**智慧场景学习平台3.0需求分析**

目录

[第一章 平台目的 3](#_Toc21147)

[第二章 业务逻辑 3](#_Toc3084)

[2.1 业务数据表 4](#_Toc19658)

[2.2 课程业务 4](#_Toc14226)

[2.3 学生业务 5](#_Toc21410)

[1) 课前 5](#_Toc32743)

[2) 课中 5](#_Toc30894)

[3) 课后 5](#_Toc22725)

[第三章 系统架构 5](#_Toc20398)

[3.1 系统框架 5](#_Toc532)

[3.1.1 硬件框架 6](#_Toc20335)

[3.1.2 电气设备层 6](#_Toc31725)

[3.1.3 感知系统层 6](#_Toc12853)

[3.1.4 控制系统层 6](#_Toc12234)

[3.1.5 智能应用层 7](#_Toc6837)

[3.2 数据需求 7](#_Toc8226)

[3.3 功能需求 8](#_Toc10682)

[3.3.1 课程接入 8](#_Toc14855)

[3.3.2 任务指导书设计与展示 8](#_Toc20950)

[3.3.3 量纲、识别点、识别链解析 8](#_Toc119)

[3.3.4 课程设备接入 9](#_Toc25051)

[3.3.5 基于AI的人员辨识、设备状态辨识、结果辨识 9](#_Toc18309)

[第四章 硬件框架 10](#_Toc17398)

[4.1 多源多模态感知框架 10](#_Toc9608)

[1) 基于 AI 的视频分析系统 10](#_Toc7825)

[2) 学习者行为分析 10](#_Toc13034)

[3) 学习过程和结果评价 10](#_Toc11635)

[4.2 全方位音视频信息采集框架需求 11](#_Toc13650)

[4.2.1 钢架结构设计 11](#_Toc7348)

[4.2.2 滑轨系统 11](#_Toc31038)

[4.2.3 云台系统 12](#_Toc15488)

[4.2.4 摄像头与麦克风集成 12](#_Toc30024)

[第五章 感知系统 13](#_Toc23485)

[5.1 感知信息分类 13](#_Toc27811)

[1) 环境感知 13](#_Toc18703)

[2) 音视频信息 13](#_Toc12963)

[4) 安全预警 13](#_Toc13520)

[5.2 设备类型分类 13](#_Toc32576)

[1) 摄像头 13](#_Toc17397)

[2) 麦克风 14](#_Toc1273)

[3) 智能手环 14](#_Toc25338)

[4) 传感器 14](#_Toc14975)

[5) 数据采集器 14](#_Toc8520)

[第六章 后台调度系统 15](#_Toc7953)

[6.1 整体系统运行流程 15](#_Toc24759)

[6.2 登录流程 15](#_Toc13944)

[6.3 实践流程 16](#_Toc16908)

[6.4 结果辨识流程 17](#_Toc8694)

[第七章 AI 部分 18](#_Toc16055)

[7.1 人员感知需求 18](#_Toc9645)

[7.1.1 学生登陆人脸识别需求 18](#_Toc8463)

[图：HOG特征可视化 19](#_Toc9802)

[7.1.2 学生信息标准需求 20](#_Toc22497)

[7.1.3 界面设计 21](#_Toc18679)

[7.1.4 访问控制 21](#_Toc18194)

[7.1.2 学生学习状态检测需求 21](#_Toc28637)

[学生情绪识别方案 24](#_Toc1689)

[7.1.3、学生语音识别需求 25](#_Toc24512)

[7.2 设备状态感知需求 27](#_Toc18985)

[7.2.1 目标识别算法 27](#_Toc32638)

[7.2.2 视觉分类算法 29](#_Toc16190)

[第八章 展示需求 30](#_Toc26739)

[1、场景展示大屏 31](#_Toc23653)

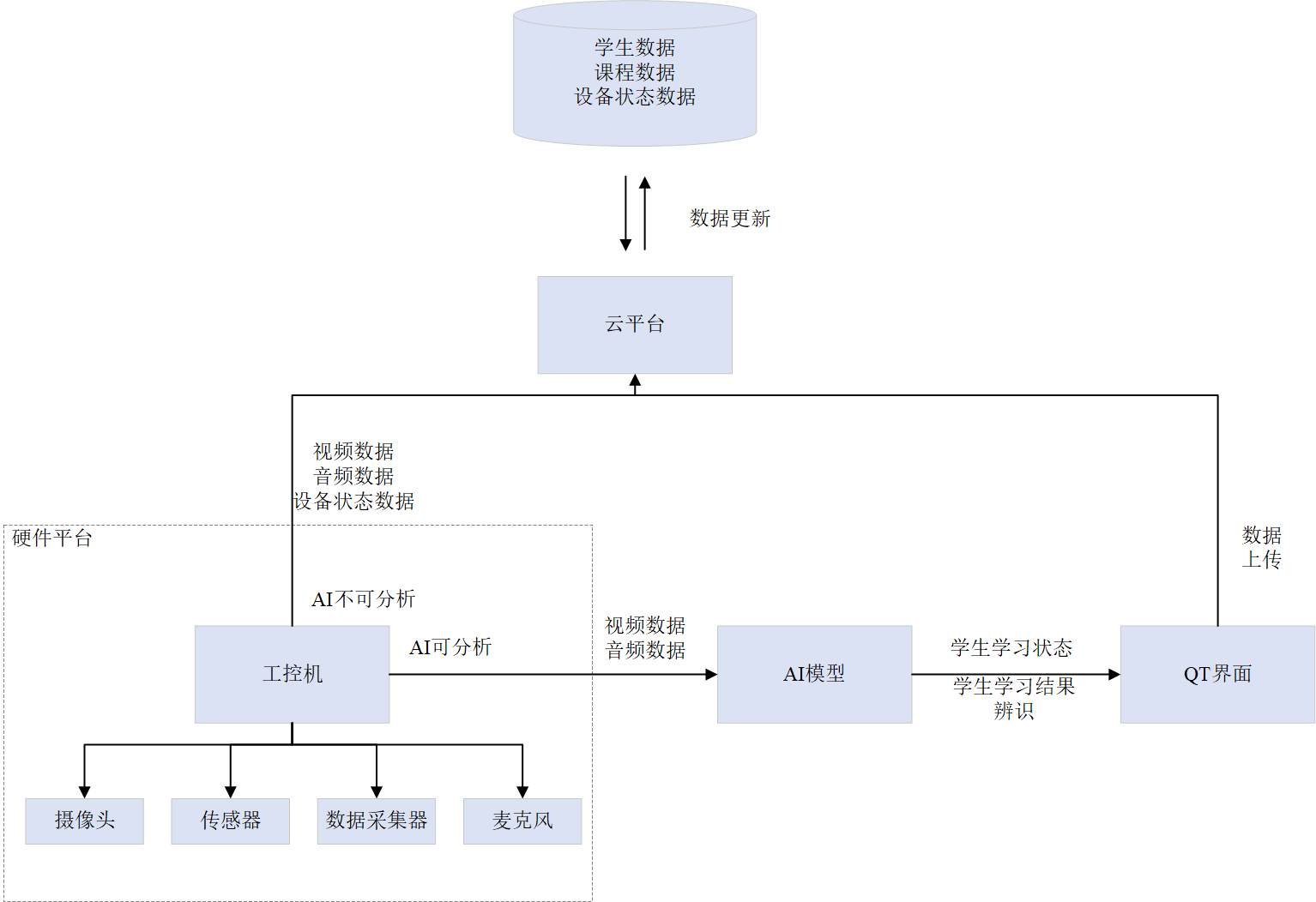
[2、用户界面 31](#_Toc30481)

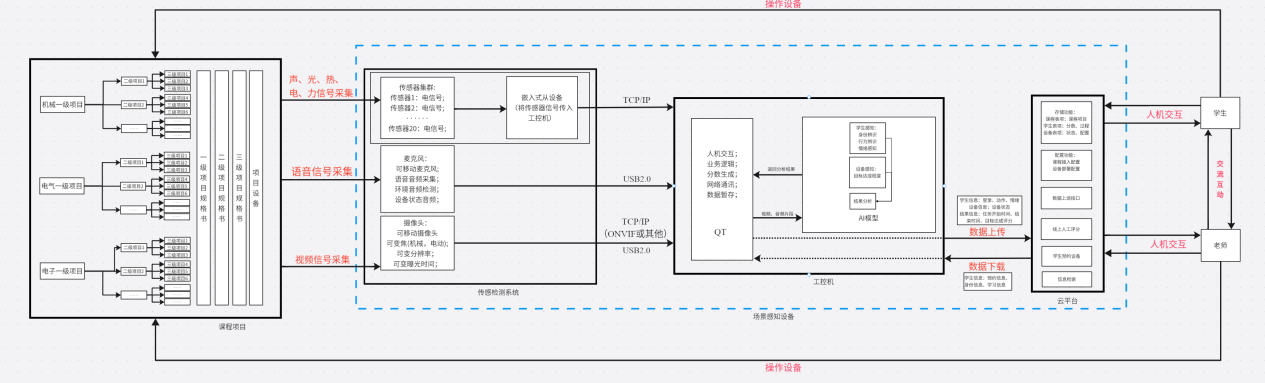
**第一章 平台目的**

智慧场景学习平台专为工科实践课程打造，旨在通过精准的任务目标导向，引领学生操作电气设备达成既定学习目标。平台融合先进感知技术，不仅实时监控设备状态，还采集现场音视频信息，构建全方位学习反馈体系。其核心目的在于提升学习成效，通过智能评价激励学生主动学习，增强学习动力与积极性。同时，有效减轻教师负担，实现高效、主动的育人模式，培养具备实践能力和创新精神的工科人才。

**第二章 业务逻辑**

本章旨在分析智慧场景学习平台服务工科实践类课程所需处理的业务需求。该平台围绕专业表、课程表、达成目标表、学生信息表、学生目标达成表及学生学习过程表等核心数据库表进行设计。通过模块化、组件化的方式，确保系统能够灵活应对不同工科实践类课程教育场景下的业务需求变化，满足通用性、可扩展性需求，同时保持系统的稳定性和高效性。





