

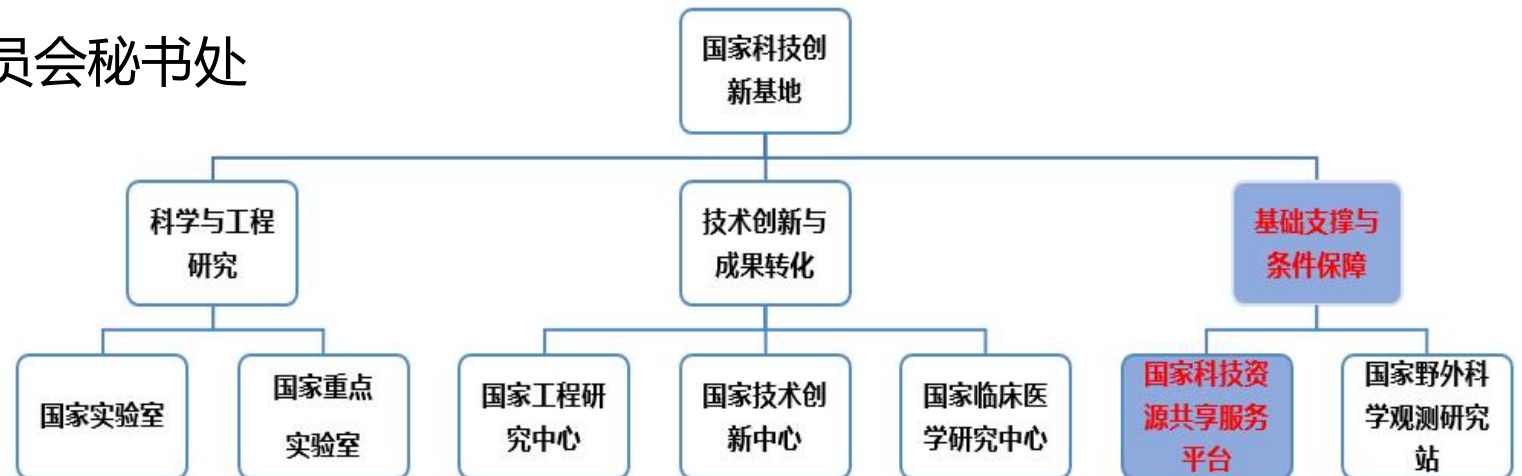
人工智能在天文研究中的应用与挑战

李长华

国家天文科学数据中心
中国科学院国家天文台

国家天文科学数据中心

- 我国首批“**国家科学数据中心**”之一，负责承担天文领域科学数据汇交管理、整编集成及开放共享的任务。
 - 国家科学数据中心始于1989年4月成立的“**世界数据中心中国中心天文学科中心**”。
 - 2011年12月，中国天文数据中心成为世界数据系统（WDS）首批正式成员。
 - 2018年10月，亚洲首个通过CoreTrustSeal认证的数据中心。
 - 全国天文学名词审定委员会秘书处



国家天文科学数据中心

■ Observation

- Telescope Proposal Submission and Management
 - 2.4m, 2.16m, Telescope Access Project (TAP), FAST

■ Data

- Data Query
- Data Archiving

■ Services

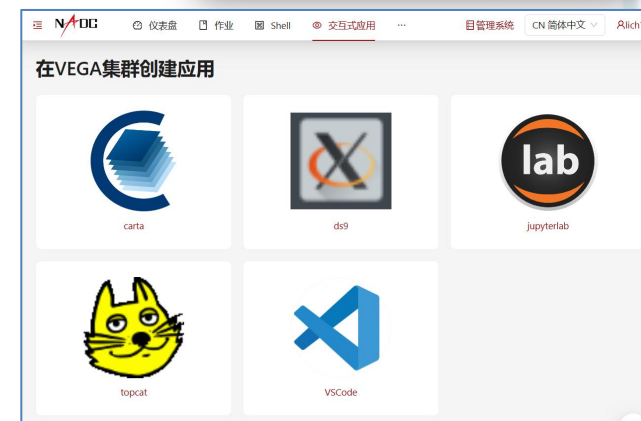
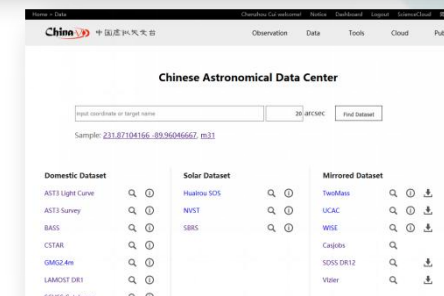
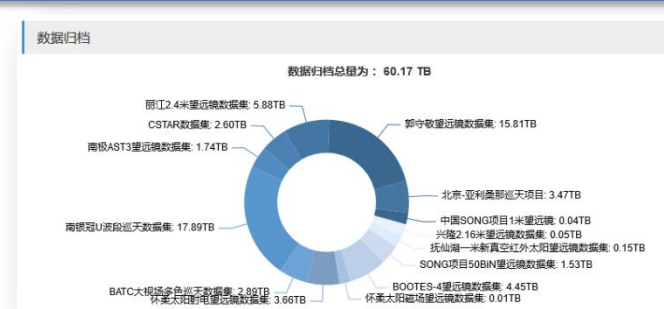
- Conferences, PaperData, Astro-term, VO tools

■ Science Platform

- VM and Containers
- data processing and analyzing environments online
- Cloud storage

■ Public channel

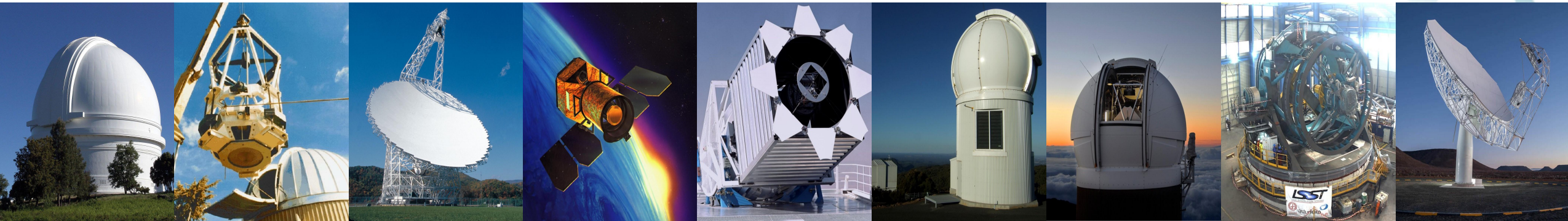
- Public Supernova search Project (PSP)
- Worldwide Telescope
- Eclipse calculator
- Citizen Science



