

# 基于机器学习的人体动作识别实用性研究

严浩宸\*, 徐爽, 张驰, 孙天凯

徐州工程学院信息工程学院(大数据学院), 江苏 徐州

收稿日期: 2023年3月26日; 录用日期: 2023年4月16日; 发布日期: 2023年4月28日

## 摘要

当前, 线上线下混合式教学成为了教育的新趋势, 就体育课程的线上教学而言, 对学生的上课情况进行智能识别一直是各大高校研究的热点问题。体育课程不同于其他学科, 线上体育课的开展具有形式单一, 考核难度大的特点。后疫情时代下的体育课程仍然具有线上线下共同授课的挑战。本文就如何有效地应对线上体育动作识别这一问题, 研究并实现了一种基于机器学习的人体动作识别的方法, 利用计算机视觉对运动图像和视频进行采集, 采用基于Kinect的人体动作识别技术进行动作规范性的检测, 根据AdaBoost算法进行面部专注性检测, 实时分析学生在运动中是否专注, 动作是否到位等问题。结果证明: 相比于传统的线上体育课教学质量评价方法, 此方法既能拥有准确的识别效果, 也能够减少评判的时间, 对提高体育教师的工作效率有着很好的推动作用。

## 关键词

机器学习, 计算机视觉, 面部专注度检测, AdaBoost算法, Kinect

# Research on the Practicability of Human Motion Recognition Based on Machine Learning

Haochen Yan\*, Shuang Xu, Chi Zhang, Tiankai Sun

College of Information Engineering (Big Data College), Xuzhou University of Technology, Xuzhou Jiangsu

Received: Mar. 26<sup>th</sup>, 2023; accepted: Apr. 16<sup>th</sup>, 2023; published: Apr. 28<sup>th</sup>, 2023

## Abstract

At present, online and offline hybrid teaching has become a new trend in education. As far as

\*通讯作者。

文章引用: 严浩宸, 徐爽, 张驰, 孙天凯. 基于机器学习的人体动作识别实用性研究[J]. 图像与信号处理, 2023, 12(2): 158-168. DOI: 10.12677/jisp.2023.122016

online teaching of physical education courses is concerned, intelligent recognition of students' classes has been a hot issue explored by universities. Physical education courses are different from other disciplines. The development of online physical education courses has the characteristics of single form and high difficulty in assessment. The physical education curriculum in the post-epidemic era still has the challenge of online and offline co-teaching. This paper studies and implements a method of human motion recognition based on machine learning on how to effectively recognize online sports motion. It uses computer vision to collect motion images and videos, uses Kinect-based human motion recognition technology to conduct normative detection of motion, uses AdaBoost algorithm to detect facial focus, and analyzes whether students are focused in motion in real time, Whether the action is in place. The result shows that compared with the traditional online PE teaching quality evaluation method, this method can not only have accurate recognition effect, but also reduce the time of evaluation, which has a very good role in promoting the work efficiency of PE teachers.

## Keywords

Machine Learning, Computer Vision, Face Concentration Detection, Adaboost Algorithm, Kinect

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

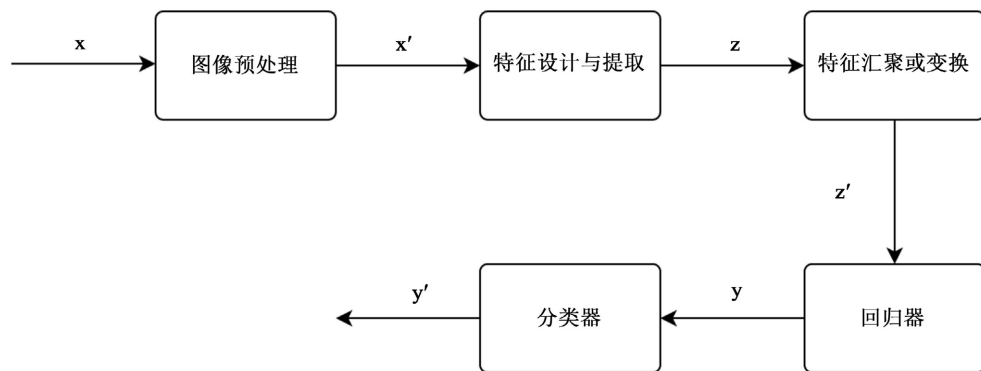
## 1. 引言

当下的线上体育课主要以教师先教授理论知识，再布置一些运动要求，然后学生被要求录制个人运动视频或通过类似“Keep”这样的APP，由系统自动对身体的动作识别分析来判断运动次数。类似的软件只对动作检测，并不能准确地识别该动作是否由学生本人完成，也不能准确识别完成的运动是否标准，这一缺陷将会使学生产生怠惰的心理，运动过程中对老师布置的任务敷衍了事。同时，教师对学生本人是否认真完成了相关的运动项目产生误判。另一方面，随着信息化的深入，大数据技术成果不断累计，它的价值等待着人们不断去发掘。科技部明确指出，人工智能与大数据技术可以有多种应用形式来构建服务于医疗健康，日常生活，乡村振兴的综合服务平台，实现对现有人工智能与大数据资源的整合和利用[1]。本研究通过面部识别、机器学习、慢特征分析等技术，获取学生运动状态信息和动作信息，将判别结果快速整合并发给教师，辅助教师判断运动训练的节奏并给出学生的运动成绩。

