2025计算机视觉基础研究生结课论文要求

结课论文：闭环式多模态视觉系统与可靠性评测（个人2 周左右完成）

形式：单人独立完成；提交论文 + 代码github链接

一、硬性要求

A. 模型覆盖（至少 3 类）

必须至少使用并跑通以下三类中的 **3 个代表**（可多但不必）：

- 视觉基础模型：DINO（特征/检索/密集表征）
- 可提示分割：SAM / SAM2
- 多模态基础：CLIP
- 多模态大模型：Qwen3-VL（或同等级 VLM，需说明版本与推理方式）

> 要求不是“都跑一下截图”，而是都要进入你的系统里承担明确功能。

B. 必须是“闭环系统”（重点）

你的方法必须包含至少 **两步以上的闭环**，并且能用实验表明闭环带来收益。
典型闭环例子（任选一种实现即可）：

- **VLM→定位/指代解析→SAM 分割→CLIP/DINO 验证→回喂给 VLM 纠错/细化→再分割**
- **CLIP 候选检索→SAM 生成多 mask→VLM 选择/解释→失败则自动换候选/换提示→直到满足约束**

闭环要能统计：每个样本平均迭代次数、失败率随迭代变化、什么时候提前停止。

C. 必须有“可靠性评测”

除主指标外，至少完成下面三项中的 **两项**：

1. **OOD 泛化**：换一个域/风格的数据（或自建小集）测性能下降
2. **扰动鲁棒性**：尺度变化、遮挡、噪声/模糊、压缩、光照（任选 2 种扰动强度）
3. **长尾/小目标分析**：按目标面积分桶（small/medium/large）或按类别频次分桶统计

D. 必须有“效率/成本分析”

至少报告并对比：

- 推理时间（每图/每问）

- 峰值显存
- 工具调用次数（SAM 调用次数、VLM 调用次数、CLIP 计算次数等）
  并给出 **至少 1 个加速策略**（缓存特征、mask 复用、batch、量化、减少候选数并保持精度等）及其对性能的影响。

二、选题主线（任选 1 条）

主线 1：闭环式指代分割 + 自检纠错（Grounded Ref-Seg Loop）

输入：图像 + 指代表达（或自然语言问题）
输出：目标 mask（可同时输出 bbox）

基线（必须实现）：

- SAM 生成候选 masks（多点/网格/box 均可）
- CLIP 对 “mask 区域—文本” 相似度排序 → Top-1 mask

闭环增强（必须）：

- VLM 读取：候选 Top-k 的可视化/统计（面积、位置、相似度）
- VLM 输出“纠错策略”：换描述、加方位约束、拆分指代、排除干扰物等
- 再调用 SAM/CLIP 进行二次选择，直到停止条件满足

你必须提出 1 个“研究性改进点”（可参考下列思路）：

- **可解释的停止条件**：置信度阈值 + 代价约束（少迭代但不降太多精度）
- **候选生成策略学习**：让候选数量/点密度随文本类型自适应（“小的/远处的/左上角”等）
- **验证器改进**：CLIP 相似度 + 结构先验（位置、面积、连通性）融合成打分函数（可用简单线性/逻辑回归，不需要大训练）
- **反事实排除**：VLM 生成“不是目标”的负提示，帮助 CLIP 拉开差距（要做消融证明）

指标：mIoU / IoU@0.5 成功率 + 平均迭代次数 + 每图耗时

主线 2：VLM 增强 VQA 的“可验证推理”（VQA with Visual Verification）

输入：图像 + 问题
输出：答案 + 证据（证据必须是可视化区域/标注）

基线（必须）：Qwen3-VL 直接答（zero-shot 或 few-shot）
闭环增强（必须）：

- VLM 先提出“需要看的证据是什么/在哪里”
- 调用 SAM 得到证据区域（或多个候选）→ 裁剪回喂 → 再答
- 必须引入一个“验证步骤”：答案必须能引用证据（例如 OCR/计数/颜色/相对位置）

你必须提出 1 个“研究性改进点”（可参考下列思路）：

- **自一致验证**：同一问题多种提示模板 + 证据一致性检查
- **证据优先**：先定位再回答 vs 说明理由后定位（比较哪种更稳）
- **失败类型归因**：读图失败/定位失败/推理失败/幻觉（必须分类统计）

指标：Accuracy/EM + 证据命中率（证据区域是否覆盖关键目标，IoU 或人工抽查）+ 成本（工具调用）

主线 3：基于基础模型的“弱监督伪标注→轻量训练→泛化评测”

这条更偏“CV 研究味”，但两周也能做（训练量控制在小头/adapter）。

任务例子：开放词汇分割 / 特定类别分割 / 简单深度头
流程（必须）：

1. 用 CLIP/VLM 生成或筛选伪标签（文本→候选区域→SAM mask）
2. 构建一个小训练集（几百到几千张）
3. 只训练轻量模块（decoder/head/LoRA）
4. 做 OOD 或长尾评测，证明“伪标注策略”确实影响泛化

你必须提出 1 个“研究性改进点”（可参考下列思路）：

- 伪标注置信度筛选策略（阈值/分桶/课程学习）
- 多来源一致性（CLIP 与 VLM 同意才采纳）
- 噪声鲁棒训练（对伪标签做边界松弛/一致性正则）

指标：主任务指标 + 伪标签质量估计（抽样人工核验 50 张也行）+ 训练成本

三、实验与论文必须包含的“加深要求”

1. 消融必须 ⩾ 4 个（比平时作业多）

至少包括：

- 无闭环（one-shot） vs 有闭环（loop）
- 无验证器（只看 CLIP） vs 有验证器（CLIP+先验/或 VLM 自检）

- 迭代上限 1/2/3 的 trade-off（精度-成本曲线）
- OOD 或扰动条件下：one-shot 与 loop 的差距

2. 必须给出置信区间或重复实验（二选一）

- 对关键指标做置信区间；或
- 固定数据划分下重复 3 次不同随机种子并报告均值±方差

3. 必须给出失败案例“分类统计表”

例：文本歧义、小目标、遮挡、背景干扰、计数、阅读文字等，并给每类至少 3 个图例。

四、提交物与论文结构（统一规范）

提交物

- 论文 PDF：8–12 页（比之前稍长一点，允许附录）
- 代码包：github链接；

论文必须包含章节

摘要、引言、相关工作（≥10 篇引用）、方法（含系统流程图/算法框）、实验设置、结果、消融、鲁棒性评测、效率评测、误差分析、局限与伦理、可复现清单。

五、评分

- 系统设计与闭环完成度：25
- 实验严谨性（消融/统计/可复现）：30
- 可靠性评测（OOD/扰动/长尾）：20
- 效率与成本分析：10
- 论文表达与规范：15
  加分（最多 +10）：工具闭环更复杂但稳（例如自动策略选择）、或做了强 baseline 对照。