

《模式识别与机器学习》 课程设计 指导书

华中科技大学
人工智能与自动化学院
二〇二六年四月

目录

一、课程设计目的.....	1
二、总体设计要求.....	1
三、课程具体步骤与时间安排.....	1
四、模式识别课程设计及具体要求.....	2
4.1 遮挡条件下遥感图像中的飞机目标检测.....	2
4.2 蜂群目标 2D 轨迹跟踪和姿态估计.....	5
4.3 鲁棒的深度检测网络对抗图像生成.....	6
4.4 面向伪装场景的视觉语言模型微调与理解输出.....	8
4.5 微小目标检测.....	11
4.6 跨域少样本目标识别.....	13
4.7 心脏疾病教师网络辅助的 ECG 单模态诊断.....	17
4.8 无人机视角下的车辆目标检测.....	19
4.9 基于视觉和大语言模型的羽毛球比赛战术分析与预测.....	21
4.10 品种无关植物计数.....	23
4.11 病理图像细胞计数.....	26
4.12 密集人群运动分析.....	28
4.13 密集视频行人流量计数.....	30
4.14 植物图像抠图.....	32
4.15 虚拟试穿风格编辑.....	33
4.16 面向植物表型视觉的异质同构多任务学习.....	35
4.17 面向视觉 ViT 模型的通用特征上采样.....	38
4.18 基于 3D Gaussian Splatting 的植物三维重建与点云分割.....	39
4.19 眨眼检测与视线估计的一体化方法与系统.....	40
4.20 基于 RGB 图像的三维交互手势估计.....	42
4.21 非受限条件下的人体眨眼检测.....	44
4.22 基于人类操作演示数据的灵巧手操作迁移.....	45
4.23 基于文本的手-物体交互动作序列生成.....	46
4.24 基于文言文高密度编码的数学推理方法.....	48
4.25 非受限条件下的跌倒检测.....	49
4.26 抗遮挡目标点跟踪.....	50
4.27 三维场景重建.....	51
4.28 红外序列图像弱小目标检测.....	52
4.29 无人机位姿估计.....	54
4.30 可见光-红外跨模态图像翻译.....	55
4.31 跨视角图像匹配.....	56
4.32 基于视觉基础模型的遥感图像目标检测与分割.....	57
4.33 图像线特征提取与匹配.....	58
4.34 海底目标检测与识别.....	59
4.35 遥感图像中的目标检测.....	61
4.36 密集场景下的行人跟踪.....	63
4.37 基于轻量化网络的麦穗目标检测.....	65

4. 38	物体 6D 姿态估计	67
4. 39	人脸表情识别	69
4. 40	3D 点云物体缺陷检测	70
4. 41	无人机视角下的目标检测与识别	72
4. 42	基于深度学习的鸟声识别	74
4. 43	视频序列中人脸微表情识别	75
4. 44	路边垃圾实时检测	76
4. 45	基于图像的饮食营养评估	78
4. 46	轻量化路面损伤检测	80
4. 47	基于结构信息挖掘的鸟类识别	82
4. 48	基于迁移学习的花卉识别	84
4. 49	花粉识别方法对比分析	85
4. 50	有限样本条件三维场景重建	86
4. 51	三维场景重建-实例分割优化方法	88
4. 52	高效的 3D 高斯重建渲染	90
4. 53	伪装目标检测	92
4. 54	增量学习目标检测	94
4. 55	小样本目标检测	96
4. 56	旋转目标检测	98
4. 57	三维场景渲染与可控编辑-目标检测一体化测试技术	100
4. 58	高分辨率时-空-谱融合计算成像	102
4. 59	低光图像饱和去模糊	103
4. 60	基于语言-视觉模型的图像复原方法	105
4. 61	事件相机高速高动态成像	106
4. 62	远距高清晰成像	108
4. 63	远距透雾成像	110
4. 64	雨雪雾霾恶劣天气图像统一增强方法	111
4. 65	基于事件相机的目标跟踪	114
4. 66	基于 3D 人体姿态估计	116
4. 67	动态场景下的 SLAM 重建	118
4. 68	基于高斯泼溅的三维人体重建	120
4. 69	基于深度学习的时序动作检测	122
4. 70	基于视觉和语言模型的视频燃烧状态描述	124
4. 71	雾天条件下的目标检测	125
4. 72	空对空视角下的红外与可见光融合反无人机识别	127
4. 73	音视频融合的网球击球动作定位与识别	129
4. 74	真实环境海面目标检测	131
4. 75	智能水下感知	133
4. 76	心电可解释性分析	134
4. 77	左心室肥大二分类检测	135
4. 78	肺炎 X 光图像诊断识别	136
4. 79	心音异常识别	138
4. 80	铣削加工刀具状态监测	139
4. 81	复杂曲面加工质量预测	140
4. 82	机载车辆目标重识别	141

4.83 面向无人机巡航的跨视角匹配定位	143
4.84 红外与可见光融合复原方法	145
4.85 典型人脸识别网络物理可实现攻击	146
4.86 排污口自动图像检测	147
4.87 血管造影图像自动诊断检测	149
4.88 堰塞坝表层颗粒物智能检测分析	150
4.89 水果新鲜度监测	152
4.90 自主竞赛选题	152
五、课程设计的考查	153
5.1 考核方式	153
5.2 成绩评定	153
六、课程设计报告撰写要求	154

一、课程设计目的

《模式识别与机器学习》课程设计作为独立的教学环节，是模式识别与智能系统专业集中实践性环节系列之一，是学习完《模式识别》、《模式识别与机器学习》课程后进行的一次全面的综合练习。其目的在于加深对模式识别与机器学习基础理论和基本知识的理解，掌握使用特征分析、特征提取及模式辨识的基本方法，提高解决实际问题、开发图像自动识别系统的实践能力。同时课程设计应充分体现“教师指导下的以学生为中心”的教学模式，以学生为认知主体，充分调动学生的积极性和能动性，重视学生自学能力的培养。

二、总体设计要求

模式识别课程设计由于涉及到的内容较多，工作量大，根据课程本身的特点，特提出如下要求：

1. 课程设计需要在课程内容讲完后给学生布置，让学生结合课堂讲授内容，并在教师的具体指导下，逐步开始分析工作。
2. 学生开展课程设计按 5 人一组进行分组，并确定一个组长，明确组员分工与协调。
3. 各组在分工的前提下制定相应的任务完成计划，并按计划开展课程设计，接受教师检查。
4. 课程设计的题目的选择可根据学生的自身理论学习体会和研究兴趣，结合实际或熟悉的课题，体现“麻雀虽小、五脏俱全”，充分练习模式识别的各个方面的内容。
5. 课程设计必须完成对模式识别与机器学习处理系统的分析与设计任务，编写相应的分析与设计报告。实施部分的内容可根据小组的实际能力决定取舍。

三、课程具体步骤与时间安排

本课程设计历时 10 周，学分 2 分。

具体步骤安排如下：

1. 课题选取，选取 1 个题目，并进行系统调研、资料整理。
2. 收集、整理、学习与设计相关的技术资料。
3. 进行相关选题的特征分析/模式分析，研究并设计实现任务的算法。可以参考复现文献中别人的研究成果，如能提出自己的见解和改进则更好。
4. 算法程序的设计和调试。
5. 按照进度对数据库数据进行实验，认真进行设计相关内容的记录, 试验结果分析。试验结果如不能满足课题要求, 需进行试验设计调整或算法调整, 再次试验, 直至试验结果满足课题要求, 或进行不同算法试验结果比对与分析。
6. 完成课程设计报告。这是出成果的阶段，要认真整理前面各阶段的成果，包括完整的研究过程，实验中的失败与经验教训等，可以一一誊写在设计报告中。报告要求文字通顺、计算准确、画图清晰整洁。注意按统一规定格式和封面，打印、装订成册。工作量大、时间紧，需要付出一定的辛苦。

时间安排及方式(每 5 人一组)

1. 课程设计任务书的布置，讲解（0.5 天）
2. 学生根据任务书的要求初步进行需求分析（第 10 周）
3. 进行方案设计，并撰写设计方案（第 11 周）
4. 指导老师审阅方案设计报告，根据意见修改设计方案（第 12 周）
5. 算法设计与实现（第 13-15 周）
6. 软件实现与算法测试（第 16-17 周）
7. 撰写课程设计报告（第 18 周）
8. 考核答辩（第 19 周）

四、模式识别课程设计及具体要求

4.1 遮挡条件下遥感图像中的飞机目标检测

任务描述：用高斯光斑模拟遥感图像中的飞机目标的部分遮挡场景，构建遮挡场景测试数据集，完成遮挡条件下遥感图像中飞机目标的自动检测，并给出目标中心（以原始无遮挡数据集中目标斜框中心为参考标注）。分析遮挡程度（可自行设计遮挡尺度指标）、

目标尺度等因素对目标检测结果的影响。

自动检测，标记出其具体区域。注意检测结果采用倾斜目标范围框(回归)的方法，如下图所示。

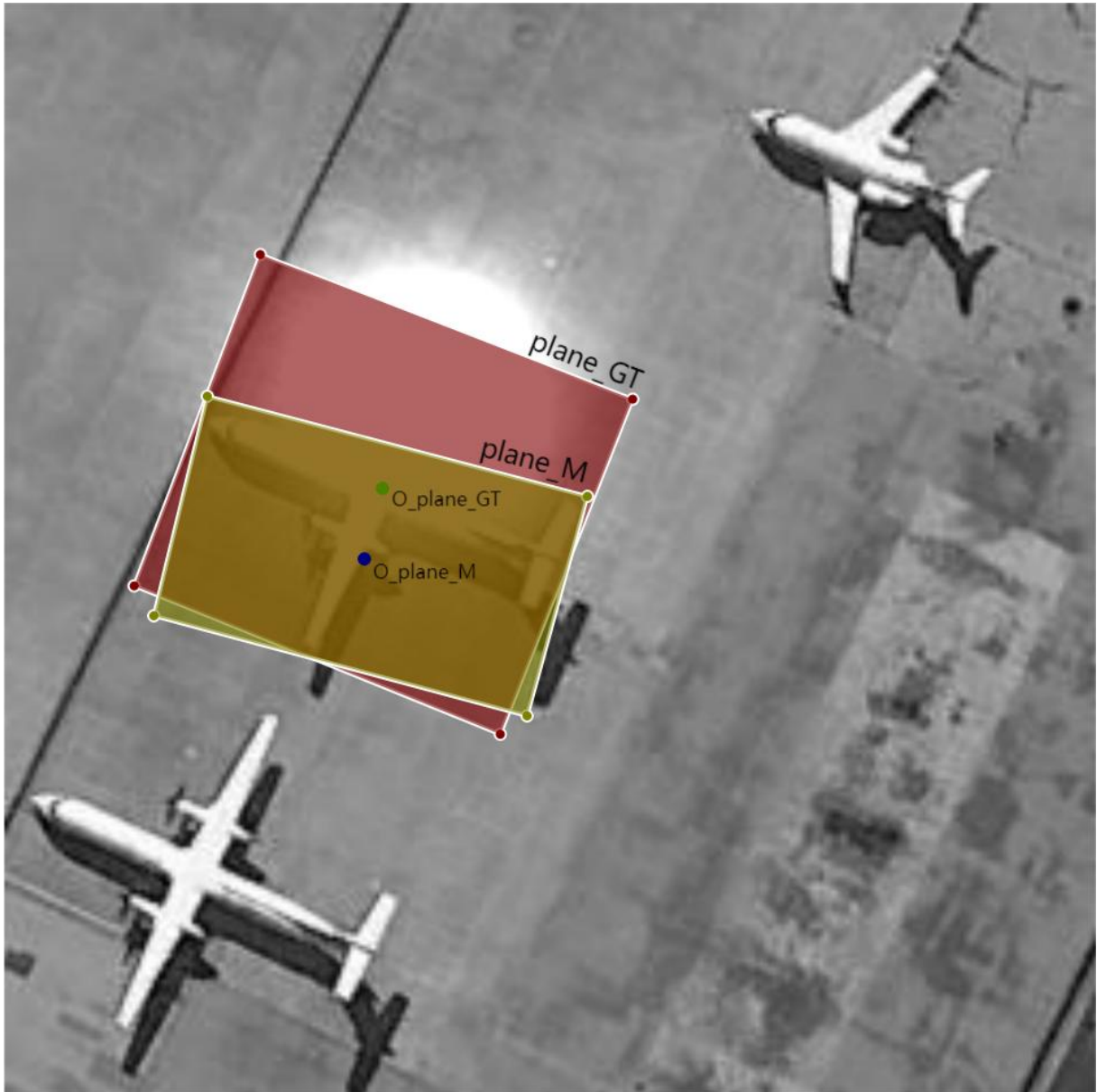
数据说明：无遮挡原始数据集采用 DOTA 数据集，选取其中飞机目标进行实验，数据集下载地址：<http://captain.whu.edu.cn/DOTAweb/>。



原图



遮挡 1



检测结果
遮挡条件下飞机目标检测结果示意图

参考文献:

- [1] You Only Look Twice—Multi-Scale Object Detection in Satellite Imagery With Convolutional Neural Networks (Part I), Adam Van Etten, 2017.
- [2] Lin Na, Feng Lirong, Zhang Xiaoqing. Aircraft Detection in Remote Sensing Image based on Optimized Faster-RCNN. Remote Sensing Technology and Application [J], 2021, 36(2): 275-284 doi:10.11873/j.issn.1004-0323.2021.2.0275
- [3] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A largescale dataset for object detection in aerial

images. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3974–3983, 2018

指导老师:胡静,by6040130@163.com, qq: 413953381

4.2 蜂群目标 2D 轨迹跟踪和姿态估计

任务描述: 实现对单视角视频中高速飞行的微小胡蜂进行 2D 多目标跟踪。胡蜂个体在图像中占据像素极少(微小目标),且飞行速度极快,导致在相邻帧间位移大、外观模糊,极易造成检测丢失和 ID 切换。传统的基于强监督学习的 MOT 方法在此类场景下面临两大瓶颈:一是获取大量精确标注的轨迹数据成本极高;二是模型在微小、快速运动目标上的泛化能力有限。本课题旨在探索不依赖或仅依赖极少量标注的跟踪范式。

请根据参考文献的弱监督检测和跟踪方法,探索单视角下,多微小目标的弱监督 2D 轨迹跟踪,要求实现以下两种方法,并在已有 2D GT 的片段(clip025, clip052, clip067)上,对比两种方法的 MOT 指标,分析两种方法的优劣。



- 1) 先基于弱监督的目标检测方法获取旋转检测框 obb,再实现多目标跟踪(TBD)。
- 2) 基于帧间差分方法,结合传统学习或深度学习获取目标旋转检测框 obb,基于此实现多目标跟踪。

最后,在上述检测跟踪的基础上,增加对胡蜂目标的躯干(或头部)检测,并进行胡蜂姿态估计,可用胡蜂三维仿真模型验证单目视觉条件下姿态估计方法的有效性。



胡蜂三维模型

数据:

Hornet3D 胡蜂观测数据集 请联系 wx: cdd11233455 获取

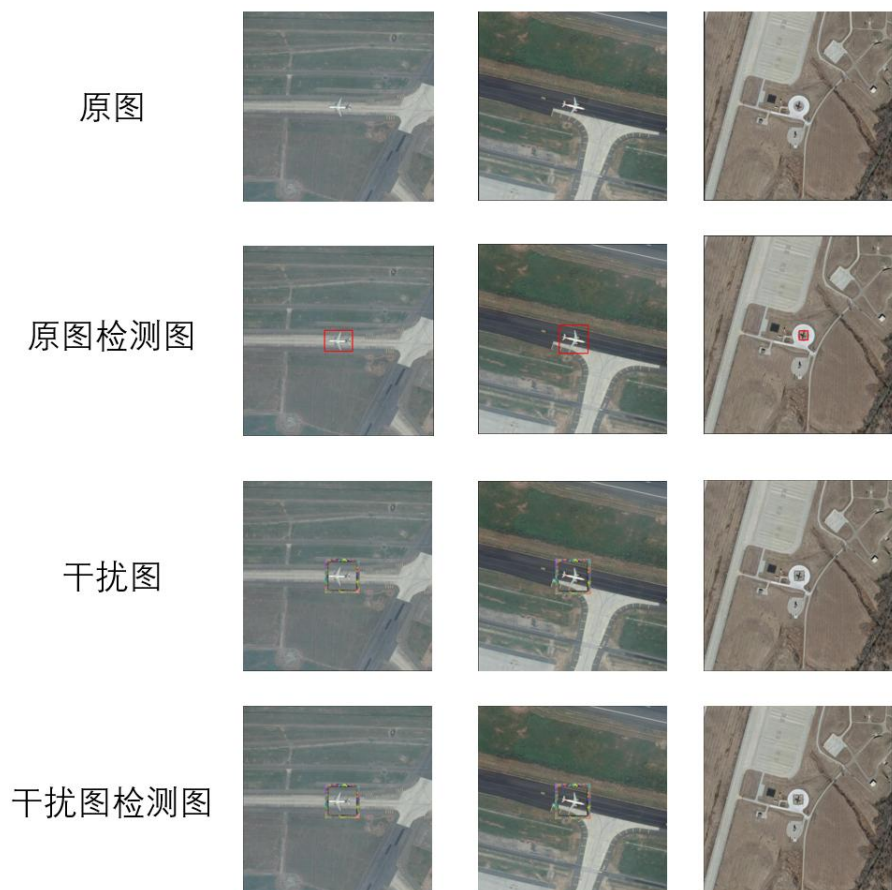
参考文献:

- [1] Ruiz, I., Porzi, L., Bulo, S. R., Kontschieder, P., & Serrat, J. (2021). Weakly supervised multi-object tracking and segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 125-133).
- [2] Li, Y. L., Lu, Y., Li, J., & Wang, H. (2023, June). Learning to reconnect interrupted trajectories for weakly supervised multi-object tracking. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.
- [3] Wang, Y., Ilic, V., Li, J., Kisačanin, B., & Pavlovic, V. (2023). ALWOD: Active learning for weakly-supervised object detection. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6459-6469).
- [4] Yoon, S., Shim, K., Park, K., & Kim, C. (2021, September). Weakly-Supervised Multiple Object Tracking Via A Masked Center Point Warping Loss. In *2021 IEEE International Conference on Image Processing (ICIP)* (pp. 1164-1168). IEEE.
- [5] He, C., Li, K., Zhang, Y., Xu, G., Tang, L., Zhang, Y., ... & Li, X. (2023). Weakly-supervised concealed object segmentation with sam-based pseudo labeling and multi-scale feature grouping. *Advances in Neural Information Processing Systems*, 36, 30726-30737.

指导老师:胡静,by6040130@163.com, qq: 413953381

4.3 鲁棒的深度检测网络对抗图像生成

任务描述: 针对 yolov8 等典型深度网络目标检测算法, 学习对抗数码图像生成技术, 考虑真实世界中的成像条件的对抗不可控性, 针对不同观测角度、不同光照条件、不同观测距离等成像条件, 研究鲁棒的对抗数码生成方法。



加干扰 patch 实验结果图

数据集: DOTA 数据集, 数据集下载地址: <http://captain.whu.edu.cn/DOTAweb/>。

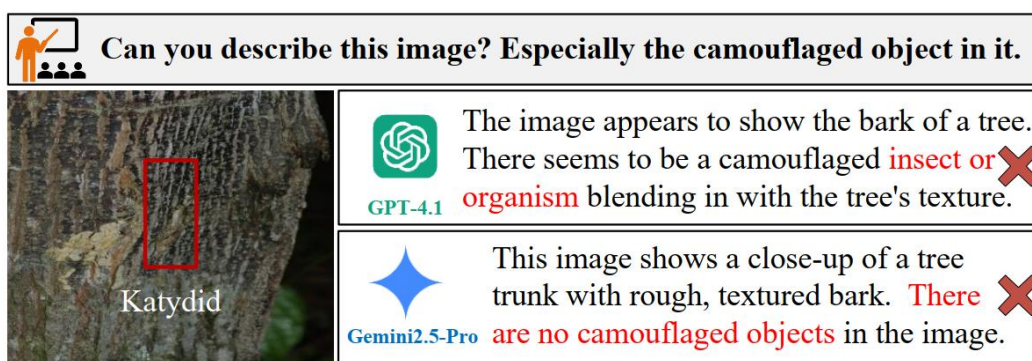
参考文献

- [1] Goodfellow I J, Shlens J, Szegedy C. Explaining and harnessing adversarial examples[J]. arXiv preprint arXiv:1412.6572, 2014.
- [2] Liu X, Yang H, Liu Z, et al. Dpatch: An adversarial patch attack on object detectors[J]. arXiv preprint arXiv:1806.02299, 2018.
- [3] Brown T B, Mané D, Roy A, et al. Adversarial patch[J]. arXiv preprint arXiv:1712.09665, 2017.
- [4] Huang L, Gao C, Zhou Y, et al. Universal physical camouflage attacks on object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 720-729.
- [5] Suryanto N, Kim Y, Kang H, et al. Dta: Physical camouflage attacks using differentiable transformation network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 15305-15314.

指导老师:胡静,by6040130@163.com, qq: 413953381

4.4 面向伪装场景的视觉语言模型微调与理解输出

任务描述：大型视觉语言模型（VLM）在通用图像描述任务上已取得显著进展，但其在伪装场景下的表现存在明显短板。伪装目标与背景的高度视觉相似性使得模型难以准确描述目标的物种类别、所处背景环境及伪装机制。本课题以伪装图像理解为核心任务，要求学生选取一个参数量在 7B 以内的开源 VLM，利用 CamoSense 数据集进行指令微调，使模型能够对输入的伪装图像自动生成一段准确的自然语言描述，内容涵盖目标物种、背景环境与伪装策略三个维度。课题同时包含微调前后的系统性评测，以验证领域数据对模型伪装理解能力的提升效果，并对 CamoSense 数据集的训练有效性进行实验验证。



现有 LVLMS 在理解伪装场景方面的表现存在局限性

数据集介绍：

1. CamoSense（主要数据集）：基于 COD10K、CAMO、NC4K、CHAMELEON 四个标准 COD 基准构建的多维属性标注数据集，共 10,513 张图像。每条样本包含物种粗粒度标签（8 类超类）与细粒度标签（135 子类）、背景环境类型（9 类）、伪装策略（background_matching/mimicry/disruptive_coloration）、难度参考标注（easy/medium/hard，仅供参考，不作为严格分组依据，详见注意事项）及模板化自然语言描述（格式为「A photo of a {物种} using {伪装策略} strategy in {背景}。」）。训练集 4,040 张，测试集 6,473 张。数据集为内部数据，已上传至百度网盘，链接联系 vx: 15727175828。



CamoSense 数据集样例

2. CAMO、COD10K、CHAMELEON、NC4K（原始图像与掩码来源，下载链接如下）

①

<https://github.com/visionxiang/awesome-camouflaged-object-detection?tab=readme-ov-file#Datasets>

② [DengPingFan/SINet: Camouflaged Object Detection, CVPR 2020 \(Oral\)](#)

③ [JingZhang617/COD-Rank-Localize-and-Segment](#)

主流数据集对照表

Dataset	Year	#Images	Mask	Species	Background	Strategy	Text	Focus/Positioning
CHAMELEON [28]	2016	76	✓	✗	✗	✗	✗	Small-scale testing
CAMO [16]	2019	1,250	✓	Coarse	✗	✗	✗	Camouflage dataset
COD10K [4]	2021	10,000	✓	Coarse	✗	✗	✗	General benchmark
NC4K [20]	2021	4,121	✓	✗	✗	✗	✗	Testing benchmark
CamoClass [41]	2025	10,523	✓	128 cls.	✗	✗	✗	Category-guided COD
CamoSense (Ours)	2026	10,513	✓	8 / 135 cls. [†]	9 cls.	3 cls.	✓	Multimodal reasoning

注意事项：

CamoSense 的 difficulty 字段（easy/medium/hard）基于人工主观判断，标注粒度和标注者间一致性未经严格验证，不一定能准确反映模型的实际识别难度。学生在按难度分层报告结果时，应将其作为参考性分析而非严格的分组实验，避免过度解读该字段所呈现的差异。建议优先以伪装策略类别（三类）作为主要分组维度进行分析，三类伪装策略之间的视觉难度差异有更充分的文献支撑。



MMCSBench^[5]的注解展示，可作为实验参考

题目要求：

基础任务为零样本能力评测。学生无需本地训练，直接调用一个开源 VLM（推荐 Qwen2.5-VL-7B 或 LLaVA-1.5-7B，量化后可在 8GB 显存笔记本运行），对 CamoSense 测试集的 200 张抽样图像进行零样本描述生成，人工归纳模型在三类伪装策略上的常见错误模式（例如将「拟态」型竹节虫描述为「树枝」而非动物，或忽略目标存在），形成基线分析报告。

进阶任务为指令微调与对比评测。利用 CamoSense 训练集的 caption 字段构建指令微调数据，格式为标准的「问题-图像-回答」三元组，对选定的开源 VLM 进行 LoRA 微调。微调完成后，在 CamoSense 测试集上与零样本基线进行全面对比，分别统计物种类别、背景环境、伪装策略三个维度的识别准确率变化，并按三类伪装策略分层汇报，检验 CamoSense 训练数据的有效性。

拓展任务为多选题形式的自动评测。参照 MMCSBench 的 MCQ 设计思路，利用 CamoSense 标注自动生成选择题（例如：「图中动物采用的伪装策略最接近以下哪项？(A) 背景匹配 (B) 拟态 (C) 破坏性着色 (D) 无伪装」），对微调前后模型的回答准确率进行全自动评测，无需人工打分，便于与 MMCSBench 基准进行横向比较。

评估方案：

结构化属性准确率为主要量化指标，从模型生成的自然语言描述中提取物种超类、背景类型、伪装策略三个字段进行关键词匹配，分别计算 Accuracy 和 Macro-F1，并按三类伪装策略分层报告。语义相似度作为生成质量的辅助指标，使用 BERTScore 衡量生成描述与 CamoSense 参考 caption 之间的语义相近程度，该指标仅需 CPU 即可计算。

人工评估作为小规模补充，每类伪装策略各抽取 20 张图像，由两名同学独立对生成描述的准确性和完整性进行 1-3 分打分，并计算评分一致性。

可行性说明：

本课题中所有深度学习实验均可通过两种途径完成，适合不同硬件条件的学生。有独立显卡（8GB 及以上）的学生可本地运行量化版 7B 模型进行推理与 LoRA 微调；显卡较弱的学生可借助 Google Colab 免费 T4 GPU 资源或调用 Qwen 开放 API（有免费额度。另提供一份资源包，仅供做此题的同学使用。兑换方式联系 vx：15727175828）完成推理评测部分，LoRA 微调部分则可使用学校服务器资源。评测阶段的 BERTScore 计算和多选题自动评测均仅需 CPU，任何环境下均可运行。

参考文献：

- [1] Fan, D.-P., et al. Concealed object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(10), 6024–6042, 2022.
- [2] Liu, H., et al. Visual instruction tuning (LLaVA). NeurIPS, 2023.
- [3] Hu, E., et al. LoRA: Low-rank adaptation of large language models. ICLR, 2022.
- [4] Zhang, T., et al. BERTScore: Evaluating text generation with BERT. ICLR, 2020.
- [5] MMCSBench: A fine-grained benchmark for large vision-language models in camouflage scenes. NeurIPS, 2025.
- [6] Liang Y, Qin G, Sun M, et al. A systematic review of image-level camouflaged object detection with deep learning[J]. Neurocomputing, 2024, 566: 127050.

指导老师:胡静,by6040130@163.com, qq: 413953381

4.5 微小目标检测

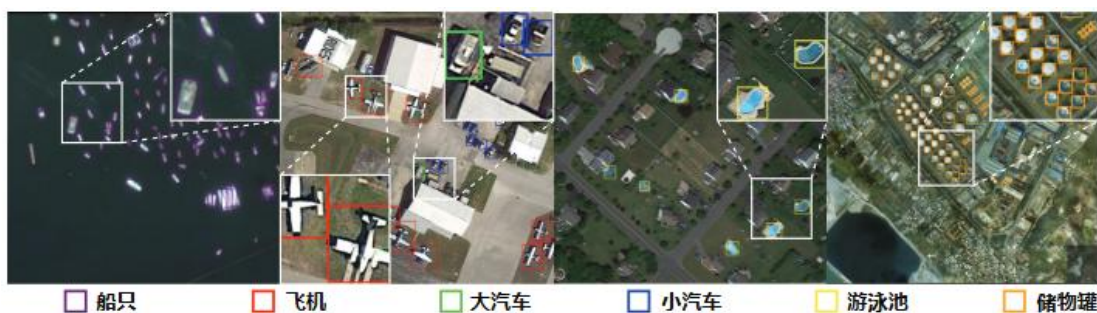
任务描述：微小目标通常指占用像素面积小于或等于 16×16 像素的目标，其通常尺寸极小、视觉信息极为匮乏，并且易受背景和噪声的干扰。本课题要求对现有微小目标检测算法进行改进，提升微小目标的检测性能。现有算法（算法、代码及相关资料联系 vx：Lucas22815）包含图像回溯模块（IR）和残差感知补偿模块（RPC），建议针对残差感知补偿模块进行改进，可从以下几个方向入手：

- （1）在 RPC 中引入 DCT (离散余弦变换) 或 Wavelet (小波变换) 作为注意力机制。
- （2）构建微型的“残差金字塔”，将残差图像下采样到与特征图相同的分辨率，然后在每一层分别做特征融合，或参考特征图结果对指定层做特征融合。
- （3）改进特征融合方式，从“加法”改为自适应门控融合。

(4) 查阅资料，采用其他特征增强及特征融合思路。

数据说明：

DOTA_{mini10}：在 13386 张图像中提供了 410305 个实例，包含 10 类目标，飞机、小车、大车、船、网球场、棒球场、游泳池、篮球场、储油罐和直升飞机。其中，训练集有 10653 张图像，328086 个实例，测试集有 2733 张图像，82219 个实例。



数据集DOTA_{mini10}中的各类目标的标注框示例

AI-TOD：在 28036 张航拍图像中提供了 700621 个实例，包含飞机、桥、油罐、船、游泳池、车辆、人和风车。经统计，AITOD 中的目标平均尺寸为 12.8 像素。



数据集 AI-TOD 中的各类目标的标注框示例

一些可供参考的数据集：

SODA 数据集：<https://shaunyan22.github.io/SODA/>

参考文献：

- [1]. G. Cheng, X. Yuan, X. Yao, K. Yan, Q. Zeng, X. Xie, et al. Towards Large-Scale Small Object Detection: Survey and Benchmarks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2023, 45(11):13467-13488.
- [2]. J. Wang, W. Yang, H. Guo, R. Zhang, G.-S. Xia. Tiny object detection in aerial images. in: Proceedings of the International Conference on Pattern Recognition (ICPR). Virtual, originally scheduled in Milan, Italy, January 10–15, 2021, IEEE, 2021: 3791–3798.
- [3]. C. Xu, J. Wang, W. Yang, H. Yu, L. Yu, G.-S. Xia. RFLA: Gaussian receptive field

based label assignment for tiny object detection. in: Proceedings of the European conference on computer vision (ECCV). Tel Aviv, Israel, October 23–27, 2022, Springer, 2022: 526–543.

- [4]. C. Xu, J. Wang, W. Yang, H. Yu, L. Yu, G.-S. Xia. Detecting tiny objects in aerial images: A normalized Wasserstein distance and a new benchmark. ISPRS Journal of Photogrammetry and Remote Sensing. 2022, 190:79–93.
- [5]. Zican Shi, Jing Hu, Jie Ren, Hengkang Ye, Xuyang Yuan, Yan Ouyang, Jia He, Bo Ji, Junyu Guo. HS-FPN: High Frequency and Spatial Perception FPN for Tiny Object Detection. in: Proceedings of the 39th Annual AAAI Conference on Artificial Intelligence (AAAI). Philadelphia, USA, February 25–March 4, 2025, AAAI Press, 2025:6896–6904.
- [6]. T. Zhang, X. Zhang, X. Zhu, G. Wang, X. Han, X. Tang, et al. Multistage enhancement network for tiny object detection in remote sensing images. IEEE Transactions on Geoscience and Remote Sensing. 2024, 62:1–12.

指导老师:胡静,by6040130@163.com, qq: 413953381

4.6 跨域少样本目标识别

任务描述: 1) 首先通过对以下域适应数据集抽样, 从而构建少样本目标域;
2) 使用非自然数据(如 officehome 的 Art, Cl 域)进行预训练, 对真实场景成像的小样本图像(如 OfficeHome 的 Rw, Pr 域)进行识别分类; 3) 使用真实场景进行预训练, 对非自然场景的小样本图像进行识别分类。

数据说明: Office31, office-home, CUFSF

进阶拓展:

- 1) 以上三个数据集都是源域和目标域类别空间相同的闭集问题, 在实际的不同成像方式之间的域适应识别任务场景中, 可能会存在目标域和源域类别无法一一对应的情况。应该考虑类别空间有重合但不完全相同的域适应识别问题, 例如将域适应少样本方法和一般跨域少样本方法结合起来, 利用部分类别获得领域之间的公共分类特征, 进而扩展到非公共的目标域类别上。
- 2) 一般来说, 标签无误的样本数量越多, 对模型的训练越有利, 但由于域适应问题中源域和目标域的样本存在着分布差异, 一些域差异较大的源域样本的引入可能

会对模型有负面影响。考虑进一步研究多源域的域适应识别方法中不同源域对模型在目标域上效果的影响。



Amazon

DSLR

Webcam

域适应目标识别数据集 Office31

Office31 数据集包括了三个域，分别是亚马逊商城（在线电商）图片，单反相机拍摄图片，网络摄像头拍摄图片，包括了 31 类数据。

下载地址：<https://faculty.cc.gatech.edu/~judy/domainadapt/>



数据集 office-home

Office-home 是域适应的一个基准数据集，包含了 4 个不同域，每个域由 65 个类别组成，分别是艺术——素描、绘画、装饰等形式的艺术形象、剪贴画——剪贴画图像的集合、产品——没有背景的对象图像和真实世界——使用常规相机拍摄的对象图像。它包含 15500 个图像，每个类平均约 70 个图像，一个类最多 99 个图像。

下载地址：<https://www.hemanthdv.org/officeHomeDataset.html>



数据集 CUFSS

CUFSF 数据集是一个中大人脸素描数据集，也可以用于人工图像和真实图像的域适应少样本目标识别研究，这个数据集包括了 1194 个人的照片，以及一张由艺术家观看

照片时绘制的带有形状夸张的草图。

下载地址：<http://mmlab.ie.cuhk.edu.hk/archive/cufsf/>

参考文献：

- [1] O. Vinyals, C. Blundell, T. Lillicrap, et al. Matching networks for one shot learning. Advances in neural information processing systems, 2016, 29
- [2] X. He, Y. Peng. Fine-grained visual-textual representation learning. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(2): 520-531
- [3] Y. Guo, N. C. Codella, L. Karlinsky, et al. A broader study of cross-domain few-shot learning. in: Proceedings of the European conference on computer vision, Glasgow, UK, 2020. Springer, 2020: 124-141
- [4] A. Krizhevsky, I. Sutskever, G. E. Hinton. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, 25
- [5] A. Nakamura, T. Harada. Revisiting fine-tuning for few-shot learning. arXiv preprint arXiv:1910.00216, 2019
- [6] G. Koch, R. Zemel, R. Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. in: Proceedings of the International conference on machine learning, Lille, France. ACM, 2015
- [7] J. Snell, K. Swersky, R. Zemel. Prototypical networks for few-shot learning. Advances in neural information processing systems, 2017, 30
- [8] P. Wang, L. Liu, C. Shen, et al. Multi-attention network for one shot learning. in: Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, 2017. IEEE, 2017: 2721-2729
- [9] B. Hariharan, R. Girshick. Low-shot visual recognition by shrinking and hallucinating features. in: Proceedings of the IEEE international conference on computer vision, Venice, Italy, 2017. IEEE, 2017: 3018-3027
- [10] Y. X. Wang, R. Girshick, M. Hebert, et al. Low-shot learning from imaginary data. in: Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, 2018. IEEE, 2018: 7278-7286

- [11] E. Schwartz, L. Karlinsky, J. Shtok, et al. Delta-encoder: an effective sample synthesis method for few-shot object recognition. *Advances in neural information processing systems*, 2018, 31
- [12] S. Ravi, H. Larochelle. Optimization as a model for few-shot learning. in: *Proceedings of the International conference on learning representations*, San Juan, Puerto Rico, 2016. Ithaca, 2016
- [13] C. Finn, P. Abbeel, S. Levine. Model-agnostic meta-learning for fast adaptation of deep networks. in: *Proceedings of the International conference on machine learning*, Sydney, Australia, 2017. PMLR, 2017: 1126-1135
- [14] Q. Sun, Y. Liu, T.-S. Chua, et al. Meta-transfer learning for few-shot learning. in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Long Beach, CA, USA, 2019. IEEE, 2019: 403-412
- [15] Y. Liu, Q. Sun, A.-A. Liu, et al. LCC: learning to customize and combine neural networks for few-shot learning. *arXiv preprint arXiv:1904.08479*, 2019

指导老师:胡静,by6040130@163.com, qq: 413953381

4.7 心脏疾病教师网络辅助的 ECG 单模态诊断

任务描述: 实现对心脏疾病的单模态心电图 (ECG) 诊断。心脏疾病诊断通常依赖多种互补模态, 如胸部 X 光 (CXR)、心电图 (ECG) 以及其他临床检查。其中, CXR 能够提供较丰富的结构与病理信息, 但获取成本较高, 且通常依赖医院影像设备; ECG 则具有获取便捷、成本低、适于大规模筛查等优势, 但其表征能力相对有限, 难以独立完成部分疾病的高精度诊断。

本课题旨在探索基于教师网络 (Teacher Network) 的心脏疾病诊断方法, 即利用训练阶段可获得的多模态信息 (如 CXR+ECG) 构建教师模型, 将高质量模态中的知识迁移到仅依赖 ECG 输入的学生模型中, 从而提升单模态 ECG 诊断能力。

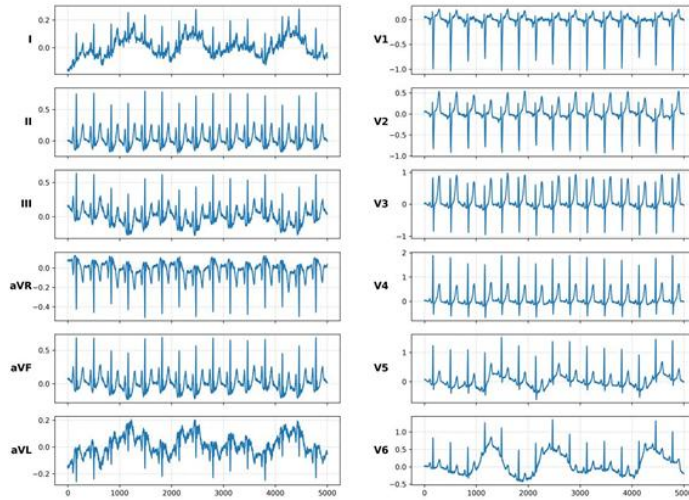
请根据参考文献, 在统一的跨模态诊断框架下, 构建 ECG 编码器与 CXR 编码器, 并利用 ECG-CXR 配对样本进行跨模态对齐, 将 CXR 模态中的疾病知识迁移到 ECG 表征中, 最终仅保留 ECG 分支完成心脏疾病诊断。在此基础上, 分别采用以下两种 ECG 自监督预训练策略进行对比:

(1) 基于对比学习的自监督预训练方法（参考 CroMoTEX^[1]）；

(2) 基于随机掩码重建的自监督预训练方法（参考 ECCL^[2]）。

在此基础上，可进一步增加对 ECG 表征的可解释性分析，例如可视化不同疾病类别的特征分布，或分析跨模态迁移前后 ECG 特征空间的变化。

数据：基于 MIMIC-IV-ECG、MIMIC-CXR-JPG 及其配对关系构建的实验数据集，请联系 wx: 18254028298 获取。



MIMIC-IV-ECG 数据集示例



MIMIC-CXR-JPG 数据集实例

参考文献：

[1] V. Punyamoorthy, A. Malusare, and V. Aggarwal, “Contrastive Cross-Modal Learning for Infusing Chest X-ray Knowledge into ECGs,” *arXiv preprint arXiv:2506.19329*, 2025.

[2] Z. Ding, Y. Hu, Z. Li, H. Zhang, F. Wu, Y. Xiang, T. Li, Z. Liu, X. Chu, and Z. Huang, “Cross-Modality Cardiac Insight Transfer: A Contrastive Learning Approach to Enrich ECG with CMR Features,” in *Medical Image Computing and Computer Assisted Intervention (MICCAI 2024)*, LNCS, vol. 15003, 2024, pp. 109–119.

[3] A. Radhakrishnan, S. F. Friedman, S. Khurshid, K. Ng, P. Batra, S. A. Lubitz, A. A. Philippakis, and C. Uhler, “Cross-modal autoencoder framework learns holistic representations of cardiovascular state,” *Nature Communications*, vol. 14, p. 2436, 2023.

[4] Y. Chen, S. Huang, Y. Cheng, P. Chen, Z. Rao, Y. Shu, B. Yang, L. Pan, and C. Guo, “AimTS: Augmented Series and Image Contrastive Learning for Time Series Classification,” *arXiv preprint arXiv:2504.09993*, 2025

指导老师:胡静,by6040130@163.com, qq: 413953381

4.8 无人机视角下的车辆目标检测

任务描述: 在无人机视角下, 由于拍摄角度较高、距离目标较远, 车辆目标在图像中呈现出多尺度、多形态的特点。具体而言, 任务要求实现对以下两类目标的精准检测与定位:

(1) 中大目标: 例如公交车、卡车等目标, 在图像中尺寸较大, 特征明显, 但在一些场景下也可能因背景复杂而受到干扰。

(2) 小目标: 例如私家车、摩托车等目标, 由于拍摄距离较远或目标本身尺寸较小, 其特征信息往往较为稀疏, 容易导致漏检。

因此, 本任务旨在设计一种能够兼顾不同尺度的车辆目标检测方法, 使得在无人机采集的复杂场景图像中能够同时实现大目标、小目标车辆的精准检测。系统要求具有高检测精度和较快的实时性, 同时对复杂背景和光照、天气等环境变化具有较好的鲁棒性。



图 1 VisDrone2019 数据集示例

数据描述:

VisDrone2019 数据集: 基准数据集共包含 8,629 张图像, 由各种无人机摄像头捕获, 覆盖范围广泛, 如图 1 所示。官方提供的数据中训练集是 6,471、验证集是 548、测试集

1,610 张，该数据集共提供了以下 10 个类，分别是：‘pedestrian’，‘people’，‘bicycle’，‘car’，‘van’，‘truck’，‘tricycle’，‘awning-tricycle’，‘bus’，‘motor’，数据集中各类目标的分布如图 2 所示。表 1 显示了每个类别中小、中和大目标的分布，其中，小目标占 VisDrone2019 数据集中目标总数的 60.49%，中大型目标占目标总数的 39.51%。

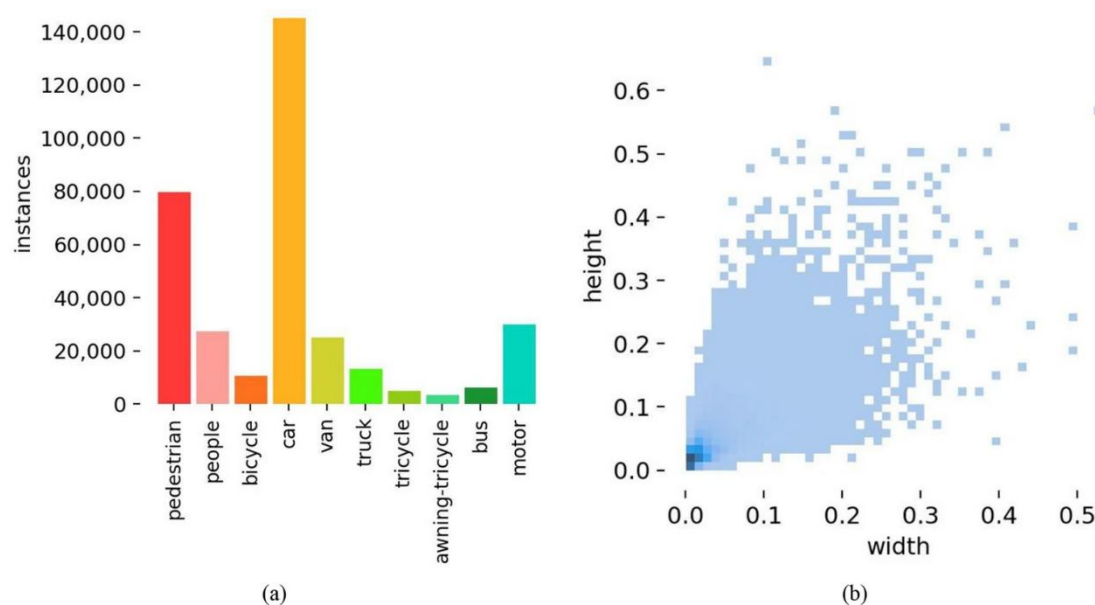


图 2 数据集中对象的分布：(a) 类别数量的分布；(b) 目标宽度和高度的分布；分布的集中程度用从浅白色到深蓝色的颜色渐变表示，表示分布越来越集中。

表 1 VisDrone2019 数据集中目标尺寸分布

Category	Small		Medium		Large	
	Number	Proportion(%)	Number	Proportion(%)	Number	Proportion(%)
Pedestrian	65227	82.22	13804	17.40	306	0.39
People	23497	86.84	3463	12.80	99	0.37
Bicycle	7111	67.85	3244	30.95	125	1.19
Car	69862	48.22	63079	43.54	11926	8.23
Van	10751	43.08	11517	46.15	2688	10.77
Truck	3982	30.93	6725	52.23	2168	16.84
Tricycle	2176	45.22	2425	50.39	211	4.38
Awning Tricycle	1360	41.90	1699	52.34	187	5.76
Bus	1661	28.03	3265	55.10	1000	16.87
Motor	21986	74.16	7395	25.00	266	0.90
Total	207613	60.49	116616	33.98	18976	5.53

参考文献：

[1] LI Yang, WU Lianquan, YANG Haitao, NIU Jinlin, CHU Xianteng, WANG Huapeng, ZOU Qinglong. A Small Target Detection Algorithm from UAV Perspective[J]. Infrared Technology, 2023, 45(9): 925.

