



## CS 330 MIP – Lab 03

# 多媒体信息处理介绍实验课- 如何做项目2

Multimedia Information Processing Introduction - Projects

Jimmy Liu 刘江

2025-03-05



# Team 1:基于MOOC的个性化教育智能体 1

# Homework 02

1

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

# Homework 02

2

目标：训练一个专精智能体，回答客户问题，给出问题相关的视频时间节点

1. 模型训练（陆博）：根据文本库对模型进行训练，使模型能够根据课程信息针对问题进行分析、回答

相关技术：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）

# Homework 02

2

目标：训练一个专精智能体，回答客户问题，给出问题相关的视频时间节点

2. 问题转义（黄少霖、王谦益）：将用户问题与文本库匹配，把问题转化成已有文本内容

相关技术：知识图谱技术、自然语言处理（文本匹配）

# Homework 02

2

目标：训练一个专精智能体，回答客户问题，给出问题相关的视频时间节点

3. 视频匹配（林枫茗）：根据已有文本内容与视频进行匹配，获得相对应的时间范围，返回给客户

相关技术：知识图谱技术、视频分析技术（视频分割与标注、关键帧提取）



## Team 2:基于MOOC的个性化教育智能体 2

# Homework 02

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





# Homework 02

2

请说明你们项目的初步设计思路和计划采用的多媒体技术

**项目设计：**学生提问、系统分析问题、检索并定位知识点、给出解答

**实现思路：**使用大模型技术，利用阿里云百炼平台调用开源大模型Qwen；  
获取Mooc教学资源，提取出课件、视频字幕等资源让大模型学习；  
调试大模型的prompt等部署参数，不断优化效果。



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



这是一行文字  
这是一行文字  
这是一行文字  
这是一行文字  
这是一行文字  
这是一行文字  
这是一行文字  
这是一行文字

这是一行文字  
这是一行文字  
这是一行文字  
这是一行文字  
这是一行文字  
这是一行文字  
这是一行文字  
这是一行文字

# 低视力视觉增强辅助： HoloLens MR 头显技术的创新应用

## Student Information:

谢景涛 12112010

金邦景 12110406

高胜寒 22510011

张阳 12110424

马鑫 12111644

助教学姐：席睿翎



# 我们的团队





# 需求分析与目标群体

- **目标群体：** 晚期视网膜色素变性患者
- **日常困难：**
  - 视野缩小，难以看清周围环境
  - 视力下降，阅读文字困难
  - 亮度感知降低，难以适应不同光照条件
- **需求：** 提高文字和场景的可读性和可见性

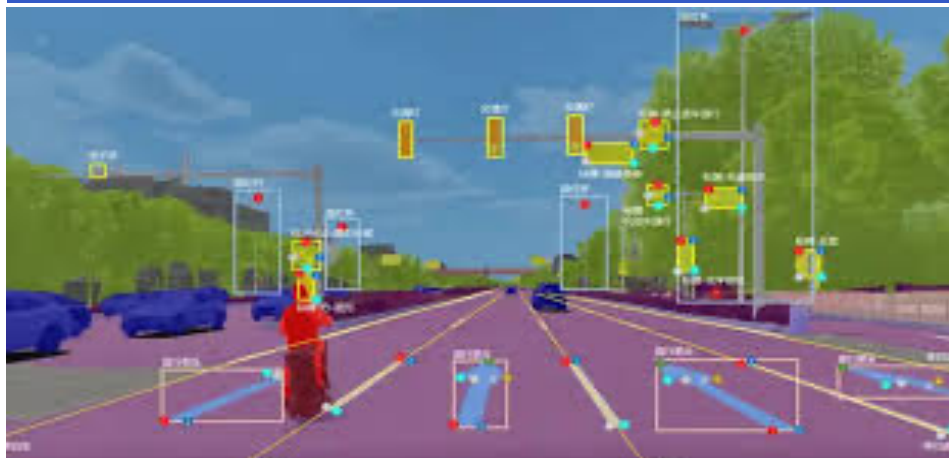
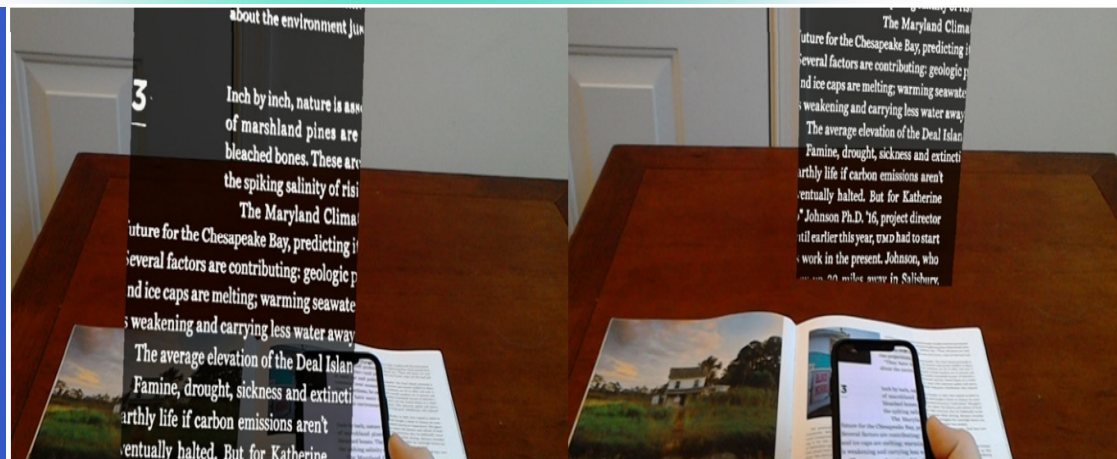


