

## 基于学生人体检测的无感知课堂考勤方法

方书雅, 刘守印\*

(华中师范大学 物理科学与技术学院, 武汉 430079)

(\* 通信作者电子邮箱 syliu@mail.ccnu.edu.cn)

**摘要:**针对基于人脸识别的课堂考勤系统漏检和低识别率的问题,采用主、从双摄像机设备,提出一种联合学生人体检测和人脸角度筛选的方法。首先通过Mask R-CNN算法检测主摄像机拍摄图中的学生人体位置;然后控制从摄像机(PTZ相机)依次获取每位学生的高质量放大图像;再通过MTCNN算法和FSA-Net算法从中检测并识别出人脸姿态,筛选出每位学生的正脸图像;最后对筛选出的学生正脸图像使用FaceNet算法提取人脸特征,用于支持向量机(SVM)分类器的训练或识别。实验结果表明,与Tiny-face算法相比,人体检测算法在重叠比(IOUS)为0.75时平均精度(AP)值提高了约36%且检测耗时减少了57%;与建立多姿态人脸数据库的方法相比,采用人脸角度筛选的方法使识别率提高了4%;多数情况下整个课堂学生识别的准确率接近100%。所提方法简化了学生注册过程,提高了人脸识别率,为解决人脸漏检问题提供了新的思路。

**关键词:**课堂考勤;目标检测;人脸姿态;人脸识别;人体检测

**中图分类号:**TP183 **文献标志码:**A

### Non-perception class attendance method based on student body detection

FANG Shuya, LIU Shouyin\*

(College of Physical Science and Technology, Central China Normal University, Wuhan Hubei 430079, China)

**Abstract:** Concerning the missed detection and low recognition rate in the class attendance system based on face recognition, a method that combines student body detection and face angle filtering was proposed by applying the master and slave dual-camera device. First, the bodies of students were detected from the photograph of master camera by the Mask R-CNN algorithm. Then, the slave camera (PTZ (Pan/Tilt/Zoom) camera) was controlled to acquire high-quality magnified image of each student in turn. Next, the face poses were detected and recognized in the magnified images through MTCNN (Multi-Task Convolutional Neural Network) algorithm and FSA (Fine-grained Structure Aggregation)-Net algorithm in order to filter the frontal face image of every student. Finally, the FaceNet algorithm was used to extract the features of the filtered student frontal face images for training or recognition of Support Vector Machine (SVM) classifiers. Experimental results showed that, compared with the Tiny-face algorithm, when the Intersection over Union (IOUS) was 0.75, the body detection algorithm had the Average Precision (AP) value increased by about 36% and the detection time reduced by 57%; compared with the method of establishing a multi-pose face database, the method of face angle filtering improved the recognition rate by 4%; and the accuracy of student recognition in the entire classroom was close to 100% in most cases. The proposed method simplifies the student registration process, improves the face recognition rate, and provides new ideas for solving the problem of face missed detection.

**Key words:** class attendance; object detection; face pose; face recognition; body detection

## 0 引言

文献[1-3]中提出高校学生上课率是影响其学习成绩的重要因素之一。近年来,文献[4-8]研究基于人脸识别的课堂考勤系统,用来确认学生是否到课堂听课,减少教师人工点名的时间开销。目前单一人脸识别技术比较成熟。但由于课堂环境特殊,人脸识别技术在应用到课堂考勤系统中存在许多难点。首先由于学生数目较多,教室的面积很大;其次因为教室采光和通风的要求,导致教室的窗户比较多,教室内光强分布极度不均匀;最后学生在上课时的姿态千变万化。这些问题都给课堂考勤中人脸检测和人脸识别的应用带来挑战。

基于人脸识别的课堂考勤系统依赖于人脸检测的性能。从一张包含众多学生、学生姿势各异、光线不均匀的视频图片中检测出所有学生的人脸,是一项技术挑战。如果不能检测出所有学生的人脸,势必会遗漏上课的学生。为了避免漏检,Balcoh等<sup>[4]</sup>使用皮肤分类技术辅助Voila-Jones人脸检测算法<sup>[9]</sup>,提高了检测过程的效率和准确性。Fu等<sup>[5]</sup>采用基于深度学习方法MTCNN (Multi-Task Convolutional Neural Network)算法<sup>[10]</sup>进行人脸检测,该算法具有良好的鲁棒性,可以检测出不同姿态的人脸,但是仍然会存在漏检。刘宇明等<sup>[11]</sup>优化了MTCNN算法<sup>[10]</sup>并引入时空特征实现对参会人员

