

# 面向智慧课堂的三位一体化辅助学习平台

肖正,朱靖宇,宋超,刘孝炎,魏鹏程

(湖南大学信息科学与工程学院,湖南长沙410082)

**摘要:**教学中,由于知识点间相互联系,随着疑惑知识点增多,学生听课的效率降低,及时为学生答疑解惑至关重要。针对此问题,考虑学生对知识理解与课堂专注度、教师授课水平、知识点难易程度之间的关联性,利用深度学习技术智能分析课堂三要素,从多维角度生成学情分析一体化数据。同时由于互联网学习资源冗余、部分内容缺乏趣味性,基于学生习惯和兴趣,将学习资源风格化处理成学生乐于接受的形式,以满足不同学生的学习需求。实践应用结果表明,课堂上人脸识别、文字识别的准确率分别为99%与92%,能够满足项目需求。其中,学生四个评估指标中,学生满意率为83.4%,表明该平台能够及时提供针对性的答疑解惑,便于学生更轻松主动地学习。

**关键词:**智慧课堂;学情分析;深度学习;个性化推荐

**DOI:** 10.11907/rjdk.212194

**中图分类号:** G434

**文献标识码:** A

**开放科学(资源服务)标识码(OSID):**

**文章编号:** 1672-7800(2022)001-0001-08



## Learning Assistant Platform Integrating Three Key Elements for Smart Class

XIAO Zheng, ZHU Jing-yu, SONG Chao, LIU Xiao-yan, WEI Peng-cheng

(College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China)

**Abstract:** In educational environments, as knowledge points are interrelated, a bunch of negative effects emerges without timely revision. Confusions will be accumulated and learning efficiency gets lower, so it is essential for student to solve their questions in time. Therefore analyze three key elements by up-to-date AI and deep learning technology, fully considering the subtle impact on knowledge digestion subjected to students' concentration, teaching quality and knowledge comprehension hardness. Integrating the implication of the three key elements above, self learning materials are generated and recommended considering students' personality, interests and habits, so as to acquire the knowledge efficiently and effectively through stylization. The experimental results show that the accuracy of text recognition is up to 92% and face recognition reaches 99%, which meet project requirements. Through survey in our university after tentative application of our platform, the average satisfaction rate is up to 83.4%. Students thought that the recommended materials were in accordance with their confusions and easy to accept.

**Key Words:** Smart Class; learning status analysis; deep learning; personalized recommendation

## 0 引言

随着互联网的迅速发展和人工智能、大数据、云计算等技术迅速发展,传统课堂教学已无法满足当代师生的需求。于是探索信息技术与教育的深度融合,实施信息化教

学<sup>[1]</sup>。学生上课产生的疑惑不仅仅与学生的课堂专注度相关,教师的授课水平与知识点的难易程度也会影响学生。因此学生专注度、教师授课水平和知识点难易程度三者密切相关。由于知识点之间相互联系,若学生困惑知识无法被及时解答,则会导致其难以听懂授课内容。在教学过程中,会产生许多数据,获取和分析课堂教学数据,可精准掌

收稿日期:2021-09-01

基金项目:湖南省自然科学基金项目(2019JJ40024)

**作者简介:**肖正(1981-),男,博士,湖南大学信息科学与工程学院副教授、博士生导师,研究方向为智能信息处理、大数据、并行与分布式系统、高性能计算;朱靖宇(1997-),男,湖南大学信息科学与工程学院硕士研究生,研究方向为计算机视觉、多模态;宋超(1995-),男,湖南大学信息科学与工程学院硕士研究生,研究方向为计算机视觉、深度学习;刘孝炎(1997-),男,湖南大学信息科学与工程学院硕士研究生,研究方向为自然语言处理;魏鹏程(1996-),男,湖南大学信息科学与工程学院硕士研究生,研究方向为数据挖掘、高性能计算。本文通讯作者:肖正。

握教师授课情况,并且,教师可针对性地修改授课形式与教学计划。因此,面对丰富的课堂学情数据,如何有效利用和分析数据至关重要。

由于互联网学习资源繁杂冗余,学生往往需要花费大量时间寻找针对性的学习资料,降低了学习效率。部分教学资源缺乏趣味性,导致学生学习兴趣降低,阅读积极性不高。面对海量互联网学习资源,如何针对性为学生推荐风格化的学习资源,提高学生学习和积极性也是急需解决的问题。

## 1 相关工作

国内很多高校在课堂教学方面仍然延续着传统教学手段,课堂智能信息化程度不高。智能技术的迅速发展为智慧课堂提供了便利,将传统课堂改造为智能教学课堂,应用智能化技术赋能,支持教师精准化教学,满足学生个性化学习。智能教学课堂可利用人工智能技术分析课堂的行为状态;其次基于大数据分析学生听课效率和教师授课效果,以全面了解课堂情况,提高课堂学情数据的利用率;最后利用大量互联网学习课程,为学生提供丰富的学习资源。