## 2. 计算机视觉及机器学习相关技术介绍

### 2.1. 计算机视觉相关技术原理

计算机视觉主要由数字图像处理，人脸检测等技术组成，它在一定程度上是以人的视觉对客观世界进行感知、识别、理解后将所理解的在计算机上表达出来。OpenCV 常被用作图像处理的函数库，在多个领域的问题处理中被经常利用，它的代码可读性好，安全性高，且文档中详细提供了多个算法的实现和定义方法[2]。只需在 Python 编译器内配置好 OpenCV 环境，便可以实现如下过程中所用到的算法的所有功能。以机器学习为基础框架的常用计算机视觉模型处理流程如图 1 所示。经过  $x$  变换， $x'$  变换，图像变为灰色图像且会被降噪，相应的视觉算法将会被应用，同时也会经历图像变换、边缘检测等过程；经过  $z$ ,  $z'$ ,  $y$ ,  $y'$  变换，图像会经历图像特征的拟合及在回归器中归一，在分类器中分类这几个变换的过程。



**Figure 1.** Processing flow of common computer vision models  
**图 1.** 常用计算机视觉模型处理流程

## 2.2. 机器学习技术处理流程

机器学习任务本质上可以用函数  $F$  来表示。对于任意输入的图像  $x$ ，有一个函数  $F$ ，它以  $\theta$  为参数，产生了由  $x$  至  $F_{\theta}(x)$  再到  $y$  的映射，使得对于实现上述任务的函数有这样的关系式，见公式(1)：

$$y = F_{\theta}(x) \quad (1)$$

对于实现上述任务的函数，它与计算机中的分治算法有异曲同工之处，主要是分治这两个核心要义。关于机器学习的常见处理流程为：自变量为输入的图像  $x$  经过一个以  $\theta$  为参数的函数  $F$  的映射。其中包括将图像进行分类和标签，将连续的变量，向量，图像进行拟合来完成如目标检测和语义分割这样的回归任务得到应变量  $y$ 。

## 2.3. 具体推荐算法执行过程

当学生上传视频后，系统将对其进行人脸识别提取其面部轮廓后提取出学生的相关信息，再与数据库中的学生图像信息进行比对从而判断是否为学生本人。若验证结果为不匹配，则将视频打回让学生重新提交，若匹配则进入分析图像阶段。

在分析图像阶段，主要根据已有的机器学习模型进行人脸检测和动作专注度与相似度比较。主要是针对眼部进行跟踪判断和根据教师已经导入上传的标准教学视频图像进行比较，分析出学生动作是否标准，是否专注。具体的推荐算法执行流程如图 2 所示。

## 3. 动作检测

### 3.1. 人体关节位置的确定

人体关节有多个区域，自上而下依次为头部、脖子、肩膀、手臂、肘、手腕、手、躯干、大腿、膝盖、脚踝、脚。确定人体关节位置数据的关键在于能否将人体图像这些区域准确归类。识别出人体各个区域之后，对于人体关节位置的确定问题，一般的方法是在人体关节的局部位置使用已知的标准深度信息得出整体结构分布区域概率质心。但是，边缘像素会严重降低这种全局估计的质量[3]。更优秀的方法是使用基于自下而上的简单聚类快速算法及基于均值漂移的精确算法和基于加权高斯核的均值漂移局部模式匹配算法。由于不同位置和类型的动作特征在视频中出现的位置通常不相同，所以需要创建一个有全局性的混合模型来有效地处理可能出现的不同序列特征来有效地描述体育运动过程中动作特征出现的各种信息及其轮廓[4]。这里以如下的估算方法来初步确定身体各个区域的密度，见公式(2)：