<https://nadc.china-vo.org/>

天文学：数据驱动的科学

■ 数据规模巨大，并快速增长



DPOSS
3TB

2MASS
10TB

GBT
20PB

GALEX
30TB

SDSS
40TB

SkyMapper
500TB

PanSTARRS
1.6PB

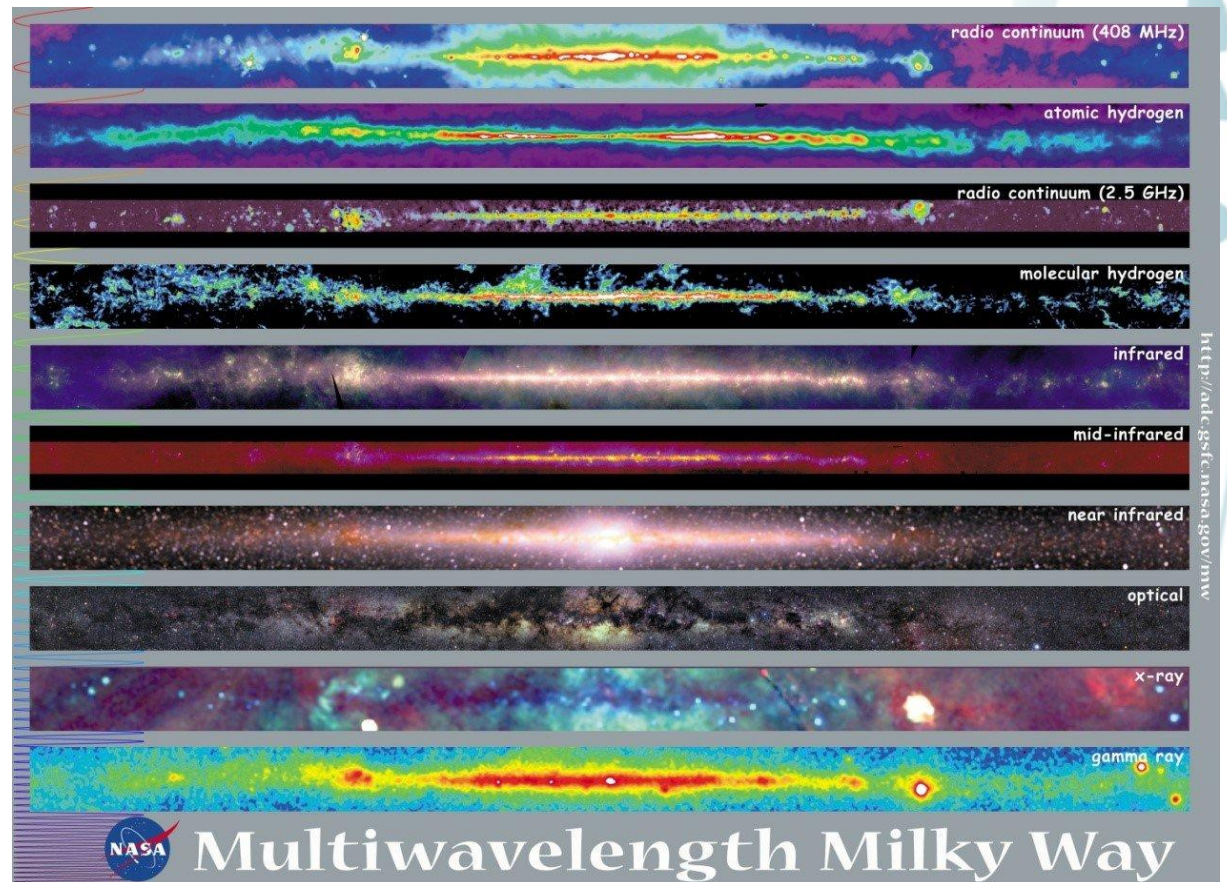
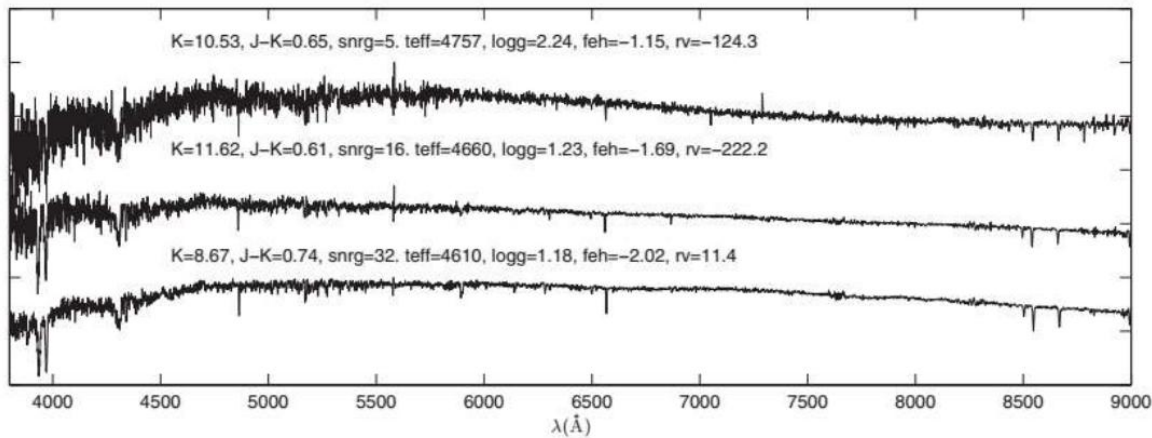
LSST
~200PB

SKA
~4EB



天文学：数据驱动的科学

■ 数据形态多样，多波段特征显著



天文学：数据驱动的科学

数据发现：搜索超新星、搜索引力透镜 ...

NADC National Astronomical Data Center
国家天文科学数据中心

申请观测 科学数据 专题服务 云资源 公众频道

欢迎加入 公众超新星搜寻项目

开始您的超新星搜寻之旅

项目简介 PSP已知目标 旧图试看 开始搜寻

15天内 提交/看图数 排名		24小时看图能手		搜寻大神	
严征宇	0.03%	毕春朋	704 张	毕春朋	793880 张
马士卓	0.04%	麦梓阳	562 张	葛迎宾1975	690820 张
麦梓阳	0.08%	张翔涛	506 张	徐建林	659201 张
侯思远	0.12%	区耀樟	324 张	高伟	463508 张
徐超	0.13%	顾少雄-福建泉州	290 张	王晓兵-湖北襄阳	447389 张
张翔涛	0.17%	孔祥慧	242 张	阮建高	445528 张

