多巡天数据和AI驱动的特殊恒星搜寻与分类研究

谈磊、王锋、邓辉、梅盈、刘超
罗杨平、柳志存、王小龙
广州大学智能天文信息处理团队
国家天文台
西华师范大学
河北师范大学

目录

CONTENTS

01
引言

02
工作进展

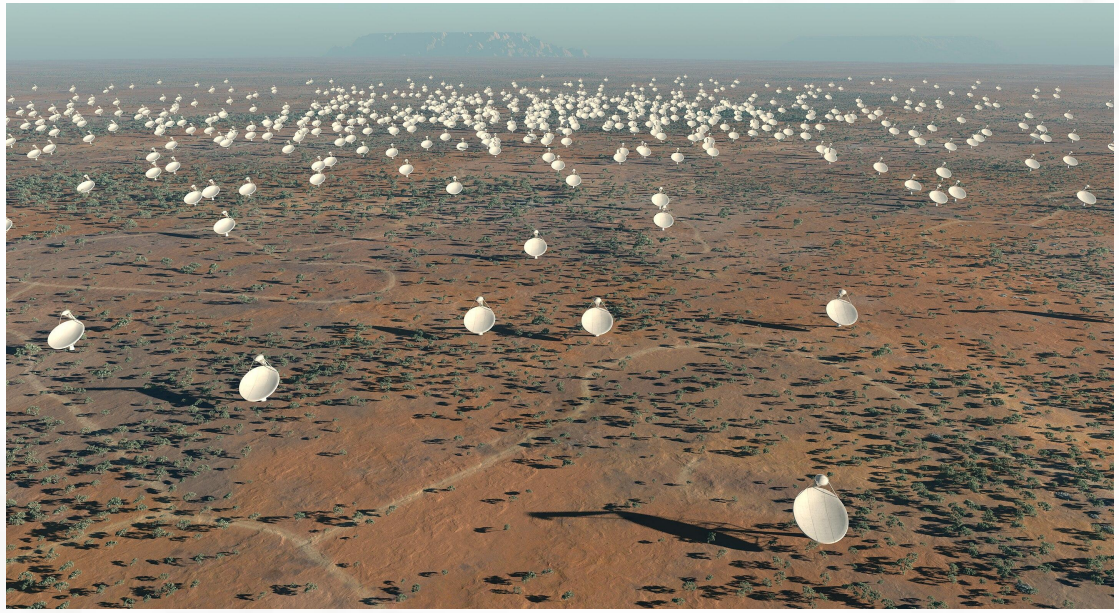
03
分析拓展

01

引言

引言





天文巡天进入“大数据时代”：

- LAMOST发布超5000万条单次曝光谱（截至DR13）。
- SDSS发布累计超500万光谱（至DR17）。
- DESI DR1发布了超1800万条光谱。

未来预期：

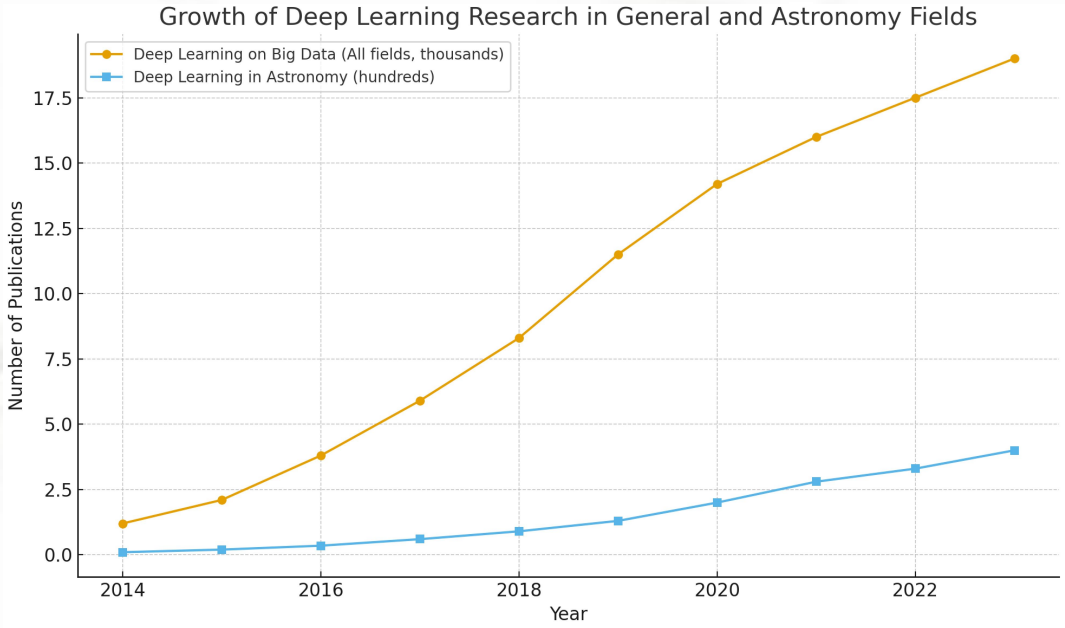
- CSST预计获取上亿天体成像和数千万光谱。
- SKA（平方公里阵列）每日数据量可达TB级。

