

# Lesson 6 作业

## Lagent & AgentLego 智能体应用搭建

### 基础作业

1. 完成 Lagent Web Demo 使用，并在作业中上传截图。文档可见 Lagent Web Demo

Installed kernelspec agent in /root/.local/share/jupyter/kernels/agent  
conda环境: agent安装成功!

=====

ALL DONE!

=====

```
• (base) root@intern-studio-40069509:~# conda activate agent
• (agent) root@intern-studio-40069509:~# lmdeploy serve api_server /root/share/new_models
/Shanghai_AI_Laboratory/internlm2-chat-7b \
> --server-name 127.0.0.1 \
> --model-name internlm2-chat-7b \
> --cache-max-entry-count 0.1

[WARNING] gemm_config.in is not found; using default GEMM algo
HINT: Please open http://127.0.0.1:23333 in a browser for detailed api usage!!!
HINT: Please open http://127.0.0.1:23333 in a browser for detailed api usage!!!
HINT: Please open http://127.0.0.1:23333 in a browser for detailed api usage!!!
INFO: Started server process [23667]
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: Uvicorn running on http://127.0.0.1:23333 (Press CTRL+C to quit)
```

```
• (base) root@intern-studio-40069509:~# conda activate agent
• (agent) root@intern-studio-40069509:~# cd /root/agent/lagent/examples
nml2_agent_web_demo.py --server.address 127.0.0.1 (agent) root@intern-stu
• n internlm2_agent_web_demo.py --server.address 127.0.0.1 --server.port 78
60
```

Collecting usage statistics. To deactivate, set browser.gatherUsageStats to false.

You can now view your Streamlit app in your browser.

URL: http://127.0.0.1:7860

模型名称:

internlm2-chat-7b

系统提示词

当开启工具以及代码时，根据需求选择合适的工具进行调用

数据分析提示词

你目前已经能够在一个有状态的Jupyter 笔记本环境中运行 Python 代码。当你向 python 发送含有 Python 代码的消息时，它将在该环

插件提示词

你可以使用如下工具：  
{prompt}  
如果你已经获得足够信息，请直接给出答案。避免不必要的工具调用!

模型IP:

127.0.0.1:23333

插件选择

ArxivSearch

127.0.0.1:7860

Langent Web Demo

请帮我搜索 InternLM2 Technical Report

```
{  "name": "ArxivSearch",  "parameters": {    "query": "InternLM2 Technical Report"  }}
```

```
{"content": "Published: 2024-03-26\nTitle: InternLM2 Technical Report\nA
```

以下是关于"InternLM2 Technical Report"的搜索结果：  
标题: InternLM2 Technical Report  
作者: Zheng Cai, Maosong Cao, Haojiong Chen, Kai Chen, Keyu Chen, Xin Chen, Xun Chen, Zehui Chen, Zhi Chen, Pei Chu, Xiaoyi Dong, Haodong Duan, Qi Fan, Zhaoye Fei, Yang Gao, Jiaye Ge, Chenya Gu, Yuzhe Gu, Tao Gui, Aijia Guo, Qipeng Guo, Conghui He, Yingfan Hu, Ting Huang, Tao Jiang, Penglong Jiao, Zhenjiang Jin, Zhikai Lei, Jiaxing Li, Jingwen Li, Linyang Li, Shuaibin Li, Wei Li, Yining Li, Hongwei Liu, Jiangning Liu, Jiawei Hong, Kaiwen Liu, Kuikun Liu, Xiaoran Liu, Chengqi Lv, Haijun Lv, Kai Lv, Li Ma, Runyuan Ma, Zerun Ma, Wenchang Ning, Linke Ouyang, Jiantao Qiu, Yuan Qu, Fukai Shang, Yunfan Shao, Demin Song, Zifan Song, Zhihao Sui, Peng Sun, Yu Sun, Huanze Tang, Bin Wang, Guoteng Wang, Jiaqi

模型名称:

internlm2-chat-7b

系统提示词

当开启工具以及代码时，根据需求选择合适的工具进行调用

数据分析提示词

你目前已经能够在一个有状态的Jupyter 笔记本环境中运行 Python 代码。当你向 python 发送含有 Python 代码的消息时，它将在该环

插件提示词

你可以使用如下工具：  
{prompt}  
如果你已经获得足够信息，请直接给出答案。避免不必要的工具调用!

