赵 洵

Xun Zhao

1998/08/17
18810903806
zhaoxun16@mails.ucas.ac.cn
GitHub: https://github.com/CarnoZhao
WeChat: Carno_Zhao
QQ: 273806108

教育经历

16.09-20.06 本科, 中国科学院大学, 生物科学, GPA：3.90/4.00

20.09-23.06 硕士, 中国科学院自动化研究所, 模式识别与智能系统, GPA：3.91/4.00

- 实验室：中国科学院分子影像重点实验室
- 研究方向：医疗影像分析

竞赛经历

2022 Kaggle（冠军，排名 1/1565）

UW-Madison GI Tract Image Segmentation

主要问题 数据集中 MRI 伪影导致标注区域与正常认知范围差异较大，非医学专业人员会产生大量误判，此类标注会对模型训练带来不稳定性。

解决方案 我们提出了粗分类-细分割的两阶段式分割方案，充分利用像素级标注的监督信息，创新性的使用分割模型执行第一阶段伪影区域分类任务，对非伪影区域采用第二阶段精细分割。

方案效果 因为在伪影检测中我们以分割代替分类，利用了更丰富的监督信息，和同样采用两阶段方案的队伍相比，即便第二阶段方案设计类似，但我们的模型在第一阶段伪影检测效果更好，最终获得冠军。

2022 天池大数据竞赛（冠军，排名 1/1149）

真实场景篡改图像检测挑战赛

主要问题 竞赛所提供的训练数据较少，且训练数据和测试数据存在一定的分布差异。另一方面，大多数队伍方案相近，均采用大模型、大尺度提升分割效果，需要从额外角度提升模型效果。

解决方案 为解决数据缺乏的问题，我们对收集的无标注外部数据采用伪标签方法，使用半监督学习流程训练模型。为了在方案设计上和其他队伍拉开差距，我们设计了多模型交叉伪标签训练的策略，利用 CNN 和 Transformer 特征提取的偏好性，互相监督以增强模型的特征提取能力。

方案效果 和其他同样采用伪标签的队伍相比，我们的交叉伪标签效果优于单类模型伪标签，在不同数据集下都有较好的鲁棒性，在决赛中长期大幅度领先其他选手，最终获得冠军。

2020 第八届 CCF 大数据与计算智能大赛（冠军，排名 1/1998）

面向数据安全治理的数据内容智能发现与分级分类赛道

主要问题 竞赛本身设计为半监督 + 无监督任务，在 10 个目标类别中，仅有 7 类提供少量标注数据，另 3 类不提供标注数据，因此需要选手针对这 3 类设计无监督算法，再针对其余 7 类设计半监督算法。

解决方案 我们依据“高维空间中的低维流形”假设，设计了完全无监督的分类策略：依靠 Bert 模型的强大的语义提取能力，仅采用 t-SNE 和 DBSCAN 对 Bert 特征进行降维聚类即可解决无监督问题，再使用伪标签技术即可解决半监督问题。

方案效果 其他选手一般使用传统 TF-IDF、人工标注等繁琐步骤解决无监督问题，在执行效率和准确性上和我们的方案存在极大的差距，这也导致了后续半监督任务我们的伪标签质量更高。因此我们最终获得了冠军。

其他奖项

- Kaggle：Sartorius - Cell Instance Segmentation （金牌）
- Kaggle：TensorFlow - Help Protect the Great Barrier Reef （银牌）
- Kaggle：Hubmap - Hacking the Kidney（银牌）
- 2021 科大讯飞 iFLYTEK A.I. 开发者大赛（三等奖）
- 2021 Sodic 全球开放数据应用创新大赛（二等奖）
- 2020 第三届“金风杯”能源创新挑战赛（特等奖）

发表论文

[1] **X. Zhao** et al., “Deep learning signatures reveal multiscale intratumor heterogeneity associated with biological functions and survival in recurrent nasopharyngeal carcinoma,”Eur J Nucl Med Mol Imaging, Apr. 2022, doi: 10.1007/s00259-022-05793-x. (IF: 10.057)