面临挑战：

- 数据量大 → 人工处理不可行。
- 数据维度高、噪声复杂、异构严重。

AI优势：

- 可自动提取特征、适应多模态、具备高识别率。
- 是实现高效识别与分类的核心技术手段。



通过机器学习、深度学习方法进行海量数据处理的论文数量逐渐上升。

引言

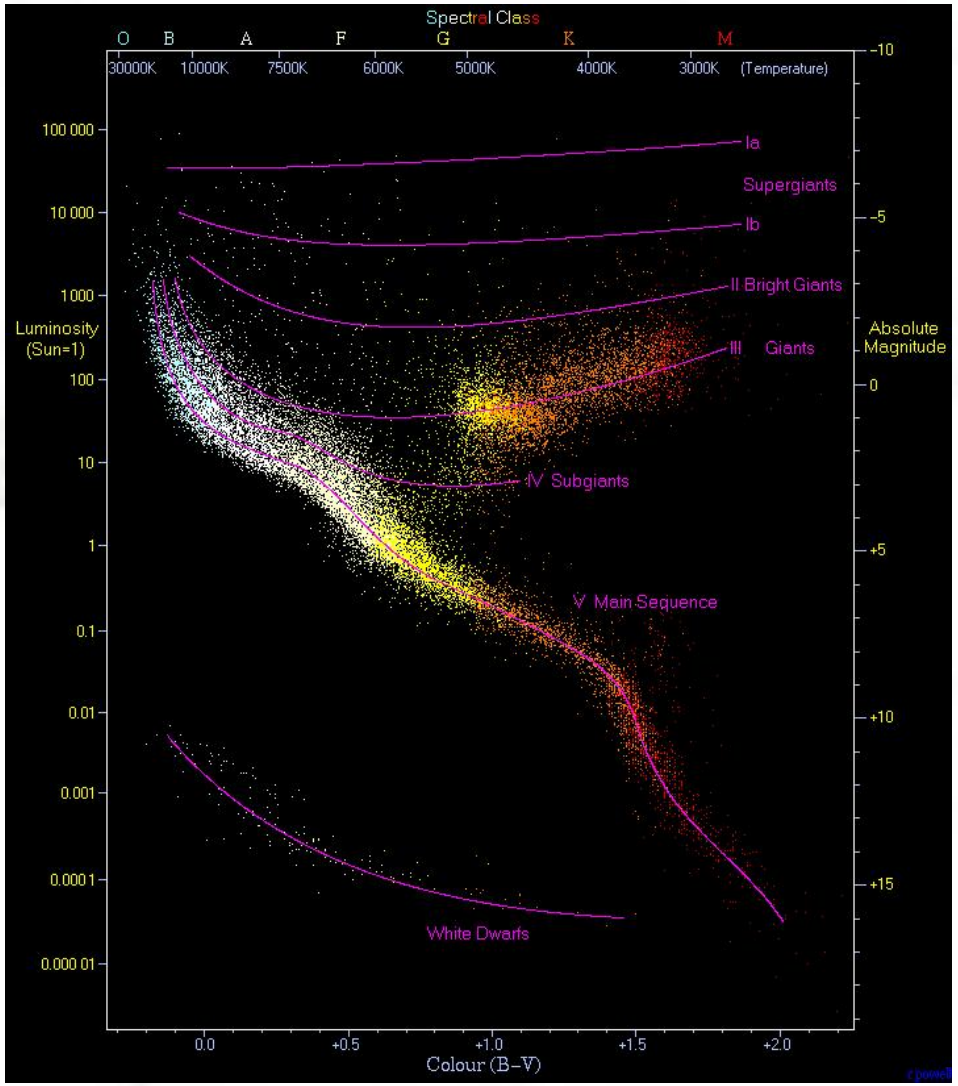
1、特殊天体探索银河系与恒星演化的关键窗口：

- 热亚矮星：与Ia超新星前身有关
- 白矮星：约束恒星生命周期，反映星族年龄
- 碳星、EMP：记录早期宇宙的化学演化
- 双星、CV等：涉及致密天体、质量交换过程
- YSO：揭示恒星形成机制与星际环境