**2.1 业务数据表**

围绕着课程、达成目标、学生等实体需要首先分析他们属性及内在联系，在达成云平台构建业务数据表。

1. 课程相关数据表

课程按照1级项目专业、2级项目课程群、3级项目课程的方式设计可兼容全部工科实践类课程的数据表结构。

1. 学生相关数据表

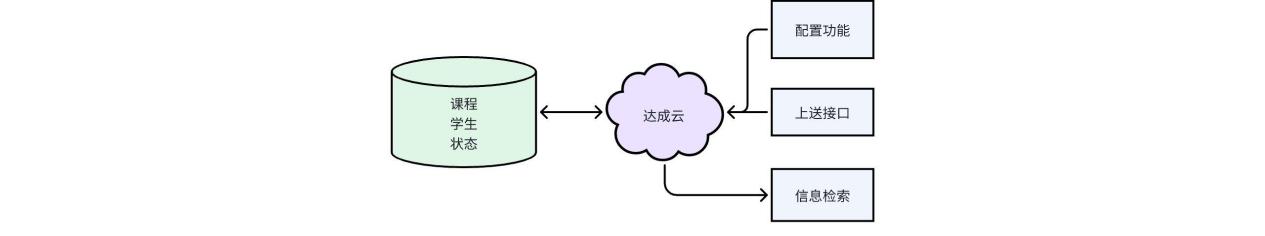
分析学生和课程之间相关关系，包括专业、课程、得分、过程成绩等设计便于管理学生上课及考评的数据表结构。

1. 设备相关数据表

分析设备与课程、场景等之间的相关关系，充分考虑设备的可扩展性，使设备便于接入智慧场景学习平台，设计相关数据表结构。

**2.2 课程业务**

数据库表确立后，针对智慧场景学习平台服务的课程，将首先确认与课程相关的表项目的具体数据。在此需求基础上，系统平台将提供勾选配置完成课程数据配置方案，为开课教师提供便捷操作。



**2.3 学生业务**

围绕着数据库表中与学生相关的数据项，智慧学习平台将通过软硬件设备对学生学习状态，学习结果进行辨识，并通过边缘计算得到处理信息，上传至达成云平台数据项中。

1. 课前

达成云平台为学生提供预约功能，也包括学生个人信息的上传包括个人照片。同时达成云平台也为教师提供审核学生预约信息的功能。相对应的表中新建关联学生的项目。

1. 课中

* 登录

AI技术的应用使得学生能够通过人脸识别登录，同时允许教师控制登录环节，包括指定具体登录人员。登录数据包括登录人员、登录时间将上传至达成云相关数据库中。

* 评分

由硬件框架、电气设备、感知系统、控制系统及AI等组成的系统将对于学生学习过程的全部达成目标进行跟踪辨识，形成对学生达成目标的评分，并上传至达成云平台课程评分数据库中。

同时系统对学生学习过程的音视频具有全程记录的功能，并通过AI计数的应用使得学生动作、状态、情绪等可以被识别，同时将这些过程信息同步至达成云平台数据库中。

1. 课后

达成云平台支持学生在课后上传自己的作业，教师也可以通过达成云平台对学生的全部课程信息包括评分信息进行审核及人工评分。

**第三章 系统架构**

本章旨在详细介绍本系统的整体架构，首先阐明课程、学生、考核等数据库表间的业务逻辑，接着从基础的硬件支撑到顶层的智能应用，全面阐述系统组成部分，核心数据的构成，各组成部分的功能、相互关系以及它们在系统中的作用。该系统采用多层次、模块化的设计思路，确保系统的稳定性、可扩展性和高效性。

**3.1 系统框架**

系统框架由外至内依次分为四个主要层次，形成了一个高度集成且协同工作的整体。外层为坚实的硬件框架，向内依次是电气设备层、传感器层、控制系统层，以及最内层的智能应用层（包含AI、前端界面、云平台）。



**3.1.1 硬件框架**

硬件框架作为整个系统的物理基础，提供了必要的计算资源、数据存储、网络通信等基础设施。它包括高性能服务器、存储设备、网络通信设备等，确保系统能够稳定运行并处理大量数据。硬件框架的设计充分考虑了可扩展性和冗余性，以适应未来可能的业务需求增长和技术升级。

**3.1.2 电气设备层**

电气设备层指的是学生在实践课程直接操作的对象，也是与系统的物理操作环境直接相关，是工科实践类课程的重要组成部分。每个课程所使用的电气设备具有专用性，需要具体设计。因此本需求分析中不对电气设备层设备做具体的需求分析。

**3.1.3 感知系统层**

感知系统层作为系统感知外部环境的关键环节，负责采集各种物理量信息，如温度、压力、位置、速度等，并将其转换为可处理的数字信号。这些传感器分布在系统的各个角落，形成了一张密集的信息采集网络，为控制系统提供实时、准确的数据支持。

**3.1.4 控制系统层**

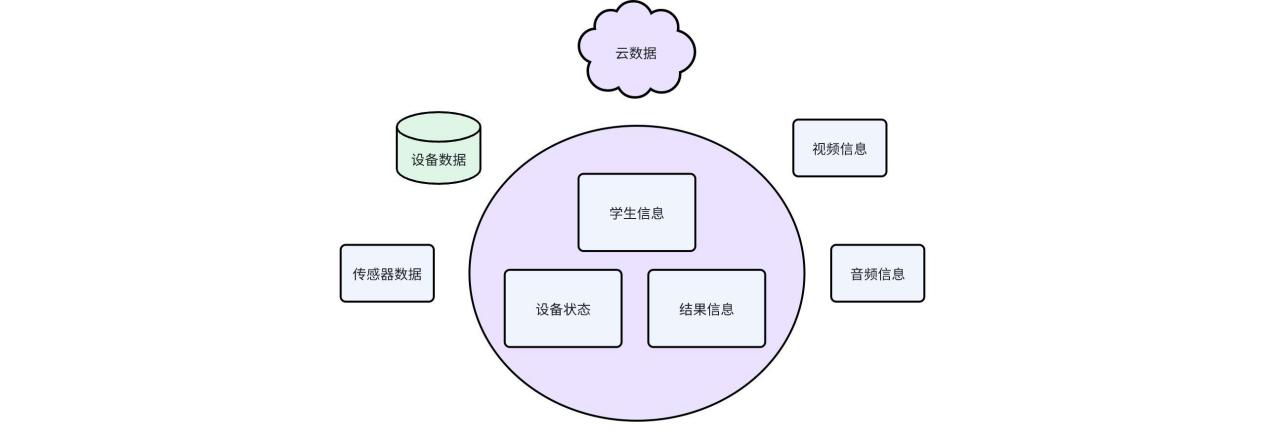
控制系统层是整个系统的中枢，负责接收传感器层的数据，通过复杂的算法进行数据处理、分析和决策，最终向电气设备层发出控制指令。该层采用先进的控制技术和算法，如PID控制、模糊控制、神经网络控制等，以实现系统的自动化、智能化运行。同时，控制系统还具备自我监测和故障诊断功能，确保系统的稳定运行。

**3.1.5 智能应用层**

智能应用层位于系统框架的最顶端，集成了AI、前端界面、云平台等先进技术。AI模块利用机器学习、深度学习等技术对海量数据进行深度挖掘和分析，提供预测性维护、优化调度等高级功能；前端界面为用户提供直观、易用的操作界面，方便用户远程监控和管理系统；云平台则提供数据存储、计算资源、应用服务等支持，实现系统的远程部署、升级和维护。

**3.2 数据需求**

智慧场景学习平台构建于三大核心数据驱动支柱之上：精准设备状态监测、深度人员及行为感知分析、以及高效结果辨识。



1. 设备状态监测

平台能够前瞻性地洞察潜在故障，实现故障预警与远程主动维护，同时紧密追踪目标任务的执行进度与达成状态，确保教学环境的持续优化与高效运行。

1. 学生状态感知

深度行为感知技术为平台赋予了敏锐的洞察力，它能够细致入微地捕捉并分析学生、教师及其他参与者的行为模式与状态变化。这一功能不仅促进了对学生学习进度、兴趣偏好的精准把握，还助力教师调整教学策略，实现个性化教学，进一步增强了教学互动的针对性和有效性。

1. 达成结果信息辨识

高效的结果辨识评价机制是平台不可或缺的一环。它基于多维度、精细化的数据分析，对学习目标达成情况进行全面、客观的评价。这一机制不仅为教学效果提供了量化依据，还促进了教学反馈的即时性与准确性，助力师生双方及时调整教学策略与学习路径，确保教学目标的高质量实现。

