

附件 2

2023 年第五届
全球校园人工智能算法精英大赛
(算法挑战赛)

赛
题
及
比
赛
规
则

全球校园人工智能算法精英大赛组委会

2023 年 9 月

赛题一：零样本下的指代表达理解任务 比赛规则

一、赛题说明

视觉-语言多模态基础大模型(如 CLIP, SAM)近年来成为了主流研究的关注点之一。类比于 ChatGPT 的成功,多模态大模型一个巨大的应用价值就是下游任务的零样本(zero-shot)泛化能力。本赛题旨在利用已有的基础大模型,通过设计有效的视觉提示(visualprompt)和其他推理策略实现高性能零样本下的指代表达理解任务(ReferringExpressionComprehension,REC),即通过语言描述来定位图像中的目标物体框,如图所示:



数据部分我们将先提供 validationset, testAset。validationset 包含图像、候选目标初始框、语言描述查询、GT,便于选手迭代算法;testAset 仅包含图像、候选目标初始框、语言描述查询。

相关数据请直接解压缩 data/val.zip 和 data/testa.zip 即可。

二、解题思路

任务是零样本设定，因此没有任何训练数据，选手也无需训练任何神经网络模型，赛题只提供测试数据。赛题拟为参赛选手提供：

- 1.由 MAttNet 为每张测试图片检测得到的候选目标初始框；
- 2.SAM-ViT Huge；
- 3.CLIP 模型的 ViT-B/32 和 RN50x16backbone 版本。

选手利用且仅可利用上述预训练基础模型及所提供的候选目标初始框，通过设计有效的视觉提示流程和其他特定策略，实现高性能零样本下的指代表达理解任务。

我们将提供 ReCLIP: A Strong Zero-Shot Baseline for Referring Expression Comprehension (ACL2022) 的源码，并给出运行和保存/提交结果的流程。赛题描述中提及的 CLIP 相关的预训练模型在代码中会自动下载，SAM 的相关模型请选手自行从 SAM 的官网 [github: https://github.com/facebookresearch/segment-anything](https://github.com/facebookresearch/segment-anything) 自行下载使用。

环境配置：`pip install -r requirements.txt`

官方提供的代码包含了生成预测框结果并保存 json 文件的代码，以及根据预测结果计算得分代码

`split`: 数据集划分，取值可以为 `val`, `testa` 和 `testb`

`save_dir`: 保存结果的路径

推理并保存/提交结果：

```
CUDA_VISIBLE_DEVICES=0pythonmain.py\  
--input_file'data/{split}.jsonl\  
--detector_file'data/dets_dict.json\  
--image_root'data/train2014\  
--clip_model'ViT-B/32,RN50x16\  
--cache_path'{save_dir}/cache\  
--results_path'{save_dir}/{split}.json'
```

计算结果得分：

```
CUDA_VISIBLE_DEVICES=0pythonval.py\  
--input_file'data/{split}.jsonl\  
--results_path'{save_dir}/{split}.json'
```

三、评价方式

记语言查询数量为 n ，每个语言查询所得到的检测框和对应 GT 如果则记为一次命中，记命中总数为 m ，最终准确率指标为 $\frac{m}{n}$ 。

四、成绩提交

初赛阶段：保存所有样本的预测框为 json 文件，文件名称保存为：队伍名.json，提交网页即可，json 文件的生成脚本见上文的解题思路部分。

json 文件参考格式：其中 $x1,y1$ 代表回归框左上角坐标， $x2,y2$ 代表右下角坐标

```
[  
  [x1,y1,x2,y2],
```

[x1,y1,x2,y2],

[x1,y1,x2,y2],

...

