

项目 7：计算机视觉技术基础

实训目标

能够利用多模态大模型完成图像描述、目标计数、文字识别与图像逻辑推理实训，对比通用大模型与专业计算机视觉系统的能力差异，认识图像生物信息采集存在的隐私风险，提高辨别 AI 伪造图像的能力。

实训环境

Windows 10 及以上系统，可联网计算机，Edge/Chrome 浏览器，本地预留存储空间用于保存各类实验图片素材，AI 平台网页端登录。清理电脑桌面空余位置，用于分类保存图 A、图 B、图 C 三类实验素材，方便实操时快速调取上传。打开浏览器，搜索并下载以下三类图片，保存到桌面。

图 A（场景）： 搜索“繁忙的十字路口”或“琳琅满目的超市”， 要求： 物体多且杂。

图 B（文字）： 搜索“中文菜单手写”或“医生处方”。要求： 字迹略显潦草。

图 C（逻辑）： 搜索“摔倒的老人”或“冒烟的厨房”。要求： 包含一种突发事件或状态。

操作步骤

任务一：基础感知

1. 登录 AI 平台，单击输入框内的“上传图片”按钮，选择图 A。

输入提示词 1：请简要描述这张图片的内容，包括主要物体、环境特征及天气



请简要描述这张图片的内容，包括主要物体、环境特征及天气状况。

这是一张高空俯拍的城市十字路口照片：

- **主要物体**：路口及周边道路上挤满了各色轿车与 SUV，形成严重拥堵；人行道和广场上有大量行人，部分在行走，部分聚集在花形地砖区域。
- **环境特征**：这是一个多方向交汇的路口，路面为深灰色沥青，斑马线清晰；左侧有蓝白相间的多层建筑，路边有行道树，人行道铺有灰色地砖，局部有红色花形装饰。
- **天气状况**：地面湿润有反光，光线柔和，推测为**阴天或雨后**，空气湿润。

图片解析

观察与分析 AI 是否识别出了主要物体（车、人），是否识别出了环境特征（天气是晴是雨？是白天是晚上？）。这背后是多模态大模型结合了视觉特征提取与语言生成技术。

任务二：检测计数

测试 AI 的计数能力，继续使用图 A。

输入提示词 2：请帮我数一下图片里一共有多少辆汽车，请按颜色分类并说明它们的大致位置。

请帮我数一下图片里一共有多少辆汽车，请按颜色分类并说明它们的大致位置。

我来帮你按颜色和位置统计一下图中的汽车数量：

1. 白色汽车（共 24 辆）

- **左上区域：** <point>56 61</point> <point>162 18</point> <point>159 180</point> <point>236 202</point> <point>226 86</point>
- **路口及周边：** <point>366 232</point> <point>397 356</point> <point>402 38</point> <point>406 696</point> <point>516 622</point> <point>539 262</point> <point>539 402</point> <point>582 466</point> <point>582 646</point> <point>706 242</point> <point>702 436</point> <point>702 502</point> <point>746 716</point> <point>856 816</point> <point>862 902</point> <point>876 56</point> <point>916 802</point> <point>956 32</point> <point>896 102</point>

2. 黑色汽车（共 10 辆）

- **路口及周边：** <point>322 276</point> <point>436 472</point> <point>526 352</point> <point>592 262</point> <point>652 332</point> <point>666 542</point> <point>742 616</point> <point>772

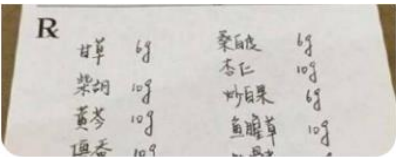
车辆计数

观察 AI 是精确计数还是模糊估算。目前的通用大模型在密集物体计数上常有误差，需对比专用检测模型之间的差异，观察是否会出现幻觉，即出现编造现象。如果 AI 计数有误，思考是物体遮挡还是物体尺寸问题，这是 CV 领域的经典难题。

任务三：OCR 文字识别

体验光学字符识别（OCR）技术，上传图 B。

输入提示词 3：请识别图片中的药方文字，将其整理成表格形式。



请识别图片中的药方文字，将其整理成表格形式。

中药处方整理表

表格		📄	📥	📏
药材名称	用量 (g)			
甘草	6			
柴胡	10			
黄芩	10			
藿香	10			

光学字符识别

观察测试 AI 对连笔字、潦草字的辨识能力。是否能将非结构化的手写图转化为可编辑的 Excel/Markdown 表格。该项技术已被广泛用于快递单识别、医疗数字化、作业自动批改等领域。

任务四：视觉推理

测试 AI 的视觉推理能力，上传图 C。

输入提示词 4：图片中发生了什么紧急情况？如果你是智能监控系统的 AI 中枢，你会发出什么级别的警报？请给出后续处理建议。



图片中发生了什么紧急情况？如果你是智能监控系统的AI中枢，你会发出什么级别的警报？请给出后续处理建议。

一、图片中的紧急情况

画面显示一名老人倒在机动车道上，身旁是侧翻的自行车，车篮内物品散落，老人处于倒地状态，疑似发生了骑行摔倒或交通事故，存在明显的人身安全风险。

二、警报级别判定

作为智能监控系统 AI 中枢，应触发一级（最高级）紧急警报，理由如下：

- 事件涉及老年群体，属于高风险人群，倒地后可能引发颅脑损伤、骨折、失温等严重后果。
- 事发地点为机动车道，存在二次碾压的重大安全隐患。
- 场景中无明显施救人员，需要快速介入。

三、后续处理建议

- 立即联动：向辖区交警、120 急救中心、属地派出所同步推送警报及精准位置，优先调度急救与交通管控资源。



视觉推理

AI 模型不能仅停留在目标检测层面，如识别出人和地面，必须具备因果推断能力。能从人体倒地姿态这一视觉特征，逻辑推导出可能存在突发疾病或摔伤的语义结论，并据此触发医疗急救类建议，而非仅仅描述物理位置。

注意事项

通用大模型目标计数容易出现误差，专业场景需要专用视觉检测模型；手写文字识别结果必须人工校对，禁止私自采集他人人脸图像开展实验。