**3.3 功能需求**

**3.3.1 课程接入**

1. 接口标准化

制定统一的API接口标准，确保不同工科实践类课程能够无缝接入智慧场景学习平台。

针对不同课程特点，开发相应的插件或模块，实现课程内容的个性化展示和交互。

1. 课程数据同步

实现课程数据（如教学大纲、教学视频、实验指导等）的自动同步与更新，确保平台上的课程内容始终与最新教学要求保持一致。

1. 用户权限管理

设置不同用户角色（如教师、学生、管理员）的权限，确保课程资源的合理使用和管理。

**3.3.2 任务指导书设计与展示**

1. 任务目标明确

在任务指导书中明确每个实践任务的目标、要求、步骤和预期成果，确保学生能够清晰理解任务要求。

1. 任务分解细化

将复杂任务分解为若干个子任务，每个子任务都有明确的达成量纲和识别点，便于学生逐步完成并检验学习成果。

1. 智能提示与反馈

在任务指导书中嵌入智能提示功能，根据学生的操作情况提供实时反馈和建议，帮助学生更好地完成任务。

**3.3.3 量纲、识别点、识别链解析**

1. 任务进度跟踪

通过智慧学习平台跟踪学生的任务完成情况，自动记录每个子任务的达成量纲。

1. 识别点与识别链构建

根据任务目标和子任务要求，构建识别点和识别链。识别点是指能够反映任务完成情况的关键点或步骤；识别链则是将多个识别点串联起来形成的完整任务路径。

利用AI技术对学生的操作行为进行分析和识别，自动判断学生是否达到识别点要求，并生成识别链。

**3.3.4 课程设备接入**

1. 设备接口标准化

制定统一的设备接口标准，确保不同实践类课程的设备能够顺利接入智慧场景学习平台。

1. 传感器选择与部署

根据课程需求选择合适的传感器（如温度传感器、压力传感器、位移传感器等），并将其部署在关键位置。

传感器应能够实时采集设备运行状态和实验数据，并将数据传输至智慧学习平台。

1. 数据传输与存储

确保传感器采集的数据能够稳定、可靠地传输至云平台。

在云平台上实现数据的专业、高效、可靠存储，以便后续的数据分析和挖掘。

**3.3.5 基于AI的人员辨识、设备状态辨识、结果辨识**

1. 人脸识别与行为识别

利用AI技术对学生的面部特征和行为进行识别和分析，确保学生在实践过程中的身份和行为符合课程要求。

1. 设备辨识与目标达成结果识别

通过图像识别或物联网技术辨识实验设备及其运行状态。

利用AI算法对实验数据进行处理和分析，自动判断学生是否达到目标达成结果的要求。

1. 智能评估与反馈

根据AI技术的识别结果对学生进行智能评估，并生成详细的评估报告和反馈意见。

评估报告应包括学生的任务完成情况、存在的问题以及改进建议等内容，帮助学生更好地提升实践能力。

**第四章 硬件框架**

系统框架设计包括一套具有采集、分析能力的计算系统，实现内核级的信号监控，自动跟踪外围系统的作业进展，在硬件层面对学习者的学习进程进行记录、分析和判断。

**4.1 多源多模态感知框架**

为适应视觉和行为感知类结果的辨识需要，多源多模态感知系统如图所示，系统包括视觉感知系统，声音感知系统，光电信号传感辅助采集等，能够完成功能：

1. 基于 AI 的视频分析系统

能够从连续多帧画面中对学习结果进行准确辨识和判断，判断学习者的学习状态和设备运行状态。

1. 学习者行为分析

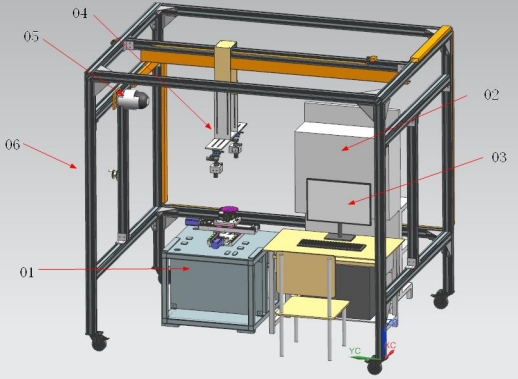
能够对参与者的参与过程，包括参与时间，参与内容进行分析判断

1. 学习过程和结果评价

对声音，或者其他类型的物理信号，包括亮暗，脉冲，纹波，频率等进行实时采集和分析处理。



传统的实践学习平台增设这些感知功能单元与系统之后，其形态如图所示，01 为智能学习岛，02 为电气控制柜，03 为控制操作界面，04 为智能学习岛视觉感知系统，05 为学习者行为感知分析系统，06 场景外框架结构。以知识应用与操作训练为主要目标的实践学习平台将升级为智慧场景学习平台。



**4.2 全方位音视频信息采集框架需求**

智慧场景学习平台将构建一套高效、灵活的全方位音视频信息采集系统，通过集成先进的机械装置与音视频采集设备（摄像头、麦克风），实现对目标区域或对象的360度无死角、高质量音视频信息采集。该系应用于工科实践类课程教学，要求具备高度的自动化、稳定性与可扩展性。

**4.2.1 钢架结构设计**

承重能力钢架需具备足够的承重能力，以支撑摄像头、麦克风及其附属设备（如云台、滑轨等）的总重量，并考虑预留一定的安全冗余。

稳定性结构设计需确保在各种环境条件下（如风力、震动）都能保持稳定，避免因摇晃而影响音视频采集质量。

可调节性：钢架应设计有高度、角度调节功能，以适应不同安装环境和采集需求。

安装便捷性：提供标准化安装接口和详细的安装指南，便于快速部署和拆卸。

**4.2.2 滑轨系统**

1. 平滑度

滑轨需确保摄像头在水平或垂直方向上的移动平滑无顿挫，以保证视频画面的连续性。

1. 精度控制

通过精密的传动机构和定位装置，实现摄像头位置的精确控制，满足高精度采集需求。

1. 速度可调

滑轨系统应具备速度调节功能，以适应不同场景下的采集速度要求。

1. 耐用性

选用高质量材料，确保滑轨系统长期运行稳定可靠，减少维护成本。

**4.2.3 云台系统**

1. 全方位旋转

云台需支持水平360度、垂直一定角度（如-90°至+90°）的连续旋转，实现无死角监控。

1. 精准定位

具备精准定位功能，可通过软件或遥控器快速调整至指定位置。

1. 防抖功能

内置或外接防抖技术，减少因机械运动产生的图像抖动。

1. 负载能力

根据所搭载摄像头和附件的重量，选择合适的云台负载能力。

**4.2.4 摄像头与麦克风集成**

1. 兼容性

确保摄像头与麦克风能够与钢架、滑轨、云台等机械装置无缝集成，实现一体化控制。

1. 高清画质

摄像头需支持高清甚至超高清视频录制，具备自动曝光、白平衡等调节功能。

1. 清晰拾音

麦克风应具备高灵敏度、低噪音特性，能够清晰捕捉目标声音，支持定向或全向拾音。

1. 远程控制

支持通过网络或专用控制器进行远程参数设置、预览和录像控制。

1. 变焦控制

使用机械运动变焦装置能够有效取代价格高昂的电子变焦摄像头

**第五章 感知系统**

在智慧场景学习平台中，感知系统层扮演着至关重要的角色，它是整个平台的数据入口和智能决策的基础。通过集成多种先进的传感器和数据采集器、摄像头、麦克风等，感知系统层能够全面、准确地捕捉工科实践类课程现场的各类信息，这些信息对于提升教学质量、优化学习体验、保障实践安全等方面具有重要意义。

**5.1 感知信息分类**

1. 环境感知

感知系统层首先需要感知并采集实践课程现场的环境参数，如温度、湿度、光照强度等，这些数据对于评估实验条件是否适宜、保护实验设备免受环境影响至关重要。

1. 音视频信息

通过高清摄像头和麦克风阵列，实时捕捉实验现场的音视频信息。视频信息可以用于远程监控、学生身份、行为、情绪分析以及教学示范的录制与回放；音频信息则可用于语音识别与交互，如指令执行问题、解答等。

1. 设备状态信息

对实验室内各种工科实践电气设备（如机床、电子仪器、计算机等）进行实时监测，获取其运行状态、故障信息、使用时长等关键数据。这有助于及时发现并处理设备故障，保障教学活动的顺利进行，同时也为设备的维护管理，及结果辨识、课程评分提供数据支持。

1. 安全预警

通过感知系统层收集的数据，结合智能算法进行风险评估，对可能存在的安全隐患进行预警，如火灾、触电、机械伤害等。这有助于迅速采取应急措施，保障师生的人身安全。

**5.2 设备类型分类**

1. 摄像头

摄像头是视觉信息采集的主要设备，用于捕捉实践课程现场的图像和视频信息。高清摄像头能够提供清晰的画面。摄像头通常具备网络传输功能，可以实时将视频数据传输至服务器或终端设备。

1. 麦克风

麦克风是音频信息采集的关键设备，用于捕捉现场的声音信息，包括教师讲解、学生提问、设备操作声音等。高质量的麦克风能够确保音频信号的清晰度和准确性，为语音识别、语音交互等应用提供可靠的数据源。特别是具有语音识别功能麦克风，能够为数据分析提供更加便捷、准确的感知信息。

1. 智能手环

智能手环是一种流行的可穿戴设备，通常戴在手腕上，具有多种功能，旨在对实践过程中学生心理状态以及生理状态进行实时监测。使用智能手环可以实时监测学生的心率，通过心率的变化来评估学生的压力水平以及学生的学习状态。

1. 传感器

* 温度传感器

NTC热敏电阻、PT100、热电偶等，用于测量设备的温度，确保设备在合适的温度范围内运行，防止过热或过冷导致的设备损坏或性能下降。

* 湿度传感器

电容式湿度传感器、阻抗式湿度传感器、电解式湿度传感器等，用于测量环境中的湿度，确保设备在适宜的湿度条件下运行，防止湿度过高或过低对设备造成损害。

1. 数据采集器

* 电流采集器

电流采集能够实时监测教学设备在运行过程中的电流变化，包括电流的大小、以及波动情况，并转换为可处理的数字信号。通过电流采集，可以及时发现电流的异常波动，如电流过大、过小、波动频繁等，从而判断设备是否存在短路、断路、过载等故障。这对于了解设备的运行状态、评估设备性能以及发现潜在故障具有重要意义。

* 电压采集器

通过电压采集，可以实时获取设备上的电压参数，监测电压是否在正常范围内波动，电压的稳定性如何，并转换为可处理的数字信号，防止事故发生及扩大，可以评估设备的运行状态，具有实时监控与状态评估、故障检测、数据分析与优化以及支持决策与规划等多重作用。

* 状态数据采集器

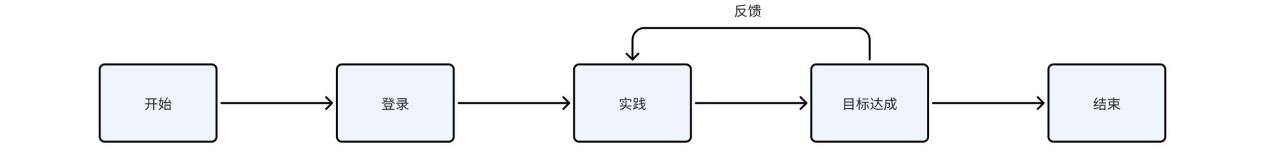
状态数据采集器在电气设备的监控和维护中扮演着至关重要的角色。这些设备能够实时或定期地从电气设备的寄存器、接口等关键部位采集指示状态的重要数据，进而通过数据分析来评估设备的运行状况、预测潜在故障，甚至实现远程监控和自动化管理。

**第六章 后台调度系统**

后台调度系统负责对感知端采集到的所有数据进行处理分析，以及应用程序的业务逻辑的实现。智慧场景学习平台从达成云端获取到课程任务书，根据任务书内容对现场设备及资源进行合理的安排和调度，方便学生开展课程学习，后台调度系统获取感知系统层后的数据进行处理分析后传输至前端进行展示。后台调度系统通过对平台资源和任务执行顺序的有效管理，提高整体系统的运行效率和性能。

