# 计算机视觉2025年秋季期末项目

## 最终项目

“ 最终项目 ”是您将课程中学到的概念和技术应用于自己感兴趣的计算机视觉问题的宝贵机会。我们强烈建议您以团队形式工作，每个团队最多可由四名成员组成。

您可以自由选择以下列出的建议项目主题。第1至3个主题侧重于经典计算机视觉任务，而第4至8个主题则探索更先进和前沿的研究方向。为了鼓励探索前沿领域，基于第4至8个主题的项目将有资格获得10个额外加分，这意味着项目的最高分数将是110分，而不是标准的100分。

我们建议选择一个不仅挑战你的技术技能，而且符合你的兴趣和未来目标的主题。

## 项目主题

### 1. 人脸识别

面部识别系统是一种能够将数字图像或视频帧中的人脸与人脸数据库进行匹配的技术。这种系统通常用于通过身份验证服务来认证用户，其工作原理是通过精确定位并测量给定图像中的面部特征。

在这个项目中，你的任务是使用以下数据集和github仓库构建一个简单的面部识别系统。你可以仅使用数据集中提供的图像作为测试和训练数据，但如果你能够输入自己的（少量的）图像用于面部识别和鉴定，我们将考虑给予额外加分。

- 数据集：VGG Face2
- Github：https://github.com/serengil/deepface

### 2. 单对象跟踪

单目标跟踪系统是一种技术，能够将视频第一帧中由人指定的目标与后续的视频序列进行匹配。这种系统通常用于安全监控，通过计算参考帧目标与搜索帧目标之间的相似度来捕捉目标。

任务：自己收集多段视频，使用标记工具在第一帧中标记需要跟踪的对象，设计跟踪算法进行跟踪并可视化跟踪结果。

- 调查：
  - https://blog.csdn.net/qq_37002417/article/details/108141409
  - https://zhuanlan.zhihu.com/p/503735985
- 数据集：http://got-10k.aitestunion.com/
- Github：

- https://github.com/visionml/pytracking
- https://github.com/heartexlabs/labelImg

## 3. 语义分割

语义分割是典型的计算机视觉问题，它涉及将一些原始数据作为输入，并将其转换为具有突出感兴趣区域的掩码，其应用包括场景理解、医学图像分析、机器人感知、视频监视、增强现实和图像压缩等。

任务：训练一个语义分割模型（不限于PASCAL VOC数据集），并自行收集一些有趣的场景，看看分割效果如何。同时，我们鼓励你尝试将模型应用于视频图像分割，看看会遇到哪些困难。

- 数据集：http://host.robots.ox.ac.uk/pascal/VOC/voc2012/index.html#devkit
- Github：https://github.com/usuyama/pytorch-unet（这是一个相对简单的方法，你可以自由选择你喜欢的方法）

---

## 4. 三维高斯喷溅中的分割⭐

由于其高效且可微分的渲染能力，3D高斯喷溅（3DGS）在3D重建、新视图合成及相关领域获得了广泛关注。值得注意的是，3DGS中的ObjectLevel分割在各种下游应用中发挥着关键作用，包括场景编辑、场景理解和体现智能。

### 项目任务

本项目由两个主要部分组成：

1. **方法复制**
   - 对象分割：对3DGS场景进行对象级分割。
   - 下游应用程序：基于分割对象，探索并实现若干可能的任务，包括：
     - 三维对象删除
     - 三维对象修复
     - 三维对象样式转移
     - 三维多对象编辑
     - 或者你提出的其他有创意和有意义的应用
   - **数据集要求：在提供的数据集中至少对三个不同的场景进行实验。**
2. **您的贡献**
   - 自定义数据验证：捕获并使用您自己的真实世界数据来测试相同的管道。

- 方法增强：对分割过程或下游应用程序进行任何改进或修改。这可以简单地是针对特定场景的参数调整——欢迎提出新想法，但不是必需的。

注：本项目重点在于尝试和验证过程，而不是达到最佳结果。
**追索权：**

- 数据集：
  - LERF-面罩：https://github.com/lkeab/gaussian-grouping/blob/main/docs/dataset.md
  - Mip-NeRF360: https://jonbarron.info/mipnerf360/
- GitHub：https://github.com/lkeab/gaussian-grouping（免费选择其他方法。）
- 3DGS查看器：https://superspl.at/editor

## 5. 图像字幕⭐

图像描述是为图像生成自然语言描述的过程。它涉及理解图像的内容，并将这种理解转化为文字。这项任务结合了
计算机视觉和自然语言处理技术。

### 项目任务

在这个项目中，你将构建一个能够为图像生成相关且准确的标题的模型。你必须使用包含图像及其相应标题的COCO数据集。目标是训练一个视觉语言模型，以生成能够准确描述图像内容的标题。