收稿日期:2020-02-04;修回日期:2020-03-02;录用日期:2020-03-10。 基金项目:湖北省高等学校省级教学研究项目(2018089)。

作者简介:方书雅(1995—),女,湖北襄阳人,硕士研究生,主要研究方向:深度学习、图像处理;刘守印(1964—),男,河南周口人,教授,博士,主要研究方向:无线通信、物联网、机器学习。

跟踪检测和人数统计的功能。Sarkar 等<sup>[6]</sup>使用 tiny-face 算法<sup>[12]</sup>用于检测不同尺度、不同姿态的人脸,极大地提高了小尺度人脸的检测率。虽然上述这些系统使用的人脸检测方法在准确率方面有大幅度提升,但是这些方法<sup>[4-8]</sup>在检测过程中仍需要学生主动面向摄像机且保证自己不被其他同学遮挡。这不仅会干扰学生和老师的正常上课,同时也不利于后续在自然状态下对学生的课堂行为和表情进行研究。

通过人脸检测方法在检测到学生人脸后,需要将人脸区域的图像裁剪出来,用于人脸识别。人脸识别的性能受人脸姿势多样性(低头、倾斜角度、遮挡等)的影响。龚锐等<sup>[13]</sup>提出建立多姿态人脸数据集的方法用于人脸识别。Surekha 等<sup>[8]</sup>将多姿态人脸数据集的方法应用到课堂领域,在注册时采集每个学生不同姿态、表情、光照下的多张图片。但这种方法的弊端是注册时需要学生配合,对每个学生拍摄视频并选择合适的多张图片。这不仅增加了额外的人工开销,而且没有在进行识别过程中考虑人脸姿态多样性的影响。

同时,人脸图像质量也是对人脸识别的准确性造成影响的因素之一。如果学生人脸在拍摄的视频图片中尺寸较小,被裁剪出的人脸图像再被放缩到统一尺寸时,会出现图像模糊、图像质量变差的问题,从而影响识别的准确率。目前,针对低分辨率图像的人脸识别技术<sup>[14-15]</sup>尚未应用到课堂识别领域。对此方冠男<sup>[16]</sup>采用 PTZ (Pan/Tilt/Zoom) 摄像机对教室进行分区定点拍摄的方法提高图像质量,但是识别性能有待提高。

为解决上述问题,在基于人脸识别的课堂考勤系统的基础上,本文提出一种联合学生人体检测和人脸角度筛选的方法,首先有效降低了漏检率,其次简化了学生人脸注册数据集的采集过程,在降低注册需要的人脸图片数量的同时提高了识别的准确率,实现无感知、低漏检、高准确的课堂学生识别功能。该方法为解决课堂学生漏检问题提供了新的思路。

## 1 相关技术

### 1.1 Mask R-CNN 目标检测算法

Mask R-CNN 算法<sup>[17]</sup>将目标检测中的 Faster R-CNN 算法<sup>[18]</sup>和语义分割领域的 FCN (Fully Convolutional Network) 算法<sup>[19]</sup>相结合,形成一个可同时进行目标检测和实例分割的多任务模型。Faster R-CNN 对每个候选对象有两个输出,一个类标签和一个边界框;Mask R-CNN 在其基础上添加了输出对象掩码的第三个分支,对每个感兴趣区域 (Region of Interest, RoI) 上采用一个小的 FCN 实现逐像素点的方法预测分割掩码。Mask R-CNN 算法的创新之处在于提出 RoIAlign 方法代替了 RoIPool<sup>[20]</sup>方法,采用双线性插值实现了像素级别的对齐。

Mask R-CNN 算法的优点是训练速度快,仅比 Faster R-CNN 增加少量开销,运算速度为 5 帧每秒 (Frames Per Second, FPS)。使用 ResNeXt-101-FPN 作为骨架的 Mask R-CNN 在 COCO 实例分割任务上获得 2016 年 MS COCO 比赛的冠军,在目标检测任务上取得了很好的效果。权衡运算速度和检测准确率,选择 Mask R-CNN 算法检测学生人体。

### 1.2 MTCNN 人脸检测算法

MTCNN 算法<sup>[10]</sup>提出了一个深度级联多任务框架,利用人脸检测与对齐的内在相关性来提高性能。该模型采用了 3 个深度卷积神经网络,并通过由粗糙到细致的方式来预测脸部整体和特征点的坐标。这 3 个级联的网络分别是快速生成候选窗口的 P-Net (Proposal Network)、进行高精度候选窗口过滤

选择的 R-Net (Refine Network) 和生成最终边界框与人脸关键点的 O-Net (Output Network)。该模型也用到了图像金字塔、边框回归、非极大值抑制等技术。MTCNN 算法准确率高、检测速度快,被广泛应用在人脸检测领域<sup>[21-22]</sup>。

