[23] | 63.27 |
| CNN + LSTM[1] | 66.53 |
| 本文算法 | 68.85 |

表 3 中对比的算法均是从该文献中选取的最好的算法结果，由结果可知，相对于目前较优异的传统的算法，本文在准确率上有较大的优势。表中对比的 CNN 与长短期记忆网络(Long Short - Term Memory,LSTM)融合的算法，虽然 LSTM 利用了前后帧的数据信息，克服了 CNN 只能提取二维空间特征的弊端，但本文网络结构由于特殊的网络堆叠方式，以及加入的 Dropout 层，使得网络深度更深，减小过拟合，并且本文中对原始的数据集进行了数据增广的处理，这在一定程度上均提高了识别率。

## 4　结束语

本文提出了一种基于 3D – CNN 网络结构的微表情识别，该网络结构是以 3D – VGG – Block 为基础实现的。该结构的主要特点就是利用小卷积核堆叠代替了大卷积核，同时能在相同的感受野下提高网络深度，更能提取出深层的更好的特征，并且明显减少参数量。针对网络训练困难问题，本文在网络中加入了 BN 层加速训练，接着加入 Dropout 层，有效缓解了过拟合问题，增强了模型的泛化性。针对目前微表情识别存在的数据集稀少、数据类别分配不均衡问题，本文在读取数据时采取随机设置起始帧的位置，提前设定每次读取帧序列的长度，循环操作，在将所有的数据都能遍历到的同时，也达到了数据增广的目的，最后实验结果最高达到 68. 85% 的识别率。

虽然本文在针对识别率低、数据集稀少等问题上得到一定的改善，但仍存在数据集中微表情类别缺乏多样性以及每个类的数据分配不均衡的问题，而且微表情持续时间非常短，鲁棒性差，在后续的工作中，将深入研究微表情数据集的建立以及研究更具鲁棒性的特征提取方法，从而使微表情识别能在日常领域更好地体现其应用价值。

## 参考文献：


[1] 唐爽.基于深度神经网络的微表情识别[J].电子技术与软件工程,2017(3):93 - 95.

[2] HAGGARD E A,ISAACS K S .Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy[M]. Heidelberg:Springer Nature,1966:154 - 165.

[3] 董晓晨,赵志刚,吕慧显,等.基于改进的局部二值模式的微表情识别方法[J].青岛大学学报(自然科学版),2018,31(3):32 - 36.

[4] 刘宇灏.微表情识别的理论和方法研究[D].南京: 东南大学,2016.

[5] 卢官明,杨成,杨文娟,等.基于 LBP - TOP 特征的微表情识别[J].南京邮电大学学报( 自然科学版) ,2017(6):1 - 7.

[6] 张轩阁,田彦涛,郭艳君,等.基于光流与 LBP - TOP 特征结合的微表情识别[J].吉林大学学报(信息科学版),2015,33(5):516 - 523.

[7] 吴进,董国豪,李乔深.基于区域卷积神经网络和光流法的目标跟踪[J].电讯技术,2018,58(1):6 - 12.

[8] 吴进,李乔深,闵育,等.一种基于 OpenCL 的 Lukas - Kanade 光流并行加速算法[J].电讯技术,2018,58(8):871 - 877.

[9] HINTON G E,OSINDERO S,TEH Y W.A fast learning algorithm for deep belief nets[J].Neural Computation,2014,18(7):1527 - 1554.

[10] HE K,GKIOXARI G,DOLLAR P,et al.Mask R - CNN[C]// Proceedings of 2017 IEEE International Conference on Computer Vision.Venice:IEEE,2017:2980 - 2988.

[11] SCHROFF F,KALENICHENKO D,PHILBIN J.Face - net:a unified embedding for face recognition and clustering[C]// Proceedings of IEEE 2015 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Boston: IEEE,2015: 815 - 823.

[12] REN S,HE K,GIRSHICK R,et al.Faster R - CNN:towards real - time object detection with region proposal networks [C]//Proceedings of 2015 International Conference on Neural Information Processing Systems. Istanbul: IEEE,2015:91 - 99.

[13] NAM H,HAN B. Learning multi - domain convolutional neural networks for visual tracking[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition.Seattle:IEEE,2016:4293 - 4302.

[14] 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. Amsterdam: ACM,2016: 382 - 386.

[15] 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.

[16] LI X,PFISTER T,HUANG X,et al.A spontaneous micro - expression database:inducement,collection and baseline[C]// Proceedings of 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition(FG).Shanghai:IEEE,2013:1 - 6.

[17] YAN W,WU Q,LIU Y,et al.CASME database:a dataset of spontaneous micro - expressions collected from neutralized faces[C]//Proceedings of 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition(FG).Shanghai:IEEE,2013:1 - 7.

[18] YAN W,LI X,WANG S,et al.CASME II:an improved spontaneous micro - expression dataset and the baseline evaluation[J].PloSoen,2014,9(1):1 - 8.

[19] KHOR H Q,SEE J,PHAN R C W,et al.Enriched long - term recurrent convolutional network for facial micro - expression recognition[C]//Proceedings of 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018).Xi´an:IEEE,2018:667 - 674.

[20] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]//Proceedings of the 32nd International Conference on Machine Learning.Lille:IEEE,2015:448 - 456.

[21] SIMONYAN K,ZISSERMAN A.Very deep convolutional networks for large - scale image recognition[C]//Proceedings of International Conference on Learning Representations 2015( ICLR 2015) .San Dieg:Computational and Biological Learning Society,2015:177 - 186.

[22] KRIZHEVSKY A,SUTSKEVER I,HINTON G .Imagenet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems,2012, 25(2):84 - 90.

[23] WANG Y D,SEE J,RAPHAEL C W,et al.Efficient sptio-temporal local binary patterns for spontaneous facial micro - expression recognition[J].Plos One,2013,10(5):11 - 12.


## 作者简介：


吴 进 女,1975 年生于江苏常州,2001 年获工学硕士学位，现为教授、硕士生导师,主要研究方向为信号与信息处理。

闵 育 女,1993 年生于陕西渭南，2017 年获工学学士学位，现为硕士研究生，主要研究方向为图像处理与行为识别。

李 聪 男,1995 年生于河南濮阳，2017 年获工学学士学位，现为硕士研究生,主要研究方向为图像处理和行为识别。

张伟华 男,1995 年生于陕西渭南，2018 年获工学学士学位，现为硕士研究生,主要研究方向为图像处理和行为识别。