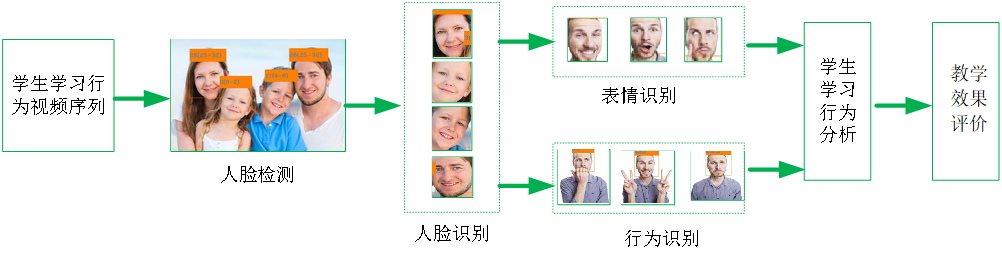
LBREM的设计思路，是首先通过学习设备和学习平台获取学生的学习行为视频信息；然后对视频中的学生人脸进行检测与识别，同时进行学生的表情识别和学习行为识别；最后实现对学生学习行为的分析，进而实现对教学效果的评价。基于此，本研究绘制了LBREM的运作流程，如图1所示。



**图1  LBREM的运作流程**

人脸检测是从图像中检测人脸并返回其位置信息，为进一步进行人脸识别做准备。传统的人脸检测方法主要有基于几何特征、肤色模型或结构部件的方法和基于统计理论的方法，这些方法性能较差、不能应用于复杂背景。而基于深度卷积神经网络人脸检测算法（Single Shot Scale-invariant Face Detector，S3FD）的单阶段检测（One-stage）方法将候选框提取、图像分类、边框坐标回归三个任务放在一个卷积神经网络中进行[4]，极大地提高了检测的速度，并达到了99.65%的检测精度，典型的相似方法还有单点无头人脸检测算法（Single Stage Headless，SSH）[5]和人脸网络识别（FaceNet）模型[6]。在人脸识别中，表情是最直接、最有效的情感识别模式。早在20世纪，Ekman等就通过跨文化调研提出了七类基础表情，分别是生气、害怕、厌恶、开心、悲伤、惊讶以及自然[7]。传统的手工设计特征乃至浅层学习特征不能很好地处理适应现实世界中光照变换、头部姿态以及面部阻挡等各种与表情无关的干扰因素，因此主流的人脸表情识别方法研究越来越多采用深度学习技术[8]。

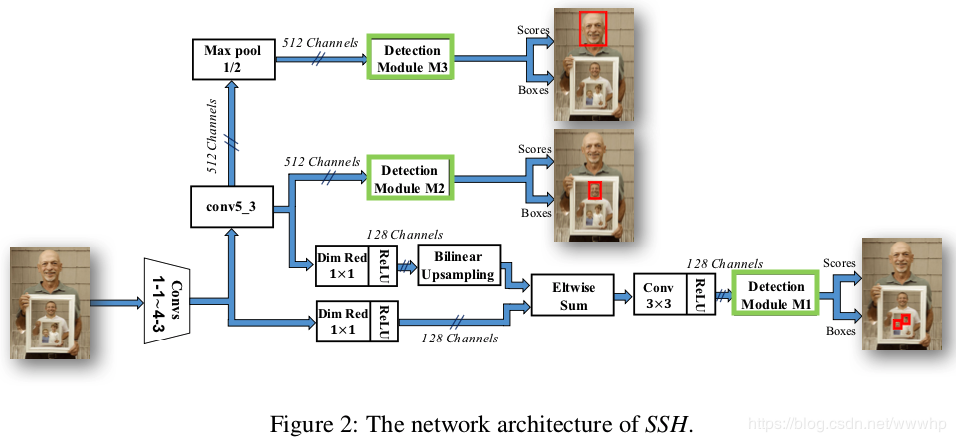
行为识别是检测学生上课状态的有力证据。关于人体行为识别的研究可以追溯到20世纪70年代，Johansson[9]提出了12点人体模型结构的描述方法，这种方法描述的人体行为，对后续算法产生了深远影响。目前，**人体行为识别可以分为两大类，即基于传统手工设计特征的分类算法和基于深度学习自动提取特征的分类算法[10]。当前关于人体行为的研究主要针对户外运动的大动作场景，而针对课堂学生学习行为的细节动作尚缺乏研究。**