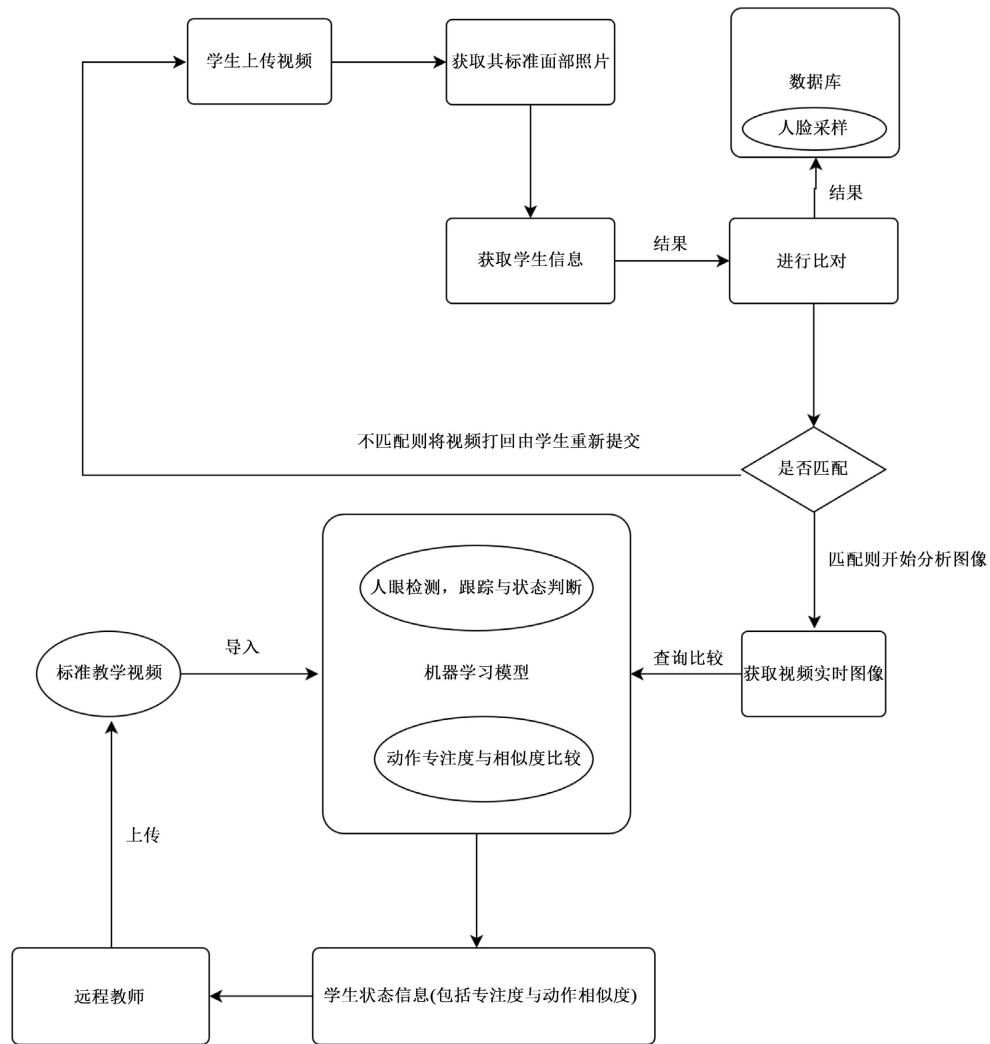


Figure 2. Execution process of specific recommended algorithms  
图 2. 具体推荐算法执行过程

$$\sum_{i=1}^N W_{ic} \exp\left(-\left\|\frac{x' - x'_i}{b_c}\right\|^2\right) \quad (2)$$

其中  $x'$  可以代表骨骼关键点对应三维空间坐标的转化,  $W_{ic}$  为像素权重,  $b_c$  为可学习偏置,  $x'_i$  为投影。其中直接影响识别结果的为像素权重, 所以可以说像素权重  $W_{ic}$  的大小在很大程度上影响了对人体骨骼部位定位的准确性, 计算方法见公式(3):

$$W_{ic} = P(c|I, x_i) * d_l(x_i)^2 \quad (3)$$

其中,  $d_l(x_i)$  描述了图片位于像素  $x_i$  的区域信息, 被称作深度信息。可以说深度信息以及经过灰度归一化的人体各区域密度这两个数据能够保证图像特征的稳定性, 当身体结构与 USB 摄像头的距离产生变化, 产生的骨骼结构也将与距离的长短有关, 但实际像素权重和深度信息是不变的[5]。如果图像出现了被遮挡或人在摄像头外的情况, 此时的深度信息值为正无穷。某个图片的像素点在水平和竖直方向会偏移一定距离并产生差值, 这个值就是深度信息  $d_l(x_i)$ , 当偏移量并非正无穷或处在一定范围内, 就可以将其

表述为人体区域[6]。因此可以这样总结：在空间直角坐标系中，深度信息这个值产生了动态平衡，具有不变性。

### 3.2. Kinect 2.0 骨骼图

Kinect 能够有效地追踪骨骼和具有在骨骼随身体移动后追踪骨骼结构变化的功能。它通过人体轮廓数据的采集完成骨骼数据的匹配并对于人体动作进行识别[7]。Kinect 2.0 骨骼图如图 3 所示：