2、探索**恒星形成与演化的边界**

3、揭示**银河系结构的历史轨迹**

4、发现**极端天体或前沿物理过程**



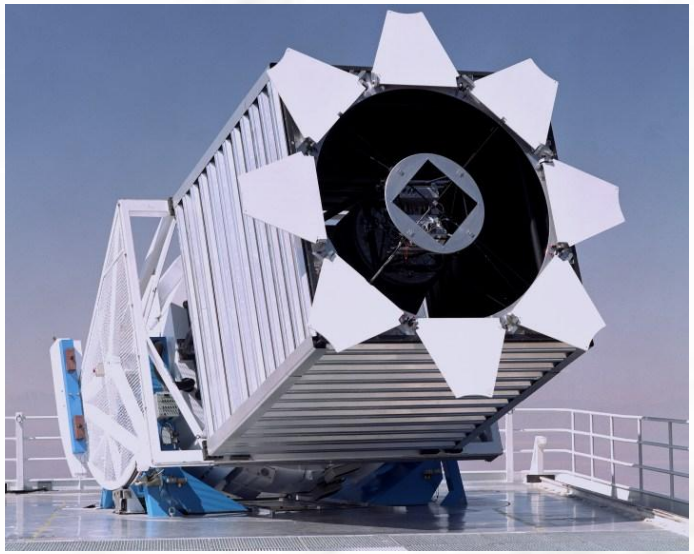
02

工作进展

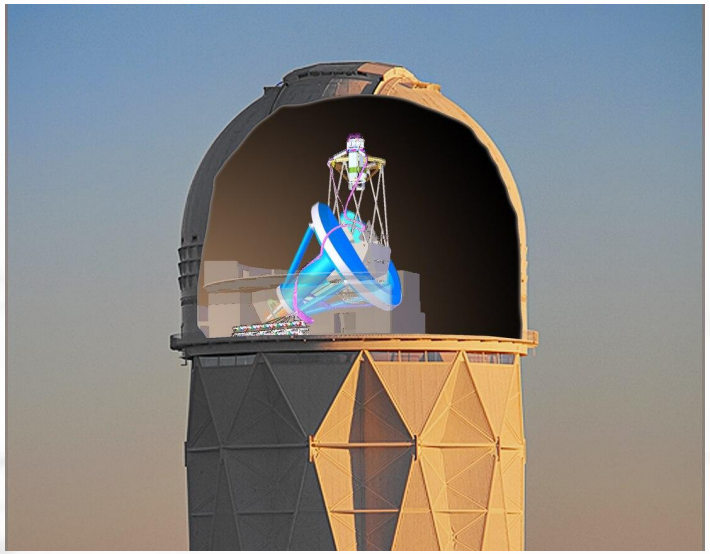
研究数据



LAMOST



SDSS



DESI

数千万光谱，覆盖3700－9000Å，低、中、高分辨率！

多源辅助数据：

- Gaia：视差、运动学、距离、光度
- ZTF、TESS、ASAS—SN：光变曲线，周期/爆发行为
- ALLWISE、Spitzer：红外颜色

光变曲线、测光参数、红外颜色等多模态数据。

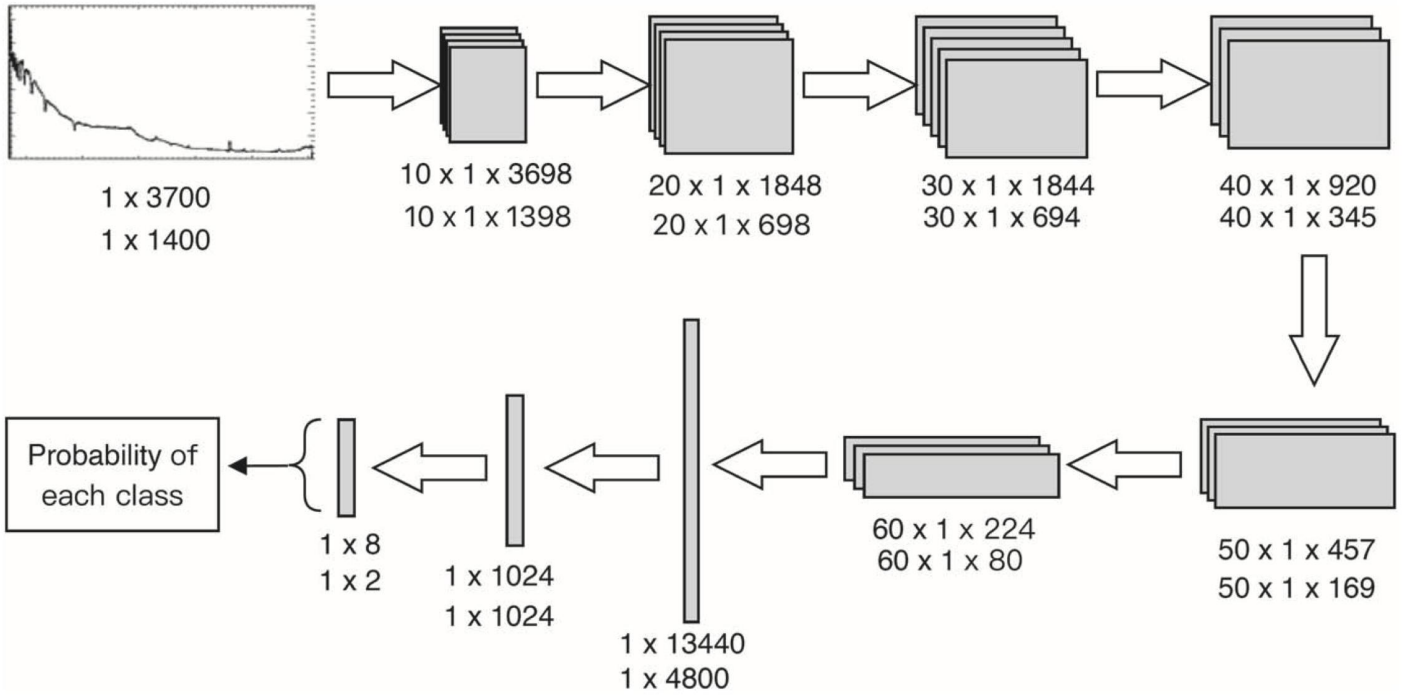
热亚矮星、碳星

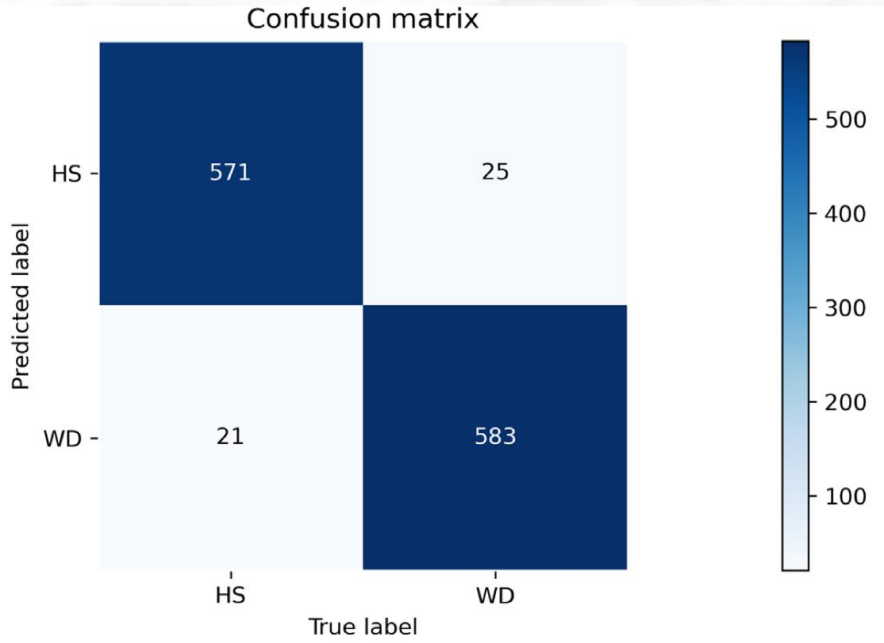
数据：LAMOST DR7-V1和LAMOST DR9 V1。热亚矮星、碳星来源于前人发布的星表。主序星来源于交叉匹配（SDSS、MKCLASS）和人眼证认，O型通过拟合黑体辐射谱扩充数据。

方法：基于CNN的模型。先对目标源进行粗筛，然后通过进一步的分类精确识别。

结果：识别到2092个热亚矮星光谱（25颗新的）、 4383 颗碳星（1197颗新的）。

Tan et al. APJS. 2022.　　He et al. APJS. 2024.