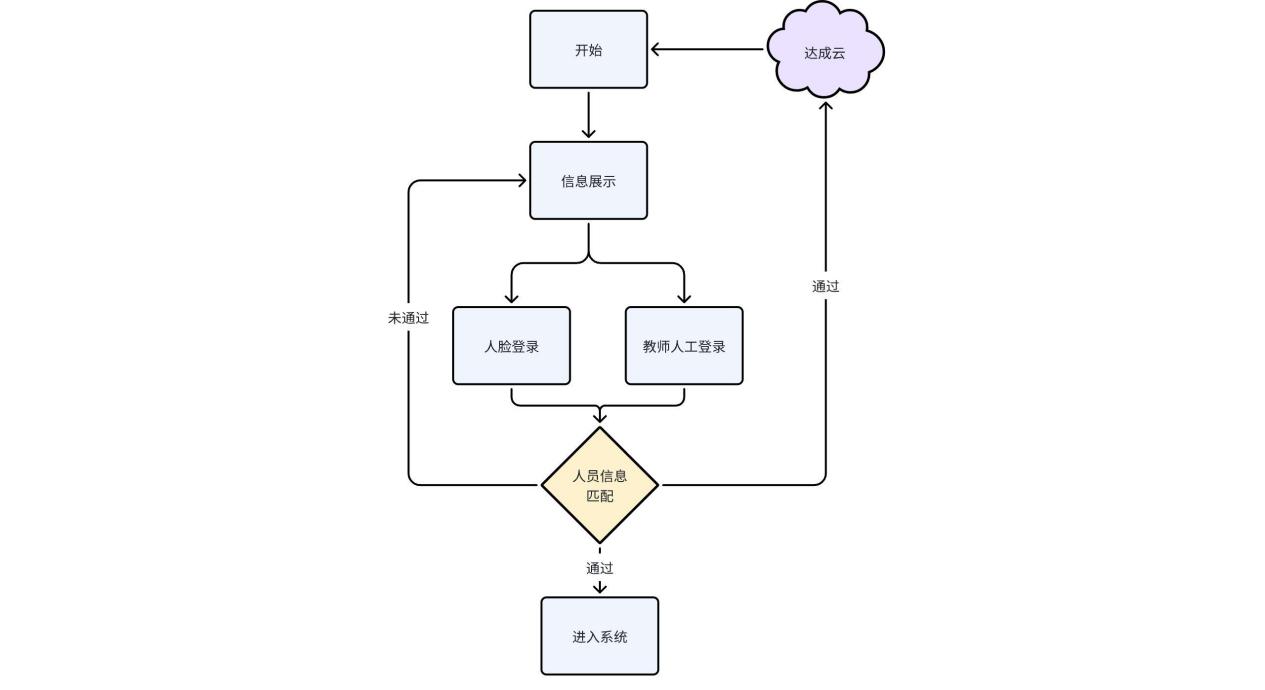
**6.1 整体系统运行流程**

使用智慧场景学习平台开始学习工科实践课，首先在系统平台登录，接着系统进入实践环节，将全程感知目标的达成度。每个任务若学生未完成，学生将重新开始。若完成则结束当前任务的感知，进入任务结束的任务，将进行相关信息保存及上送达成云。



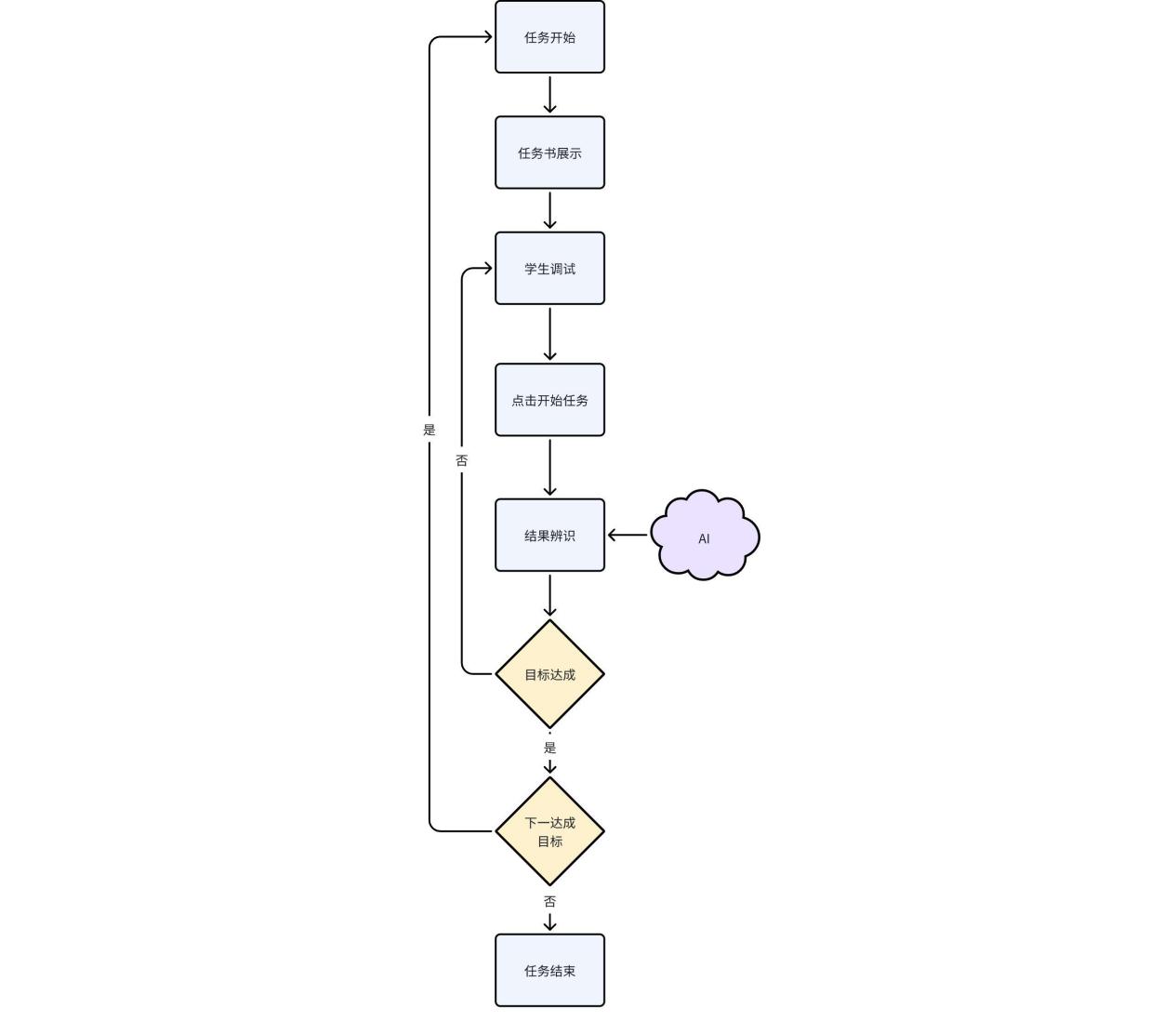
**6.2 登录流程**

学生使用智慧场景学习平台登录系统时，系统首先与达成云平台同步信息。之后进入信息展示页面，展示预约相关学生信息。之后可以选择人脸识别自动登录，或者是现场老师人工选择学生登录。当所需匹配的信息一致时，将进入系统平台的实践课程管理部分，如果信息匹配不一致时，则返回信息展示环节。



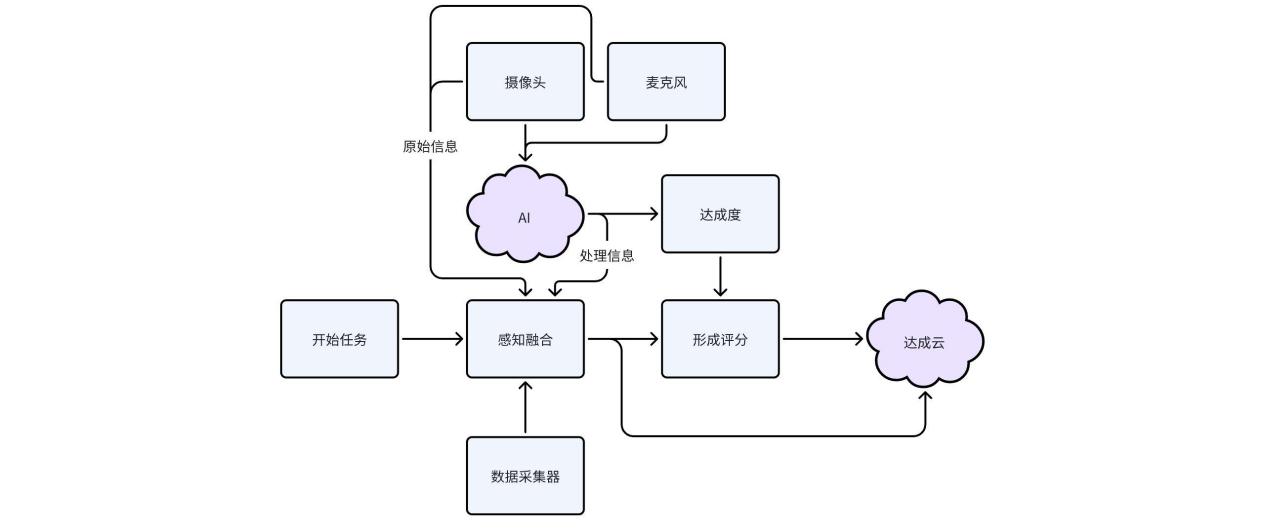
**6.3 实践流程**

学生完成系统登录后，系统进入实践课程管理部分。学生在系统平台选择需要完成的任务，平台首先展示任务相关的信息包括任务指导书等。学生可以操作场景内的电气设备，测试自己的代码等。当学生点击开始任务时，系统将全程感知场景内数据流通过AI系统对学生状态、设备状态、结果状态进行辨识，并自动对目标达成度进行评估，形成任务的评分。当任务完成后学生可以选择任务结束，也可以选择下一个任务，系统平台将进入下个任务开始环节。