模型IP:

127.0.0.1:23333

插件选择

ArxivSearch

☐ 数据分析

清空对话

Deploy

Yingfan Hu, Ting Huang, Tao Jiang, Penglong Jiao, Zhenjiang Jin, Zhikai Lei, Jiaxing Li, Jingwen Li, Linyang Li, Shuaibin Li, Wei Li, Yining Li, Hongwei Liu, Jiangning Liu, Jiawei Hong, Kaiwen Liu, Kuikun Liu, Xiaoran Liu, Chengqi Lv, Haijun Lv, Kai Lv, Li Ma, Runyuan Ma, Zerun Ma, Wenchang Ning, Linke Ouyang, Jiantao Qiu, Yuan Qu, Fukai Shang, Yunfan Shao, Demin Song, Zifan Song, Zhihao Sui, Peng Sun, Yu Sun, Huanze Tang, Bin Wang, Guoteng Wang, Jiaqi Wang, Jiayu Wang, Rui Wang, Yudong Wang, Ziyi Wang, Xingjian Wei, Qizhen Weng, Fan Wu, Yingtong Xiong, Chao Xu, Ruiliang Xu, Hang Yan, Yirong Yan, Xiaogui Yang, Haochen Ye, Huaiyuan Ying, Jia Yu, Jing Yu, Yuhang Zang, Chuyu Zhang, Li Zhang, Pan Zhang, Peng Zhang, Ruijie Zhang, Shuo Zhang, Songyang Zhang, Wenjian Zhang, Wenwei Zhang, Xingcheng Zhang, Xinyue Zhang, Hui Zhao, Qian Zhao, Xiaomeng Zhao, Fengzhe Zhou, Zaida Zhou, Jingming Zhuo, Yicheng Zou, Xipeng Qiu, Yu Qiao, Dahua Lin

摘要: 该报告介绍了 InternLM2，一种开源的大型语言模型，它在多个方面和基准测试中表现出色。该模型采用了创新的预训练和优化技术，并详细介绍了其预训练过程。InternLM2还采用了 Supervised Fine-Tuning和一种名为COOL RLHF的新型强化学习技术，以解决人类偏好和奖励作弊问题。该报告还提供了不同训练阶段和模型大小的InternLM2模型，以供社区参考。

标题: Cumulative Reports of the SoNDe Project July 2017

作者: Sebastian Jaksch, Ralf Engels, Günter Kemmerling, Codin Gheorghe, Philip Pahlsson, Sylvain Désert, Frederic Ott

摘要: 该报告是SoNDe探测器项目截至2017年7月的累积报告。报告内容包括：  
1x1模块技术演示器的报告  
所用材料的报告  
组件辐射硬度的报告  
潜在附加应用的报告  
2x2模块技术演示器的报告  
2x2技术演示器测试结果的报告

标题: Banach Couples. I. Elementary Theory

作者: Jaak Peetre, Per Nilsson

摘要: 该报告是Jaak Peetre的"Banach Couples. I. Elementary Theory"的精确副本，最初于1971年在吕勒（Lund）作为技术报告发布。报告中还添加了更多最近的一般参考资料，并对一些参考资料进行了更新。

