



CS 330 MIP – Lab 05

多媒体信息处理介绍实验课- 如何做项目4

Multimedia Information Processing Introduction - Projects

Jimmy Liu 刘江

2025-03-19



Team 1:基于MOOC的个性化教育智能体 1

成员

基于MOOC的个性化教育智能体
助教：王星月老师



林枫茗

陆博

王谦益

黄少霖

白楚焓

项目描述

背景

在数字化教育迅猛发展的时代背景下，MOOC以其开放性、普惠性和可扩展性，成为全球教育革新的重要载体。然而，当前在线MOOC教育存在"重资源供给、轻学习支持"的问题。比如，学生的问题不能及时得到解答；视频作为MOOC核心载体，知识点定位困难，学生需反复跳转寻找目标片段，耗时低效，学习效率低下。

项目描述

目标

开发多模态在线教育智能体，通过融合视频、课件、讲稿等多源数据，构建具备实时答疑与知识点定位的精智能体，实现实时回答问题，给出相关的视频时间节点的功能

项目描述

步骤

1. 数据处理：将现有视频、PPT、文字稿进行处理，提取文字、图片信息，构建课程相关知识库
2. 模型训练：根据已有数据库对开源模型进行训练，使模型能够通过课程相关信息针对问题进行分析、回答

项目描述

步骤

3.问题转义：用户问的问题可能会比较泛泛，我们需要对用户的问题进行专属，将问题与数据库进行语义匹配，把问题转化成已有的文本内容

4.视频匹配：将已有数据库中的文本信息与视频进行一一匹配，给每个文本一个时间标注，当检索到符合用户需求的文本之后，可以借此返回一个时间范围

项目描述

步骤

5.搭建框架：将上述功能综合起来，搭建一个精美、实用的前端框架

可能用到的多媒体技术

1. 数据处理

2. 模型训练

- 多模态大模型、自然语言处理（问答系统）、机器学习与数据挖掘（个性化推荐）
- BERT等相关模型

可能用到的多媒体技术

3.问题转义

- 知识图谱技术、自然语言处理（文本匹配）
- BM25、SBERT、Cross-Encoders，混合检索系统（Hybrid Retrieval System），命名实体识别（NER）、问题分类

4.视频匹配

- 知识图谱技术、视频分析技术（视频分割与标注、关键帧提取）
- 视频片段定位（Temporal Sentence Grounding, TSG）

5.搭建框架

每个人的分工

基于MOOC的个性化教育智能体
助教：王星月老师



林枫茗 陆博 王谦益 黄少霖 白楚焓
视频匹配 模型训练 问题转义 问题转义 视频匹配



Team 2:基于MOOC的个性化教育智能体 2

Homework 04

1

请列出组员、助教名字、项目名称，并附上本周课上拍的组员照片

项目名称：基于MOOC的教育智能体

助教：张佳璐，王星月

组员：杨祎勃，何逸沣，神远洋，刘玉林



项目描述

项目名称：基于MOOC的教育智能体

在MOOC学习场景中，学生常面临**知识碎片化定位困难与练习反馈滞后**两大痛点。本项目以刘江老师主讲的《人工智能多媒体计算》课程为实践载体，构建垂直化教育智能问答系统，致力于实现**教学资源精准溯源与自适应学习路径引导**，助力学生建立系统化知识网络。

核心功能实现：

- 智能知识锚定**：针对课程视频、课件PPT、实验手册等多元教学资源，系统在回答学生问题时自动关联知识点坐标。例如，当学生提问“卷积神经网络的特征提取原理”时，答案将标注“详见【3.2节 深度学习基础】课件第8页图表”及“参考【实战案例解析】视频15:30–18:10演示”，实现“即问即溯”的高效复习；
- 渐进式能力培养**：基于认知负荷理论，根据当前问题复杂度动态推荐阶梯式练习题。如对“图像分类模型调参”的提问，系统将依次推送基础概念选择题、TensorFlow代码补全题、跨数据集泛化分析题，形成“理解→应用→创新”的能力跃迁路径。