]

后续阶段：要求提交可编译运行的作品源代码以及运行命令说明文件

五、联系方式

参赛 QQ 群：833020799

赛题二：电力大模型问答挑战赛比赛规则

一、赛题说明

当前大语言模型在众多任务上都展现出强大的问答能力，但在知识密集型领域往往表现欠佳，如何提升大模型在知识密集型领域的的能力是一个有趣的问题。电力是一个典型的知识密集型领域，同时涉及电子、物理、化学、数学等学科知识，是一个良好的衡量大模型知识密集型能力的场景，一个优秀的电力问答大模型需要同时掌握多个学科背景知识，并具备在此基础上进行归纳和推理的能力。

本次比赛要求参赛选手以 ChatGLM2-6B 模型为核心制作一个问答系统，回答电力领域相关的问题。题目类型包含单选题、多选题和自由问答，下面是题目样例：

单选题样例：

题目：电流对人体的效应由生理参数和电气参数决定。

15~100Hz 正弦交流电流反应阈的通用值为（ ）。

选项 A：1.5mA

选项 B：2mA

选项 C：0.1mA

选项 D：0.5mA

答案：D

多选题样例：

题目：下列选项中，电力负荷应该为三级负荷的是（）。

选项 A：中断供电将在经济上造成较大损失的负荷

选项 B：中断供电影响重要用电单位正常工作的负荷

选项 C：一般货梯和自动扶梯

选项 D：不属于一级和二级的电力负荷

答案：CD

自由问答样例：

题目：如果一台 p 对磁极的单叠绕组，其元件电阻为 r_a ，电枢电流为 I_a ，若把它改接为单波绕组，并保持支路电流不变。试问电枢电阻和电枢电流变为多少？

答：因为 p 对极的单叠绕组有 $2p$ 条支路并联，所以每条支路电阻为 $2pr$ ， $2p$ 条支路串联后的总电阻为 $4p^2r_a$ 。改为单波绕组时，则仅有两条支路并联，每条支路电阻为 $\frac{1}{2}(4p^2r_a) = 2p^2r_a$

电枢电阻为： $\frac{1}{2}(2p^2r_a) = p^2r_a$ ，增加到原来的 p^2 倍，单波绕组时，支路电流为 $\frac{I_a}{2p}$ ，改为单波绕组时，支路电流不变，但仅有两条支路，所以总电流为 $2 \times \frac{I_a}{2p} = \frac{I_a}{p}$ ，减少到了原来的 $\frac{1}{p}$ 。

除上述样例外，本次比赛所提供数据均为测试集，用于测试大模型的能力。

比赛不允许使用其他的大语言模型，但可以使用其他大语

言模型生成的数据，也可以使用其他公开访问的外部数据来微调模型。选手需要严格使用大模型进行作答并提交答案，严禁人工作答，违者取消参赛资格。

二、解题思路

一个简单的解题方式是将题目直接输入到 ChatGLM2-6B 模型，并构建指令让模型针对对应的题目类型进行作答。升级的做法是收集电子、物理、化学、数学等学科的背景知识，以检索的方式检索出和当前问题相关的知识，并作为背景知识输入到模型，帮助模型进行作答。也可以采用预训练、指令学习将学科背景知识融入模型，从而提升模型的电力问答能力。

三、评价方式

选手报名成功后，举办方提供问答测试数据，参赛队伍在本地调试算法。获得模型预测的答案后，参赛队伍在线提交结果，系统会实时评估并返回成绩，按照评估指标从高到低进行排序。每个队伍每天的提交次数限制为 3 次。

实时排行榜评分采用 AB 榜形式。比赛期间排行榜仅显示 A 榜成绩，B 榜成绩作为最终成绩排名。比赛将会选择选手的最后一次提交以及选手的 A 榜最高分的模型作为选手 B 榜分数的计算依据，以这两次提交在 B 榜的最高分进行排名。

评价指标：

单选题得分与多选题得分均按下式计算：

其中 N 表示的是测试集中单选题/多选题对应题型的总数，

表示单选题/多选题的单题分值，其中表示的是单题选项命中率，表示模型回答的选项集合，表示正确选项集合。需要注意的是，指示函数表示回答选项必须在正确选项中才计分。选项可以少答，不能多答，多答该题计 0 分。