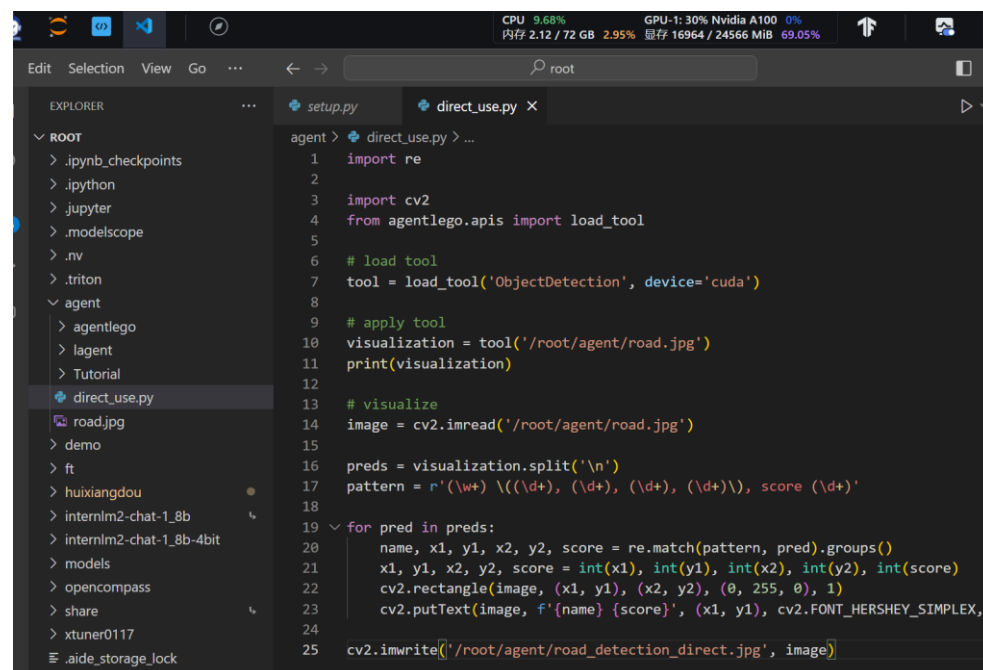
希望这些信息对您有所帮助！

2. 完成 AgentLego 直接使用部分，并在作业中上传截图。文档可见 直接使用 AgentLego。

```
(base) root@intern-studio-40069509:~# cd /root/agent
(base) root@intern-studio-40069509:~/agent# wget http://download.openmmlab.com/agentlego/road.jpg
--2024-04-18 23:40:38-- http://download.openmmlab.com/agentlego/road.jpg
Resolving proxy.intern-ai.org.cn (proxy.intern-ai.org.cn)... 172.18.128.194
Connecting to proxy.intern-ai.org.cn (proxy.intern-ai.org.cn)|172.18.128.194|:50000... connected.
Proxy request sent, awaiting response... 200 OK
Length: 284676 (278K) [image/jpeg]
Saving to: 'road.jpg'

road.jpg          100%[=====>] 278.00K  --.-KB/s    in 0.006s

2024-04-18 23:40:38 (46.0 MB/s) - 'road.jpg' saved [284676/284676]
```