- [2] Zhou, S., Zhou, H. & Qian, L. A multi-scale small object detection algorithm SMA-YOLO for UAV remote sensing images. *Sci Rep* 15, 9255 (2025).
- [3] 刘宸, 李士心, 孟范润, 陈范凯. 基于深度学习的无人机目标检测研究综述[J]. *计算机科学与应用*, 2023, 13(5): 1092-1099.
- [4] K. Ou et al., "Drone-TOOD: A Lightweight Task-Aligned Object Detection Algorithm for Vehicle Detection in UAV Images," in *IEEE Access*, vol. 12, pp. 41999-42016, 2024.
- [5] Qiu, Z.; Bai, H.; Chen, T. Special Vehicle Detection from UAV Perspective via YOLO-GNS Based Deep Learning Network. *Drones*.2023, 7, 117.
- [6] Xu, H.; Zheng, W.; Liu, F.; Li, P.; Wang, R. Unmanned Aerial Vehicle Perspective Small Target Recognition Algorithm Based on Improved YOLOv5. *Remote Sens.* 2023,15 , 3583.
- [7] D. Du et al., "VisDrone-DET2019: The Vision Meets Drone Object Detection in Image Challenge Results," 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South), 2019, pp. 213-226.
- [8] Liu, Jianzheng, et al. "Design of UAV target detection network based on deep feature fusion and optimization with small targets in complex contexts." *Neurocomputing* (2025): 130207.
- [9] Xue, B., Zhang, B. & Cheng, Q. Experiment study on UAV target detection algorithm based on YOLOv8n-ACW. *Sci Rep* 15, 11352 (2025).
- [10] Sui, Jiacheng, et al. "A new algorithm for small target detection from the perspective of unmanned aerial vehicles." *Ieee Access* 12 (2024): 29690-29697.
- 指导老师:胡静,by6040130@163.com, qq: 413953381

4.9 基于视觉和大语言模型的羽毛球比赛战术分析与预测

难度: ☆☆☆☆

任务描述: 结合目标检测、目标跟踪、三维重建和人体骨架等算法得到的数据和战术序列文本标签, 研究大语言推理模型(例如 DeepSeek-R1)的思维链(CoT)的生成方式, 生成带 CoT 的 Prompt 数据集并微调 DeepSeek-R1 实现对羽毛球比赛中球员所采用战术的分析与预测。

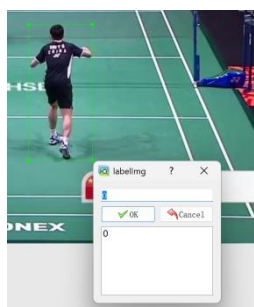
对羽毛球比赛视频运用目标检测和目标跟踪技术, 获得运动员和羽毛球在 2D 图像上的

坐标，再结合三维重建算法，得到运动员和羽毛球在 3D 空间的位置和速度数据，并且运用人体骨架算法提取运动员的姿态，将这些数据综合成 Prompt 数据集的 Question，将未来的战术序列文本标签作为数据集的 Response，研究复杂思维链（Complex_CoT）的生成方式，并利用整合而成的 Prompt 数据集微调 DeepSeek-R1 推理模型，实现战术分析与预测。

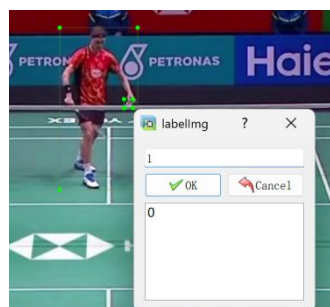
(1) 数据标注（运动员、羽毛球）

下载 labeling，教程链接如下：

<https://blog.csdn.net/knighthood2001/article/details/125883343>



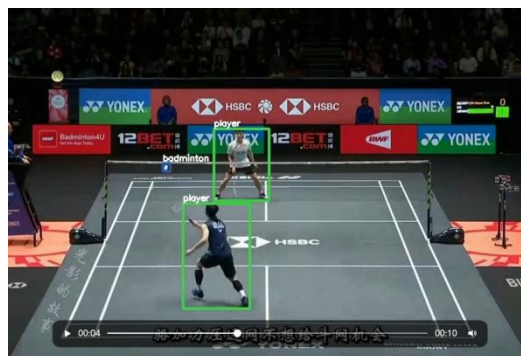
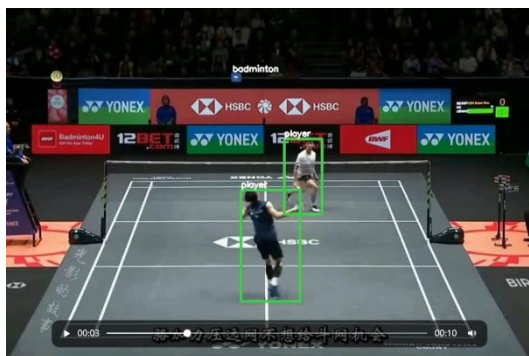
标注球员



标注羽毛球

(2) 运动员和羽毛球的检测与跟踪

搭建和训练目标检测与跟踪模型，实现对运动员和羽毛球的检测与跟踪：



(3) 基于大语言推理模型实现战术分析与预测

结合三维重建算法得到运动员和羽毛球的三维坐标和速度信息，结合人体骨架算法得到运动员的姿态信息，生成 Prompt 数据集，研究复杂思维链（Complex_CoT）的生成方式以更好地微调 DeepSeek-R1 进行战术分析与预测。

Question

运动员0在后场起跳，未来1s内的战术序列是什么？

Complex CoT



思考中... ^



Response

未来1s内的战术序列：
运动员0杀球
运动员1防守

参考文献：

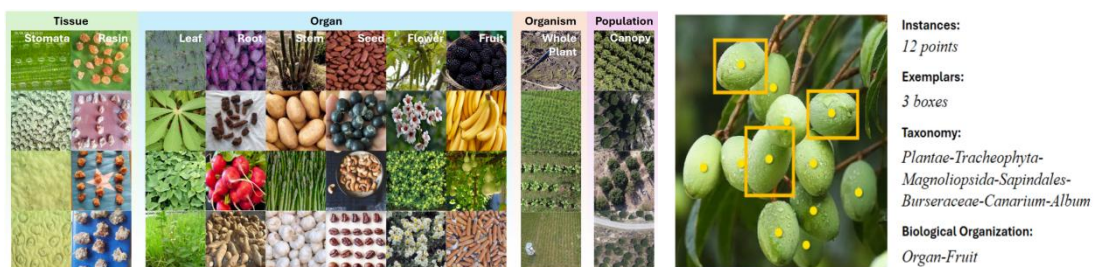
[1] Gao R, Wang L. MeMOTR: Long-term memory-augmented transformer for multi-object tracking[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 9901-9910.

[2] Tian Y, Ye Q, Doermann D. Yolov12: Attention-centric real-time object detectors[J]. arXiv preprint arXiv:2502.12524, 2025.

指导老师:肖阳,Yang_Xiao@hust.edu.cn, wx: hustcowboy

4.10 品种无关植物计数

任务描述：植物具有极高的生物多样性，在植物分类学中，包含界、门、纲、目、科、属、种等复杂的类别层级。实现面向不同品种的精准植物计数是现代植物表型分析的核心。然而，由于植物品种繁多、形态各异，且每年不断选育新品种，传统的特定类别计数算法往往面临严重的灾难性遗忘或泛化瓶颈，无法适用所有品种。近年来兴起的类别无关计数 [4-5] 打破了这一限制，它不依赖于待计数对象的类别先验，而是通过少样本示例专注于学习如何计数这一通用能力。但现有的通用的类别无关计数模型直接应用于植物类别无关计数任务时往往表现不佳，植物的极端跨尺度观测角度、密集生长带来的严重遮挡、生长的剧烈时空变异性等使得现有的通用特征匹配和相似度学习机制在植物领域大打折扣。本课题要求基于大规模植物分类学体系构建的植物计数数据集 TPC-268 [1]，给定 RGB 的植物图像和少量计数对象示例输入，实现可适用真实开放环境下跨尺度、跨场景、跨时空的高效精准的品种无关植物计数。



不同的植物计数目标（左）与数据集中一个样本的标注示例（右）

题目要求：

1) 算法基准与结果分析：小组成员需在 TPC-268 数据集上运行（#小组成员）个现有的主流类别无关计数基线模型（要求近三年的非论文中提到的基线，至少包含一个文本引导的 0-shot 计数方法）。重点结合测试结果，深入分析现有模型在植物图像上表现不佳的深层原因，可结合可视化错误案例和定量评估指标进行详尽的失效模式分析。

2) 数据拓展：从互联网和已有公开数据集中收集并标注新的植物计数数据，要求为 TPC-268 中没有的类别，小组每位成员需至少收集 10 个类别，每个类别至少包含 20 张图像，数据标注方式可参考论文 [1]。

3) 算法改进：针对上述分析中暴露出的痛点与植物固有特性，对基线模型进行针对植物特点的算法改进（可参考传统植物计数方法 [3]，针对植物的类别无关计数方法 [2]），切实提升模型在复杂植物场景下的计数精度与泛化能力，以 MAE、rMAE、RMSE、rRMSE、R2、FLOPS 等指标评估计数模型性能。

4) 交互系统设计与 Demo 演示：仿照 SAM 实现一个人工交互界面，将训练好的最优改进模型集成至系统中。实现用户通过框选、点选计数对象，或文本对任意输入植物图像的实时计数，在验收时进行完整的 Demo 演示，以验证算法的实用性。SAM 交互界面：

[Segment Anything | Meta AI \(segment-anything.com\)](https://segment-anything.com/)

参考文献：

[1] Xu, J., Hu, T., Hu, X., Zhou, L., Cao, S., Zhang, M., & Lu, H. (2026). Plant Taxonomy Meets Plant Counting: A Fine-Grained, Taxonomic Dataset for Counting Hundreds of Plant Species. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*.

[2] Hu, X., Li, X., Xu, J., Adan, A. D., Zhou, L., Zhu, X., ... & Lu, H. (2026). TasselNetV4: A vision foundation model for cross-scene, cross-scale, and cross-species plant counting. *ISPRS Journal of Photogrammetry and Remote Sensing*, 231, 745-760.

[3] Lu, H., Cao, Z., Xiao, Y., Zhuang, B., & Shen, C. (2017). TasselNet: counting maize tassels in the wild via local counts regression network. *Plant methods*, 13(1), 79.

[4] Ranjan, V., Sharma, U., Nguyen, T., & Hoai, M. (2021). Learning to count everything. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3394-3403).

[5] Wang, Z., Xiao, L., Cao, Z., & Lu, H. (2024). Vision transformer off-the-shelf: A surprising baseline for few-shot class-agnostic counting. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 6, pp. 5832-5840).

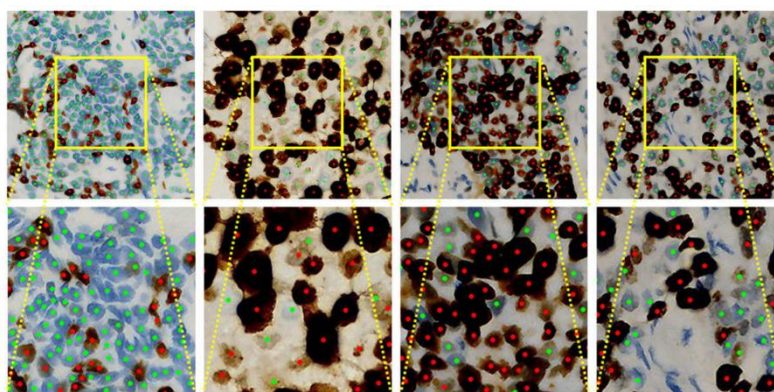
指导老师: 陆昊, e-mail: hlu@hust.edu.cn

难度系数: ★★★★★

建议专业: 人工智能

4.11 病理图像细胞计数

任务描述：病理图像分析是临床诊断的基础，而精准的细胞计数为其核心任务。病理图像中的细胞具有极高的复杂性，最大的痛点在于细胞形态极不规则，不同类型的细胞在大小、形状和边缘特征上差异巨大；其次，细胞经常呈现高度密集的聚集与重叠状态，边界极其模糊；此外，由于不同实验室的染色工艺差异，图像之间存在强烈的视觉域偏移。如下图所示，本题要求基于真实公开的病理图像数据集，给定组织切片图像，实现在复杂病理环境下高效、精准的细胞计数。



不同染色切片下形态各异、不规则分布的病理细胞图像与计数示例

题目要求：

- 1) 在三个病理细胞切片公开数据集 (BCData[1], CoNIC[2], MoNuSeg[3]) 上复现对比例如密度图回归(如 Steerer [4])、点监督定位匹配(如 PET[5])、实例分割(如 HoVer-Net[6]) 等至少 (#小组人数) 的细胞计数与定位估计方法，应至少包含三种不同类型的方法(如密度图回归、点监督定位、实例分割三类，基线要求近三年方法)，且至少需要同时包含医学领域与通用领域的计数模型，在选定数据集上报告对比性能指标和效率指标；
- 2) 细胞计数的宏观定量分析需以 MAE、MSE 等为指标进行评估，微观定位的定量分析需以 Precision、Recall、F1-score 等为指标，效率指标包括模型参数量、模型计算量、推理时间等；
- 3) 基于综合效果最好的计数或定位估计模型，针对病理切片数据集中细胞形态极其不规则、密集堆叠导致边界粘连等挑战进行改进，并基于该数据集分析细胞分布，实现对不规则细胞的精准识别与计数。要求可视化细胞的预测位置或密度分布，输出不同区域细胞的数量，并以散点图或热力图的形式与真实标签区分显示，并展示改进前后的评测指标对比结果。

数据集介绍：

- **BCData (Breast Cancer Cell Dataset) [1]**: 专为基于点标注的细胞计数和检测设计，包含大量具有严重重叠和形态不规则的乳腺癌细胞图像。
- **CoNIC (Colon Nuclei Identification and Counting) [2]**: 结肠核识别和计数挑战赛数据集，包含极其复杂的细胞形态和多类别细胞标签，非常适合验证模型对不规则细胞的特征提取能力。
- **MoNuSeg [3]**: 多器官细胞核分割与计数数据集，涵盖了多种器官的不同组织形态，染色差异极大，适合验证模型的鲁棒性。

参考文献:

- [1] Huang, Z., Ding, Y., Song, G., Wang, L., Geng, R., He, H., ... & Chen, J. (2020). Bcdata: A large-scale dataset and benchmark for cell detection and counting. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 289-298).
- [2] Graham, S., Vu, Q. D., Jahanifar, M., Weigert, M., Schmidt, U., Zhang, W., ... & CoNIC Challenge Consortium. (2024). CoNIC Challenge: Pushing the frontiers of nuclear detection, segmentation, classification and counting. *Medical image analysis*, 92, 103047.
- [3] Kumar, N., Verma, R., Anand, D., Zhou, Y., Onder, O. F., Tsougenis, E., ... & Sethi, A. (2019). A multi-organ nucleus segmentation challenge. *IEEE transactions on medical imaging*, 39(5), 1380-1391.
- [4] Han, T., Bai, L., Liu, L., & Ouyang, W. (2023). Steerer: Resolving scale variations for counting and localization via selective inheritance learning. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 21848-21859).
- [5] Liu, C., Lu, H., Cao, Z., & Liu, T. (2023). Point-query quadtree for crowd counting, localization, and more. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1676-1685).
- [6] Graham, S., Vu, Q. D., Raza, S. E. A., Azam, A., Tsang, Y. W., Kwak, J. T., & Rajpoot, N. (2019). Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 58, 101563.

指导老师: 陆昊, e-mail: hlu@hust.edu.cn

难度系数: ★★★★★

建议专业: 人工智能

4.12 密集人群运动分析

任务描述：基于视频监控的人群运动分析是公共安全管理中的关键环节。本任务不仅要求对行人的运动状态（速度、方向）进行精准估计，更要求实现对场景内多方向流量的定量统计。任务面向开阔场景中的密集人群，要求对武汉地铁（WuhanMetroCrowd）数据集进行标注，利用多种 Baseline 方法（光流、跟踪、轨迹预测）进行运动估计。在此基础上，要求基于自行标注的地铁数据和公开数据集，重点解决多向流量计数问题，即统计穿过特定检测区域或界面的各方向（上、下、左、右）行人数量，并评估各方向上的统计误差。



以光流法进行运动分析（左）与行人运动分析效果图（右）

题目要求：

1) 在自标注的 WuhanMetroCrowd 数据集和公开数据集（如 MovingDroneCrowd[1]）上对比例如光流法（如 RAFT[2]）、多目标跟踪或运动滤波（如 DeepSORT[3]）、人群流计数（如 [4]）、轨迹预测（如 SocialVAE[5]）等至少 2 种运动估计方法，且为不同类型的方法（如光流法、多目标跟踪算作两类），并在该数据集上报告评价指标（各个方向上的计数误差）及对比结果；其中，WuhanMetroCrowd 数据标注要求共标注不少于 300* 小组人数的总帧数，baseline 算法要求为近三年最新的算法模型。

2) 多向流量统计与误差分析：

- 实现对视频中上、下、左、右四个主要方向流过人数的自动计数。

- 定量指标评估：计算各方向计数的 MAE（平均绝对误差）、MSE（均方误差）及 WRAE（加权相对绝对误差）。

3) 基于一个合适的运动估计模型，针对 WuhanMetroCrowd 数据集的挑战进行改进，并基于该数据集分析人群运动，实现人群运动方向的识别效果。要求可视化不同方向的人群运动场（光流向量或运动热图），并按方向（上下左右）区分显示实时计数值，展示改进前后的指标对比；

4) 集成算法完成典型场景演示，输出人群运动场、多向流量统计面板，且应保证算法在主流 GPU 上的实时性。

数据描述： MovingDroneCrowd 是首个专门针对移动无人机视角下的视频个体计数（Video Individual Counting, VIC）任务的大规模真实场景数据集。它包含了由无人机在不同高度、俯仰角度和运动轨迹（如平移、旋转、变焦）下拍摄的高清视频序列，涵盖了极其密集的真实人群场景。下载链接：<https://pan.baidu.com/s/13NsJehHNw5IfGZy7qE2R6w?pwd=1234>。WuhanMetroCrowd 是从武汉地铁站采集的真实数据，包含了不同监控视频视角下丰富的人群运动场景。本数据集具有良好的运动方向区分度，适合使用光流算法进行处理，该数据集上的定量评测需要实现部分数据标注工作。本题目的最终演示测试将基于该数据，请联系指导老师获取。

参考文献：

- [1] Fan, Yaowu, et al. "Video individual counting for moving drones." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2025.
- [2] Teed Z, Deng J. Raft: Recurrent all-pairs field transforms for optical flow, ECCV: 402-419,2020.
- [3] Wojke, Nicolai, Alex Bewley, and Dietrich Paulus. "Simple online and realtime tracking with a deep association metric." 2017 IEEE international conference on image processing (ICIP). IEEE, 2017.
- [4] Liu, Weizhe, Mathieu Salzmann, and Pascal Fua. "Counting people by estimating people flows." IEEE Transactions on Pattern Analysis and Machine Intelligence 44.11 (2021): 8151-8166.
- [5] Xu, Pei, Jean-Bernard Hayet, and Ioannis Karamouzas. "Socialvae: Human trajectory prediction using timewise latents." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022.

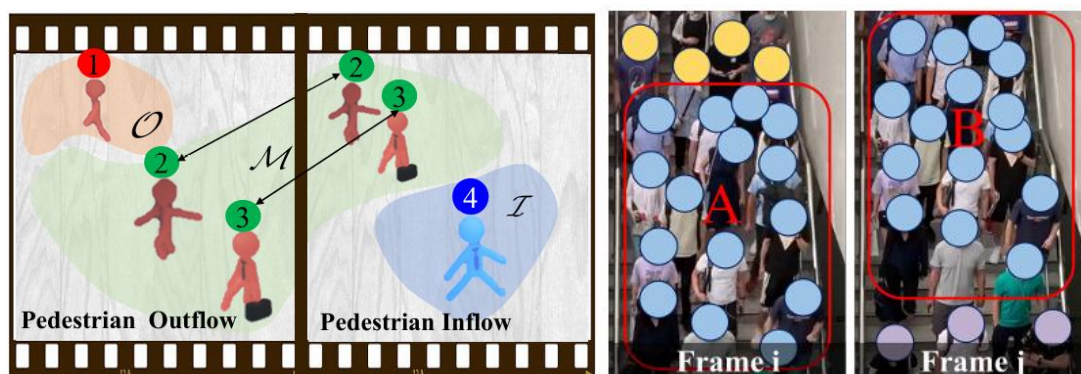
指导老师: 陆昊, e-mail: hlu@hust.edu.cn, QQ: 370875068

难度系数:★★★★★

建议专业: 人工智能

4.13 密集视频行人流量计数

任务描述: 本任务针对地铁视频监控场景统计行人流量。元旦、国庆等节假日期间, 武汉江汉路、中南路等关键地铁换乘站常出现严重拥堵情况, 造成了重大的公共安全隐患, 交通部门迫切需要一个实时检测人群流量的有效手段。在计算机视觉中, 研究人员已提出了大量基于图像的人群计数与定位算法[1]。然而, 流量是一段时间内某一场所中不同行人的数量, 无法仅凭计数和定位算法无法直接得到。为此, 基于多目标跟踪任务的研究工作, DRNet[2]引入了视频行人流量计数的新任务, 旨在对视频序列中出现的不同行人进行计数。然而, 现有算法在密集场景上的表现不尽人意, 为此, 本题目要求以最新的方法[3-7]作为 Baseline, 基于自行标注的地铁数据和公开数据集, 针对密集场景中严重的遮挡、背景干扰及个体消失重现等问题, 从物理约束和时空一致性出发进行算法创, 实现对视频行人流量的精准、稳定估计。



视频行人流量计数任务 (左) 与流量标注方法 (右)

题目要求: 针对 WuhanMetroCrowd 数据集, 解决密集场景下的视频行人流量计数问题, 具体而言:

1) 选取若干具有代表性的视频片段, 完成人群的流量数据标注工作, 并分析所标注数据的人群分布状况 (如每帧的人群平均密度等); 其中, WuhanMetroCrowd 数据标注要求共标注不少于 $300 \times$ 小组人数的总帧数。

2) 在 WuhanMetroCrowd 数据集和公开数据集 (如 VSCrowd[8]、UAVVIC[9]) 上训练

和测试至少 3 种与流量相关的算法，对比不同算法对流入、流出行人数量预测的准确性。评价时，需以 MAE、MSE、WRAE 为评价指标，详细公式见[2]；其中，baseline 算法要求为近三年最新的算法模型。

3) 在 WuhanMetroCrowd 数据集上，针对物理约束和时空一致性问题，完成模型的改进与微调，通过评价指标和可视化展示改进后模型在密集场景的优越性；

4) 完成演示 demo。

数据描述：WuhanMetroCrowd 是在武汉地铁采集的真实数据，包含了多种视频监控视角下的密集人群场景。本数据集能有效展现当前学术上最新的算法迁移到真实场景下会面临的若干问题。请联系指导老师获取相关数据。

参考文献：

[1] Liu, C., Lu, H., Cao, Z., & Liu, T. (2023). Point-query quadtree for crowd counting, localization, and more. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1676-1685).

[2] Han, Tao, et al. "Dr. vic: Decomposition and reasoning for video individual counting." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[3] Liu, Xinyan, et al. "Weakly Supervised Video Individual Counting." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

[4] Wan, Chang-Lin, Feng-Kai Huang, and Hong-Han Shuai. "Density-Based Flow Mask Integration via Deformable Convolution for Video People Flux Estimation." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024.

[5] Li, Rui, et al. "Prototype-Guided Dual-Transformer Reasoning for Video Individual Counting." Proceedings of the 32nd ACM International Conference on Multimedia. 2024.

[6] Fan, Yaowu, et al. "Video individual counting for moving drones." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2025.

[7] Huang, Feng-Kai, et al. "Flowing Crowd to Count Flows: A Self-Supervised Framework for Video Individual Counting." Proceedings of the 33rd ACM International Conference on Multimedia. 2025.

[8] <https://github.com/HopLee6/VSCrowd-Dataset>

[9] <https://github.com/streamer-AP/CGNet>

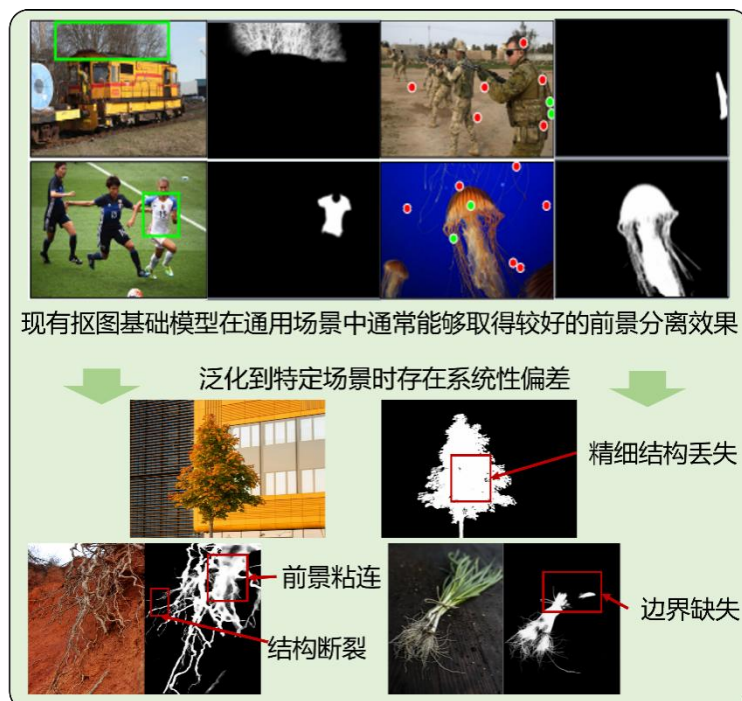
指导老师: 陆昊, e-mail: hlu@hust.edu.cn

难度系数: ★★★★★

建议专业: 人工智能

4.14 植物图像抠图

任务描述: 图像抠图是底层视觉域底层图像处理关键任务之一, 旨在从复杂背景中精确分离感兴趣的前景目标, 例如人像中的发丝等细粒度结构。近年来, 现代深度抠图技术发展焦点已由解决特定领域特定场景抠图问题, 逐步转向对象更泛化、技术更统一的抠图基础模型。尽管这类抠图基础模型已在多种评估协议下展现出较强的泛化能力, 但其泛化能力边界仍缺乏研究, 存在系统的针对特定视觉模式的显著性能退化。基于这一观察, 本题拟聚焦植物对象这一具有复杂边界、细粒度结构和显著分布差异的典型场景, 探索并建立一套面向抠图基础模型性能退化的标准化评估方法。



题目要求:

- 1) 系统评估部分主流抠图模型 (包括 zero-shot 模型 ZIM、基于视觉基础模型驱动的统一式抠图方法 MatAnything, 专用抠图模型等) 在植物场景中的失效现象与性能退化规律;
- 2) 结合已有数据集与适量新增标注数据, 设计一套针对植物场景的系统化、标准化评

测方法，重点刻画复杂边界、自遮挡、细长结构、前背景相似、空洞结构等困难因素对模型性能的影响；

3) 针对现有模型失效情况，尝试提出改进方法，提升模型在分布域外的抠图性能。

4) 探索基于抠图结果在植物计数中的应用价值，分析抠图质量与下游任务性能之间的关系。

代码和文献资源:

[1] <https://github.com/wchstrife/Awesome-Image-Matting>

[2] <https://github.com/zixuan-ye/SmartMatting>

[3] <https://github.com/SHI-Labs/Matting-Anything>

[4] <https://github.com/naver-ai/ZIM>

[5] <https://github.com/facebookresearch/segment-anything>

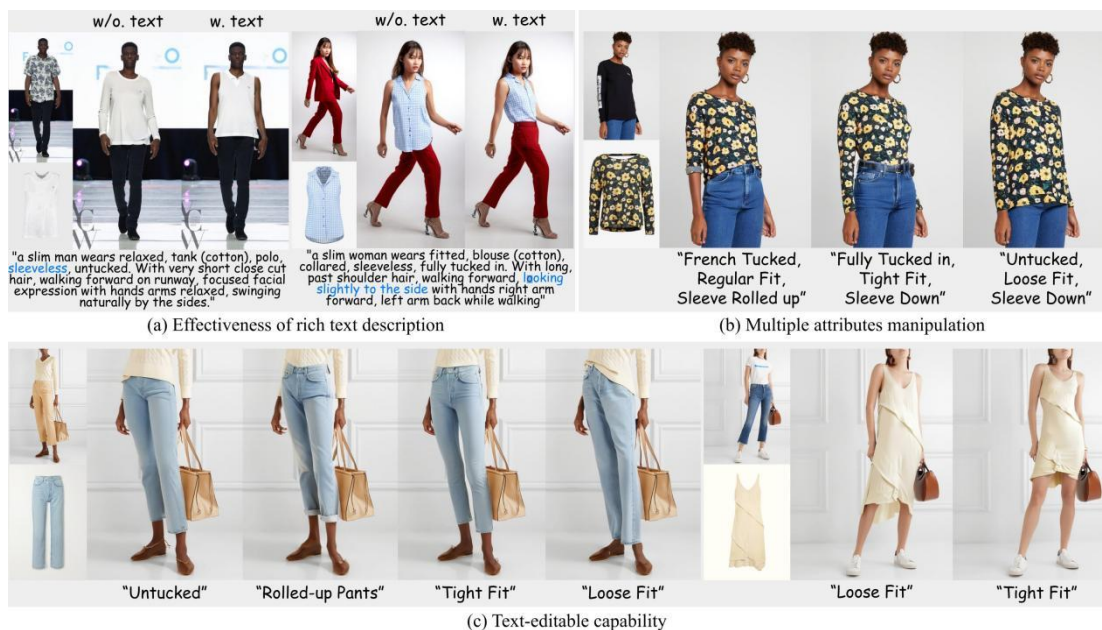
指导老师: 陆昊, e-mail: hlu@hust.edu.cn

难度系数: ★★★★★

建议专业: 人工智能

4.15 虚拟试穿风格编辑

任务描述:在电子商务模式下，服装商家需拍摄大量模特展示图以呈现商品，消费者亦期望获得更直观的服装交互体验。为降低商家因频繁更换模特拍摄所带来的高昂成本，同时提升消费者的购物体验，虚拟试衣技术逐渐成为研究热点。当前虚拟试衣算法多聚焦于“如何穿上衣服”，而忽略了“如何穿得好看”这一关键问题。其中，风格搭配是影响整体美观度与用户交互体验的重要因素。本课题主要探究基于文本描述的穿搭风格编辑方法，如下图所示。



基于文本对试穿风格编辑示例

题目要求:

- 1) 分析影响穿搭风格的关键因素，并选取（#小组人数）个风格维度展开课题研究。
- 2) 思考实现基于文本的风格编辑任务所需数据格式，设计数据收集与过滤方法，探索基于大模型实现数据合成与标注的自动化流程。
- 3) 基于所选取的风格维度，设计测试集与评估方法（如用户调研），对 PromptDresser / Qwen image edit / FLUX.1-Kontext-dev 等（#小组人数）个基线模型进行评估。
- 4) 尝试基于所选风格维度构建训练集，选取一个基线模型进行微调，并对其表现进行评估。
- 5) 实现人机交互界面，进行 demo 演示。

参考文献:

- [1] Kim J, Jin H, Park S, et al. Promptdresser: Improving the quality and controllability of virtual try-on via generative textual prompt and prompt-aware mask[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2025: 16026-16036.
- [2] Labs B F, Batifol S, Blattmann A, et al. FLUX. 1 Kontext: Flow Matching for In-Context Image Generation and Editing in Latent Space[J]. arXiv preprint arXiv:2506.15742, 2025.
- [3] Wu C, Li J, Zhou J, et al. Qwen-image technical report[J]. arXiv preprint arXiv:2508.02324, 2025.

指导老师: 陆昊, e-mail: hlu@hust.edu.cn

难度系数: ★★★★★

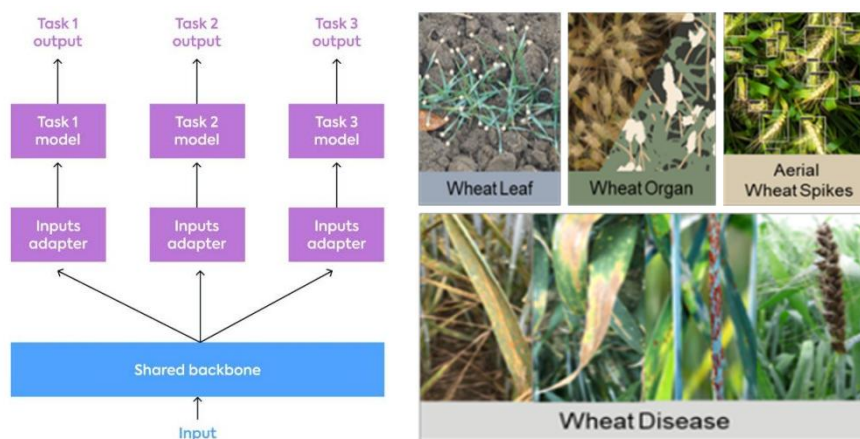
建议专业: 人工智能

4.16 面向植物表型视觉的异质同构多任务学习

任务描述: 在植物表型视觉分析中, 分割、检测、计数和分类是四大核心任务, 分别用于植物结构解析、器官定位、种群数量估计和病害识别。现有方法通常针对单一任务或共享输入数据的多任务学习进行建模, 然而在实际应用中, 不同任务往往依赖于不同数据集与不同图像输入, 导致传统多任务学习方法难以直接应用。

近年来, 多任务学习 (Multi-Task Learning, MTL) 在提升模型泛化能力和减少计算开销方面展现出巨大潜力, **但在跨数据集 (cross-dataset) 场景下的研究仍较为有限**。如何在不同任务数据不共享输入的情况下, 实现统一建模与知识共享, 是当前亟需解决的问题。

本题以植物视觉分析为研究对象, 基于分割、检测、计数、分类四类任务, 构建一个统一的跨数据集多任务学习框架。通过引入 MTDNN[1] 的数据调度方式, 实现不同任务间的协同训练, 并对多种经典及前沿多任务学习方法进行适配与对比分析。



多任务学习框架 (左) 与农业数据集图像示例 (右)

题目要求:

1) 多任务方法复现与适配

复现至少 (#小组人数) 种多任务学习方法, 并将其适配到统一的数据调度框架下进行训练。建议方法包括:

- MTLORA [2]
- TADFormer [3]

- DiTASK [4]
- TaskPrompter [5]
- PGT [6]

要求：尽量保持各方法的网络结构不变，在统一的训练流程和数据调度框架下完成适配。以 Swin Transformer 为骨干的模型请使用 tiny 或 small 量级的预训练权重。对于原方法中未包含的任务，需补充相应的任务头：检测任务采用 FCOS 检测头[7]，分类任务采用 MLP 分类头，分割任务采用 DPT 分割头[8]，计数任务参考 Fomo4wheat[9]和 PET[10]的计数头设计。

2) 数据集

本题目使用 Fomo4wheat 相关的植物视觉数据集。Fomo4wheat 是专门面向作物监测的视觉基础模型，其预训练使用了 ImAg4Wheat 数据集——目前最大、最多样化的小麦图像数据集，包含 250 万张高分辨率图像，涵盖 2010-2024 年间来自全球 30 个站点、超过 2000 种基因型、500+种环境条件的数据。

本题将使用包含以下任务的植物数据集：

- 分割任务：小麦器官分割数据（Wheat Organ Segmentation）
- 检测任务：小麦穗部检测数据（Wheat Head Detection From Ground-based Imagery）
- 计数任务：叶尖计数数据（Wheat Leaf Counting）
- 分类任务：病害分类数据（Wheat Disease Classification）

具体数据集的获取请联系指导老师。

3) 多任务训练与评估：

在统一框架下完成多任务训练，并报告各任务性能指标：

- 分割：mIoU、mAcc
- 检测：AP、AP50
- 计数：MAE、RMSE、 R^2
- 分类：mAP、BA

参考文献：

[1] Liu X, He P, Chen W, et al. Multi-task deep neural networks for natural language understanding[C]//Proceedings of the 57th annual meeting of the association for computational linguistics. 2019: 4487-4496.

- [2] Agiza, A., Neseem, M. and Reda, S., 2024. Mtlora: Low-rank adaptation approach for efficient multi-task learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* pp.16196-16205.
- [3] Baek, S., Lee, S., Jo, H., Choi, H. and Min, D., 2025. Tadformer: Task-adaptive dynamic transformer for efficient multi-task learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 14858-14868).
- [4] Mantri, K.S.I., Schönlieb, C.B., Ribeiro, B., Baskin, C. and Eliasof, M., 2025. Ditask: Multi-task fine-tuning with diffeomorphic transformations. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 25218-25229).
- [5] Ye, H. and Xu, D., 2023, May. Taskprompter: Spatial-channel multi-task prompting for dense scene understanding. In *The Eleventh International Conference on Learning Representations*.
- [6] Lu, Y., Sirejiding, S., Ding, Y., Wang, C. and Lu, H., 2024. Prompt guided transformer for multi-task dense prediction. *IEEE Transactions on Multimedia*, 26, pp.6375-6385.
- [7] Tian, Z., Shen, C., Chen, H. and He, T., 2020. FCOS: A simple and strong anchor-free object detector. *IEEE transactions on pattern analysis and machine intelligence*, 44(4), pp.1922-1933.
- [8] Ranftl, R., Bochkovski, A. and Koltun, V., 2021. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 12179-12188).
- [9] Han, B., Zhu, C., Han, D., Yu, R., Cao, S., Wu, J., Chapman, S., Wang, Z., Zheng, B., Guo, W. and Weiss, M., 2025. FoMo4Wheat: Toward reliable crop vision foundation models with globally curated data. *arXiv preprint arXiv:2509.06907*.
- [10] Liu, C., Lu, H., Cao, Z. and Liu, T., 2023. Point-query quadtree for crowd counting, localization, and more. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1676-1685).

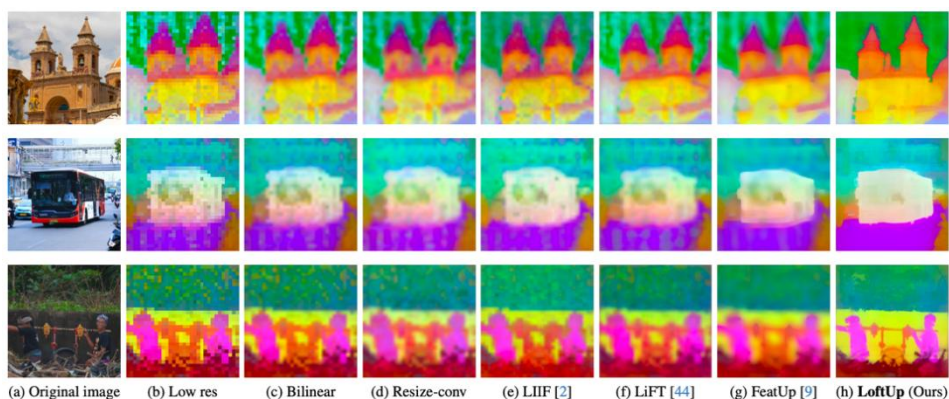
指导老师: 陆昊, e-mail: hlu@hust.edu.cn

难度系数: ★★★★★

建议专业: 人工智能

4.17 面向视觉 ViT 模型的通用特征上采样

任务描述：当前的视觉基础模型架构 DINOv2、DiT 等多以 ViT 架构主。然而，ViT 具有固有的空间分辨率缺陷，难以满足高分辨率输出的视觉稠密预测任务的需求。近年来，Model-Agnostic Feature Upsampling 的概念开始兴起，如何在不改变预训练模型参数的前提下，提升特征空间分辨率以及特征质量成为当前学界关注的重点。本题要求复现当前若干模型无关特征上采样算子，并评估其在垂直领域图像（如农业）中的效果。



模型无关特征上采样方法的特征可视化

题目要求：

- 1) 复现几种代表性的模型无关特征上采样方法，在论文所示的公开数据集中汇报相关性能，需与公开文献中的数值相近。
- 2) 选取（#小组人数）个下游任务，包括但不限于农业/遥感图像（如 LoveDA 数据集）、医学图像（如 ISIC 皮肤病变或局部 CT/MRI），评估相关上采样方法在下游任务中的有效性，验证上采样方法在对边界细节敏感的精细场景中的表现。
- 3) 基于复现的结果，针对现有方法在特定下游领域中容易出现的边缘分割不够锐利、特征图中存在噪点等问题，选取代表性的算法（如 NAF）进行探索性改进，通过下游任务的 mIoU、Acc、边界 F1、FPS 等指标验证改进效果，并给出分割预测图和特征可视化对比图。

参考文献：

- [1] Chambon, L., Couairon, P., Zablocki, E., Boulch, A., Thome, N., & Cord, M. (2025). NAF: Zero-Shot Feature Upsampling via Neighborhood Attention Filtering. arXiv preprint arXiv:2511.18452.
- [2] Wimmer, T., et al. "AnyUp: Universal Feature Upsampling." Proceedings of the International Conference on Learning Representations. 2026.

[3] Seo, M., Hamilton, M., & Kim, C. (2025). Upsample Anything: A Simple and Hard to Beat Baseline for Feature Upsampling. arXiv preprint arXiv:2511.16301.

[4] Liu, X., & Fan, H. (2026). DiveUp: Learning Feature Upsampling from Diverse Vision Foundation Models. arXiv preprint arXiv:2603.13571.

[5] Wang, J., et al. LoveDA: A Remote Sensing Land-Cover Dataset for Domain Adaptive Semantic Segmentation. Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks. 2021.

指导老师：陆昊，e-mail: hlu@hust.edu.cn

难度系数：★★★

建议专业：人工智能、自动化

4.18 基于 3D Gaussian Splatting 的植物三维重建与点云分割

任务描述：植物三维重建与表型提取是智能农业与植物表型组学中的重要研究内容。相比传统的 SfM/MVS 三维重建流程，3D Gaussian Splatting (3DGS) 在渲染速度和视图合成质量方面具有明显优势，但在植物场景中仍面临诸多挑战，例如叶片细长且边界薄弱、植株自遮挡严重、叶面纹理重复、少视角条件下重建不稳定等。这些问题会导致标准 3DGS 在植物重建中出现浮点、模糊、锯齿、结构膨胀以及几何表达不准确等现象，从而影响后续叶片分割与表型测量的可靠性。

本题要求系统调研并复现多种 3DGS 及其改进方案，分析不同改动对植物重建效果的影响，并思考哪些改进更适合植物这一类细粒度、薄结构、强遮挡的目标，实现植物叶片的分割和表型参数的计算。强调从“可视化重建”走向“可测量重建”。



输入图像（左）；3DGS 渲染图像细节（中）；分割后的叶片点云（右）

题目要求：

1) 在给定植物数据集上（至少选取三个植株数据集），对标准 3DGS 及若干代表性改进

方法进行调研、复现与对比，方法数量不少于（#小组人数）个，并分析不同方法在植物场景中的优缺点；

2) 基于重建得到的稠密植物点云，采用 2D-to-3D 分割方法(即利用二维分割模型在图片中分割出植物再反投影到三维空间中，需要标注)或其他三维分割方法完成植物叶片分割，至少实现叶片的语义分割，并进一步尝试完成单叶实例分离；

3) 结合实验结果，分析不同方案渲染质量的变化是由哪些改动引起的，分析植物场景下三维重建与分割的主要难点

数据描述:

给定数据集为植物多视角图像数据，包含不同作物、不同生长阶段及不同拍摄条件下的单株植物样本。本数据集能有效展现真实场景下植物三维重建可能出现的各项问题。数据集请咨询拷贝。

代码和文献资源:

[1] <https://github.com/MrNeRF/awesome-3D-gaussian-splatting>

[2] <https://mrnerf.github.io/awesome-3D-gaussian-splatting>

[3] Kundu, Abhijit, et al. "Virtual multi-view fusion for 3d semantic segmentation." *European conference on computer vision*. Cham: Springer International Publishing, 2020.