自由问答的得分按照 ROUGE-L 计算：

其中 M 表示的是测试集中问答题的总数，表示模型给出的回答，表示参考回答，LCS 即 Longest Common Subsequence(最大公共子串)，表示 QA 题的单题分值。

最终得分的加权公式：

该公式将单选、多选、自由问答的总得分归约到 100 分。

四、成绩提交

参赛队伍自行调试算法，在线提交结果，结果文件命名为"参赛队名称_result"，以 utf-8 编码格式保存。文件每行是一个 json 串，包含"ID"，"answer"。比赛期间不需要提交模型。其中，单选题回答"A"- "D"中的一项（例如，A），多选题回答选项以"、"分割（例如，A、C），问答题的答案为一段连续文本。

赛题成绩提交截止日期后，入围队伍需要上交代码、模型、说明文件（详细描述模型的复现过程）。只有模型复现成功的队伍可以进入比赛的下一阶段。对于结果无法复现的队伍，将取消比赛资格，并按照排行榜顺次选取队伍入围。

结果提交网站及后续相关通知详见 QQ 群通知。

五、省赛说明

为了确保比赛的公平和公正，参赛选手需要在省赛截止日期之前提交结果文件。赛事主办方将核实问答结果的可复现性，未通过验证的团队将被取消成绩。赛题成绩提交截止日期后，对所有通过验证的队伍按照同一省赛区内参赛团队比赛成绩排名确定省赛一、二、三等奖获奖名单（以参赛团队队长所在高校为准）。若某省赛赛区参赛团队数量少于 50 支则合并后统一按成绩进行排名。

省赛一、二等奖参赛团队进入全国总决赛。

六、总决赛

进入全国总决赛团队最终提交成绩、代码及解题思路说明（未通过可复现性验证的团队将被取消成绩），经大赛组委会组织线上/线下答辩，最终参赛选手总决赛成绩，并按照成绩排名（不划分省赛区）确定全国赛一、二、三等奖最终获奖名单。大赛组委会根据赛事实际情况，举办全国总决赛颁奖典礼。

七、联系方式

参赛 QQ 群：852305141

赛题三：道路病害目标检测

比赛规则

一、赛题背景

道路病害的及时发现和处理道路病害对于保障道路的安全、舒适和使用寿命具有重要意义。道路病害检测的课题是道路评级，道路桥梁工程验收，路政部门养护等实际业务被广泛需要。

