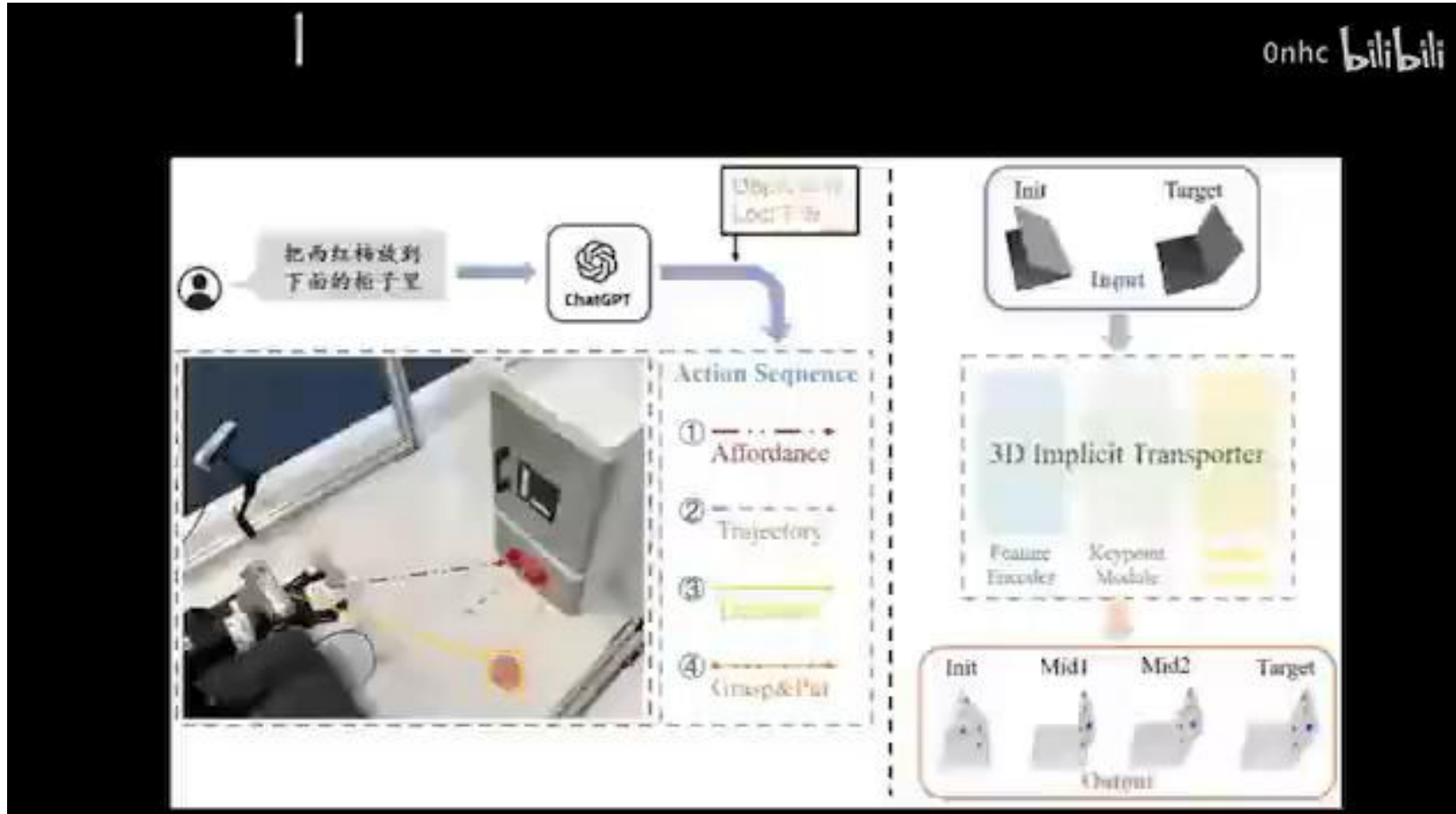


# A Brief Introduction for the Course Project

2024/09/09

# Interactive Robotic Bin Picking

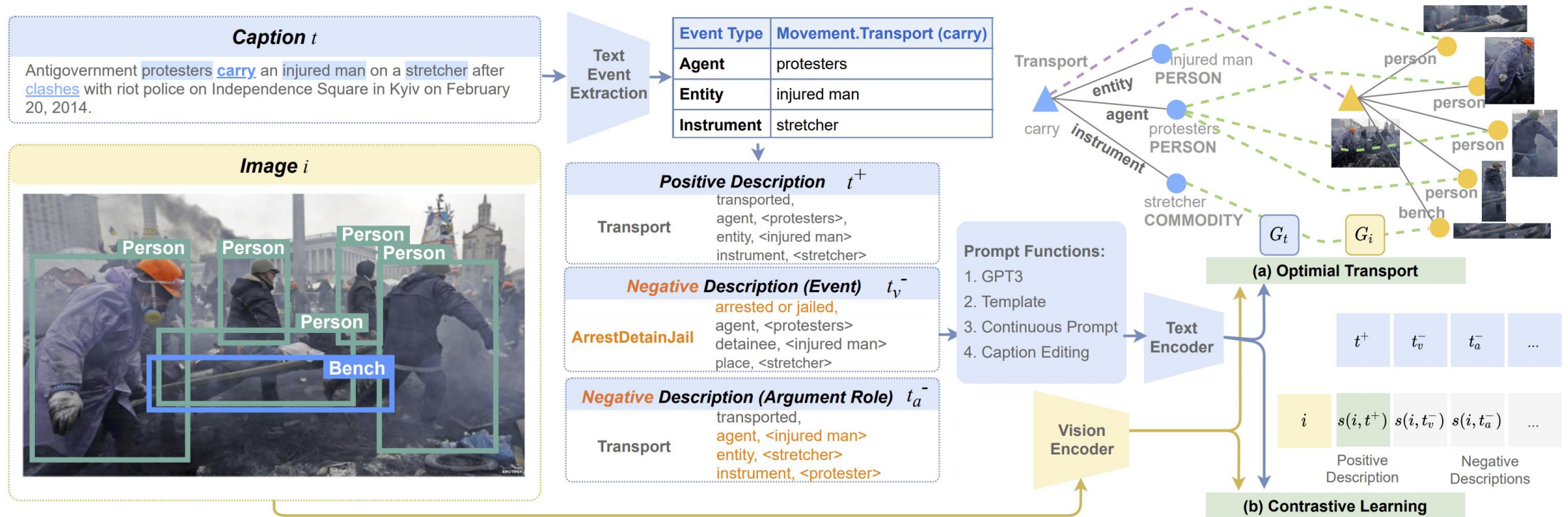
## --- Cognition & Execution for Trans-Modality Media



# Interactive Robotic Bin Picking

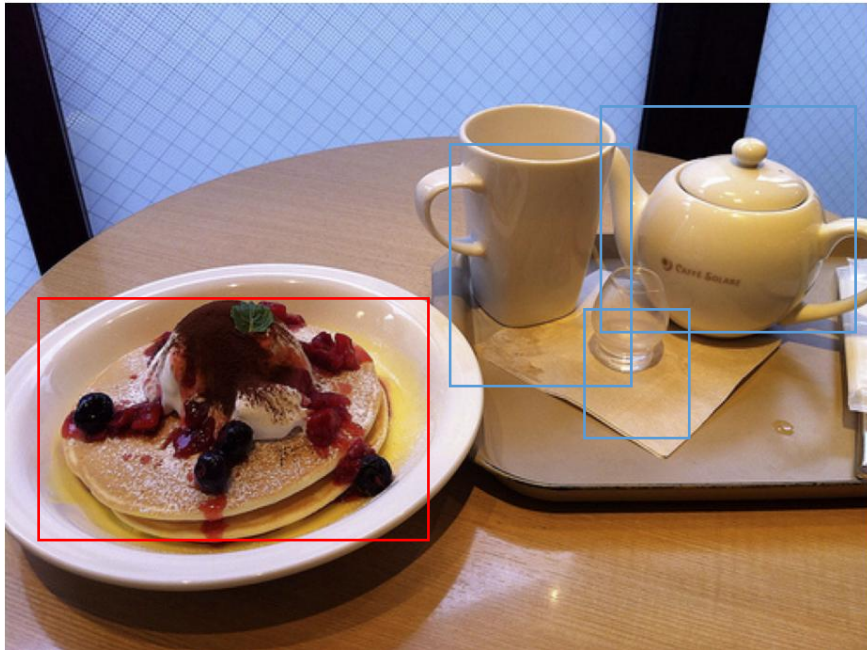
- Stage 1 – Trans-Modality Media

--- Natural Language meets Visual Perception



# Objective for Stage 1

- To find out the exact object according to a prompt.
  - E.g., Prompts 1:
    - Pass me the **pancake!**
  - E.g., Prompts 2:
    - Show me the **blue pack!**