LENSFINDER

首页 教程 寻找 成果 关于 个人中心

欢迎来到引力透镜搜寻

开始寻找引力透镜
Find Gravitational Lens

了解更多 操作指南 意见反馈

什么是引力透镜

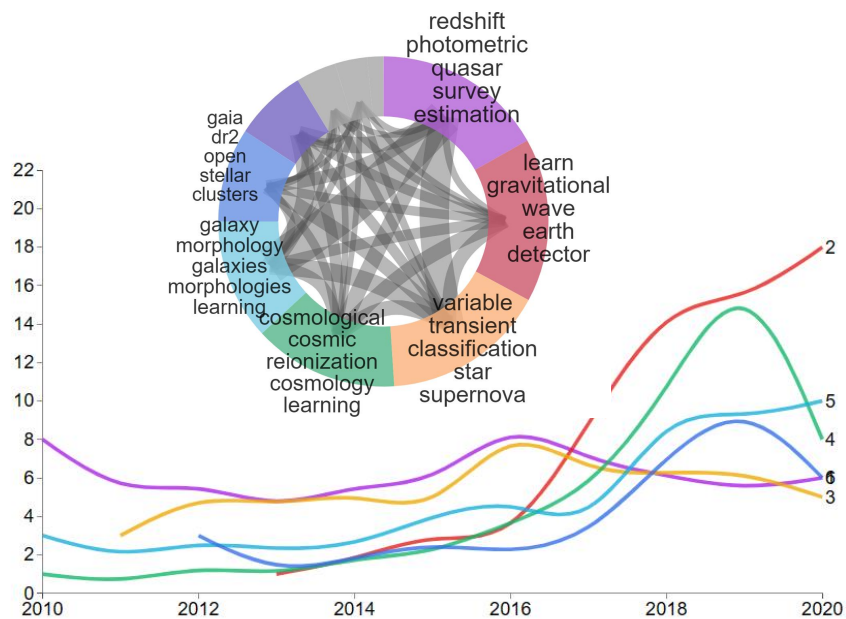
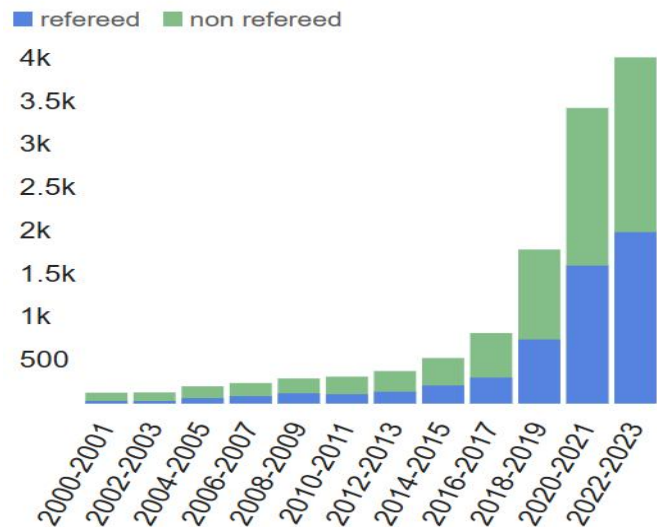
简单地说，引力透镜效应是指光线围绕着由大质量星系产生的引力发生弯曲。这通常会导致看到多个相同的光图像，或高度扭曲的条图像(完美的，将会有个爱因斯坦环)。

查看更多



天文学：数据驱动的科学

- 在论文标题、摘要和关键词中出现 **machine learning / deep learning / deep neural network / convolutional neural network** 的论文数量 (2000年-2023年) (来源于ADS, <https://ui.adsabs.harvard.edu/>)
 - 机器学习与深度学习在天文学中的应用在2014-2015年之前一直稳步增加, 在此之后, 迎来大爆发。



典型应用：天体识别与分类

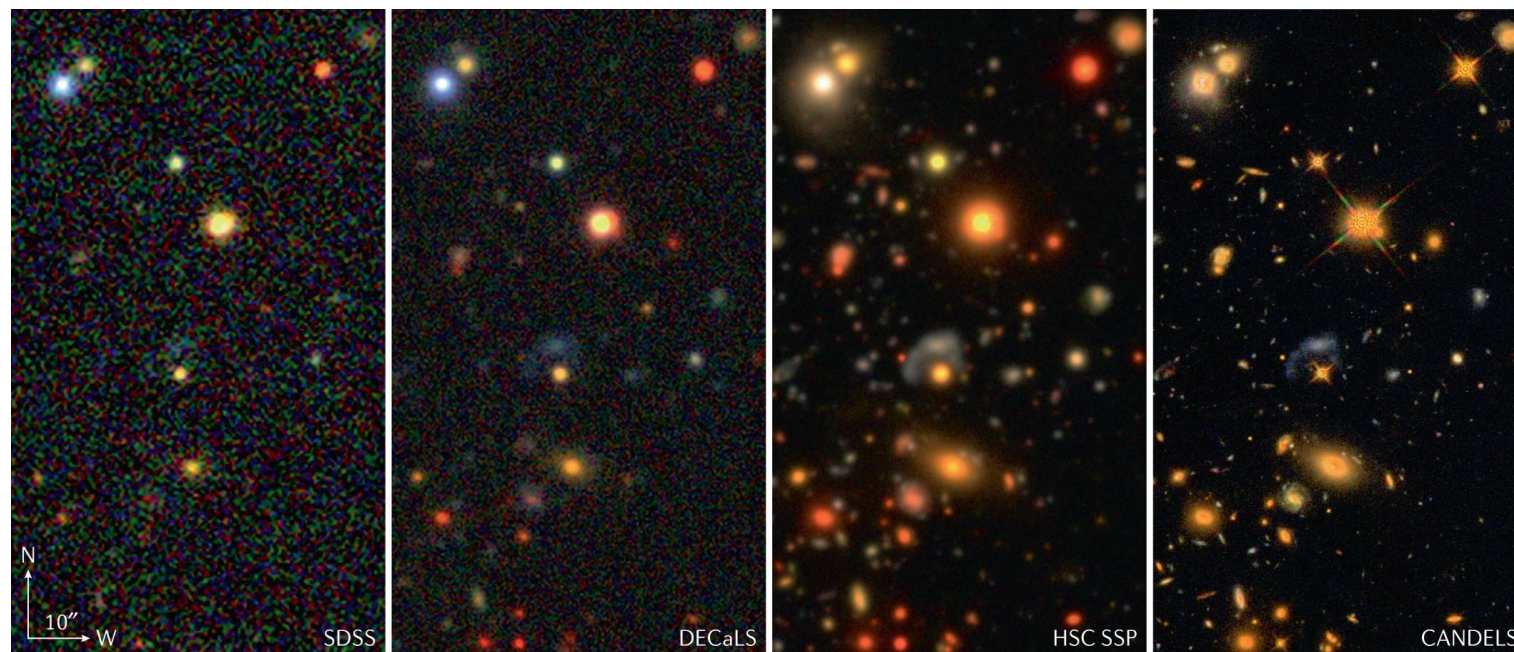
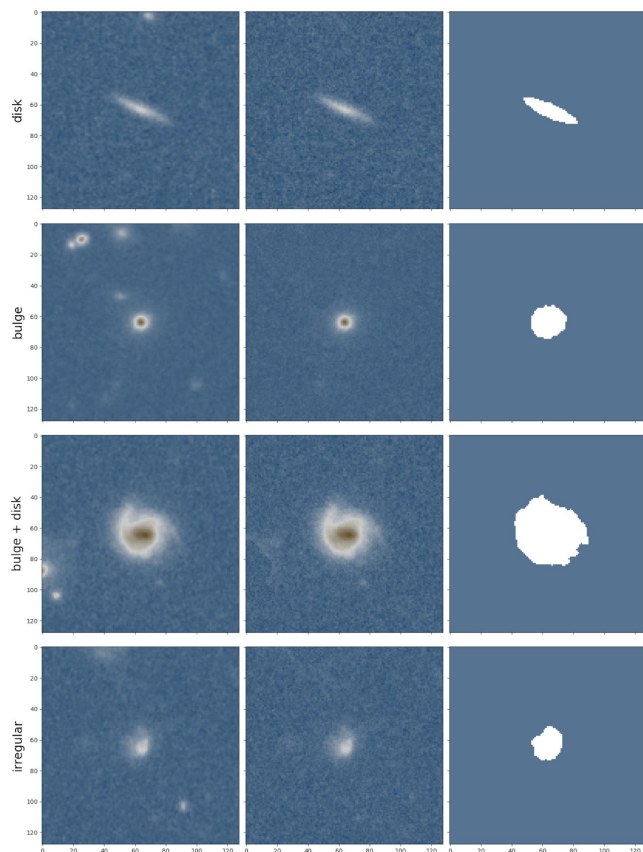