可能用到的多媒体技术

多媒体技术	用途说明
语音识别（视频字幕提取）、PyMuPDF（课件解析）	提取视频时间轴标签、课件章节结构
大语言模型模型、自定义Prompt模板库	生成带定位标注的答案，控制输出格式规范
FFmpeg视频切片、OpenCV关键帧提取	MOOC课程片段定位
时间戳反向检索技术、PDF文本坐标定位	将模型输出映射到具体视频节点/课件坐标

每个人的分工

成员	主要职责
杨祎勃	阿里云资源配置、Qwen API对接、模型调试
神远洋	Prompt主模板开发（知识定位模块）
杨博乔	教学资源分析与标注（视频切片/字幕文件/课件坐标记录）
张玉柱	练习题建设与难度分级规则实施
刘玉林	系统集成测试、输出结果校验



Team 3: 低视力视觉增强辅助: HoloLens MR 头显技术的创新应用

项目描述

我们的项目聚焦于利用 Hololens 技术为视野缺失疾病患者提供高效导航支持。视野缺失疾病种类多样，如青光眼，会因眼压升高损害视神经，逐渐导致视野缺损；视网膜色素变性则会使视网膜细胞进行性退变，造成视野范围不断缩小；黄斑病变也会影响中心视力，致使部分视野出现异常。我们通过在 Hololens 设备中添加小地图功能，将患者所处环境的关键信息直观呈现，帮助他们更精准地识别方向、避开障碍，有效提升出行的安全性与自主性，改善生活质量。

可能用到的多媒体技术

- 1 动画路径规划
- 2 增强现实技术
- 3 图形表达技术
- 4 图形渲染技术
- 5 图形用户界面
- 6 图像目标检测

每个人的分工

对照上课拍的照片写下名字、分工



每个人的分工

Unity 开发：谢景涛，高胜寒

场景搭建：张阳，马鑫

计算机视觉：金邦景，张维宁



Team 4:低视力人群室内导航系统

1.