孙曙光等<sup>[2]</sup>利用微服务器、云计算等技术手段,依据学生行为构建动态学习数据分析和云端应用智慧课堂信息环境,实现教与学的有效互动。刘邦奇<sup>[3]</sup>通过创设智慧型学习环境、基于动态学习数据的收集和分析,以数据化方式呈现学生学习的全过程及效果,使教学过程从依赖教师的教学经验转向客观数据,依靠数据精准掌握学情,提高教学过程的数据分析及资源推送能力。同时,作为科技行业巨头的腾讯、百度、科大讯飞等公司对智慧教育提出了新的解决方案。百度智慧课堂<sup>[4]</sup>通过教育大数据实现校园数据互通、资源多端同步、学情智能分析等功能,为师生提供智能化的课堂解决方案。虽然智慧课堂在教学应用中取得了一定的效果,但经过分析调研表明,还有许多改进空间,主要体现在以下2点:

(1)孙曙光等<sup>[2-3]</sup>缺乏对学生、教师授课水平、知识点难度的综合分析。学生对课堂知识的掌握情况与教师授课水平和知识点的难易程度具有关联性,课堂数据分析不够深入,可能会导致学习诊断不准确。

(2)百度智慧课堂<sup>[4]</sup>等所包含的互联网资源繁杂,缺少针对性的知识点和资源推荐,然而提供的学习资源与学生自身需求缺乏相关性,部分资源枯燥乏味,导致学生课后学习兴趣与积极性不高。

基于以上考虑,本文设计了基于智慧课堂的三位一体化辅助学习平台,通过人工智能技术与课堂结合,采集和分析整个教学过程的数据,全面展示课堂授课情况,同时推荐满足学生个性化需求的学习资源,及时为学生提供针对性的答疑解惑。本文的主要工作如下:

(1)考虑到学生对知识点的疑惑程度不仅与自身专注

度和教师授课水平有关,而且与课堂知识点的难易程度具有关联性。通过基于多任务卷积神经网络(Multi-task Convolutional Neural Network, MTCNN)<sup>[5]</sup>技术将检测的人脸进行静态表情识别<sup>[6]</sup>,然后与人眼张合度相结合判断学生上课的专注度。对于教师生动程度的判断,基于Openpose模型<sup>[7]</sup>检测教师姿态,通过教师授课身体摆动的幅度对生动度进行判断。最后,在文字识别中应用注意力机制,精准识别教师所述内容,准确无误地提取知识点信息。

(2)由于知识点之间是相互联系的,如果知识点不能被及时解答,疑惑知识堆积,学生听课难度上升,就会导致上课效率降低。因此本文基于准确的学情分析,为学生推荐针对性的学习资源,并及时解答学生的疑问。

(3)基于TextRank<sup>[8]</sup>算法将学习资源内容风格化,根据学生学习习惯、兴趣爱好及场景,生成学生喜闻乐见的风格化学习资源。例如:知识的摘要生成、图文并茂、符合学者文学风格的阅读内容等学习资源。

最后,通过湖南大学大学物理课程实践调研表明:模型人脸识别准确率达到99%;文字识别准确率达到92%;学生的4个评估指标中,非常满意平均占比为31.4%,满意平均占比为51.9%。学生普遍认为三位一体化的辅助学习平台效果显著,针对性的学习资源推荐能够为他们提供学习指导,减少疑惑知识点,提高听课效率。而且风格化学习资源能让学生轻松愉快的学习,提高学习兴趣与积极性。

## 2 三位一体辅助学习平台框架

本文中展示了两大系统智慧课堂系统、APP资源推荐端和六大功能。其中,包括学生专注度分析、教师生动度分析、黑板纸知识点提取、针对性资源推荐、知识点摘要生成、风格化资源生成以及学生、教师、家长、管理者四大用户。平台框架如图1所示,具体介绍如下:

(1)学生专注度分析通过捕捉学生表情及人眼张合度分析学生课堂专注情况。

(2)教师生动度分析主要根据教师鼻子、眼睛、手肘、肩膀等关键点位置摆动的幅度进行分析判断。此外结合多模态技术,添加姿态、表情和声音数据,以准确判断教师生动度。

(3)黑板知识点提取通过准确识别黑板上的文字并结合知识库提取知识点,并分析知识点的难易程度。

(4)针对性资料推送基于课堂三要素作出全面的学情诊断,然后针对学生疑问,基于知识图谱知识点匹配和推荐,为学生推荐符合需求的学习资源。

(5)知识点摘要式总结是风格化学习资源的一部分,主要将冗余、过长的学习内容,处理成简洁明了的摘要文档供学生阅读。

(6)风格化学习资源生成要考虑学生的学习习惯、兴趣爱好及所处场景,为学生生成乐于接受,符合学生需求的个性化学习资源。



Fig. 1 Framework of learning assistant platform integrating three key elements

图 1 三位一体辅助学习平台框架

本文所提出的基于智慧课堂的三位一体化辅助学习平台框架如图 2 所示。该平台分为数据分析与数据服务两部分。数据分析部分,基于深度学习平台与 InsightFace<sup>[9]</sup>模型的人脸识别和基于 MTCNN 的静态表情识别模型,结合改进后的文字识别技术对黑板板书内容进行识别及知识匹配,分析课堂数据。同时基于 Flink 流处理引擎加速深度学习平台的实时处理,通过 Web 端动态展示课堂教学情况。数据服务基于全面的课堂数据分析,结合多标签分类、知识图谱、推荐算法等技术为学生推荐针对性的风格化学习资源。