China-VO

典型应用：天体识别与分类

源检测及解混合

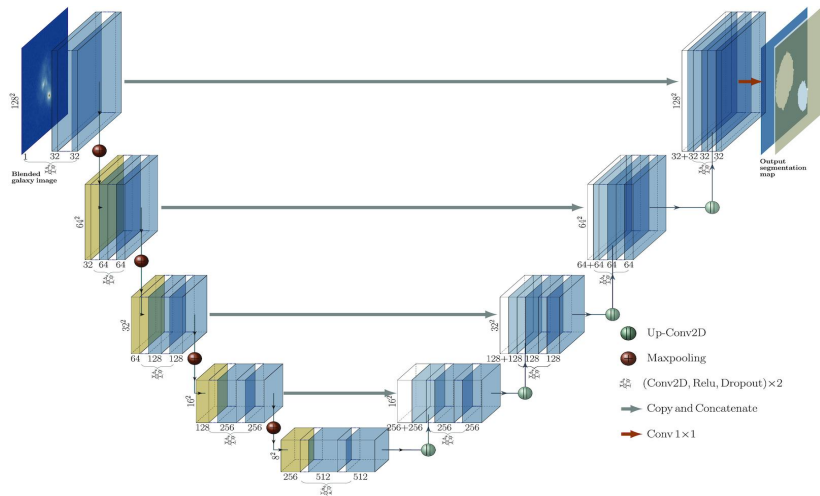
SExtractor



随着望远镜性能提升，源的密度不断增加，混淆问题变得更加突出，大量的星系受到重叠或混合的影响。

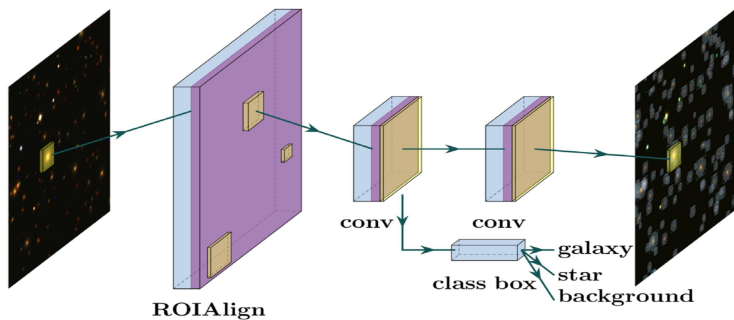
典型应用：天体识别与分类

U-Net架构引入目标检测 > SExtractor

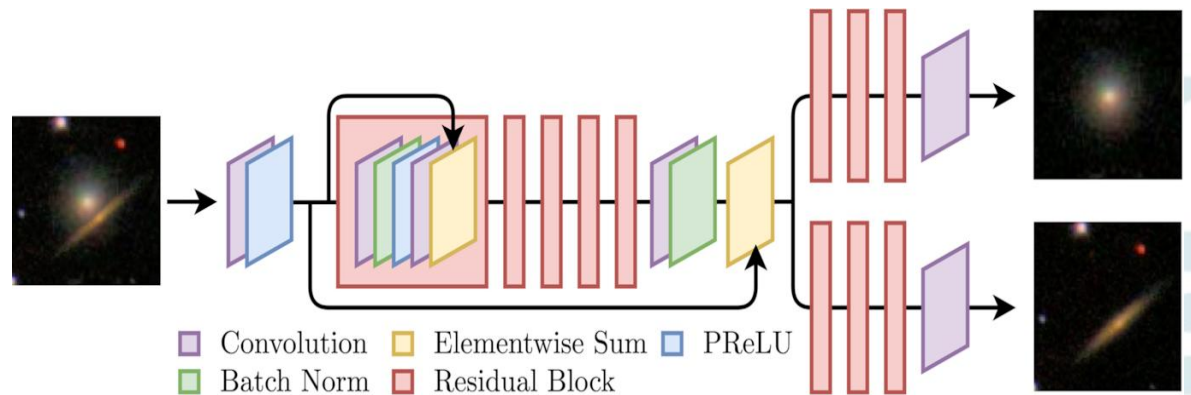


(Boucaud et al. 2020)