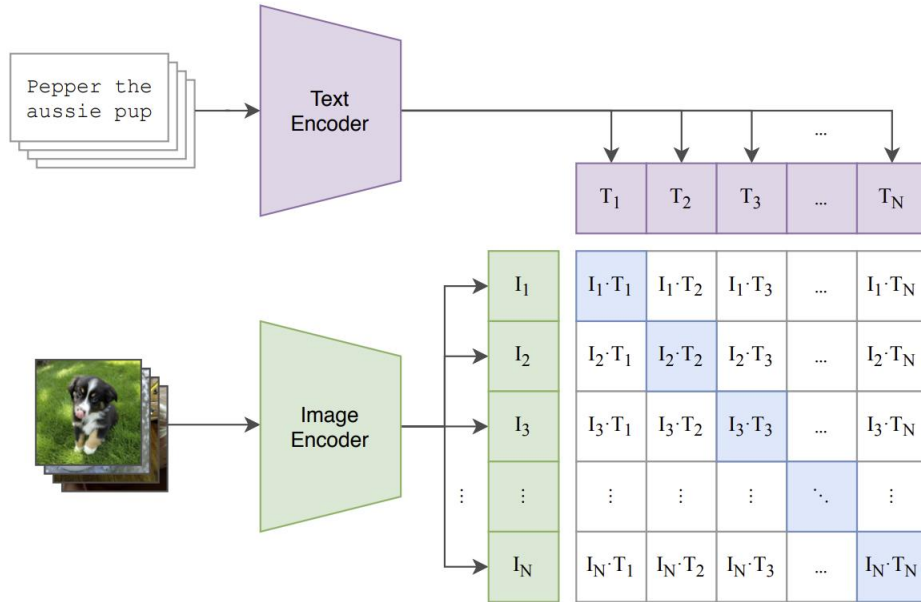


Drew. A. et. al, "GQA: A New Dataset for Real-World Visual Reasoning and Compositional Question Answering". <https://arxiv.org/pdf/1902.09506.pdf>

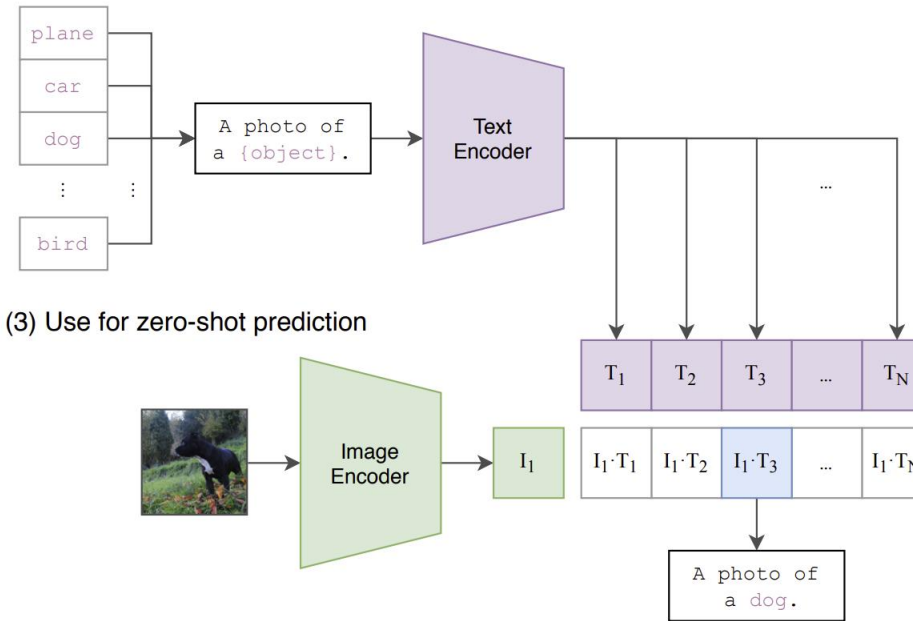
← This could also be a dataset candidate for your own algorithm training.