### 1.3 FSA-Net 头部姿态识别算法

FSA (Fine-grained Structure Aggregation)-Net 算法<sup>[23]</sup>提出一种直接使用回归方法的、无标志的姿态估计模型,用于单张图像的头部姿态估计。模型的核心思想是将特征图在像素级别上的特征组合成一组空间信息编码的特征,作为聚合的候选特征。该方法的目的是学习寻找细粒度的结构映射,以便将像素级别的特征在空间上进行分组以形成更多的幂域级特征。

FSA-Net 算法框架如图 1 所示。输入的图像通过两个支流、三个阶段进行提取特征。每个支流在每个阶段提取一个特征图。在每个阶段,将两个支流提取的特征图融合在一起。对每个阶段融合的特征先计算它的注意力图,然后将特征图和注意力图一起被输入到细粒度结构特征聚合模块中。该模块通过对特征图中的像素级别特征进行空间加权的方式来编码特征,然后将这些特征聚合以生成用于回归的最终代表特征集。最后将这些输出代入软阶段回归网络 (Soft Stages Regression network, SSR) 函数以获得姿态估计。

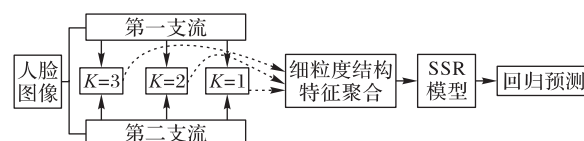


图 1 FSA-Net 框架

Fig. 1 FSA-Net framework

FSA-Net 算法在 AFLW2000 (Annotated Facial Landmarks in the Wild 2000) 数据集和 BIWI (Biwi Kinect Head Pose Database) 数据集上的结果优于现有的方法 (基于地标的方法和无地标的方法), 而其模型尺寸是以前方法的 1/100 左右。因此本文选用 FSA-Net 算法进行头部姿态识别。

## 2 无感知课堂考勤系统及方法

本系统增加学生人体检测模块和人脸角度筛选模块,采用主、从双摄像机设备,分别实现教室全景图的拍摄和定点图像缩放拍摄的功能。主摄像机拍摄的全景图像作为学生人体检测模块的输入,计算分析课堂实到学生的数量和位置信息。根据学生位置信息,系统控制从摄像机对每个学生定点缩放拍摄,以获取高清人脸图像。最后对获取的人脸图像先通过人脸角度筛选,选出合格的正面人脸图像后再进行识别。

相较于其他人脸考勤系统,本系统有如下 4 个优点:

- 1) 解决了在单次或有限次数的人脸识别过程中,因人脸信息缺失造成的漏检问题。
- 2) 可以获取到每位学生高清放大的人脸图像。
- 3) 在学生无感知的情况下获取学生高清的正面人脸图像用于注册,极大简化了传统的注册过程。
- 4) 在提高识别率的同时,减少了每个学生注册时需要的图像数量。

### 2.1 系统框架

系统在学生无感知、不影响学生正常上课的情况下实现学生注册和出勤记录功能。注册阶段不需要额外的设备,也不需要课下逐个采集学生图像信息。具体框架如图 2 所示,系统在注册和识别阶段采用相同的处理方式,都经过学生人

人体检测、云台参数估计和人脸角度筛选的步骤。不同点在于,注册过程在初次课程教学时完成;识别阶段则在后续课程教学时实时进行。下面对图2中的硬件和软件进行详细介绍。

系统的硬件部分由设备端和服务端组成。设备采用主、从双摄像机实现图像和视频的采集。其中主摄像机是定焦广角摄像机,负责拍摄教室全景图;从摄像机是PTZ摄像机,可通过参数控制从摄像机对感兴趣的区域定点拍摄,并能对区域的图像进行缩放处理。系统通过控制主、从双摄像机,可以获得教室内每位学生高清放大的人脸图像。

软件功能在服务器端实现,包括学生人体检测、云台参数估计、人脸角度筛选、注册和识别部分。首先,采用Mask R-CNN算法对主摄像机拍摄的教室全景图进行学生人体检测,确定实到学生数量和每个学生的位置;然后将学生位置信息输入到多元线性回归模型中,估计出云台在垂直方向和水平方向上的转动角度和变焦倍数这3个参数。通过参数控制云台转动,使从摄像机对准目标学生。接着对从摄像机拍摄的区域图进行基于FSA-Net算法的人脸角度筛选,筛选出清晰、无遮挡的正面人脸。最后根据情况选择进入注册或者识别阶段。两个阶段均采用FaceNet算法<sup>[24]</sup>对筛选的人脸图像进行特征提取,在注册阶段使用提取的特征训练支持向量机(Support Vector Machine, SVM)的分类器;在识别阶段将提取的特征输入到训练好的SVM模型中进行预测以确定学生身份。当每个学生位置依次进行识别后,系统输出本节课出勤学生名单。