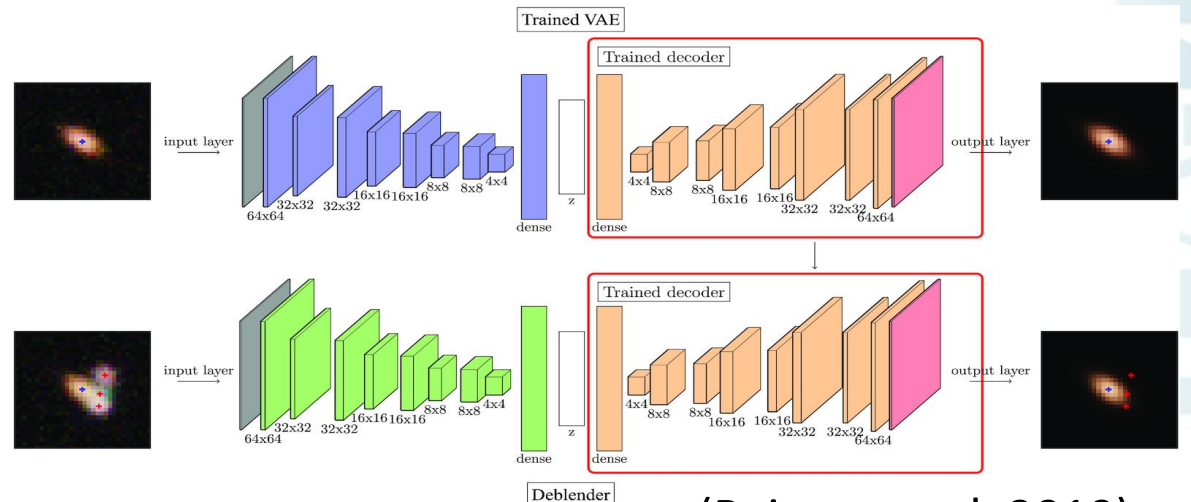
Mask-RCNN框架



(Burke et al. 2019)

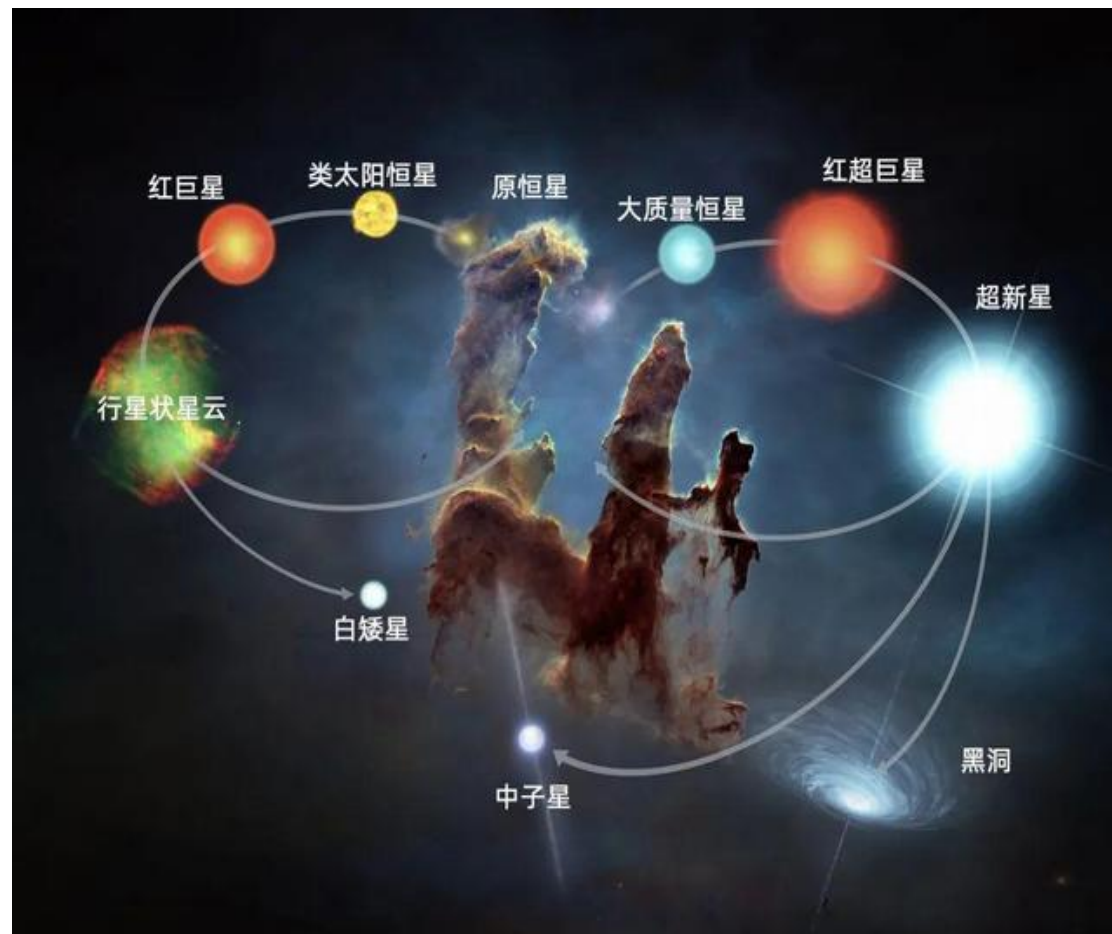
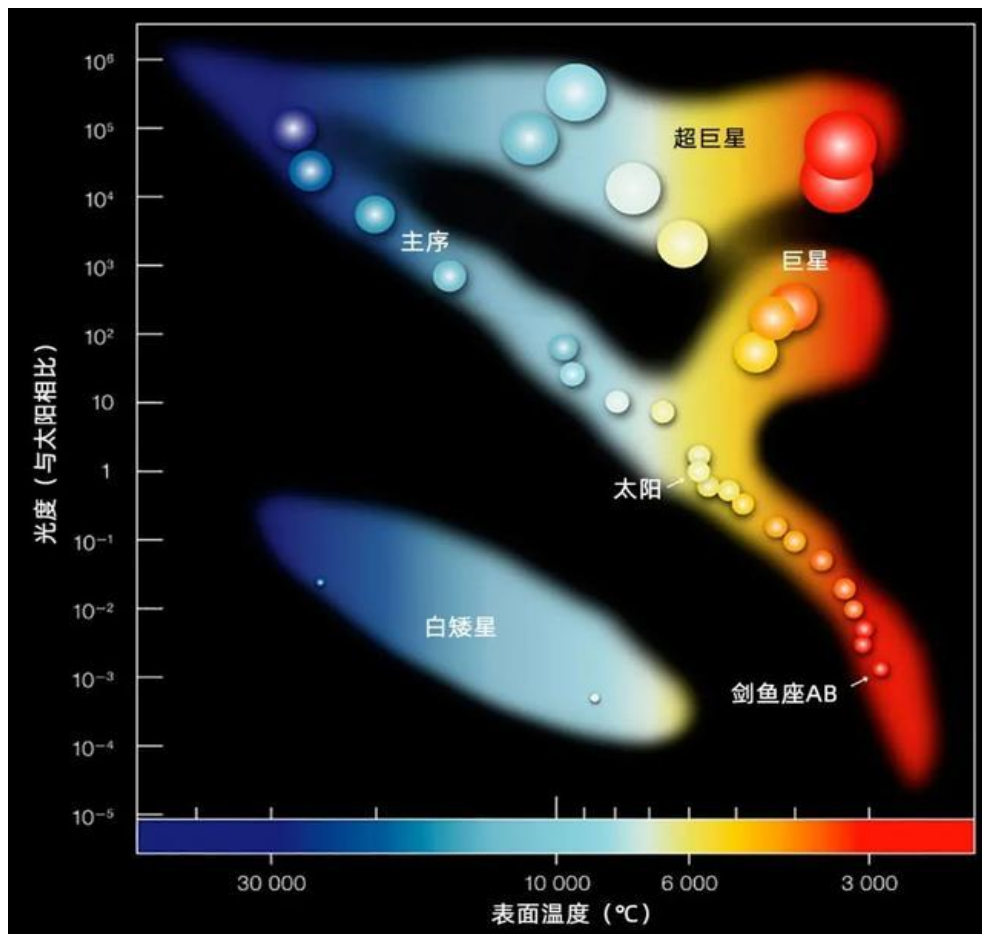