- 组员：张立远，岳雪骋，李镓璇，李亭翰
- 助教：韩载道，张颖麟
- 项目名称：低视力人群室内导航系统



低视力人群室内导航 项目及技术简述

李镓璇 岳雪聘 **张立远** 周沁儒 李亭瀚

CONTENTS

目 录



项目介绍



多媒体技术

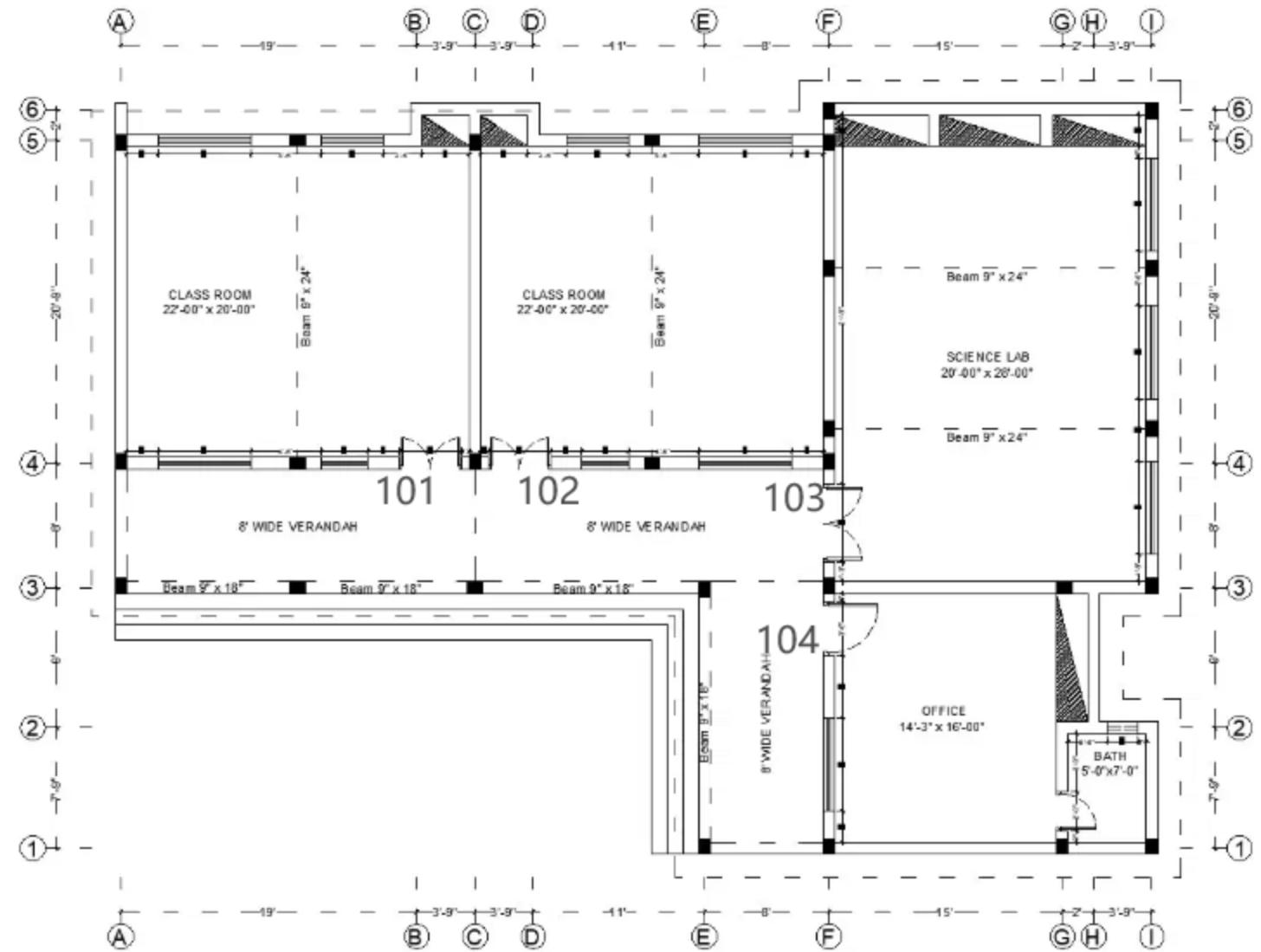


组员分工

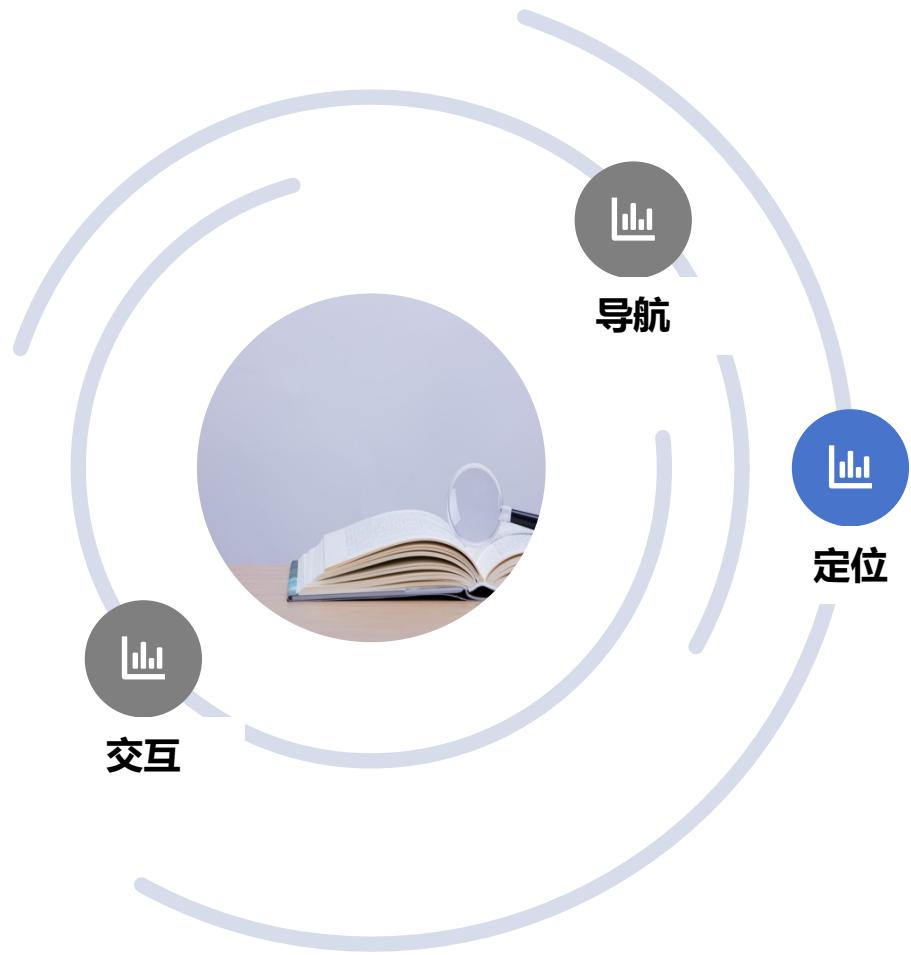
1

项目介绍

在室外应用场景中，基于GPS或北斗导航系统的定位与导航技术已相当成熟，并深度融入我们的日常生活，例如百度地图、高德地图等工具已成为不可或缺的出行助手。然而在大型商场、学校建筑等室内环境中，由于室内地图精度不足、定位技术受限以及环境复杂性等因素，目前尚未出现真正实用且普及的室内导航解决方案。对于视力正常人群，他们可通过视觉标识、电子指引或现场问询快速规划路径，但低视力群体却面临巨大挑战——无论是识别指示牌还是自主沟通都异常困难。为此，我们致力于开发一款适配智能手机的智能导航系统，用户可通过自然语言交互直接输入需求，例如询问洗手间怎么走，系统将实时解析并转化为清晰的多维度引导，通过语音播报与触觉反馈结合的方式，为低视力人群提供无障碍、高精度的室内导航服务。



项目介绍



1 交互

通过语音输入输出进行交互，低视力人群也能轻松使用

2 导航

路径规划与导航指令生成

3 定位

WIFI定位与OCR图像定位相结合

2

多媒体技术

多媒体技术

通过用户终端接收到的来自各个无线接入点发射的信号强度，通过Android SDK读取到RSSI数据进行定位

WiFi指纹定位

提前建立场景的视觉特征数据库用户输入摄像头采集的图像，经过特征提取，和数据库中的图像特征向量进行相似度匹配

视觉场景特征匹配

图像采集获取门牌号图像，OCR识别图像中的门牌号文字信息，结合平面图判断定位

OCR识别门牌号



定位方法

3

组员分工

组员分工



周沁儒：安卓端app开发

李亭翰：交互

张立远，岳雪聘：定位

李镓璇：路径规划



Team 5: BADAPPLE -BAsed on multimodality : A novel visual interpretation PiPeLinE