指导老师：陆昊，e-mail: hlu@hust.edu.cn

难度系数: ★★★

建议专业: 人工智能、自动化

4.19 眨眼检测与视线估计的一体化方法与系统

难度: ☆☆☆☆☆

任务描述: 眨眼是上下眼睑快速闭合与张开的过程，在疲劳检测、人机交互、注意力分析等领域具有重要应用价值。视线估计（Gaze Estimation）旨在通过眼部图像推断人的注视方向，是视觉交互的关键技术。本任务要求设计一个一体化系统，同时实现眨眼检测与视线方向估计两项功能。系统以普通摄像头采集的视频为输入，检测眨眼事件并估计注视方向，且在眨眼发生时能合理处理视线估计结果（如图 1 所示）。图 2 展示了经典的眼部纵横比（EAR）眨眼检测原理。

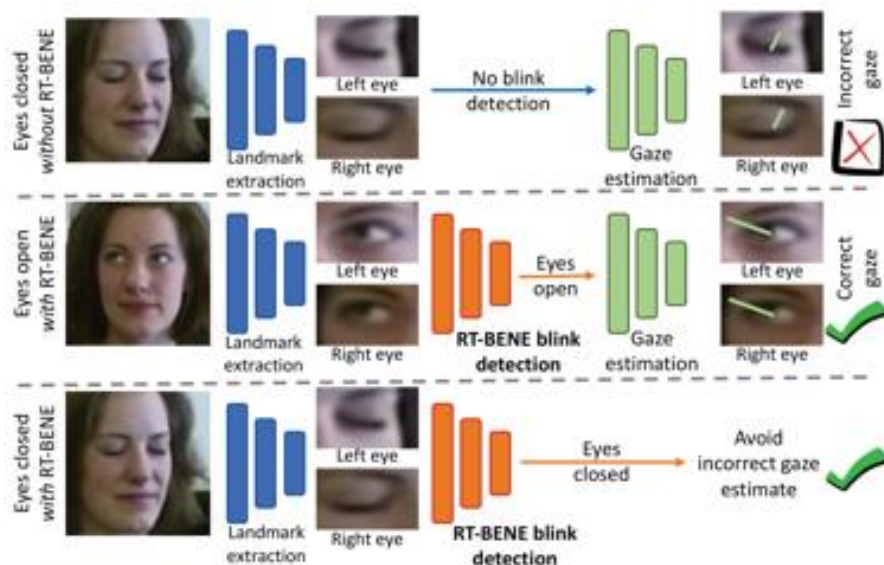


图 1 视线估计任务示意

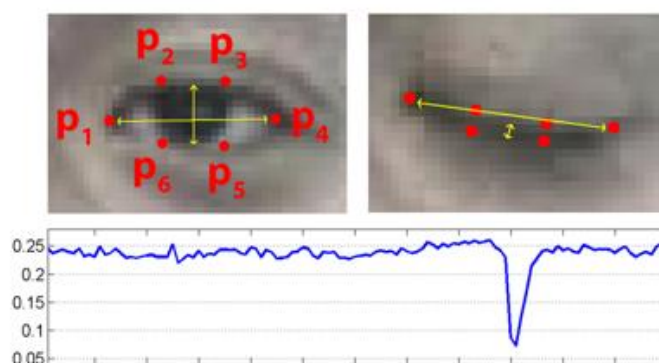


图 2 眨眼检测任务示意

参考代码:

https://github.com/Tobias-Fischer/rt_gene

<https://github.com/thorhu/Eyeblick-in-the-wild>

<https://github.com/swook/GazeML>

<https://github.com/TadasBaltrusaitis/OpenFace>

数据描述: 可选用以下公开数据集进行训练和评估, 也鼓励自行采集少量数据用于演示。

MPIIGaze 数据集[2]: 包含 213,659 张来自 15 名笔记本用户的注视图像, 涵盖多种光照和外观条件。下载地址:

<https://www.mpi-inf.mpg.de/departments/computer-vision-and-machine-learning/research/gaze-based-human-computer-interaction/appearance-based-gaze-estimation-in-the-wild>

GazeCapture 数据集[4]: 通过众包方式从约 1,500 人收集的大规模移动设备注视数据集。

下载地址: <https://gazecapture.csail.mit.edu/>

【自采数据】: 使用普通摄像头录制包含自然眨眼的短视频用于系统功能演示, 具体方式不做限定。

评价指标:

眨眼检测: 汇报 Precision、Recall 及 F1 指标。

视线估计: 汇报平均角度误差 (Mean Angular Error, 单位: 度)。

系统演示: 展示实时处理效果, 说明一体化处理的策略。

参考文献:

[1] K. Cortacero, T. Fischer, Y. Demiris. RT-BENE: A dataset and baselines for real-time blink estimation in natural environments. IEEE International Conference on Computer Vision Workshops (ICCVW), 2019.

[2] X. Zhang, Y. Sugano, M. Fritz, A. Bulling. Appearance-based gaze estimation in the wild. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 4511-4520.

[3] T. Soukupová, J. Čech. Real-time eye blink detection using facial landmarks. Computer Vision Winter Workshop (CVWW), 2016.

[4] K. Krafska, A. Khosla, P. Kellnhofer, et al. Eye tracking for everyone. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 2176-2184.

[5] D. E. King. Dlib-ml: A machine learning toolkit. Journal of Machine Learning Research, 2009, 10: 1755-1758.

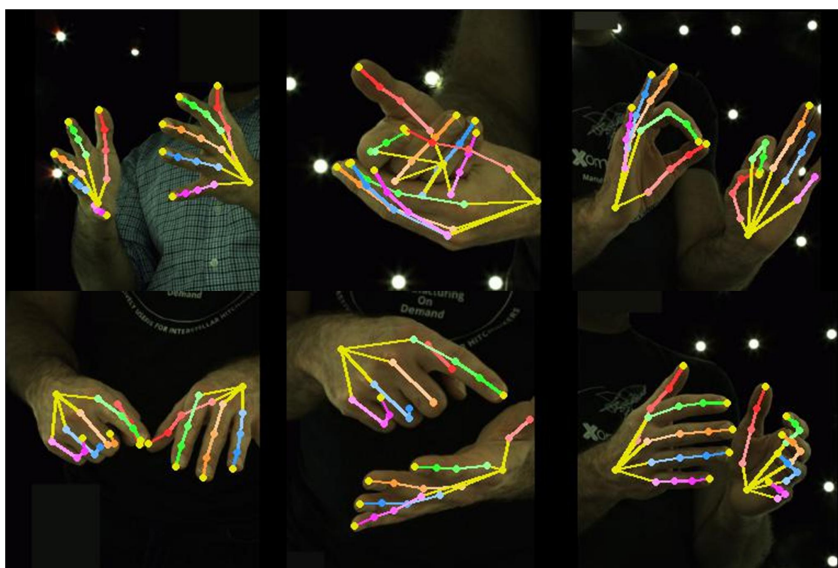
[6] C. Lugaresi, J. Tang, H. Nash, et al. MediaPipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172, 2019.

指导老师: 肖阳, Yang_Xiao@hust.edu.cn, wx: hustcowboy

4.20 基于 RGB 图像的三维交互手势估计

难度: ☆☆☆☆☆

任务描述: 三维交互手部姿态估计任务是对交互中的手部关节点进行估计, 而基于 RGB 图像则要求输入数据为 RGB 图像。请设计算法, 遵循指定数据集的基准进行训练和测试, 汇报性能指标。



基于 RGB 图像的三维交互手势估计结果图

参考代码: [GitHub - facebookresearch/InterHand2.6M: Official PyTorch implementation of "InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose Estimation from a Single RGB Image", ECCV 2020](https://github.com/facebookresearch/InterHand2.6M)

参考代码基于 PyTorch, 具有使用基础网络对 InterHand2.6M 数据集进行训练、测试以及 demo 演示的功能。

数据描述:

InterHand2.6M 数据集介绍及下载:

[InterHand2.6M dataset](#) | [InterHand2.6M \(mks0601.github.io\)](https://mks0601.github.io)

注意, 该数据集的总量较大, 为方便实验进行, 请下载以下百度网盘链接中的文件并仔细阅读 Readme.txt 文件, 使用该文件中的小型数据集进行实验。

链接: https://pan.baidu.com/s/1SSf6v_UI5OpaTeNUH_2LiQ

提取码: 4z8s

根据作者所提出的评价基准, 汇报 MPJPE 的性能指标。

参考文献:

[1]Moon G, Yu S I, Wen H, et al. Interhand2. 6m: A dataset and baseline for 3d interacting hand pose estimation from a single rgb image[C]//European Conference on Computer Vision. Springer, Cham, 2020: 548-564.

[2]Zimmermann C, Brox T. Learning to estimate 3d hand pose from single rgb images[C]//Proceedings of the IEEE international conference on computer vision. 2017: 4903-4911.

指导老师:肖阳,Yang_Xiao@hust.edu.cn, wx: hustcowboy

4.21 非受限条件下的人体眨眼检测

难度：☆☆☆☆

任务描述：眨眼为上下眼睑快速的闭合以及远离过程，直观上表现为眼部快速的睁眼-闭眼-睁眼的状态变化（如下图所示）。本任务旨在在非受限场景下实现眨眼检测。如图所示，在非受限场景下，人物面部属性、姿态、光照条件等方面均有巨大差异。请设计算法，遵循指定数据集的基准进行训练和测试，汇报性能指标。



眨眼示例



非受限场景实例

参考代码：

<https://github.com/thorhu/Eyeblick-in-the-wild>

<https://github.com/rezaghoddoosian/Early-Drowsiness-Detection>

数据描述：

HUST-LEBW 数据集[1]：<https://thorhu.github.io/Eyeblick-in-the-wild/>

- (1) 对裁剪短视频的眨眼识别：输入短视频段，预测眨眼/非眨眼的二分类任务。请汇报 Recall, Precision 以及 F1 指标。
- (2) 对未裁剪长视频的眨眼检测：输入长视频，定位每次眨眼的开始和结束帧位置。请汇报 AP_{50} 指标，指标计算方法请参考提供的示例代码：

链接：<https://pan.baidu.com/s/1ggeZkIPpAKgip68SGNv2w> 提取码：tkc2

参考文献：

- [1] G. Hu, Y. Xiao, Z. Cao, L. Meng, Z. Fang, J. T. Zhou, and J. Yuan. Towards real-time eyeblick detection in the wild: Dataset, theory and practices. IEEE Transactions on Information Forensics and Security, 15:2194–2208, 2020.
- [2] Roberto Daza, Daniel DeAlcala, Aythami Morales, Ruben Tolosana, Ruth Cobos, and Julian Fierrez. Alebk: Feasibility study of attention level estimation via blink detection applied to e-learning. arXiv preprint arXiv:2112.09165, 2021.

[3] Tereza Soukupov'a and Jan Cech. Real-time eye blink detection using facial landmarks. In Proc. Computer Vision Winter Workshop (CVWW), pages 1–8, 2016.

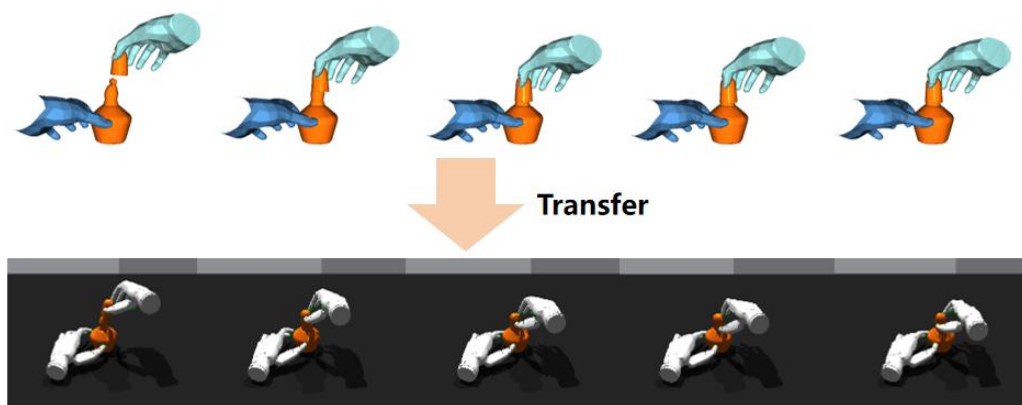
[4] R. Ghoddoosian, M. Galib, and V. Athitsos, “A realistic dataset and baseline temporal model for early drowsiness detection,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 0–0

指导老师:肖阳,Yang_Xiao@hust.edu.cn, wx: hustcowboy

4.22 基于人类操作演示数据的灵巧手操作迁移

难度：☆☆☆☆☆

任务描述：本任务聚焦于具身智能与机器人技能学习领域，旨在面向帮助机器人从人类专家演示中学习灵巧操作技能这一核心研究要点，设计人工智能模型。如图所示，输入标准采集的人体操作动作序列数据与对应场景任务下的物体运动轨迹数据，借助模型将人体动作迁移到与人手存在异构的多指灵巧手上，输出指定灵巧手的关节运动数据，完成操作技能从人手向灵巧手迁移的过程。任务要求设计一套方法框架，能够针对指定动作序列与灵巧手模型上完成训练与推理，获取具备物理约束、实现任务功能的相应灵巧手动作序列，并针对开源交互数据集中常见的几种灵巧操作动作进行实验。本任务可以基于模仿学习、强化学习等范式进行实现，在 OAKINK2、GRAB 等数据集上选取动作数据在仿真环境中进行实验，需在物体旋转误差、物体平移误差、指尖位置误差等指标上比较不同类型方法性能，并进行清晰地动作结果可视化，具体指标评估可以参考相关方法的实现。



基于人类操作序列数据的灵巧操作迁移过程

数据描述：不限定具体使用的数据集，可从 OAKINK2 数据集（<https://oakink.net/v2/>）、GRAB 数据集（<https://grab.is.tue.mpg.de/>）、H2O 数据集（<https://taekwon.com/projects/h2o/>）等开源数据集中选取实验数据，考虑到算力情况，可以使用较小体量的数据进行实验。本任务主要使用其中的手部关节数据与物体轨迹数据，可以使用文本、接触信息等作为辅助。

参考文献：

[1] Zhan X, Yang L, Zhao Y, et al. Oakink2: A dataset of bimanual hands-object manipulation in complex task completion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 445-456.

[2] Chen Y, Wang C, Yang Y, et al. Object-centric dexterous manipulation from human motion data[J]. arXiv preprint arXiv:2411.04005, 2024.

[3] Li K, Li P, Liu T, et al. Maniptrans: Efficient dexterous bimanual manipulation transfer via residual learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2025: 6991-7003.

[4] Zhou B, Yuan H, Fu Y, et al. Learning diverse bimanual dexterous manipulation skills from human demonstrations[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2026, 40(34): 28919-28927.

指导老师:肖阳,Yang_Xiao@hust.edu.cn, wx: hustcowboy

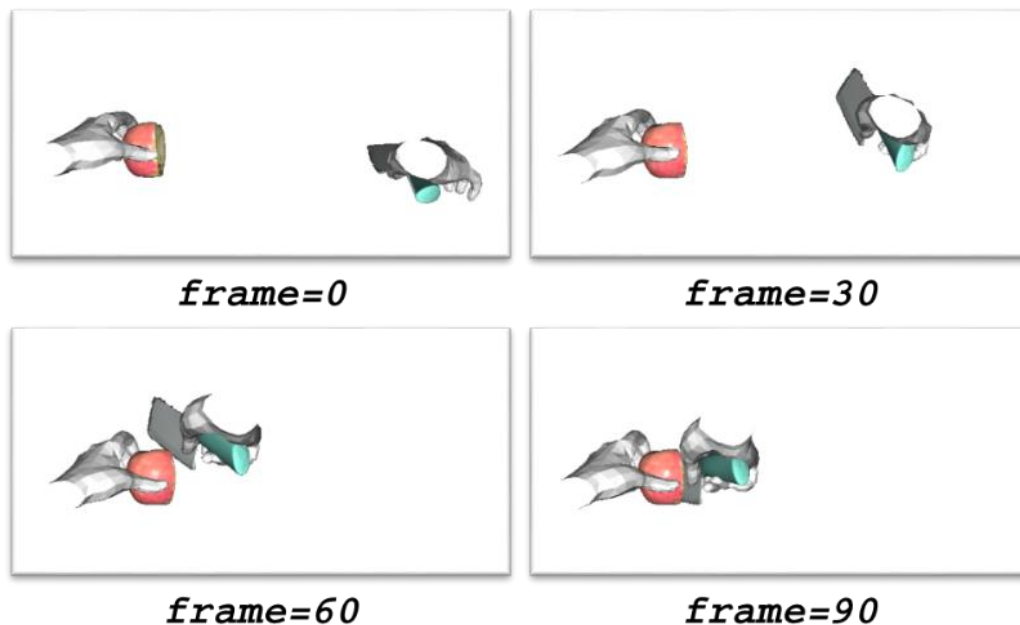
4.23 基于文本的手-物体交互动作序列生成

难度：☆☆☆☆

任务描述：本任务聚焦于具身智能领域，面向机器人灵巧手自主操作任务，旨在设计人工智能模型理解文本语义意图，并根据交互物体运动轨迹生成相应的手部操作动作序列。如图 1 所示，本任务以自然语言任务描述和物体运动轨迹作为模型输入，生成相应的符合语义以及物理规律的手部动作序列。任务要求设计多模态融合算法，将文本语义与物体轨迹联合建模，生成精确抓取、操作等手部姿态参数序列，同时保证手-物体交互的物理合理性，如提高物理稳定性、防止手-物体穿模。本任务可以采用回归、Diffusion、VAE 以及自回归等多种类型方法，在 OAKINK2 数据集上进行实验，需在接触率(contact ratio)、固体交叉体积(solid intersection volume)、FID、PSKL-J，MPJPE 等指标上比较不同

类型算法性能，性能指标评估参考 <https://github.com/oakink/OakInk2-TaMF>。

"Cut the apple with the white cleaver knife."



基于文本的手-物体交互动作序列生成结果

数据描述:

OAKINK2 数据集格式参考链接: <https://github.com/oakink/OakInk2#dataset-format>.

本任务不使用图像数据，模型输入包括文本描述、交互物体点云、交互物体 6d pose 序列，模型输出交互手的 MANO 参数序列。

考虑到计算资源受限，本任务采用部分数据集进行实验，数据集百度网盘链接:

<https://pan.baidu.com/s/11r6c0ejbcD7KperxwUeWHw?pwd=wmae>, 提取码: wmae



参考文献:

[1] Zhan, Xinyu, et al. "Oakink2: A dataset of bimanual hands-object manipulation in complex task completion." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024. ([oakink/OakInk2-TaMF: \[CVPR 2024\] OakInk2 baseline model: Task-aware Motion Fulfillment \(TaMF\) via Diffusion](#)) (Diffusion 方法)

[2] Cha, Junuk, et al. "Text2hoi: Text-guided 3d motion generation for hand-object interaction." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024. ([JunukCha/Text2HOI: Text2HOI: Text-guided 3D Motion Generation for Hand-Object Interaction](#)) (Diffusion 方法)

[3] Zhang, Jianrong, et al. "Generating human motion from textual descriptions with discrete representations." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2023. ([Mael-zys/T2M-GPT: \(CVPR 2023\) Pytorch implementation of "T2M-GPT: Generating Human Motion from Textual Descriptions with Discrete Representations"](#)) (人体动作序列生成领域自回归方法)

指导老师:肖阳,Yang_Xiao@hust.edu.cn, wx: hustcowboy

4.24 基于文言文高密度编码的数学推理方法

难度: ☆☆☆☆

任务描述: 文言文（尤其是先秦至魏晋时期的古汉语）在书面表达的信息密度上，是人类自然语言的“天花板”。构建“白话文-文言文”平行数学应用题数据集。基于 Qwen3.5-0.8B 小尺寸模型展开微调，在受限上下文窗口下，测试高密度文言文对缓解思维链（CoT）截断的作用，并可以进一步推广到其它类型的推理任务与应用上。

1、数据构建与量化

参考《九章算术》等中国古代算书的文本风格，选取逻辑链清晰的数学应用题，剔除冗余场景描述，构建“原题-文言文-带 CoT 的 Response”三元组平行数据集。调用 Qwen Tokenizer，量化对比白话文与文言文的 Token 长度与信息密度差异。

2、Qwen3.5-0.8B 微调与测试

基于 Qwen3.5-0.8B，分别使用两组数据集进行 LoRA 对比微调。设定极小的最大推理窗口，测试在高压比文本下，模型思维链的完整性与最终推理准确率。

本任务的数据集需要自己构建，参考数据如图所示，组织多个同学协助数据收集。最终在构建的数据集上划分训练测试集，并进行算法的准确率评测。



人体跌倒行为 RGB 图像序列。

参考文献：

- [1] Ma, X., Wang, H., Xue, B., Zhou, M., Ji, B., & Li, Y. (2014). Depth-based human fall detection via shape features and improved extreme learning machine. *IEEE journal of biomedical and health informatics*, 18(6), 1915-1922. (现有数据集参考)
- [2] Wang, X., Ellul, J., & Azzopardi, G. (2020). Elderly fall detection systems: A literature survey. *Frontiers in Robotics and AI*, 7, 71. (跌倒检测系统综述，了解目前主流方法和问题)
- [3] Nogas, J., Khan, S. S., & Mihailidis, A. (2020). Deepfall: Non-invasive fall detection with deep spatio-temporal convolutional autoencoders. *Journal of Healthcare Informatics Research*, 4(1), 50-70.

(基于深度图异常检测的方法，代码地址：<https://github.com/JJN123/Fall-Detection>)

指导老师:肖阳,Yang_Xiao@hust.edu.cn, wx: hustcowboy

4.26 抗遮挡目标点跟踪

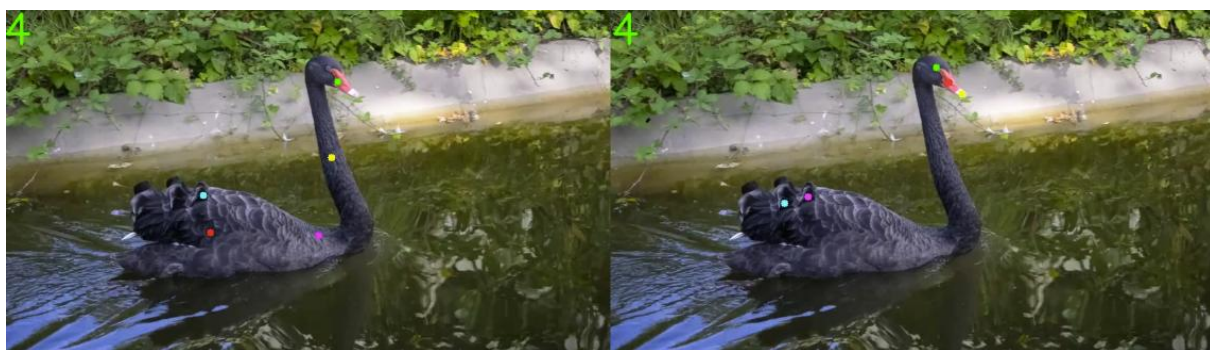
任务描述：视频点跟踪（Tracking Any Point）是指给定视频某帧中的任意一个像素点，在后续所有帧中持续估计该点的精确像素位置及其可见性。与传统的边界框目标跟踪不同，点跟踪要求像素级精度，且需判断目标点是否被遮挡。遮挡是导致点轨迹漂移和丢失的主要因素，设计能在遮挡后准确恢复点位置的跟踪算法是本任务的重点。要求对比

至少两种点跟踪方法在公开基准上的性能，评价指标包括 AJ (Average Jaccard)、位置精度、遮挡预测精度等，重点分析各方法在遮挡场景下的表现差异，并对典型成功与失败案例进行可视化分析。

数据集： TAP-Vid (含 DAVIS、Kinetics 子集)、PointOdyssey 等公开点跟踪基准数据集

CoTracker: <https://github.com/facebookresearch/co-tracker>

TAP: <https://github.com/google-deepmind/tapnet>



目标点跟踪示意图

参考文献：

- [1] Karaev N, Rocco I, Graham B, et al. CoTracker: It is Better to Track Together. ECCV, 2024.
- [2] Karaev N, Makarov I, Wang J, et al. CoTracker3: Simpler and Better Point Tracking by Pseudo-Labeling Real Videos. ECCV, 2024.
- [3] Doersch C, Yang Y, Veber M, et al. BootsTAP: Bootstrapped Training for Tracking-Any-Point. arXiv:2402.00847, 2024.
- [4] Zholus A, Doersch C, Yang Y, et al. TAPNext: Tracking Any Point (TAP) as Next Token Prediction. arXiv:2504.05579, 2025.
- [5] Aydemir G, Cai X, Xie W, et al. Track-On: Transformer-based Online Point Tracking with Memory. ICLR, 2025.

指导老师：杨卫东, yangwd@hust.edu.cn, qq14223185

4.27 三维场景重建

任务描述： 三维场景重建是计算机视觉的核心任务，在数字城市、文化遗产保护、自动驾驶等领域具有广泛应用。近年来，基于神经辐射场 (NeRF) 及 3D 高斯泼溅 (3DGS)

等新型场景表征方法取得了突破性进展，能够从多视角图像实现高保真的新视角合成与三维重建。本任务要求利用多视角图像进行三维场景重建与新视角渲染，对比至少两种不同的重建方法，采用 PSNR、SSIM、LPIPS 等指标进行定量评估，并对渲染结果进行可视化分析。鼓励使用自采集数据（如手机拍摄校园场景）进行测试。

数据： Mip-NeRF360、Tanks and Temples、自采集数据，红外图像数据等。



三维高斯场景重建

参考文献：

- [1] Mildenhall B, Srinivasan P P, Tancik M, et al. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. ECCV, 2020.
- [2] Kerbl B, Kopanas G, Leimkühler T, et al. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM TOG (SIGGRAPH), 2023.
- [3] Yu Z, Chen A, Huang B, et al. Mip-Splatting: Alias-Free 3D Gaussian Splatting. CVPR, 2024.
- [4] Lu T, Yu M, Xu L, et al. Scaffold-GS: Structured 3D Gaussians for View-Adaptive Rendering. CVPR, 2024.
- [5] Fu Y, Liu S, Kulkarni A, et al. COLMAP-Free 3D Gaussian Splatting. CVPR, 2024.

指导老师：杨卫东，yangwd@hust.edu.cn, [qq14223185](tel:14223185).

4.28 红外序列图像弱小目标检测

任务描述： 红外弱小目标检测技术是反无人机探测、低空安防及预警系统的关键技术之一。在云层遮挡、海面耀斑等动态复杂背景下，单帧图像中的弱小目标极易与高亮杂波

或随机噪声混淆，导致单帧检测算法虚警率居高不下。多帧红外序列包含了目标连续运动的时空相关性与航迹信息，能够有效区分真实的运动目标与静态/随机闪烁背景。然而，实际场景中平台自身的运动、目标剧烈的形态波动以及极低的信杂比，导致帧间目标的时空特征关联极其困难。设计能充分提取并融合时空信息、适应动态复杂环境的多帧（序列）红外弱小目标检测算法是本任务的重点。

IRDST 数据集下载链接：<https://pan.baidu.com/s/1icmYAeRfPdL5n0YcUtuNhg?pwd=kk6h>

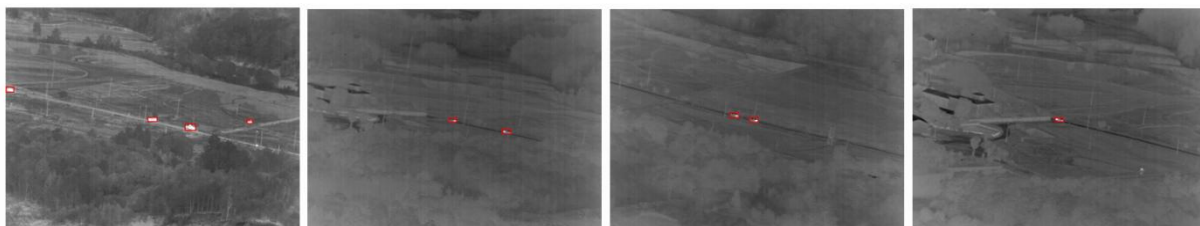
NUDT-MIRSDT 数据集下载链接：

<https://pan.baidu.com/share/init?surl=pSN350eurMafLiHBQBnrPA&pwd=5whn>

TSIRMT 数据集下载链接：

<https://drive.google.com/drive/folders/1aWDNdUWkTOuV3fLbgLDEqM2N2erW05n?usp=sharing>

数据：IRDST、NUDT-MIRSDT、TSIRMT



多帧序列红外小目标检测

参考文献：

- [1] C. Deng et al., "Learning Global Dynamic Query for Large-Motion Infrared Small Target Detection," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 64, pp. 1-16, 2026, Art no. 5002016, doi: 10.1109/TGRS.2026.3657842.
- [2] Chen, Shengjia, Luping Ji, Jiewen Zhu, Mao Ye and Xiaoyong Yao. "SSTNet: Sliced Spatio-Temporal Network With Cross-Slice ConvLSTM for Moving Infrared Dim-Small Target Detection." in *IEEE Transactions on Geoscience and Remote Sensing* 62 (2024): 1-12.
- [3] H. Qin, T. Xu, Y. Tang, F. Xu and J. Li, "OSFormer: One-Step Transformer for Infrared Video Small Object Detection," in *IEEE Transactions on Image Processing*, vol. 34, pp. 5725-5736, 2025, doi: 10.1109/TIP.2025.3598426.
- [4] F. Li, P. Rao, W. Sun, Y. Su and X. Chen, "A New Motion Feature-Enhanced Multiframe Spatial-Temporal Infrared Target Detection Network," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1-19, 2025, Art no. 5006819, doi:

10.1109/TGRS.2025.3603784.

- [5] Chen, S., Ji, L., Duan, W., Peng, S., & Ye, M. (2025). Motion Prior Knowledge Learning with Homogeneous Language Descriptions for Moving Infrared Small Target Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(2), 2186-2194. <https://doi.org/10.1609/aaai.v39i2.32217>

指导老师：杨卫东, yangwd@hust.edu.cn, qq14223185

4.29 无人机位姿估计

任务描述：在低空安防领域，对无人机的姿态估计是实现精准识别与稳定跟踪的重要途径。旨在从单目视频图像中估计无人机目标的三维姿态（3个旋转自由度）。例如通过检测无人机关键点（如旋翼中心等），结合几何求解或学习方法估计 6DoF 位姿。本任务要求实现从无人机检测到位姿估计的完整流程，并进行定量评估，对不同飞行姿态下的估计结果进行可视化分析。

数据：MMAUD、2DroneKey、3DronePose 等公开无人机位姿数据集，目标红外图像。

DroneKey: <https://github.com/kkanuseobin/DroneKey>



图1 无人机位姿估计

参考文献：

- [1] Hwang S-B, Kim Y-J, Choi A J. DroneKey: Drone 3D Pose Estimation in Image Sequences using Gated Key-representation and Pose-adaptive Learning. arXiv:2508.17746, 2025.
- [2] Jung M, Choi A J. Enhanced Real-Time 6D Pose Estimation for Automatic Recovery of In-Flight UAVs Using Distance-Aware Keypoint Heatmaps. *Scientific Reports*, 2025.
- [3] Wen B, Yang W, Kautz J, et al. FoundationPose: Unified 6D Pose Estimation and Tracking

of Novel Objects. CVPR, 2024.

[4] Xu L, Qu H, Cai Y, et al. 6D-Diff: A Keypoint Diffusion Framework for 6D Object Pose Estimation. CVPR, 2024.

[5] Peng S, Liu Y, Huang Q, et al. PVNet: Pixel-wise Voting Network for 6DoF Pose Estimation. CVPR, 2019.

指导老师：杨卫东，yangwd@hust.edu.cn, qq14223185.

4.30 可见光-红外跨模态图像翻译

任务描述：红外图像数据采集需要专用红外成像设备，成本高昂且标注困难，导致红外领域训练数据稀缺，制约相关算法的发展。利用易获取的可见光图像数据，通过跨模态翻译生成高质量红外图像，是缓解红外数据不足问题的有效途径。本任务要求基于生成模型，实现可见光图像到红外图像的跨模态翻译。以可见光图像作为条件输入，生成与之语义一致的红外风格图像。要求实现至少一种生成模型的训练与测试，并与不同种类方法（如基于GAN的方法，基于扩散的方法等）进行对比分析，评估对下游检测任务的影响。



可见光场景 真实红外 生成结果

红外图像样本生成

数据集：LLVIP数据集：包含16836对严格配准的可见光-红外行人图像对。下载地址：

<https://bupt-ai-cz.github.io/LLVIP/>

KAIST多光谱行人数据集：包含白天和夜晚场景的可见光-红外配对图像。下载地址：

<https://soonminhwang.github.io/rgbt-ped-detection/>

M3FD数据集：包含多种场景的红外-可见光配对图像及目标检测标注。下载地址：

<https://github.com/JinyuanLiu-CV/TarDAL>

参考文献：

[1] Ran L, et al. DiffV2IR: Visible-to-Infrared Diffusion Model via Vision-Language Understanding. arXiv:2503.19012, 2025.

- [2] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. NeurIPS, 2020.
- [3] Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models. CVPR, 2022.
- [4] Mao F, et al. PID: Physics-Informed Diffusion Model for Infrared Image Generation. CVPR, 2024.
- [5] Saharia C, Chan W, Saxena S, et al. Palette: Image-to-image diffusion models. SIGGRAPH, 2022.

指导老师：杨卫东, yangwd@hust.edu.cn, qq14223185

4.31 跨视角图像匹配

任务描述：跨视角目标定位是指给定一张参考图像（模板），在不同视角或光照条件下的查询图像中自动定位该目标的精确位置。本任务要求实现从特征提取、匹配到目标定位的完整流程并可视化结果。对比至少 2 种不同类型的特征匹配方法，分析其在匹配数量、内点率和定位精度上的差异。

数据：

HPatches 数据集。下载地址：<https://github.com/hpatches/hpatches-dataset>

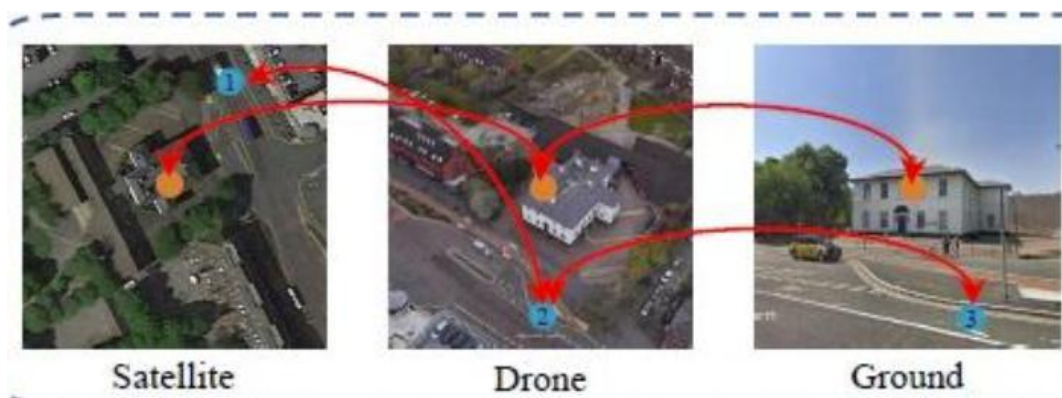


图1 跨视角图像

参考文献：

- [1] Edstedt J, Nordström D, Zhang Y, et al. RoMa v2: Harder Better Faster Denser Feature Matching. arXiv preprint arXiv:2511.15706, 2025.
- [2] Edstedt J, Sun Q, Bökman G, et al. RoMa: Robust Dense Feature Matching. CVPR, 2024.
- [3] Lindenberger P, Sarlin P-E, Pollefeys M. LightGlue: Local Feature Matching at Light Speed. ICCV, 2023.
- [4] Wang Y, He X, Peng S, et al. Efficient LoFTR: Semi-dense local feature matching with

sparse-like speed[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024: 21666-21675.

[5] DeTone D, Malisiewicz T, Rabinovich A. SuperPoint: Self-Supervised Interest Point Detection and Description. CVPR Workshops, 2018.

[6] Balntas V, Lenc K, Vedaldi A, et al. HPatches: A Benchmark and Evaluation of Handcrafted and Learned Local Descriptors. CVPR, 2017.

指导老师：杨卫东，yangwd@hust.edu.cn，[qq14223185](https://github.com/yangwd)

4.32 基于视觉基础模型的遥感图像目标检测与分割

任务描述：遥感图像目标检测与分割是遥感智能解译的核心任务。传统方法需要大量标注数据训练专用检测器，成本高昂。本任务要求利用视觉基础模型构建零样本/少样本的遥感目标检测与分割，具体要求：（1）以文本提示实现对遥感典型目标（如飞机、船舶、车辆等）的检测与分割；（2）对比不同提示方式（框提示、点提示、组合提示）对分割结果的影响；（3）分析基础模型在遥感场景下对小目标、密集目标等的适应性与不足，讨论可能的改进方向。

预期效果：输入遥感图像，首先输出目标检测结果（检测框+类别），然后基于 SAM 输出每个目标的精细分割掩码。可视化展示检测框与分割掩码的对比效果，分析不同提示策略下的分割精度差异。

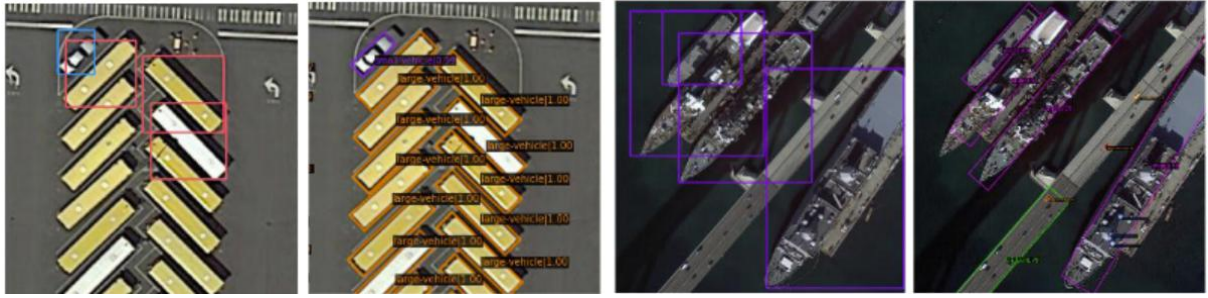
数据：

NWPU VHR-10 数据集：包含 10 类遥感目标的高分辨率遥感图像，共 800 张。下载地址：https://github.com/chaozhong2010/VHR-10_dataset_coco

DIOR 数据集：包含 20 类遥感目标的 23463 张图像和 190288 个实例标注。下载地址：<https://gcheng-nwpu.github.io/#Datasets>

iSAID 数据集：大规模遥感实例分割数据集，包含 15 类目标和 655451 个实例。下载地址：<https://captain-whu.github.io/iSAID/>

Grounded SAM 2：<https://github.com/IDEA-Research/Grounded-SAM-2>



遥感图像检测效果示意图

参考文献:

- [1] Kirillov A, Mintun E, Ravi N, et al. Segment Anything. ICCV, 2023.
- [2] Liu S, Zeng Z, Ren T, et al. Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection. ECCV, 2024.
- [3] Ren T, Jiang Q, Liu S, et al. Grounding DINO 1.5: Advance the "Edge" of Open-Set Object Detection. arXiv:2405.10300, 2024.
- [4] Ravi N, Gabeur V, Hu Y-T, et al. SAM 2: Segment Anything in Images and Videos. arXiv:2408.00714, 2024.
- [5] Chen K, Liu C, Chen H, et al. RSPrompter: Learning to Prompt for Remote Sensing Instance Segmentation Based on Visual Foundation Model. IEEE TGRS, 2024.

指导老师: 杨卫东, yangwd@hust.edu.cn, qq14223185.

4.33 图像线特征提取与匹配

任务描述: 图像匹配在图像检索、无人飞行器导航及遥感测绘领域应用广泛。道路网是遥感图像中常见的元素之一,可用于飞行器的目标识别、跟踪和定位校准等应用中。与图像相比,矢量图存储更高效,与道路网表征契合,适合大范围内的飞行器导航应用。利用样本数据,构建路网等线特征提取网络,设计相似性度量计算准则,提高特征匹配准确率,分析特征的有效性和场景适应性。

数据: 红外、可见光图对数据集。



线特征提取与匹配结果

参考文献:

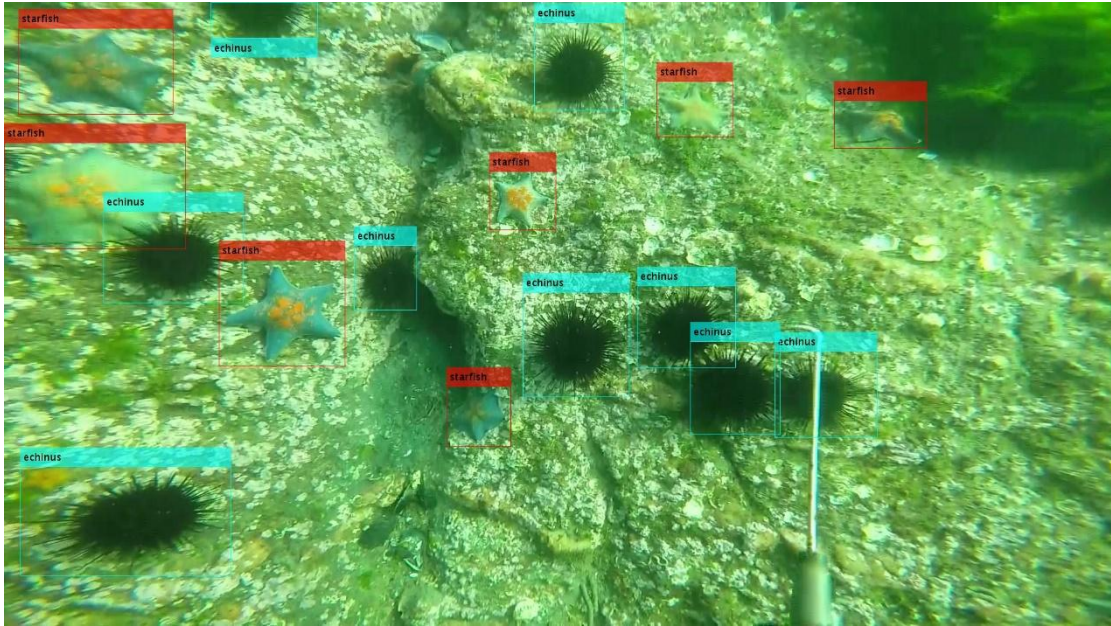
- [1] Xu Z, Liu Y, Gan L, et al. RNGDet: Road Network Graph Detection by Transformer in Aerial Images. IEEE TGRS, 2022, 60: 1-12.
- [2] Pautrat R, Lin J-T, Larsson V, et al. DeepLSD: Line Segment Detection and Refinement with Deep Image Gradients. CVPR, 2023.
- [3] Xu Z, Liu Y, Gan L, et al. RNGDet++: Road Network Graph Detection by Transformer with Instance Segmentation and Multi-scale Features. IEEE TGRS, 2024.
- [4] Pautrat R, Barath D, Larsson V, et al. GlueStick: Robust Image Matching by Sticking Points and Lines Together. ICCV, 2023.
- [5] Lindenberger P, Sarlin P-E, Pollefeys M. LightGlue: Local Feature Matching at Light Speed. ICCV, 2023.

指导老师: 杨卫东, yangwd@hust.edu.cn, qq14223185.