(Arcelin et al. 2021)

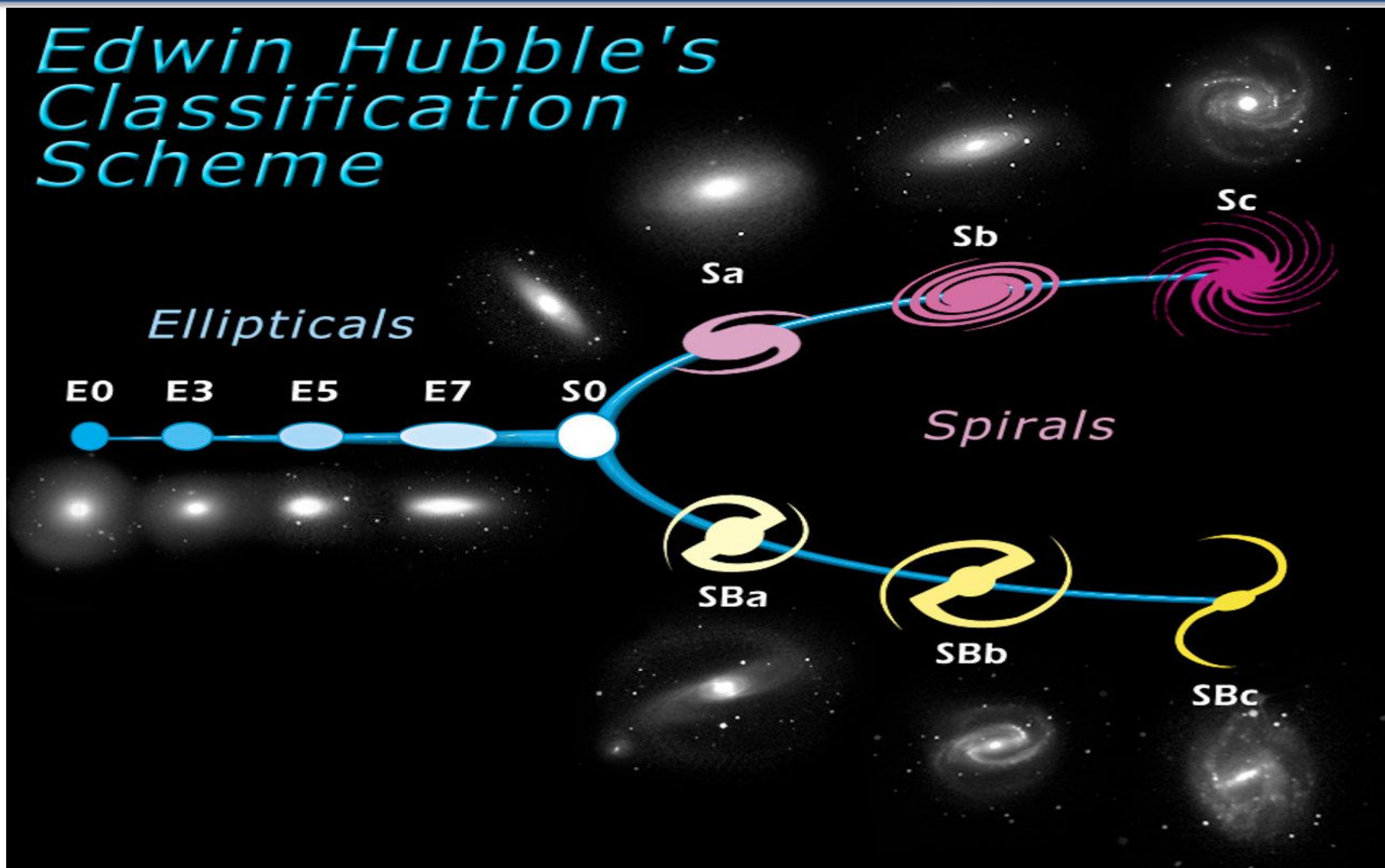


(Reiman et al. 2019)

典型应用：天体识别与分类



典型应用: 天体识别与分类



China-VO

典型应用：天体识别与分类

■ 类星体搜索

- BASS DR3 (~30 millions)
- SDSS+LAMOST
- XGBoost、RF



	Point Source		Extended Sources	
Accuracy	Precision	Recall	Precision	Recall
98.67%	98.39%	98.35	98.86%	98.89%

	Star		Quasar	
Accuracy	Precision	Recall	Precision	Recall
99.15%	99.00%	99.46%	99.33%	98.77%

Classifier	objective	Input pattern	Class
Classifier 1 st	binary:logistic	$(\Delta g, \Delta r, \Delta z, g - r, g - z, r, r - z, g)$	(point and extended sources)
Classifier 2 nd	binary:logistic	$(\Delta g, \Delta r, \Delta z, g - r, r - z, g - z, r, g, z)$	(stars and quasars)
Classifier 3 rd	multi:softmax	$(\Delta g, \Delta r, \Delta z, g - r, g - z, r, r - z, z)$	(galaxies, stars and quasars)
Classifier 4 th	binary:logistic	(Pattern I)	(point and extended sources)
Classifier 5 th	binary:logistic	(Pattern II)	(stars and quasars)
Classifier 6 th	multi:softmax	(Pattern III)	(galaxies, stars and quasars)

^a Extended sources represent galaxies while point sources for stars and quasars.

^b Pattern I represents $\Delta g, \Delta z, g - W1, W1 - W2, z - W1, \Delta r, g - z, z - W2, g - r, r - z, W1, r, g, z, r - W2$.

^c Pattern II represents $z - W2, W1 - W2, g - z, g - r, z - W1, r - z, \Delta z, r, r - W2, z, \Delta g, g - W1, \Delta r, W1, g - W2$.