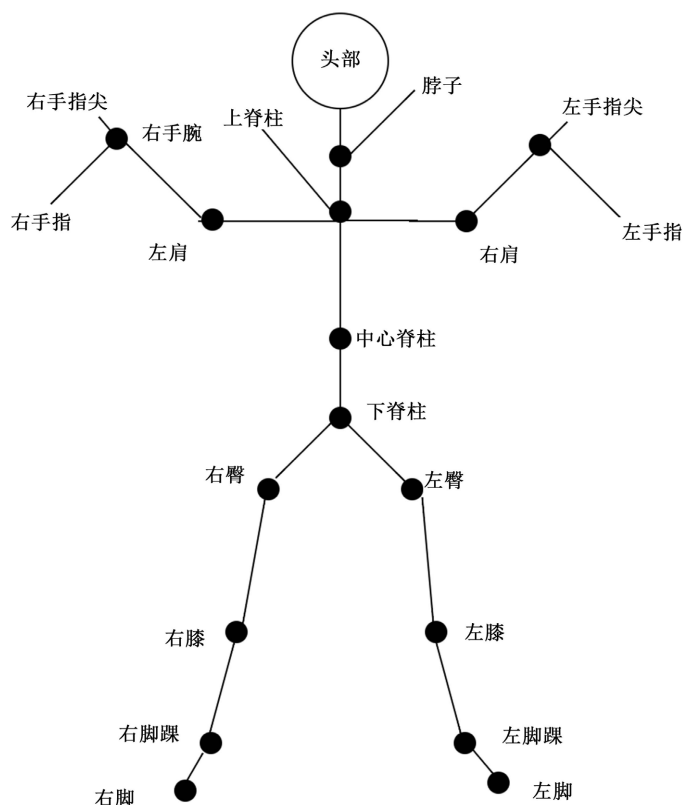


Figure 3. Kinect 2.0 Bone Map  
图 3. Kinect 2.0 骨骼图

基于 Kinect 的人体动作识别技术会以感应器为坐标原点，感应器一面为 x 轴方向，垂直感应器一面为 y 轴方向，以感应器的深度这一平面为 z 轴构建空间直角坐标系，追踪 USB 摄像头中的骨骼关节点。总的来说人体关节点共有 25 个，分别为：脊柱、中心脊柱、上脊柱、头部、脖子、左右两边从肩、到肘到手腕再到手，手指尖，拇指，同样在人下肢骨骼处从左至右，从上至下依次为臀、膝、脚踝、脚[8]。

## 4. 实时视频中的图像处理

### 4.1. 图像预处理

实时图像中的学生背景可能因光线不足，有遮挡等多种原因造成后续专注度检测和动作识别的错误，这里针对图像边缘检测进行主要介绍。

#### 4.1.1. 边缘检测

边缘是刻画图像中较为基础的一个属性，是图像局部区域与整体图像灰度值有显著反差的像素集合。

边缘检测便是通过发生急剧变化的区域边界来表示图像的灰度变换[2]。OpenCV 内 CV 自建库中蕴含了许多算法模型，调用 `cv2.Canny` 函数作为边缘检测的函数很合适用来检测灰度变化，从而进行边缘检测。运行后的效果见图 4。

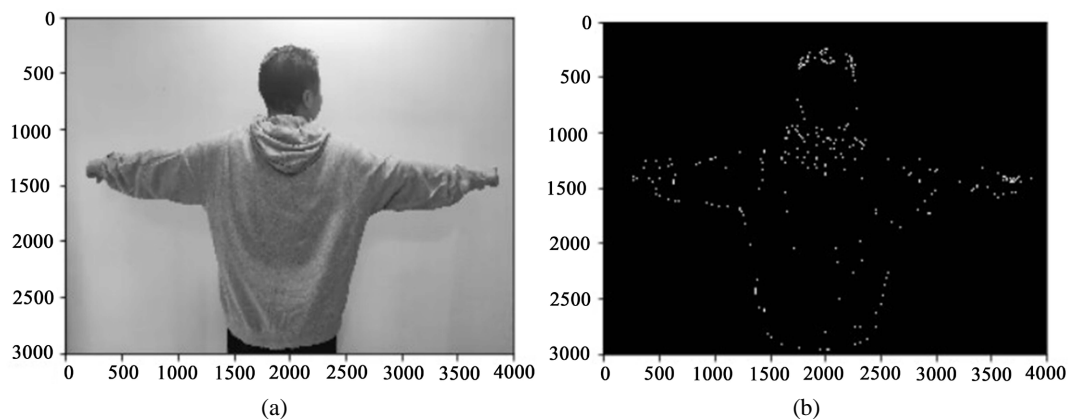


Figure 4. Example of edge detection

图 4. 边缘检测示例图

#### 4.1.2. 灰度值

灰度值一般指由图像构成的三维坐标内点的颜色深度，范围是[0, 255]。我们可以将黑白图像整体看作由三个二维向量 RGB 在三维坐标内组成，这三个二维向量便是 Red, Green, Blue。灰度值在图像识别领域是一个很好判断图像轮廓矫正度的属性值，对于 Canny 算子中的最优化边缘检测算子，它能将一个图像中三个二维向量的灰度值映射到 0~255 之间，极大的降低了处理器的计算量。经过回归化的图像灰度值，计算方式见公式(4)：

$$N(i, j) = \frac{I(i, j) - \text{Min}}{255 * (\text{Max} - \text{Min})} \quad (4)$$