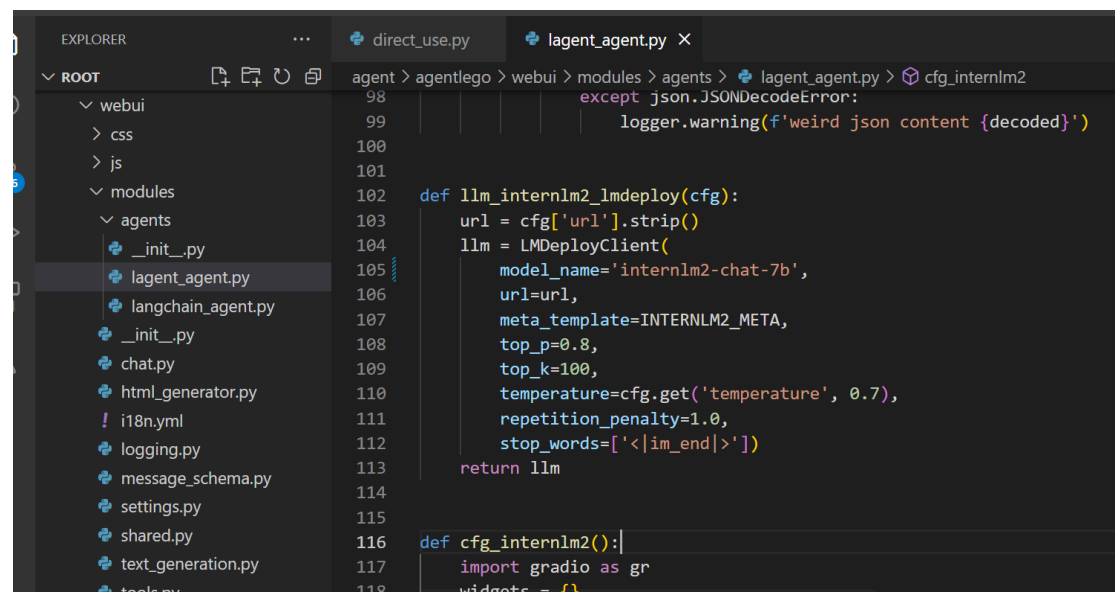


```
agent > direct_use.py > ...
1  import re
2
3  import cv2
4  from agentlego.apis import load_tool
5
6  # load tool
7  tool = load_tool('ObjectDetection', device='cuda')
8
9  # apply tool
10 visualization = tool('/root/agent/road.jpg')
11 print(visualization)
12
13 # visualize
14 image = cv2.imread('/root/agent/road.jpg')
15
16 preds = visualization.split('\n')
17 pattern = r'(\w+) \((\d+), (\d+), (\d+), (\d+)\), score (\d+)'
18
19 for pred in preds:
20     name, x1, y1, x2, y2, score = re.match(pattern, pred).groups()
21     x1, y1, x2, y2, score = int(x1), int(y1), int(x2), int(y2), int(score)
22     cv2.rectangle(image, (x1, y1), (x2, y2), (0, 255, 0), 1)
23     cv2.putText(image, f'{name} {score}', (x1, y1), cv2.FONT_HERSHEY_SIMPLEX,
24                 0.5, (0, 0, 0))
25 cv2.imwrite('/root/agent/road_detection_direct.jpg', image)
```

```
return _VF.meshgrid(tensors, **kwargs) # type:
ignore[attr-defined]
Inference
truck (345, 428, 528, 599), score 83
car (771, 510, 837, 565), score 81
car (604, 518, 677, 569), score 75
person (866, 503, 905, 595), score 74
person (287, 513, 320, 596), score 74
person (964, 501, 999, 604), score 72
person (1009, 503, 1047, 602), score 69
person (259, 510, 279, 575), score 65
car (1074, 524, 1275, 691), score 64
person (993, 508, 1016, 597), score 62
truck (689, 483, 764, 561), score 62
bicycle (873, 551, 903, 602), score 60
person (680, 523, 699, 567), score 55
bicycle (968, 551, 996, 609), score 53
bus (826, 482, 930, 560), score 52
bicycle (1011, 551, 1043, 617), score 51
```