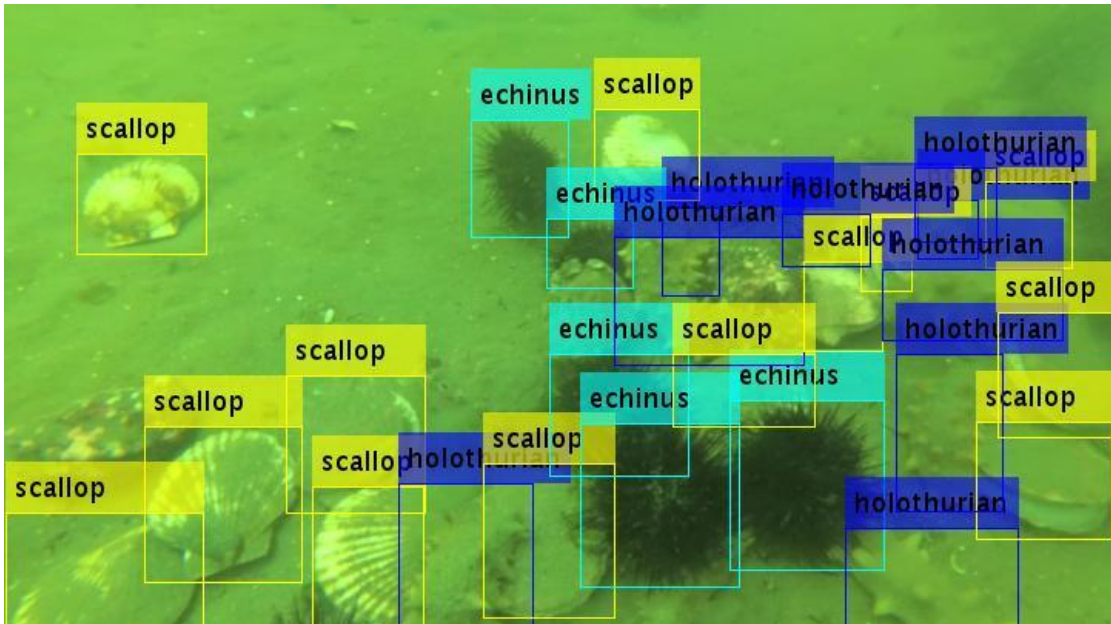
4.34 海底目标检测与识别

任务描述: 在真实海底图片数据中检测出不同目标(海参、海胆、扇贝、海星)的位置。

预期效果如下图所示:



海底检测图



海底目标检测图

数据说明：数据集由 5000 张训练图像以及 2600 张测试图像组成；这些图片之间不存在帧间连续性。图片路径示例如下：`train/image/ c000001.jpg`，其对应的目标检测标注真值位于路径 `train/box/ c000001.xml` 文件中，该文件包含了对应图像中所有物体的类别以及目标框参数（位置和尺寸）。本届比赛需检测的目标类别包括海参“holothurian”，海胆“echinus”，扇贝“scallop”和海星“starfish”四类。训练数据真值中可能存在水草“waterweeds”这一类别，请忽略这一类。

数据集下载地址：<https://pan.baidu.com/s/1gQyf2F0FhtjstI0gg44n-Q?pwd=1dkg>

提取码：1dkg

注意：若电脑性能有限，可以从训练集和测试集中抽取部分数据（例如 2000 张训练集，

1500 张测试集) 进行实验。

关于标注:

1、训练图片中存在一定的标注噪声, 如有需要选手可自行清洗数据。其中, 前缀 c 为基本干净标注数据, 前缀 u 为含噪声标注数据。

2、关于扇贝的标注歧义性较大, 测试评估时基本以“前缀 c 中的扇贝标注标准”为准。
(必要时可忽略前缀为 u 的标注数据中的“扇贝”标注结果)

参考资料:

水下目标检测介绍: https://blog.csdn.net/qq_41088475/article/details/106764923

YOLOv1 原理: <https://zhuanlan.zhihu.com/p/70387154>

YOLOv1-YOLOv4 发展: <https://zhuanlan.zhihu.com/p/139128562>

YOLOv5 原理: <https://blog.csdn.net/WZZ18191171661/article/details/113789486>

YOLOv5 训练自己的数据集: <https://blog.csdn.net/oJiWuXuan/article/details/107558286>

特别注意: 本课设很可能用到 NVIDIA GPU (显存不小于 6G)。请检查电脑或者服务器是否配备独立显卡以及显卡显存大小, 建议熟悉 Anaconda 安装 Pytorch 和其它环境的流程, 不满足实验条件的同学请谨慎选择。

参考文献:

[1] YOLOv5 code : <https://github.com/ultralytics/yolov5>

[2] [Review On Deep Learning Technique For Underwater Object Detection ,](https://arxiv.org/abs/2209.10151)
(<https://arxiv.org/abs/2209.10151>)

[3] [SWIPENET: Object detection in noisy underwater images ,](https://arxiv.org/abs/2010.10006)
(<https://arxiv.org/abs/2010.10006>)

[4] RoIMix: Proposal-Fusion among Multiple Images for Underwater Object Detection ,
(<https://arxiv.org/abs/1911.03029>)

[5] A high-precision underwater object detection based on joint self-supervised deblurring and improved spatial transformer network, (<https://arxiv.org/abs/2203.04822>)

[6] Achieving Domain Generalization in Underwater Object Detection by Domain Mixup and Contrastive Learning, (<https://arxiv.org/abs/2104.02230>)

指导老师: 邹腊梅, zlmhust@163.com, qq: [156685941](https://www.qq.com)

4.35 遥感图像中的目标检测

任务描述： 实现对于遥感图像中的目标进行自动检测，并标记出其具体区域。

数据说明： LEVIR 数据集由 800×600 像素和 0.2m-1.0m /像素的高分辨率 Google Earth 图像组成，所有图像总共标记了 11k 个独立边界框，包括 4724 架飞机，3025 艘船和 3279 个油罐。【若训练时间过久，可以选择只检测某一类目标】

参考资料：

数据下载地址：<http://pan.baidu.com/s/1geTwAVD>

YOLOv1 原理：<https://zhuanlan.zhihu.com/p/70387154>

YOLOv1-YOLOv4 发展：<https://zhuanlan.zhihu.com/p/139128562>

YOLOv5 原理：<https://blog.csdn.net/WZZ18191171661/article/details/113789486>

YOLOv5 训练自己的数据集：<https://blog.csdn.net/oJiWuXuan/article/details/107558286>

预期效果如下图所示：



检测结果示意图

特别注意： 本课设需用到 NVIDIA GPU（显存不小于 3G）。建议笔记本或台式机的 GPU

性能够用，建议熟悉 Anaconda 安装 Pytorch 和其它环境的流程，不满足实验条件的同学请谨慎选择。

参考文献：

[1] YOLOv5 code : <https://github.com/ultralytics/yolov5>

[2] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[3] Van Etten A. You only look twice: Rapid multi-scale object detection in satellite imagery[J]. arXiv preprint arXiv:1805.09512, 2018.

[4] Zou Z, Shi Z. Random access memories: A new paradigm for target detection in high resolution aerial remote sensing images[J]. IEEE Transactions on Image Processing, 2017, 27(3): 1100-1111.

指导老师: 邹腊梅, zlmhust@163.com, qq: 156685941

4.36 密集场景下的行人跟踪

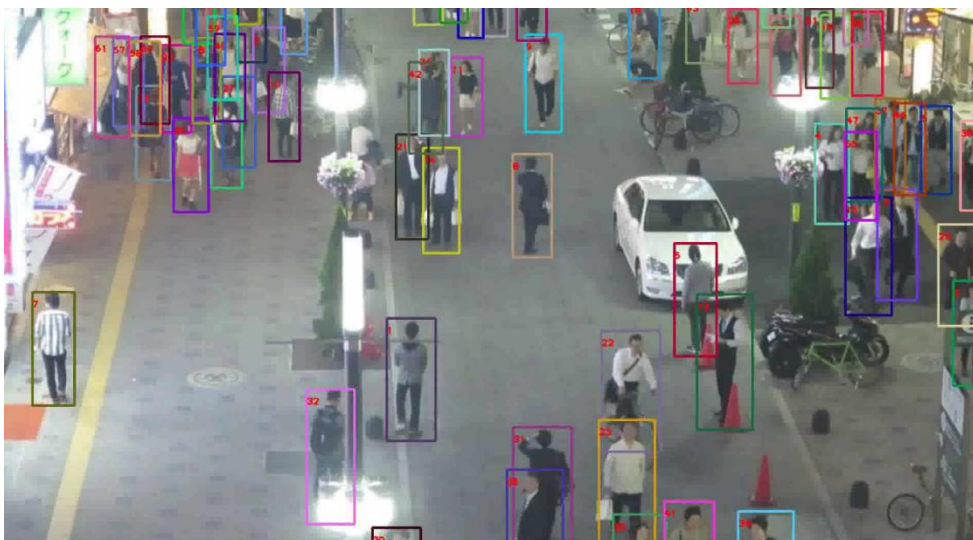
任务描述：利用深度学习算法实现复杂环境下拥挤的任务目标定位，并且能够准确跟踪人物对象。

数据说明：MOT15 数据集是多目标跟踪的基准，旨在评估算法在复杂环境中同时追踪多个目标（如行人）的能力，包含 jpeg 格式的图像和 CSV 格式的目标信息与轨迹信息标注文件。每个目标由 10 个数值描述，包括帧数、轨迹 ID、bounding box 坐标、置信度分数和 3D 位置。ground truth 文件使用置信度分数（第 7 个值）来决定是否考虑某个目标，值为 0 表示忽略，值为 1 表示活动目标。

数据集下载：

[MOT Challenge - Data](https://motchallenge.net/data/MOT15/) (<https://motchallenge.net/data/MOT15/>)

预期效果如下图所示：



特别注意：本课题会用到 NVIDIA GPU。建议笔记本或台式机的 GPU 性能够用，建议熟悉 Anaconda 安装 Pytorch 和其它环境的流程，不满足实验条件的同学请谨慎选择。

参考文献：

- [1] Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, and Wenyu Liu. Fairmot: On the fairness of detection and re-identification in multiple object tracking. *International Journal of Computer Vision*, 129:3069 – 3087, 2021.
- [2] Chao Liang, Zhipeng Zhang, Xue Zhou, Bing Li, Shuyuan Zhu, and Weiming Hu. Rethinking the competition between detection and reid in multiobject tracking. *IEEE Transactions on Image Processing*, 31:3182 – 3196, 2022
- [3] Xiao D ,Wang Z ,Shen Z , et al. A FairMOT approach based on video recognition for real-time automatic incident detection on expressways[J]. *Signal, Image and Video*

Processing, 2024, 18(10): 7333-7348.

- [4] Dong Q ,Zhangang W .Improved FairMOT multi-target tracking method for complex scenes[J].Journal of Physics: Conference Series,2022,2303(1):

其它相关参考资料:

- [1] <https://github.com/ifzhang/FairMOT>
- [2] <https://arxiv.org/abs/2004.01888>
- [3] <https://arxiv.org/abs/2010.12138>
- [4] <https://zhuanlan.zhihu.com/p/345818662>
- [5] https://blog.csdn.net/zhangzhao147/article/details/137052708?ops_request_misc=&request_id=&biz_id=102&utm_term=fairmot&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-3-137052708.142^v102^pc_search_result_base2&spm=1018.2226.3001.4187
- [6] https://blog.csdn.net/maohule/article/details/133710512?ops_request_misc=&request_id=&biz_id=102&utm_term=fairmot&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-4-133710512.142^v102^pc_search_result_base2&spm=1018.2226.3001.4187
- [7] https://blog.csdn.net/mzpmzk/article/details/126186658?ops_request_misc=%257B%2522request%255Fid%2522%253A%2522be60306d473b8252c24fb060f1994f59%2522%252C%2522scm%2522%253A%252220140713.130102334..%2522%257D&request_id=be60306d473b8252c24fb060f1994f59&biz_id=0&utm_medium=distribute.pc_search_result.none-task-blog-2~all~baidu_landing_v2~default-5-126186658-null-null.142^v102^pc_search_result_base2&utm_term=fairmot&spm=1018.2226.3001.4187

指导老师: 邹腊梅, zlmhust@163.com, qq: 156685941

4.37 基于轻量化网络的麦穗目标检测

任务描述: 实现对于小麦图像中的麦穗目标进行自动检测, 并标记出其具体区域。

数据说明: 数据集由 3432 张图像组成, 只包含“麦穗”一个类别, 图片大小为 1024×1024。

数据下载地址: <https://www.kaggle.com/c/global-wheat-detection/data>

MobilenetV1-V3: <https://zhuanlan.zhihu.com/p/70703846>

YOLOv1 原理: <https://zhuanlan.zhihu.com/p/70387154>

YOLOv1-YOLOv4 发展: <https://zhuanlan.zhihu.com/p/139128562>

YOLOv5 原理: <https://blog.csdn.net/WZZ18191171661/article/details/113789486>

YOLOv5 训练自己的数据集: <https://blog.csdn.net/oJiWuXuan/article/details/107558286>

YOLOv5 修改网络结构: <https://blog.csdn.net/waltzy/article/details/114492726>

注意事项: 下载的文件中包含 train.zip(训练图像)和 test.zip (测试图像), 由于 test.zip 没有标签, 所以本次课设只使用 train.zip 中的数据。

首先对 train 文件夹下面图片根据名称进行排序(第一张图片为: 00b5c6764.jpg), 然后从第一张图片开始选取 700 张图片(第 700 张图片为: 7fcfe0ae1.jpg)作为测试集(即这 700 张图片不参与训练, 只作为训练完成后测试模型准确度的数据); 剩余图片作为训练集。

评判标准: 使用 mAP (mean Average Precision) 作为评判标准。

模型参考: 可以考虑使用 Mobilenet-V3 作为特征提取的 Backbone, 使用 YOLOv5 的检测部分输出具体定位框。

预期效果如下图所示:



果示意图

特别注意: 本课设很可能用到 NVIDIA GPU (显存不小于 4G)。建议笔记本或台式机的 GPU 性能够用, 建议熟悉 Anaconda 安装 Pytorch 和其它环境的流程, 不满足实验条件的同学请谨慎选择。

参考文献:

[1] YOLOv5 code : <https://github.com/ultralytics/yolov5>

[2] David E, Madec S, Sadeghi-Tehran P, et al. Global Wheat Head Detection (GWHD) dataset: a large and diverse dataset of high-resolution RGB-labelled images to develop and benchmark wheat head detection methods[J]. Plant Phenomics, 2020.

[3] Howard A, Sandler M, Chu G, et al. Searching for mobilenetv3[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 1314-1324.

指导老师: 邹腊梅, zlmhust@163.com, qq: 156685941

4.38 物体 6D 姿态估计

任务描述: 利用深度学习算法实现对图像中目标物体的 6D 姿态估计, 即估计物体相对于相机的姿态变换矩阵, 从而获取目标物体的位置及姿态信息。

数据说明: 数据集 LM-O 为 BOP 的七个常用数据集之一。LM-O 中包含 8 个常见工业物体 (如胶水罐、电源适配器等)。LM-O 专注于解决物体在严重遮挡场景下的姿态估计问题, 它是对原始 LineMOD 数据集的扩展, 增加了复杂的遮挡条件, 更贴近真实世界的应用场景。

LM-O 数据集中每个类别都有约 1300 个训练数据 (无遮挡场景, 用于模型预训练), 其中训练数据包括以下几项: RGB 图像: 分辨率为 640x480; 深度图: 与 RGB 对齐的深度信息 (可选); 姿态标注: 每个物体的 6D 姿态 (3D 平移 + 3D 旋转矩阵); 遮挡标注: 目标物体被遮挡的像素级掩码。

而测试集约有 200 张图像 (含复杂遮挡, 用于评估), 其中每张图像中的目标物体都有独立的真值标注。

官方数据集地址: [BOP: Benchmark for 6D Object Pose Estimation](#)

(<https://bop.felk.cvut.cz/datasets/>)

数据集官方说明: [项目文件预览 - bop_toolkit:A Python toolkit of the BOP benchmark for 6D object pose estimation. - GitCode](#)

LM-O 数据集下载地址: [bop-benchmark/lmo at main](#)

预期效果如下图:



特别注意：本课设会用到 NVIDIA GPU。建议笔记本或台式机的 GPU 性能够用，建议熟悉 Anaconda 安装 Pytorch 和其它环境的流程，不满足实验条件的同学请谨慎选择。

参考文献：

- [1] Wen B, Yang W, Kautz J, et al. Foundationpose: Unified 6d pose estimation and tracking of novel objects[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 17868-17879.
- [2] Lin J, Liu L, Lu D, et al. Sam-6d: Segment anything model meets zero-shot 6d object pose estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 27906-27916.
- [3] Li Y, Mao Y, Bala R, et al. Mrc-net: 6-dof pose estimation with multiscale residual correlation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 10476-10486.
- [4] Nguyen V N, Groueix T, Salzmann M, et al. Gigapose: Fast and robust novel object pose estimation via one correspondence[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 9903-9913.

指导老师：邹腊梅, zlmhust@163.com, [qq: 156685941](https://www.qq.com/)

4.39 人脸表情识别

任务描述：利用深度学习方法对人脸图像中的表情特征进行建模与分析，实现人脸表情的自动识别，并输出对应的表情类别。

数据说明：RAF-DB 是一个常用的人脸表情识别数据集，包含约 30,000 张从互联网收集的具有丰富的场景变化（如姿态、光照、遮挡及人群属性差异等）的人脸图像。RAF-DB 包含两个子集：基本表情子集（包含 7 类基本情绪：惊讶、恐惧、厌恶、高兴、悲伤、愤怒和中性）、复合表情子集（包含 12 类复合情绪）。数据集图片示例：



在实际研究中，通常在基本表情子集上进行性能对比。基本表情子集包含 12,271 张训练样本和 3,068 张测试样本。除表情标签外，每张图像还提供：5 个精确关键点位置、37 个自动检测关键点位置、人脸边界框、种族、年龄范围和性别属性标注。可根据需求，灵活利用上述标注信息进行算法设计与模型优化。

RAF-DB 数据集下载地址： <http://www.whdeng.cn/RAF/model1.html>（官方地址）

<https://aistudio.baidu.com/datasetdetail/48204>（非官方地址）

另外，还有 **AffNet 数据集**可作为验证或补充的数据集。

AffNet 数据集介绍及下载地址：<https://blog.csdn.net/SneakateRter/article/details/154121745>

特别注意：本课题会用到 NVIDIA GPU。建议笔记本或台式机的 GPU 性能够用，建议熟悉 Anaconda 安装 Pytorch 和其它环境的流程，不满足实验条件的同学请谨慎选择。

参考文献：

- [1] Mao J, Xu R, Yin X, et al. Poster++: A simpler and stronger facial expression recognition network[J]. Pattern Recognition, 2025, 157: 110951.
- [2] Wang C, Chen L, Wang L, et al. QCS: Feature refining from quadruplet cross similarity for facial expression recognition[C]//Proceedings of the AAAI conference on artificial intelligence. 2025, 39(7): 7563-7572.

[3] Zhao Z, Cao Y, Gong S, et al. Enhancing zero-shot facial expression recognition by llm knowledge transfer[C]//2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). IEEE, 2025: 815-824.

[4] So J, Han Y. Facial Landmark-Driven Keypoint Feature Extraction for Robust Facial Expression Recognition[J]. Sensors, 2025, 25(12): 3762.

指导老师：邹腊梅，zlmhust@163.com qq: 156685941

4.40 3D 点云物体缺陷检测

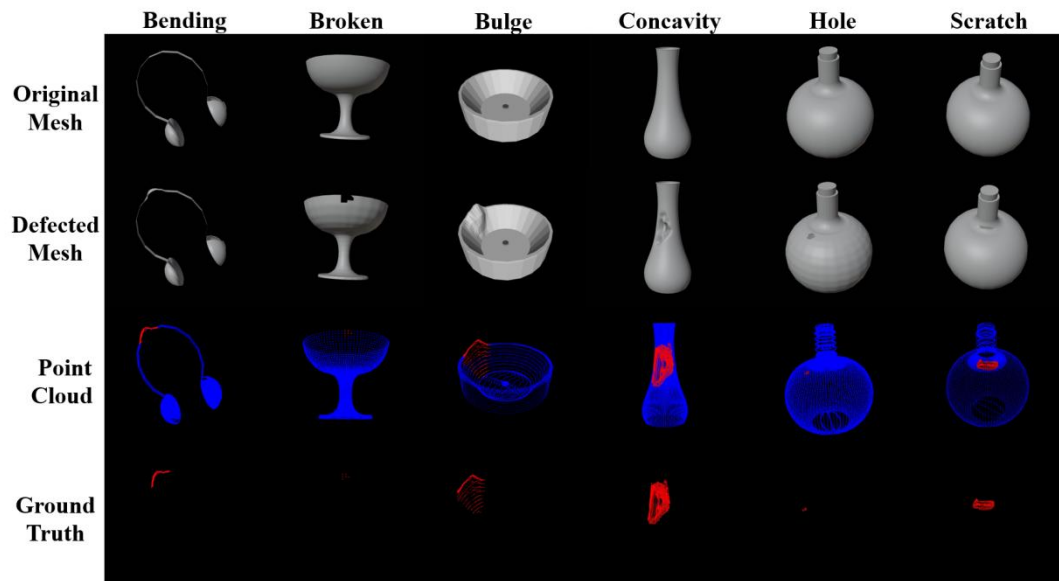
任务描述：利用深度学习算法实现对 3D 点云中的缺陷和异常进行像素级/对象级的异常/缺陷检测定位。通过构建高效的三维特征提取网络，提升复杂场景下物体异常/缺陷的识别精度与鲁棒性。

数据说明：Anomaly-ShapeNet 是专为 3D 点云异常检测任务构建的合成基准数据集，基于 ShapeNetCoreV2 子集构建，覆盖丰富类别与多样异常形态。其构建流程分为三步：先对 ShapeNet 原始正常模型进行网格细分以优化拓扑结构，再通过工业设计软件 Blender 雕刻各类真实工业缺陷生成异常样本，最后借助 CloudCompare 精准生成逐像素级与对象级的异常标注。

该数据集训练集包含 160 个样本，点数范围为 17422-157824，单样本平均点数为 55429.26，每类别包含 4 个正常样本。测试集包含 1312 个样本，点数范围为 17640-152842，单样本平均点数为 57035.88，每类别包含 28-40 个覆盖正常与异常的实例。

数据集地址：<https://github.com/Chopper-233/AnomalyShapeNet>

数据集图片示例：



特别注意： 本课设会用到 NVIDIA GPU。建议笔记本或台式机的 GPU 性能够用，建议熟悉 Anaconda 安装 Pytorch 和其它环境的流程，不满足实验条件的同学请谨慎选择。

参考文献：

- [1] Li W, Xu X, Gu Y, et al. Towards scalable 3d anomaly detection and localization: A benchmark via 3d anomaly synthesis and a self-supervised learning network[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024: 22207-22216.
- [2] Zha Y, Yuerong X, Fan C, et al. Casl: Curvature-augmented self-supervised learning for 3d anomaly detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2026, 40(15): 12340-12348.
- [3] Cheng Y, Cao Y, Wang D, et al. Boosting global-local feature matching via anomaly synthesis for multi-class point cloud anomaly detection[J]. IEEE Transactions on Automation Science and Engineering, 2025.
- [4] Ye J, Zhao W, Yang X, et al. Po3ad: Predicting point offsets toward better 3d point cloud anomaly detection[C]//Proceedings of the Computer Vision and Pattern Recognition Conference. 2025: 1353-1362.
- [5] Zhou Z, Wang L, Fang N, et al. R3d-ad: Reconstruction via diffusion for 3d anomaly detection[C]//European conference on computer vision. Cham: Springer Nature Switzerland, 2024: 91-107.

指导老师：邹腊梅， zlmhust@163.com qq: 156685941

4.41 无人机视角下的目标检测与识别

任务描述：任务描述：无人机已广泛应用于各种应用，例如监视、自主检测、车队导航和农业。尽管自然图像如 COCO 数据集上的目标检测取得了显著进展，但检测无人机视角图像中的目标仍然具有挑战性，这主要源于小尺度和极端尺度变化。本课题基于深度学习方法，实现对于无人机航拍视角下的图片进行目标检测，能够区分不同类别目标并给出目标的准确位置。

数据说明：数据集为 VisDrone2019 数据集，其中图像分辨率为 2000*1500，包含 6471 张训练图像，548 张验证图像和 1610 张测试图像，且都具有标签文件。数据集包含十类不同类型目标如车辆，行人等等。

VisDrone 数据集：GitHub - VisDrone/VisDrone-Dataset: The dataset for drone based detection and tracking is released, including both image/video, and annotations.

VisDrone 数据集（COCO 格式）：GitHub - PuAnysh/UFPMP-Det: The official implementation of UFPMP-Det

YOLOv8 原理：<https://zhuanlan.zhihu.com/p/668516241>

YOLOv8 训练 VisDrone 数据集：<https://dhexx.cn/news/show5056287.html?action=onClick>

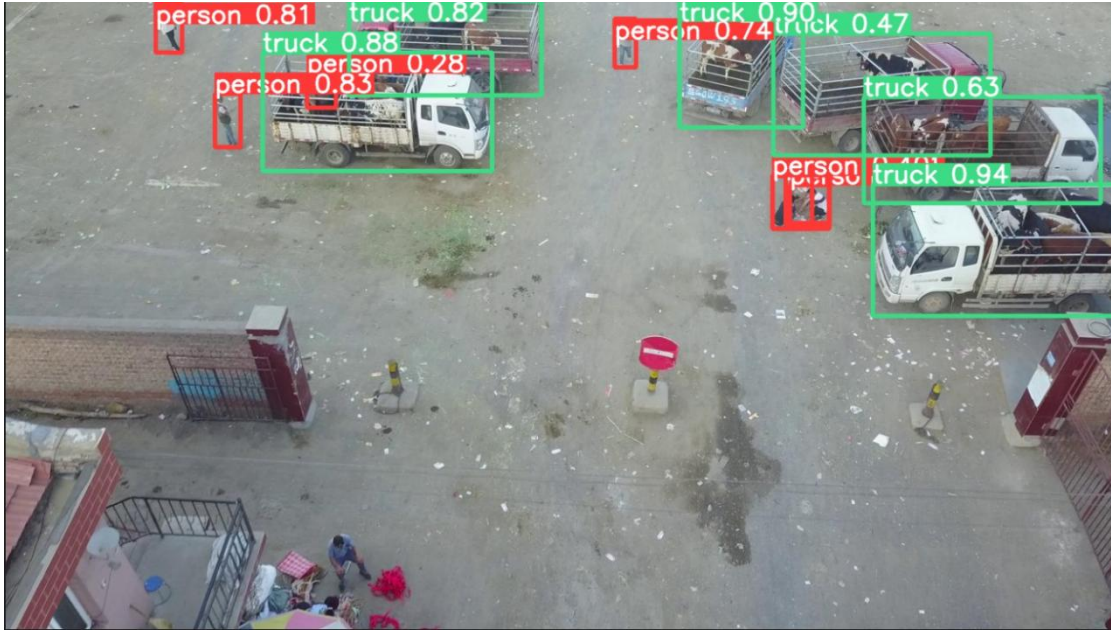
YOLOv8 修改网络结构：<https://blog.csdn.net/wahaha987656789/article/details/134141589>

注意事项：下载的文件中包含四个压缩文件，除了 VisDrone2019-DET-test-challenge.zip 外都具有标签。其他标签格式数据集也提供了下载方式。

注意，虽然测试集也有标签文件，但是禁止将其加入训练或验证中，否则最后测试较为麻烦。

评判标准：使用 mAP（mean Average Precision）作为评判标准。

预期检测效果如下图所示：



检测结果示意图

特别注意：本课设需用到 NVIDIA GPU（显存不小于 4G）。建议笔记本或台式机的 GPU 性能够用，建议熟悉 Anaconda 安装 Pytorch 和其它环境的流程，不满足实验条件的同学请谨慎选择。

参考文献：

- [1] YOLOv8 code : <https://github.com/ultralytics/ultralytics>
- [2] Zhao L L, Zhu M L. MS-YOLOv7: YOLOv7 Based on Multi-Scale for Object Detection on UAV Aerial Photography[J]. Drones, 2023, 7(3): 188.
- [3] Du D, Zhu P, Wen L, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results[C]//Proceedings of the IEEE/CVF international conference on computer vision workshops. 2019: 0-0.
- [4] Huang Y, Chen J, Huang D. UFPMP-Det:Toward Accurate and Efficient Object Detection on Drone Imagery [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(1): 1026–1033.
- [5] Wang G, Chen Y, An P, et al. UAV-YOLOv8: a small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J]. Sensors, 2023, 23(16): 7190.

指导老师: 邹腊梅, zlmhust@163.com, qq: 156685941

4.42 基于深度学习的鸟声识别

任务描述: 设计合适的深度学习网络, 实现从音频录音中辨识鸟类个体及其物种, 利用降噪、数据增强等方法提高识别性能, 并与其他文献方法的性能进行比较分析。

数据说明: BirdCLEF 2023 公开数据集, 训练数据由 xenocanto.org 网站的用户慷慨上传的单个鸟类鸣叫的短音频组成。这些文件已被下采样至 32 千赫兹, 并转换为 ogg 格式。train_metadata.csv 文件中为训练数据提供了广泛的元数据, 最直接相关的字段包括: 1) primary_label: 鸟类物种的代码。可以通过将代码添加到

<https://ebird.org/species/> 来查看有关鸟类代码的详细信息, 例如对于美洲乌鸦 (American Crow), 链接为 <https://ebird.org/species/amecro>。2) latitude & longitude: 录制音频的经纬度坐标。有些鸟类物种可能有当地的鸣叫 “方言”, 所以您可能希望在训练数据中寻找地理上的多样性。

3) author: 提供音频记录的用户。4) filename: 相关音频文件的名称。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	primary_label	secondary_label	type	latitude	longitude	scientific_name	common_name	author	license	rating	url	filename		
2	abethr1	[]	[song]	4.3906	38.2788	Turdus	teq:African	Bai	Rolf A. de	Creative C	4	https://www.abethr1/XC128013.ogg		
3	abethr1	[]	[call]	-2.9524	38.2921	Turdus	teq:African	Bai	James Brac	Creative C	3.5	https://www.abethr1/XC363501.ogg		
4	abethr1	[]	[song]	-2.9524	38.2921	Turdus	teq:African	Bai	James Brac	Creative C	3.5	https://www.abethr1/XC363502.ogg		
5	abethr1	[]	[song]	-2.9524	38.2921	Turdus	teq:African	Bai	James Brac	Creative C	5	https://www.abethr1/XC363503.ogg		
6	abethr1	[]	[call, 'song]	-2.9524	38.2921	Turdus	teq:African	Bai	James Brac	Creative C	4.5	https://www.abethr1/XC363504.ogg		
7	abethr1	[rbsrob1]	[song]	-2.9524	38.2921	Turdus	teq:African	Bai	James Brac	Creative C	3.5	https://www.abethr1/XC379322.ogg		
8	abethr1	[]	[call, 'song]	-2.9965	37.6244	Turdus	teq:African	Bai	isaac kilust	Creative C	3	https://www.abethr1/XC432639.ogg		
9	abethr1	[]	[song]	-4.0904	37.8807	Turdus	teq:African	Bai	Peter Erics	Creative C	5	https://www.abethr1/XC467121.ogg		
10	abethr1	[]	[song]	-4.0904	37.8807	Turdus	teq:African	Bai	Peter Erics	Creative C	5	https://www.abethr1/XC467122.ogg		

下载链接: [BirdCLEF 2023 | Kaggle](#) 或 [BirdCLEF 2023_数据集-飞桨AI Studio星河社区](#)

参考文献:

- [1] Daidai Liu, Hanguang Xiao, Kai Chen. Research progress in bird sounds recognition based on acoustic monitoring technology: A systematic review, *Applied Acoustics*, 2025(228): 110285.
- [2] René Heinrich, Lukas Rauch, Bernhard Sick, et al. AudioProtoPNet: An interpretable deep learning model for bird sound classification. *Ecological Informatics*, 2025(87), 10381.
- [3] Xie JJ, Zhong YJ, Zhang JG, et al. A review of automatic recognition technology for bird vocalizations in the deep learning era. *Ecological Informatics*, 2023(73), 101927.
- [4] Carvalho S, Gomes EF. Automatic classification of bird sounds: Using MFCC and Mel spectrogram features with deep learning. *Vietnam Journal of Computer Science*, 2023(10), 39–54.
- [5] Tang Q, Xu LM, Zheng BC, et al. Transound: Hyper-head attention transformer for birds sound recognition. *Ecological Informatics*, 2023(75), 102001.

[6] Zabidi MM, Wong KL, Sheikh UU, et al. Bird sound detection with binarized neural networks. Journal of Electrical Technology, 2022(21), 48–53.

[7] Zhang CY, Chen YH, Hao ZZ, et al. An efficient time-domain end-to-end single-channel bird sound separation network. Animals, 2022(12), 3117.

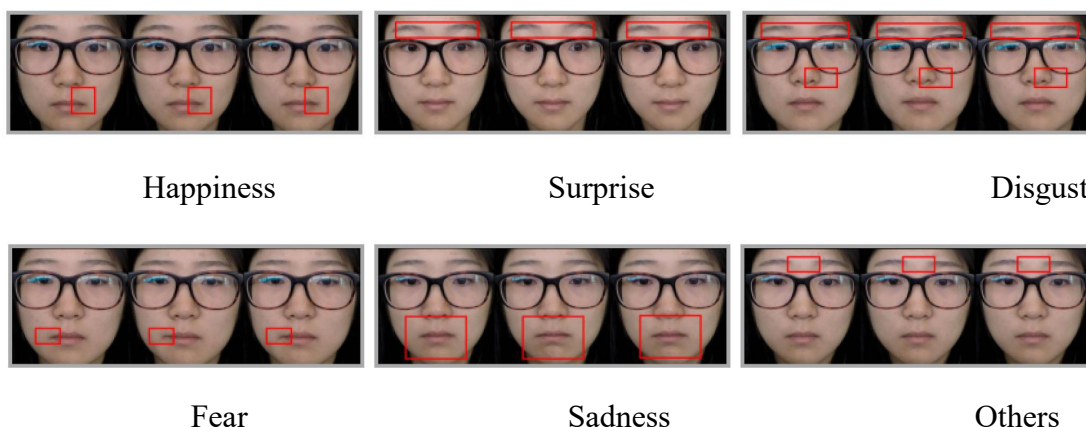
[8] Zhang FY, Zhang LY, Chen HX, et al. Bird species identification using spectrogram based on multi-channel fusion of DCNNs. Entropy, 2021(23), 1507–1518.

指导老师：胡若澜 huruolan@hust.edu.cn qq:20536555

4.43 视频序列中人脸微表情识别

任务描述：选择一种方法实现对视频序列中的人脸微表情进行识别，并给出识别性能指标，与其他文献方法的性能进行比较。

数据说明：MMEW 数据集，micro-and-macro expression warehouse 2021 年发布，在实验室环境下，通过观看情感视频诱惑微表情，同时试图保持中性表情，该数据集中包含表情视频样本 300，宏表情视频样本 900，受试者 36，平均年龄 22.35，亚洲人，样本标签包含 7 种情感类型，Happiness (36), Anger (8), Surprise (89), Disgust (72), Fear (16), Sadness (13) and Others (66)，视频分辨率为 1920*1080，每秒帧数 (fps) 200，视频中人脸尺寸 400*400。下载链接 [GitHub - benxianyeteam/MMEW-Dataset: Micro-and-Macro Expression Warehouse \(MMEW\) Dataset](https://github.com/benxianyeteam/MMEW-Dataset: Micro-and-Macro Expression Warehouse (MMEW) Dataset)



MMEW 数据集示例

参考文献：

[1] M. Verma, S. K. Vipparthi and G. Singh, "Deep Insights of Learning-Based Micro Expression Recognition: A Perspective on Promises, Challenges, and Research Needs,"

in *IEEE Transactions on Cognitive and Developmental Systems*, vol. 15, no. 3, pp. 1051-1069, Sept. 2023, doi: 10.1109/TCDS.2022.3226348.

[2] X. Ben *et al.* Video-Based Facial Micro-Expression Analysis: A Survey of Datasets, Features and Algorithms[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9), 2022: 5826-5846, doi: 10.1109/TPAMI.2021.3067464

[3] Z. Shang, J. Liu and X. Li. Micro-Expression Recognition Based on Spatio–Temporal Capsule Network. *IEEE Access*, 11, 2023: 13704-13713, doi: 10.1109/ACCESS.2023.3242871.

[4] X. Cai, H. Tang and L. Chai. Micro Expression Recognition based on Graph Convolutional Networks with LSTM[C]. *2023 35th Chinese Control and Decision Conference (CCDC)*, Yichang, China, 2023: 5449-5453, doi: 10.1109/CCDC58219.2023.10327083.

[5] N. Kim, S. Cho, C. H. Ahn, et al. Facial Micro-Expression Recognition in Video using Squeezed Landmark Feature Maps[C]. *2021 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju Island, Korea, Republic of, 2021: 1107-1110, doi: 10.1109/ICTC52510.2021.9620973.

[6] S. Indolia, S. Nigam, R. Singh, et al. Micro Expression Recognition Using Convolution Patch in Vision Transformer[J]. *IEEE Access*, 11, 2023: 100495-100507.

[7] P. Gupta. MERASTC: Micro-Expression Recognition Using Effective Feature Encodings and 2D Convolutional Neural Network[J]. *IEEE Transactions on Affective Computing*, 14(2) 2023: 1431-1441, doi: 10.1109/TAFFC.2021.3061967.

[8] Y. Li, X. Huang and G. Zhao. Joint Local and Global Information Learning With Single Apex Frame Detection for Micro-Expression Recognition[J]. *IEEE Transactions on Image Processing*, 30, 2021: 249-263, doi: 10.1109/TIP.2020.3035042.

指导老师：胡若澜 huruolan@hust.edu.cn qq:20536555

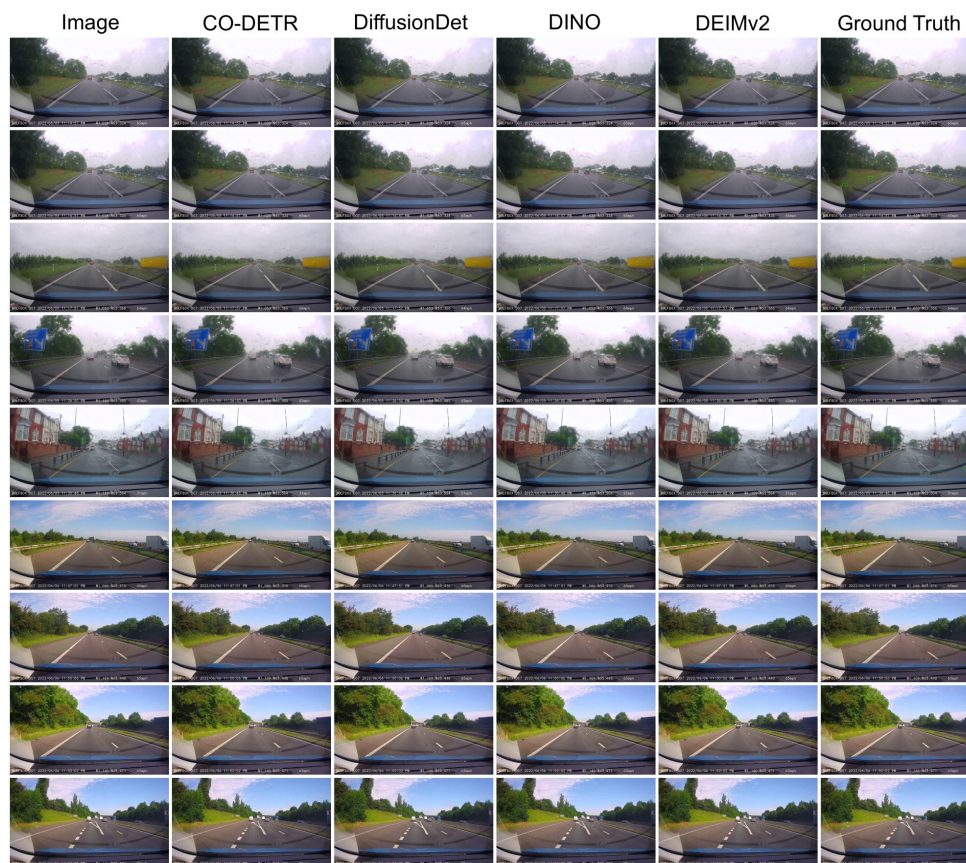
4.44 路边垃圾实时检测

任务描述：选择一种方法实现路边垃圾的实时检测，与其他文献方法的性能进行比较分析。

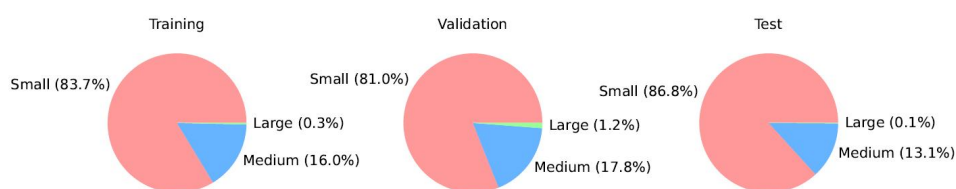
数据说明：RoLID-11K 数据集是首个基于行车记录仪的大规模路边垃圾数据集，包括

乡村道路、郊区街道、双车道主干道和城市环境，多样的天气和光照条件（晴朗、多云、低光和阴影环境）涵盖超 1.1 万标注图像，呈现显著的长尾分布和小目标分布特征。

下载链接：<https://github.com/xq141839/RoLID-11K>



RoLID-11K 数据集检测结果示例



RoLID-11K 数据集目标尺寸统计

参考文献：

- [1] Tao Wu, Qing Xu, Xiangjian He, et al. RoLID-11K: A Dashcam Dataset for Small-Object Roadside Litter Detection. arXiv:2601.00398v1 [cs.CV] 1 Jan 2026.
- [2] Juan Carlos Arbelaez-Estrada, Paola Vallejo, Jose Aguilar, et al. A systematic literature review of waste identification in automatic separation systems[J]. Recycling, 8(6), 2023.
- [3] Tianheng Cheng, Lin Song, Yixiao Ge, et al. Yolo-world: Real-time open-vocabulary object detection[C]. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2024, 16901–16911.
- [4] S. Chen, P. Sun, Y. Song and P. Luo. DiffusionDet: Diffusion Model for Object

Detection[C]. 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2023: 19773-19786.

[5] Shihua Huang, Yongjie Hou, Longfei Liu, Xuanlong Yu, and Xi Shen. Real-time object detection meets dinov3. arXiv preprint arXiv:2509.20787, 2025.

[6] Zhuofan Zong, Guanglu Song, and Yu Liu. Detsr with collaborative hybrid assignments training[C]. In Proceedings of the IEEE/CVF international conference on computer vision, 2023: 6748–6758.

[7] Shenlin Liu, Ruihan Chen, Minhua Ye, et al. EcoDetect-YOLO: A Lightweight, High-Generalization Methodology for Real-Time Detection of DomesticWaste Exposure in Intricate Environmental Landscapes[J]. Sensors 2024, 24, 4666.

[8] Jing Su, Ruihan Chen, Mingzhi Li, et al. EcoDetect-YOLOv2: A High-Performance Model for Multi-ScaleWaste Detection in Complex Surveillance Environments[J]. Sensors 2025, 25, 3451.

指导老师：胡若澜 huruolan@hust.edu.cn qq:20536555

4.45 基于图像的饮食营养评估

任务描述： 选择一种方法实现图像中的食物类别识别，同时估计食物的体积，给出图像中食品的营养评估。

数据说明： Food2K 是一个大型食品识别数据集，包含 2000 个类别和超过 100 万张图像。数据集下载链接 <http://123.57.42.89/FoodProject.html>



FOOD2K 数据集示例

参考文献:

[1] F. S. Konstantakopoulos, E. I. Georga, D. I. Fotiadis. A Review of Image-Based Food Recognition and Volume Estimation Artificial Intelligence Systems. *IEEE Reviews in Biomedical Engineering*, 17, 2024: 136-152, doi: 10.1109/RBME.2023.3283149.

[2] Weiqing Min, Zhiling Wang, Yuxin Liu, et al. Large Scale Visual Food Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 45(8), 2023: 9932-9949, DOI:10.1109/TPAMI.2023.3237871.

[3] Sergio Romero-Tapiador, Ruben Tolosana, Aythami Morales, et al. Leveraging Automatic Personalised Nutrition: Food Image Recognition Benchmark and Dataset based on Nutrition Taxonomy. 14, 2022.

[4] Chi-Sheng Chen, Guan-Ying Chen, Dong Zhou, et al. Res-VMamba: Fine-Grained Food Category Visual Classification Using Selective State Space Models with Deep Residual Learning. 24, 2024.

[5] H. He, F. Kong, J. Tan. DietCam: Multiview Food Recognition Using a Multikernel SVM[J]. *IEEE Journal of Biomedical and Health Informatics*, 20(3), 2016: 848-855, doi: 10.1109/JBHI.2015.2419251.

[6] B. Magid *et al.* CalorieMe: An Image-based Calorie Estimator System[C]. *2023 Eleventh International Conference on Intelligent Computing and Information Systems (ICICIS)*, Cairo, Egypt, 2023: 555-560, doi: 10.1109/ICICIS58388.2023.10391113.

[7] E. D. Cherpanath, P. R. Fathima Nasreen, K. Pradeep, et al. Food Image Recognition and Calorie Prediction Using Faster R-CNN and Mask R-CNN[C]. *2023 9th International Conference on Smart Computing and Communications (ICSCC)*, Kochi, Kerala, India, 2023: 83-89, doi: 10.1109/ICSCC59169.2023.10335053.

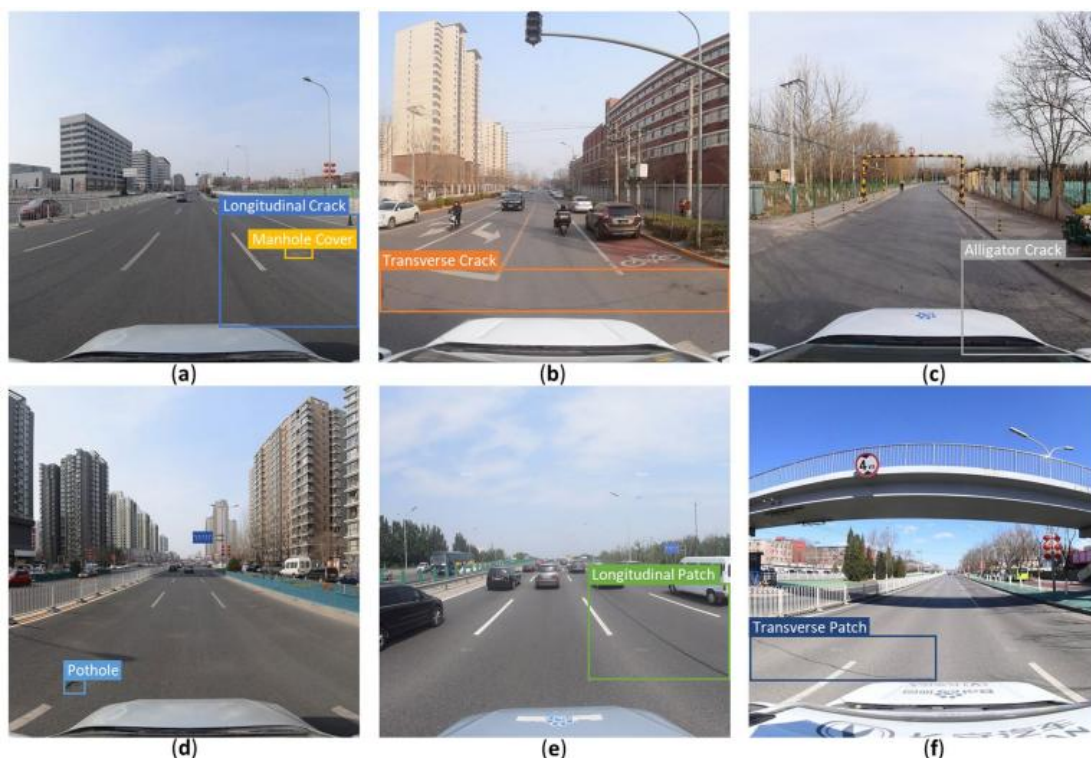
[8] K. Moumane, I. El Asri, T. Cheniguer, et al. Food Recognition and Nutrition Estimation using MobileNetV2 CNN architecture and Transfer Learning[C]. *2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA)*, Casablanca, Morocco, 2023: 1-7, doi: 10.1109/SITA60746.2023.10373725.