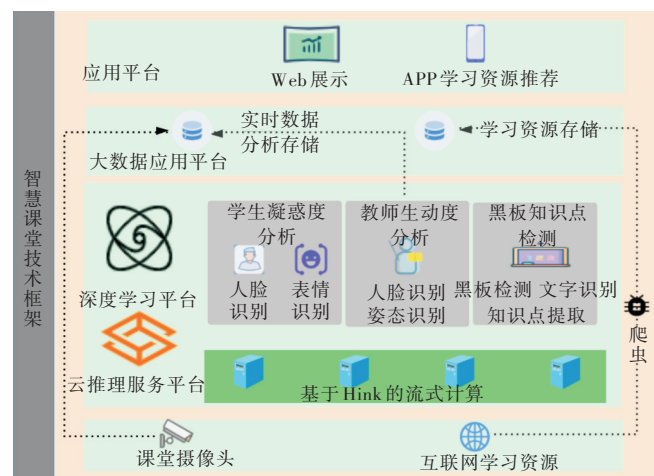


Fig. 2 Technology framework of Smart Class

图 2 智慧课堂技术框架

## 2.1 深度学习平台

为满足本文学生人脸识别、表情识别、教师姿态识别、黑板板书内容识别等需求,本文搭载了多个深度学习框架,包括 Tensorflow、Caffe、MXNet、Keras、Torch、Pytorch 等及用于人脸识别的 InsightFace 模型、基于 MTCNN 的静态表情识别模型、用于教师姿态识别的 Openpose 模型、基于 YOLOV3<sup>[10]</sup> 的黑板检测和基于 AOA<sup>[11]</sup> (Attention on Attention) 注意力机制的文字识别模型。提供了从模型开发、训练、部署的完整流程和工具,支持高性能 GPU 和分布式训练功能,同时配备良好的深度学习硬件平台,全面提高 AI 应用开发,满足开发者对各模型搭建的需求,为快速融入后续

模型接口提供便利。通过 AI 赋能课堂,分析学生上课的专注度、教师授课生动度和黑板知识点提取所组成的三要素课堂数据,产生一体化学情数据,为应用平台提供海量数据支撑。

## 2.2 云推理服务平台

为提高效率,需要对上课教学视频进行加速处理,能否实时、快速对学生进行表情识别、教师姿态识别和黑板板书内容识别则格外重要。但经过深度学习训练后模型网络结构较深,体积较大,处理数据所需时间较长。因此,如何快速、准确处理视频数据,缩短处理时间则成为目前亟待解决的问题。选择使用基于流处理的推理服务平台,在智慧课堂中主要负责 2 方面:①基于 Flink 流处理引擎,配合深度学习模型缓存和模型划分、放置策略,综合降低视频帧处理延迟;②平台负责维护、管理和更新智慧课堂所使用的深度学习模型,包括人脸识别模型、人体姿态识别模型、文字识别模型等,为用户提供一体化服务。本文使用 Flink 作为计算引擎,计算效率更高,方便系统部署、管理和维护,有效降低了推理延迟。

## 2.3 大数据应用平台

平台通过融合多标签分类与知识点图谱技术,实现学习资源分类,减少人力标注成本,提升推荐资料实用性和相关性。针对知识点在爬取相关网络学习资源时,由于知识点之间具有层级关系,学习资源分类问题难以解决,例如物理中“电学—电流—电荷”三个标签具有层级关系,在为学生推荐学习资源时,如何提高推荐资源实用性及相关性则至关重要。

文本多标签分类技术在智慧课堂系统中应用于各学科学习资源的分类。首先,构建多学科知识库,将各学科知识点按照树状结构组合,确定知识点间的关系。随后,将文本形式的学习资源归类为树状知识点结构,实现学习资源分类,减少人力标注成本。

对于知识图谱应用,旨在通过建立知识点数据间的关联性形式知识库,让数据更容易被人和机器理解和处理,便于课堂知识点匹配。同时,构建课程知识之间的关联性图谱帮助梳理整个学科的知识脉络,为学生推荐实用和相关性更强的学习资源。

### 3 智慧课堂关键技术

项目的关键是需要可行性较高的技术分析课堂三要素。随着人脸表情识别、姿态识别、文字识别等研究和深度学习模型愈发成熟,为项目提供了一定的理论基础和技术支持。对于检测学生专注度,首先基于人脸表情识别模型检测学生基本表情,然后辅助眼睛和鼻子的张合度进行矫正;对于教师生动度,基于 Openpose 模型以及多模态技术分析教师的生动度;黑板知识点基于知识图谱匹配知识点;学习资源风格化则基于 TextRank 算法生成简洁明了的短文摘要。