本项目由两个主要部分组成：

1. **构建和培训VLM**
   - 从预训练的视觉编码器（ResNet、ViT等）和预训练的大语言模型（LLaMA、Qwen2/Qwen2.5等）开始，使用连接器（默认为MLP）形成一个视觉-语言模型。
   - 准备数据集管道，并在COCO标题数据集上训练视觉语言模型。
     建议使用huggingface的transformer包来构建项目，更多细节可以参考llava的实现风格。
   - 在COCO测试集上评估训练好的模型，报告BLEU分数和Cider分数。可以使用pycocoevalcap包计算这两个指标。
2. **消融研究**
   - 探索不同视觉编码器架构（例如CNN、基于变压器、基于mamba）对最终性能的影响，并报告您的发现。
   - 探索不同连接器对最终性能的影响，可以简单地替换MLP，也可以替换其他模块，如Q-former或其他你想要的模块。如果你设计了新的模块，请解释你的设计原理。

- （可选，加分项）探索基于变压器解码器的语言模型和基于Mamba的语言模型对最终性能的影响。此外，比较它们在不同输入序列长度下的时间消耗，绘制时间-序列长度曲线，并探讨可能的原因。

注：本项目重点在于尝试和验证过程，而不是达到最佳结果。

提示：如果计算资源有限，可以先从较小的视觉模型和语言模型开始，例如GPT-2、Mamba-130M、Qwen3-0.6B等。

**资源：**

- 对于COCO标题注释，你可以在HuggingFace上找到，也可以直接下载预处理的COCO2014标题注释文件。注意，你仍然需要下载图像文件。
- LLaVA风格的实现：https://github.com/haotian-liu/LLaVA
- 变压器教程：https://huggingface.co/docs/transformers/index

# 6. 人到人形运动重定向⭐

人到类人机器人运动重定向旨在将3D人类运动序列（从amass等数据集中捕获）转移到类人机器人(例如。Unitree H1)具有不同的运动结构，使机器人能够在保持物理稳定的同时模仿自然的人类动作。这项任务需要解决诸如关节自由度不匹配、动态平衡和实时控制等挑战。

## 项目任务

设计一种优化算法，将基于SMPL的人体运动参数映射到人形机器人关节控制信号。使用人形机器人在Isaac Gym平台验证生成的运动。

本项目需要按照以下步骤完成：

1. 数据获取：下载AMASS数据集，了解SMPL数据和刚体运动变换的相关知识
2. 算法设计：设计一个人体运动数据到人形机器人数据的映射算法，可以参考MimickingBench这篇文章了解重定向方法
3. 可视化显示：在Isaac Gym平台上使用Unitree H1或其他类人机器人重新定向后显示数据
4. 实验评价：设计相应的指标来评价算法的准确度等

- 数据集：
  - AMASS：https://amass.is.tue.mpg.de包含SMPL参数的大规模人体运动捕捉数据集。
- Github：
  - UH-1: https://github.com/sihengz02/UH-1
  - InterMimic:https://github.com/Sirui-Xu/InterMimic

- 艾萨克Gym：https://developer.nvidia.com/isaac-gym

# 7. R1推理⭐

DeepSeek-R1在解决复杂问题时展现了令人印象深刻的推理能力，激发了基于大语言模型推理领域的研究兴趣。在这个项目中，你将探索R1风格的推理——在倒计时任务中，给定一个目标数字和一组N个数字，目标是生成一个有效的方程，使用提供的数字和运算达到目标。
注：建议至少访问一个具有≥40GB内存的GPU。

## 项目任务

本项目由两个主要部分组成：

**1. 方法复制**

- 倒计时推理：重现解决倒计时任务的训练流程。你的实现应该能够使模型逐步推理，生成有效的方程来解决目标。
- 评估与比较：将复制模型的性能与基线变体（例如，没有推理特定训练的预训练模型）进行比较。

**2. 您的贡献**

- 结果分析：分析培训后对倒计时任务模型性能的影响。这包括：
  - 定量分析（例如，准确度、奖励指标）
  - 定性分析（例如，生成的输出、错误模式）

    由于主要目标是探索推理能力而非最大化准确率，因此您对学习趋势和失败模式的见解尤其有价值。如果性能没有提升，请讨论可能的原因，例如数据稀缺或训练不稳定。我们鼓励您从训练日志中可视化趋势（例如，随时间变化的奖励进度）。
- **自定义任务扩展：**
  - 自定义任务：可选择探索模型在超出倒计时任务范围内的推理能力。您可以使用简单的符号推理任务、算术思维链或其它结构化逻辑任务。
  - 多模态推理：如果时间充裕，计算资源充足，可以考虑将推理扩展到多模态任务。例如，Geometry3K就涉及对图表和问题陈述的解释。