## 进阶作业

1. 完成 AgentLego WebUI 使用，并在作业中上传截图。



```
EXPLORER
ROOT
  webui
    css
    js
    modules
      agents
        _init_.py
        agent_agent.py
        langchain_agent.py
        _init_.py
        chat.py
        html_generator.py
        i18n.yml
        logging.py
        message_schema.py
        settings.py
        shared.py
        text_generation.py
        tools.py
  agent > agentlego > webui > modules > agents > agent_agent.py > cfg_internlm2
    98
    99
    100
    101
    102 def llm_internlm2_lmdeploy(cfg):
    103     url = cfg['url'].strip()
    104     llm = LMDeployClient(
    105         model_name='internlm2-chat-7b',
    106         url=url,
    107         meta_template=INTERNLM2_META,
    108         top_p=0.8,
    109         top_k=100,
    110         temperature=cfg.get('temperature', 0.7),
    111         repetition_penalty=1.0,
    112         stop_words=['<|im_end|>'])
    113     return llm
    114
    115
    116 def cfg_internlm2():
    117     import gradio as gr
    118     widgets = {}
```

```
(agent) root@intern-studio-40069509:~#
(agent) root@intern-studio-40069509:~# lmdeploy serve api_server /root/share/new_models/Shanghai_AI_Laboratory/internlm2-chat-7b \
>
> --server-name 127.0.0.1 \
> --model-name internlm2-chat-7b \
> --cache-max-entry-count 0.1

[WARNING] gemm_config.in is not found; using default GEMM algo
HINT: Please open http://127.0.0.1:23333 in a browser for detailed api usage!!!
HINT: Please open http://127.0.0.1:23333 in a browser for detailed api usage!!!
HINT: Please open http://127.0.0.1:23333 in a browser for detailed api usage!!!
INFO: Started server process [69167]
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: Uvicorn running on http://127.0.0.1:23333 (Press CTRL+C to quit)
```

```
(base) root@intern-studio-40069509:~# conda activate agent
(agent) root@intern-studio-40069509:~# cd /root/agent/agentlego/webui
(agent) root@intern-studio-40069509:~/agent/agentlego/webui# python one_click.py

Looking in indexes: https://pypi.tuna.tsinghua.edu.cn/simple
Collecting gradio>=4.13.0
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/67/48/d71c7bbb92e72b30921cdb7034a437930695e7d7fd83290b766974792d31/gradio-4.26.0-py3-none-any.whl (17.1 MB)
    17.1/17.1 MB 250.2 kB/s eta 0:00:00
Collecting langchain
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/ed/3e/93045d37eba24e0b5eb05312e30cd9e12805ea5flae9ba51ec8a7d2f5372/langchain-0.1.16-py3-none-any.whl (817 kB)
    817.7/817.7 kB 402.2 kB/s eta 0:00:00
Collecting langchain-openai
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/a2/d2/141e675257d82dbd39d5b999f090e7d491db4a88f7312715f2ca9ddaa009/langchain_openai-0.1.3-py3-none-any.whl (33 kB)
Collecting aiofiles<24.0,>=22.0 (from gradio>=4.13.0)
  Using cached https://pypi.tuna.tsinghua.edu.cn/packages/c5/19/5af6804c4cc0fed83f47bfff6e413a98a36618e7d40185cd36e69737f3b0e/aiofiles-23.2.1-py3-none-any.whl (15 kB)
Requirement already satisfied: altair<6.0,>=4.2.0 in /root/.conda/envs/agent/lib/python3.10/site-packages (from gradio>=4.13.0) (5.3.0)
```

```
Successfully installed SQLAlchemy-2.0.29 aiofiles-23.2.1 aiohttp-3.9.5 aiosignal-1.3.1 async-
timeout-4.0.3 dataclasses-json-0.6.4 ffmpeg-0.3.2 frozenlist-1.4.1 gradio-4.26.0 gradio-client
-0.15.1 greenlet-3.0.3 importlib-resources-6.4.0 jsonpatch-1.33 langchain-0.1.16 langchain-co
mmunity-0.0.33 langchain-core-0.1.44 langchain-openai-0.1.3 langchain-text-splitters-0.0.1 la
ngsmith-0.1.48 marshmallow-3.21.1 multidict-6.0.5 mypy-extensions-1.0.0 openai-1.22.0 orjson-
3.10.1 packaging-23.2 pydub-0.25.1 python-multipart-0.0.9 ruff-0.3.7 semantic-version-2.10.0
shellingham-1.5.4 tomlkit-0.12.0 typer-0.12.3 typing-inspect-0.9.0 websockets-11.0.3 yar1-1.9
.4
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting beha
viour with the system package manager. It is recommended to use a virtual environment instead
: https://pip.pypa.io/warnings/venv
[00:29:16] INFO Loading tool `Calculator`
Running on local URL: http://127.0.0.1:7860
```