三 学生的学习行为分析

为了对学生的学习行为进行快速检测和识别，LBREM采用人脸检测与人脸识别的技术方案：人脸检测采用SSH算法，实现对学生人脸的检测和定位；人脸识别采用FaceNet算法，实现对学生身份的识别——将SSH算法与FaceNet算法密切结合，对学生人脸检测和身份识别具有很好的效果。之后，LBREM根据人脸表情和身体姿态开展学生学习状态与学习行为的识别：**人脸表情采用卷积神经网络（Convolutional Neural Networks，CNN）识别方法，而身体姿态采用三维卷积神经网络（3D ConvNet）进行学习行为识别[11]——两者配合，能够很好地开展学生学习状态与行为的识别。**

1 学生人脸快速检测和身份识别

人脸检测中采用的SSH算法，能较好地实时检测学生人脸并具有较高的精度。SSH的网络结构如图2所示，是将VGG16网络中的全连接层全部去掉，只使用了卷积层，并在不同层次的卷积层输出特征图后面加了由分类器和回归器组成的检测模块M1、M2、M3，通过对不同尺度的特征图进行分析，从而实现多尺度的人脸检测。具体来说，对于VGG16网络中conv4\_3的输出使用M1进行检测，stride为8，检测小尺度人脸目标；对于VGG16网络中conv5\_3的输出使用M2进行检测，stride为16，检测中尺度人脸目标；而对于VGG16网络中conv5\_3池化后的输出使用M3进行检测，stride为32，检测大尺度人脸目标。



为了确保身份识别的精度，LBREM采用目前性能最好的FaceNet网络模型。FaceNet通过卷积神经网络学习将图像映射到欧几里得空间，空间距离直接和图片相似度相关：同一个人的不同图像在空间距离很小，不同人的图像在空间中有较大的距离。因此，只要该映射确定下来，相关的人脸识别任务就变得很简单了。

2 学生学习状态与行为识别

LBREM采用CNN开展表情识别，而表情可以很好地表达人类的情感活动，因此可以通过学生的表情识别来分析其学习行为。**LBREM将人脸表情分为七种，分别是愤怒、厌恶、恐惧、快乐、悲伤、惊讶、自然。其中，愤怒、厌恶、恐惧、悲伤表现为消极的学习行为，而快乐、惊讶、自然表现为积极和专注的学习行为**。

传统的人脸表情识别方法采用手工特征如LBP、BoW、HoG、SIFT等[12]，但由于在真实环境中光照和姿势是多种多样的，这使得传统的方法识别效果较差。考虑到表情识别的速度和准确度，LBREM采用基于CNN的表情识别方法，其人脸表情识别体系结构如图3所示。