白矮星

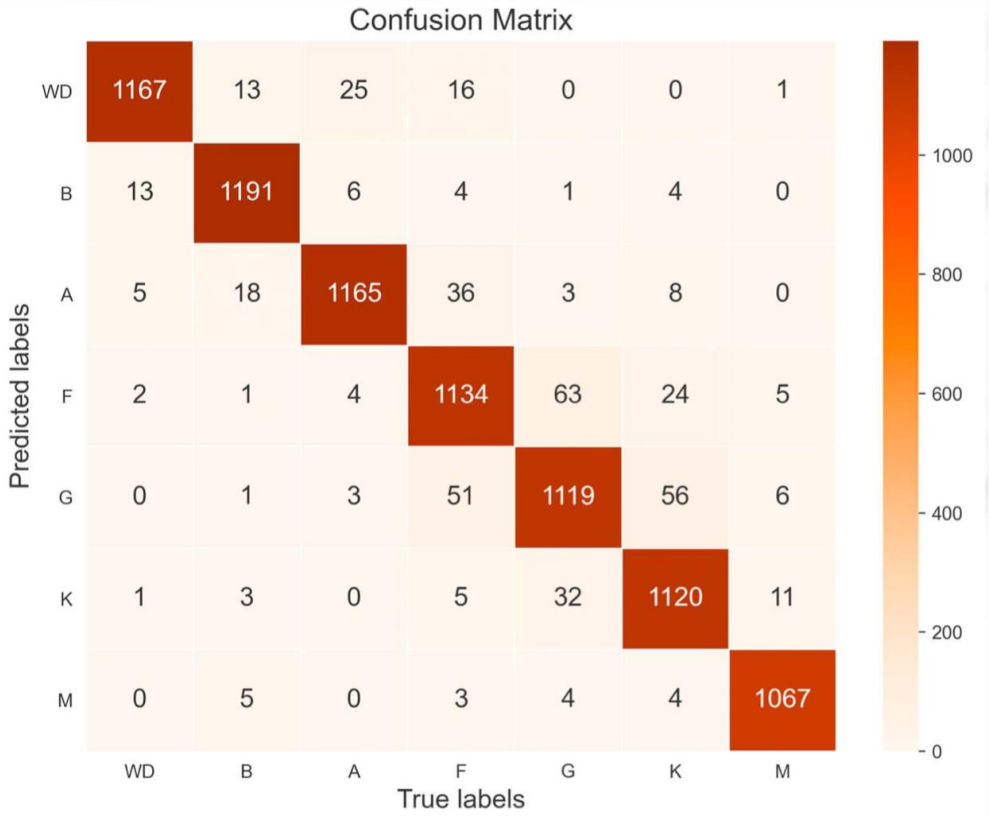
数据：LAMOST DR9 V1、DESI DR1。

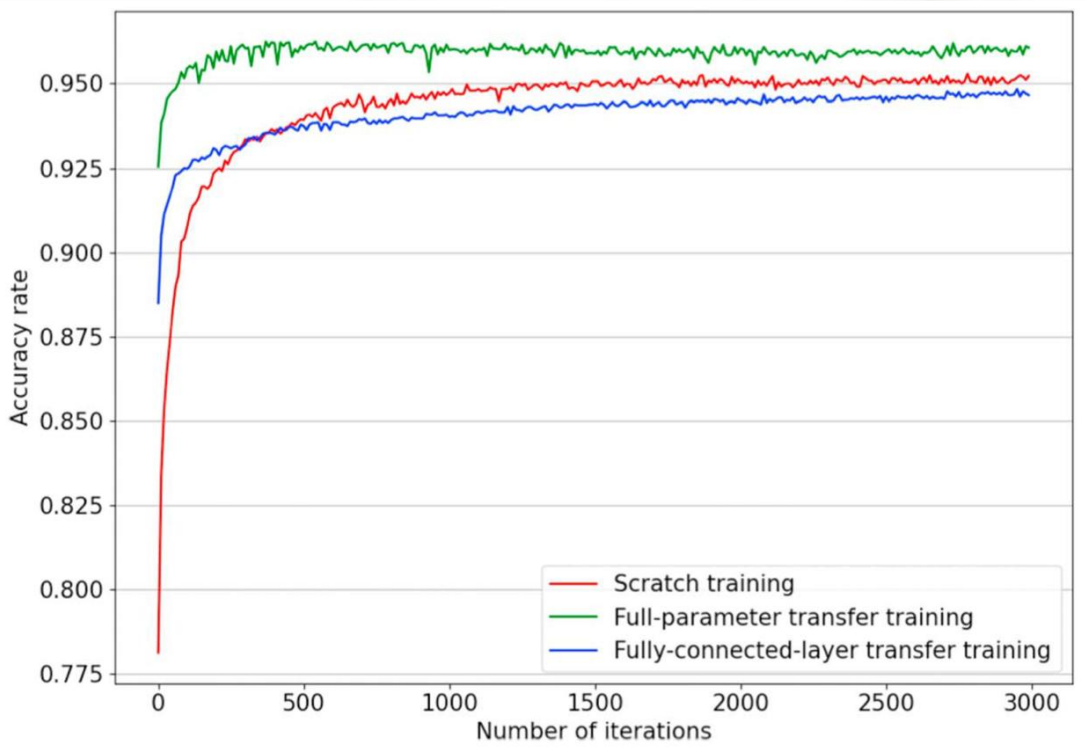
方法：基于CNN、LSTM的迁移学习模型。

结果：在LAMOST中识别到5503颗（489颗新的）白矮星，在DESI中识别到27423颗（1765颗新的）。