指导老师：胡若澜 huruolan@hust.edu.cn qq:20536555

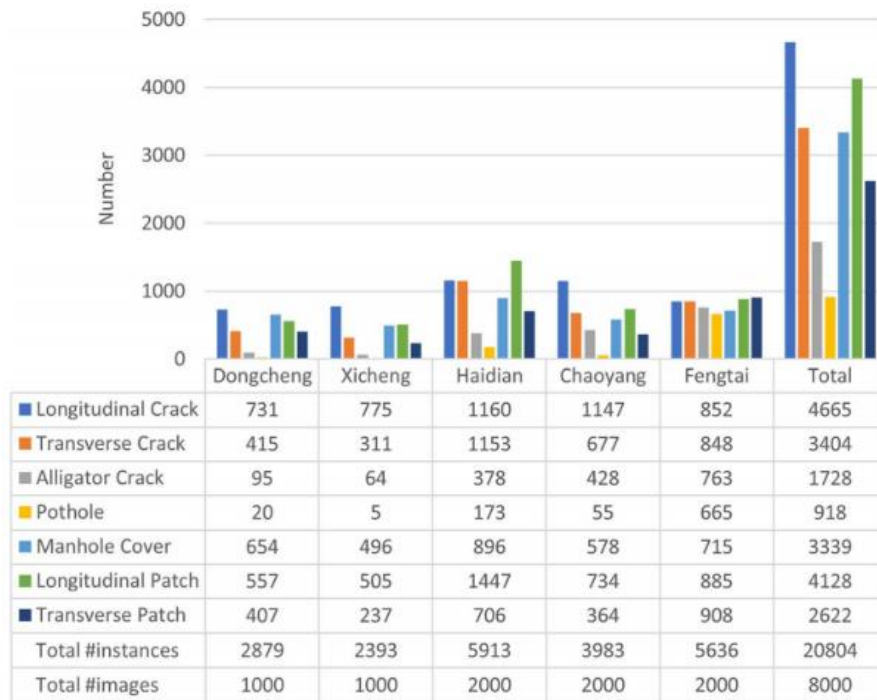
4.46 轻量化路面损伤检测

任务描述：选择一种方法实现路面损伤检测，在此基础上进行轻量化设计，实现实时路面损伤检测，与其他文献方法的性能进行比较分析。

数据说明：SVRDD 数据集，基于从百度地图获取的 8000 张街景图像图像，视觉识别并标注了 20804 个损伤实例，这些实例分布在北京市东城、西城、海淀、朝阳和丰台五个行政区，涉及的路面损伤类别包括纵向裂缝、横向裂缝、网状裂缝、坑洞、纵向修补、横向修补和井盖。数据集下载链接：<https://zenodo.org/records/10100129>



SVRDD 数据集中不同路面损伤类别示例



SVRDD 数据集中路面损伤类别及数量

参考文献:

[1] D Arya, H Maeda, SK Ghosh. From global challenges to local solutions: A review of cross-country collaborations and winning strategies in road damage detection. *Advanced Engineering Informatics*[J],60,2024,102388.

[2] Zhao, M., Su, Y., Wang, J. et al. MED-YOLOv8s: a new real-time road crack, pothole, and patch detection model[J]. *J Real-Time Image Proc* 21, 2024, 26.

[3] Zeng, J., Zhong, H. YOLOv8-PD: an improved road damage detection algorithm based on YOLOv8n model[J]. *Sci Rep* 14, 2024, 12052.

[4] Xueqiu Wang, Huanbing Gao, Zemeng Jia, Zijian Li. BL-YOLOv8: An Improved Road Defect Detection Model Based on YOLOv8[J]. *Sensors* 2023, 23(20), 8361.

[5] Miao Ren, Xiangfeng Zhng, Xiao Chen, Bo Zhou, Ziyuan Feng. Yolov5s-M: A deep learning network model for road pavement damage detection from urban street-view imagery[J]. *International Journal of Applied Earth Observation and Geoinformation*, 120, 2023, 103335.

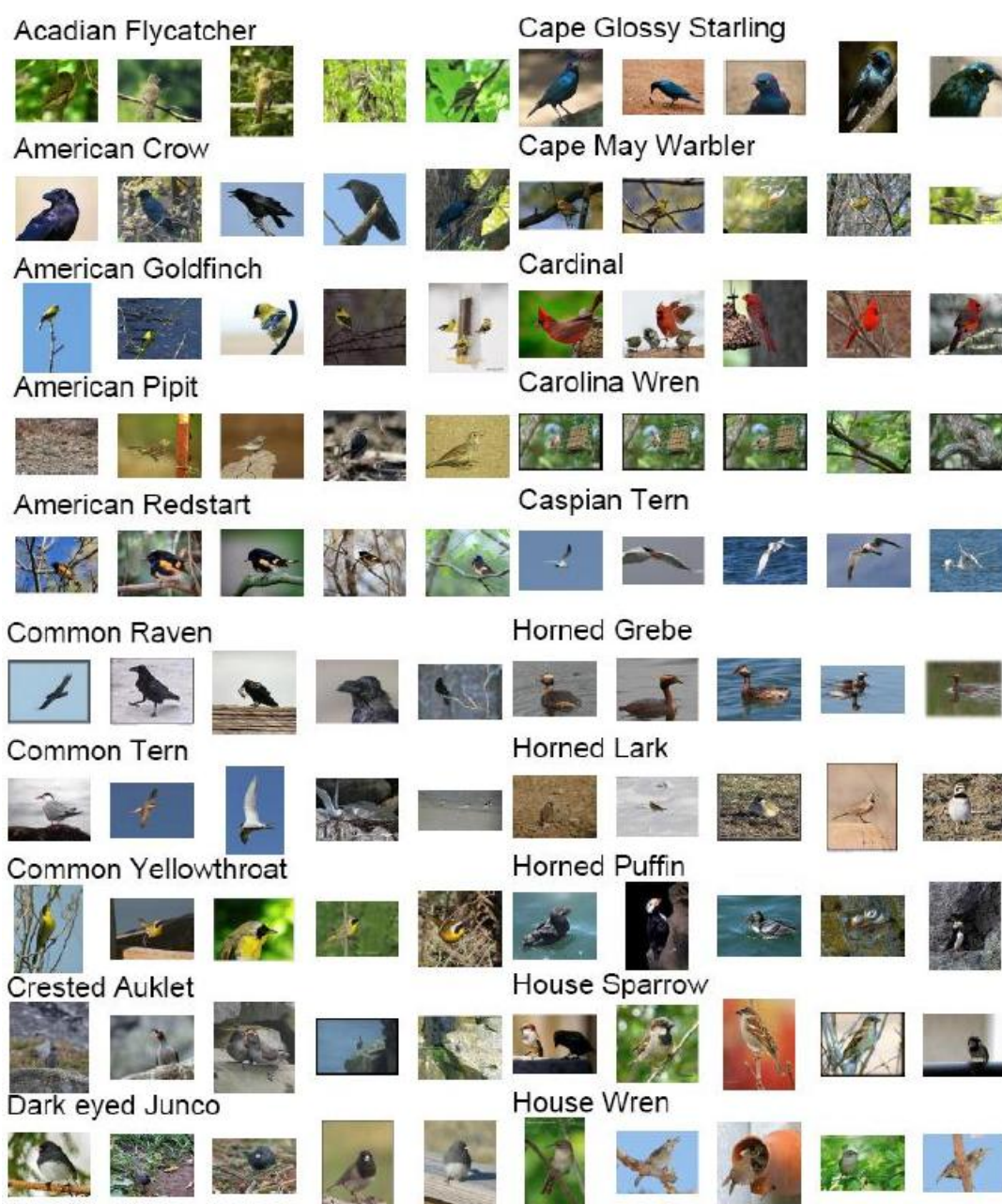
[6] Ren, M., Zhang, X., Zhi, X. et al. An annotated street view image dataset for automated road damage detection[J]. *Sci Data*,11, 2024,407.

指导老师: 胡若澜 huruolan@hust.edu.cn qq:20536555

4.47 基于结构信息挖掘的鸟类识别

任务描述: 设计一种方法对图像中鸟的种类进行识别, 能够挖掘图像中用于细粒度视觉分类的结构信息, 提高鸟类识别性能, 并与其他文献方法的性能进行对比。

数据说明: CUB-200-2011 数据集, 加州理工大学与 2010 年提出的公开鸟类数据集, 也是目前细粒度图像识别领域研究的基准数据集, 该数据集包括了北美 200 种鸟类, 11788 张鸟类图像, 训练图像 5994 张, 测试图像 5794 张, 每一种鸟类 30 张左右训练数据以及测试数据, 提供了图像级别的分类标签, 鸟类边界框信息, 鸟的关键点区域以及鸟类的属性信息。下载链接 [CUB-200-2011 \(caltech.edu\)](http://caltech.edu)



CUB-200-2011 数据集示例

参考文献:

- [1] Xiu-Shen Wei, Yi-Zhe Song, Oisín Mac Aodha, et al. Fine-Grained Image Analysis with Deep Learning: A Survey. *Computer Vision and Pattern Recognition*, arXiv: 2111.06119 [cs.CV]. <https://doi.org/10.48550/arXiv.2111.06119>
- [2] Wah, C., Branson, S., Welinder, P., Perona, P., & Belongie, S. (2022). CUB-200-2011 (1.0) [Data set]. CaltechDATA. <https://doi.org/10.22002/D1.20098>
- [3] H. Sun, X. He, J. Xu and Y. Peng. SIM-OFE: Structure Information Mining and Object-Aware Feature Enhancement for Fine-Grained Visual Categorization. *IEEE Transactions on Image Processing*, 2024(33): 5312-5326.
- [4] Asish Bera, Zachary Wharton, Yonghuai Liu, et al. SR-GNN: Spatial Relation-aware Graph Neural Network for Fine-Grained Image Categorization. CVPR, 2022.
- [5] U. E. Akpudo, X. Yu, J. Zhou, Y. Gao. What EXACTLY are We Looking at?: Investigating for Discriminance in Ultra-Fine-Grained Visual Categorization Tasks[C]. *2023 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Port Macquarie, Australia, 2023: 129-136, doi: 10.1109/DICTA60407.2023.00026.
- [6] B. -S. Wang, J. -W. Hsieh, Y. -K. Hsieh, et al. COFENet: Co-Feature Neural Network Model for Fine-Grained Image Classification[C]. *2022 IEEE International Conference on Image Processing (ICIP)*, Bordeaux, France, 2022: 3928-3932, doi: 10.1109/ICIP46576.2022.9897463.
- [7] S. Li, S. Wang, Z. Dong, et al. PSBCNN: Fine-grained image classification based on pyramid convolution networks and SimAM[C]. *2022 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCCom/CyberSciTech)*, Falerna, Italy, 2022: 1-4, doi: 10.1109/DASC/PiCom/CBDCCom/Cy55231.2022.9927801.
- [8] M. Ju, H. Ryu, S. Moon, et al. GAPNet: Generic-Attribute-Pose Network For Fine-Grained Visual Categorization Using Multi-Attribute Attention Module[C]. *2020 IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates, 2020: 703-707, doi: 10.1109/ICIP40778.2020.9190875.

指导老师: 胡若澜 huruolan@hust.edu.cn qq:20536555

4.48 基于迁移学习的花卉识别

任务描述: 识别图像中的花卉类别, 针对花卉图像样本少的问题进行样本增强, 采用迁移学习方法提高模型性能, 给出所实现方法的性能分析。

数据说明: 中国植物主题数据库的五类花卉数据集, 下载链接

http://download.tensorflow.org/example_images/flower_photos.tgz



daisy



dandelion



rose



sunflowers



tulips

中国植物主题数据库的五类花卉数据集

牛津大学视觉几何组提出的 Oxford-17 flowers 数据集和 Oxford-102 flowers 数据集, 下载链接 [Visual Geometry Group - University of Oxford](http://www.robots.ox.ac.uk/~vgg/data/flowers/)

参考文献:

- [1] R. Xiao and R. Wang. Transfer Learning-Based Flower Image Classification: Leveraging The Pre-Trained Alexnet Model[C]. *2023 4th International Symposium on Computer Engineering and Intelligent Communications (ISCEIC)*, Nanjing, China, 2023 : 519-522, doi: 10.1109/ISCEIC59030.2023.10271209.
- [2] P. Shourie, V. Anand and S. Gupta. Flower Classification using a Transfer-based Model[C]. *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India, 2023: 1-6, doi: 10.1109/ICAAIC56838.2023.10140969.
- [3] G. Yifei, Q. Chuxian, X. Jiexiang, et al. Flower image classification based on improved convolutional neural network[C]. *2022 12th International Conference on Information Technology in Medicine and Education (ITME)*, Xiamen, China, 2022: 81-87, doi: 10.1109/ITME56794.2022.00028.
- [4] M. He, H. Zhu, Y. Li, et al. Flower Image Classification Based on Multi-scale Dense Residual Network[C]. *2021 6th International Conference on Image, Vision and Computing (ICIVC)*, Qingdao, China, 2021: 144-148, doi: 10.1109/ICIVC52351.2021.9526970.
- [5] F. Hu, F. Yao and C. Pu. Learning Salient Features for Flower Classification Using

Convolutional Neural Network[C]. *2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIS)*, Dalian, China, 2020: 476-479, doi: 10.1109/ICAIS49377.2020.9194931.

[6] F. R. Siregar and W. F. Al Maki. Hybrid Method for Flower Classification in High Intra-class Variation[C]. *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia, 2020: 73-78, doi: 10.1109/ISRITI51436.2020.9315379.

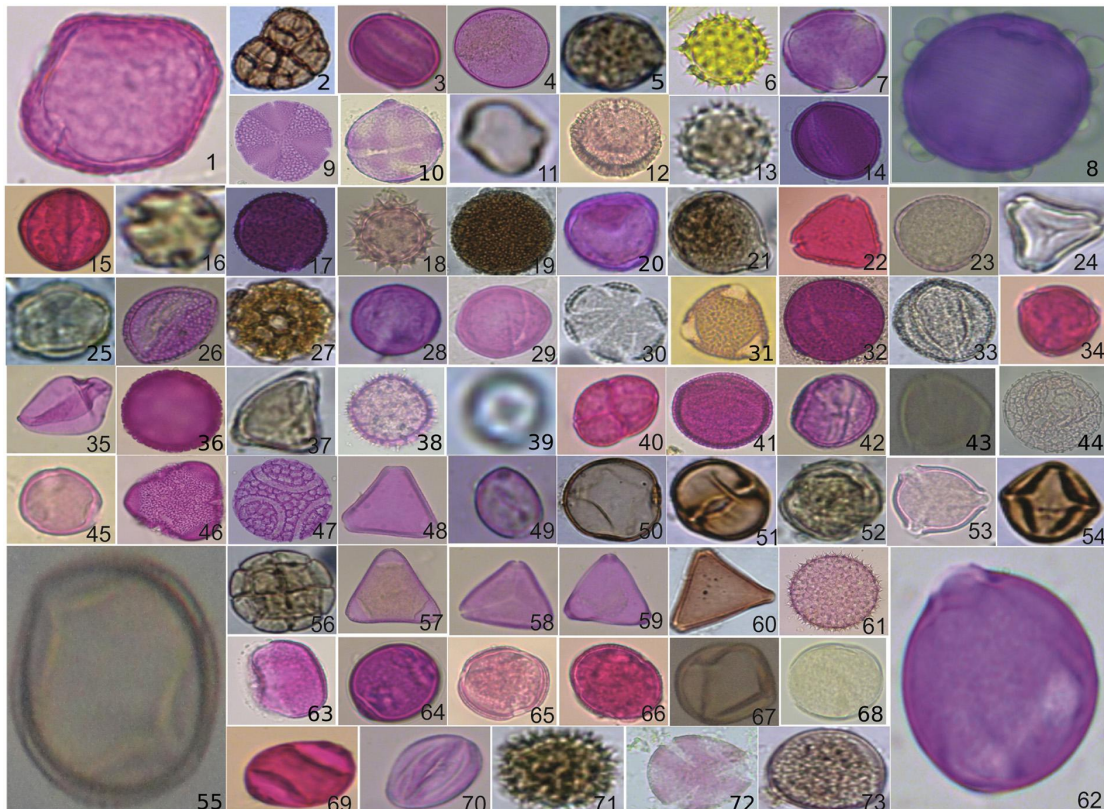
指导老师：胡若澜 huruolan@hust.edu.cn qq:20536555

4.49 花粉识别方法对比分析

任务描述：选择两种方法实现图像中花粉类别的识别，比较分析两种方法的性能。

数据说明：

POLLEN73S 花粉图像数据集由 Astolfi 于 2020 年提出，包含了 73 种花粉类别共 2523 张图像。该数据集采集巴西南马托格罗索州首府坎波格兰德市城区 20°23'16.8"S 54°36'36.3"W 坐标 1.5 公里半径范围内的开花物种样本。下载链接：[POLLEN73S \(figshare.com\)](https://figshare.com)



参考文献:

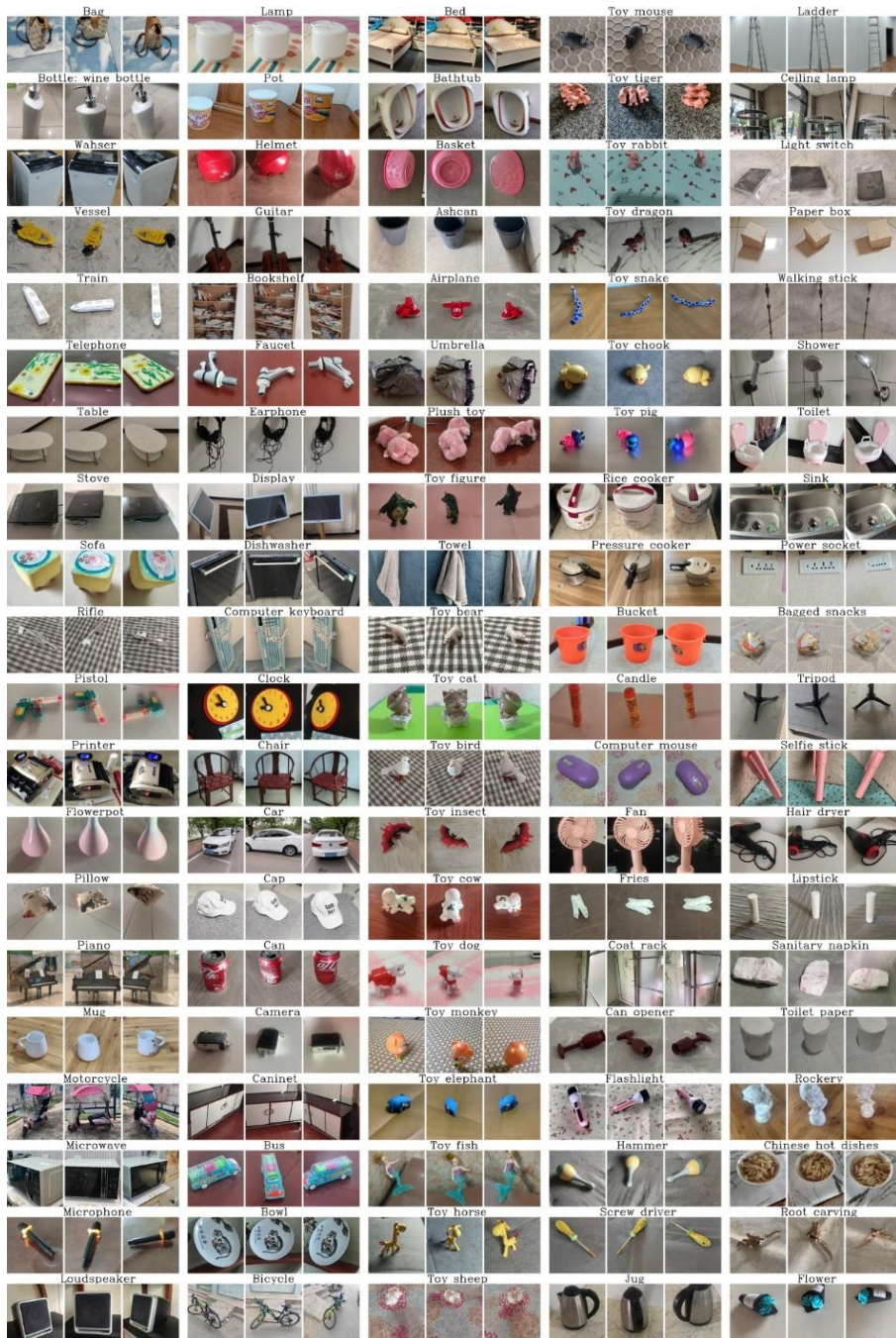
- [1] GONÇALVES A B, SOUZA J S, SILVA G G, et al. Feature extraction and machine learning for the classification of Brazilian Savannah pollen grains[J]. *PloS One*, 2016, 11(6): 1-20.
- [2] ASTOLFI G, GONCALVES A B, MENEZES G V, et al. POLLEN73S: An image dataset for pollen grains classification[J]. *Ecological Informatics*, 2020, 60: 101165-101173.
- [3] N. Khanzhina, M. Kashirin and A. Filchenkov. New Bayesian Focal Loss Targeting Aleatoric Uncertainty Estimate: Pollen Image Recognition[C]. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Vancouver, BC, Canada, 2023: 4253-4262, doi: 10.1109/CVPRW59228.2023.00448.
- [4] K. Duan, S. Bao, H. Zhou, et al. Adaptive Global Token Pooling Transformer for Microscope Pollen Image Recognition[C]. *2023 8th International Conference on Computational Intelligence and Applications (ICCI)*, Haikou, China, 2023: 198-204, doi: 10.1109/ICCI59741.2023.00043.
- [5] W. Cheng *et al.* Attention to Contour: A Contour-Guided Deep Network for Pollen Classification[C]. *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Prague, Czech Republic, 2022: 809-814, doi: 10.1109/SMC53654.2022.9945495.
- [6] S. Battiato, A. Ortis, F. Trenta, et al. Detection and Classification of Pollen Grain Microscope Images[C]. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, 2020: 4220-4227, doi: 10.1109/CVPRW50498.2020.00498.

指导老师: 胡若澜 huruolan@hust.edu.cn qq:20536555

4.50 有限样本条件三维场景重建

任务描述: 三维场景重建是计算机视觉和人工智能领域近年来研究的热点主题之一,能应用于建立数字资产、样本生成等多种。然而在实际应用场景中可能获取的样本数量有限,此外不同样本可能采集于不同时间、受光照、分辨率、部分遮挡、气象等因素影响。本课题的研究目的在于探索不同的三维场景重建方法,并提升其在有限样本条件下的效

果。本题目要求制作的样本数量不少于 10 个，评估标准包括 PSNR 和 LPIPS，重建的方法包括：SparseNeRF、SparseGS、LRM、NoPoSplat、TRELIS 等。



数据描述:

MVImgNet 数据集可从 <https://gaplabor.cuhk.edu.cn/projects/MVImgNet/> 获取。在制作新样本时可以拍摄大量照片，然后进行 colmap 稀疏重建，在测试时选取少量的图片进行评估。

参考文献:

[1] Yu, Alex, et al. "pixelnerf: Neural radiance fields from one or few images." Proceedings

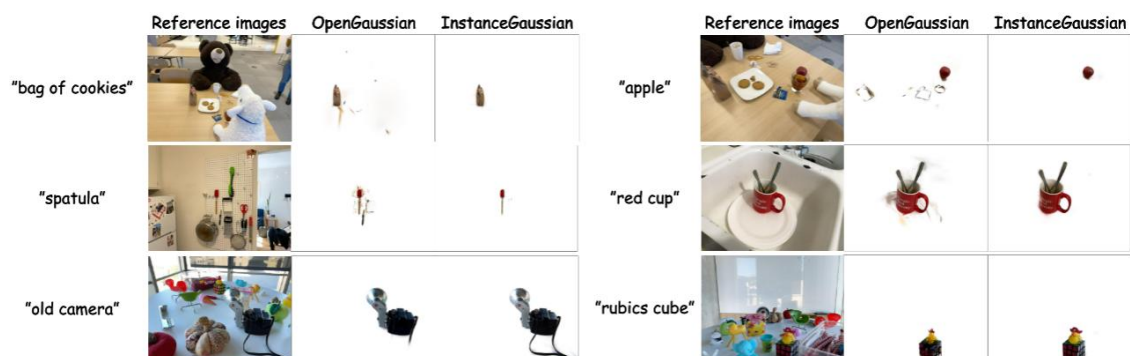
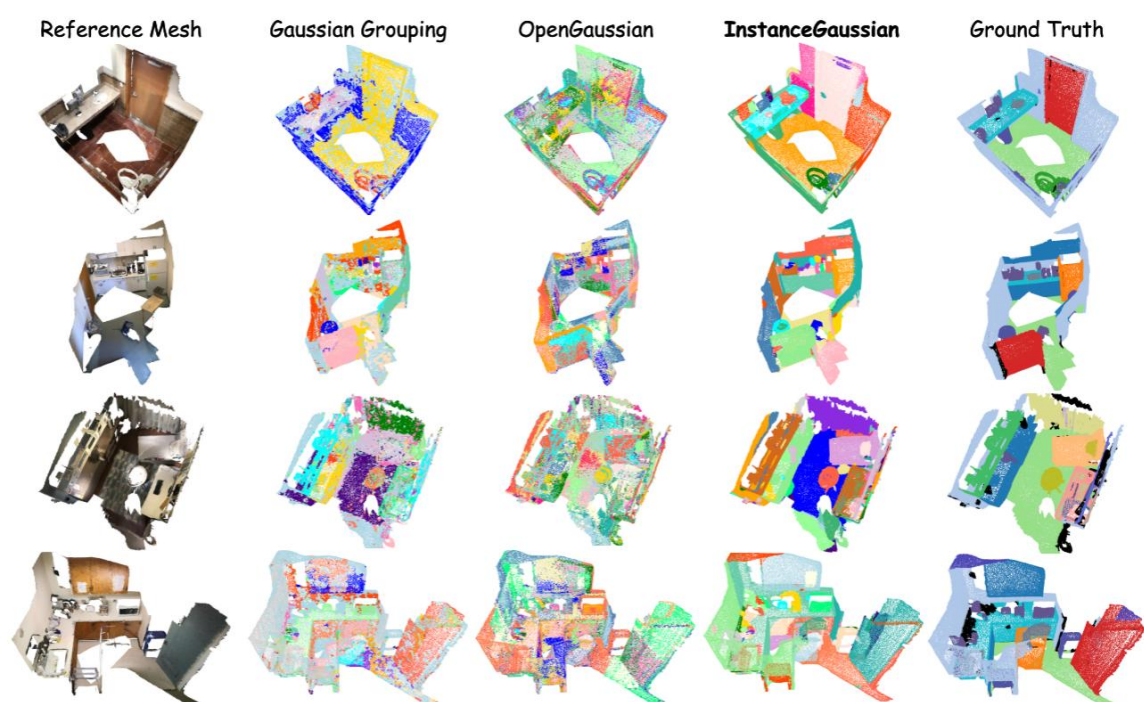
- of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
- [2] ain, Ajay, Matthew Tancik, and Pieter Abbeel. "Putting nerf on a diet: Semantically consistent few-shot view synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
- [3] Wang, Guangcong, et al. "Sparsenerf: Distilling depth ranking for few-shot novel view synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.
- [4] Yu, Xianggang, et al. "Mvimngnet: A large-scale dataset of multi-view images." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023.
- [5] Schonberger, Johannes L., and Jan-Michael Frahm. "Structure-from-motion revisited." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [6] H. Xiong, S. Muttukuru, H. Xiao, R. Upadhyay, P. Chari, Y. Zhao, et al. SparseGS: Sparse View Synthesis using 3D Gaussian Splatting. International Conference on 3D Vision, 2025, Singapore, IEEE, 2025: 1-10
- [7] Y. Hong, K. Zhang, J. Gu, S. Bi, Y. Zhou, D. Liu, et al. LRM: Large Reconstruction Model for Single Image to 3D, in: International Conference on Learning Representations, Vienna, Austria, 2024, International Conference on Learning Representations, 2024: 1-25
- [8] B. Ye, S. Liu, H. Xu, X. Li, M. Pollefeys, M. Yang, et al. No Pose, No Problem: Surprisingly Simple 3D Gaussian Splats from Sparse Unposed Images, in: International Conference on Learning Representations, Singapore Expo, Singapore, 2025, International Conference on Learning Representations, 2025: 1-25
- [9] J. Xiang, Z. Lv, S. Xu, Y. Deng, R. Wang, B. Zhang, et al. Structured 3D Latents for Scalable and Versatile 3D Generation. ArXiv, 2024, 2412.01506

指导老师: 邹旭, zoux@hust.edu.cn, qq: 275666087

4.51 三维场景重建-实例分割优化方法

任务描述: 三维场景重建与细粒度语义理解是计算机视觉领域的关键任务,在自动驾驶、具身智能及 AR/VR 等场景具有广泛的应用前景。然而,现有的三维重建方法往往难以兼顾高质量的新视角生成与精准的物体级感知,且前沿的三维场景理解模型在实际部署时

常面临高维语义特征显存开销大、物体几何边界易粘连等瓶颈。本课题的研究目的在于探索三维场景重建、实例分割与开放语义查询的联合优化方法。在具体任务执行上，本课题需要梳理基于三维高斯溅射（3DGS）的场景智能理解技术的发展脉络，熟悉并复现 2~3 个当前流行的 3D 实例分割或开放语义查询算法模型（例如 InstanceGaussian、OpenGaussian、LangSplat 等）。最终要求完整跑通模型的训练与推理流程，对场景渲染质量（如 PSNR）以及语义查询/分割精度（如 mIoU 或 mAP）等核心性能指标进行定量的汇报与对比分析。



数据描述:

采用 ScanNet (<http://www.scan-net.org/>)、LERF-OVS (<https://www.lerf.io/>)、Replica (<https://github.com/facebookresearch/Replica-Dataset>) 等数据集。

参考文献:

- [1] Lu, T.; Yu, M.; Xu, L.; Xiangli, Y.; Wang, L.; Lin, D.; and Dai, B. 2024. Scaffold-GS: Structured 3D Gaussians for View-Adaptive Rendering. In CVPR.
- [2] Cen, J.; Fang, J.; Yang, C.; Xie, L.; Zhang, X.; Shen, W.; and Tian, Q. 2024. Segment Any 3D Gaussians. In ECCV.
- [3] Ye, M.; Danelljan, M.; Yu, F.; and Ke, L. 2024. GaussianGrouping: Segment and Edit Anything in 3D Gaussians. In ECCV.
- [4] Li H, Wu Y, Meng J, et al. Instancegaussian: Appearance-semantic joint gaussian representation for 3d instance-level perception[C]//Proceedings of the Computer Vision and Pattern Recognition Conference. 2025: 14078-14088.
- [5] Wu, Y.; Meng, J.; Li, H.; Wu, C.; Shi, Y.; Cheng, X.; Zhao, C.; Feng, H.; Ding, E.; and Wang, J. 2024. OpenGaussian: Towards Point-Level 3D Gaussian-based Open Vocabulary Understanding. In NeurIPS.
- [6] Zhou, S.; Chang, H.; Jiang, S.; Fan, H.; Meng, Z.; Qu, W.; and Zhang, Y. 2024. Feature 3DGS: Supercharging 3D Gaussian Splatting to Enable Distilled Feature Fields. In CVPR.
- [7] Qin, M.; Li, W.; Zhou, J.; Wang, H.; and Pfister, H. 2024. LangSplat: 3D Language Gaussian Splatting. In CVPR.
- [8] Kerr, J.; Kim, CM; Goldberg, K.; Kanazawa, A.; and Tancik, M. 2023. LERF: Language Embedded Radiance Fields. In ICCV.

指导老师: 邹旭, zoux@hust.edu.cn, qq: 275666087

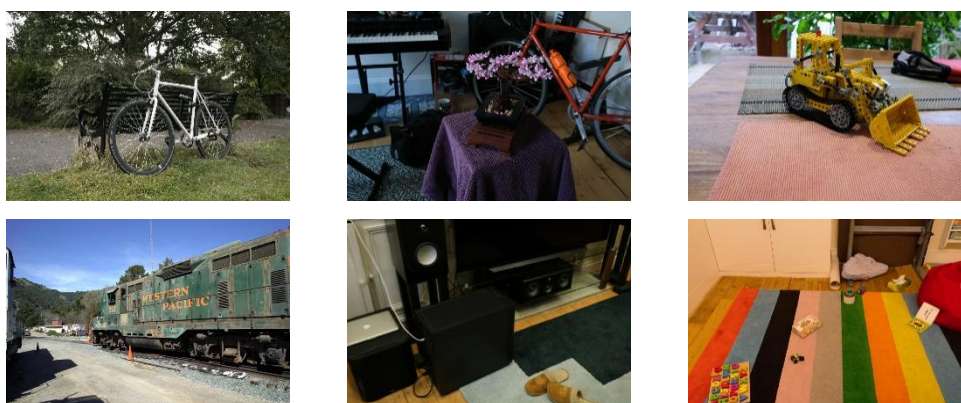
4.52 高效的 3D 高斯重建渲染

任务描述: 针对三维场景重建与新视角渲染任务, 现有方法多数仍以神经隐式表示为主, 即需要较长时间的训练过程和较高的计算资源开销, 使得模型能够完成高质量的场景表达与渲染。近年来, 基于 3D Gaussian Splatting 的方法通过显式建模三维高斯分布, 在保证重建质量的同时显著提升了渲染效率, 具有广阔的应用前景。高效 3D 高斯重建与渲染是一个相对新兴的研究方向, 大多数相关方法仅在近两年内提出。本课题需要梳理 3D 高斯重建方法的发展脉络, 熟悉 2~3 个当前主流的高效 3D 高斯重建算法模型, 并将其应用到具体数据集上, 对重建质量与渲染效率等性能指标进行汇报与分析。同时,

重点从高斯自适应致密化、冗余高斯的剪枝以及渲染加速机制三个方面分析其“高效性”的体现，其中高斯自适应致密化在高误差区域动态增加高斯以加快收敛速度，冗余高斯的剪枝体现在去除低贡献或无效高斯以降低模型复杂度，渲染加速机制则体现在提升渲染帧率和降低推理计算开销。

数据描述:

采用 Mip-NeRF 360、Tanks and Temples 和 Deep Blending 数据集进行分析测试（参考：<https://github.com/graphdeco-inria/gaussian-splatting>）。Mip-NeRF 360 数据集包含多种室内与室外复杂场景，具有大视角变化和非受限拍摄轨迹；Tanks and Temples 数据集主要为大规模户外场景，用于评估方法在复杂几何结构与远距离视角下的重建性能；Deep Blending 数据集为高质量室内场景数据，适用于评估方法在精细结构与光照条件下的渲染效果。实验中采用统一的数据划分与训练设置，对不同方法在各数据集上的表现进行对比分析，采用 PSNR、SSIM 和 LPIPS 作为重建质量评价指标，同时结合渲染帧率（FPS）和训练时间（train_time）作为效率评价指标。



参考文献:

- [1] Kerbl B, Kopanas G, Leimkühler T, et al. 3D Gaussian splatting for real-time radiance field rendering[J]. ACM Transactions on Graphics, 2023, 42(4): 1-14
- [2] Hanson A, Tu A, Lin G, et al. Speedy-splat: Fast 3D Gaussian splatting with sparse pixels and sparse primitives[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2025: 21537-21546
- [3] Fan Z, Wang K, Wen K, et al. LightGaussian: Unbounded 3D Gaussian compression with 15x reduction and 200+ FPS[J]. Advances in Neural Information Processing Systems, 2024, 37: 140138-140158

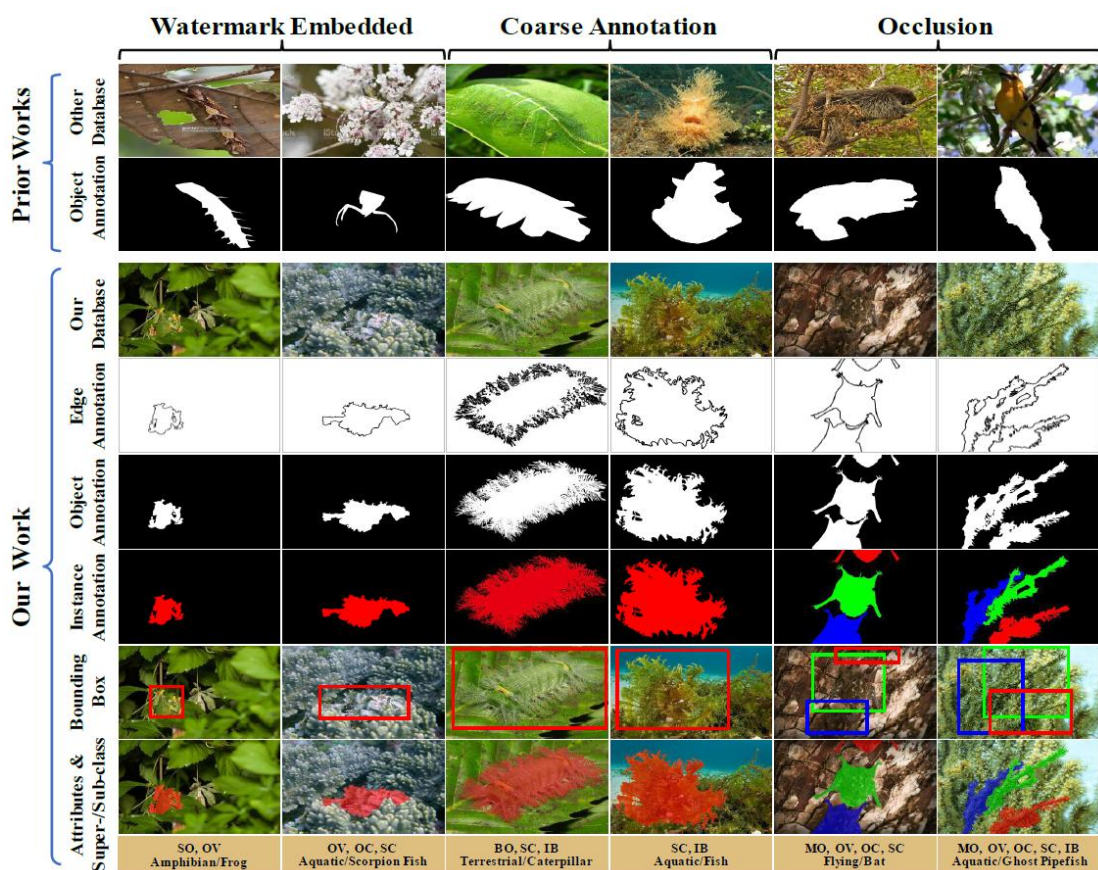
- [4] Lu T, Yu M, Xu L, et al. Scaffold-GS: Structured 3D Gaussians for view-adaptive rendering[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 20654-20664
- [5] Yu Z, Chen A, Huang B, et al. Mip-splatting: Alias-free 3D Gaussian splatting[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 19447-19456
- [6] Huang B, Yu Z, Chen A, et al. 2D Gaussian splatting for geometrically accurate radiance fields[C]. ACM SIGGRAPH 2024 Conference Papers, 2024: 1-11
- [7] Lee J C, Rho D, Sun X, et al. Compact 3D Gaussian representation for radiance field[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 21719-21728
- [8] Chen Y, Wu Q, Lin W, et al. HAC: Hash-grid assisted context for 3D Gaussian splatting compression[C]. European Conference on Computer Vision, 2024: 422-438

指导老师：邹旭，zoux@hust.edu.cn，qq：275666087

4.53 伪装目标检测

任务描述：自然界中许多生物都具有“伪装”的本领，将自身融入周围环境避免被识别和捕食。与通用目标检测以及显著性目标检测不一样，伪装目标检测关注将伪装目标和背景进行二分类分割。具体而言，本题目需要熟悉 2~3 个当前流行的伪装检测算法模型（如 SINet、ANet、PopNet 等），并将其应用到本任务中，对检测性能指标（如 S-measure、E-Measure、weighted F-measure、MAE）进行汇报并分析提升检测性能和造成检测失败的原因。

数据描述：可以采用 COD10K、CAMO 等数据集进行公平比较和分析。COD10K(可参考：<https://github.com/DengPingFan/SINet?tab=readme-ov-file#4-proposed-cod10k-datasets>)数据集包含 6000 个训练样本和 4000 个测试样本。



CAMO 数据集 (<https://github.com/ltnghia/CAMO>) 包含 2000 个训练样本和 500 个测试样本。



参考文献:

- [1] D.-P. Fan, G.-P. Ji, G. Sun, M.-M. Cheng, J. Shen, and L. Shao, "Camouflaged object detection," CVPR, 2020.
- [2] Y. Lyu, J. Zhang, Y. Dai, A. Li, B. Liu, N. Barnes, and D.-P. Fan, "Simultaneously localize, segment and rank the camouflaged objects," CVPR, 2021.
- [3] M. Zhang, S. Xu, Y. Piao, D. Shi, S. Lin, and H. Lu, "Preynet: Preying on camouflaged objects," ACM MM, 2022.
- [4] Y. Sun, S. Wang, C. Chen, and T.-Z. Xiang, "Boundary-guided camouflaged object

detection,” in IJCAI, 2022.

[5] Y. Pang, X. Zhao, T.-Z. Xiang, Z. Lihe, and H. Lu, “Zoom in and out: A mixed-scale triplet network for camouflaged object detection,” CVPR, 2022.

[6] Z. Wu, D. P. Paudel, D.-P. Fan, J. Wang, S. Wang, C. Demonceaux, R. Timofte, and L. V. Gool, “Source-free depth for object pop-out,” ICCV, 2023.