The screenshot shows the LMStudio web interface in a browser window. The address bar displays '127.0.0.1:7860'. The interface has three tabs: 'Chat', 'Agent', and 'Tools', with 'Agent' currently selected. Under the 'Agent' tab, the 'Agent' dropdown menu is set to 'internlm2'. To the right of this dropdown are three buttons: a blue square icon, a trash can icon, and a 'Load' button. Below this, the 'Agent class' is set to 'agent.InternLM2Agent', with a 'Save' button to its right. A status message on the right says 'Successfully loaded internlm2.'. The 'URL' section has a text input field containing 'http://127.0.0.1:23333'. Below the URL is a slider for 'Max number of turns' set to 6. Another slider for 'Temperature' is set to 0.7. The 'System prompt' section contains the text '当开启工具以及代码时，根据需求选择合适的工具进行调用'. The 'Plugin prompt' section contains the text '你可以使用如下工具：{prompt} 如果你已经获得足够信息，请直接给出答案。避免不必要的工具调用！同时注意你可以使用的工具，不要随意捏造！'.

Chat Agent Tools

Agent

internlm2

Agent class

agent.InternLM2Agent Save

URL

The internlm2 server url of LMDeploy, like 'http://127.0.0.1:23333'

http://127.0.0.1:23333

Max number of turns

6

Temperature

What sampling temperature to use.

0.7

System prompt

当开启工具以及代码时，根据需求选择合适的工具进行调用

Plugin prompt

你可以使用如下工具：  
{prompt}  
如果你已经获得足够信息，请直接给出答案。避免不必要的工具调用！同时注意你可以使用的工具，不要随意捏造！

Successfully loaded internlm2.

127.0.0.1:7860

Chat Agent Tools

Tools

New Tool

Tool class

ObjectDetection

Save

Import from tool server

Confirm

Name

ObjectDetection

Description

The tool can detect all common objects in the picture.

☒ Enable

Device

cuda:0


Initialize arguments


Loading tool ObjectDetection...