其中  $I(i, j)$  通常用以表示初始的灰度值， $N(i, j)$  通常表示变换后的灰度值，Max 和 Min 用以表示原图像灰度的最大，最小值。

#### 4.2. 轮廓检测与矫正

实时图像中，如何确定人物占据图片哪一部分十分重要。整个图片的 dpi 大，作为输入不合适，十分考虑计算机的算力，所以定位人物所在区间，确定人物轮廓十分重要。详细的轮廓检测与矫正流程为：首先是去除图像噪声；其次是利用模糊半径以水平，垂直，斜对角三个因素来考量和计算图像的梯度；最后要判断在梯度的垂直接收器向上和平行接收器向右的方向上是否能找到某一像素的最大值。如果能满足这个条件，则说明此处的人物轮廓不能符合识别标准，需要进行矫正；最后要用设定一个阈值(本文中为 0.5)用于选择边缘，检查产生人体动作的轮廓边缘是否大于这个阈值，即达到了可以输出的标准，如果达到了这个标准，则将所有筛选出的不符合要求的边缘去除来确定最终输出的图像[9]。

### 5. 面部专注性检测

#### 5.1. 人脸检测

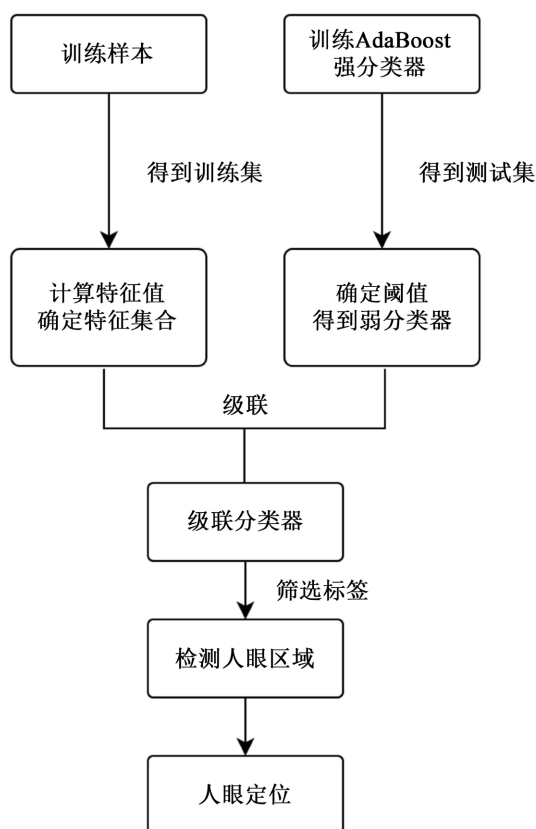
进行人脸检测主要需要普通 USB 摄像头、计算机和相应程序组成。图像的采集与预处理本文已提及。



首先,由 USB 摄像头对学生的个人图像信息进行实时采集,并用计算机对视频流进行处理。接入视频图像信息后,应用强分类器训练得到的测试集(弱分类器)和训练样本得到的训练集计算特征值,确定特征集合,进行人体骨骼脸部区域的初定位[10],之后在系列图像中检测到人脸这个元素,并从大量的背景中提取处理。原始的深度图像中存在许多物体,由于摄像头像素和网络等原因,图像边缘十分粗糙[11]。因此采用 Gabor 滤波器对图像进行去除噪点的操作十分有必要,这样会使三维像素图像在视频中各个物体的边缘上更加清楚可见,有效地解决了色彩失真的问题,并且能显著减少了图像模糊之处,这样更能够准确的识别人脸[12]。

## 5.2. 人眼检测跟踪与对齐

要做到人眼检测与跟踪,也就是要根据测试集对人眼特征分类器进行测试,然后对所有输入的待检目标进行扫描进而确定人眼的位置。由多层强分类器组合构成的 AdaBoost 算法可帮助训练出人眼检测器[13]。图 5 是 AdaBoost 算法训练人眼检测器的主要步骤。



**Figure 5.** Main steps of AdaBoost algorithm for training human eye detectors

**图 5.** AdaBoost 算法训练人眼检测器的主要步骤

正如机器学习将不同种类事物区分开的训练过程一样,AdaBoost 算法仍然会将若干张图片分为测试集和训练集两个主要部分。当图像被输入时,会有无数大小尺寸不一的检测框对整个图像进行扫描,再按照选定的比例进行放缩后再继续扫描,这是创建出正确的分类的过程。在特征集合和阈值的确定过程中,利用 K-Means 算法对每一个检测框都进行特征计算,得出所有数据点的值及其之间的距离,将其打