项目描述

本项目旨在通过人工智能技术实现**小说文本到动态短视频动画的自动化转换**，打造面向碎片化阅读用户群体的全新阅读体验。

项目核心围绕“文本、图像、语音、视频”四大模块展开，利用大型语言模型（LLM）对小说内容进行语义分段和场景分析，提取关键剧情节点，形成结构化剧情大纲。

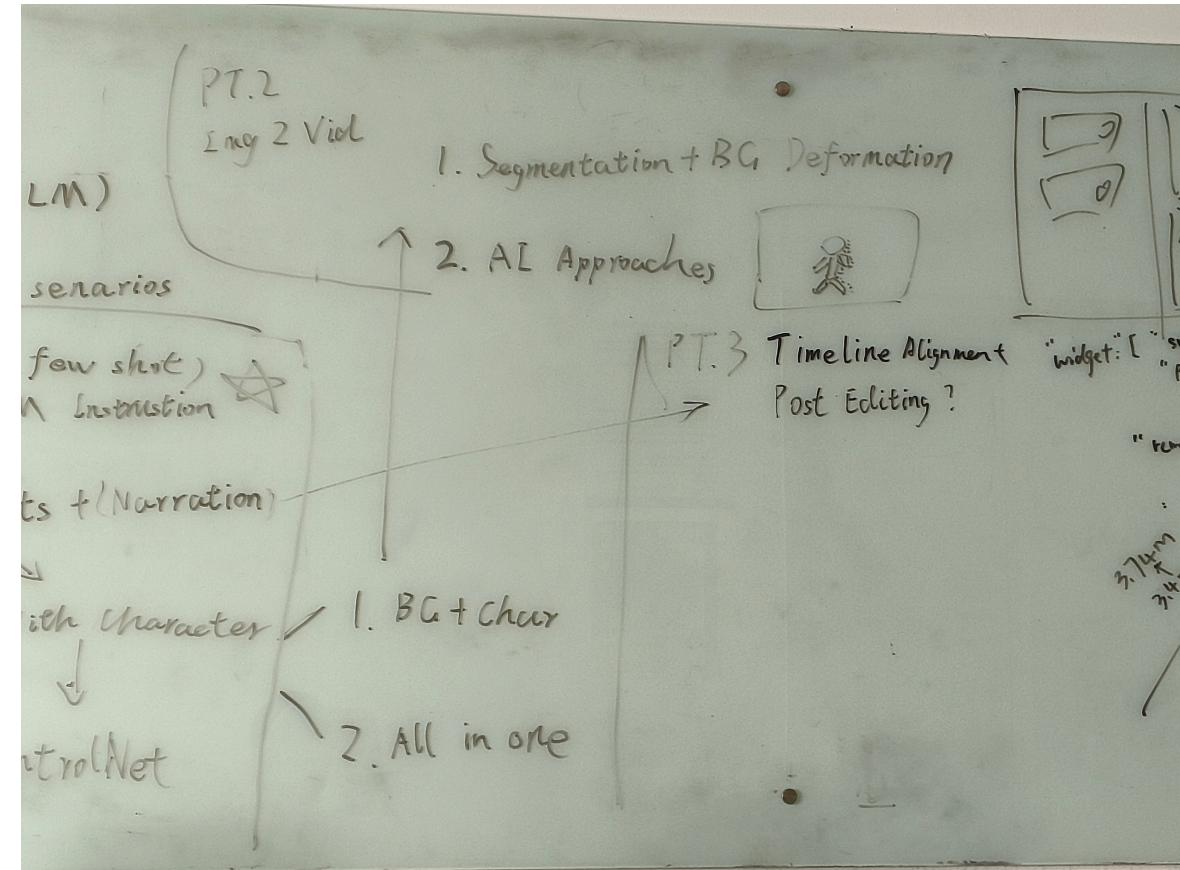
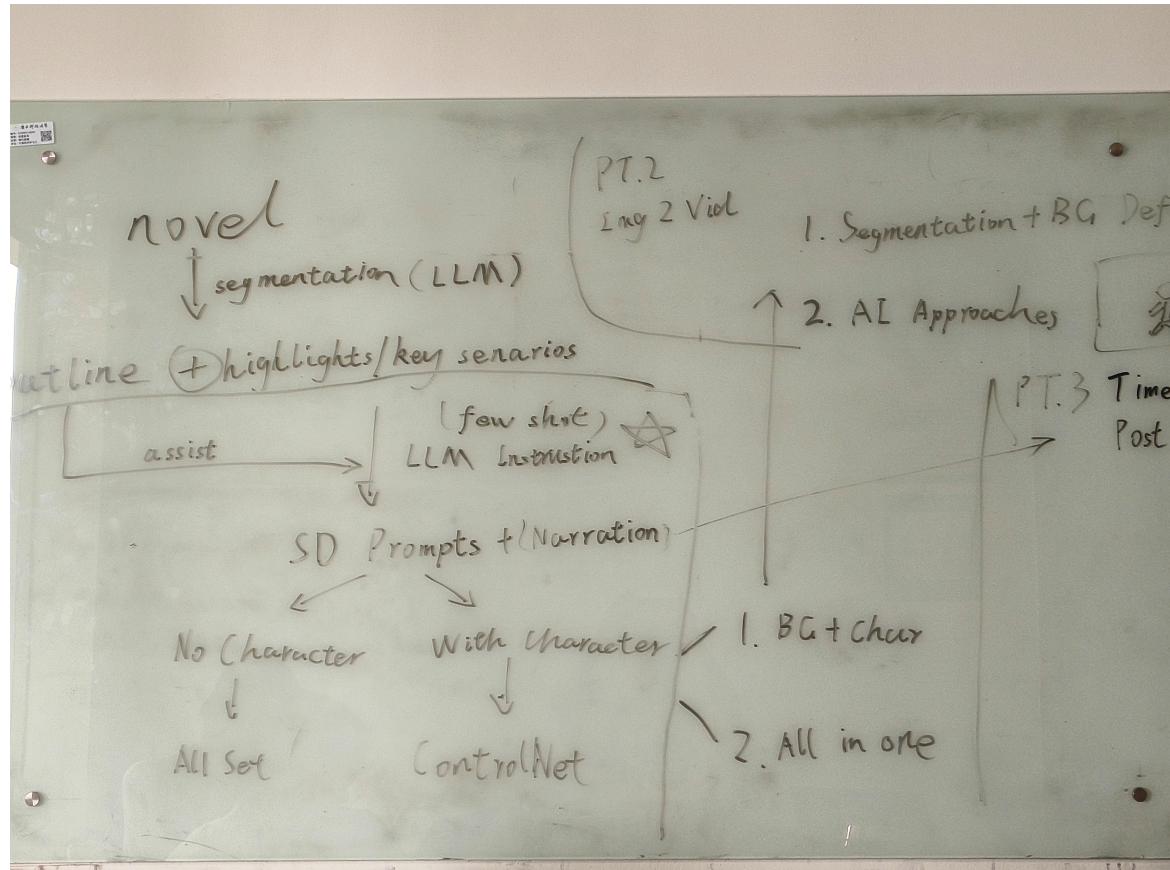
随后借助Stable Diffusion（SD）及ControlNet等图像生成技术，自动生成符合剧情背景和角色设定的高清动画画面。通过AI语音合成技术，使用MARS5为角色配音并生成旁白，实现自然流畅的语言交互。