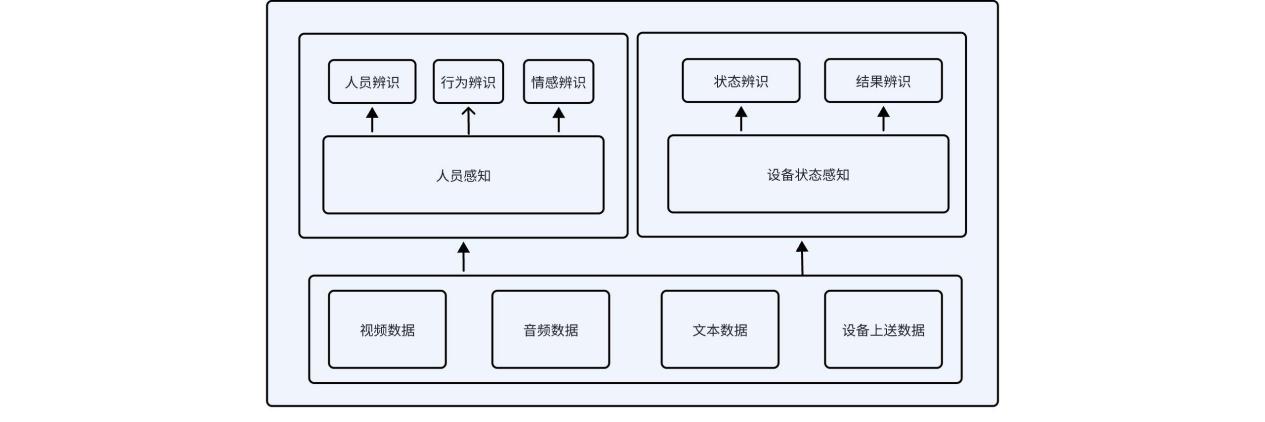
**6.4 结果辨识流程**

结果辨识流程是完成整个实践课程任务流程中的重要组成部分。系统平台需要获得全部感知设备的感知数据，通过AI对所有感知端数据进行融合分析，形成最终的课程评分，并将这个过程中产生的阶段数据和最终分数上传至达成云平台。



**第七章 AI 部分**

由传感器、数据采集器、摄像头、麦克风等感知设备得到学习场景下的感知信息，较复杂的信息处理将由AI部分完成。主要分为人员感知类及设备状态类两类信息的处理。其中人员感知类信息即需要分析出学生身份、行为、情绪等。设备状态感知类信息需要分析出设备的运行状态、结果达成的评价等。



**7.1 人员感知需求**

**7.1.1 学生登陆人脸识别需求**

**7.1.1.1 人员辨识需求**

方式一：

学生通过人脸识别，得到匹配的人员信息显示在登录页面中。

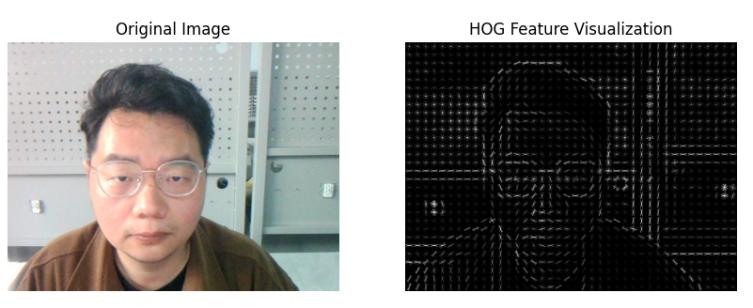
方式二：

在场老师通过账号密码选定学生登录

**7.1.1.2 人脸识别技术方案**

* A. 人脸检测

Dlib通过结合方向梯度直方图（Histogram of Oriented Gradients, HOG）特征和级联分类器的方法，实现了高效的人脸检测。HOG特征能够有效地描述图像中的边缘和纹理信息，而级联分类器则通过多层弱分类器的级联，提高了检测的准确性和效率。



图：HOG特征可视化

Dlib中的级联分类器采用多层结构，每一层都是一个弱分类器，通过逐步筛选和排除非人脸区域，最终得到人脸检测结果。这种结构不仅减少了计算量，还提高了检测速度。

HOG特征（Histogram of Oriented Gradients）是一种通过计算和统计图像局部区域的梯度方向直方图来构成的特征，常用于目标检测和图像识别

* B. 人脸识别

构建SENet50分类模型得到脸部信息的2048维特征向量，通过比较在库人员特征向量的欧式距离，从而确认登录者的身份。

SENet50是一种结合了Squeeze-and-Excitation（SE）模块的深度残差网络，它在ResNet50的基础上引入了SE模块，通过动态地重新标定特征通道间的响应强度，增强了模型对重要特征的敏感度。在脸部信息提取中，SENet50能够捕捉到人脸图像中的复杂特征，并将其编码为2048维的特征向量，这些特征向量包含了丰富的脸部信息，如表情、姿态、光照变化等。

维度选择与性能：选择2048维作为特征向量的维度是权衡了计算效率与表征能力之后的结果。这个维度的特征向量既能够保留足够的脸部信息以供后续的分类、识别等任务使用，又能够在一定程度上控制计算量和模型复杂度，确保在实际应用中具有较好的性能表现。

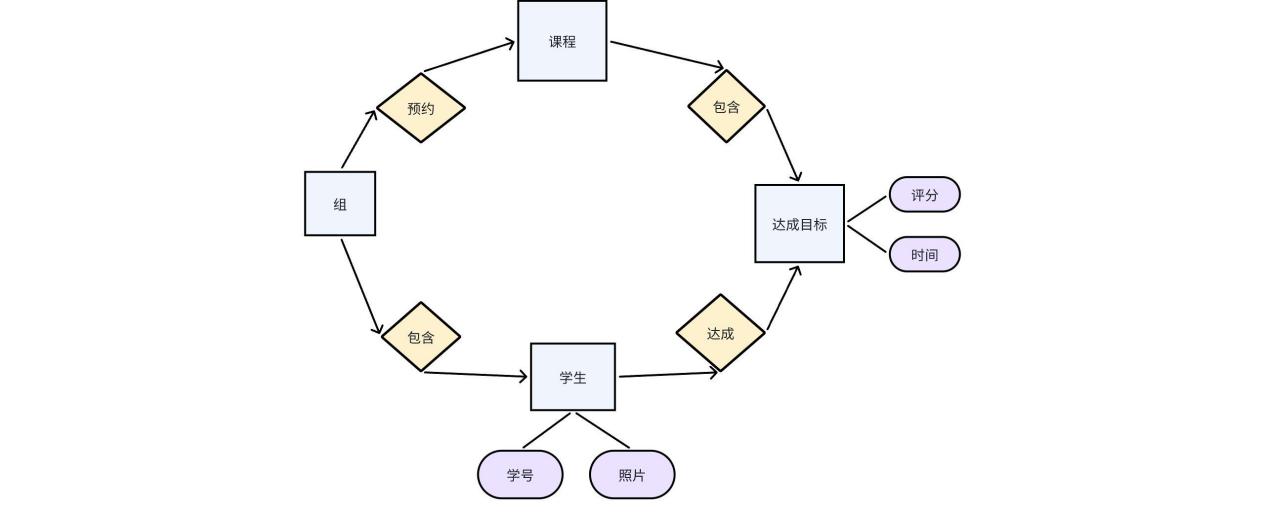
直接计算两张人脸图片的特征向量（通常是多维数组表示的embedding）之间的欧式距离。这种方法直观且易于实现，计算过程简单明了，适合大规模数据集的处理。基于欧式距离的人脸识别系统通常具有较高的识别率，尤其在较小的人脸数据集上表现更加出色。此外，欧式距离能够较好地抵抗光照、表情等因素的干扰，因此在复杂环境下表现较为稳定。



图：人脸识别效果

**7.1.2 学生信息标准需求**

云平台数据库：



* 学生表

学生号（主键） 学生姓名 照片

* 组表

组号（主键）组名 学生号(按逗号隔开)

* 课程预约表

预约号（主键） 预约时间 组号

* 课程登录记录表

（预约号 组号 学生号)(主键) 登录时间

* 达成目标表

达成目标号（主键）

* 课程成绩记录表

达成目标号 预约号 组号 学生号 评分 开始时间 达成时间 达成用时

**7.1.3 界面设计**

欢迎界面需求：通过查询数据库，欢迎界面展示了该场景通过在线预约小组的信息，点击进入则可以进行登录。

登录界面需求：界面左侧用于实时显示人脸识别画面，右侧则用于动态展示识别到的人员结果。

老师验证登录需求：界面中出现需要勾选的学生，老师账号输入框，老师密码输入框

识别画面速率需求：画面FPS不小于8。每秒更新8次以上画面，可以提供流畅的视觉体验。

**7.1.4 访问控制**

平台首先在小组登录时根据预约信息得到该时间段内小组的组员信息，组员人脸信息。人脸身份信息从数据库读入本地。

登录时得到人员辨识结果，记录登录时间。

登录平台后平台会显示组员中已登录的学生和未登录的学生。

课程时间内允许退出补充登录人员，并记录补充登录人员信息，已登录人员状态不变。

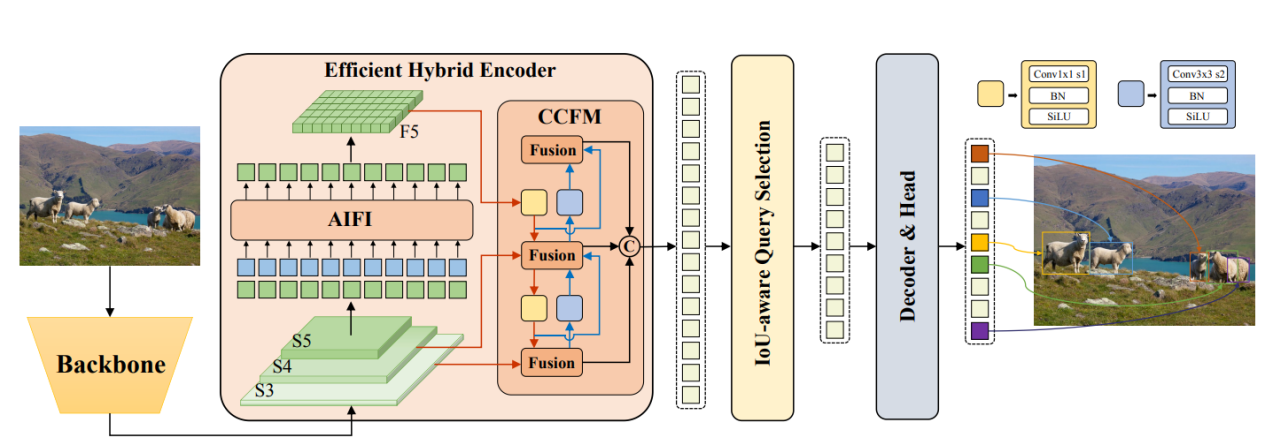
允许学生重新上传自己的照片。

**7.1.2 学生学习状态检测需求**

**7.1.2.1 学生行为检测需求分析**

通过摄像头拍摄学生在实践过程中的行为，设计行为检测算法对学生的学习行为进行检测分类，分析学生在实践过程中中各类行为的时间占比，从而实现学生在实践过程中的学习状态的检测与分析。