# 初步设计思路

## 1. 功能模块设计

### 文字增强模块

自动检测并转换文字背景和字体颜色  
(如白底黑字转黑底白字)  
提高文字可读性。



### 场景增强模块

通过深度感知技术增强环境对比度  
标注障碍物和重要物体  
帮助用户更好地感知周围环境。



# 初步设计思路

## 2. 用户界面设计

**简洁直观：**设计简洁的用户界面，支持手势和语音交互，方便低视力用户操作。

**个性化设置：**允许用户根据自己的视觉需求调整对比度、字体大小等参数



# 计划采用的多媒体技术



## HoloLens硬件与API

利用HoloLens的深度摄像头和显示系统实现文字和场景的实时增强。



## Unity开发环境:

使用Unity XR Interaction Toolkit实现交互逻辑。

利用Unity的图像处理功能增强文字对比度和环境亮度。



## MRTK (Mixed Reality Toolkit)

使用MRTK的UX组件（如Bounds Control和Data Binding）构建交互界面。

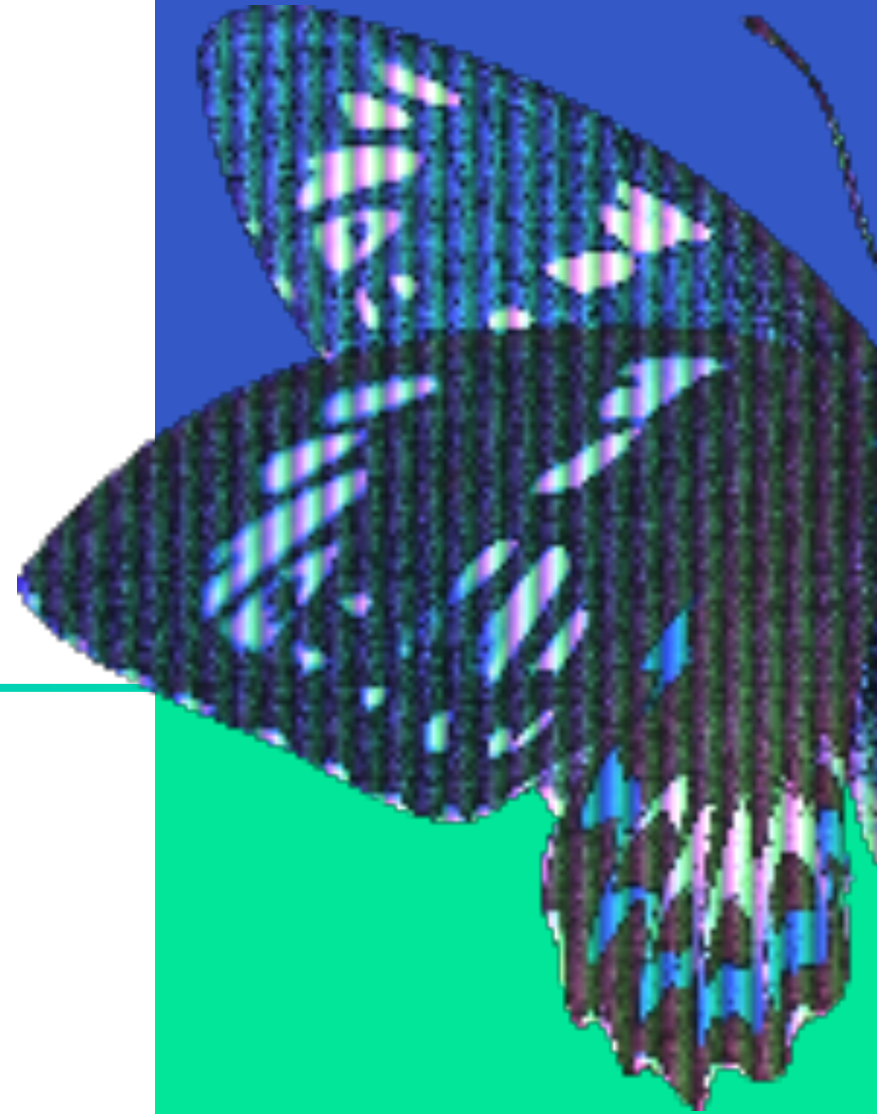
利用MRTK的Input Simulator测试手势和语音交互功能。







谢谢聆听





# 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

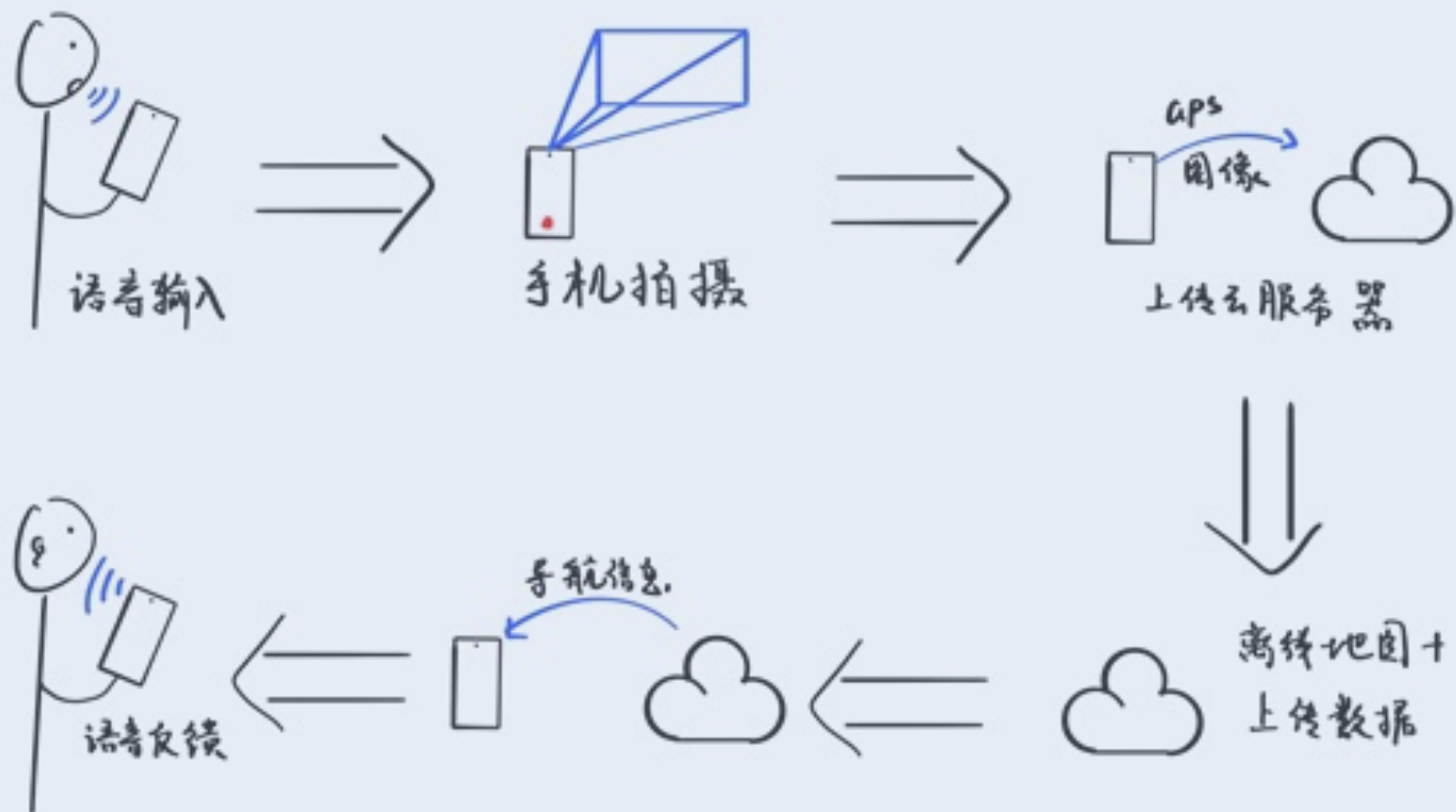


组员照片：（黑衣闭眼的是我。。。李镓璇）

## 2.

- 项目名称：低视力人群室内导航系统
- 项目介绍：在室外应用场景下，基于GPS或北斗导航系统的定位和导航技术（如百度地图、高德地图等）已经相当成熟，并且在我们的日常生活中变得不可或缺，然而，在室内场景，如大型商场、学校建筑设施等，由于室内地图的不精确性和室内定位的困难性等等因素，目前还没有一个好用的室内导航系统，对于视力正常的人群，他们可以轻松通过根据各种指引或询问有关人员来获取到达目的地的路径，然而对于低视力人群，以上操作显得困难得多，因此需要一个可以将目的地语音输入（如洗手间怎么走）经过处理转化为语音导航提示（往前走30m，右拐）的可以运行在智能手机上的导航系统。
- 多媒体技术：
  - GPS信息读取和处理：需要调用智能手机的GPS定位数据