在本次“道路病害目标检测挑战赛”中选手将使用人工智能技术地对道路病害检测，结合实际生产数据设计模型算法，提升自己的算法开发水平。

二、任务描述

（一）任务描述

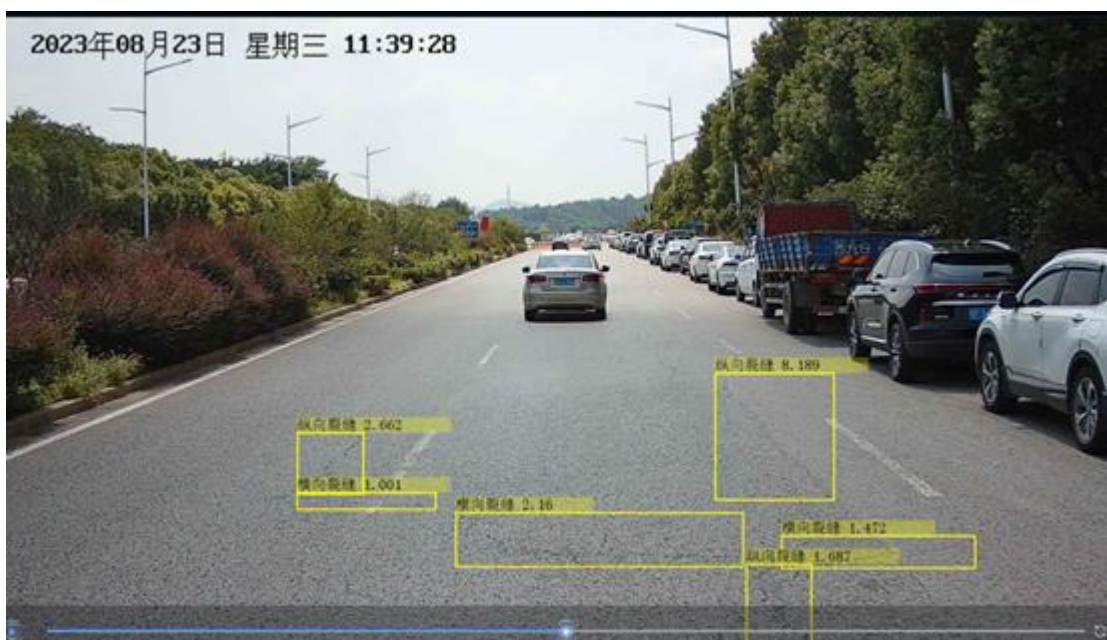
- 1.构建算法模型检测出图片中横向裂缝、纵向裂缝、块状裂缝、龟裂、坑槽、修补网状裂缝、修补裂缝、修补坑槽的目标。
- 2.利用训练数据构建算法模型。
- 3.构建算法模型的 **docker** 镜像并提供容器运行说明手册，要求说明算法程序如何读取输入图像数据生成预测文件。
- 4.初赛和决赛在两批数据上进行选拔，将从提交作品的团队中，筛选 9 支队伍参加决赛。

（二）训练数据简介

2000 张使用 **txt** 文件标注图片，标注文件和图片文件文件名一一对应，仅文件名不同后缀，类别编号为：横向裂缝 0、纵向

裂缝 1、块状裂缝 2、龟裂 3、坑槽 4、修补网状裂缝 5、修补裂缝 6、修补坑槽 7

(三) 图片数据示例



标注数据示例及说明：一个标注文件中会存在大于等于 1 行的标注信息，如下图所示，信息由空格分分格，各位置字段含义分别为：目标类别， x_center , y_center , w , h ，其中 x_center , y_center 表示目标中心点坐标占图片长和高的比例； w , h 表示目标长和高占图片的比例。

```
j2_0.txt - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)
2 0.097808 0.650167 0.172890 0.283335
2 0.301638 0.535417 0.221782 0.218168
2 0.444519 0.442868 0.059206 0.120834
2 0.516854 0.410368 0.085466 0.099167
2 0.588799 0.363484 0.082792 0.077634
2 0.645292 0.314976 0.063503 0.081667
2 0.691367 0.294768 0.031035 0.055417
2 0.713684 0.276744 0.040284 0.057192
2 0.742985 0.254971 0.033627 0.055655|
```

(四) 预测程序要求

输入参数 1 为预测图片文件夹路径，输入参数 2 为生成预测文件文件夹路径，要求程序执行完毕后在预测文件文件夹路径中生成每一个预测图片对应的预测文件，预测文件的构成与标注文件一致。

(五) 考核要求

1. 镜像程序可运行且功能正确

2. 程序在 500 张测试集上总预测时间不超过 12 小时

3. 使用 Macro-f1 评价各队伍成绩， $\text{Macro-f1-score} = \frac{1}{N} \sum_{i=1}^N \text{f1}_i$ ，其中， N 为类别个数， $\text{recall}(\text{召回率}) = \frac{\text{检测正确的目标数量}}{\text{检测正确的数量} + \text{漏检目标的数量}}$ ； $\text{precision}(\text{精确率}) = \frac{\text{检测正确的目标数量}}{\text{检测正确目标的数量} + \text{误检目标的数量}}$ 。