Tan et al. APJS. 2023.,

Tan et al. Prepare for submission.





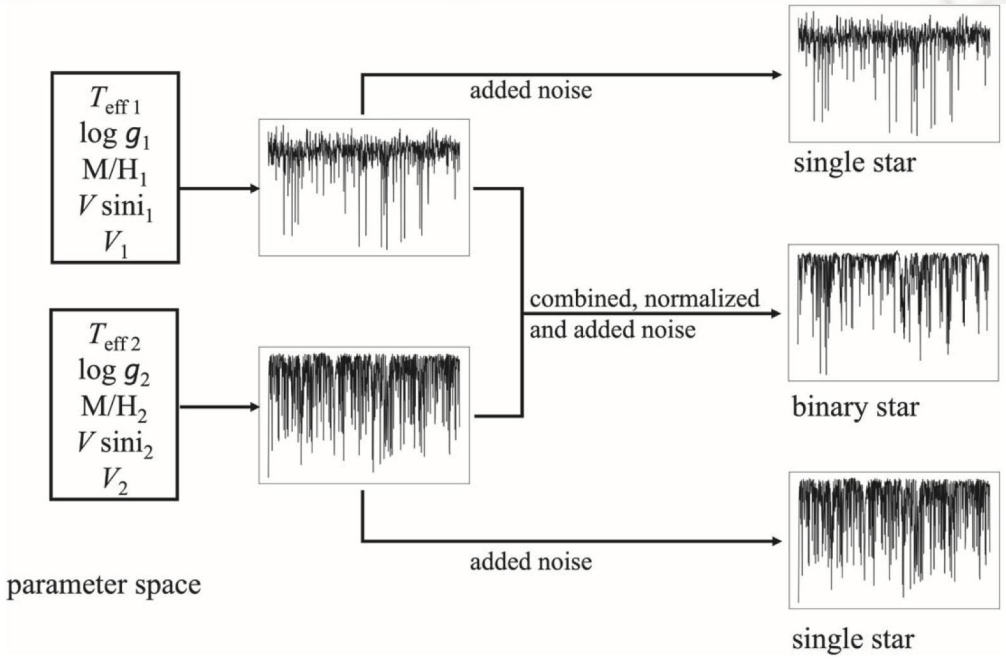
双星

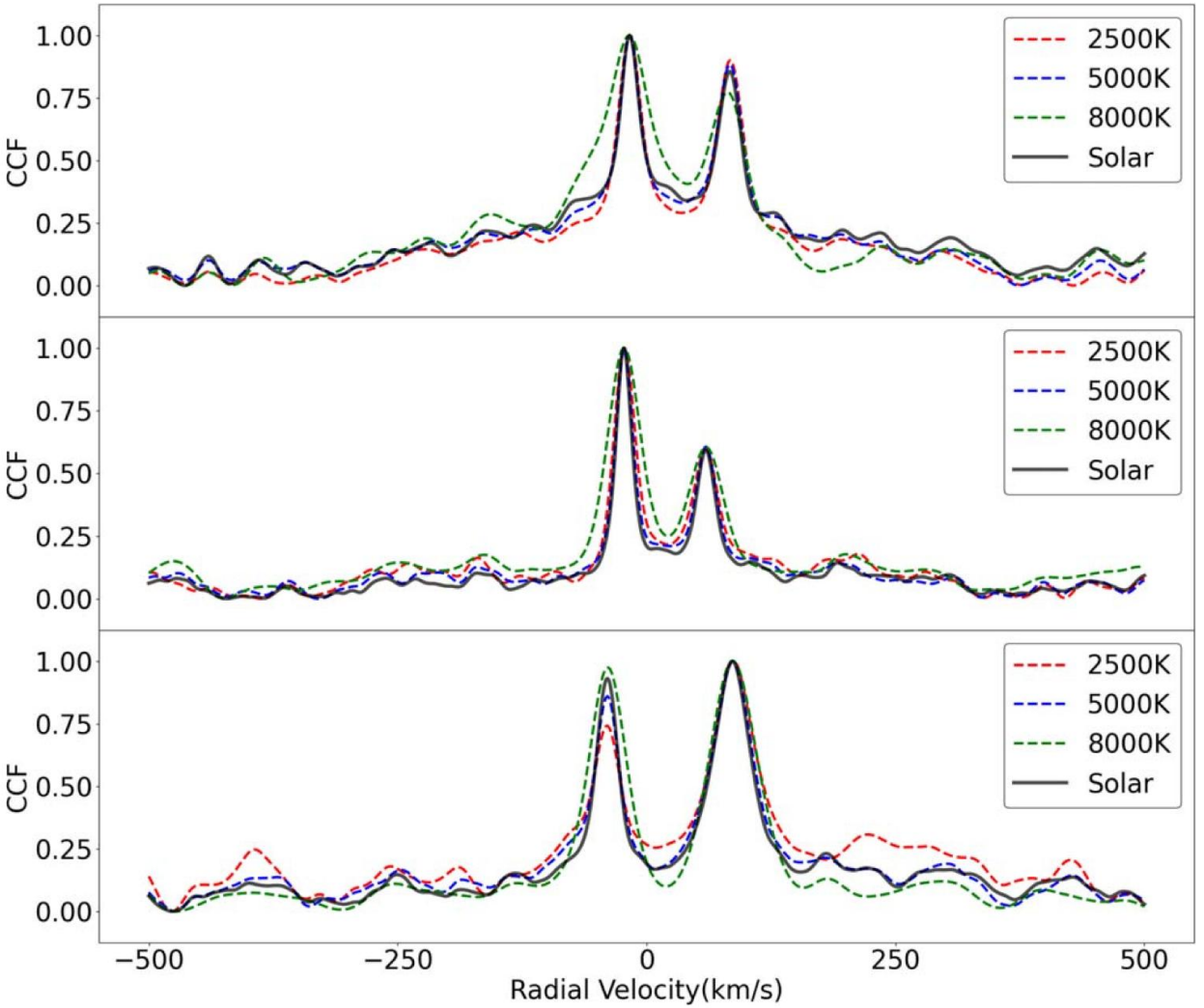
数据：LAMOST DR9中分辨率、 iSpec spectral synthesis software仿真数据。