# Potential Road-map for Stage 1

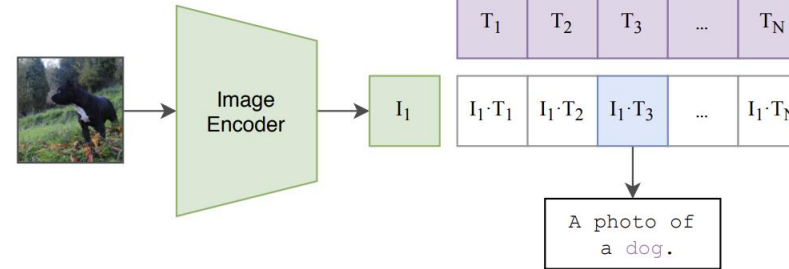
(1) Contrastive pre-training



(2) Create dataset classifier from label text



(3) Use for zero-shot prediction



- Jointly train an **image encoder** and a **text encoder** to predict the correct pairings of a batch of (image, text) training examples.

Radford, Alec, et al. "Learning transferable visual models from natural language supervision." International conference on machine learning. PMLR, 2021.



# Evaluation Datasets & Deliverables

- Input: 'all the **containers**' + image



- Input: 'all **green** blocks' + image

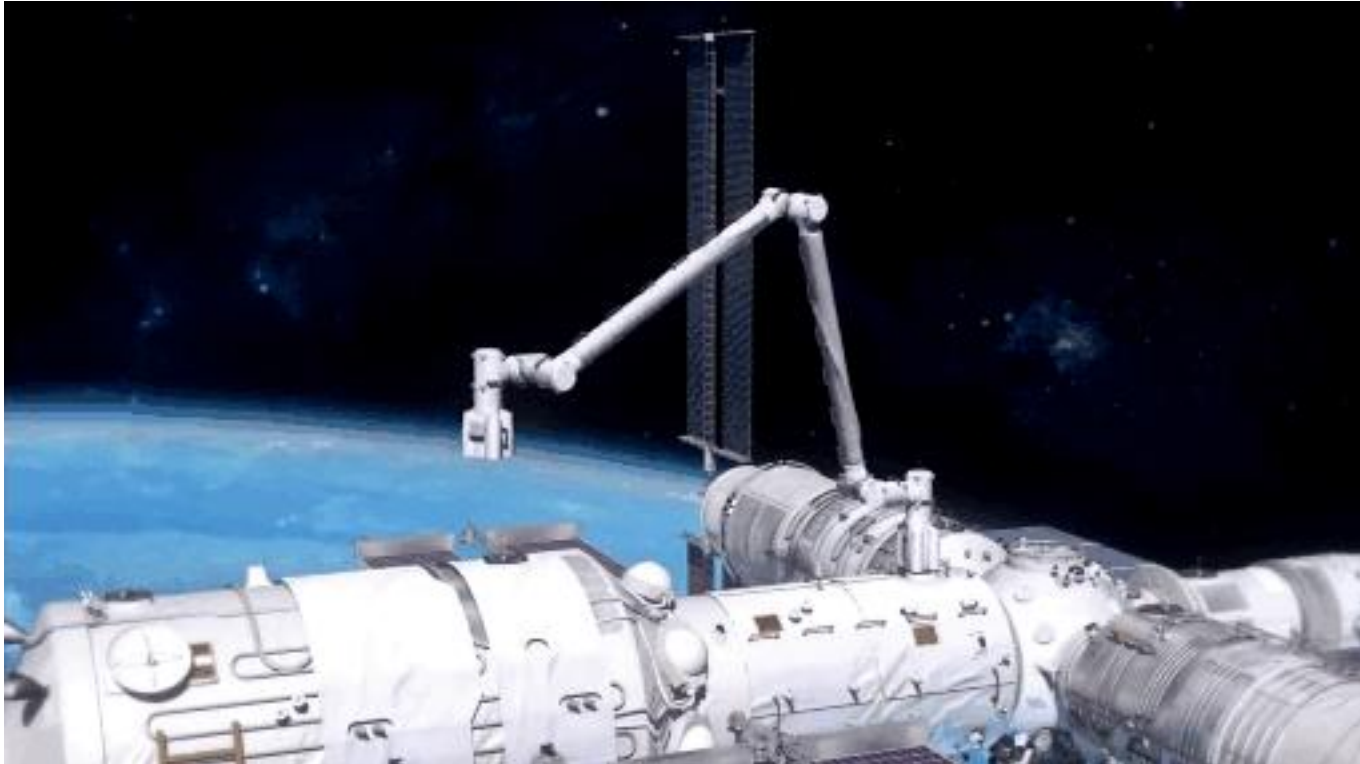
- Output: bbox for desired objects



- ~20 examples will be available for algorithm testing
- Models will be evaluated with similar cases (not public accessible)

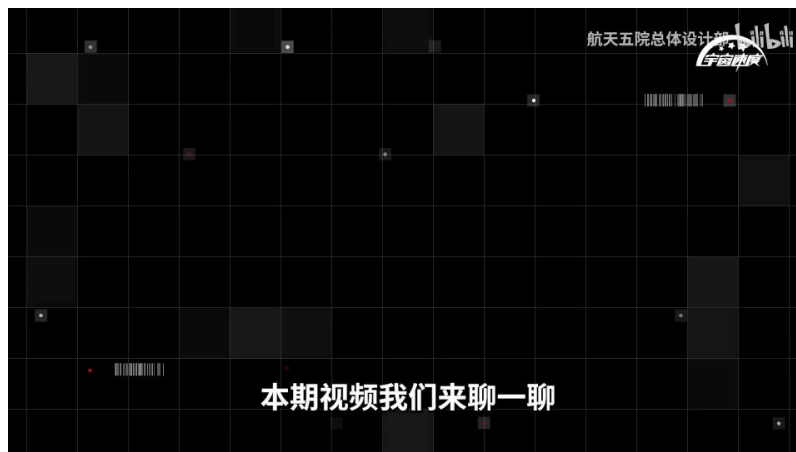
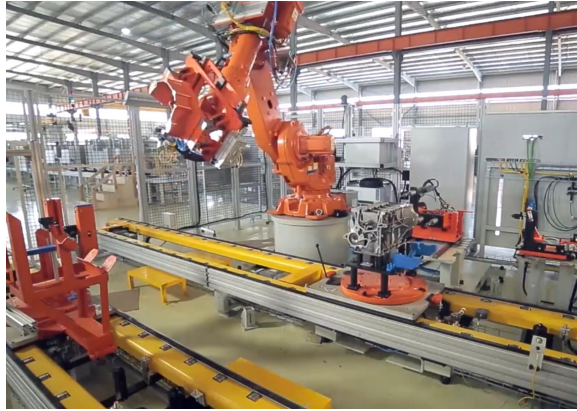
# Interactive Robotic Bin Picking

- **Stage 2 – Cognition-Based Execution**
  - Maneuvering Intelligent Robots upon Questioning