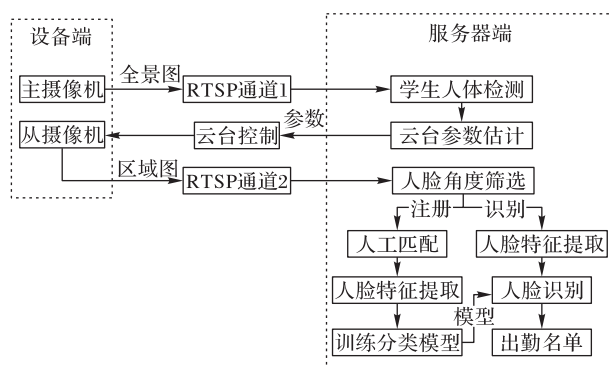


图2 系统框架

Fig. 2 System framework

## 2.2 无感知课堂考勤方法

### 2.2.1 学生人体检测

现有课堂考勤系统均采用基于人脸的检测方法,其缺点是检测过程中完全依赖人脸信息。一旦出现学生人脸信息缺失的情况(如学生低头看书或者记笔记、学生侧脸、回头或者被其他学生遮挡等情况),学生即使出现在课堂上也会被判定为缺勤。而相较于人脸,人体面积更大,包含的信息更多,被全部遮挡住的概率更小。故本系统在人脸检测的基础上增加学生人体检测模块,目的是先确定整个课堂实到学生的数量和位置信息,避免因人脸信息缺失造成的漏检。

学生人体检测采用目标检测中的Mask R-CNN算法。输入图片 $I_j$ ,Mask R-CNN算法的标准输出为 $\{(b_j, mask_j, c_j, score_j)\}_j$ , $j$ 是检测到的目标索引,向量 $b_j$ 是边界框, $mask_j$ 是掩码, $c_j$ 是类标签, $score_j$ 是得分。其中, $b_j=[y_{1j}, x_{1j}, y_{2j}, x_{2j}]$ 分别代表边界框的左上顶点和右下顶点在 $y$ 轴和 $x$ 轴方向上的图像坐标。Mask R-CNN算法可以检测80种不同类别的物体实例。学生人体检测方法将教室全景图作为输入,筛选Mask R-CNN输出中类标签为“person”的边界框集合 $B=\{b_1, b_2, \dots, b_i, \dots, b_N\}$ 作为学

生人体检测的结果, $N$ 表示检测到的学生个数。

### 2.2.2 从摄像机控制

通过学生人体检测确定到课学生的数量和位置后,需要对学生身份进行识别。为获取学生的高质量图像以提高识别的准确率,系统根据学生位置信息控制从摄像机,依次定位到每个学生,并对学生图像进行缩放,获取高清的视频图片。

通过设置垂直和水平方向上的转动角度以及变焦倍数这三个参数,系统可以控制从摄像机转动到指定位置并对图像进行缩放处理。为了根据学生的位置信息得到控制从摄像机运动的三个参数,本文建立了多元线性回归模型,其表达式和展开形式如式(1)所示。其中 $i$ 是学生的索引号,每个学生的人体边界框记为向量 $b_i$ ,矩阵 $A$ 为回归系数项,向量 $c$ 为随机误差项,模型输出为从摄像机在垂直方向和水平方向上的转动角度 $p_i$ 和 $t_i$ ,以及变焦倍数 $z_i$ 。

$$\begin{bmatrix} p_i \\ t_i \\ z_i \end{bmatrix} = A \cdot b_i^T + c = \begin{bmatrix} a_{11} & a_{11} & a_{11} & a_{11} \\ a_{11} & a_{11} & a_{11} & a_{11} \\ a_{11} & a_{11} & a_{11} & a_{11} \end{bmatrix} \cdot \begin{bmatrix} y_{1i} \\ x_{1i} \\ y_{2i} \\ x_{2i} \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \quad (1)$$

通过这3个参数控制云台转动,使从摄像机对准本张全景图中的第 $i$ 个学生,并在后续过程中识别该学生。

### 2.2.3 人脸角度筛选

本文通过实验发现上课时学生不会一直抬头看向黑板或者教师(面朝摄像机),会出现侧脸、低头或前后排学生遮挡。在对每个学生进行识别的过程中,若采取实时检测,则会造成严重的系统负担;若随机选择一张或者多张图像进行人脸检测,则无法保证图像包含有效的人脸信息。因此,在人脸识别前,系统会对检测到的人脸进行角度识别筛选出正面人脸图像。具体流程如图3所示。