  - 全局地图信息处理：结合手机GPS定位和拉取到的地图信息，给出一个用户的粗定位
  - 光学字符识别：手机摄像头采集用户周边图像，并提取文字信息（如商店招牌、指示牌等），结合地图上已有的地点标签进行匹配，给出用户的精确定位
  - 路径规划和导航：运用各种导航算法，找出用户所在位置和目的地之间的最优路径
  - 语音处理和生成：需要将用户的语音输入转换成文字并提取其中的目的地信息，同时将计算得到的导航路径转化成用户可以接收的形式并生成语音指令，播放给用户
  - 智能手机应用程序开发：如Android、iOS、harmonyOS等手机操作系统的应用开发，手机应用客户端和远程服务端的通信和同步

3.





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



# Homework 02

**Leader:** Shengding Liu

**Group Member:** Zhanwei Zhang, Jianan Xie, Kangrui Chen, Ran Wang

TA: Lingxi Zeng, Jiaqi Wei

Our project topic is **BADAPPLE -Based on multimoDality: A novel visual interpretation PiPeLinE**

**Group Photo:**



# Homework 02

2

## 请说明你们项目的初步设计思路和计划采用的多媒体技术

### Goal:

本项目旨在将小说文本自动转换为动态的短视频动画，通过 AI 生成的 图片、语音、动画 等多媒体技术，使小说内容更易被接受，以适应碎片化阅读的趋势，并探索AI 生成内容（AIGC）在短视频与动漫领域的商业价值

### Motivation:

- 碎片化阅读，没有时间和精力去读书，短视频之风盛行，所以利用短视频的形式采用一种喜闻乐见的形式去让大家去阅读
- AI生成爽文动漫盈利，很有钱景
- 与课程目标适配，text,image,voice,video真正做到多媒体多模态

### Pipeline:

项目主要包括 文本处理、图像生成、语音合成、动画合成 四个核心模块,最终生成完整的动漫化短视频。



# Homework 02

## (1) 文本处理与分解

- 分割小说文本