**追索权**

- 基本代码和数据集：TinyZero：https://github.com/Jiayi-Pan/TinyZero
- 培训方法参考：GRPO：https://arxiv.org/abs/2402.03300
- 多模式培训代码：https://github.com/hiyouga/EasyR1

# 8. 多视图立体SLAM⭐

多视图立体（MVS）将被动立体的原理扩展到多个视角，旨在从一系列已知相机参数的图像中重建场景的密集3D模型。多视图立体SLAM在此基础上进一步发展，在重建过程中逐步估计相机参数。

任务：在本项目中，你需要运行一个MVS-SLAM流水线，接受单目视频作为输入，并输出3D场景点云。在公开数据集上测试你的结果，并使用SUSTech场景（教室或实验室）构建实时演示。

额外提示：你可以考虑使用NVS方法（如3D高斯散射）来构建更佳的视觉效果。

- 数据集：
  - 7-场景：https://www.microsoft.com/en-us/research/project/rgb-d-dataset-7-scenes/（如果遇到问题，复制下载链接并粘贴到搜索栏中进行下载）。
- Github：https://github.com/rmurai0610/MASt3R-SLAM

注意：我们不会限制您使用的数据集和github存储库，您可以使用您找到的资源来完成任务。

# 组建小组

- 群组最多可包含4个成员。
- 报告应单独提交，并应突出每个团队成员在论文部分的贡献。
- 请在腾讯文档中完成主题选择和团队成员信息：https://docs.qq.com/sheet/DTm5QcEZoWndxVnB3？tab=BB08J2 VFZRXFV？tab=BB08J2。分组截止时间为2025.5.24。

# 报告

我们提供了一个模板来帮助指导您的最终项目报告，但请放心，只要结构清晰且组织良好，您可以自行选择。我们也推荐使用Latex撰写报告。

**关于报告：**

- 每个小组应提交一份报告。
  - 报告应包括所有合作者的姓名。
  - 您可以使用word、markdown或Latex来形成您的报告。
- 应提交PDF文件。

你应该描述并评估你在项目中所做的工作，这可能并不一定是你最初希望完成的内容。一个小成果如果描述和评估得当，会比一个雄心勃勃但没有做好任何方面的工作获得更多的分数。准确地描述你试图解决的问题。详细解释你的方法，并说明你采取了哪些简化或假设。同时展示你的方法的局限性。它在什么情况下不起作用？为什么？如果你继续研究，你会采取哪些步骤？确保添加对您审查或使用过的所有相关工作的参考。

您可以提交任何您认为对评估您的工作很重要的补充材料，但我们不能保证我们会审查所有这些材料，您不应假设这一点。报告应该是自成一体的。

提交：将报告以pdf文件形式提交至黑板，文件名命名为groupid_final.pdf。提交任何补充材料（例如视频）时，请将其作为单个压缩文件提交，文件名命名为groupid_sup.zip。添加一个说明文件，描述补充内容。将代码作为单个压缩文件提交，文件名命名为groupid_code.zip。

# 分级政策

- 报告（60）
  - 导言（10）
  - 相关工作(5)
  - 方法（10）
  - 实验结果（20）
  - 结论（五）
  - 参考文献(5)
  - 报告的整体清晰度(5)
- 各成员的贡献（例如：xxx为35%，xxx为25%）：我们将增加或减少• 演示文稿（40）
  - 每组15 min+3 min Q&A
- 选题奖励（10）：只要选择4到8的题目并且完成项目，你将获得10分奖励。但是，如果作品质量太低，将不会获得奖励。

# 码头交货

- 问：这个主要任务需要算法上的创新吗？
- A：我们非常鼓励学生在算法创新方面下功夫。然而，由于时间限制，这并不是强制性的。你可以将其视为一个应用导向的项目，重点展示你的工程能力和团队合作。当然，如果有学生提出在标准数据集上表现良好的创新算法，我们将额外加分。
- 问：TA将如何确定贡献水平？
- 答：对于在GitHub上开源的团队，我们可以通过提交来大致判断贡献水平。但是，我们主要还是依赖于你们报告中的贡献部分。

# 由于

- 报告将于2025年6月21日结束时提交。• **演讲时间是**

**2025年6月21日下午2点到4点。**