三、作品提交

1. 提交内容：程序镜像文件及容器运行说明手册；

2. 发送到：yewenwu.js@chinatelecom.cn；

3.邮件主题：题目+队长姓名+学校+电话（学生信息需与注册信息一致）；

4.邮件附件：请使用 QQ 邮箱的文件中转站并从 QQ 邮箱发出为佳；

5.提交次数：每位参赛队伍每周最多成功提交 1 次。

四、资源下载

资源下载链接通过 QQ 群发布。

五、大赛 QQ 群及线上答疑

1.所有注册参赛选手请按照“姓名+学校名”的格式提交入大赛 QQ 群；

2.数据下载链接等赛事信息将通过 QQ 群发布；

3.赛事或赛题相关问题请发送至 QQ 群，主办方技术专家或工作人员将及时解答；

六、联系方式

参赛 QQ 群：136366826

赛题四：AIGC 驱动的小样本分类 比赛规则

一、赛题背景

近年来，随着 GAN、扩散模型等生成模型的提出，人工智能生成内容（AIGC）技术引起了广泛关注。在这个领域中，漫画生成是一个重要的应用方向，在日常生活中几乎随处可见，因此对漫画图像的识别具有重要的实际应用价值。然而，受制漫画绘制的复杂性和困难性，通常只能收集少量的漫画图像可供于训练，这对于传统的深度学习算法而言有较大的困难。因此，小样本学习成为解决这一问题的有效方法。相对于传统的大数据量下的机器学习分类问题，小样本学习算法因其能够仅利用有限数据实现快速适应新任务的特性而备受关注，这弥补了深度学习对数据的高度依赖。为了满足实际任务需求，本赛题使用漫画人脸数据集，要求参赛者利用小样本学习相关技术，在 5-way 5-shot 和 5-way 1-shot 两个经典小样本场景下完成漫画人脸的识别任务。

二、赛题任务

参赛者需利用小样本学习相关技术，在 5-way 5-shot（5 个类别，每个类别 5 个训练样本）和 5-way 1-shot（5 个类别，每个类别 1 个训练样本）两个小样本任务场景分别建立一个高效的漫画人脸识别模型。模型应能够根据测试集中的漫画人脸图像，准确地将它们分类到对应的漫画类别。

三、数据简介

数据来源于 WebCaricature 数据集，原数据集共包括属于 252 个人物的 6042 幅漫画照片。我们将数据集分割为训练集，验证集和测试集，其中训练集有 200 个类别，验证集有 32 个类别，测试集暂不提供。同时由于所有的漫画图片均来自网络爬虫爬取，因此数据集中的漫画的艺术风格是多种多样。此外，该数据集中也并不对光照条件、姿态、表情、遮挡和年龄等信息进行约束。

四、数据说明

WebCaricature/images	训练集图片文件夹，包含数据集中所有的图片
WebCaricature/train.csv	含有数据训练集信息的表格文件，其中每行有两列，分别为 filename 和 label，对应 images 文件夹下的图片名和所属类别
WebCaricature/val.csv	含有数据验证集信息的表格文件，内容格式与 train.csv 一致

五、提交文件格式

参赛队员需仿造参考资源中提供的开源库，分别提交 5-way 5-shot 和 5-way 1-shot 两个场景下训练完成后的模型权重，赛方会基于该开源库提供的测试接口对模型的算法能力进行评估。

六、测评标准

在评测阶段，赛方会对两个场景下的算法能力进行评估。具体而言，赛方会将测试集图像进行划分，在 5-way 5-shot 和 5-way 1-shot 两个小样本场景下随机模拟生成若干个任务，计算在

每个任务的 Top-1 分类精度。最终使用同一个场景下所有任务的 Top-1 分类精度的平均值作为该场景下的分类精度。

七、参考资料

数据集获取以及小样本学习相关算法实现可参考

<https://github.com/RL-VIG/LibFewShot>

八、联系方式

参赛 QQ 群：815232826