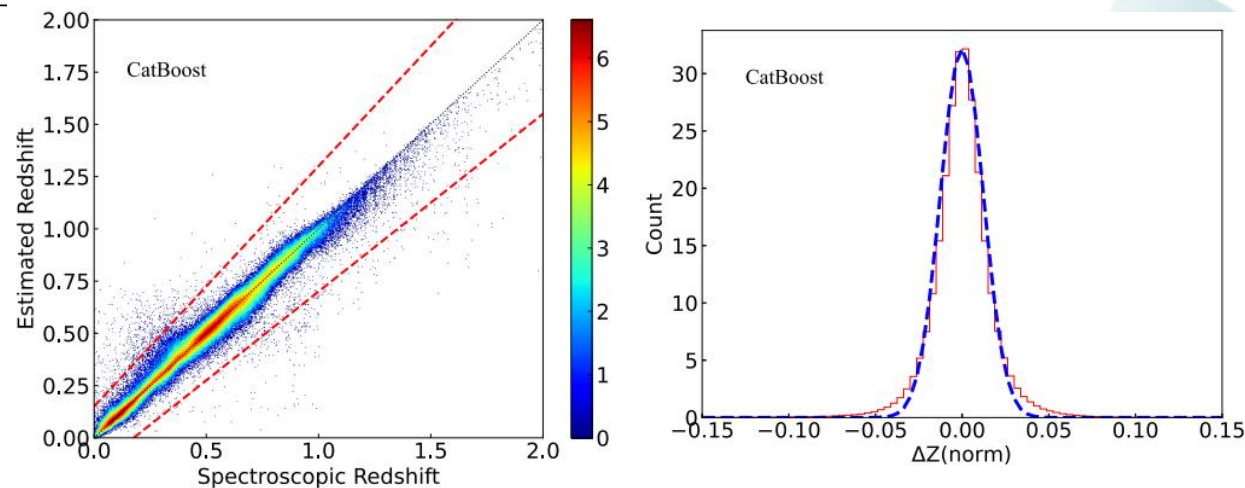
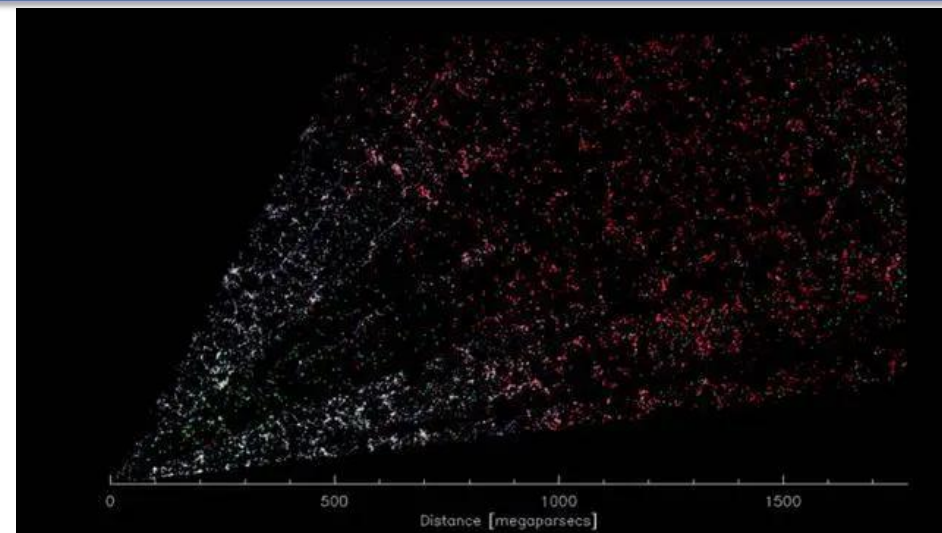
^d Pattern III represents $z - W2, \Delta z, W1 - W2, \Delta r, g - r, z - W1, \Delta g, g - z, r - W2, r - z, r$.

典型应用：天体识别与分类

■ 测光红移

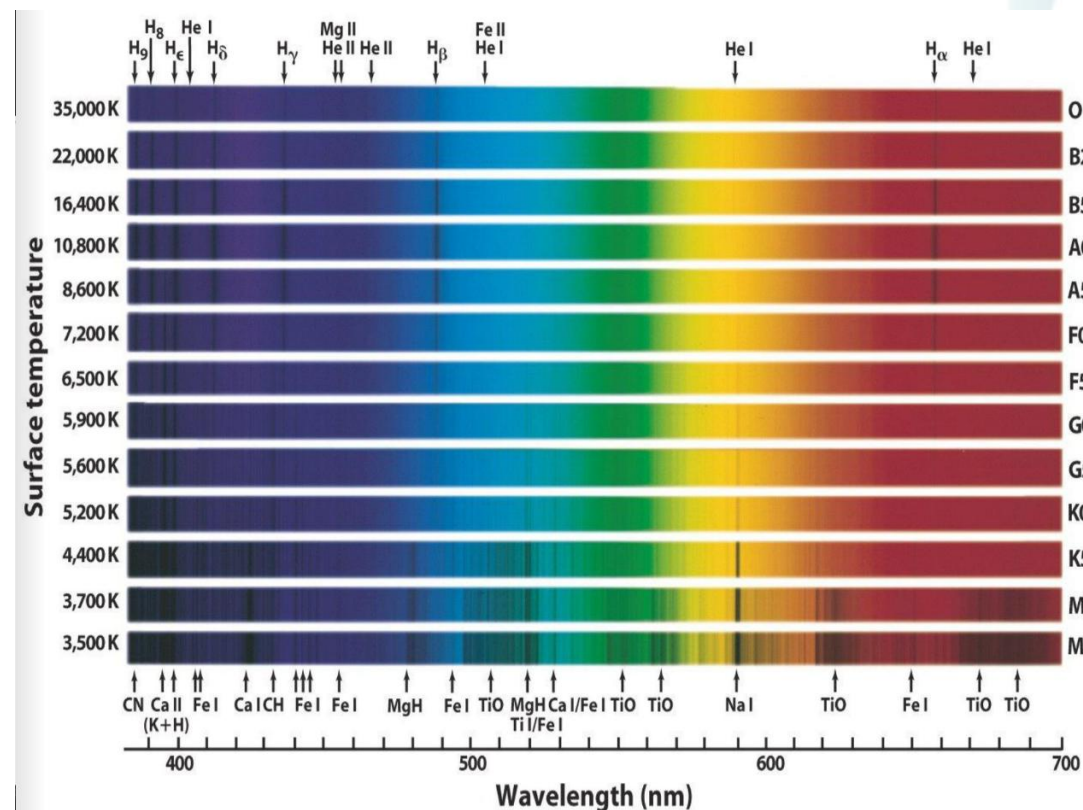
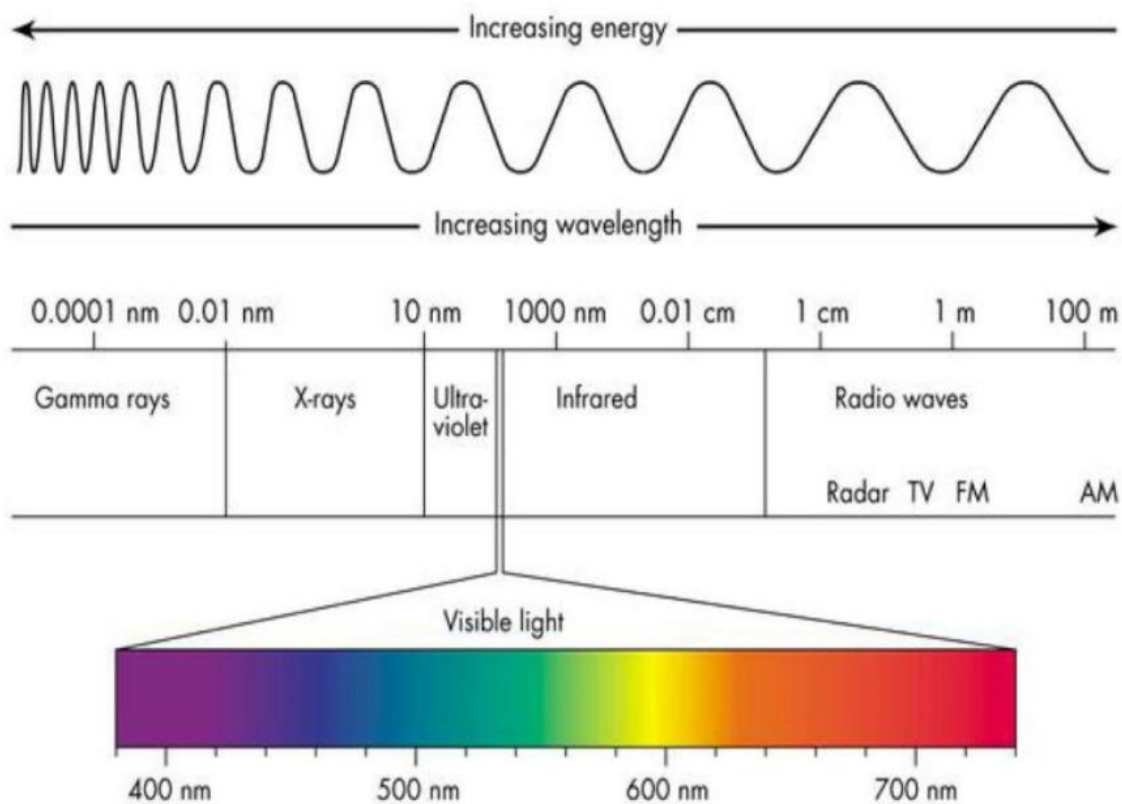
- 红移定义：
$$z = \frac{\lambda - \lambda_0}{\lambda_0}$$
- 光谱观测（昂贵，能力有限）
- 测光观测（无法直接测量准确值）
 - 模版匹配
 - 机器学习