#### 3.1 学生专注度分析

分析学生上课的专注度,为学生推荐针对性的学习资源及与该知识点相关的学习资源,让学生多角度、深层次

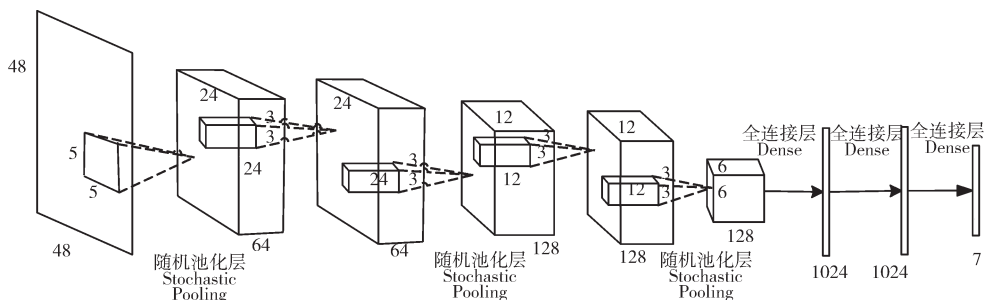


Fig. 3 Static facial expression recognition model

图3 静态表情识别模型

在课堂环境下,由于学生眼睛的张合度与其上课状态密切相关,如果仅通过基本表情评判学生专注度分析结果,准确度不高,可将人眼张合度纳入学生专注度状态评价指标之一。由于人眼呈现椭圆形,为了方便图形处理将眼睛面积视为长方形,张合度为人眼长方形中宽度和长度的比值,比值越大则表示越专注。但在实际训练中,课堂学生较为密集,单张人脸分辨率较低,人眼面积更小,故考虑使用双眼和鼻子3个关键点角度关系代替眼睛张合度评价指标。

经过数据处理发现可知,鼻子的位置相对固定。在摄像头位置不变的情况下,学生抬头、低头、侧身等动作都会影响实际截取人脸中双眼和鼻子的角度关系。例如若学生低头,则监控摄像截取人眼面积较小,鼻子位置相对较高,造成夹角C相对较大;若学生抬头,摄像头截取人眼面积仍然较小,鼻子位置相对不变,夹角C也会变大;若学生侧脸,则某一只眼睛和鼻子的距离就会变小,从而导致夹角C变大。如图4所示,采取三点检测法可大幅度降低计算量,并提升系统疑惑度分析的准确度。

通过分析已有数据,将角度C的阈值设置为66°最为适合,则夹角余弦的阈值为0.41,大于0.41则可判定为专注状态,否则为不专注状态。

最后本模块结合专注度评判准则则修正最基本的静态表情判断。若情绪判断为疑惑状态且眼鼻关键点角度正

进行学习,减少疑惑知识点堆积。教师可及时了解本堂课学生的上课效率及对于各知识点的理解情况,便于教师及时修改教学计划,提高教学质量。

学生表情识别使用了 Lopes 团队基于卷积神经网络的面部表情识别分类器<sup>[6]</sup>,该模型的基本结构如图3所示。其中包含了5个卷积层,6个随机池层和6个完全连接层,级联了6个 state-of-the-art 人脸检测算法,从而保证人脸检测的正确性。此处自适应地为每个网络分配不同权重,即学习集合权重设置为 $w$ 。独立训练多个不同初始化的 CNN 并输出训练响应。在加权的集合响应上定义了训练损失,通过 $w$ 优化以最小化该损失。在测试中,学习的 $w$ 也被用以计算整体测试响应,架构最后有P个 Dense,通过P个扰动样本输出结果,作为图像的预测值。在本文模型中,为了能在一定程度上实现数据增强,将图像尺寸归一化、直方图均衡化并去均值除方差。

常,则修正为正常状态;若眼鼻关键点异常,则仍然为疑惑。若情绪判断为正常且眼鼻关键点角度正常,则为专注,否则为正常态。最终的结果以正常、专注、疑惑三种状态呈现。



正脸状态 低头状态 侧脸状态

Fig. 4 Eye and nose angle evaluation

图4 人眼鼻子角度评价

#### 3.2 教师生动度分析

在课堂上,教师的行为状态包含了许多信息,例如从教师的授课语气、授课姿态变化中,可反应当前所讲知识点的重要程度,这些信息可用来辅助智慧课堂的建设。

如图5所示,OpenPose<sup>[7]</sup>人体姿态识别是美国卡耐基梅隆大学基于卷积神经网络和监督学习并以 caffe 框架开发的开源库。可实现人体动作、面部表情、手指运动等姿态估计,适用性广泛,具有极高的鲁棒性。针对教师姿态,采用 OpenPose 模型能满足项目的基本需求,对教师鼻子、眼睛、肩膀、耳朵等关键点进行检测识别,根据每一帧关键点的位置,采用欧氏距离计算相似度以计算教师授课的生