# Variant kinds of Robotic Arms

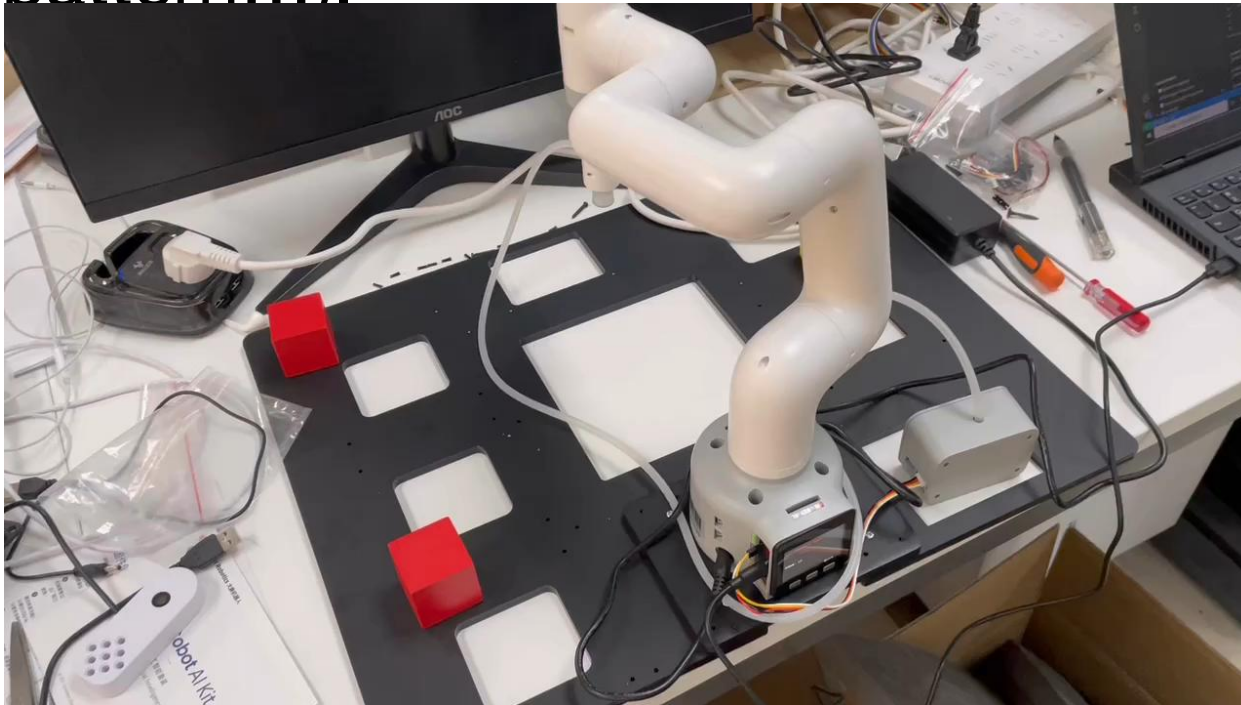
- extreme environment temperature, communication efficiency ...
- High precision requirements, speed ...





# Objective for Stage 2

- → Deploy a RGB camera for scene acquisition;
- → Form a complete robotic arm system;
- → Objects sorting upon request (open vocabulary, color, pattern...).



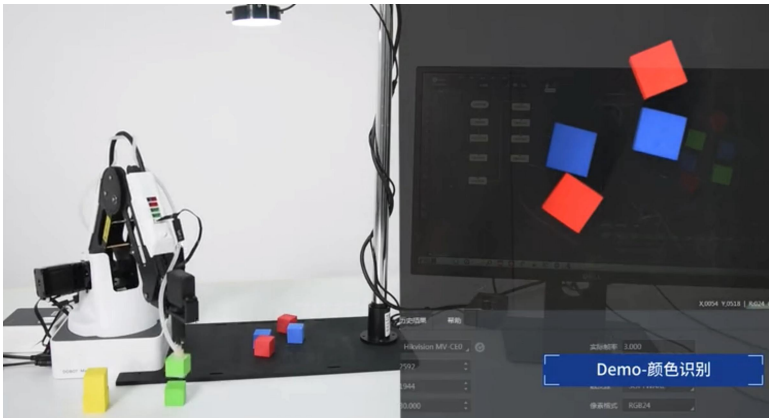
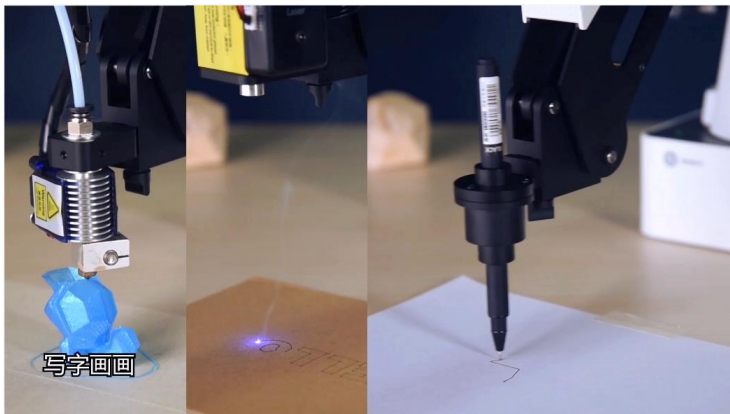
- E.g., Input prompts:  
**Put the lower red block to the center!**