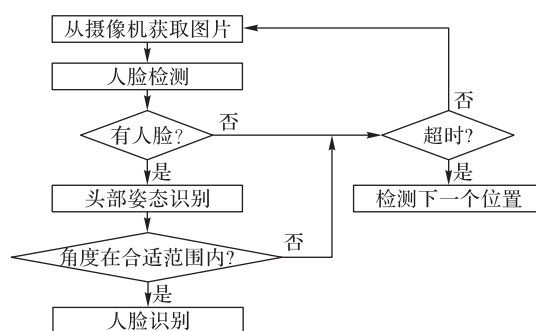


图3 人脸角度筛选的流程

Fig. 3 Flowchart of face angle filtering

系统对从摄像机获取的第 $i$ 个学生的图像依次进行人脸检测和人脸角度筛选,将获取到的第 $i$ 个学生的第 $r$ 图像记为 $X_{ir}$ 。先采用MTCNN算法对图像 $X_{ir}$ 进行人脸检测,输出 $160 \times 160$ 维的人脸图像 $F_{ir}$ ;再通过FSA-Net头部姿态识别算法对人脸图像 $F_{ir}$ 进行角度值的预测,输出偏航角 $y_{ir}$ 、俯仰角 $p_{ir}$ 、翻滚角 $r_{ir}$ 。如式(2):

$$F_{ir} = MTCNN(X_{ir})$$

$$[y_{ir}, p_{ir}, r_{ir}] = FSA-Net(F_{ir}) \quad (2)$$

如果3个角度的取值范围满足式(3)或(4)的条件,则认为第 $i$ 个学生的第 $r$ 人脸图像 $F_{ir}$ 是合适的正面人脸图像,可用于后续的训练或者识别:

$$|p_{ir}| < 20 \text{ 且 } |y_{ir}| < 20 \text{ 且 } |r_{ir}| < 15 \quad (3)$$

$$|20 < p_{ir}| < 28 \text{ 且 } |y_{ir}| < 20 \text{ 且 } |r_{ir}| < 10 \quad (4)$$



### 3 实验及分析

为验证学生人体检测方法性能优于主流的人脸检测算法,设计了学生人体检测实验。针对人脸角度筛选方法能够降低注册数据量、提高识别准确率的说法,设计了人脸角度筛选实验。最后对系统进行整体测试。三个实验采用不同的评价指标,验证基于学生人体检测的无感知课堂考勤方法在不同评价指标上的性能表现和提升。

#### 3.1 评价指标

学生人体检测实验将比较人脸检测算法和Mask R-CNN算法的性能。因二者都属于目标检测领域,故该实验采用目标检测算法最常用的指标:检测速度和MS COCO目标检测挑战中使用的 $AP_{coco}$ 、 $AP_{coco}^{IOU=0.5}$ 、 $AP_{coco}^{IOU=0.75}$ 、评估精度。 $AP_{coco}^{IOU=0.5}$ 、 $AP_{coco}^{IOU=0.75}$ 分别表示IOU设置的阈值为0.5和0.75时的平均精度(Average Precision, AP)。 $AP_{coco}$ 是IOU设置的阈值从0.5~0.95,以0.05的步长,分别计算AP最后将所有结果取的均值。重叠比(Intersection Over Union, IOU)的计算公式如下:

$$IOU(b, b_g) = \frac{area(b \cap b_g)}{area(b \cup b_g)} \quad (5)$$

设学生人体检测或人脸检测预测的任意一个边界框为向量 $b$ ,其对应的真实边界框为向量 $b_g$ 。

人脸角度筛选实验和系统验证实验分别采用平均精度均值(Mean Average Precision, MAP)和准确率 $\eta$ 作为评价指标,如式(6)和(7)所示:

$$MAP = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FP_i} \quad (6)$$

$$\eta = \frac{TP + TN}{N} \quad (7)$$

其中: $N$ 表示学生个数, $TP$ (True Positive)是正确检测到的出勤学生数量, $FP$ (False Positive)是检测错误的学生, $TN$ (True Negative)是正确检测到的缺席学生数量。

#### 3.2 学生人体检测实验

实验对比了本文采用的Mask R-CNN算法和课堂考勤系统中常用的三种人脸检测算(Voila-Jones<sup>[9]</sup>、MTCNN<sup>[10]</sup>、Tiny-face<sup>[12]</sup>)在平均精度和检测速度上的差异。数据采集了同一间教室12个课堂的200张全景图片。采集方式为在学生正常上课的情况下,每2 min保存一张全景图片,并对每一张全景图内的学生人脸和人体分别进行人工标注作为验证集。实验结果如图4和表1所示。