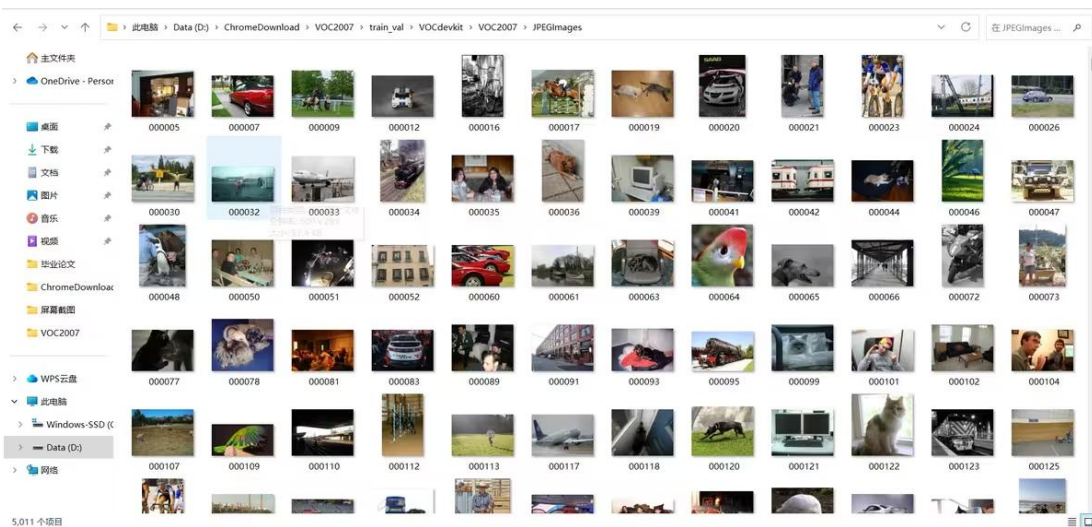
指导老师：邹旭，zoux@hust.edu.cn，qq：275666087

4.54 增量学习目标检测

任务描述：目标检测是计算机视觉和人工智能领域最热点的研究主题之一。然而现有方法大多依赖于有限类别的海量高质量标注数据，而实际开放应用场景中会不断的采集到新类别的新样本数据。本课题的研究目的在于探索不同的增量目标检测方法，并提升其在增量学习过程中兼顾新增类别可塑性的同时对旧数据保持抗遗忘的能力。。

数据描述：采用 Pascal VOC 2007 和 MS-COCO 数据集。PASCAL VOC 2007 数据集包含 20 个类别的 9,963 张图像，其中 50%作为训练集和验证集，剩余半数留作测试集；MS-COCO 是一个更具挑战性的数据集，涵盖 80 个物体类别。在类增量目标检测中，我们将数据集划分为不同的增量任务序列。PASCAL VOC 2007 被划分为两阶段与多阶段增量任务设定。

- 两阶段增量设定中，任务被划分为 19-1 类、15-5 类、10-10 类和 5-15 类四种模式，分别对应每次新增 1、5、10 及 15 个类别的学习过程。
- 多阶段增量设定则进一步细分为 10-5 类、5-5 类、10-2 类、15-1 类和 10-1 类五种模式，分别包含 3、4、6、6 和 11 个任务。



参考文献:

- [1] Nader Asadi, MohammadReza Davari, Sudhir Mudur, Rahaf Aljundi, and Eugene Belilovsky. Prototype-sample relation distillation: towards replay-free continual learning. In International Conference on Machine Learning (ICML), 2023
- [2] rancisco M Castro, Manuel J Marín-Jiménez, Nicolás Guil, Cordelia Schmid, and Karteek Alahari. End-to-end incremental learning. In European Conference on Computer Vision (ECCV), 2018
- [3] Fabio Cermelli, Antonino Geraci, Dario Fontanel, and Barbara Caputo. Modeling missing annotations for incremental learning in object detection. In Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- [4] Guobin Chen, Wongun Choi, Xiang Yu, Tony Han, and Manmohan Chandraker. Learning efficient object detection models with knowledge distillation. Advances in neural information processing systems (NeurIPS), 2017.
- [5] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 44 (7):3366–3385, 2021.
- [6] Dipam Goswami, Yuyang Liu, Bartłomiej Twardowski, and Joost van de Weijer. Fecam: Exploiting the heterogeneity of class distributions in exemplar-free continual learning. Advances in Neural Information Processing Systems (NeurIPS), 2024
- [7] Akshita Gupta, Sanath Narayan, KJ Joseph, Salman Khan, Fahad Shahbaz Khan, and

Mubarak Shah. Ow-detr: Open-world detection transformer. In Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

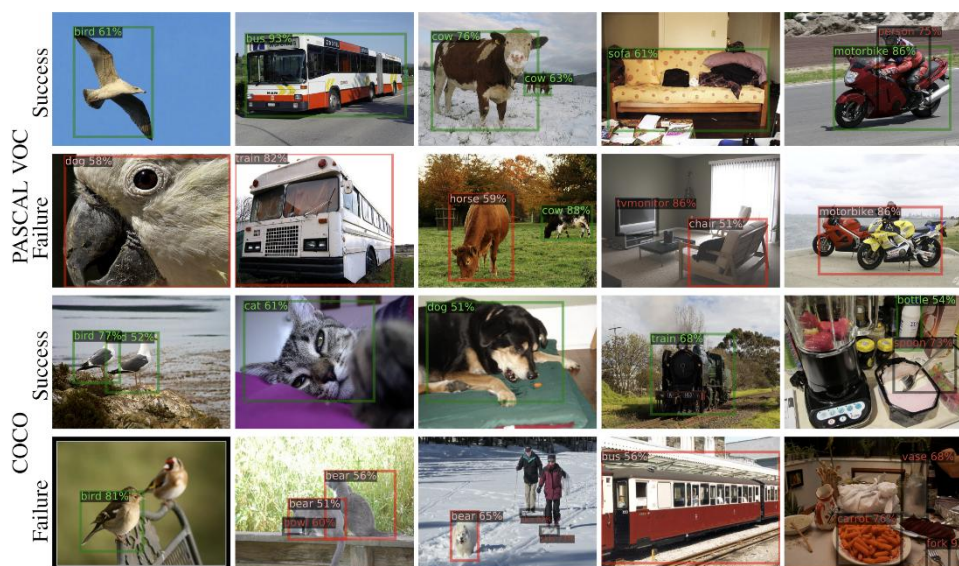
- [8] KJ Joseph, Salman Khan, Fahad Shahbaz Khan, and Vineeth N Balasubramanian. Towards open world object detection. In Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

指导老师：邹旭，zoux@hust.edu.cn，qq：275666087

4.55 小样本目标检测

任务描述：针对目标检测任务，现有深度学习方法多数仍以数据驱动，即需要成千上万的类别实例训练，使得模型能够“认识”类别的新实例。小样本目标检测通过对大量公开可用数据(基类)进行预训练检测模型后，旨在仅对新类目标做少量实例标注后就能快速实现目标检测，具有广阔应用前景。小样本目标检测是一个相对新兴的研究领域，大多数方法仅在过去两年内发表。本课题需要梳理小样本目标检测的发展脉络，熟悉2~3个当前流行的小样本目标检测算法模型，并将其应用到本任务中，对检测性能指标进行汇报并分析。

数据描述：采用 PASCAL VOC 数据集和 MS-COCO 数据集进行分析测试（参考 <https://github.com/ucbdrive/few-shot-object-detection>）。PASCAL VOC 数据集包括三个不同的数据集划分，每个划分随机地将 20 类目标分为 15 个基类目标和 5 个新类目标。对于每个新类，1、2、3、5、10 个样本用于微调的设置被采用，采用 AP(IoU=0.5)的 nAP50 作为新类目标的检测性能评价指标。MS-COCO 数据集包括 80 个类别，其中 20 个为新类目标，60 个为基类目标，对于每个新类，1、2、3、5、10、30 个样本用于微调的设置被采用，采用 COCO 风格的 AP 作为目标检测性能评价指标。



参考文献:

- [1] Xin Wang, Thomas E Huang, Trevor Darrell, Joseph E Gonzalez, and Fisher Yu. Frustratingly simple few-shot object detection. In ICML, 2020.
- [2] Limeng Qiao, Yuxuan Zhao, Zhiyuan Li, Xi Qiu, JiananWu, and Chi Zhang. Defrcn: Decoupled faster r-cnn for few-shot object detection. In ICCV, 2021.
- [3] Shuang Wu, Wenjie Pei, Dianwen Mei, Fanglin Chen, Jiandong Tian, and Guangming Lu. Multi-faceted distillation of base-novel commonality for few-shot object detection. In ECCV, 2022.
- [4] Jingyi Xu, Hieu Le, and Dimitris Samaras. Generating features with increased crop-related diversity for few-shot object detection. In CVPR, 2023.
- [5] Jiaming Han, Yuqiang Ren, Jian Ding, Ke Yan, and Gui-Song Xia. Few-shot object detection via variational feature aggregation. AAAI, 2023.
- [6] Adrian Bulat, Ricardo Guerrero, Brais Martinez, and Georgios Tzimiropoulos. Fs-detr: Few-shot detection transformer with prompting and without re-training. In ICCV, 2023.
- [7] Xiaonan Lu, Wenhui Diao, Yongqiang Mao, Junxi Li, Peijin Wang, Xian Sun, and Kun Fu. Breaking immutable: information-coupled prototype elaboration for few-shot object detection. In AAAI, 2023.
- [8] Qi Fan, Chi-Keung Tang, and Yu-Wing Tai. Few-shot object detection with model calibration. In ECCV, 2022.

指导老师: 邹旭, zoux@hust.edu.cn, qq: 275666087

4.56 旋转目标检测

任务描述：相比通用目标检测（水平框检测），旋转目标检测在通用目标检测的基础上同时输出目标的朝向角，能实现对目标更准确地定位。本课题需要对目标旋转角度的表达方法进行调研、梳理和分析，熟悉4~6个当前流行的旋转目标检测算法模型（如RoI Trans、Gliding Vertex、CSL、ReDet、GWD、KLD等），并将其应用到本任务中，对检测性能指标进行汇报并分析。

数据描述：采用DOTA数据集进行评测。DOTA数据集（参考<https://captain-whu.github.io/DOTA/dataset.html>）包括2806个航拍图像（图像尺寸从800×800到4000×4000），包括188282个目标实例和15个目标类别。HRSC2016数据集（参考）包括436张训练图像、181张验证图像和444张测试图像。



参考文献:

- [1] J. Ding, N. Xue, Y. Long, G.-S. Xia, and Q. Lu, “Learning RoI transformer for oriented object detection in aerial images,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 2849–2858.
- [2] X. J. Pan et al., “Dynamic refinement network for oriented and densely packed object detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2020, pp. 11207–11216.
- [3] Y. Xu et al., “Gliding vertex on the horizontal bounding box for multioriented object detection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 4, pp. 1452–1459, Apr. 2021.
- [4] X. Yang, J. Yan, Q. Ming, W. Wang, X. Zhang, and Q. Tian, “Rethinking rotated object detection with Gaussian Wasserstein distance loss,” in Proc. 38th Int. Conf. Mach. Learn., 2021, pp. 11830–11841.
- [5] X. Yang et al., “Learning high-precision bounding box for rotated object detection via Kullback–Leibler divergence,” in Proc. Adv. Neural Inf. Process. Syst., vol. 34, 2021, pp. 18381–18394.
- [6] X. Yang et al., “Detecting rotated objects as Gaussian distributions and its 3-D generalization,” IEEE Trans. Pattern Anal. Mach. Intell., early access, Aug. 8, 2022, doi: 10.1109/TPAMI.2022.3197152.
- [7] X. Yang and J. C. Yan, “Arbitrary-oriented object detection with circular smooth label,” in Proc. Eur. Conf. Comput. Vis., 2020, pp. 677–694.
- [8] X. Yang, L. Hou, Y. Zhou, W. Wang, and J. Yan, “Dense label encoding for boundary discontinuity free rotation detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 15819–15829.
- [9] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, “Oriented R-CNN for object detection,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2021, pp. 3520–3529.
- [10] Y. Wang, Z. Zhang, W. Xu, L. Chen, G. Wang, L. Yan, S. Zhong, X. Zou, “Learning Oriented Object Detection via Naïve Geometric Computing”, IEEE Transactions on Neural Networks and Learning Systems, 2023.

指导老师: 邹旭, zoux@hust.edu.cn, qq: 275666087

4.57 三维场景渲染与可控编辑-目标检测一体化测试技术

任务描述：开放场景下，目标感知模型一方面需持续适应样本分布与任务需求的动态演化，另一方面则需基于当前算法模型在不同复杂干扰条件下的效能评估进行实时闭环反馈。依托高精度三维场景重建与可控编辑技术，本课题研究的一体化测试技术能够实现对虚拟环境的细粒度几何与外观操控。平台不仅支持对场景内特定目标进行精准的增删、位姿操控与材质替换，还能实时动态生成高度拟真的全局环境干扰（如雨雾弥漫、强光眩目、夜间低照度、沙尘暴等）并实现多变背景的无缝自适应融合，从而为目标检测算法测试提供接近真实世界且具备高度对抗性的挑战场景。具体而言，本课题需要熟悉从前端三维场景交互式编辑、渲染到后端目标检测的全流程一体化平台构建路径，将当前流行的目标检测（包括伪装、增量、小样本等）算法部署到该平台中，完成“目标自适应生成-场景渲染与可控编辑-感知检测-结果反馈”的动态闭环构建，并尝试融入新的小功能（如特定目标的局部纹理替换或光影重绘），最终对感知模型在各类编辑场景下的性能指标进行汇报与对比分析。





数据描述:

采用提供的多谱试验场扫描图像集、东校区扫描图像集等数据，或者自建一定范围场景下的扫描图像集。

参考文献:

[1] Chen, Y.; Chen, Z.; Zhang, C.; Wang, F.; Yang, X.; Wang, Y.; Cai, Z.; Yang, L.; Liu, H.; and Lin, G. 2024. GaussianEditor: Swift and Controllable 3D Editing with Gaussian Splatting. In CVPR.

[2] Ye, M.; Danelljan, M.; Yu, F.; and Ke, L. 2024. GaussianGrouping: Segment and Edit Anything in 3D Gaussians. In ECCV.

[3] Guédon, A.; and Lepetit, V. 2024. SuGaR: Surface-Aligned Gaussian Splatting for

Efficient 3D Mesh Reconstruction and High-Quality Mesh Editing. In CVPR.

[4] Xie, T.; Zong, Z.; Qiu, Y.; Li, X.; Feng, Y.; Yang, Y.; and Jiang, C. 2024. PhysGaussian: Physics-Integrated 3D Gaussians for Generative Dynamics. In CVPR.

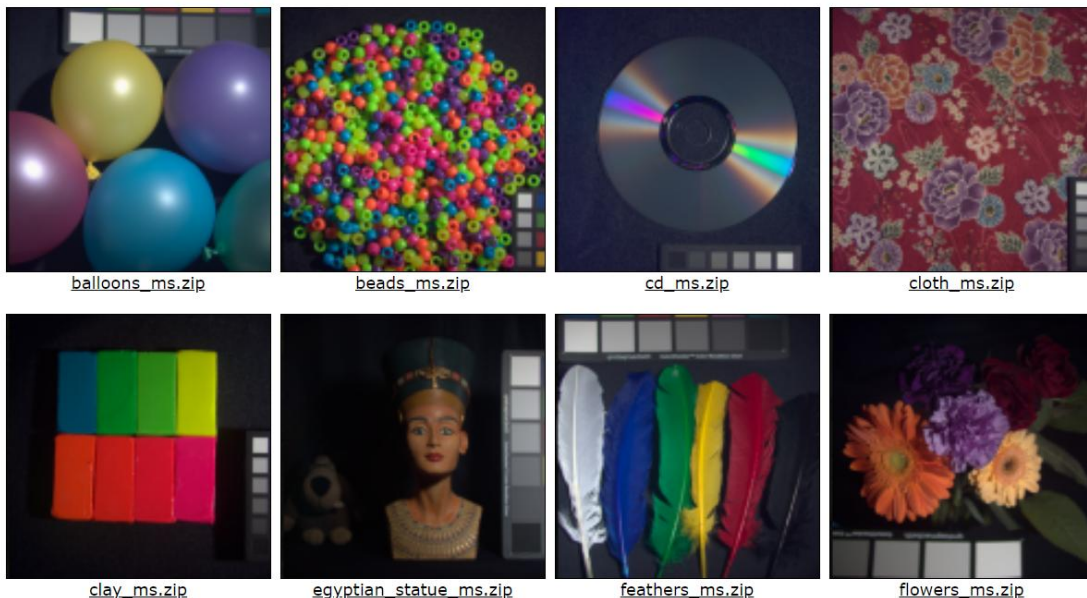
[5] Fang, J.; Wang, J.; Zhang, X.; Xie, L.; and Tian, Q. 2024. GaussianEditor: Editing 3D Gaussians Delicately with Text Instructions. In CVPR.

指导老师：邹旭，zoux@hust.edu.cn，qq：275666087

4.58 高分辨率时-空-谱融合计算成像

问题描述：高分辨率时空谱成像在生物学、航天遥感领域发挥着关键作用。传统‘所见即所得’图像获取方式受限于传感器性能和传输带宽，难以同时获得高时间、空间、光谱分辨率的图像。计算成像技术应运而生，通过前端光学调制、后端计算解码实现高分辨率成像。本题要求完成多光谱-高光谱双模融合成像相关文献梳理形成综述文档，重点研究多光谱 + 高光谱数据的双模融合计算成像方法，并对 MHF-Net [5]算法重新训练并测试，从性能、速度、泛化性等维度评价。

数据描述：CAVE 数据集 <https://www.cs.columbia.edu/CAVE/databases/multispectral/>



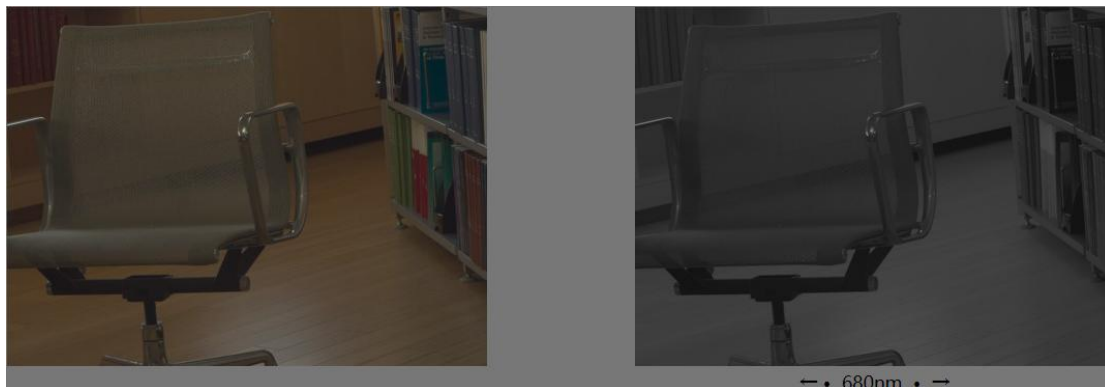
The database consists of 32 scenes, divided into 5 sections. Each scene has an associated zip file. These zip files include full spectral resolution reflectance data from 400nm to 700nm at 10nm steps (31 bands total). Each band is stored as a 16-bit grayscale PNG image. Image filenames are of the format 'object_ms_01.png', where the '01' at the end signifies that this is

the first image (captured at 400nm). Thus, '02' corresponds to 410nm, and so on, until '31' for 700nm. Each scene also contains a single representative color image, displayed using sRGB values rendered under a neutral daylight illuminant (D65).

Harford 数据集 <https://vision.seas.harvard.edu/hyperspec/d2x5g3/>

50 Indoor & outdoor images under daylight: CZ_hbdb.tgz (5.3GB)

27 Indoor images under artificial & mixed illumination: CZ_hbdbi.tgz (2.2GB)



参考文献:

- [1] Lizhi Wang, Zhiwei Xiong, et al. High-Speed Hyperspectral Video Acquisition By Combining Nyquist and Compressive Sampling [J]. IEEE transactions on pattern analysis and machine intelligence, 2019.
- [2] Yang Liu, Xin Yuan, et al. Rank Minimization for Snapshot Compressive Imaging [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
- [3] Ming Cheng, Zhan Ma, et al. A Dual Camera System for High Spatiotemporal Resolution Video Acquisition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
- [4] Yibo Xu, Liyang Lu, et al. A Compressive Hyperspectral Video Imaging System using a Single-pixel Detector [J]. Nature Communications, 2024.
- [5] Qi Xie, Minghao Zhou, et al. MHF-Net: An interpretable deep network for multispectral and hyperspectral image fusion [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.

指导老师: 昌毅, yichang@hust.edu.cn, qq 458449613

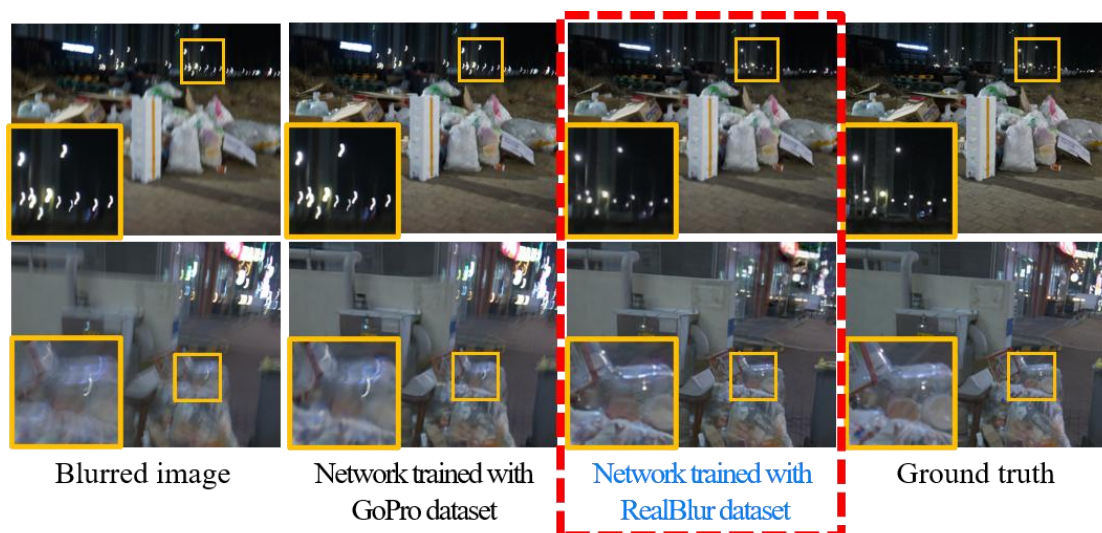
4.59 低光图像饱和去模糊

问题描述: 结构噪声(线状、斑状)与模糊广泛存在与各类成像系统中, 成像质量严重

下降，探测目标遭到破坏，制约图像后续使用。解决结构噪声干扰下图像去模糊问题，可提升图像质量，保障图像高层任务性能。本题要求完成低光饱和图像去模糊相关文献梳理形成综述文档，重点研究饱和情况下图像去模糊复原方法，掌握 OID[3]实现细节，从性能、速度、泛化性等维度评价。

数据描述: RealBlur Dataset: <http://cg.postech.ac.kr/research/realblur/>

This work present a large-scale dataset of real-world blurred images and ground truth sharp images for learning and benchmarking single image deblurring methods. To collect the dataset, the authors build an image acquisition system to simultaneously capture geometrically aligned pairs of blurred and sharp images, and develop a post-processing method to produce high-quality ground truth images.



RealBlur: 1) training set : 3,758 image pairs of 182 different scenes. 2) test set : 980 image pairs of 50 different scenes.

参考文献:

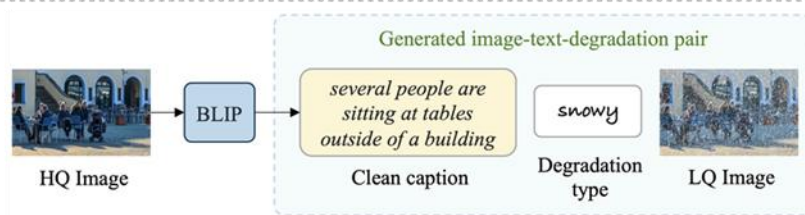
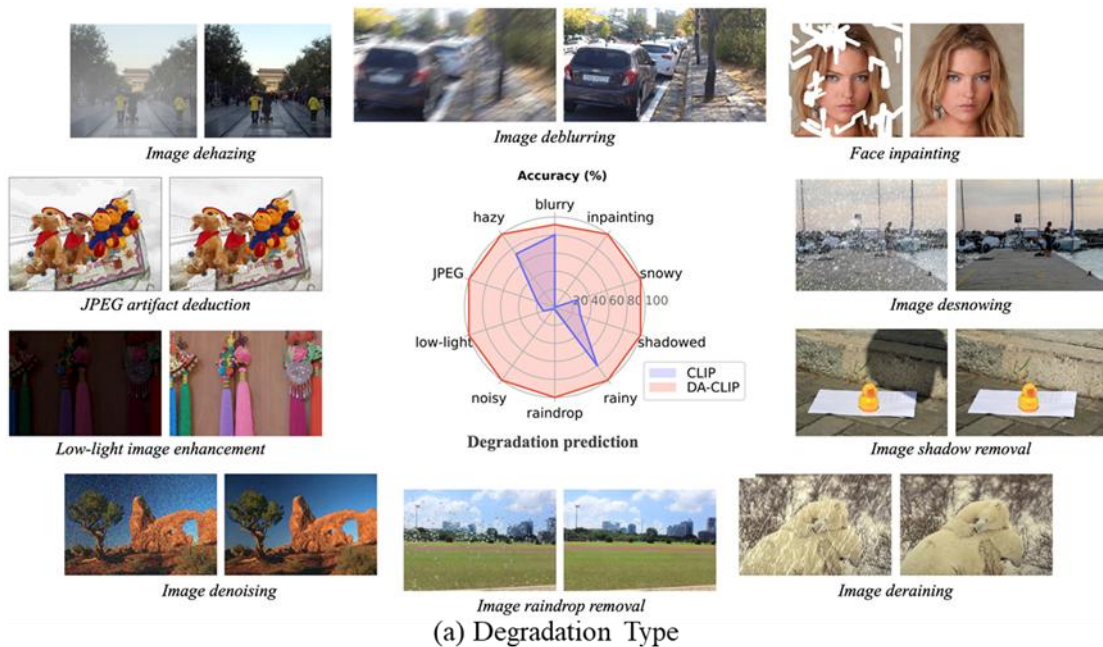
- [1] Qi Shan, Jia Jiaya, Aseem Agarwala. High-quality Motion Deblurring From A Single Image. ACM Transactions on Graphics, 2008, 27(3): 557-566.
- [2] Cao Shuning, Fang Houzhang, Chen Linqun, et al. Robust blind deblurring under stripe noise for remote sensing images. IEEE Trans. on Geoscience and Remote Sensing, 2022, 60: 1-17.
- [3] Liang Chen , Faming Fang, et al. Outlier Identifying and Discarding in Blind Image Deblurring. European Conference on Computer Vision, 2020, 2478-2486.

指导老师: 昌毅, yichang@hust.edu.cn, qq 458449613

4.60 基于语言-视觉模型的图像复原方法

问题描述: 实际场景中, 可见光图像易受到雨天、雾天、雪天等恶劣天气退化效应干扰, 现有方法多集中于利用图像的单一模态信息进行复原, 忽视了其他模态的潜在作用。本课题旨在探索语言文本这一额外模态在恶劣天气图像复原中的作用, 借助语言-视觉模型, 利用语言描述提供的语义线索与图像内容的协同, 提升恶劣天气图像视觉质量, 进而提升计算机视觉系统对恶劣天气退化的鲁棒性。

数据描述: <https://github.com/Algolzw/daclip-uir>



DA-CLIP 所采用的混合退化图像数据集是基于多个已有公开数据集收集整理而成, 旨在提升模型在多种图像退化场景下的图像复原能力。该数据集主要来源于广泛应用的视觉图像数据集, 包括 BSD、GoPro、LOL、RESIDE、Synthetic Rain 数据集及 AWMM-100k 等。这些数据集提供了大量成对的高质量图像与其对应的退化图像, 涵盖了包括雨、雪、雾霾等天气相关退化, 以及运动模糊、JPEG 压缩、噪声干扰、阴影遮挡与图像缺失等

多种退化类型，具备高度的多样性与代表性。数据集中所涉及的图像覆盖了丰富的场景内容，包含室内与室外环境、不同的光照条件以及动态与静态目标等。如雾霾退化图像主要选自 I-Haze 和 O-Haze 数据集，分别针对室内与室外场景，每个子集均包含约 45 对图像，图像分辨率为 4560×3040 像素；雨天退化图像主要来源于 RainCityscapes 数据集，包含 9437 张带有真实雨天效果的图像，分辨率为 2048×1024 像素；而降雪类图像则主要使用 Snow100K 数据集，涵盖约 100,000 张含雪景图像，分辨率为 1280×720 像素。为实现文本引导的图像修复，每对高质量与低质量图像均配备由 BLIP 框架自动生成的图像描述文本，进而构建图像-文本-退化类型的三元组训练样本。这种跨模态配对方式不仅增强了模型对退化类型的识别能力，也为引导式图像复原任务提供了有效的数据支持。

参考文献：

- [1] Z. Luo, F. Gustafsson, Z. Zhao, et al. Controlling Vision-Language Models for Multi-Task Image Restoration. International Conference on Learning Representation. 2024
- [2] H. Yang, L. Pan, Y. Yang, et al. Language-driven All-in-one Adverse Weather Removal. IEEE Conference on Computer Vision and Pattern Recognition. 2024: 24902-24912.
- [3] Y. Wei, Y. Zhang, K. Li, et al. Leveraging vision-language prompts for real-world image restoration and enhancement. Computer Vision and Image Understanding, 2024: 104222.
- [4] X. Xu, S. Kong, T. Hu, et al. Boosting Image Restoration via Priors from Pre-trained Models. IEEE Conference on Computer Vision and Pattern Recognition. 2024: 2900-2909.
- [5] B. Gella, H. Zhang, R. Upadhyay, et al. WeatherProof: Leveraging Language Guidance for Semantic Segmentation in Adverse Weather. arXiv preprint:2403.14874, 2024.

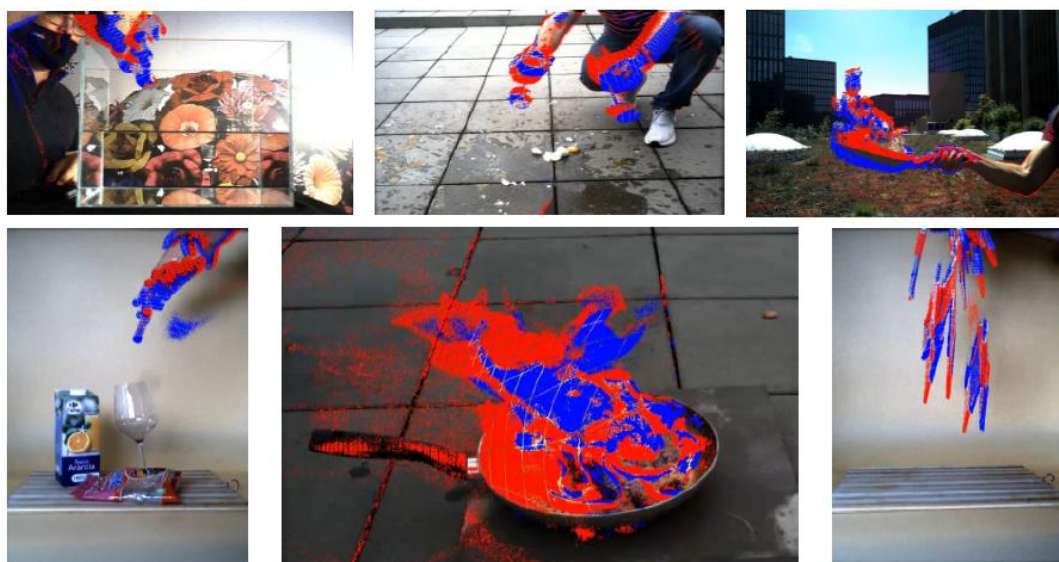
指导老师：昌毅，yichang@hust.edu.cn，qq 458449613

4.61 事件相机高速高动态成像

问题描述：高速高动态范围成像可以为视觉感知任务提供更丰富的时空信息。传统帧相机采用全局同步曝光的方式积分成像，难以同时适应场景中局部过亮和过暗区域对于曝光的要求，导致其动态范围较低；且积分成像需要一定时间累积能量，难以在无补光的自然场景高速成像。而事件相机各像素异步差分触发，时间分辨率达到 1 微秒，动态范

围达到 120dB，具有同时进行高速高动态成像的能力。本课题要求完成基于事件相机的高速高动态成像相关文献梳理形成综述文档，重点研究事件 + 图像的双模融合计算成像方法，对算法[5]重新训练并测试，并以合理的方式评价融合成像的质量。

数据描述：BS-ERGB 数据集 <https://uzh-rpg.github.io/timelens-pp/>



The authors built an experimental setup with a global shutter RGB Flir 4096x2196 camera and a Prophesee Gen4M 1280x720 event camera arranged with the beam splitter. The dataset consists of 123 diverse and challenging scenes with varying depth. And the dataset is divided into 78 training scenes, 19 validation scenes, and 26 test scenes.

ERF-X170FPS 数据集 <https://github.com/intelpro/CBMNet>



The authors build the large-scale dataset named ERF-X170FPS (High Resolution Events and RGB Frames with eXtreme Motions at 170FPS) with the beam-splitter-based camera rigs comprising of the Prophesee Gen4 event cameras (1280×720) and Blackfly-S global shutter cameras (1440×1080, maximum 226fps). These two cameras are hardware-level synchronized

using a micro-controller (the external trigger). This dataset contains a total of 36 sequences containing 990 frames per sequence.

参考文献:

[1] Tulyakov S, Gehrig D, Georgoulis S, et al. Time lens: Event-based video frame interpolation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 16155-16164.

[2] Tulyakov S, Boicchio A, Gehrig D, et al. Time lens++: Event-based frame interpolation with parametric non-linear flow and multi-scale fusion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 17755-17764.

[3] Kim, Taewoo, et al. Event-based video frame interpolation with cross-modal asymmetric bidirectional motion fields. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

[4] Han J, Zhou C, Duan P, et al. Neuromorphic camera guided high dynamic range imaging[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 1730-1739.

[5] Wenming Weng, Yueyi Zhang, Zhiwei Xiong. Event-based Blurry Frame Interpolation under Blind Exposure. CVPR, 2023.

指导老师: 昌毅, yichang@hust.edu.cn, qq 458449613

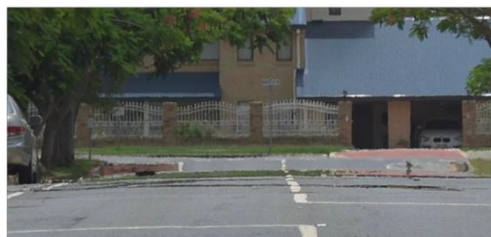
4.62 远距高清晰成像

问题描述: 远距成像是大气科学、天文观测、军事国防等领域不可或缺的关键技术手段。然而,长距离观测不可避免会受到大气湍流的干扰。具有不同尺度涡流的湍流介质在光波的传播过程中会引起折射率的波动,从而对原有的空间关系、相位关系和光程产生干扰,带来成像模糊、畸变扭曲的退化。本题要求完成大气湍流退化图像复原方法相关文献梳理形成综述文档,重点研究可见光数据的大气湍流校正方法,并对 DATUM [3]算法重新训练并测试,从算法原理实测性能、推理速度、泛化性等维度评价算法。

数据描述: ATD 数据集 <https://pan.baidu.com/s/1vErExH0U5oRDU3-n04S5mg?pwd=vakg#list/path=%2F>

To fully demonstrate the performances of the different models in neutralizing the impact

of turbulence, we constructed a dataset composed of 5,201 real-world video sequences with 78,015 frames containing thermally driven turbulence. The real-world data were composed of two parts. The first contained 4,303 video sequences shot in Australia and 898 clips containing turbulence sampled from a series of documentaries. In Brisbane, Australia, we used a Nikon D750 camera to capture images in street view when the local temperature was ~ 93 °F to 99 °F. Among the collected data, 4,303 video sequences involving typical scenes (including buildings, vehicles and pedestrians) were sampled and constitute the second part of the dataset. These were mainly recorded in Queen Street ($153^{\circ} 2'E, -27^{\circ} 28'N$), Victoria Bridge and Pinelands Road ($152^{\circ} 45'E, -27^{\circ} 35'N$). The related camera parameter settings were from auto mode, with shutter speed of 1/250–1/200, aperture size of 1/11–1/9 and ISO of 100–300. The resolution of the original video was $1,920 \times 1,080$, and the sequences used in the dataset were cropped to an appropriate size. Real-world turbulence, which is influenced by various complicated factors as well as moving objects at different distances, could further validate the efficacy of the proposed and comparison models.



TMT 数据集 <https://github.com/xg416/TMT?tab=readme-ov-file>

ATSyn 数据集 <https://xg416.github.io/DATUM/>

参考文献:

[1] Jin D, Chen Y, Lu Y, et al. Neutralizing the impact of atmospheric turbulence on complex scene imaging via deep learning[J]. Nature Machine Intelligence, 2021, 3(10): 876-884.

[2] Zhang X, Mao Z, Chimmitt N, et al. Imaging through the atmosphere using turbulence mitigation transformer[J]. IEEE Transactions on Computational Imaging, 2024.

[3] Zhang X, Chimitt N, Chi Y, et al. Spatio-Temporal Turbulence Mitigation: A Translational Perspective. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024.

指导老师：昌毅，yichang@hust.edu.cn，qq 458449613

4.63 远距透雾成像

问题描述：远距成像过程中不可避免会受到严重大气散射效应影响，导致远距场景对比度以及纹理细节极度劣化。现有单帧图像去雾方法主要依赖手工先验或者数据驱动方式，从退化图像内寻找线索从而恢复出无雾的清晰图像。但是此类方法无法处理远距成像散射效应导致的严重场景信息丢失。为了找回丢失的信息，受益于近红外波段天然的透雾能力，部分方法提出将雾天图像增强问题转化为可见光+近红外图像融合问题，从额外的近红外波段找回可见光中丢失的场景内容以及纹理细节信息。本题要求完成雾天图像复原/增强以及双模图像融合相关文献梳理形成综述文档，重点研究可见光+近红外双模融合远距雾天图像增强方法，对 SwinFusion [3]和 CDDFuse[4] 算法重新训练并测试，并从性能、速度、泛化性等维度评价算法。

数据描述：RGB-NIRScene 数据集

<https://www.epfl.ch/labs/ivrl/research/downloads/rgb-nir-scene-dataset/>



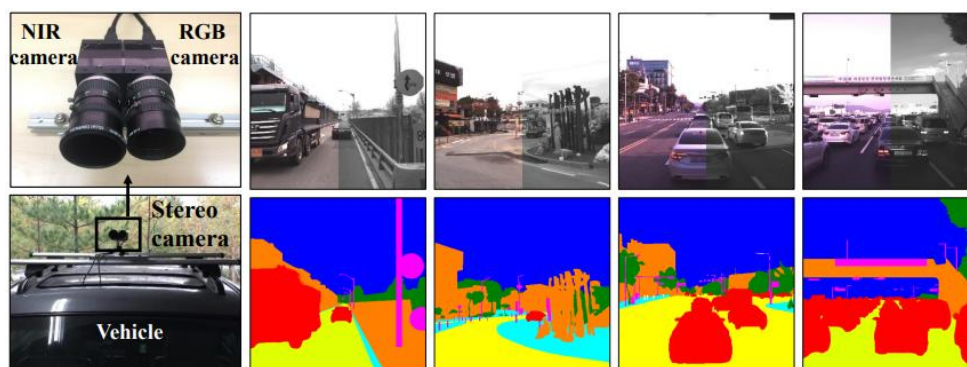
This dataset consists of 477 images in 9 categories captured in RGB and Near-infrared (NIR). The images were captured using separate exposures from modified SLR cameras, using visible and NIR filters. For more info on NIR photography, see the references below. The scene categories are: country, field, forest, indoor, mountain, oldbuilding, street, urban, water.

RANUS 数据集

https://drive.google.com/file/d/1WJ7rcCeMBPy9Qb2c_pwI0rUIIFgotkzT/view

Database contains 40k spatially-aligned RGB-NIR scenes captured from a vehicle driven

across various urban, rural and campus roads consisting of buildings and natural landscapes.



参考文献:

- [1] Schaul L, Fredembach C, Süsstrunk S. Color image dehazing using the near-infrared. 2009 International Conference on Image Processing. 2009: 1629-1632.
- [2] Liu L, Wang F, Jung C. LRINet: Long-range imaging using multispectral fusion of RGB and NIR images. Information Fusion, 2023, 92: 177-189.
- [3] Ma J, Tang L, Fan F, et al. SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer. IEEE/CAA Journal of Automatica Sinica, 2022, 9(7): 1200-1217.
- [4] Zhao Z, Bai H, Zhang J, et al. Cddfuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 5906-5916.

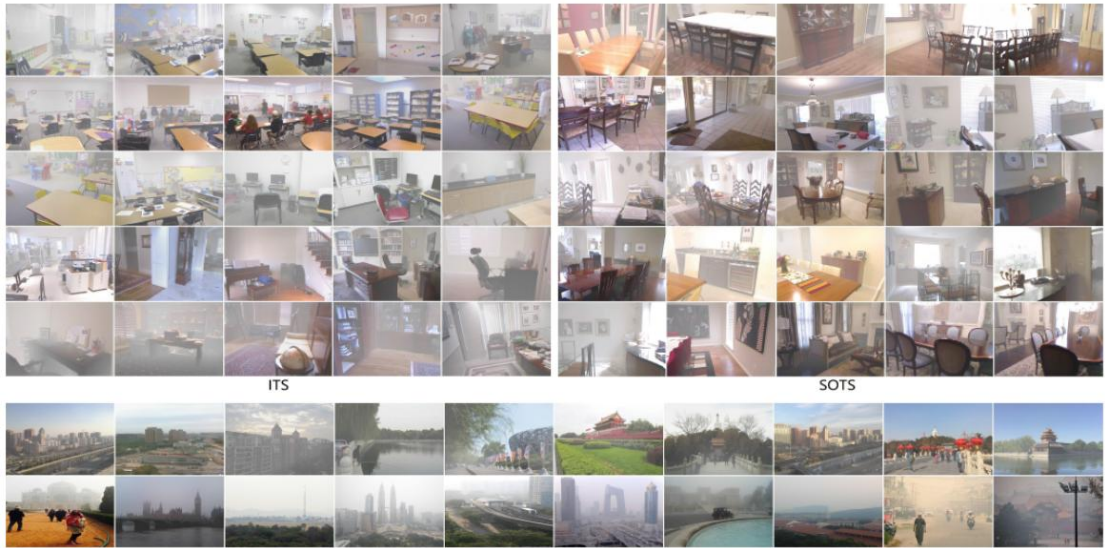
指导老师: 昌毅, yichang@hust.edu.cn, qq 458449613

4.64 雨雪雾霾恶劣天气图像统一增强方法

问题描述: 恶劣天气图像增强对提升计算机视觉系统在不同天气环境下的适应性具有重要作用, 在自动驾驶、侦察监视领域具有极大应用前景。现有图像增强方法往往专注于增强某一类特定恶劣天气退化(如雨、雾、雪)。然而, 实际场景中不同种类退化可能不定时出现。因此, 能够同时处理多种退化的恶劣天气图像统一增强模型具有重大研究意义。本题要求完成恶劣天气图像增强相关文献梳理形成综述文档, 重点研究恶劣天气图像统一增强模型与方法, 对 WeatherDiff[5]算法重新训练并测试, 从性能、速度、泛化性等维度评价算法。

数据描述: RESIDE 数据集

<https://sites.google.com/view/reside-dehaze-datasets/reside-standard>



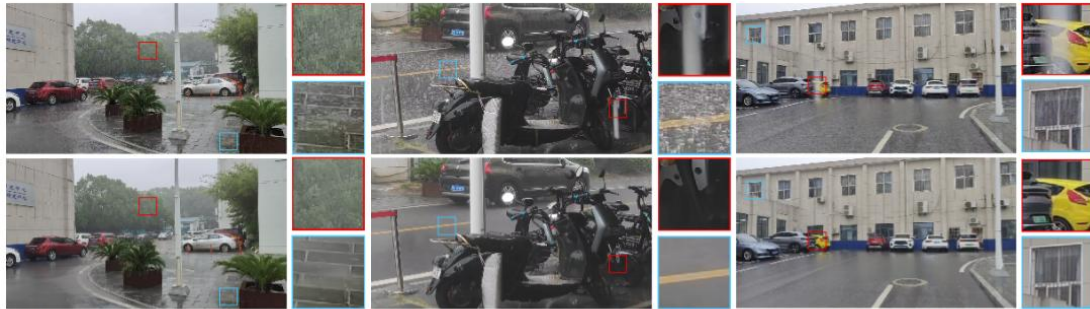
HSTS. Top row: 10 synthetic hazy images; Bottom row: 10 realistic hazy images.

RESIDE(Standard)				
Subset	Number of Images	real/synthetic	indoor/outdoor	annotations
Indoor Training Set (ITS)	13,990	synthetic	indoor	No
Synthetic Objective Testing Set (SOTS)	500	synthetic	indoor	No
Hybrid Subjective Testing Set (HSTS)	20	real	outdoor	No

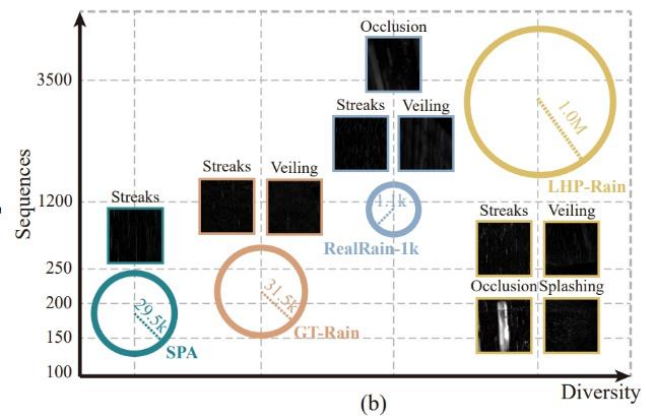
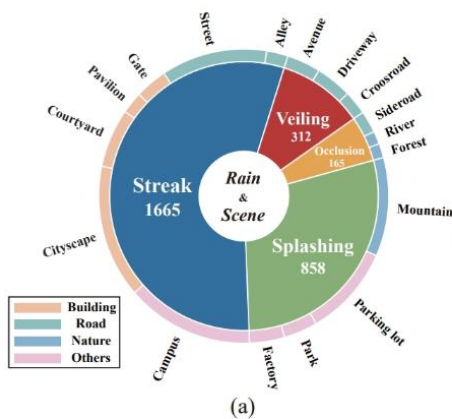
RESIDE- β				
Subset	Number of Images	real/synthetic	indoor/outdoor	annotations
Outdoor Training Set (OTS)	72,135	synthetic	outdoor	No
Real-world Task-driven Testing Set (RTTS)	4,322	real	outdoor	Yes

LHPRain 数据集

<https://yunguo224.github.io/LHP-Rain.github.io/>



LHP-Rain dataset contains various rain patterns, such as streaks, veiling, occlusion, and the first dataset to claim and tackle **ground splashing** problem.



Snow100K 数据集

<https://sites.google.com/view/yunfuliu/desnownet>



This dataset consists of 1) 100k synthesized snowy images, 2) corresponding snow-free ground truth images and 3) snow masks, and 4) 1,329 realistic snowy images. The snow-free and snowy realistic images 2) and 4) are downloaded from Flickr, and we manually validate each sample to see whether it is snow-free or not. The largest size of the image boundary is 640 pixels. This dataset consists of three subsets as per the variations inside single image. 1) Snow100K-S: Samples in this subset are only synthesized with small snow particles. 2) Snow100K-M: It contains the samples that are of both snow particles in small and medium sizes. 3) Snow100K-L: Snow particles of sizes small, medium, and large are all used for synthesizing samples. Each subset contains around 33k images.

参考文献:

- [1] Zhu Y, Wang T, Fu X, et al. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 21747-21758.
- [2] Valanarasu J, Yasarla R, Patel M. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 2353-2363.
- [3] Patil W, Gupta S, Rana S, et al. Multi-weather Image Restoration via Domain Translation. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 21696-21705.
- [4] Kulkarni A, Phutke S, Murala S. Unified transformer network for multi-weather image restoration. European Conference on Computer Vision. 2022: 344-360.

[5] Ozan Özdenizci and Robert Legenstein. Restoring Vision in Adverse Weather Conditions with Patch-based Denoising Diffusion Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.