127.0.0.1:7860


Agent Tools

Current agent: internlm2





请检测图中的物体



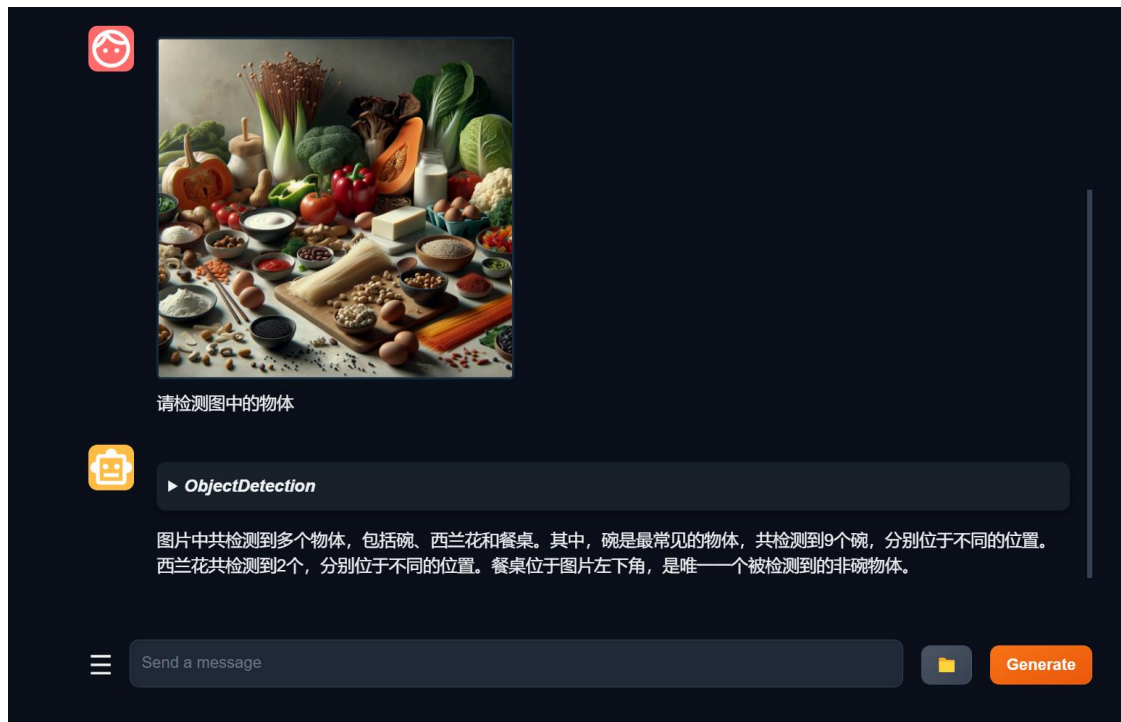
▼ ObjectDetection

Args: image=path

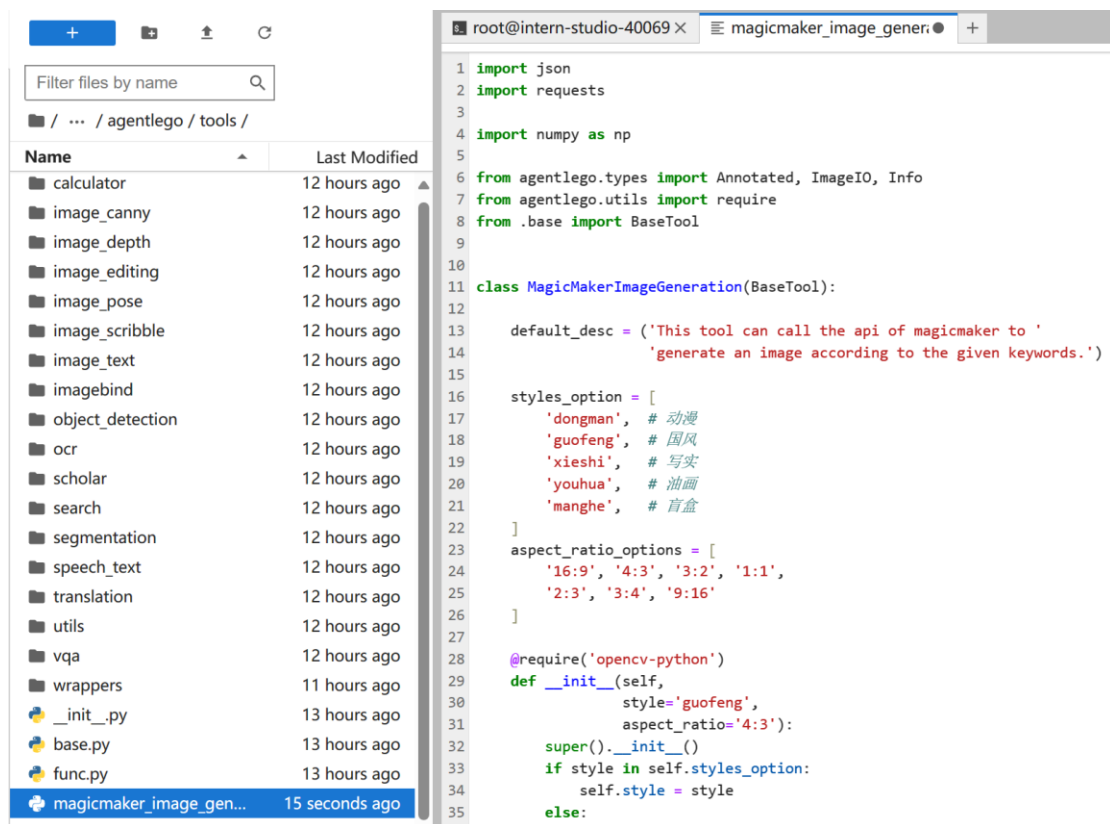
Response: airplane (1, 35, 508, 474), score 59 sports ball (342, 395, 399, 448), score 51