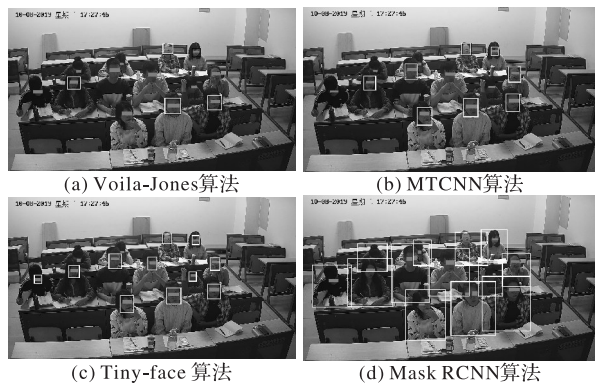


图4 针对某图的4种算法检测结果

Fig. 4 Detection results of four algorithms of a certain photograph

表1是4种算法根据评价指标在200张数据集上的实验结果。除Voila-Jones算法,其他3种深度学习的算法均使用GPU进行加速。从检测速度上看,传统的Voila-Jones算法在不适用GPU加速时的速度明显比深度学习算法检测速度上高一个数量级。但是在检测精度上,深度学习算法要明显优于Voila-Jones算法。

表1 四种算法结果

Tab. 1 Results of four algorithms

| 算法          | FPS  | $AP_{coco}$ | $AP_{coco}^{IOU=0.5}$ | $AP_{coco}^{IOU=0.75}$ |
|-------------|------|-------------|-----------------------|------------------------|
| Voila-Jones | 14.6 | 18.5        | 39.9                  | 9.9                    |
| MTCNN       | 1.9  | 42.2        | 76.0                  | 41.9                   |
| Tiny-face   | 0.6  | 48.5        | 91.7                  | 44.5                   |
| Mask R-CNN  | 1.4  | 65.2        | 91.8                  | 80.2                   |

在3种人脸检测算法中,Tiny-face算法在 $AP_{coco}$ 、 $AP_{coco}^{IOU=0.5}$ 、 $AP_{coco}^{IOU=0.75}$ 三个指标上都明显高于其他2种人脸检测算法。与MTCNN人脸检测算法相比,Tiny-face算法在 $AP_{coco}^{IOU=0.5}$ 指标上具有明显优势,高出约15个百分点。虽然Tiny-face算法在 $AP_{coco}^{IOU=0.5}$ 指标上与Mask R-CNN算法效果相同,但是Mask R-CNN算法能够兼顾在不同IOU阈值下的AP值、 $AP_{coco}$ 、 $AP_{coco}^{IOU=0.75}$ 指标上分别提升了约17个百分点和36个百分点。这是由于人体比人脸面积更大,在一定程度上降低了检测难度。与Tiny-face算法相比,Mask R-CNN算法耗时减少了57%。

图4是对同一张教室全景图采用不同算法的检测结果,其中的矩形框是检测到的人脸或人体的边界框。为保护学生的隐私权,对结果图片进行了人工打码。图4(a)、(b)、(c)三种人脸检测算中,检测效果最好的是Tiny-face算法,只漏检了最后一排低头和被遮挡的两个学生。Voila-Jones算法和MTCNN算法分别漏检了8个和5个学生。图4(d)采用Mask R-CNN人脸检测算法的检测结果优于Tiny-face算法,所有学生都被检测到。

因此Mask R-CNN算法在精度和检测速度上都优于Tiny-face人脸检测算法。从实验结果可知,本文提出的学生人体检测方法可以减少因为人脸信息缺失造成的学生漏检。

#### 3.3 人脸角度筛选实验

人脸角度筛选方法可以用在注册阶段和识别阶段。本实验的目的是验证在两个阶段均采用人脸角度筛选方法会提高最终识别的准确率,并减小注册时的数据量。本实验的对象为某课堂的14名学生。以该课堂第一次上课时采集的数据作为训练集,后续三周的数据作为人脸识别的测试集。

在注册阶段,采用人脸角度筛选方法为14名学生每人筛选出10张正面人脸图像,共计140张,制作成筛选训练集Train-With10。在后续三周的课堂上,使用同样的方式采集每名学生的正面人脸图像共计420张,制作成筛选测试集Test-With10。为进行对照实验,在注册和识别阶段,通过MTCNN检测算法对每名学生的视频流进行裁剪,再由人工为每名学生筛选出20张不同角度的人脸图像,制作成多姿态人脸数据集Train-Without20和Test-Without20。数据集的详细介绍如表2所示。因为经过人脸角度筛选后的人脸图片比多姿态人脸图片少,所以两种类型的数据集的大小不同。