**7.1.2.2 学生学习行为检测技术方案**



RT-DETR网络结构图

针对学生的学习行为检测，提出使用RT-DETR进行实时学习者学习行为检测

从结构上来看，RT-DETR可以分为三大块：主干网络、颈部网络以及解码器。我们分别来说一下这三大块。

**a. 主干网络**

对于主干网络，RT-DETR采用CNN网络，如流行的ResNet系列，或者百度自家研发的HGNet。当然，使用ViT系列的主干网络如SwinTransformer及其后续诸多变体也是可以的，因为就特征提取而言，虽然研究初期ViT系列来势汹汹，颇有山雨欲来风满楼的大变动之驾驶，但后来的诸多CNN工作如ConvNeXt等工作则证明了“ViT的训练trick可能是关键之处”，由此抹平了CNN架构和ViT架构在视觉表征学习上的性能差距。但毋庸置疑，就当下的硬件而言，CNN架构无疑是快于ViT架构的，因此，从实时性的角度出发，选择CNN架构来做特征提取还是有助于提高DETR系列的实时性，进而提升实用性。



RT-DETR主干网络，红色虚线框

**b. 颈部网络**

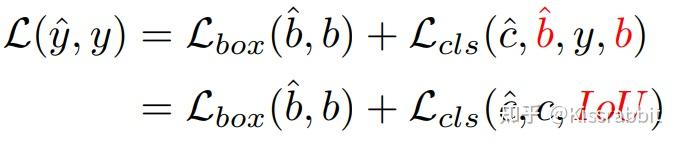
对于颈部网络，RT-DETR采用了一层Transformer的Encoder，只处理主干网络输出的 𝑆5 特征，即图6中所展示的AIFI（Attention-based Intra-scale Feature Interaction）模块。尽管这个模块的名字起得很是那么一回事，但其实他就是一个很普通的Transformer的Encoder层，包含标准的MSAH（或者Deformable Attention）和FFN。

**c. 标签匹配与损失函数**

最后，再说一下RT-DETR的assignment和loss两大部分。尽管RT-DETR的前大半部分保留了太多的CNN框架的痕迹，但对于最终的检测头，或者说解码器，RT-DETR还是选择了基于cross attention的Transformer decoder，并连接若干MLP作为检测头，因此，RT-DETR无疑是DETR架构。具体来说，RT-DETR选择DINO的decoder，使用了具体线性复杂度的deformable attention，同时，在训练阶段，使用到了DINO的“去噪思想”来提升双边匹配的样本质量，加快训练的收敛速度。整体来看，RT-DETR的检测头几乎就是把DINO的照搬了过来，当然，其中的一些边边角角的操作给抹掉了，尽可能达到“精简”的目的。

随后，在训练的label assignment环节，还是用到了DETR系列万年不变的“匈牙利匹配”，当然，这里的实现还是参考了DINO的实现，考虑到了一些噪声样本的匹配。

而在训练的loss上，回归损失还是GIoU损失和L1损失，而类别损失却做了一点改动——引入“IoU软标签”，如下图所：



RT-DETR的优化目标

所谓的“IoU软标签”，就是指将预测框与GT之间的IoU作为类别预测的标签。熟悉YOLO工作的读者一定对此不会陌生，其本质就是已经被广泛验证了的IoU-aware。在最近的诸多工作里，比如RTMDet、DAMO-YOLO等工作中，都有引入这一理念，去对齐类别和回归的差异。

之所以使用IoU软标签，是因为按照以往的one-hot方式，完全有可能出现“当定位还不够准确的时候，类别就已经先学好了”的“未对齐”的情况，毕竟类别的标签非0即1。但如果将IoU作为类别的标签，那么类别的学习就要受到回归的调制，只有当回归学得也足够好的时候，类别才会学得足够好，否则，类别不会过快地先学得比回归好，因此后者显式地制约着前者。

在使用了这个技巧后，显然训练过程中，类别的标签不再是此前的0和1离散值，而是0~1的连续值，那么就不能再使用标准的focal loss，因而采用了variable focal loss（VFL），这一损失函数也被用在了YOLOv6。

不过，本以为匈牙利匹配的类别代价也会用IoU软标签，但看过其代码后，发现还是用01离散标签（使用focal loss计算类别代价），或许，这里使用01连续值标签也许会有好处？笔者对此不敢妄言。

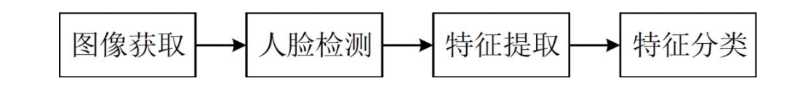
综上所述，RT-DETR的Transformer decoder基本可以看作是DINO的transformer decoder的“精简实时版”，并在此基础上做了一些小幅改动，以便在不破坏推理速度的前提下，适当提升模型的性能。

**7.1.2.3 学生情绪识别需求分析**

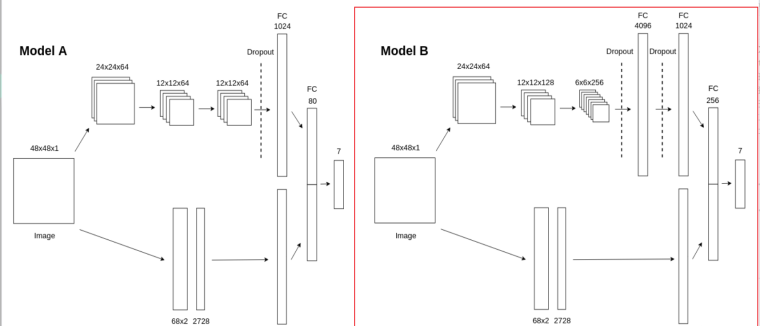
通过摄像头拍摄学生在实践过程中的面部情绪，设计情绪识别算法对学生的学习情绪进行识别，分析学生在实践过程中的情绪状态，从而实现学生在实践过程中的学习状态的检测与分析。

**学生情绪识别方案**

人类的面部表情至少有21种，除了常见的高兴、吃惊、悲伤、愤怒、厌恶和恐惧6种，还有惊喜（高兴＋吃惊）、悲愤（悲伤＋愤怒）等15种可被区分的复合表情。学生情绪识别系统如图所示，主要由图像获取、人脸检测、特征提取、特征分类四部分组成。



下图我们可以看出，在 Model B 的卷积部分，输入图片 shape 为 48X48X1，经过一个3X3X64卷积核的卷积操作，再进行一次 2X 2的池化，得到一个 24X24X64 的 feature map 1（以上卷积和池化操作的步长均为1，每次卷积前的padding为1，下同）。将 feature map 1经过一个 3X3X128 卷积核的卷积操作，再进行一次2X2的池化，得到一个 12X12X128 的 feature map 2。将feature map 2经过一个 3X3X256 卷积核的卷积操作，再进行一次 2X2 的池化，得到一个 6X6X256 的feature map 3。卷积完毕，数据即将进入全连接层。进入全连接层之前，要进行数据扁平化，将feature map 3拉一个成长度为 6X6X256=9216 的一维 tensor。随后数据经过 dropout 后被送进一层含有4096个神经元的隐层，再次经过 dropout 后被送进一层含有 1024 个神经元的隐层，之后经过一层含 256 个神经元的隐层，最终经过含有7个神经元的输出层。一般再输出层后都会加上 softmax 层，取概率最高的类别为分类结果。



通过数据的前向传播和误差的反向传播来训练模型了。在此之前，还需要指定优化器（即学习率更新的方式）、损失函数以及训练轮数、学习率等超参数。在本项目中，采用的优化器是SGD，即随机梯度下降，其中参数weight\_decay为正则项系数；损失函数采用的是交叉熵；可以考虑使用学习率衰减。

**7.1.3、学生语音识别需求**

**7.1.3.1 学生语音识别需求分析**

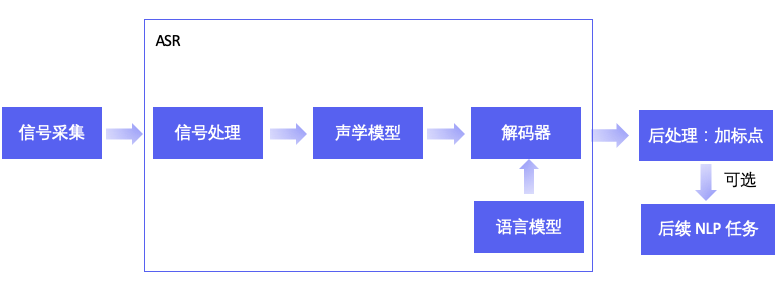
通过麦克风录制学生在实践过程中的语音，设计语音识别算法对学生的语音进行识别，方便学生进行人机交互。

**7.1.3.2 音频识别技术方案**

PP-ASR 是一个 提供 ASR 功能的工具。其提供了多种中文和英文的模型，支持模型的训练，并且支持使用命令行的方式进行模型的推理。 PP-ASR 也支持流式模型的部署，以及个性化场景的部署。 PP-ASR支持多种预训练模型：released\_model。 其中效果较好的模型为支持流式 ASR 的 Conformer 模型。

特点

语音识别的基本流程如下图所示：



PP-ASR 的主要特点如下：

提供在中/英文开源数据集 aishell （中文），wenetspeech（中文），librispeech （英文）上的预训练模型。模型包含 deepspeech2 模型以及 conformer/transformer 模型。