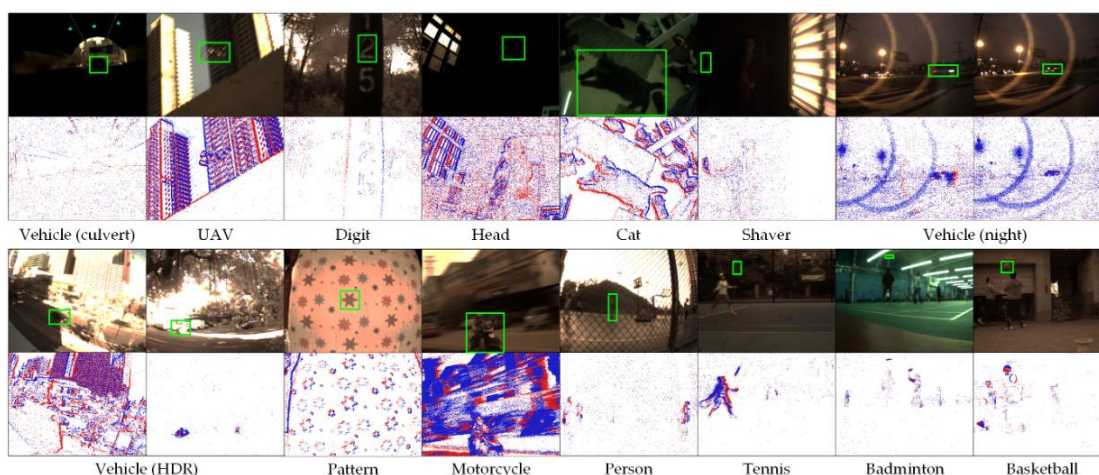
指导老师：昌毅，yichang@hust.edu.cn，qq 458449613

4.65 基于事件相机的目标跟踪

问题描述：在复杂动态场景中，目标跟踪是视觉感知系统中的核心问题之一。传统基于帧相机的目标跟踪方法依赖固定帧率采样与积分成像机制，易受到运动模糊、遮挡干扰以及光照剧烈变化的影响，难以在高速运动或高动态范围条件下保持稳定性能。同时，帧式数据存在时间冗余与信息延迟的问题，不利于实时性要求较高的应用场景。相比之下，事件相机采用像素级异步触发机制，仅在亮度发生变化时输出事件流，具备微秒级时间分辨率与高动态范围特性，为目标的连续时空建模提供了全新的数据表达形式。因此，如何充分挖掘事件数据的时空稀疏结构，并与传统图像信息进行有效融合，成为提升目标跟踪鲁棒性与实时性的关键问题。本课题要求完成基于事件相机的目标跟踪相关文献调研并形成综述报告，重点研究极端光照或高速场景下的目标跟踪方法，探索事件与图像融合的跟踪框架设计，对 ViPT[1]或 SDSTrack[2]算法进行复现与训练测试，并结合评价指标（如精度、成功率及鲁棒性）对跟踪性能进行系统分析。

数据描述：

VisEvent[4]数据集 https://github.com/wangxiao5791509/VisEvent_SOT_Benchmark

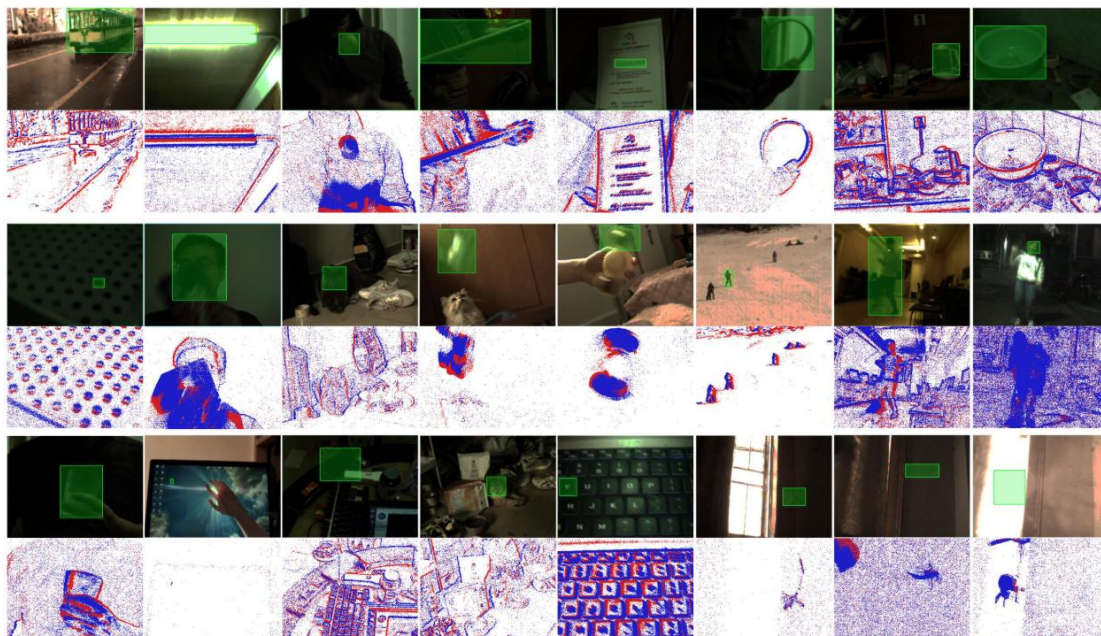


The VisEvent is developed to provide a dedicated platform for the training and evaluation of Visible-Event tracking algorithms. It collects 820 video pairs with an average of 450 frames

for each video. This dataset contains 709 and 111 videos for short-term and longterm tracking which will be beneficial for constructing a robust and flexible tracker. The resolution of dual modalities is 346×260 . The target objects are UAV, Hand, Pen, Bottle, Tank, Toy, Car, Tennis, Pedestrian, Badminton, Basketball, Book, Plant, Shoes, Phone, Laptop, Bag and Cat.

COESOT[5]数据集 <https://github.com/Event-AHU/COESOT>

COESOT is a general dataset for Color-Event camera based single object tracking. It contains 1354 color-event videos with 478,721 color frames and corresponding event streams. It splits sequences into a training and testing subset, which contains 827 and 527 videos, respectively. The videos are collected from both outdoor and indoor scenarios (e.g., street, zoo, and home) using the DVS346 event camera with a zoom lens.



参考文献:

- [1] Zhu J, Lai S, Chen X, et al. Visual prompt multi-modal tracking. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2023: 9516-9526.
- [2] Hou X, Xing J, Qian Y, et al. Sdstrack: Self-distillation symmetric adapter learning for multi-modal visual object tracking. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2024: 26551-26561.
- [3] Sun C., Zhang J., Wang Y., et al. Exploring Historical Information for RGBE Visual Tracking with Mamba. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2025: 6500-6509.
- [4] Wang X, Li J, Zhu L, et al. Visevent: Reliable object tracking via collaboration of frame

and event flows. IEEE transactions on cybernetics. 2023, 54(3):1997-2010.

[5] Tang C, Wang X, Huang J, et al. Revisiting color-event based tracking: A unified network, dataset, and metric. Pattern Recognition. 2025, 7:112718.

指导老师：昌毅，yichang@hust.edu.cn，qq 458449613

4.66 基于 3D 人体姿态估计

任务描述：本任务旨在深入研究单目视频中的 3D 人体姿态估计技术。该技术是计算机视觉的核心挑战之一，旨在从单个摄像头的 2D 视频中恢复出人体在三维空间中的关节点的精确时序位置。本研究将首先复现一篇经典或前沿论文的开源实现，在标准数据集上进行全面评测；进而探索并实现一种或多种性能提升策略，最终在更具挑战性的真实场景中进行定性验证，展示模型的泛化能力。

任务需求和步骤：

1. 复现论文[1]的开源内容，并在在 H3.6M 中选取 5 种以上动作类型数据进行测试分别报告定量结果。
2. 尝试在此基础上提升 3D 姿态估计方法的性能。
3. 自行从网上收集一段动作视频 (>10S) 进行 3D 姿态估计，并从至少三个不同视角展示提取的 3D 动作

评价指标：

1. 定量指标：平均关节点误差 (MPJPE)，正确关节点百分比 (PCK)。
2. 定性指标：从 3 种可视化视频中的动作。

备注：为提升性能可自主收集新的数据训练方法，显存要求约 10G。

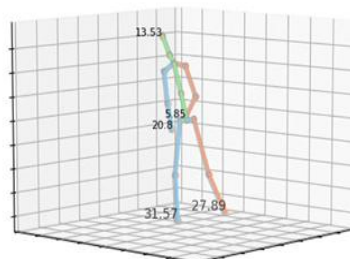
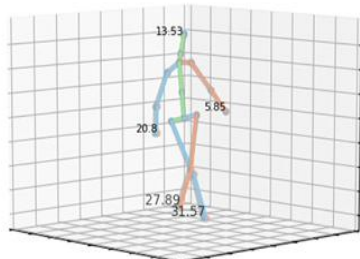
数据说明：

本次作业使用以下公开数据集，**要求至少完成在 H3.6M 数据集上的实验**，鼓励在更多样化的数据集上进行训练与验证。

(1) Human3.6M 数据集

- Human3.6M 是一个在室内场景下采集的 3D 人体姿态数据集，包含 11 位演员共 15 种动作类型。
- **内容：**超过 3.6M 个视频帧，包含丰富的户外活动。

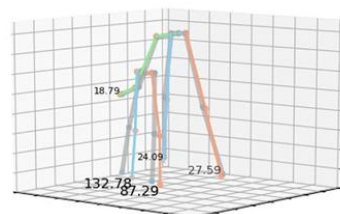
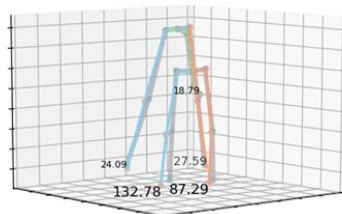
- **标注:** 提供了图像、相机参数，以及每个人物的 3D 关节位置。
- **用途:** 主要用于测试与评估。



数据集地址 [Human3.6M Dataset](#)

(2) MPI-INF-3DHP (选做)

- MPI-INF-3DHP 包含 3 种场景
- **内容:** 包含 8 位演员，8 类动作用于训练。7 类动作用于测试
- **标注:** 提供了图像、相机参数，以及每个人物的 3D 关节位置。



数据集地址 [Monocular 3D Human Pose Estimation In The Wild Using Improved CNN Supervision, 3DV 2017](#)

参考文献

[1] Tang Z, Qiu Z, Hao Y, et al. 3d human pose estimation with spatio-temporal criss-cross attention[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 4790-4799.

[2] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(7):1325–1339, 2013.

[3] Liu Y, Qiu C, Zhang Z. Deep learning for 3D human pose estimation and mesh recovery: A survey[J]. Neurocomputing, 2024, 596: 128049.

[4] Dushyant Mehta, Helge Rhodin, Dan Casas, Pascal Fua, Oleksandr Sotnychenko, Weipeng Xu, and Christian Theobalt. Monocular 3d human pose estimation in the wild using improved cnn supervision. In 3DV, 2017. 5

指导老师：谭毅华，yhtan@hust.edu.cn，[wx:Hust18971386510](https://weibo.com/Hust18971386510)

4.67 动态场景下的 SLAM 重建

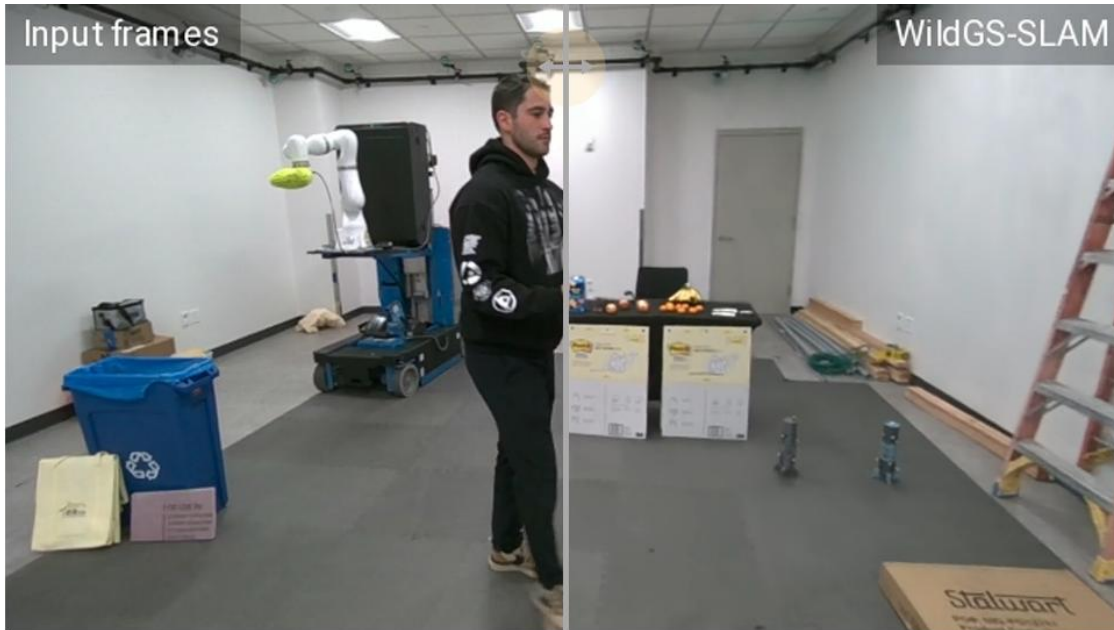
任务描述：在现实世界中，视觉传感器不可避免地会观测到移动的行人、车辆等动态物体。传统的即时定位与建图（SLAM）和三维重建算法通常建立在“场景是绝对静态”的强假设之上，动态物体的出现会导致严重的特征误匹配、相机轨迹漂移甚至系统崩溃。

本课程设计要求针对包含复杂动态干扰的公开单目视频数据集，设计并实现一个鲁棒的视觉 SLAM 与三维重建系统。典型数据集包括：Wild-SLAM、DROID-W。



任务和需求：

视动态物体为“噪声”。系统需准确估计相机位姿，并在建图过程中识别并彻底剔除动态物体，最终输出一个纯净、无伪影的静态背景三维地图，示例如下所示：



在实现过程中，需重点解决以下核心问题：

1. **前端特征与深度处理：** 提取鲁棒的图像特征（如光流特征），或利用预训练模型先验获取单目深度图。
2. **动态区域的掩码与不确定性估计：** 设计算法（如基于多层感知机预测、几何一致性检验或光流残差）来评估像素级的置信度，精准定位动态区域。
3. **现代三维场景表示：** 探索使用点云或基于 3D 高斯溅射（3D Gaussian Splatting）等前沿技术进行场景表达与渲染。
4. **进阶探索（加分项）：** 探索系统在真实复杂环境或计算受限平台上的工程落地。例如，梳理复杂的编译依赖关系并在 Linux/WSL 环境下成功构建系统，或尝试将位姿输出接口与 ROS 系统桥接，为无人机等智能体的下游导航任务提供支持。

在实现过程中，需重点解决以下问题：

1. 图像特征的提取与处理（如利用预训练的 DINOv2 等视觉大模型提取鲁棒特征）。
2. 像素级不确定性的估计（如设计轻量级的多层感知机 MLP 预测不确定性图）。
3. 基于 3D 高斯溅射的场景几何表示与实时渲染更新机制。
4. 不确定性感知的相机位姿追踪。
5. 动态物体剔除与纯静态 3D 高斯地图的联合优化。

备注： 计算资源大约 8GB GPU。

参考文献：

- [1] Li Z, Tucker R, Cole F, et al. Megasam: Accurate, fast and robust structure and motion from casual dynamic videos[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2025: 10486-10496.
- [2] Zheng J, Zhu Z, Bieri V, et al. Wildgs-slam: Monocular gaussian splatting slam in dynamic environments[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2025: 11461-11471.
- [3] Li M, Zhu Z, Pollefeys M, et al. DROID-SLAM in the Wild[J]. arXiv preprint arXiv:2603.19076, 2026.
- [4] Bescos B, Fàcil J M, Civera J, et al. DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes[J]. IEEE robotics and automation letters, 2018, 3(4): 4076-4083.
- [5] Yu C, Liu Z, Liu X J, et al. DS-SLAM: A semantic visual SLAM towards dynamic environments[C]//2018 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, 2018: 1168-1174.

指导老师：谭毅华， yhtan@hust.edu.cn， [wx:Hust18971386510](https://github.com/yhtan)

4.68 基于高斯泼溅的三维人体重建

任务描述：

三维人体重建是计算机视觉与计算机图形学领域的重要研究方向，其目标是从单视角或多视角视频中恢复出逼真的人体几何形状与外观纹理。近年来，三维高斯泼溅（3D Gaussian Splatting, 3DGS）作为一种新颖的显式辐射场表示方法，凭借其极高的渲染质量和实时渲染速度，在三维重建领域取得了突破性进展。

本任务要求针对公开的人体视频数据集，设计并实现一个基于高斯泼溅的动态三维人体重建系统。系统应能够处理包含复杂动作的视频数据，通过引入参数化人体模型（SMPL）作为先验结构，在模型表面或顶点初始化高斯点，并自动优化高斯参数，最终实现高质量的新视角合成与动态人体重建。

任务和需求：

1. 复现 Gauhuman 项目论文中的结果（<https://github.com/skhu101/GauHuman>）。

2. 尝试在此基础上提升性能。

在实现过程中，需重点解决以下问题：

1. 人体先验模型的获取与对齐(如从图像特征中提取或利用预计算数据生成精确的 SMPL 模型)；

2. 三维高斯点的初始化策略(在 SMPL 模型顶点上初始化高斯点的位置、缩放、旋转、不透明度和球谐函数)；

3. 结合可微渲染框架的高斯参数优化(利用目标图像作为监督信号更新高斯属性)；

4. 动态人体形变的处理(利用 SMPL 的蒙皮权重或形变场，驱动高斯点随人体运动发生合理的非刚性形变)；

5. 探索更高效的网络架构或正则化策略，实现高质量、抗遮挡、高帧率的三维人体渲染。

要求完成一个完整的重建流程，并能够对重建结果进行可视化展示(如生成任意视角的视频、对比 SMPL 骨架与渲染结果的对齐情况等)。此外，应分析不同因素(如输入视角数量、高斯点初始化密度、动作复杂程度等)对重建性能和渲染质量的影响。

指标：

1. 定量指标:PSNR、SSIM、LPIPS
2. 定性可视化结果

备注：算力需求大约 GPU 24GB。

数据说明

本实验使用以下公开三维人体数据集：

(1) ZJU_MoCap 数据集(基本)

ZJU_MoCap 是动态三维人体重建和新视角合成领域的经典多视角基准数据集，适用于验证高斯人体模型在多视角下的有效性。



- 包含多个受试者执行复杂动作(如跳舞、武术、热身等)的视频
- 提供精确的相机内参和外参矩阵
- 提供预计算的 3D 人体关键点、SMPL 参数(形状与姿态)以及前景掩码
- 具有标准的多视角训练集与新视角测试集划分

数据集下载及参考地址：<https://chingswy.github.io/Dataset-Demo/>

(2) DNA_rendering 数据集（扩展）

DNA_rendering 数据集包含多目视角的动态人体视频，适用于拓展研究服装更具挑战性的人体重建任务。



- 包含海量的高精度多视角人体动态视频，涵盖多样的日常服装、复杂的表演服饰（如宽松的裙摆、汉服等）以及丰富的肢体动作。
- 提供高分辨率图像（由几十个视角的专业相机阵列同步拍摄）、高精度的相机内外参校准数据，以及配套的 SMPL/SMPL-X 模型参数和高质量的前景掩码。
- 极大地考验三维重建系统在处理复杂服装拓扑（如衣物褶皱、非刚性动态形变）以及恢复高保真纹理细节时的特征表示与渲染优化能力。

数据集下载：<https://dna-rendering.github.io/>

参考文献

[1] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1 - 248:16, 2015.

[2] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graphics*, 42(4):139 - 1, 2023.

指导老师：谭毅华，yhtan@hust.edu.cn, [wx:Hust18971386510](https://weibo.com/Hust18971386510)

4.69 基于深度学习的时序动作检测

任务描述：时序动作检测（Temporal Action Detection, TAD）是视频理解领域的重要研究方向，其目标是在未裁剪视频（untrimmed video）中自动定位动作发生的时间区间，并识别对应的动作类别。与视频分类任务不同，时序动作检测不仅需要判断“视频中是否存在某种动作”，还需要精确给出动作的起止时间，具有更高的挑战性。

本任务要求针对公开视频数据集，设计并实现一个基于深度学习的时序动作检测系统。系统应能够在包含复杂背景、动作持续时间变化较大、动作边界模糊等情况下，自动检测视频中的多个动作实例，并输出每个动作的类别标签、时间区间（起始时间与结束时间）及对应置信度。

任务和需求：

1. 复现 TriDet 项目论文中的结果 (<https://github.com/dingfengshi/tridet>)
2. 尝试在此基础上提升性能

在实现过程中，需重点解决以下问题：

1. 视频特征的提取与表示（如基于 3D CNN 或预提取特征）；
2. 时序信息建模（如基于卷积、RNN 或 Transformer 的结构）；
3. 动作候选区间生成（proposal generation）或直接回归动作边界；
4. 多实例检测与重叠动作的处理；
5. 在得到动作的类别、起止时间之外，探索如何获取执行该动作的 subject
6. 探索更高效的网络架构实现**端到端、实时处理**。

要求完成一个完整的检测流程，并能够对检测结果进行可视化展示（如在时间轴上标注动作区间、执行动作的 subject 等）。此外，应分析不同因素（如动作时长、视频复杂度、特征分辨率等）对检测性能的影响。

数据说明

本实验使用以下公开视频数据集：

（1）THUMOS14 数据集（基本）

THUMOS14 是时序动作检测领域经典基准数据集之一，包含大量未裁剪视频及精确的时间标注。

- 包含约 400 余个视频样本
- 涉及 20 类动作（主要为体育类动作）
- 提供训练集、验证集与测试集划分
- 每个动作实例均标注时间起止位置

数据集下载地址：<http://www.crev.ucf.edu/THUMOS14/>

（2）ActivityNet 数据集（扩展）

ActivityNet 是更大规模的视频理解数据集，适用于拓展研究：

- 约 2 万个视频

- 200 类动作
- 视频内容更加多样化

数据集下载地址：<http://activity-net.org/>

指标：

1. 定量指标: mAP（不同 IoU 阈值）
2. 定性可视化结果



备注：算力需求大约 GPU 16GB。

参考文献

- [1] Shou Z., Wang D., Chang S.F. Temporal Action Localization in Untrimmed Videos via Multi-stage CNNs. CVPR, 2016.
- [2] Idrees H. et al. The THUMOS Challenge on Action Recognition for Videos in the Wild. 2016.
- [3] Heilbron F.C. et al. ActivityNet: A Large-Scale Video Benchmark for Human Activity Understanding. CVPR, 2015.
- [4] Lin T. et al. BMN: Boundary-Matching Network for Temporal Action Proposal Generation. ICCV, 2019.
- [5] Zhang C.L. et al. ActionFormer: Localizing Moments of Actions with Transformers. ECCV, 2022.
- [6] Wang B. et al. Temporal Action Localization in the Deep Learning Era: A Survey. IEEE TPAMI, 2024.

指导老师：谭毅华，yhtan@hust.edu.cn，wx:Hust18971386510

4.70 基于视觉和语言模型的视频燃烧状态描述

任务描述：识别炉内视频中的燃烧状态对于降低工业成本具有重要意义。早期方法通常采用事件分类模型，将燃烧状态预定义为若干固定类别。然而，由于燃烧过程复杂、专业性且状态边界模糊，人工划分类别不仅困难，还难以捕捉细粒度特征。随着多模态大模型的发展，使用自然语言对燃烧状态进行描述成为可能。本课题旨在构建面向燃烧

状态的文本描述数据集，并结合视频大语言模型的监督微调和思维链生成方法，对 Qwen3.5 或 Qwen3-VL 进行微调，实现对炉内视频燃烧状态的精准描述与分析。

(1) 数据标注

原始视频请联系向璇渝（QQ：2542024726），标注 1-5K 视频描述训练样本。包含以下内容：

1. 燃烧现象描述：火焰呈明亮黄色，波动剧烈，局部有黑烟产生…
2. 状态判断：当前为不完全燃烧状态/当前存在刺火/当前有车杆介入…
3. 原因分析（思维链）：
 - a) 火焰偏黄 → 碳颗粒未完全燃烧
 - b) 出现黑烟 → 氧气不足
 - c) 火焰波动大 → 燃料供给不稳定

(2) 微调视频语言大模型实现燃烧状态描述

结合标注好的训练数据，对 Qwen3.5 或 Qwen3-VL 进行微调，对视频进行准确且自然的自然语言描述。



视频语言大模型

开始时，熔炉内火焰处于旺盛燃烧状态，呈现出持续而稳定的橙红色，充满整个炉腔空间。随后，在02:50:50至56秒期间，火焰保持活跃但无剧烈异常波动；接着，画面整体未显示出刺火、车杆操作或其他外部干预迹象，火焰始终维持在均匀且高强度的喷射形态。最后，视频片段结束前，炉内仍保持同样的稳定燃烧态势，无明显衰退或变化。

备注：该课题对 GPU 算力需求较大，团队成员可联系向璇渝（QQ：2542024726）借用实验室算力。

参考文献：

[1] Bai S. et al, Qwen3-VL Technical Report, arXiv, 2025.

指导老师：谭毅华，yhtan@hust.edu.cn, [wx:Hust18971386510](https://weixin.qq.com/w/Hust18971386510)

4.71 雾天条件下的目标检测

任务描述：现实世界中的目标检测常常会遭遇各种各样的大气扰动，尤其是恶劣的天气状况，在这种情况下，现有的检测器性能往往会急剧下降。鉴于全球气候波动的日益加剧，雾天的出现频率不断增加，这已成为一个重大挑战。在雾天环境中提高检测的稳健性不仅对于自动驾驶的安全至关重要，对于弥合理论模型与实际现实应用之间的差距也

极为关键。

具体方案:

1. 基于图像修复的检测方案: 首先利用图像去雾模型恢复雾图的可视性, 随后对复原后的图像进行检测;
2. 基于多任务学习的检测方案: 将图像去雾模型与目标检测模型进行并联, 共享骨干网络参数, 通过并行子网同时学习图像去雾和目标检测能力;
3. 基于领域自适应的检测方案: 利用域适应方法解决雾天条件的域偏移, 提高检测性能;
4. 从上述方案中选择一个进行实现, 计算测试集上的检测指标并展示可视化检测结果;
5. 加分项: 提出自己的改进方法, 在基础实现上进一步提升性能。

数据说明:

VOC_hazy: 基于 PASCAL VOC 数据集, 人工合成相应的雾图数据集

Foggy Cityscapes: 基于 Cityscapes 数据集, 官方通过估计的深度图合成的雾图数据集

RTTS: 真实采集的雾图数据集



备注: 本课题对 GPU 显存需求 < 11GB。

参考文献:

- [1] Li, Boyi, et al. "Aod-net: All-in-one dehazing network." *Proceedings of the IEEE international conference on computer vision*. 2017.
- [2] Li, Chengyang, et al. "Detection-friendly dehazing: Object detection in real-world hazy scenes." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.7 (2023): 8284-8295.
- [3] Huang, Shih-Chia, Trung-Hieu Le, and Da-Wei Jaw. "DSNet: Joint semantic learning for object detection in inclement weather conditions." *IEEE transactions on pattern analysis and machine intelligence* 43.8 (2020): 2623-2633.
- [4] Liu, Yajing, et al. "Unbiased faster r-cnn for single-source domain generalized object

1. mAP@0.5:0.95 (尤其关注小目标类别的 AP_S)
2. FPS (Frames Per Second)

数据说明:

数据集 Anti-UAV: 数据集链接 <https://github.com/ucas-vg/Anti-UAV>

DroneVehicle: 数据集链接: <https://github.com/VisDrone/DroneVehicle>

ESCVehicle: <https://github.com/sjm2001-rslab/>

https://pan.baidu.com/s/1_Ahx2ji8jpbQhDv78J97tA?pwd=1109

备注: 可以选择基于 YOLO 框架的方法, GPU 显存需求 ≤ 10 GB

参考资料与可入手的关键技术点:

- 弱配准鲁棒融合 (OAFA): CVPR 2024 提出的 OAFA 方法, 显式建模模态间偏移, 利用可变形对齐解决非严格像素级对齐问题。[参考文献 3]
- 条件感知动态融合 (ATR-UMOD): ICCV 2025 最新进展, 根据光照、天气、角度等条件动态调整双模态权重, 避免“一刀切”融合。[参考文献 12]:
- 模态缺失鲁棒性 (Scarf-DETR): 针对传感器掉帧或失效场景, 实现单/双模态自适应切换的检测框架。[参考文献 4]
- 轻量化实时检测 (TAF-YOLO / DVIF-Net): 针对机载嵌入式部署, 优化 YOLO 架构, 引入频域分解与轻量融合模块, 平衡精度与速度。[参考文献 10]
- 端到端联合优化 (UniFusOD): 将图像融合与目标检测任务统一建模, 以检测性能为导向反向优化融合特征。

参考文献:

- [1] Wang, Y., et al. "How Far are Modern Trackers from UAV-Anti-UAV? A Million-Scale Benchmark and New Baseline." UAV-Anti-UAV Benchmark, 2025.
- [2] Li, X., et al. "TMRGBT-D2D: A Temporal Misaligned RGB-Thermal Dataset for Drone-to-Drone Target Detection." IEEE Transactions on Circuits and Systems for Video Technology, 2024.
- [3] Zhang, H., et al. "Weakly Misalignment-free Adaptive Feature Alignment for UAVs-based Multimodal Object Detection (OAFA)." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.
- [4] Chen, L., et al. "On Modality Incomplete Infrared-Visible Object Detection: An Architecture Compatibility Perspective (Scarf-DETR)." arXiv preprint arXiv:2501.xxxxx,

2025.

[5] Jia, X., et al. "LLVIP: A Visible-infrared Paired Dataset for Low-light Vision." Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021.

[6] Liu, Z., et al. "Fusion Meets Diverse Conditions: A High-diversity Benchmark and Baseline for UAV-based Multimodal Object Detection with Condition Cues." Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2025.

[7] Zhao, K., et al. "Rethinking Multi-modal Object Detection from the Perspective of Mono-Modality Feature Learning (M2D-LIF)." IEEE Transactions on Image Processing, 2025.

[8] Sun, Q., et al. "Infrared-Visible Image Fusion Meets Object Detection: Towards Unified Optimization for Multimodal Perception (UniFusOD)." Proceedings of the AAAI Conference on Artificial Intelligence, 2025.

[9] Wu, J., et al. "A Small-Target Detection Network for UAV Aerial Images Based on Visible and Infrared Fusion (DVIF-Net)." Remote Sensing, 2025.

[10] Guo, R., et al. "TAF-YOLO: A Small-Object Detection Network for UAV Aerial Imagery via Visible and Infrared Adaptive Fusion." IEEE Geoscience and Remote Sensing Letters, 2025.

[11] Yang, S., et al. "Frequency-Driven Feature Decomposition Network for Infrared-Visible Object Detection (FD2-Net)." Pattern Recognition, 2025.

[12] Rethinking Multi-modal Object Detection from the Perspective of Mono-Modality Feature Learning. ICCV, 2025

指导老师：谭毅华，yhtan@hust.edu.cn，[wx:Hust18971386510](https://weibo.com/Hust18971386510)

4.73 音视频融合的网球击球动作定位与识别

任务描述：网球击球动作（如发球、正手、反手）识别是体育分析的核心任务。传统纯视觉方法需对视频逐帧处理，计算量大且难以实时。击球瞬间伴随清脆的撞击声，音频信号具有更高的时间分辨率。本课题拟研发一种音视频融合策略：利用音频快速定位击球时刻，仅在关键时间窗内进行视觉分析，实现高效率、高精度的动作识别。

具体方案：

1. 音频事件检测模块：从视频中分离音频，检测击球声事件，输出精确的时间戳。

2.关键帧采样策略：根据音频时间戳，在击球前后（如 -0.5s 至 +0.5s）进行密集帧采样，替代全视频处理。

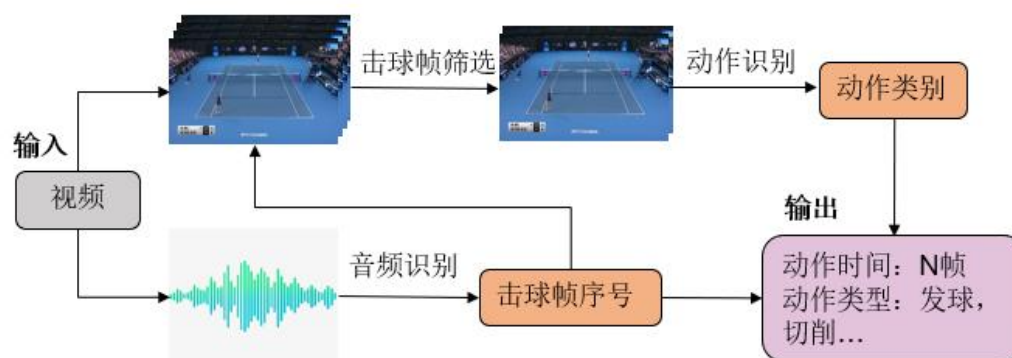
3.视觉动作识别模块：基于采样的关键帧序列，识别具体的击球动作类别（六分类）。

4.构建基于音视频融合的动作定位与分类算法框架：输入视频数据，输出动作发生的时间和动作类别。

数据说明：F3set 是一个专为分析视频中快速、频繁、细粒度事件而设计的基准数据集，其网球分部包含了 152 段职业比赛视频，并精确标注了每个回合中击球发生的帧序号以及六种具体击球动作，包括平击、削球、截击、挑高球、放小球和高压球。该数据集的标注高度适合用于验证音视频融合方案，其帧级精确的时间戳可直接用于模拟音频事件检测的输出，从而允许在已知的击球时刻前后进行密集的关键帧采样，以构建视觉动作识别模块所需的时序输入。在准备数据的过程中，需要从 [F3Set/data/f3set-tennis/videos.csv at main · F3Set/F3Set](https://github.com/F3Set/F3Set/blob/main/data/f3set-tennis/videos.csv) 中下载对应的 youtube 视频，与 github 上提供的标签文件进行对应，并基于训练集优化算法，在验证集中进行验证。



F3set 数据集示例



算法流程示例

备注：

1. https://github.com/qiuqiangkong/audioset_tagging_cnn（音频事件分类，显存<2GB）
2. <https://github.com/ultralytics/ultralytics>(动作分类，可用 YOLOv8-pose+GRU，显存<2GB)

参考文献：

- [1] Liu, Zhaoyu, Kan Jiang, Murong Ma, Zhe Hou, Yun Lin, and Jin Song Dong. "F³ Set: Towards Analyzing Fast, Frequent, and Fine-grained Events from Videos." arXiv preprint arXiv:2504.08222 (2025).
- [2] Yan, F., Kittler, J., Windridge, D., Christmas, W., Mikolajczyk, K., Cox, S. and Huang, Q., 2014. Automatic annotation of tennis games: An integration of audio, vision, and learning. *Image and Vision Computing*, 32(11), pp.896-903.
- [3] Liu, Zhaoyu, Xi Weng, Lianyu Hu, Zhe Hou, Kan Jiang, Jin Song Dong, and Yang Liu. "TennisExpert: Towards Expert-Level Analytical Sports Video Understanding." arXiv preprint arXiv:2603.13397 (2026).
- [4] Baughman, Aaron, Eduardo Morales, Gary Reiss, Nancy Greco, Stephen Hammer, and Shiqiang Wang. "Detection of tennis events from acoustic data." In *Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports*, pp. 91-99. 2019.

指导老师：谭毅华，yhtan@hust.edu.cn，wx:Hust18971386510

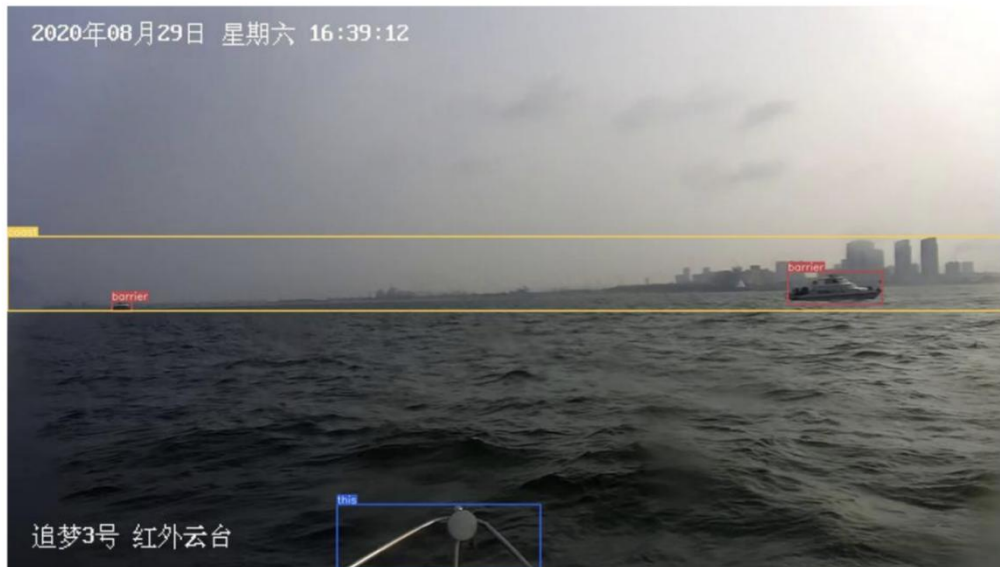
4.74 真实环境海面目标检测

任务描述：在近年实海域航行中，出现了诸如：海面反光导致目标检测虚警、波浪抨击导致跟踪目标突然丢失、飞溅液滴悬挂导致光学识别失效等特殊场景的人工智能算法应用问题。海洋智能装备与系统最典型的技术特征就是具有智能演进能力。有鉴于此基于实海域试验场，积累大量测试数据，持续更新迭代算法参数与架构，实现海洋装备智能演进，是智能船艇发展的关键路径。通过真实环境下的水面实景视频（训练集为海面船艇海试过程中录制的影像视频），检测出不同海上目标（如船只、海岸等）在画面中的位置。

数据描述：需检测的目标类别包括障碍物“barrier”，海岸“coast”两类。训练数据真值中可能存在 **this“船体”** 这一类别，这一类不需要检测。需自行处理数据缺失、异常和少量的质量问题。

数据集请于南一楼程骏处拷贝。联系方式：微信 15673628667

数据标注样例：



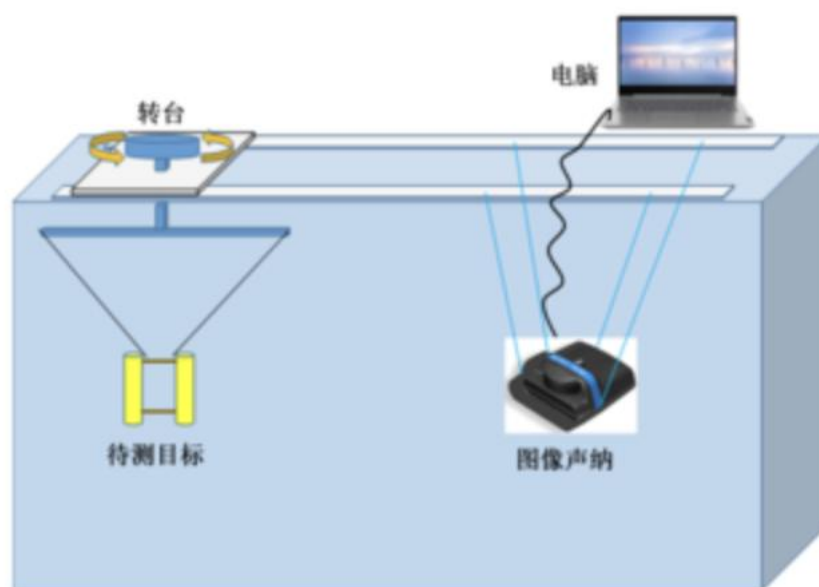
参考文献:

- [1] Y. Huang, H. Wang, J. Ma, J. Lou and H. Yi, "Research and Practical Exploration of Test and Validation Technologies Applied on Unmanned Surface Vehicle Optical Recognition," *2021 IEEE*
- [2] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation.
- [3] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs.
- [4] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation.

指导老师: 程骋, c_cheng@hust.edu.cn, qq:286652621

4.75 智能水下感知

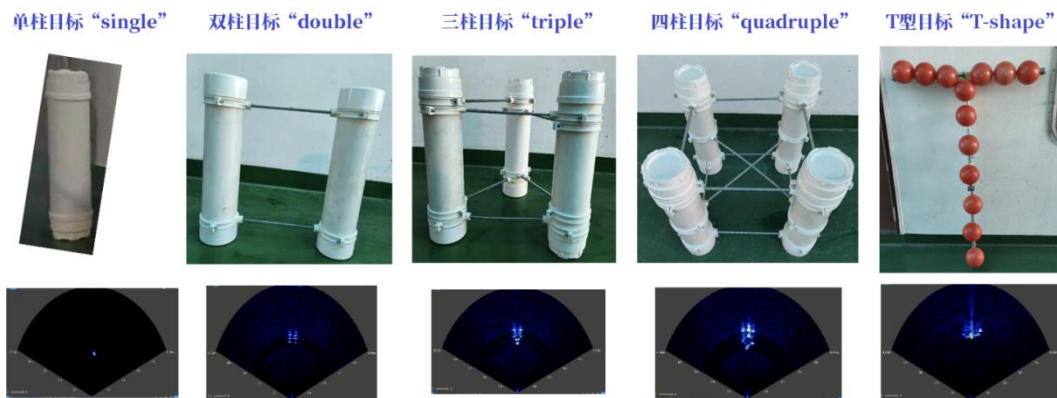
任务描述：声呐是一种声学探测设备，利用声波在水中的传播和反射特性，通过电声转换和信息处理进行探测、测距和导航，是水声学中应用最广泛、最重要的一种装置，广泛应用于搜救打捞、海洋工程、码头检查、环境监测、现场清理和水下调查等领域。将待测目标悬挂在实验转台下方，目标旋转 360°，获取目标全方位的声图像；将图像声呐吊放在目标同一深度，在目标旋转过程中，记录目标全方位的声图像。通过提供的真实环境下的全方位水下声呐图像，对不同种类水下目标进行分类标注。可以自行补充数据集以提高识别的泛化能力。



试验系统示意图

数据描述：需识别的目标类别包括单柱目标“single”，双柱目标“double”，三柱目标“triple”，四柱目标“quadruple”，T型目标“T-shape”五类。部分图像中有鱼游过，须自行处理。

训练集：240 张用于训练的图像，包含了五个类别目标的全方位探测数据；有 1000 张用于测试的图像；



备注：如遇数据缺失、异常和少量的质量问题，需自行处理。

数据集请于南一楼程骏处拷贝。联系方式：微信 15673628667

参考文献：

- [1] Y. Huang, H. Wang, J. Ma, J. Lou and H. Yi, "Research and Practical Exploration of Test and Validation Technologies Applied on Unmanned Surface Vehicle Optical Recognition," 2021 IEEE
- [2] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation.
- [3] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs.
- [4] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation.

指导老师：程骋, c_cheng@hust.edu.cn, qq:286652621

4.76 心电可解释性分析

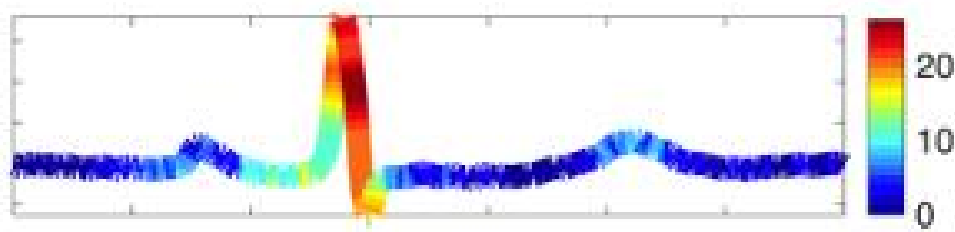
任务描述：心电图是心脏产生的电活动。先后有序的电激动的传播会引起一系列电位改变，形成心电图上的相应波段。主要有五个重要的波段：P波、Q波、R波、S波、T波，加上心房去极化导致的心房收缩前的小转折点。心电图当中比较重要的特征包括有：P波、PR间隔、QRS波群、ST段、T波和QT周期。标准的心电图称为12导联

心电图，它是由 12 根导线、10 个电极获得。通过分析心电图波形和间隔是否正常可以进行心电图解释。

数据描述：提供左心室肥大和正常两类 12 导联心电信号的 excel 数据。自行对信号提取特征然后基于决策树等模型进行分类和分析。

数据集在南一楼中 315 江一诺处拷贝。联系方式: QQ 25643385

数据可视化样例（单导联下重要性分析）：



参考文献：

- [1] Friesen G M, Jannett T C, Jadallah M A, et al. A comparison of the noise sensitivity of nine QRS detection algorithms[J]. IEEE Transactions on biomedical engineering, 1990, 37(1): 85-98.
- [2] Pan J, Tompkins W J. A real-time QRS detection algorithm[J]. IEEE transactions on biomedical engineering, 1985 (3): 230-236.
- [3] Li Q, Rajagopalan C, Clifford G D. A machine learning approach to multi-level ECG signal quality classification[J]. Computer methods and programs in biomedicine, 2014, 117(3): 435-447.
- [4] Song Y Y, Ying L U. Decision tree methods: applications for classification and prediction[J]. Shanghai archives of psychiatry, 2015, 27(2): 130.

指导老师：程骋, c_cheng@hust.edu.cn, qq:286652621

4.77 左心室肥大二分类检测

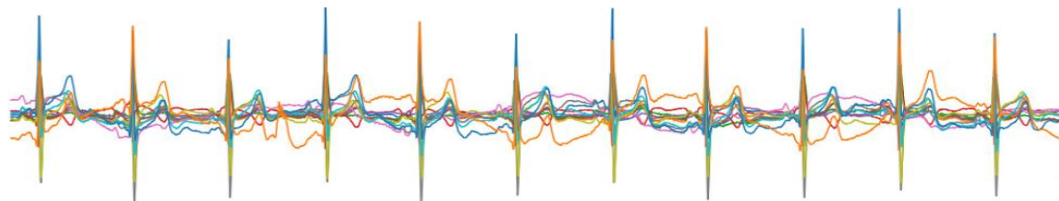
任务描述：左心室肥大在心内科临床诊断治疗中非常常见，可能导致高血压等疾病。因此，对左心室肥大进行及时有效的诊断，让患者早发现早治疗，具有重要的临床意义。临床上心电图和心脏彩超是心室肥大诊断的常用手段，但心脏彩超检查价格昂贵并容易受到设备和检查时间的限制，而心电图诊断作为一种经济实用、简单有效的方法，在心

脏类疾病诊断中应用广泛。利用深度学习进行左心室肥大心电图诊断，可以有效提取心电图的复杂特征，摆脱传统人工诊断的局限性。

数据描述：提供左心室肥大和正常两类心电信号的 excel 数据，数据维度为（5000，12）。所提供的数据类别即为 B 超诊断的标签，利用心电信号进行二分类诊断。

数据集在南一楼中 315 江一诺处拷贝。联系方式: QQ 25643385

数据可视化样例（12 导联心电信号）：



参考文献：

[1] Zhu H, Cheng C, Yin H, et al. Automatic multilabel electrocardiogram diagnosis of heart rhythm or conduction abnormalities with deep learning: a cohort study[J]. The Lancet Digital Health, 2020, 2(7): e348-e357.