上标签与之对应，再经过级联进行统一决策：即进行阈值的确定，得出一个能够被各个层级强分类器都筛选出的检测框，此为人眼的弱分类器，则输出该检测框为人眼区域。

要做到人脸的对齐，需要利用到以 C++ 为编程语言的 Dlib 模型，它不仅可以有效输出人脸的关键点(眼睛、眉毛、鼻子、嘴、人脸轮廓)，还能通过这些关键点描述人的信息[14]。由 Dlib 得到的关键点的几何位置信息后，便可以通过 Procrustes 变换实现人脸在平面内以任意形式变换来完成人脸的调整，使其进行对齐，不被初始图片干扰，做到人脸形状的归一[15]。

### 5.3. 状态判断

与人眼检测和人脸检测相比，人外观特征的计算难度较大，尤其是人的专注度这一特征，往往需要结合多张特征，将其降维并组合成一起作为图片特征。这样便能产生局部特征而并非需要计算机付出大量算力将整个面部特征一起计算并筛选眼部特征。在局部特征的确定中，

算法能够通过特征多样的 Gabor 滤波器卷积得到面部特征图片的纹理[16]。对于体育课线上教学有效监管的质量评价，需要从人体在线上体育运动中的面部和动作方面进行对比反映，从而获得结果。这里利用到 FACS 作为人脸状态判断的编码系统，它结合了状态判断中的 Gabor 滤波器特征提取和人眼检测中 AdaBoost 算法进行人眼区域确定这两个方法，并最终产生了面部正面，中性，负面三个动作单元[17]。

## 6. 实验过程

### 6.1. 训练集的清洗及筛选

模型在搭建完成后，将通过训练集对其进行训练，训练集内的数据集部分源自网络，其余由笔者通过自己拍摄获取。若干数据集还需进一步对其进行筛选与清洗，从而获得其相应数据集，以用于模型的训练过程。图 6 是清洗及筛选后的训练集。