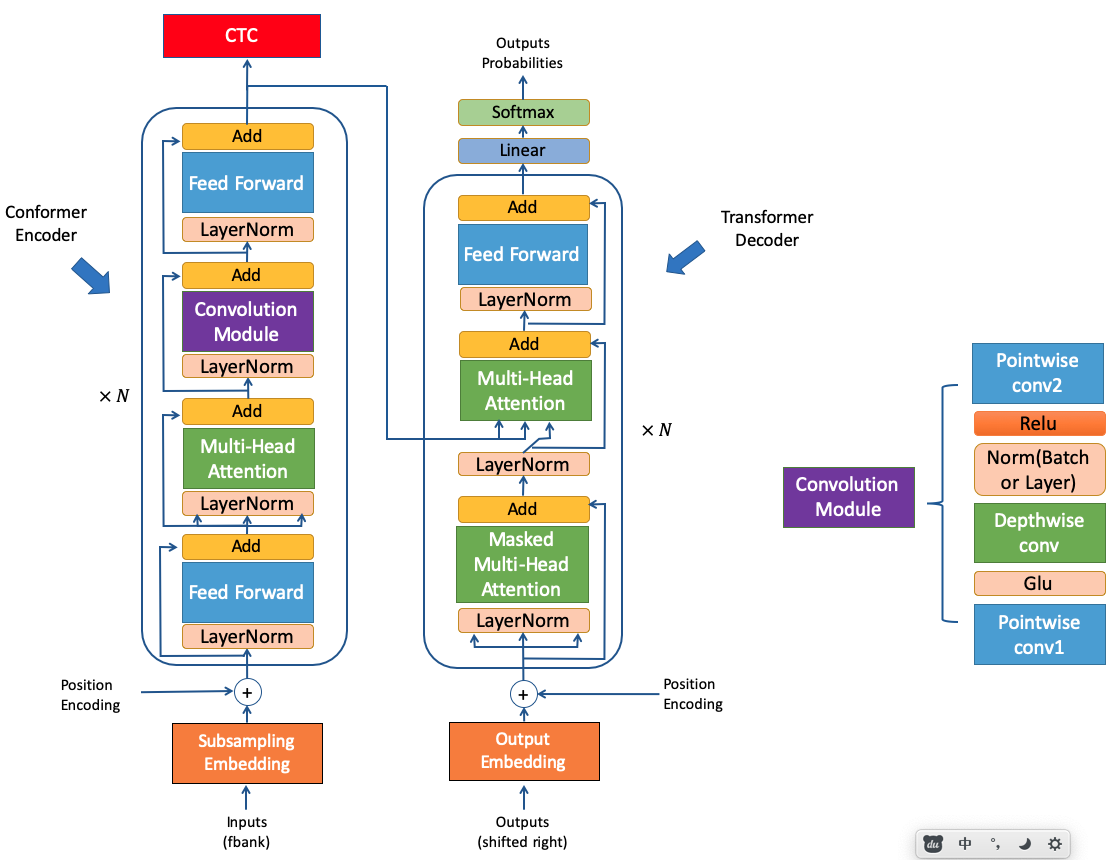
支持中/英文的模型训练功能。

支持命令行方式的模型推理，可使用 paddlespeech asr --model xxx --input xxx.wav 方式调用各个预训练模型进行推理。

支持流式 ASR 的服务部署，也支持输出时间戳。

支持个性化场景的部署。

1. 流式 Conformer 模型原理



Conformer 主要由 Encoder 和 Decoder 两个部分组成，整体的模型结构和 Transformer 非常相似。

Conformer 和 Transformer 有着相同的 Decoder，主要的区别有2点：

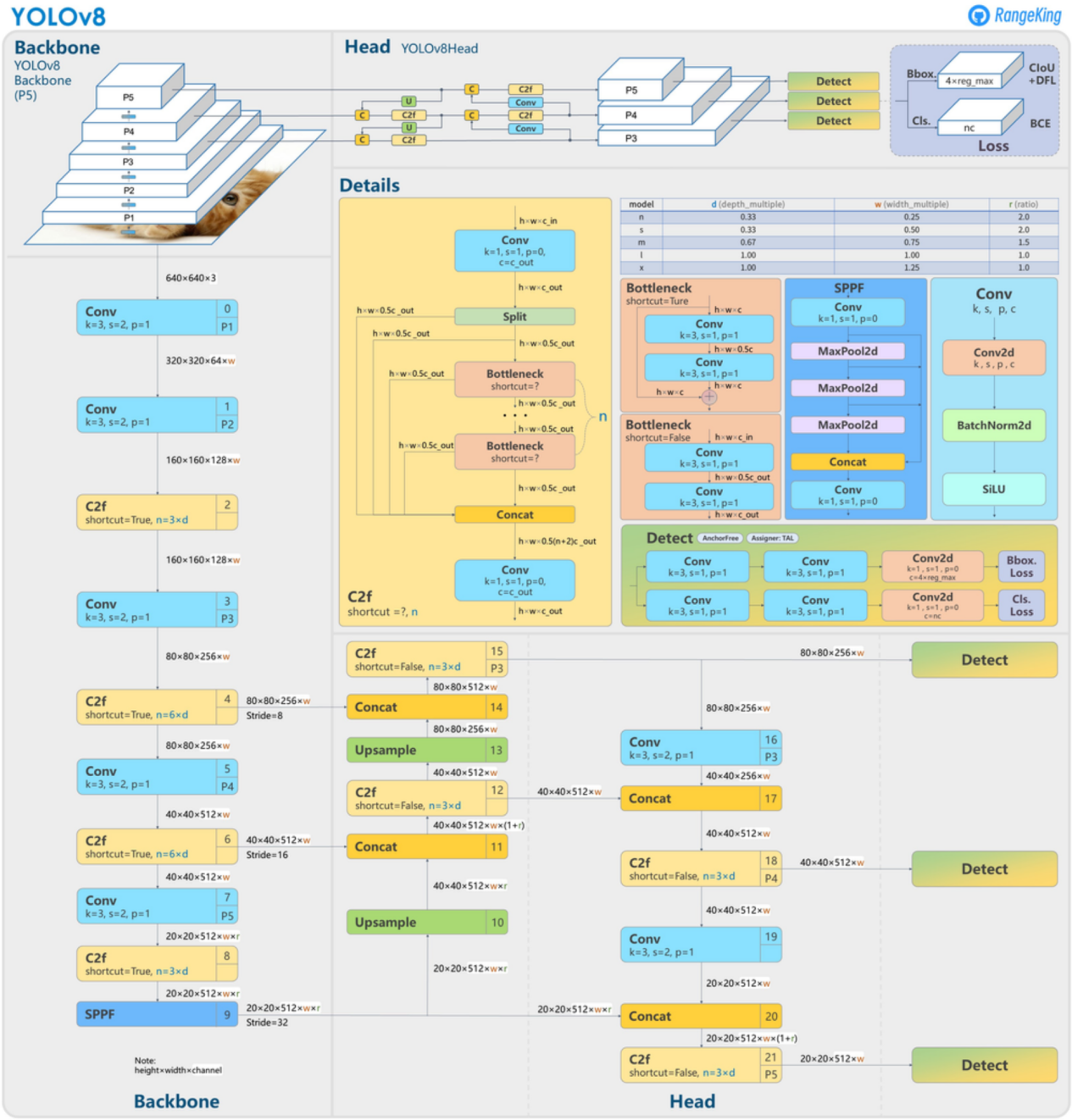
1. Conformer 的 Encoder 中包含了 conv 模块。该 conv 模块由 pointwise conv，GLU层，Depthwith conv， RELU层，以及第二层 pointwise conv， 共5个部分组成。
2. Conformer 的 Encoder 使用了2层 FeedForward，分别位于每层 encoder的头和尾，并且设置每层输出的权重设置为0.5，整体类似于一个汉堡的结构。

**7.2 设备状态感知需求**

通过AI算法可以辨识设备各组件的位置信息，从而得到设备状态。还可以通过分类网络感知正确的状态信息从而给出结果辨识评价，最终形成学生达成目标的分数。

**7.2.1 目标识别算法**

YOLOv8 是最新最先进的 YOLO 模型，可用于对象检测、图像分类和实例分割任务。YOLOv8 由 Ultralytics 开发，该公司还创建了具有影响力和行业定义的 YOLOv5 模型。YOLOv8继承了YOLO系列模型的高效性和实时性，并在多个方面进行了优化。YOLOv8的网络结构主要由三个大部分组成：Backbone、Neck和Head。



1. Backbone（主干网络）

Backbone部分负责特征提取，采用了一系列卷积和反卷积层，同时使用了残差连接和瓶颈结构来减小网络的大小并提高性能。YOLOv8的Backbone部分采用了C2f模块作为基本构成单元，与YOLOv5的C3模块相比，C2f模块具有更少的参数量和更优秀的特征提取能力。具体来说，C2f模块通过更有效的结构设计，减少了冗余参数，提高了计算效率。此外，Backbone部分还包括一些常见的改进技术，如深度可分离卷积和膨胀卷积，以进一步增强特征提取的能力。

1. Neck（颈部网络）

Neck部分负责多尺度特征融合，通过将来自Backbone不同阶段的特征图进行融合，增强特征表示能力。YOLOv8的Neck部分采用了类似于YOLOv5的PAN-FPN结构，这是一种高效且速度快的特征融合方式。Neck部分还包括SPPF（Spatial Pyramid Pooling Fast）模块，用于不同尺度的池化操作，将不同尺度的特征图拼接在一起，提高对不同尺寸目标的检测能力。此外，YOLOv8还可能使用了PAA（Probabilistic Anchor Assignment）模块来智能地分配锚框，以及PAN（Path Aggregation Network）模块来增强特征图的表达能力。

1. Head（头部网络）

Head部分负责最终的目标检测和分类任务，包括一个检测头和一个分类头。检测头包含一系列卷积层和反卷积层，用于生成检测结果，预测每个锚框的边界框回归值和目标存在的置信度。分类头则采用全局平均池化对每个特征图进行分类，通过减少特征图的维度，输出每个类别的概率分布。值得注意的是，YOLOv8采用了无锚点（Anchor-Free）检测机制，这减少了锚框的超参数设置，并通过直接预测目标的中心点来简化训练过程。