方法：基于CCF、CNN。

结果：识别到了728个SB2星（281颗新的）。

Zheng et al. APJS. 2023.,





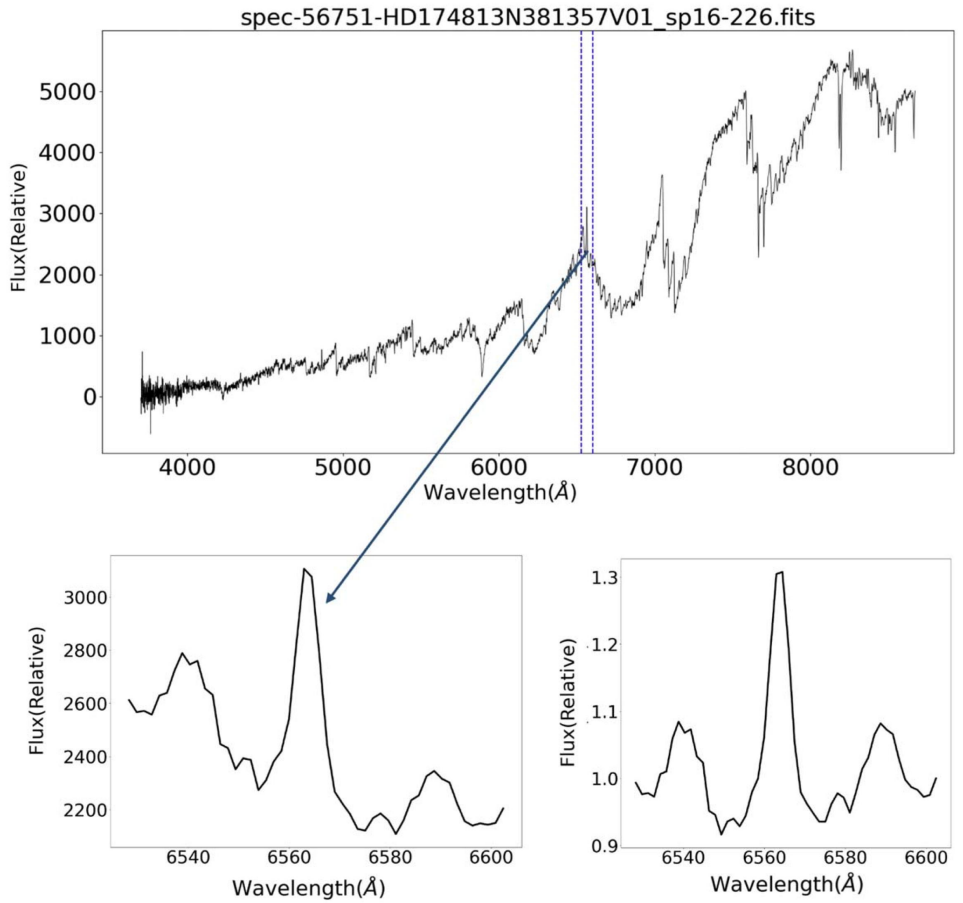
低分辨率YSO

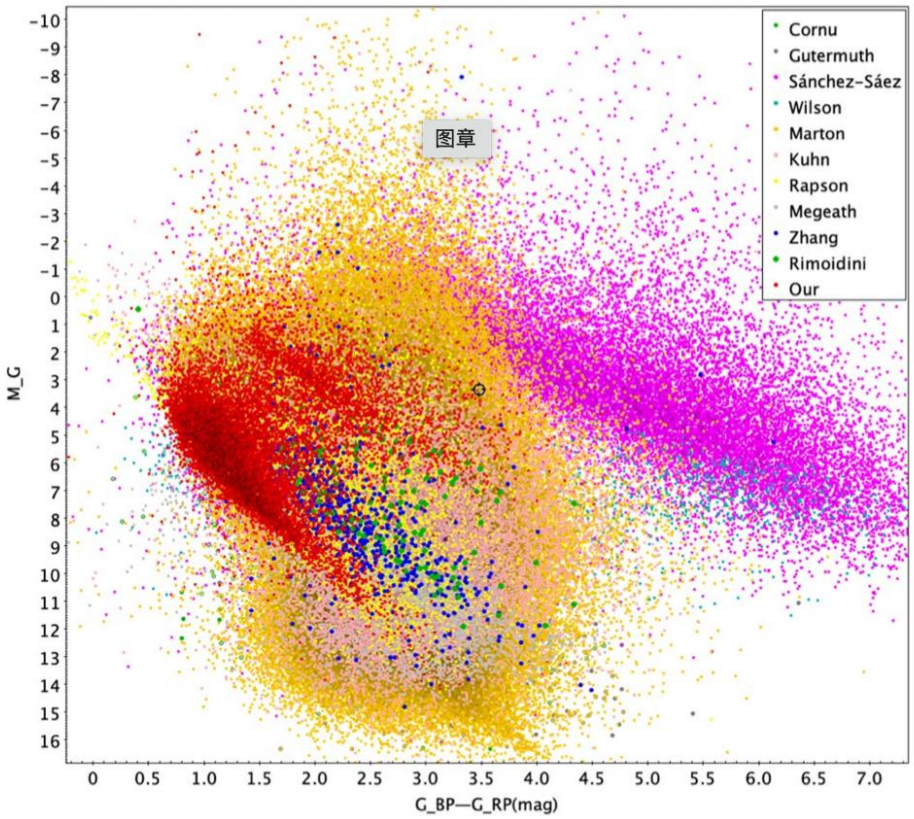
数据：LAMOST DR11 低分辨率。

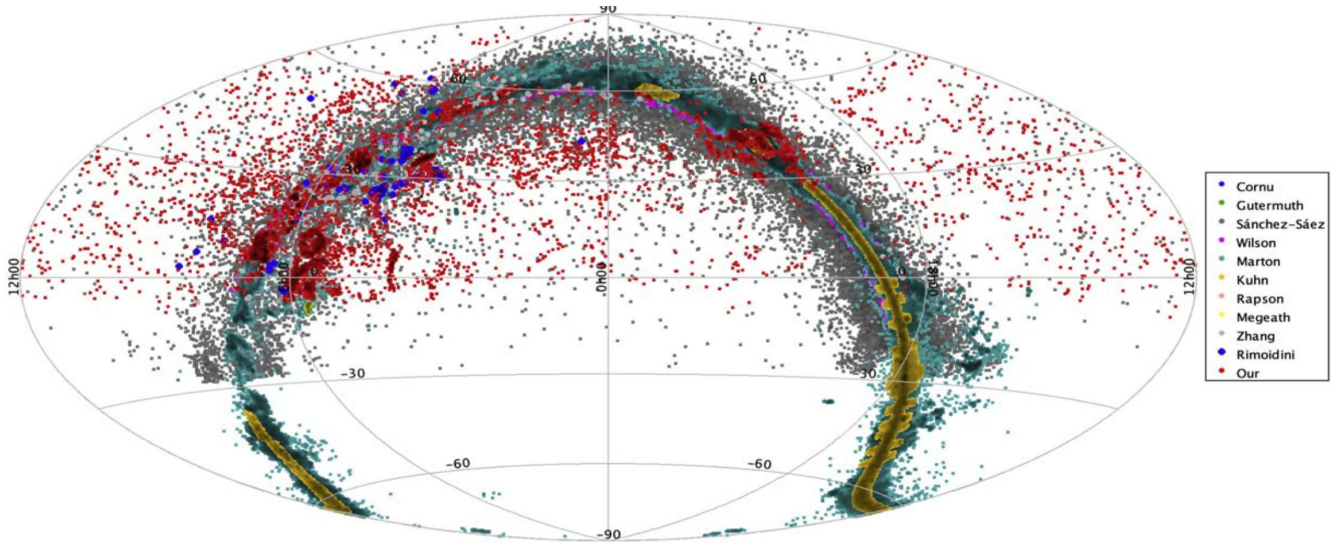
方法：基于 LSTM和CNN的混合模型。

结果：识别到了20530个YSO候选体。

Tan et al. APJS. 2024.