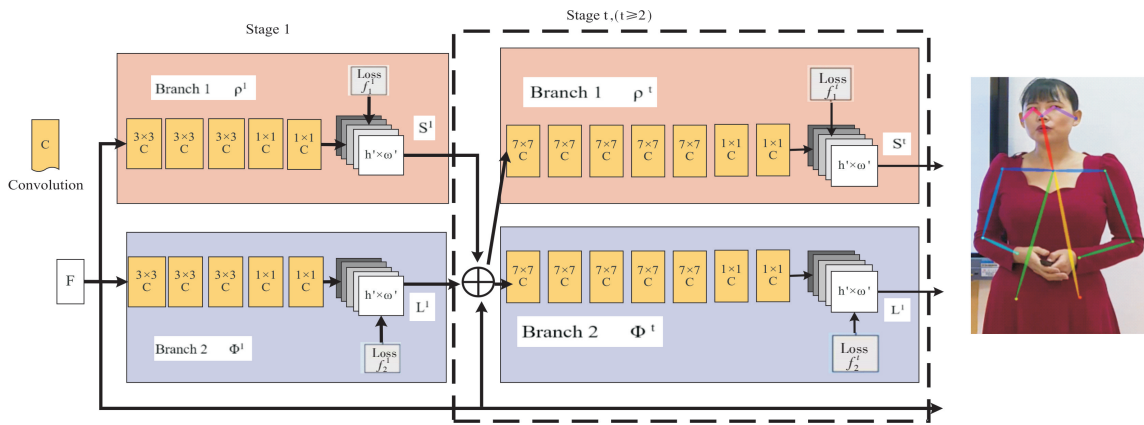


Fig. 5 Openpose model structure diagram

图 5 Openpose 模型结构

动度。

计算教师动作相似度方法参考 KNN<sup>[12]</sup>分类算法,使用类似于基于距离图像相似性度量方法进行分类或匹配。由于输入视频帧的大小不一致,视频帧中人的大小也不一致,因此需要归一化处理不同大小视频帧中大小不一人的相似度计算值。之后,针对教师姿态、表情及上课语音,利用多模态技术对三者数据进行模态融合,对最终结果进行判断从而辅助建设智慧课堂。

### 3.3 黑板知识点提取

知识点的难易程度因人而异,难度越大的知识点,学生的疑问也随之越多。因此,需要针对疑问知识点为学生推荐学习资源。目前考虑基于深度学习技术提取黑板的板书内容,接着对数据进行清洗、切词和去停用词操作,对比分析处理后的数据与已有知识点,生成词频矩阵。最后,计算二者的余弦相似度,将余弦值大于一定阈值且与原知识点库对比后不存在的知识点插入知识点库中。

CTPN<sup>[13]</sup> (Connectionist Text Proposal Network) 是一种将 CNN 和 LSTM (Long Short-Term Memory, LSTM) 结合的深度神经网络,能有效检测复杂场景中横向分布的文字。但课堂环境特殊,受到光照、中文字符较多、识别图像的分辨率等因素影响,导致 CTPN 在课堂环境下检测效果一般,识别率不高。同时,由于注意机制广泛应用于编码、解码器框架中,在每个时间步长生成编码向量加权平均值以指导解码过程。但解码器几乎无法查询向量与输入向量是否相关,因此可能会使解码器产生误导。为了解决上述问题,通过引入“Attention on Attention” (AoA) 模块扩展常规的注意力机制,以确保注意力结果和查询之间的相关性。

本模块提出了端到端的基于编码器—解码器架构的文本识别模型解决场景文本识别中存在的问题,模型由转换 (Trans)、特征提取 (Feat)、优化模块、序列建模 (Seq)、预测 (Pred) 五部分组成,模型的结构如图 6 所示。

(1) 将图片作为输入,首先使用空间转换器网络标准化输入文本图像,将原始弯曲或者倾斜的图像进行标准化,简化下游任务,使下游特征提取阶段不需要学习图像文本的几何形状不变性表示。

(2) 将标准化文本图像通过 CNN 进行特征提取,重点关注与字符识别相关的属性,抑制图像中不相关的特征,比如字体、大小、颜色和背景。

(3) 优化模块中,使用 AoA 优化模块提取特征。通过多头注意力机制对特征向量进行建模,构建图像中字符间的关系,然后应用 AoA 确定其关联程度。

(4) 将优化后的特征向量送入序列建模阶段,该模块基于 BiLSTM 通过捕捉下一阶段字符序列的上下文信息,从而更稳健地预测每个字符。

(5) 最后,使用 AoA 模块预测结果输出。通过注意力机制帮助特征对齐,自动捕获输入序列内的信息流以预测输出序列。其中,扩展注意力机制模块可过滤无关或误导的注意力结果,仅保留有用信息。