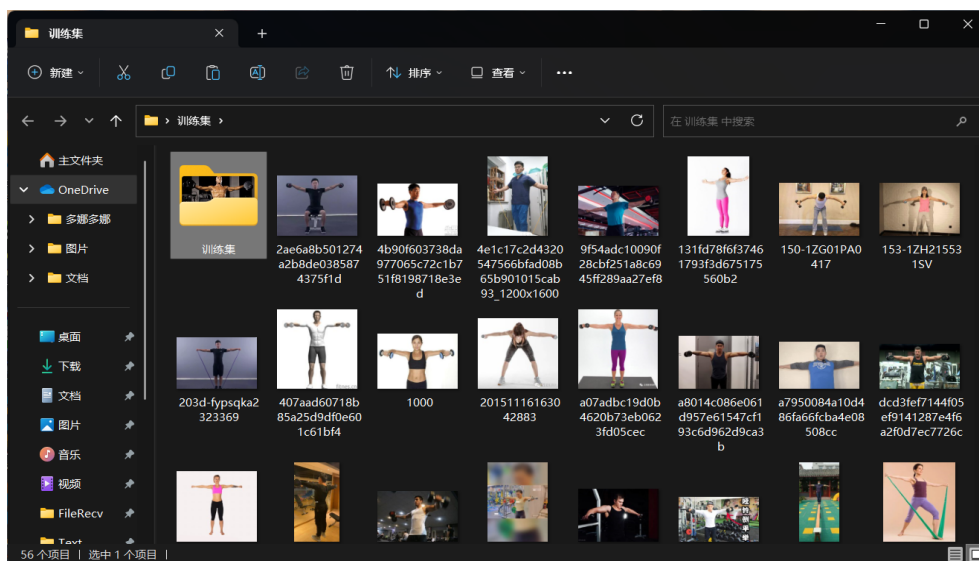


Figure 6. Training set after cleaning and screening

图 6. 清洗及筛选后的训练集

### 6.2. 模型训练

将清洗和筛选后的训练集导入模型，同时以百度 AI 程序为接口的 EasyDL 平台进行模型训练。其过



程如图 7 所示，每次训练后使用测试集对其进行模型系统的准确度测试，通过机器判断以及人工的调整干涉，由此不停测试迭代至更优的结果，最终获得更加准确的模型系统。

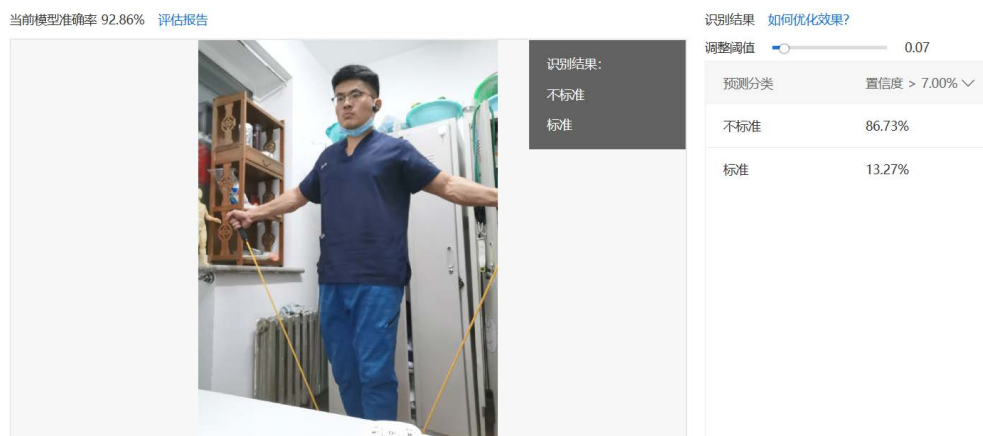


Figure 7. Model training process

图 7. 模型训练的过程

### 6.3. 模型导出与图像识别

在训练完模型之后，借助于百度以及 EasyDL 的接口，可为系统搭建一套简易的小程序，再通过小程序选出已经训练完成的所需要的模型，并通过上传动作图片对其进行识别与判断。

如图 8，当选择“动作识别 V2”这一训练好的模型并点击“上传图片并开始识别”之后，即可得到计算的结果。

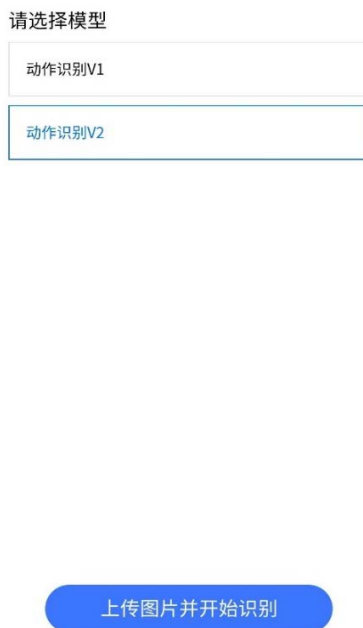
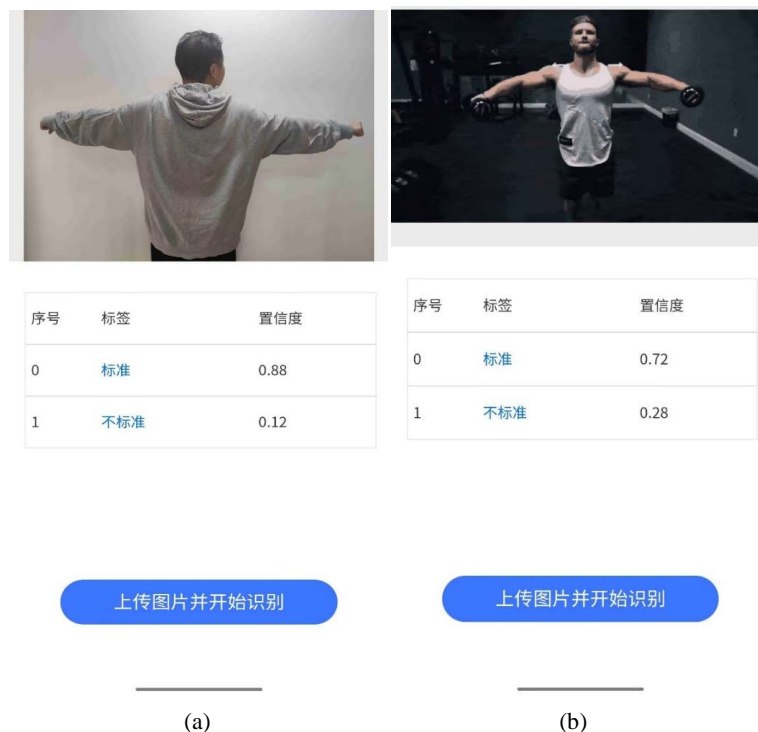


Figure 8. Model selection for action recognition in applets

图 8. 小程序中动作识别的模型选择

结果分为标准与不标准两个选项，以置信度 0.5 为标准基线，超过 0.5 的为当前动作结果。置信度数值的高低则表示图片中所展示动作接近其选项的程度，置信度越高，其动作越标准，反之则不标准，据此对动作进行相应的判断与评价。图 9 为人体结构形态相反情况时模型的计算结果。