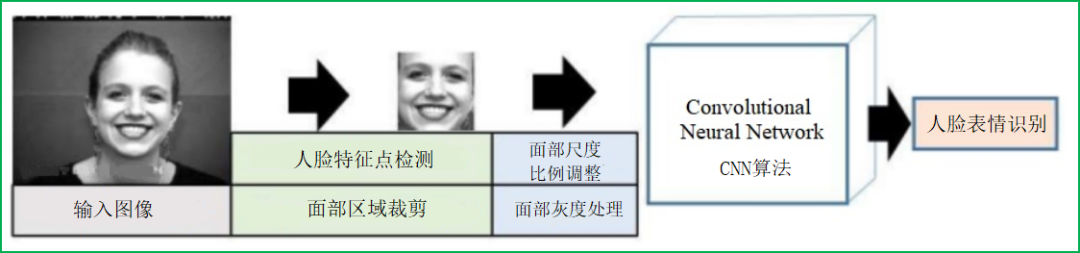


图3  表情识别方法的人脸表情识别体系结构

学生的学习行为表现是教学效果评价的重要组成部分，通过识别和评价学生的学习行为，形成有效的反馈信息和教学导向，可以有效促进教学质量的提升。通过观察大量的课堂学习行为，LBREM将学生的学习行为分为积极学习行为和消极学习行为两类，共囊括7种学习行为：积极学习行为包括端坐、书写、举手、起立，而消极学习行为包括趴桌、左顾右盼、玩手机。为了保证学生学习行为识别的准确性和实时性，本研究在现有人体行为识别算法的基础上对人体行为识别方法进行了改进和应用，即针对学生学习行为的视频信息，充分利用时序信息和时空特性，采用3D ConvNet进行人脸检测和行为识别。

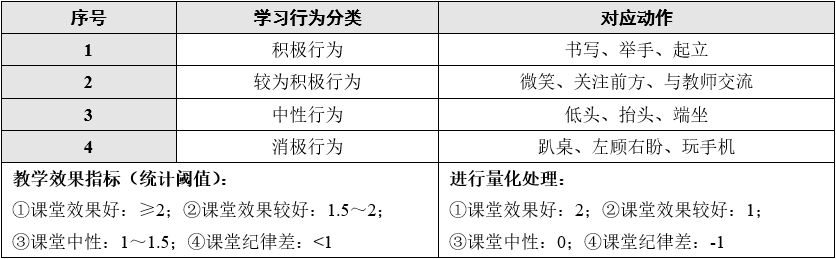
3 LBREM的识别准确率验证

为了验证LBREM的有效性，本研究在国际汉语互动课堂教学场景下的多人数据集上进行训练。该数据集包含400条5秒长度的视频，视频的原始分辨率为1920×1080；视频包含9名学生，按照预设规则交替完成7个动作：端坐、书写、趴桌、左顾右盼、举手、起立、玩手机，如图4所示。由于样本数据量较小，为了防止训练过拟合，**本研究模拟教室中光照条件的变化，采用亮度和对比度调节等方法进行了数据增广。另外，当前大多数人体行为识别方法都是面向单人目标进行研究，而课堂中需要识别的人体行为涉及多个目标。因此，为了满足课堂多人场景的应用，LBREM借鉴目标识别算法中的边界框预测方法，将每个学生的边界框划分出来，并对每个目标进行肢体关键点的识别，以更加准确地判断出学生当前的学习行为类别。**

1 TEM方法设计

如前文所述，LBREM将学生的学习行为分为积极学习行为和消极学习行为两类，其中积极学习行为包括端坐、书写、举手、起立四种，而消极学习行为包括趴桌、左顾右盼、玩手机三种。在此基础上，TEM方法将学生的学习行为进一步细化，分为积极学习行为、较为积极行为、中性行为和消极学习行为四类：书写、举手、起立为积极学习行为；微笑、关注前方、与教师交流为较为积极行为；低头、抬头、端坐为中性行为；而趴桌、左顾右盼、玩手机为消极学习行为。由此，本研究得到TEM方法中的教学效果分析指标，如表4所示。

表4  学生学习行为分析指标

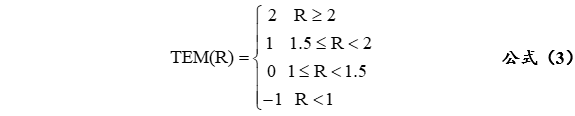


IMG_257

按照公式（1）统计所有学习行为指标的次数Np，分别对应积极学习行为、较为积极行为、中性行为和消极学习行为。其中，fp(k)表示第K帧图像中对应的学习行为指标的人数， fp(k)=S。