[2] Al Rahhal M M, Bazi Y, Al Zuair M, et al. Convolutional neural networks for electrocardiogram classification[J]. Journal of Medical and Biological Engineering, 2018, 38(6): 1014-1025.

[3] Zubair M, Kim J, Yoon C. An automated ECG beat classification system using convolutional neural networks[C]. in: 2016 6th international conference on IT convergence and security (ICITCS). IEEE, 2016: 1-5.

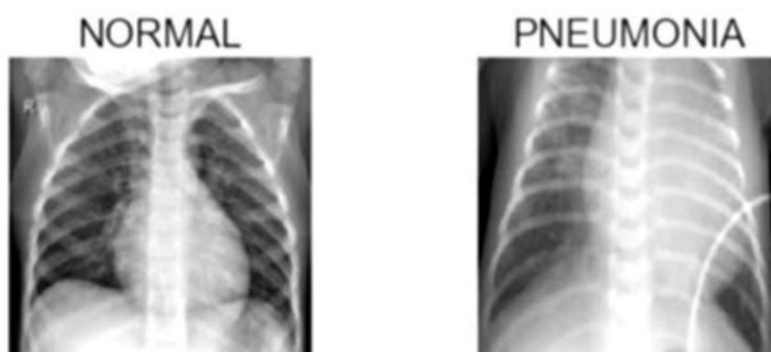
[4] Hannun A Y, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network[J]. Nature medicine, 2019, 25(1): 65-69.

指导老师：程骋, c_cheng@hust.edu.cn, qq:286652621

4.78 肺炎 X 光图像诊断识别

任务描述：胸部 X 光图像是肺炎辅助诊断的重要依据。肺炎患者在 X 光图像中通常会出现肺部阴影增强、纹理异常等表现，而正常样本的肺部结构相对清晰。通过对胸部 X 光图像进行分析，可以提取图像特征，并据此建立分类模型，实现肺炎与正常样本的自动识别。本任务提供肺炎和正常两类胸部 X 光图像数据。自行对图像进行预处理，然后

利用深度学习（EfficientNetV2、ViT 等）进行二分类。可进一步结合特征重要性、关键区域可视化等方法，对分类结果进行解释。



数据说明： 提供肺炎和正常两类胸部 X 光图像公开数据。推荐使用以下数据集：Chest X-Ray Images (Pneumonia) 。数据集包含 5863 张 JPEG 胸部 X 光图像，分为 PNEUMONIA 和 NORMAL 两类，目录中已按训练集、测试集和验证集组织。联系方式：陈煜璇（QQ）2513611861

参考文献：

- [1] Litjens G, Kooi T, Bejnordi B E, et al. A survey on deep learning in medical image analysis[J]. Medical Image Analysis, 2017.
- [2] Rajpurkar P, Irvin J, Zhu K, et al. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning[J]. arXiv preprint arXiv:1711.05225, 2017.
- [3] Efficient pneumonia detection using Vision Transformers on chest X-rays[J]. Scientific Reports, 2024.

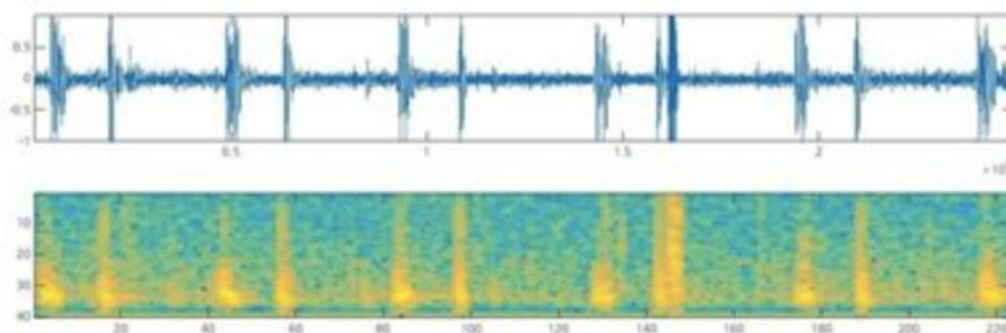
指导老师：程骋, c_cheng@hust.edu.cn, qq:286652621

4.79 心音异常识别

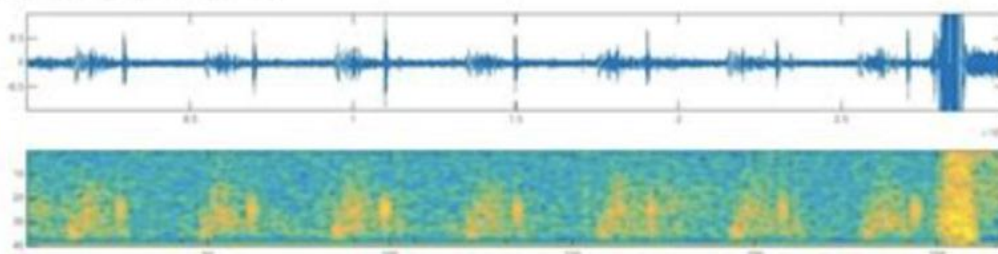
任务描述：心音是心脏活动过程中产生的声音信号。正常心音通常具有较稳定的节律和波形结构，而异常心音中常伴随杂音、额外心音或频率分布变化。通过对心音信号进行分析，可以提取相关特征，并据此建立分类模型，实现正常与异常心音的自动识别。

本任务提供正常心音和异常心音两类公开数据。要求学生对信号进行预处理和特征提取，并采用传统机器学习方法（如决策树、SVM 等）或深度学习方法（如 RNN、LSTM 等）进行分类和分析，完成正常/异常心音的二分类识别任务。

(A) Normal



(B) Aortic stenosis



数据说明：

提供正常心音和异常心音两类公开数据集：PhysioNet/Computing in Cardiology Challenge 2016。该数据集面向正常/异常心音分类任务，训练集和测试集中的录音被划分为 normal 和 abnormal 两类；正常录音来自健康受试者，异常录音来自经临床确诊存在心脏疾病的患者。联系方式：陈煜璇（QQ）2513611861。

参考文献：

- [1] Liu C, Springer D, Li Q, et al. An open access database for the evaluation of heart sound algorithms[J]. Physiological Measurement, 2016.
- [2] Clifford G D, Liu C, Moody B, et al. Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge 2016[J]. 2016.

[3] Noman F, Ting C M, Salleh S H, et al. A CNN-based classification of heart sound recordings[J].

指导老师：程骋, c_cheng@hust.edu.cn, qq:286652621

4.80 铣削加工刀具状态监测

任务描述：铣削加工过程中刀具状态监测是一项重要任务，刀具更换不及时会导致损坏昂贵工件，过早地更换刀具则会导致生产效率的下降。因此，刀具状态监测成为智能制造中重要的一项任务。目前刀具状态主流方法为数据驱动方法，即根据目标刀具的历史加工信号与当前加工信号预测当前一次加工完成后的刀具后刀面磨损宽度。其中主要任务为时间序列表征以及刀具状态识别，要求熟悉时间序列表征方法以及非自回归任务常用框架。

数据说明：PHM2010 数据集：包括 3 把铣刀全寿命周期铣削信号（切削力，振动，声发射）与后刀面磨损宽度，每把铣刀使用 300 次。数据集联系助教拷贝，QQ: 935359421。

HMOTP 数据集：包括 3 把铣刀全寿命周期铣削信号（切削力，振动，力矩）与后刀面磨损宽度，每把铣刀使用 100 次。下载链接：

https://drive.google.com/drive/folders/1eBhTSXMd8jVITfQoJifC50fE_NIFqb89?usp=sharing



三轴加工中心与数据采集系统

参考文献：

[1] Pimenov D Y, Bustillo A, Wojciechowski S, et al. Artificial intelligence systems for tool condition monitoring in machining: Analysis and critical review[J]. Journal of Intelligent Manufacturing, 2023, 34(5): 2079-2121.

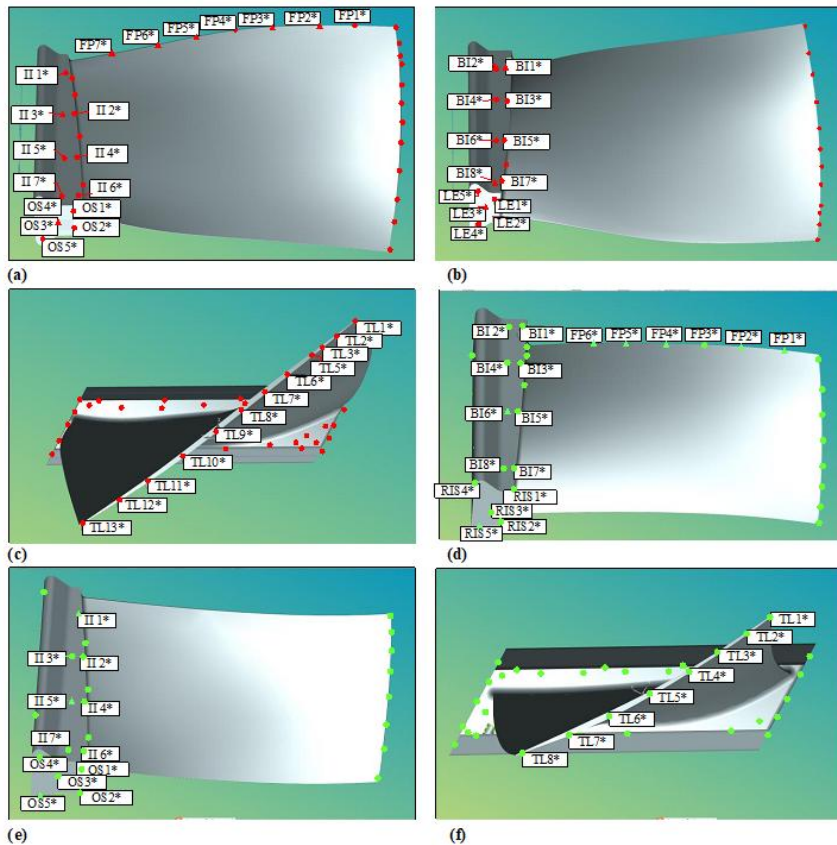
[2] Wang R, Song Q, Peng Y, et al. Toward digital twins for high-performance manufacturing: Tool wear monitoring in high-speed milling of thin-walled parts using domain knowledge[J]. Robotics and Computer-Integrated Manufacturing, 2024, 88: 102723.

[3] Hao C, Mao X, Ma T, et al. A novel deep learning method with partly explainable: Intelligent milling tool wear prediction model based on transformer informed physics[J]. Advanced Engineering Informatics, 2023, 57: 102106.

指导老师：程骋, c_cheng@hust.edu.cn, qq:286652621

4.81 复杂曲面加工质量预测

任务描述：复杂曲面加工过程中存在小样本、工序耦合复杂、误差溯源困难的问题。基于生产需求，加工复杂曲面产线为柔性产线，即在同一条产线上加工多种类型的复杂曲面。为了充分利用样本，要求建立的加工质量预测模型能够在不同域（对应不同尺寸外形的叶片）之间进行域适应学习。主要任务为根据曲面中间工序状态预测最终加工质量，运用数据补全方法，并要求使用域适应学习方法，如 MMD，CORAL 等。



数据说明：包括四种复杂曲面加工的两种中间工序（铣叶根、综合铣）目标点位坐标，以及最终加工质量数据。数据集联系助教拷贝，QQ：935359421。

参考文献：

[1] Li X, Zhang Z, Gao L, et al. A new semi-supervised fault diagnosis method via deep CORAL and transfer component analysis[J]. IEEE Transactions on emerging topics in computational intelligence, 2021, 6(3): 690-699.

[2] Li X, Zhang Z, Gao L, et al. A new semi-supervised fault diagnosis method via deep CORAL and transfer component analysis[J]. IEEE Transactions on emerging topics in computational intelligence, 2021, 6(3): 690-699.

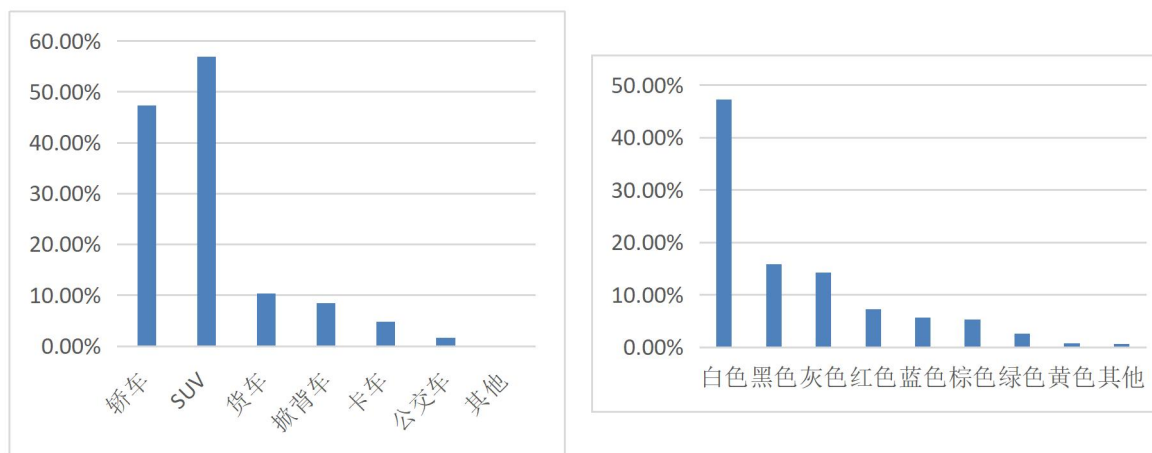
[3] Peng X, Bai Q, Xia X, et al. Moment matching for multi-source domain adaptation[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1406-1415.

指导老师：程骋，c_cheng@hust.edu.cn, qq:286652621

4.82 机载车辆目标重识别

任务描述：车辆重识别是指从多个相机所拍摄的图像中检索目标车辆的技术，在智能交通、车辆追踪、车辆围捕等众多领域具有广阔的应用前景。针对机载车辆重识别任务中车辆尺度不一、角度多变等问题展开研究，设计机载车辆目标重识别算法，测试分析算法的性能。

数据说明：目前专门针对无人机视角的车辆重识别数据集主要有 VeRi-UAV、UAV-VeID 和 VRAI 等，其中 VRAI（Vehicle Re-identification for Aerial Image）数据集是西北工业大学王鹏团队发布的一个大规模、多视角的车辆重识别基准数据集，使用两架 DJI Phantom4 无人机在 15 米至 80 米的高度同时在两个位置拍摄而来。在数据采集过程中，研究团队采用了多种飞行模式，如悬停、巡航和旋转等，以模拟实际应用中无人机的不同飞行状态。这些多样化的飞行模式有效地丰富了图像数据，使得数据集中包含了车辆在不同动态状态下的外观变化，例如车辆的动态模糊、视角变换以及光照条件的变化等。整个数据集包括 137613 张图像，共有 13022 辆不同的车辆。除了车辆 ID 标注外，VRAI 还提供车辆颜色、类型等细粒度标注信息。由于其拍摄过程中无人机与车辆间的相对运动更加复杂，数据集规模也更大，使其对于机载车辆重识别任务也更具有挑战性。



VRAI 数据集中车辆颜色和类型分布



参考资料：

[1] Bingliang Jiao, Lu Yang, Liying Gao, Peng Wang, Shizhou Zhang, Yanning Zhan

- g. "Vehicle Re-Identification in Aerial Imagery: Dataset and Approach." IEEE Transactions on Circuits and Systems for Video Technology, vol. 34, no. 3, pp. 1586-1603, March 2024.
- [2] X. Pan, P. Luo, J. Shi, X. Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. in: 2018 European Conference on Computer Vision . Munich, Germany, 8-14 Sept. Springer, 2018: 464-479.
- [3] Y. Pu, Y. Wang, Z. Xia, Y. Han, Y. Wang, W. Gan, et al. Adaptive rotated convolution for rotated object detection. in: 2023 IEEE/CVF International Conference on Computer Vision . Paris, France, 2-6 Oct. IEEE, 2023: 6589-6600.
- [4] N. Dong, S. Yan, H. Tang, J. Tang, L. Zhang. Multi-view information integration and propagation for occluded person re-identification. Information Fusion. 2024, 104: 102201.

指导老师：黎云，liyun@hust.edu.cn, QQ: 1158125469

4.83 面向无人机巡航的跨视角匹配定位

任务描述：GNSS 拒止环境下，基于图像匹配的视觉定位方法作为一种有效的补充手段，得到了广泛的研究和应用。针对无人机航拍图像与卫星基准图像之间图幅差距大、视角不一致、成像差异显著等问题，研究航拍图像和卫星基准图像跨视角匹配方法（SP+LG），实现倾斜视角下无人机光轴瞄准点精准定位。

数据说明：COCO 和 SUES-200 数据集、仿真生成或实拍无人机航拍和卫星基准图。

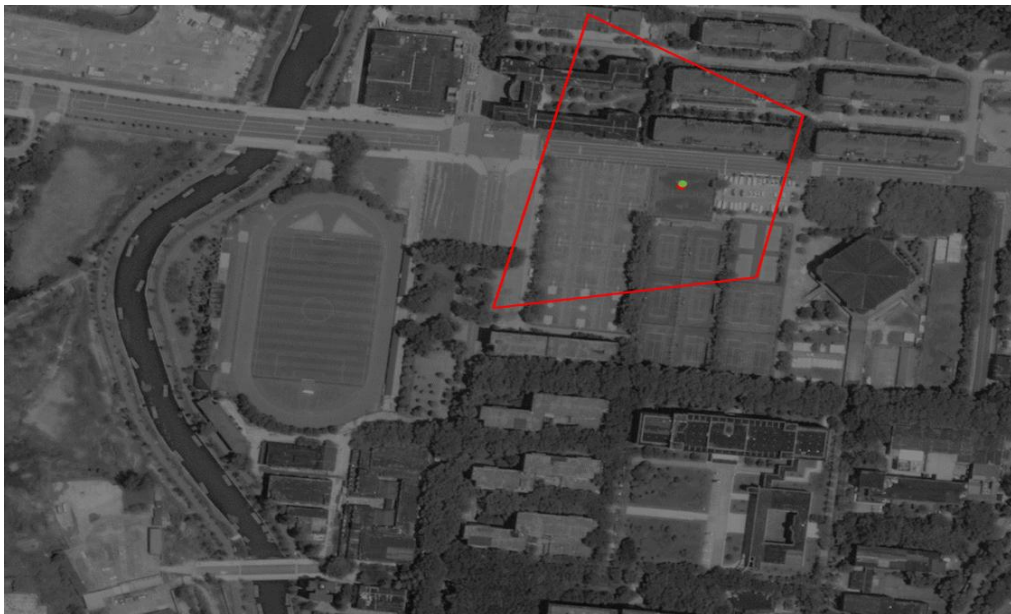
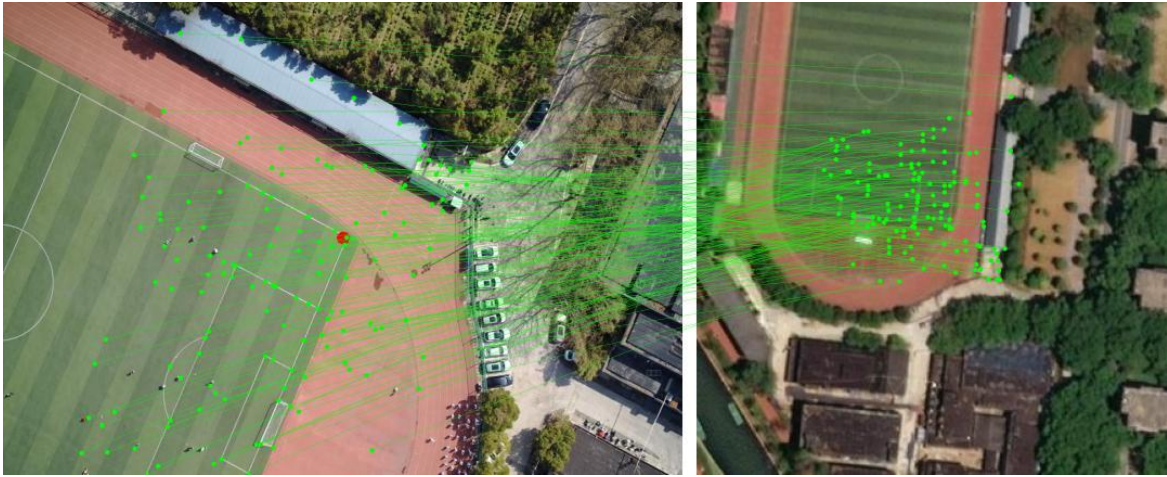
COCO：Microsoft Common Objects in Context，公开数据集可用于特征点检测和描述。

官方网站：<https://cocodataset.org>

下载地址：<http://images.cocodataset.org> (可通过官方工具下载)

SUES-200：A Multi-height Multi-scene Cross-view Image Benchmark Across Drone and Satellite，上海工程技术大学 2023 年发布，首个系统性地考虑无人机飞行高度差异的跨视角数据集。24,120 张无人机图像 + 对应卫星视图，无人机在 4 个不同高度采集（15m、30m、50m、70m），可研究高度变化对匹配的影响，覆盖多种场景。

<https://github.com/Reza-Zhu/SUES-200-Benchmark>



跨视角匹配和光轴定位示例

参考资料:

- [1] D. DeTone, T. Malisiewicz, A. Rabinovich. Superpoint: Self-supervised interest point detection and description. in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. Salt Lake City, USA, 18-22 Jun. 2018, IEEE, 2018: 224-236
- [2] P. Lindenberger, P. E. Sarlin, M. Pollefeys. Lightglue: Local feature matching at light speed. in: Proceedings of the IEEE/CVF International Conference on Computer Vision. Vancouver, Canada, 18-22 Jun. 2023, IEEE, 2023: 17627-17638
- [3] P. E. Sarlin, D. DeTone, T. Malisiewicz, A. Rabinovich. Superglue: Learning feature matching with graph neural networks. in: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition(CVPR). Seattle, USA, 14-19 Jun. 2020, IEEE, 2020: 4938-4947

- [4] T. Y. Lin, M. Maire, S. Belongie. Microsoft coco: Common objects in context. in: Computer Vision–ECCV 2014: 13th European Conference. Zurich, Switzerland, 6-12 Sep. 2014, Springer International Publishing, 2014: 740-755

指导老师：黎云，liyun@hust.edu.cn, QQ: 1158125469

4.84 红外与可见光融合复原方法



任务描述：红外与可见光融合技术通过整合红外图像热辐射信息与可见光图像纹理细节信息，实现不同模态的信息互补，为下游视觉任务提供有力支撑，在目标检测、自动驾驶、智能安防、军事侦察和遥感监测等领域发挥了重要作用，其性能也直接决定了下游任务对于场景理解的可靠性。实际应用中各种退化场景导致融合图像关键信息丢失、引入干扰信息，研究退化条件下红外与可见光融合复原方法，提升融合图像的质量，为下游任务的性能提供有效支撑。

数据说明：(1)MSRS 与 FMB 带有语义分割标签的融合数据集。MSRS(Multi-Spectral Road Scenarios)是由武汉大学基于 MFNet 数据集构建的一个高质量多光谱数据集。包含 1444 对严格配准的红外-可见光图像对，其中包含 715 对昼间场景，729 对夜间场景，分辨率为 480×640 像素。相较于原始 MFNet，MSRS 具有更高的图像质量与标注精度，其丰富的场景多样性及模态互补特性，为融合算法在真实道路场景下的性能验证提供了可靠的基准。面向全天候复杂环境的视觉感知需求，大连理工大学刘日升教授团队构建了全时段多模态基准数据集 FMB。包含 1500 对经过严格时空配准的红外与可见光图像，图像分辨率为 800×600。数据集覆盖多种复杂环境，如浓雾、大雨和低光照条件，适合用于图像融合和语义分割任务。

(2)AWMM-100k、DeMMI-RF，复杂退化条件下红外可见光融合领域数据集。

<https://pan.baidu.com/s/12bYsY3E15aG5skiQ3Tz9zQ?pwd=whv3>

<https://pan.baidu.com/s/1KbaGUXzuOW6ej4maHN5ZcQ?pwd=TGEC#list/path=%2F>

序号	数据集	介绍	可见光图像	红外图像
1	AWMM-100k	包含 187,699 张图像，涵盖雨、雾和雪等天气退化状况，支持在恶劣天气条件下进行图像融合研究。		

2	DeMMI-RF	包含超过3万对红外可见光图像, 涵盖噪声、模糊、雾霾与条纹噪声等多种复杂退化类型, 覆盖地面与无人机视角。		
---	----------	---	--	---

参考资料:

- [1] Li X, Liu W, Li X, et al. All-weather multi-modality image fusion: Unified framework and 100k benchmark[J]. Information Fusion, 2026: 104130.
- [2] Sun Y, Li X, Zhu P, et al. Task-gated multi-expert collaboration network for degraded multi-modal image fusion[C] International Conference on Machine Learning. PMLR, 2025: 57571-57586.
- [3] Yi X, Xu H, Zhang H, et al. Text-if: Leveraging semantic text guidance for degradation-aware and interactive image fusion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 27026-27035

指导老师: 黎云, liyun@hust.edu.cn, QQ: 1158125469

4.85 典型人脸识别网络物理可实现攻击

任务描述: 物理可实现的人脸识别攻击, 指的是能在真实物理世界中通过实物 (而非单纯修改数字图像文件) 来欺骗人脸识别系统的攻击方式。这类攻击的核心挑战在于, 它们必须克服从数字仿真到物理打印/部署之间的"模拟-现实"鸿沟, 并能在不同角度、光照等复杂现实条件下持续生效。选择一到两种典型的人脸检测识别网络进行物理贴片攻击方法实现, 测试贴片的实际攻击效果。

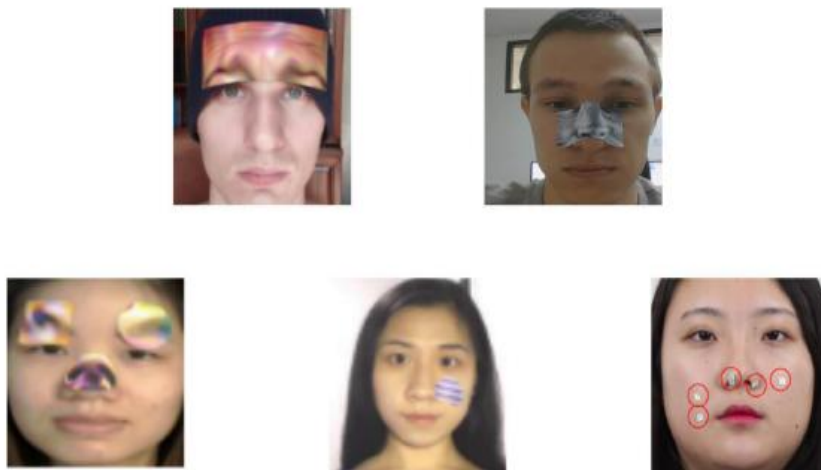
注: GaP (Gaussian Patch) 是 2025 年提出的一种针对人脸识别系统的黑盒物理对抗攻击方法。简单来说, 它是一种通过贴在额头上的特定灰度图案, 就能让先进的人脸识别系统认不出你的技术。

数据说明: (1) LFW (Labeled Faces in the Wild) 数据集是一个专门为研究非受限场景下的人脸识别问题而设计的公开图像集, 由美国马萨诸塞大学阿默斯特分校计算机视觉实验室于 2007 年发布, 用于测试补丁在真实世界人脸识别系统 (如 ArcFace) 上的黑盒攻击成功率及物理可迁移性 (欺骗 FaceNet)。

(2) OULU-NPU 是一个在人脸防伪 (Face Anti-Spoofing) 和呈现攻击检测 (Presentation

Attack Detection, PAD) 领域最具权威性和影响力的基准数据集之一, 这类包含真实物理传感器数据的数据集, 可以验证打印出的补丁在真实摄像头下是否依然有效。

注: 可以自拍人脸攻击测试图像



物理贴片攻击示例

参考资料:

- [1] Wang M S, Zhou J C, Li T L, et al. A survey on physical adversarial attacks against face recognition systems[J]. arXiv preprint, 2024, arXiv:2410.16317.
- [2] Vakhshiteh F, Nickabadi A. Threat of adversarial attacks on face recognition: A comprehensive survey[J]. arXiv preprint, 2020, arXiv:2007.11709
- [3] Cui C X. A review of deep face adversarial sample generation techniques[C] Proceedings of the International Conference on Computer Science and Electronic Information Technology (CSEIT 2025). ITM Web of Conferences, 2025, 78: 02013
- [4] 蔡楚鑫, 王宇飞, 章烈剡, 等. 物理域中针对人脸识别系统的对抗样本攻击方法[J]. 信息安全学报, 2023, 8(2): 127-137.
- [5] A. Kaznacheev, M. Mikhalchuk, A. Kuznetsov, A. Petiushko, and A. Razzhigaev, "Real-World Transferable Adversarial Attack on Face-Recognition Systems," arXiv, Sep. 2025. doi: 10.48550/arXiv.2509.23198.

指导老师: 黎云, liyun@hust.edu.cn, QQ: 1158125469

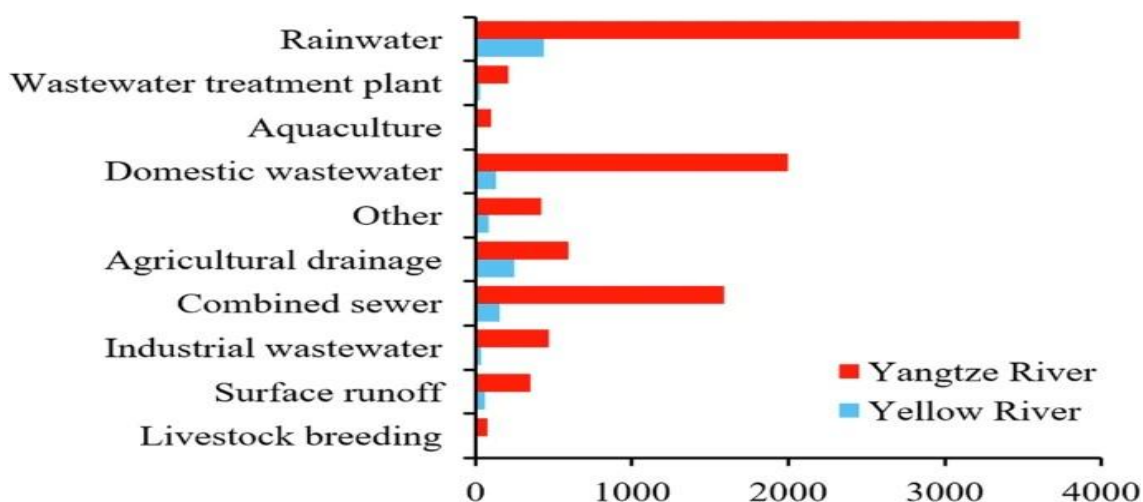
4.86 排污口自动图像检测

任务描述: 利用无人机航拍巡检图像（可见光、红外），对排污口进行自动检测，实现水环境污染源的自动识别与定位。

数据说明: iSOOD (images for Sewage Outfalls Objective Detection 排污口目标检测图像数据集, 清华大学环境学院) 10,481 张, 中国长江流域 (9,285 张) + 黄河流域 (1,196 张), Zenodo 平台发布。

<https://doi.org/10.1038/s41597-024-03574-9>

<https://github.com/Daniel00ll/iSOOD-code>



a. Near the center

b. Close to the edge

排污口自动图像检测示例

参考资料:

[1] Tian Y Q, Deng N, Xu J, Wen Z G. A fine-grained dataset for sewage outfalls objective

detection in natural environments[J]. *Scientific Data*, 2024, 11: 724

[2] Song X Y, Zuo G G, Wang X F, Xie J C. Automatic recognition of environmental hazards in river and lake ecosystems using deep learning[J]. *Frontiers in Environmental Science*, 2025, 13: 1657930.

[3] 冯磊, 崔胜涛. 无人机遥感技术在海域监测陆源排污口中的应用[J]. *测绘与空间地理信息*, 2019, 42(5): 107-109.

[4] Chen, J., Zhang, Y., Wang, Y., & Wang, L. (2022). *An Improved YOLOv5-Based Method for Real-Time Detection of Pollutants in Water Surfaces*. *Sensors*, 22(12), 3456.

指导老师: 黎云, liyun@hust.edu.cn, QQ: 1158125469

4.87 血管造影图像自动诊断检测

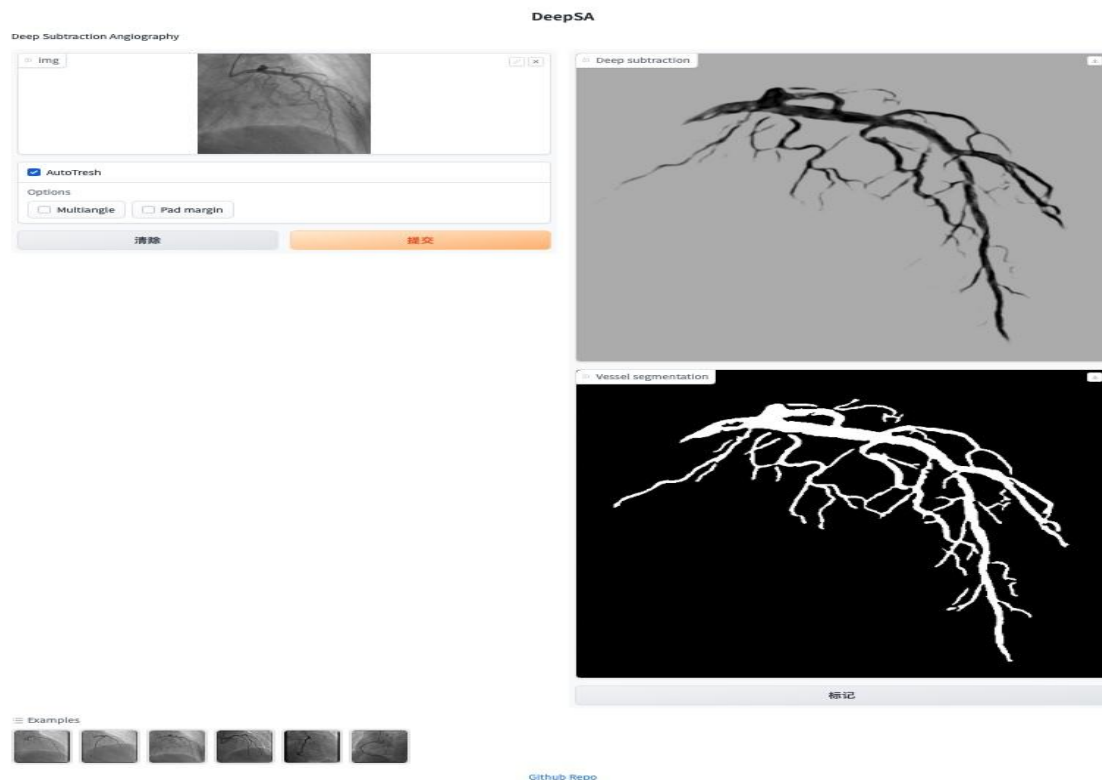
任务描述: 选择一种血管造影成像方式和监测部位, 如冠状动脉血管 DSA (数字减影血管造影) 图像, 进行血管健康状态自动诊断检测, 可涵盖血管分割、狭窄检测、SYNTAX 评分评估等多项任务。

数据说明: (1) ARCADE 数据集 (最全面, 2024 年发布) 数据规模与构成: 总计 3,000 张 X 射线冠状动脉造影图像。冠状动脉血管分类 - 1,500 张, 标注冠状动脉节段; 狭窄检测 - 1,500 张, 标注斑块位置。一个公开但受控访问的数据集, 需要申请。

(2) XCAD 数据集和 FS-CAD 数据集 (X-ray Angiography Coronary Artery Disease dataset X 射线冠状动脉造影数据集), 公共测试集。包含 126 张手动标注的冠脉造影图像, 主要用于分割任务评估 DSA 冠脉分割数据集, 需联系作者。

(3) ImageCAS : 大规模冠脉分割基准数据集
<https://github.com/XiaoweiXu/ImageCAS-A-Large-Scale-Dataset-and-Benchmark-for-Coronary-Artery-Segmentation-based-on-CT>

(4) LM-CAD 预训练数据集大规模预训练数据集 (17,398 张无造影剂帧+38,930 张有造影剂帧), 来自《Scientific Reports》2024 论文, 配套自监督预训练模型 DeepSA (2024 年)
<https://github.com/newfyu/DeepSA>



血管分割提取示例

参考资料:

- [1] Zhang Y X, Zhang X L, He Y L, et al. Coronary p-Graph: Automatic classification and localization of coronary artery stenosis from Cardiac CTA using DSA-based annotations[J]. Computerized Medical Imaging and Graphics, 2025: 102537
- [2] Li Y, et al. Automatic coronary artery segmentation and diagnosis of stenosis by deep learning based on computed tomographic coronary angiography[J]. European Radiology, 2022, 32: 6037-6049.
- [3] Labrecque Langlais et al. (2024). Dataset for Automatic Region-based Coronary Artery Disease Diagnostics Using X-Ray Angiography Images. Scientific Data, 11, 20.
- [4] Song Z, et al. Automatic coronary artery segmentation of CCTA images with an efficient feature-fusion-and-rectification 3D-UNet[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(8): 4044-4055

指导老师：黎云，liyun@hust.edu.cn, QQ: 1158125469

4.88 堰塞坝表层颗粒物智能检测分析

任务描述: 利用无人机影像，实现对堰塞坝表层颗粒物的智能检测和粒径分析，为坝

体稳定性快速评估提供依据。

数据说明:

(1) 台湾滑坡堆积体粒径分布数据集

<https://doi.org/10.4211/hs.ade683be61e54fa5b60da97418a5f3df>

(2) VIPA 岩土颗粒破碎仿真数据集 ,

<https://github.com/doujiang-zheng/GNN-For-Particle-Crushing>

(3) 无人机视角滑坡泥石流分割/检测数据集, firc-dataset



堰塞坝表层颗粒物质的智能检测

参考资料:

[1] 基于深度学习的堰塞坝表层颗粒物质识别研究[J/OL]. CSDN 博客, 2025.

<https://huanghm.blog.csdn.net/article/details/155990321>

[2] Mair D, Witz G, Do Prado A H, et al. Automated detecting, segmenting and measuring of grains in images of fluvial sediments: the potential for large and precise data from specialist deep learning models and transfer learning[J]. *Earth Surface Processes and Landforms*, 2024, 49(3): 1099-1116.

[3] Automated particle size characterization and gradation evaluation of rock fragments for dam construction using close-range photogrammetry[J]. *Journal of Rock Mechanics and Geotechnical Engineering*, 2025.

[4] 一种堰塞体表面粒组分区智能识别方法[P]. 中国专利, 2023.

指导老师: 黎云, liyun@hust.edu.cn, QQ: 1158125469

4.89 水果新鲜度监测

任务描述: 日常水果的新鲜度自动监测, 涉及新鲜、腐烂分类, 成熟度评估, 化学污染检测(如福尔马林)等任务。

数据说明: (1) FruitVision 数据集, 包括 5 种水果(苹果、香蕉、芒果、橙子、葡萄)和三种新鲜度类别(新鲜、腐烂、福尔马林混合)

<https://data.mendeley.com/datasets/xkbjx8959c/2>

(2) FruitQ 多水果新鲜度数据集, 包括 11 种水果和 3 种新鲜度类别

<https://www.kaggle.com/datasets/sholzz/fruitq-dataset>

FruitVision 数据集分布

水果	新鲜	腐烂	福尔马林混合
苹果	765 → 6120	630 → 5040	643 → 5144
香蕉	749 → 5992	632 → 5056	660 → 5280
葡萄	770 → 6160	630 → 5040	610 → 4880
芒果	763 → 6104	630 → 5040	616 → 4928
橙子	753 → 6024	656 → 5248	647 → 5176

参考资料:

[1] Bijoy M H I, Hasan M M, Jahan N, et al. FruitVision: A benchmark dataset for fresh, rotten, and formalin-mixed fruit detection[J]. Data in Brief, 2025, 61: 111752.

[2] Fruit and vegetable quality detection using SENet and EfficientNet[C]//Proceedings of IEEE Conference. Chennai, India, 2025.

[3] 李涛, 王哲, 陈明, 等. (2023). 基于多模态感知的水果新鲜度智能评估方法. *计算机学报*, 46(8), 1453–1467.

[4] Li, T., Chen, Z., Wang, Y., & Liu, J. (2023). *FruitFreshNet: A Multimodal Framework for Non-Invasive Fruit Freshness Assessment Using UAV and Ground Sensors*. Computers and Electronics in Agriculture, 2023, 112345.

指导老师: 黎云, liyun@hust.edu.cn, QQ: 1158125469

4.90 自主竞赛选题

任务描述: 从以下网址中自主选择公开竞赛项目作为本课程设计题目, 以竞赛排名作为

验收的测试依据。（本题选题小组数量不限）

<https://www.kaggle.com/competitionshttps://aistudio.baidu.com/aistudio/competition>

指导老师：可选择胡静、杨卫东、肖阳、邹腊梅、陆昊、胡若澜，邹旭、昌毅，谭毅华，程骋，黎云老师中的任意一位作为指导老师。（指导老师指导小组数量设置上限，总指导小组不超过 9 组）

五、课程设计的考查

5.1 考核方式

由平时检查和结题答辩时的提问抽查、现场演示、课题难易程度和工作量饱满程度，以及设计报告撰写情况等几方面综合起来考虑。具体包括学习与设计态度的认真性，课堂知识理解掌握的深入程度，常用工具软件应用的熟练程度，设计方案的正确性或合理性，图文的质量效果，是否独立完成，是否具有独立分析解决问题的能力 and 创新精神等。

5.2 成绩评定

课程设计成绩的评价依据 = 工作量（15%）+个人工作表现（10%）+创新性（20%）+实验与分析（35%）+报告写作（10%）+报告答辩及成果展示（10%）。

评分内容	综合表现		设计报告			答辩
	工作量	个人工作表现	创新性	实验与分析	报告写作	报告答辩及成果展示
评价比重	0.15	0.10	0.20	0.35	0.10	0.10

***切记要独立完成，不得抄袭他人成果。一旦发现抄袭者，课程成绩一律按不及格处理。**

***由于答辩时间有限，报告答辩时的成果展示建议采用视频展示的方式进行，包括具体的算法运行、性能指标计算过程、报告的 AI 撰写占比及评价等自证内容，不超过 1 分 30 秒。**

课程设计成绩：按答辩时的评审老师评分平均值计算，采用百分制评定。95-100 为优秀，90-94 为良好，70-89 为中等，60-69 为及格，低于 60 分的为不及格。

优秀：能独立完成设计要求所规定的全部内容，设计方案正确、基本概念清楚，有独到的见解或创造性，与当前主流研究成果相比较效果更优异。每个组员的工作量饱满，贡献突出。

良好：能较好完成设计要求所规定的全部内容，设计方案正确，分析问题正确、基本概念清楚，运用多种当前主流技术和算法进行了任务实验、性能对比和深入分析。

中等：能完成设计要求规定的全部内容，设计方案基本正确，基本概念清楚。

及格：基本完成设计要求规定的内容，设计方案基本合理，基本概念较清楚。

不及格：未完成设计要求规定的内容，设计方案不合理，或有较严重缺陷，基本概念不清楚。

六、课程设计报告撰写要求

课程设计说明书是课程设计工作的总结，它应该反映出学生在课程设计过程中所做的主要工作和取得的主要成果。学生必须以积极认真、严谨求实的态度完成课程设计说明书的撰写。

课程设计说明书写作的具体内容包括：

1、封面

包括题目，指导老师，组员信息（姓名，学号，班级，每个人的工作占比）

2、摘要：要求写出同一内容的中文和英文的摘要。摘要应说明本设计的中心思想和主要内容，突出设计中的新见解新方法，说明该设计方案的理论根据及现实意义。摘要力求简明扼要，字数为 300 字左右。

3、目录：目录是整个设计的提纲，也是设计的重要组成部分，它方便评阅教师了解设计的整体结构。目录以章、节两级目录为宜。

4、正文：课程设计说明书的正文一般可按章、节的格式来书写，正文常常包括如下几个部分：

1) 第一章 课题概述

简要介绍所选课题现行研究现状、存在的主要问题，说明选题的意义及必要性。

2) 第二章 算法分析

2.1 需求分析

2.2 研究方案设计

2.3 试验算法选择与分析（或算法设计）

- 2.4 特征提取算法研究（可选）
- 2.5 特征分析算法研究（可选）
- 2.6 模式分类/识别算法研究（可选）
- 3) 第三章 试验系统设计(包含程序流程图、程序功能介绍)
 - 3.1 系统总体结构设计
 - 3.2 代码设计
 - 3.3 输入/输出设计
 - 3.4 模块功能与处理过程设计
- 4) 第四章 软件实施与实验运行
 - 4.1 软件系统实施（编程、调试、运行）
 - 4.2 数据库测试
 - 4.3 试验结果与分析
- 5) 第五章 结束语
 - 研究结论, 通过课程设计对模式识别学科的认识与体会, 每人的工作划分。
- 6) 附录 主要模块代码（注释代码比例不得低于 1：1）
- 7) 参考文献（不少于 6 篇）

以上内容供学生编写设计说明书时参考，学生可根据实际系统开发情况及指导教师的具体要求进行内容的增删或章节的调整。