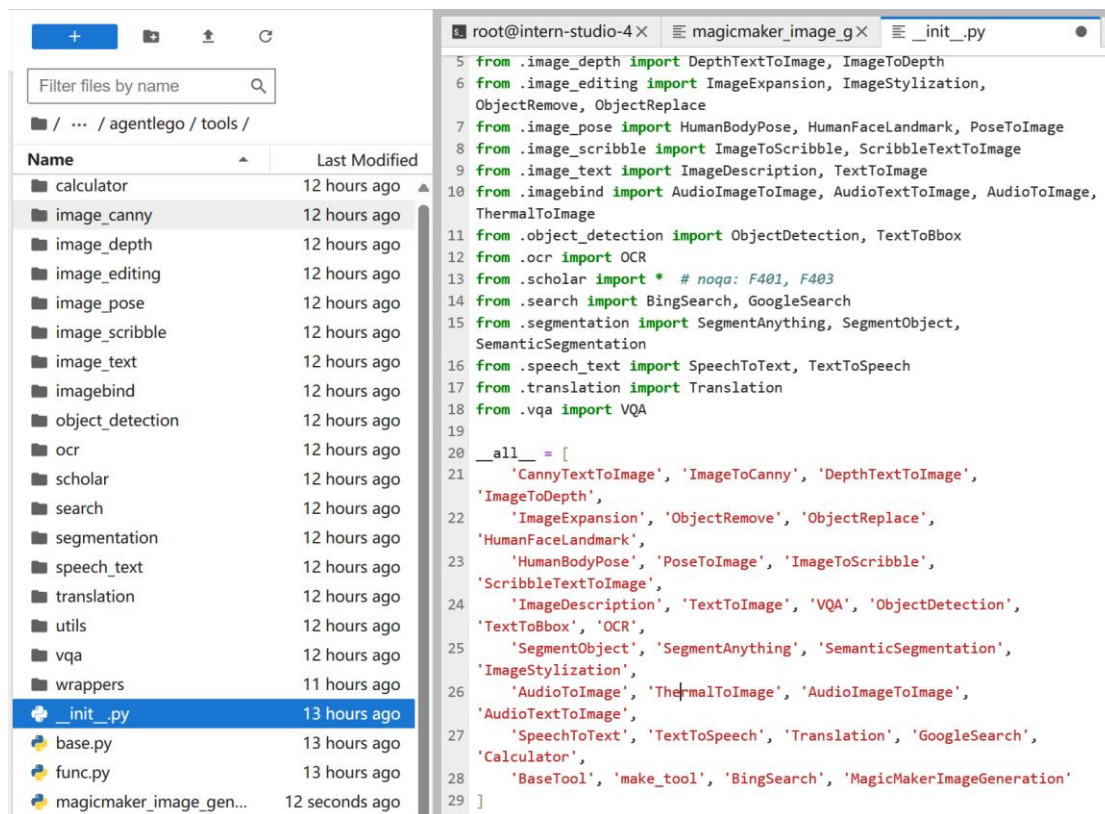
图片中共检测到两个物体，分别是飞机和体育球。飞机位于图片左上角的位置，坐标为(35, 474)，体育球位于图片右下角的位置，坐标为(399, 448)。





2. 使用 Lagent 或 AgentLego 实现自定义工具并完成调用, 并在作业中上传截图。





```

(base) root@intern-studio-40069509:/opt/jupyterlab# conda activate agent
(agent) root@intern-studio-40069509:/opt/jupyterlab# lmdeploy serve api_server /root/s
hare/new_models/Shanghai_AI_Laboratory/internlm2-chat-7b \
>
> --server-name 127.0.0.1 \
>
> --model-name internlm2-chat-7b \
>
> --cache-max-entry-count 0.1
[WARNING] gemm_config.in is not found; using default GEMM algo
HINT: Please open http://127.0.0.1:23333 in a browser for detailed api usage!!!
HINT: Please open http://127.0.0.1:23333 in a browser for detailed api usage!!!
HINT: Please open http://127.0.0.1:23333 in a browser for detailed api usage!!!
INFO: Started server process [54252]
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: Uvicorn running on http://127.0.0.1:23333 (Press CTRL+C to quit)

```