**7.2.2 视觉分类算法**

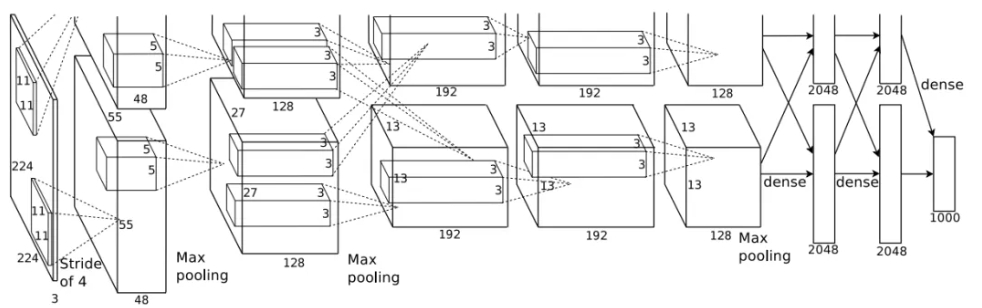
在设备状态感知系统中，图像分类算法扮演着至关重要的角色，它用于从采集到的图像数据中自动识别和分类出不同的设备状态或特征。视觉图像分类网络其目标是将输入的图像分配到预定义的类别中。深度学习中，多种网络模型和架构被广泛应用于图像分类任务，其中一些简单且常用的模型包括卷积神经网络（Convolutional Neural Network, CNN）。以下是图像分类中最流行的深度学习网络模型、架构及框架的详细介绍：

1. AlexNet

AlexNet采用低采样率把每张图片的分辨率降为256×256，具体方法就是给定一张矩形图像，首先重新缩放图像，使得较短边的长度为256，然后从结果图像的中心裁剪出256×256大小的图片。

AlexNet结构分为上下两个部分，每部分含有五层卷积层和三层全连接层，之所以分为两部分是为了方便在两片GPU上进行训练，只在第三层卷积层和全连接层处上下两部分可以交互。由于上下两部分完全一致，分析时一般取一部分即可。

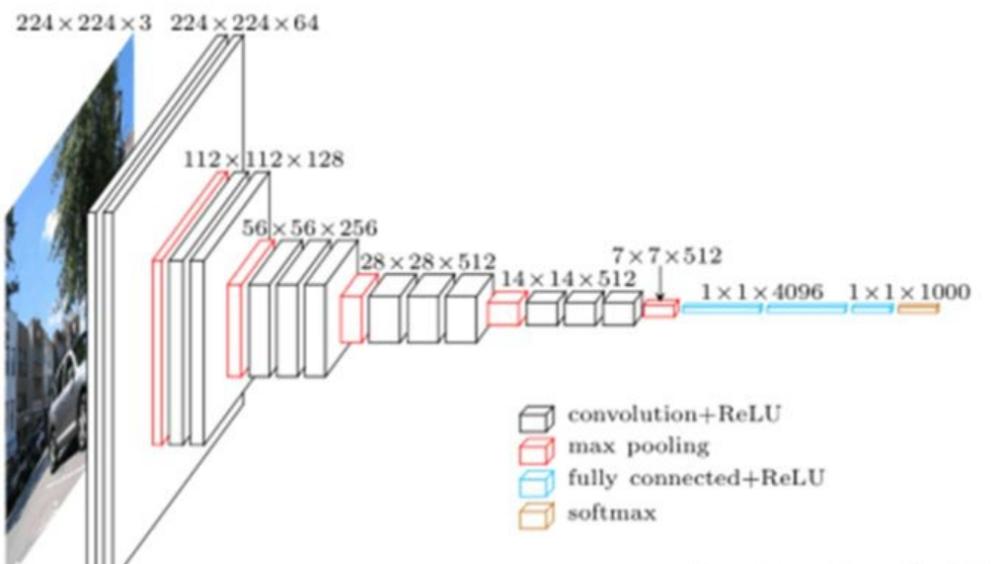
在AlexNet中使用ReLU函数作为激活函数，这种非饱和的非线性函数在梯度下降的时候要比饱和的非线性函数快得多，如tan函数。



1. VGGNet

VGGNet是由牛津大学计算机视觉组（Visual Geometry Group）提出的一种深度卷积神经网络模型，其特点在于层数较深且参数规模较大。

VGGNet使用了连续的卷积层堆叠，每个卷积层都使用3x3的卷积核，并采用ReLU激活函数进行非线性映射，通过多次卷积操作提取图像的特征。这种设计策略减少了模型参数量，并增加了网络的深度。VGGNet使用了最大池化（Max Pooling）操作，采用2x2的池化窗口进行子采样。VGGNet的全连接层设置了几个较大的隐藏层，最后连接一个softmax输出层得到分类结果，将卷积层输出的特征图展开为一维向量，并通过多个全连接层进行分类。



1. ResNet

ResNet（残差网络）是一种深度神经网络架构，由微软研究院的研究员在2015年提出。ResNet的提出解决了深度神经网络在训练过程中出现的梯度消失或梯度爆炸问题，使得训练更深层的网络成为可能。

ResNet的核心思想是通过引入“残差学习”（Residual Learning）来构建深度网络。传统的网络（如VGG）是直接学习输入到输出的映射，而ResNet则是学习输入与输出之间的残差（即差异），即学习一个恒等映射（identity mapping）加上一个残差函数。这种设计使得网络在深层时，如果当前层的输出已经足够好，那么残差映射可以被推向零，此时网络就退化为恒等映射，保持了浅层网络的表现力，从而避免了网络退化问题。

**第八章 展示需求**

智慧场景学习平台为了满足展示及交互需求，提供了场景展示大屏以及支持操作交互的用户界面。场景展示大屏集中展示学生信息、学习状态、实时数据、设备概况等基本信息，同时展示学习过程实时视频流，场景展示大屏只负责信息展示，不具备交互功能；用户界面展示实时数据和分析结果，同时学生可通过用户界面操作平台相关设备，实现人机交互功能。

**1、场景展示大屏**

场景展示大屏界面风格与工业展示界面类似，可见下图。可视化大屏集中展示学生信息、设备和系统的实时数据，通过将数据可视化的展示在大屏上，可以实时监控学生学习状态，以便于及时反映学习进度和学习效率。



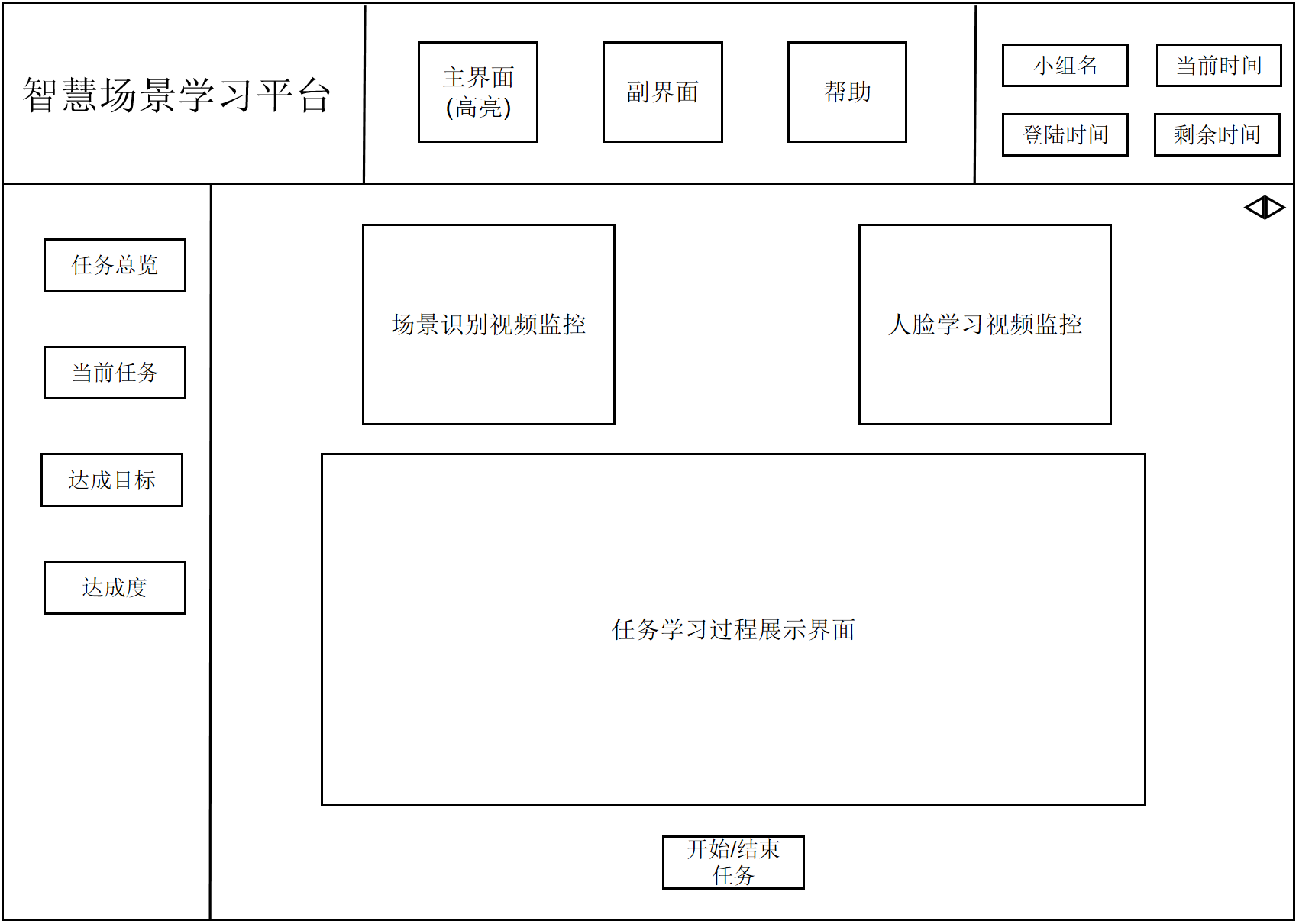
**2、用户界面**

用户界面运行于PC或工控机上，该界面上具备系统登陆、实时数据展示、人机交互等功能。按界面功能可分为以下几个界面。

a.登陆界面

|  |  |
| --- | --- |
|  |  |

b.主界面



c.帮助界面