最后结合时间线和动画合成技术，将图像与语音动态融合，生成完整的动漫化短视频。

该系统通过**全流程AIGC**技术降低了内容制作成本，大幅提升了生成效率。

项目不仅契合当前“短视频+爽文”赛道的商业化趋势，具备显著的变现潜力，同时符合多模态、多媒体课程设计要求，探索文本、图像、语音、视频等多模态AI生成内容（AIGC）的跨领域融合应用。

System Design



可能用到的多媒体技术

1. Python爬虫和文本分割

2. Prompt Engineering and Few Shots Design

3. 基于AI和Prompt生成图片

- DALLE 3 / Stable Diffusion / MidJourney：生成动漫风格或写实风格图片。
- ControlNet (Stable Diffusion 插件)：控制人物动作，使不同帧保持一致。
- ComfyUI 或 深度学习模型 LoRA：微调模型风格，使得角色形象稳定。
- 使用 Inpainting 技术进行局部重绘

4. 基于AI和Narration生成语音

- MARS 5

5. 2D Image to 3D Motion

- Deploy and Call Leiapix AI API

每个人的分工

对照上课拍的照片写下名字、分工



成员	分工
刘圣鼎	Novel Segmentation、SD Prompt Design、 2D Image to 3D Motion
张展玮	Novel Segmentation、SD 3.5 Large Deployment、SD Prompt Design
陈康睿	Novel Segmentation、SD 3.5 Large Deployment、SD Prompt Design
谢嘉楠	Novel Segmentation、SD Prompt Design、 MARS 5 Deployment
王然	Novel Segmentation、SD Prompt Design



CS 330 MIP – Lab 05

多媒体信息处理介绍实验课 - 如何做项目4

Multimedia Information Processing Introduction - Projects

Jimmy Liu 刘江

2025-03-19