根据学习行为统计情况建立课堂教学效果评价方法TEM，评价依据是学习行为指标与学习行为总数的比例R，R的计算如公式（2）所示。

则课堂教学效果评价方法可以定义为TEM(R)，具体评价得分规则如公式（3）所示。其中，TEM(R)=2表示课堂学习效果好，1表示较好，0表示一般，而-1表示较差。



3 教学效果实证

为验证以上模型的实证效果，实证使用25堂国际汉语教育的实时课堂总结1125分钟的视频动作进行分析，同时邀请授课教师、学生、同行和领导对课堂采用多种主体主观评价指标和LBREM方法进行对比评估分析，对25堂课程分别选取5、10、15、20、25堂课程对行评价对比分析，统计结果如表5所示，从该表可以看出两种方法在课堂效果积极因素方面（课堂效果好与课堂较好累计）表现高度一致。

表5  评价方法实证数据列表



LBREM方法定义了统计教学效果评价方法TEM，通过课堂学习中学生学习行为的分布统计实现课堂教学效果评价，评价效果绘制雷达图（Teaching Effect View，TEV）。假设一次课堂学习时间是T，听课学生人数为S，为了保证学生学习行为识别和统计的高效性，不需要对学习视频每帧图像进行识别和统计，而是采取等间隔时间进行识别和统计，设识别的间隔时间为t，则识别帧数F=t/T。

绘制两种方法在对应25堂课程时的雷达图TEV，如图6（a）所示，两种在不同的课堂效果TEV图形高度叠加，表现也高度一致，其中LBREM方法稳定性更好些，对于LBREM方法，以阈值为100进行数据量化，绘制雷达图TEV的结果分析如图6（b）、图6（c）、图6（d）、图6（e）所示，雷达构成的面积为正数值，值越大课堂效果越好，反之会越差。一般来说，高质量的课堂学习可以更好的吸引学生，从而产生积极的学习行为，反之将产生消极的学习行为。

五 结论与展望

与传统课堂教学相比，基于远程开放教育学生互动课堂学习已成为今后教育发展的趋势，其教学过程更加复杂，学生学习具有四个特性：时空不限性，学习的自主性，学习路径的非线性，学习材料的多媒化，因此这种教育学习行为分析和教学效果评价会更加多样性和综合性，使用的人工智能技术也就更加重要。本研究提出的LBREM方法能够快速、准确的识别学习者学习行为和教学效果评价，其中教学效果评价TEM方法与多种主体评价方法表现效果高度一致，但效率更高，LBREM方法主要依赖于视觉行为，对于智慧课堂、远程学习、移动学习、MOOC等远程互动课堂和在线学习的教学效果评价和个性化推荐，具有很好的借鉴意义。考虑到未来教育中，基于教育的趋势及学生学习的特性，学习行为还需考虑学生互动状态、网络点击行为和学习日志等网络行为，下一步将从海量、高维、异构、分布式特征的新型维度空间开放教育数据中挖掘有效信息，引入学习者学习行为的因素数据构建行为预测模型，同时研究学校评价指标形成一个有效的教育学习行为分析与评价模型，有效解决目前多样化的新维度空间下教育资源的合理管理问题，提高未来教育学生学习的效率。

下周找导师两次，12月份想办法跑出个结果，一月份初准备初稿，一月中结束初稿

论文的一些解决方案--改为占比，变成对教师的课堂评价

1、从视频中提取照片

2、人脸表情识别和定位

3、学生行为识别

4、课堂状态分析

|  |  |  |
| --- | --- | --- |
| **问题** | **解决方法（算法）** | **备注** |
| 从视频中提取照片并标记 |  |  |
| 人脸表情识别和定位 | 深度卷积神经网络人脸检测算法（Single Shot Scale-invariant Face Detector，S3FD） | 将候选框提取、图像分类、边框坐标回归三个任务放在一个卷积神经网络中进行 |
| 学生行为识别 |  | 积极行为（抬头）、中性行为（低头）、消极行为（转头） |
| 课堂状态分析 | 对老师的分析（占比） |  |
|  |  |  |