# Detail of the Robotic Arm (which will be used)



Desktop robotic arm

- 4 degrees of freedom
- | Axis                      | Range        | Max Speed<br>(250g workload) |
|---------------------------|--------------|------------------------------|
| Joint 1 base              | -120° ~+120° | 320° / s                     |
| Joint 2 rear arm          | -5° ~+90°    | 320° / s                     |
| Joint 3 forearm           | -15° ~+90°   | 320° / s                     |
| Joint 4 rotation<br>servo | 140° ~+140   | 480° / s                     |
- Development Platform: Python、C++



# Operations Required for Stage 2

**Localize objects using RGB images & prompts in the real world.**

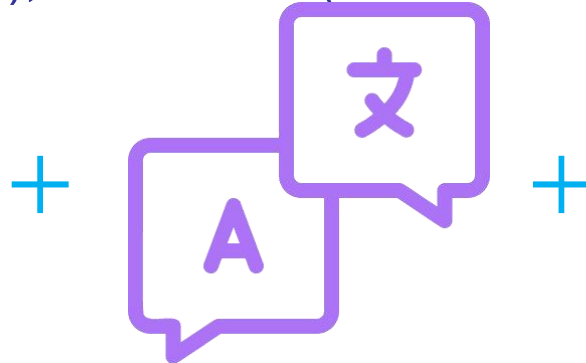
1. RGB images: perception of the real scenario.
2. Prompts: locate the exact position of the target object.

**Control the robotic arm to sort the target objects.**

1. Control the robotic arm to reach the correct position.
2. Catch the object to the specified target location.

## \* Evaluation metrics

Accuracy (prompts understanding),  
efficiency (speed), robustness(extreme cases), etc.



# Grading Criteria \*

- **Stage 1: Trans-Modality Media:** 50% in total
  1. Accuracy:
    1. Object detection (bounding boxes): 20%
    2. Prompts translation & understanding: 20%
  2. Coding accessibility: 5%
  3. Bonus: 5% (algorithm novelty, efficiency, ...)
- **Stage 2: Cognition-Based Execution:** 50% in total
  1. System design: 5%
  2. Accuracy and efficiency:
    1. Recognition: 10%
    2. Localization: 10%
    3. Operation: 10%
  3. Coding accessibility: 5%
  4. Bonus: 10% (extreme cases, occlusion, interruption, ...)

*Work in **teams of 3 to 4 students**.  
End-of-term presentations and  
assessments.*

\* Note: This criteria may vary according to your practice.



# Contact Us

- FIT building 4-405
- 19:00 – 22:00 every Tuesday, Thursday, Saturday
- Please fill the table for team registrations & appointment
- Contact T.A. through WeChat for more details



群聊: Media & Cognition  
2024-2025

	A	B	C	D	E	F	G	H	I
1	分组名单:								
2	注意事项: 分组最少2人, 最多4人, 请填写协商好的每位成员的姓名与学号。								
3									
4									
5	组序号	成员1-姓名	成员1-学号	成员2-姓名	成员2-学号	成员3-姓名	成员3-学号	成员4-姓名(可选)	成员4-学号(可选)
6	0(示例)	管海洋	202xxxxxx	王光宇	20xxxxxxx	徐智昊	20xxxxxxx		
7	1								
8	2								
9	3								
10	4								
11	5								
12	6								
13	7								
14	8								
15	9								
16	11								
17	12								
18									
19									
20									
21	机械臂操作时间预约:								
22	注意事项: 为保证合理的人员密度, 我们建议每时段不超过4组同时进行								
23	每天分为早中晚三个时段, 请在对应日期与时段的空格中填入自己的组序号								
24	来之前一天请提前与助教确认时间, 以便我们安排机械臂, 多谢								
25									
26	日期	晚间19pm-22pm	预约组别1	预约组别2	预约组别3	预约组别4			
27	10月17日	周二							
28	10月19日	周四							
29	10月21日	周六							
30	10月24日	周二							
31	10月26日	周四							
32	10月28日	周六							
33	10月31日	周二							



***Good luck on the grand  
tour!***

• Q & A

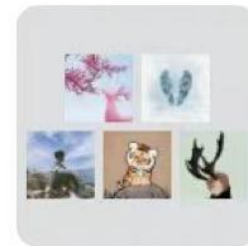
2024-2025媒体与认知分组



使用 金山文档 查看文档



金山文档 | WPS云文档



群聊: Media & Cognition  
2024-2025