- 关键元素提取
- 文本优化:转换成适合 AI 生成图像和语音的 Prompt
- 输入小说片段: "王朝的公主在黑夜中逃亡, 身后燃烧的宫殿照亮了她的脸庞。"
- 优化后 Prompt: "一位身穿华丽金色长裙的公主, 在夜晚快速奔跑, 背景是一座燃烧的东方宫殿, 火光映照着她惊恐的脸庞。"

## (2) 基于AI和Prompt生成图片

- 模型选择
- DALL·E 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。
- 设定画面风格 (二次元、油画、科幻等)。
- 使用 AI 生成静态场景 (如宫殿、森林、城市夜景)。
- 生成角色立绘, 并利用姿势控制 (Pose Control) 确保角色动作合理。
- 使用 Inpainting 技术进行微调, 使人物在不同镜头中的形象保持一致。

# Homework 02

## (3)基于AI和文本生成语音

-技术选择

- Edge-TTS / Coqui-TTS / ElevenLabs: 生成高质量角色配音。
- 音色匹配: 为不同角色选择不同的语音风格 (如冷酷男主、活泼女主)。
- 背景音效: 使用 AIVA / ChatGPT 生成音效提示词, 合成背景音乐和环境音。

-角色台词输入语音合成模型, 生成符合角色情绪的音频。

-对语音进行 情感增强处理 (如 Whisper + Prosody Tuning), 使其更生动。

-添加背景音乐、环境音, 增强沉浸感 (如风声、战斗音效)。

## (4) 合成动画

-生成动画序列: 利用 EbSynth 让 AI 生成的帧间图片平滑过渡。

-补充动态效果: 使用 Stable Diffusion Deforum 让场景更加生动, 如烟雾、火焰、光影变化。

-同步语音: 调整动画节奏, 使画面与语音配合得更自然。

-后期编辑: 加入特效、背景音乐、字幕等, 生成完整短视频。

使用的多媒体技术包括但不限于 文本分割、NLP、图像生成、语音合成、动画合成、视频编辑等



## CS 330 MIP – Lab 03

# 多媒体信息处理介绍实验课- 如何做项目2

Multimedia Information Processing Introduction - Projects

Jimmy Liu 刘江

2025-03-05