**Figure 9.** Calculation results of the model when the human body structure and morphology are opposite

**图 9.** 人体结构形态相反情况时模型的计算结果

#### 6.4. 实验结果及分析

本实验的实验环境是在操作系统 Windows10，内存 16.0 GB，Intel(R)Core(TM) i7-10510U 的 CPU 下进行，其中历次实验结果的统计见表 1。

**Table 1.** Statistics of experimental results

**表 1.** 实验结果统计

序号	置信度	
	标准	不标准
1	0.88	0.12
2	0.72	0.28
3	0.94	0.06
4	0.76	0.24
5	0.82	0.18

可以看到，判定结果为标准的置信度均在阈值 0.5 以上，说明图片中所展示的动作已经十分接近标准这个程度，且图像训练集中图片数量多，产生的结果不具有偶然性。综上所述，该实验结果具有普遍

性, 模型稳定, 可以很好的判断动作的标准程度, 可利用性高。

## 7. 结论

综上所述, 本文建立了一个依据实时图像进行判断的基于机器学习的人体动作识别实用性研究的系统, 它通过学生在运动过程中动作的标准程度, 运动时的人脸表情, 人眼的状态来判断学生线上体育训练时的运动状态, 并将这些信息同步到教师终端, 告知教师学生的运动质量。有效地解决了当今部分软件通过对身体的动作捕捉分析来判断运动次数从而不能准确地识别学生的运动状态和运动标准情况这一问题。此方法主要通过机器学习、计算机视觉、面部检测、AdaBoost 算法、Kinect 等技术来获取学生运动状态信息和动作信息, 并能快速整合分析发给教师, 辅助教师判断运动训练节奏和学生运动效果。由于肢体动作的复杂程度及多种原因, 本系统仍需不断测试和更新完善, 争取创建一个更稳定的模型, 以适应不同的场景和更好地完成动作识别。

## 致 谢

本文为徐州工程学院大学生创新创业训练计划项目(xcx2022191)的阶段性成果之一。

## 参考文献

- [1] 林子雨. 大数据技术原理与应用[M]. 第3版. 北京: 人民邮电出版社, 2021: 326-327.
- [2] 黄界生. 基于深度学习的计算机视觉中图像检索算法研究[J]. 信息技术与信息化, 2022(9): 181-182.
- [3] 张亮. 基于 Kinect 的人体动作识别算法研究与系统设计[D]: [硕士学位论文]. 徐州: 中国矿业大学, 2019: 5-6.
- [4] 刘俊来. 基于融合不变性特征与混合核方法的体育视频动作识别[J]. 沈阳工业大学学报, 2022, 44(2): 198-202.
- [5] 郭欣, 孙侨, 张燕. 基于 IMU 和 Kinect 的人体下肢动作捕捉算法的研究[J]. 控制工程, 2023, 30(1): 169-176.
- [6] 杜宇宁, 刘其文. 深度学习在计算机视觉的应用[J]. 中国安防, 2022(5): 34-40.
- [7] 谈伟峰. 基于深度学习的视频动作识别方法研究与应用[D]: [硕士学位论文]. 南京: 南京邮电大学, 2022: 3-6.
- [8] 吴响军, 吴超, 江鑫富, 刘森林. 基于计算机视觉的电子血压计数字识别算法研究与设计[J]. 医疗装备, 2022, 35(9): 5-8.
- [9] 周原锐, 张逸群, 曹远航, 孙慧慧. 基于 OpenCV 视觉库的手写数字识别系统[J]. 吉林大学学报, 2021, 39(5): 602-608.
- [10] 汪磊, 孙瑞山. 基于面部特征识别的管制员疲劳监测方法研究[J]. 中国安全科学学报, 2012, 22(7): 66-71.
- [11] 查晶晶. 基于 Kinect 视频的图像增晰与检测算法研究[J]. 太原学院学报(自然科学版), 2022, 40(3): 65-70.
- [12] 徐凌风. 联合肤色分割的 Adaboost 算法人脸快速检测[J]. 网络安全技术与应用, 2022(8): 34-36.
- [13] 曹衡. 基于深度学习的人眼检测及应用研究[D]: [硕士学位论文]. 北京: 北方工业大学, 2020: 14-15.
- [14] 王秀, 周枫晓, 刘保罗, 霍文怡. 基于 Dlib 库的驾驶员疲劳驾驶检测系统[J]. 物联网技术, 2021, 11(12): 26-29.
- [15] 李婷. 基于 FACS 和多模态融合的网民情感分类研究[D]: [硕士学位论文]. 南京: 南京理工大学, 2020: 12-14.
- [16] 王慧敏, 孙建武. 基于视觉注意力的 Gabor 滤波器跟踪方法[J]. 装甲兵学报, 2022, 1(6): 92-100.
- [17] 随海亮, 马军山, 李丽莹. 基于生成对抗网络与 FACS 的面部表情合成研究[J]. 软件导刊, 2020, 19(6): 235-239.