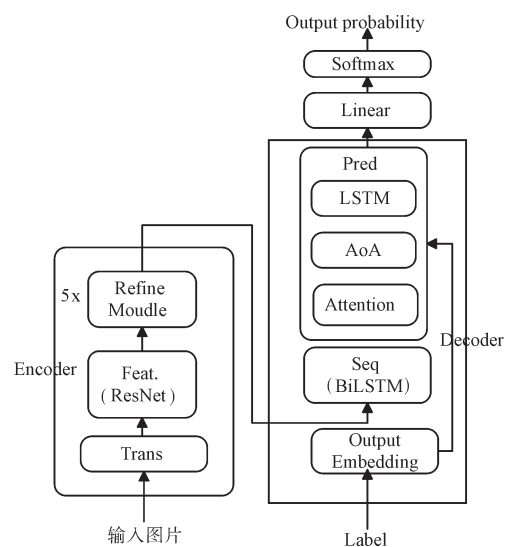


Fig. 6 Text recognition model

图 6 文字识别模型

### 3.4 学习资源风格化

为了满足不同学生的学习兴趣,体现学习的个性化,通过处理推送的互联网学习资源,将学习资源生成摘要式文本。TextRank 算法<sup>[8]</sup>是一种基于图的文本排序算法,基本思想来源于谷歌的 PageRank<sup>[14]</sup>算法,通过把文本分割成

若干组成单元(单词、句子)后建立图模型,利用投票机制对文本的重要成分进行排序,仅利用单篇文档本身的信息即可实现关键词提取、文摘。TextRank简洁有效,无需事先对多篇文档进行学习训练,通过使用TextRank算法进一步处理互联网获取的学习资源,可为学生生成简洁明了的学生资源,便于阅读。这是风格化学习资源的一部分,此外还考虑了结合学生的兴趣爱好、阅读习惯,将学习内容进一步处理成为学生喜欢阅读的方式。

### 4 辅助学习平台的实践

将平台分为4个部分进行介绍,依次为硬件平台配置、系统实现效果图、实验结果分析和实践效果。基于湖南大学信息科学与工程学院三个班级大学物理课程进行实践。此外,以问卷调查的形式获取学生反馈并作出简要的分析。

#### 4.1 硬件平台配置

硬件平台主要分为深度学习硬件平台和大数据硬件平台配置,深度学习的机器配置为:GeForce GTX1080显卡和48GB内存,以及两台配备NVIDIA TITAN V和128GB内存的服务器构成,配置如表1所示。

Table 1 Deep learning hardware configuration  
表1 深度学习硬件配置

主机名	CPU型号	内存	显卡型号	显存
DL1	Intel(R)Core(TM)i9-7900XCPU @ 3.3	48GB	GeForce GTX 1080Ti	24GB
DL2	Intel(R)Core(TM)i9-7900XCPU @ 3.3	128GB	TITAN V	48GB
DL3	Intel(R)Core(TM)i9-7900XCPU @ 3.3	128GB	TITAN V	48GB

通过集成Caffe、Tensor-Flow、PyTorch、Keras等主流深度学习框架作为技术支持,全力加速深度学习领域的人工

智能开发。大数据硬件平台主要存储教学视频的流式数据和互联网爬取的学习资源与课堂数据分析结果。通过5台曙光服务器进行试验测试,配置如表2所示。

Table 2 Big data storage platform configuration  
表2 大数据存储平台配置

主机名	CPU型号	内存	硬盘
Master1	Intel(R) Xeon(R) Gold 5118 CPU @2.30G	64GB	12TB
Master2	Intel(R) Xeon(R) Gold5118 CPU @ 2.30G	64GB	12TB
Master3	Intel(R) Xeon(R) Gold 5118CPU @ 2.30G	128GB	12TB
Master4	Intel(R) Xeon(R) Gold5118 CPU @ 2.30G	128GB	12TB
Master5	Intel(R) Xeon(R) Gold 5118CPU @ 2.30G	128GB	12TB

#### 4.2 系统实现效果

Web端和APP端均有相应的技术人员进行开发和维护。如图7所示,首先在Web端展示课堂(学生、教师和黑板知识点)三要学的学情分析结果。通过动态展示学生上课专注度、教师授课生动度情况、黑板板书内容,学生能及时了解到自己对知识点的疑惑程度,教师和管理者也能全面了解课堂的整体学情,并根据反馈信息及时修改教学方案。

接下来,介绍为学生推荐个性化学习资源的App端。其中,第一部分为针对疑惑知识点的学习资源推送(含有百度百科、博客、PPT、视频等);第二部分为趣味学习资源。通过对知识点资料的内容凝炼,实现了知识点摘要生成功能,但这仅属于实现风格化生成的开始。相信随着技术发展,风格化学习资源的生成功能会越来越完善,可满足不同学习者的需求。

#### 4.3 实验结果分析

##### 4.3.1 实验分析

Deng等<sup>[9]</sup>所提出的InsightFace模型与Yu等<sup>[6]</sup>所提出的静态表情识别模型对人脸和表情进行识别。通过实验