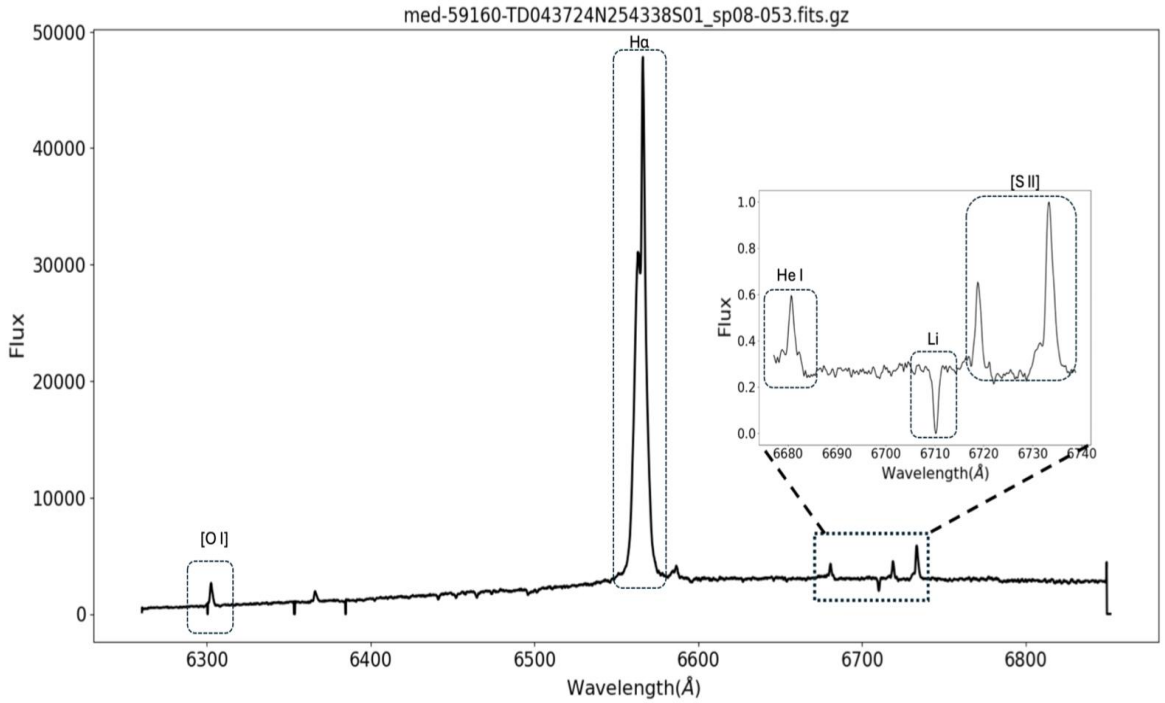
中分辨率YSO

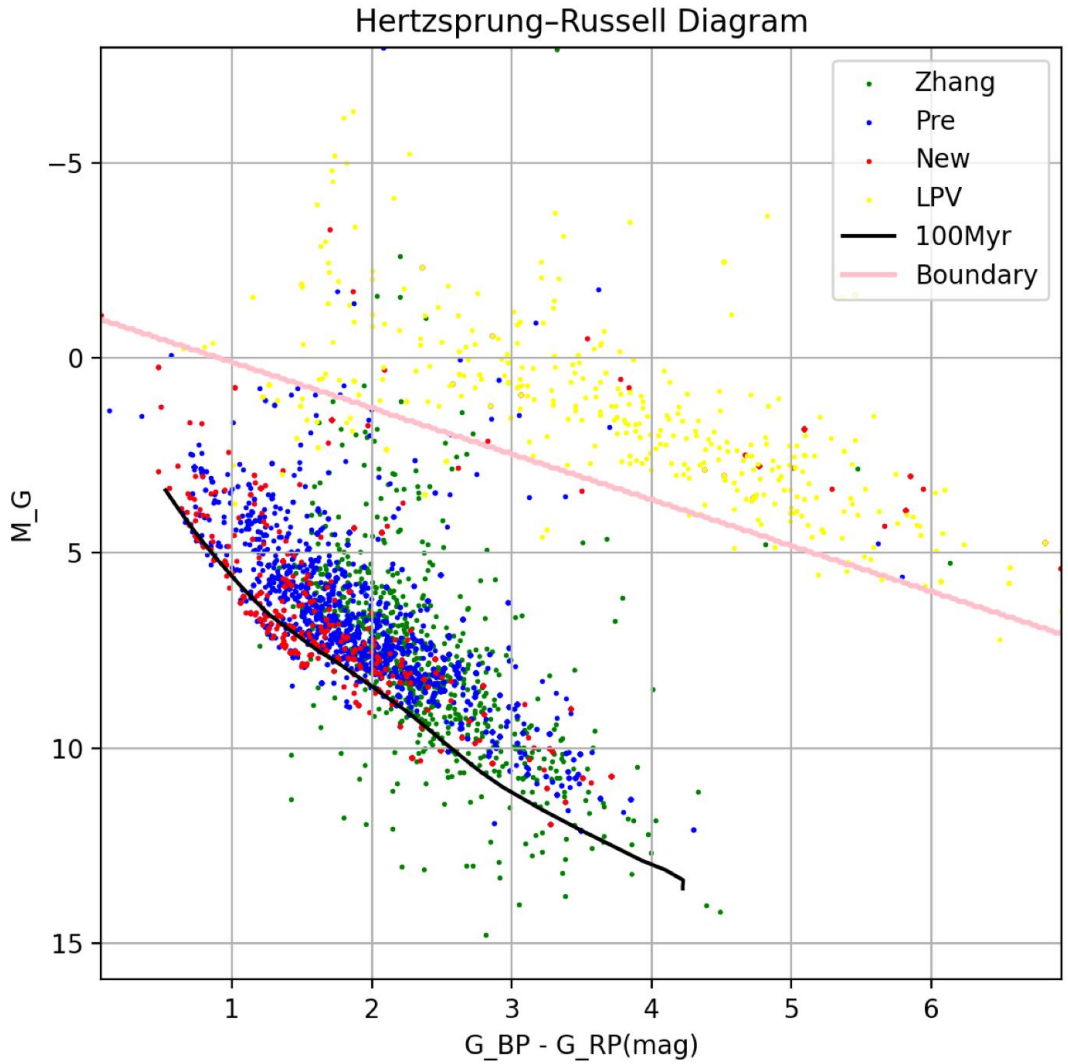
数据：LAMOST DR11 中分辨率。YSO截取Ha发射线的特征、结合Li吸收线。

方法：基于LSTM和CNN的混合模型。

结果：识别到了46867个新的Ha发射线星和785个YSO。

Tan et al. APJS. 2025.