图5为某学生的8张人脸图片:第一行的4张图片来自于

筛选数据集;第二行的4张图片来自于多姿态人脸数据集。由图可知,筛选数据集中的图像为正面人脸图像,不同图像中的人脸角度差别较小;而在多姿态人脸数据集中,有回头、侧脸和低头的人脸图像,人脸角度差异较大。

表2 数据集介绍

Tab. 2 Introduction of datasets

| 数据集             | 每人每节课图片数 | 共计张数 |
|-----------------|----------|------|
| Train-Without20 | 20       | 280  |
| Train-With10    | 10       | 140  |
| Test-With10     | 10       | 420  |
| Test-Without20  | 20       | 840  |

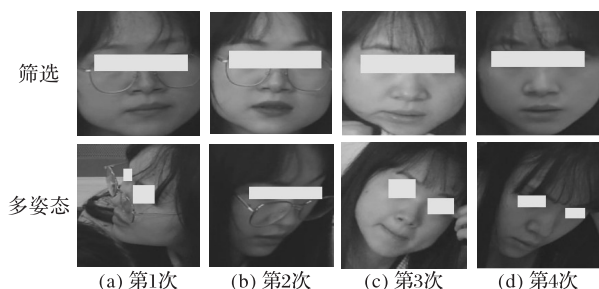


图5 两种数据集图片样例

Fig. 5 Sample images of two datasets

实验使用训练集 Train-With10 和 Train-Without20 分别采用 Facenet 算法<sup>[24]</sup>训练对应的人脸分类模型,并在 Test-With10 和 Test-Without20 这两个测试集上测试其性能。实验结果如表3所示,两个模型在筛选过的 Test-With10 测试集上的 MAP 值均高于不筛选的 Test-Without20 测试集,分别提高了1个百分点和5个百分点。在筛选过的测试集上,使用筛选过的训练模型(Train-With10 模型)的 MAP 值比不筛选的模型(Train-Without20 模型)提高了3个百分点。

表3 模型测试结果

单位: %

Tab. 3 Model test results

unit: %

| 模型                 | Test-With10 | Test-Without20 |
|--------------------|-------------|----------------|
| Train-Without20 模型 | 91          | 90             |
| Train-With10 模型    | 94          | 89             |

在表3中, Train-Without20 模型在 Test-Without20 测试集上的准确率为90%,其物理含义是在注册和识别时均采用多姿态人脸方法的准确率为90%;同理, Train-With10 模型在 Test-With10 测试集上的准确率表示在注册和识别时均采用人脸角度筛选方法的准确率为94%。由此可知,在注册和识别过程中均采用人脸角度筛选方法,与多姿态人脸注册和识别的方法相比,可以将识别准确率提高4个百分点。

为证实采用人脸角度筛选的方法可以减少学生注册时的数据量,将训练集 Train-With10 中每个人的图片从10张减少至5张,记为训练集 Train-With5。训练集 Train-With10 和 Train-With5 分别用于模型训练并在 Test-With10 测试集上进行测试。两个模型在测试集上的 MAP 值均为94%。因此在保证识别率的同时,可以减少数据量,对每个同学只采集5张图片制作训练集。

### 3.4 系统验证实验

为测试整个系统的性能,对3.3节的同一实验对象进行

实时课堂识别。一共检测了6节课,每次上课时间相隔一周。6次识别中5次识别的准确率为100%;1次为92.8%,即一位同学被误判。从系统实验结果分析,本系统识别的准确率较高。

## 4 结语

本文提出一种联合学生人体检测和人脸角度筛选的方法,改进现有系统的整体框架和学生注册方案。研究表明人体检测方法能够有效缓解因人脸信息缺失造成的人脸漏检问题;在注册和识别阶段均采用人脸角度筛选方法可以减少注册需要的数据量,提高识别准确率,在课堂应用场景下优于建立多姿态人脸数据集的方法。本文为课堂学生出勤率检测提供了一种新的解决思路,具有一定的实践意义。在后续课堂人脸识别的研究中,可以考虑使用超分辨率图像重建技术提高人脸识别的准确率、降低硬件成本。