Fig. 7 Smart Class platform

图7 智慧课堂数据展示平台

证明:人脸识别准确率达到 95%,基本能够满足项目的需求;表情识别只有 61.29%,有待改进。下一步,将研究新的模型以更精确识别学生的表情特征。

#### 4.3.2 改进后的模型实验分析

##### (1) 黑板检测

###### ① 数据集

此方面工作较少,缺乏标注完整的数据集,因此通过网络采集、现场拍摄、同学收集等形式共制作 300 张图片。选取的标准为图片中清晰、完整的黑板,并且黑板的大小至少占据图片的四分之一,黑板的位置在图片的中部或者两侧。在收集完毕之后,利用 LabelImg 软件标注图片。最终,将 300 张图片随机分为 train.txt、val.txt、test.txt,数量分别为 240,30,30。在训练过程中,首先利用开源的 ImageNet 数据集进行预训练,然后利用标注好的数据集进行训练。

###### ② 黑板检测实验分析

由于要应对课堂这一特殊环境,需要对 YOLOv3 网络模型进行修改。原因有以下 3 点:黑板目标具有位置相对单一、类别相对固定的特点,检测难度不高;针对课堂黑板中单一的目标检测,可通过压缩网络,去除复杂的残差层,进一步提升训练效率和检测速度;课堂黑板一般位于视频中央或中央偏两侧位置,因此黑板会占据整体视频图像中较大的面积,可以暂时忽略检测到的小物体。

Table 3 Blackboard detection experimental results

表 3 黑板检测实验结果

评估指标	YOLO Layer0	YOLO Layer1
输入大小	15	30
损失	0.343 193	0.476 856
准确率	100%	100%
置信度	95%	96%

实验结果如表 3 所示:当 Grid\_size 为 15 时,损失为 0.34,判断物体种类准确率为 100%,置信度准确度为 0.947;当 Grid\_size 为 30 时,损失为 0.48,判断物体种类准确率为 100%,置信度准确度为 0.959,总体准确率达到 95%,能够精准检测黑板位置,基本能够满足项目需求。

##### (2) 文字识别

###### ① 数据集

通过使用人工合成的数据集,包括 SynthText (ST) 和 MJSynth (MJ) 两个主流的人工合成文本识别数据集,分别从中取 5 500 000 张和 8 900 000 张图片用于实验。对于测试集而言,使用了 3 个真实场景数据。其中,SVT 数据集包含了 647 张图片,图片来自于 Google 街景采集的户外街道,包含了一些低分辨率、模糊的图像;IC03 数据集包含了 867 张图片,来自于 ICDAR 2003 健壮阅读比赛中用阅读相机捕捉的场景文本;IC13 数据集包含了 1015 张图片,在继承 IC03 的大多数图像的基础上扩充了来自于 ICDAR 2013 健壮阅读比赛捕捉的图片。

###### ② 实验分析

本文所提出的文字识别模型与现有模型进行对比,使

用相同的训练集训练,在不同的数据集上测试,结果如表 4 所示。其中,本文提出的模型在不同数据集上均具有良好的性能,在多个数据集上总体准确率为 93%,针对性上的提升,分析原因如下:通过引入 AoA 模块扩展了常规的注意力机制,以确定注意力结果和查询之间的相关性。构建了基于编码器-解码器架构的模型以解决文本识别中存在的问题;在编码器阶段,通过 CNN 提取图片特征之后,未直接将其送入解码器中,而是构建了一个包含 AoA 的模块优化网络,有助于更好地建模图像中不同字符之间的关系,优化特征表示;在解码器阶段,应用 AoA 过滤无关的注意力结果,只保留有用信息。

从表 4 可见,本文提出的模型相比现有模型,在文字识别准确率方面有着一定的提升,在课堂中取得了很好的效果。

Table 4 Text recognition experimental results

表 4 文字识别实验结果

Model	TrainData	SVT647	IC03860	IC13857	Total Accuracy
CRNN	MJ+ST	0.816	0.931	0.911	0.886
RARE	MH+ST	0.858	0.939	0.926	0.907
R2AM	MJ+ST	0.824	0.922	0.902	0.883
STARNet	MJ+ST	0.869	0.944	0.928	0.914
GRCNN	MJ+ST	0.837	0.935	0.909	0.894
Rosetta	MJ+ST	0.847	0.934	0.909	0.897
Four-sta STR	MJ+ST	0.875	0.949	0.936	0.92
Our Model	MJ+ST	0.896	0.947	0.95	0.93

#### 4.4 辅助学习平台实践效果

在 2020 年度下学期的教学实践中,对湖南大学信息科学与工程学院计算机专业学生的物理课程(3 个班级)采用了基于智慧课堂的三位一体辅助学习平台。通过问卷调查(78 份有效问卷)包括 3 个班级的 78 名学生获取反馈信息,问卷调查数据统计如表 5 所示。

Table 5 Results of student questionnaire survey

表 5 学生问卷调查结果

评价指标	非常满意	满意	不满意
推荐资源相关性	52 人 66.7%	18 人 23.1%	8 人 10.2%
学生疑问是否及时得到解答	19 人 24.4%	50 人 64.1%	9 人 11.5%
风格化学习资源是否有效	12 人 15.5%	46 人 59%	20 人 25.6%
学习兴趣与积极性是否提高	15 人 19.2%	48 人 61.6%	15 人 19.2%