结合Gaia数据和Parsec模型对数据进行证认

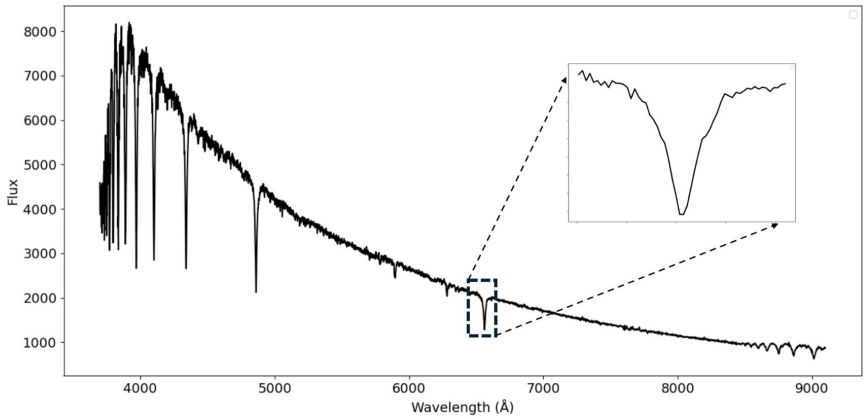
OB、Be

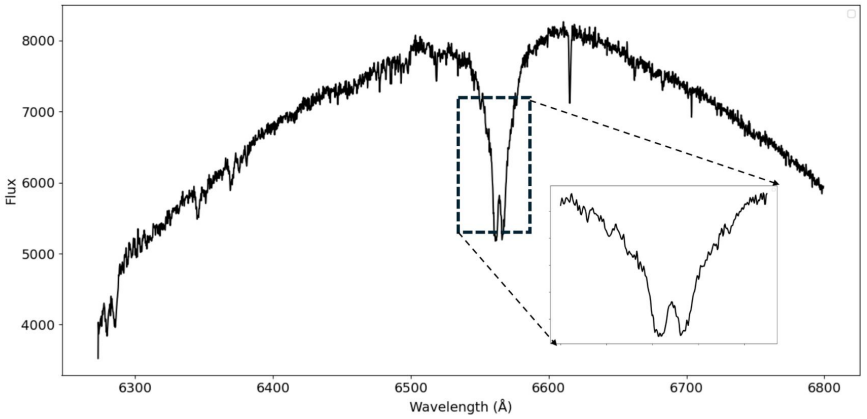
数据：LAMOST DR11 低分辨率。

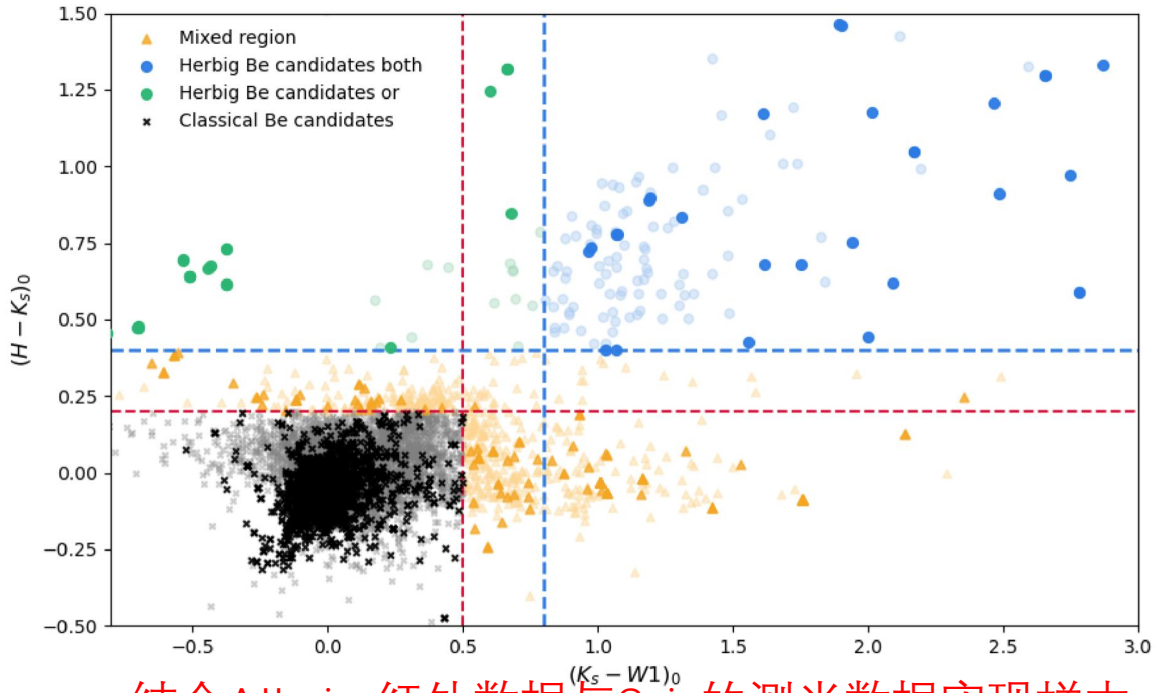
方法：基于Bi-LSTM、CNN和Transformer的混合模型。

结果：识别到了40223个B星和4512个Be星。

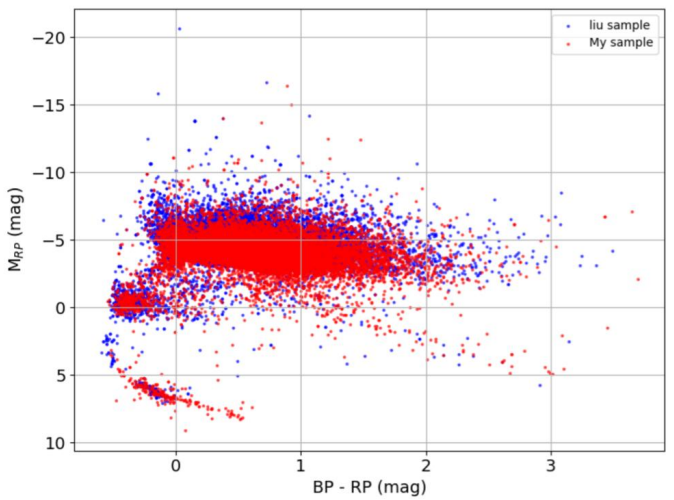
Tan et al. Submitted to APJS.



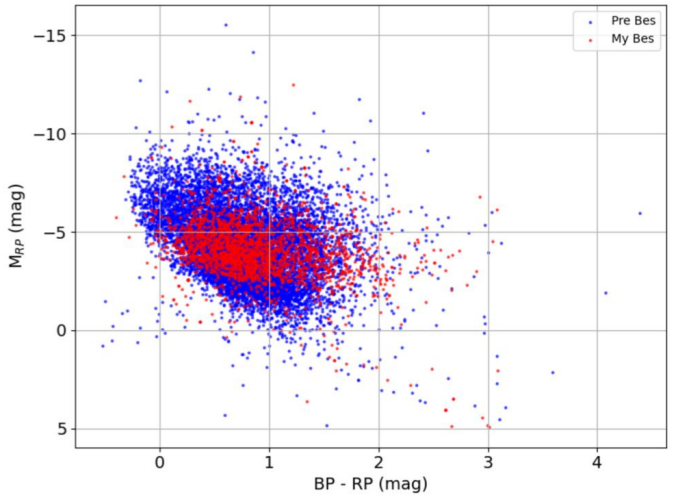




结合AllWise红外数据与Gaia的测光数据实现样本的证认和分类



(a)



(b)

03

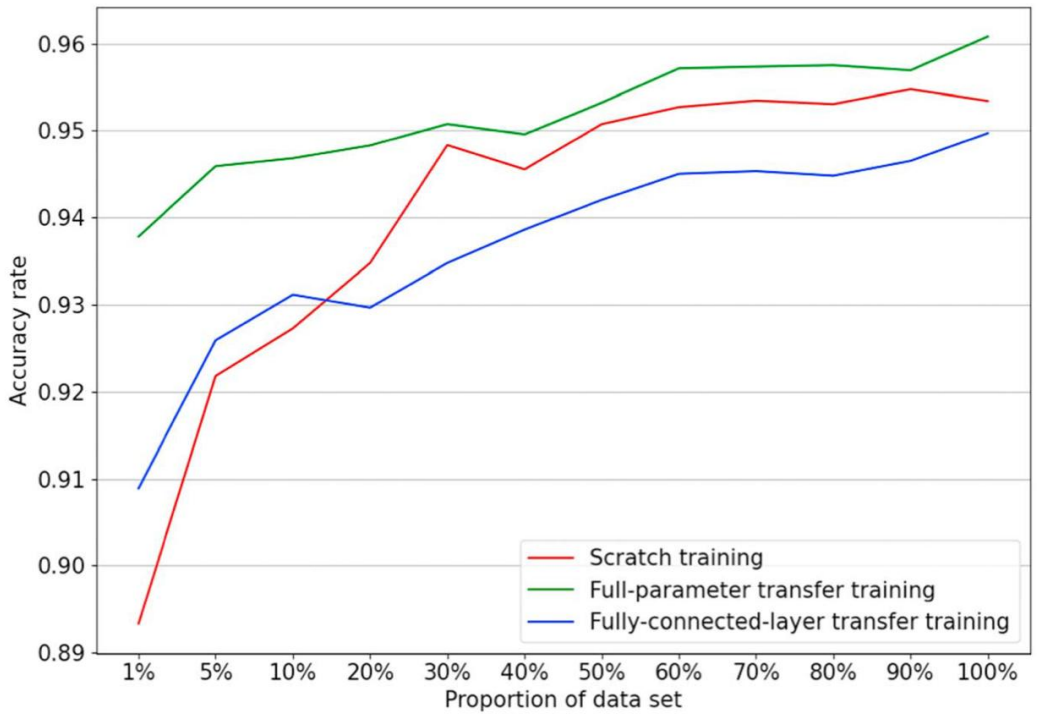
分析拓展

研究点：

- 基于不同望远镜数据，多波段、多模态数据融合
- 光谱、光变曲线、测光参数联合建模
- 面向CSST，建立先导模型

技术挑战：

- 数据异构性、复杂信噪比
- 稀有样本不均衡、标签不完备
- 宇宙线/缺失值，多模态数据质量问题
- 计算量与效率要求高



测试迁移学习的性能，为后续CSST数据进行模型储备

研究计划：

- 改进数据准备与模型搭建流程，形成系统全面的数据处理管线。
- 对主要的特殊星构建高正确率模型，实现自动化分类和参数分析。
- 多模态融合与分析，采用多波段、多模态数据对候选体星表进行证认和分析。
- 完善分类系统，在CSST仿真数据上进行测试。

预期成果：

- 数据处理管线，自动识别模型及流程工具包，构建起特殊恒星智能识别的技术体系。
- 高置信度候选体星表（白矮星、热亚矮星、YSO、Be等）。
- 面向CSST等未来巡天，具备重要支撑作用。
- 期待在星系结构、恒星演化等研究中挖掘新发现。

感 谢 您 的 聆 听

THANK YOU