### 参考文献 (References)

- [1] NORDMANN E, CALDER C, BISHOP P, et al. Turn up, tune in, don't drop out: the relationship between lecture attendance, use of lecture recordings, and achievement at different levels of study[J]. Higher Education, 2019, 77(6): 1065-1084.
- [2] 万强,王海晖,周剑秋. 基于大数据的教学分析[J]. 中国科技信息, 2017(5): 40-43. (WAN Q, WANG H H, ZHOU J Q. Teaching analysis based on big data [J]. China Science and Technology Information, 2017(5): 40-43.)
- [3] TEIXEIRA A A C. The impact of class absenteeism on undergraduates' academic performance: evidence from an elite Economics school in Portugal [J]. Innovations in Education and Teaching International, 2016, 53(2): 230-242.
- [4] BALCOH N K, YOUSAF M H, AHMAD W, et al. Algorithm for efficient attendance management: face recognition based approach [J]. International Journal of Computer Science Issues, 2012, 9(4): 146-150.
- [5] FU R, WANG D, LI D, et al. University classroom attendance based on deep learning [C]// Proceedings of the 10th International Conference on Intelligent Computation Technology and Automation. Piscataway: IEEE, 2017: 128-131.
- [6] SARKAR P R, MISHRA D, SUBHRAMANYAM G R K S. Automatic attendance system using deep learning framework [M]// TANVEER M, PACHORI R. Machine Intelligence and Signal Analysis, AISC 748. Singapore: Springer, 2019: 335-346.
- [7] MERY D, MACKENNEY I, VILLALOBOS E. Student attendance system in crowded classrooms using a smartphone camera [C]// Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2019: 857-866.
- [8] SUREKHA B, NAZARE K J, VISWANADHA RAJU S, et al. Attendance recording system using partial face recognition algorithm [M]// DEY N, SANTHI V. Intelligent Techniques in Signal Processing for Multimedia Security, SCI 660. Cham: Springer, 2017: 293-319.
- [9] VIOLA P, JONES M J. Robust real-time face detection [J]. International Journal of Computer Vision, 2004, 57(2): 137-154.
- [10] ZHANG K, ZHANG Z, LI Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks [J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.

- [11] 刘宇明,凌志祥,吴强,等. 基于多任务卷积网络的参会人员人数统计算法[J]. 计算机应用, 2018, 38(S2):51-54. (LIU Y M, LING Z X, WU Q, et al. People counting algorithm for participants based on multi-task convolutional network[J]. Journal of Computer Applications, 2018, 38(S2): 51-54. )
- [12] HU P, RAMANAN D. Finding tiny faces[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 1522-1530.
- [13] 龚锐,丁胜,章超华,等. 基于深度学习的轻量级和多姿态人脸识别方法[J]. 计算机应用, 2020, 40(3): 704-709. (GONG R, DING S, ZHANG C H, et al. Lightweight and multi-pose face recognition method based on deep learning [J]. Journal of Computer Applications, 2020, 40(3): 704-709. )
- [14] 卢涛,杨威,万永静. 基于图像超分辨率学习机的极低分辨率人脸识别[J]. 计算机应用, 2016, 36(2):580-585. (LU T, YANG W, WAN Y J. Face recognition with extremely low resolution based on image super-resolution extreme learning machine [J]. Journal of Computer Applications, 2016, 36(2): 580-585. )
- [15] ZANGENEH E, RAHMATI M, MOHSENZADEH Y. Low resolution face recognition using a two-branch deep convolutional neural network architecture [J]. Expert Systems with Applications, 2020, 139; No. 112854.
- [16] 方冠男. 基于视频流人脸识别的课堂考勤系统的设计与实现[D]. 武汉:华中师范大学, 2018:30-40. (FANG G N. Design and implementation of class attendance system based on face recognition with real-time video [D]. Wuhan: Central China Normal University, 2018:30-40. )
- [17] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2980-2988.
- [18] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]// Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2015: 91-99.
- [19] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 3431-3440.
- [20] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [21] XIANG J, ZHU G. Joint face detection and facial expression recognition with MTCNN [C]// Proceedings of the 4th International Conference on Information Science and Control Engineering. Piscataway: IEEE, 2017: 424-427.
- [22] WANG Y, YUAN G, ZHENG D, et al. Research on face detection method based on improved MTCNN network [C]// Proceedings of the 11th International Conference on Digital Image Processing, 11179. Bellingham, WA: SPIE, 2019: No. 1117 91C.
- [23] YANG T Y, CHEN Y T, LIN Y Y, et al. FSA-Net: learning fine-grained structure aggregation for head pose estimation from a single image [C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 1087-1096.
- [24] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 815-823.

This work is partially supported by the Provincial Teaching Research Project of Universities in Hubei Province (2018089).

**FANG Shuya**, born in 1995, M. S. candidate. Her research interests include deep learning, image processing.

**LIU Shouyin**, born in 1964, Ph. D., professor. His research interests include wireless communication, internet of things, machine learning.