在这四个评估指标中满意度取平均值,学生非常满意率为 31.5%、满意率为 51.9%、不满意率为 16.6%。从统计数据可见,绝大多数学生认为个性化辅助学习平台能够为学生及时答疑解惑,提供针对性的学习资源,并且学习内容丰富,趣味性强,满足学生并的学习需求。不仅提高了学生学习的效率和积极性,而且明显改善了当前教学效果。当然,也有少部分学生存在不满意的情况,证明平台

仍有待完善和改进。一方面,需加强学习资源风格化技术的研究,满足不同学生的个性化学习需求以及提高对于学生专注度、教师生动度等关键部分评判的准确性;另一方面,加强学生在课堂中的实践性与互动性,提高学生解决问题的能力 and 创新能力。随着科学技术的迅速发展,后期将持续对平台进行升级,为师生提供更优质的课堂服务。

## 5 结论

本文提出的面向智慧课堂的三位一体化辅助学习平台,通过实时智能分析课堂三要素,从多维角度产生一体化学情数据,及时为学生答疑解惑,并考虑知识点之间的关联性,让学生查缺补漏。提供针对性的风格化学习资源,充分调动了学生学习兴趣与积极性,提高学生课后学习的主动性和效率。主要工作和创新点如下:

(1)本文提出了一种新理念,综合考虑学生对于知识点的理解与学生自身专注度、教师授课水平以及知识点的难易程度三者息息相关,从多维角度产生一体化学情数据,使课堂分析更加全面,为学生推荐更加准确的学习资源。

(2)利用人工智能技术赋能课堂,结合云推理服务、大数据技术,初步实现了面向智慧课堂的三位一体化辅助学习平台。

(3)在实际应用过程中,研究了黑板检测和文字识别,通过实验证明测试的准确率分别为95%与93%,能基本满足项目需求。

同时该平台也存在一些不足,要继续深入研究关键技术,提高模型精度。今后将持续丰富和完善平台的各项功能,真正为学生和教师提供便利,提高学习效率和教学质量。

### 参考文献:

- [1] LIU J. Smart class: new approach for the develop of future school class in the era of "Internet +" [J]. China Educational Technology, 2017 (7): 14-19.  
刘军.智慧课堂:"互联网+"时代未来学校课堂发展新路向[J].中国电化教育,2017(7):14-19.
- [2] SUN S H, LIU B Q, LI X Y. The construction and application of smart classrooms in the era of big data [J]. China Information Technology Ed-

ucation, 2015(Z1): 112-114.

孙曙辉,刘邦奇,李新义.大数据时代智慧课堂的构建与应用[J].中国信息技术教育,2015(Z1):112-114.

- [3] LIU B Q. Research on design and implementation strategy of Smart Class teaching in the era of "Internet +" [J]. China Educational Technology, 2016(10): 51-56, 73.  
刘邦奇."互联网+"时代智慧课堂教学设计与实施策略研究[J].中国电化教育,2016(10):51-56,73.
- [4] Shanghai Iresearch Consulting. Iresearch consulting group series research report (no.3, 2020) [R]. 2019 China AI+ Education Industry Development Research Report, 2020: 45.  
上海艾瑞市场咨询有限公司.艾瑞咨询系列研究报告(2020年第3期)[R].2019年中国AI+教育行业发展研究报告,2020:45.
- [5] YIN X, LIU X. Multi-task convolutional neural network for pose-invariant face recognition [J]. IEEE Transactions on Image Processing, 2017, 27(2): 964-975.
- [6] YU Z, ZHANG C. Image based static facial expression recognition with multiple deep network learning [C]//Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, 2015: 435-442.
- [7] CAO Z, SIMON T, WEI S E, et al. Realtime multi-person 2d pose estimation using part affinity fields [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7291-7299.
- [8] MIHALCEA R, TARAU P. Text rank: bringing order into text [C]//Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 2004: 404-411.
- [9] DENG J, GUO J, XUE N, et al. Arcface: additive angular margin loss for deep face recognition [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 4690-4699.
- [10] REDMON J, FARHADI A. YOLOv3: an incremental improvement [DB/OL]. <https://arxiv.org/abs/1804.02767>.
- [11] OCASIO W. Attention on attention [J]. Organization Science, 2011, 22(5): 1286-1296.
- [12] DOU X F. Summary of KNN algorithm [J]. Telecom World, 2018 (10): 273-274.  
窦小凡.KNN算法综述[J].通讯世界,2018(10):273-274.
- [13] TIAN Z, HUANG W, HE T, et al. Detecting text in natural image with connectionist text proposal network [C]. European Conference on Computer Vision, 2016: 56-72.
- [14] BIANCHINI M, GORI M, SCARSELLI F. Inside PageRank [J]. ACM Transactions on Internet Technology (TOIT), 2005, 5(1): 92-128.  
(责任编辑:刘嘉文)