很难构建完备样本



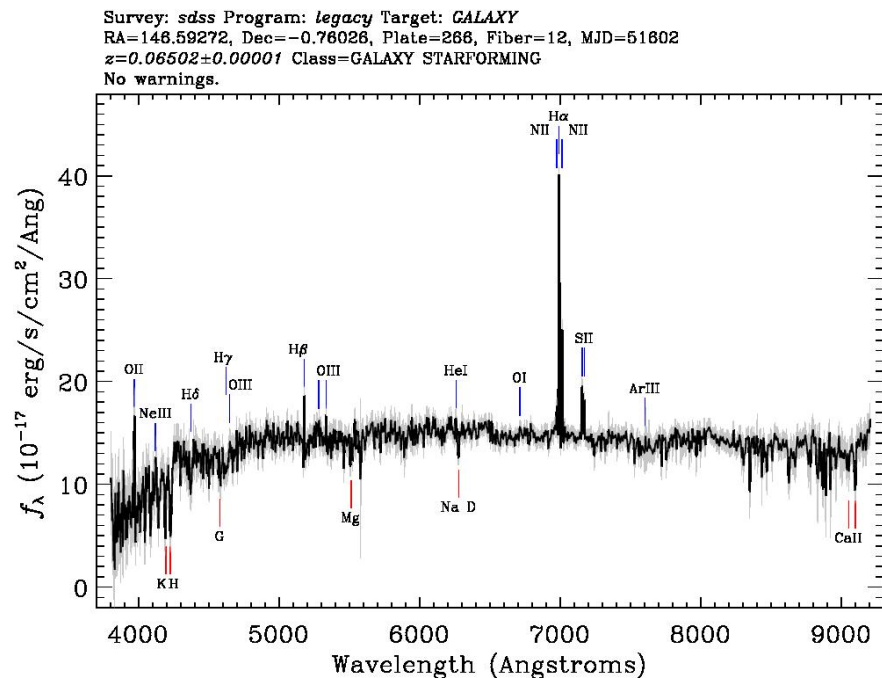
典型应用：天体光谱分析

- 天体光谱蕴含丰富的物理信息，能够确定天体的物理性质和化学组成



典型应用：天体光谱分析

■ 光谱分类



传统方法：

- 依赖于精确的发射线测量
- 不能处理高红移和低SNR的光谱
- 需要人工处理、难以适应大规模巡天数据

机器学习方法：

- 特征筛选依赖专家知识
- 准确率需进一步提升



典型应用：天体光谱分析

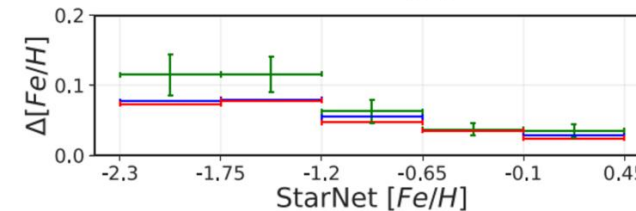
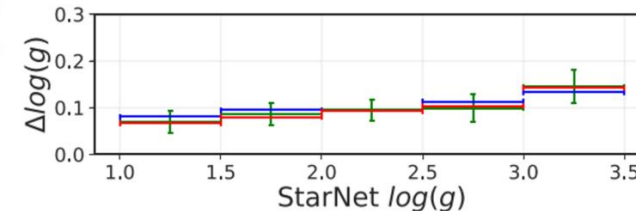
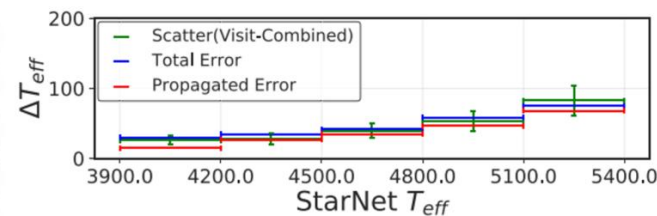
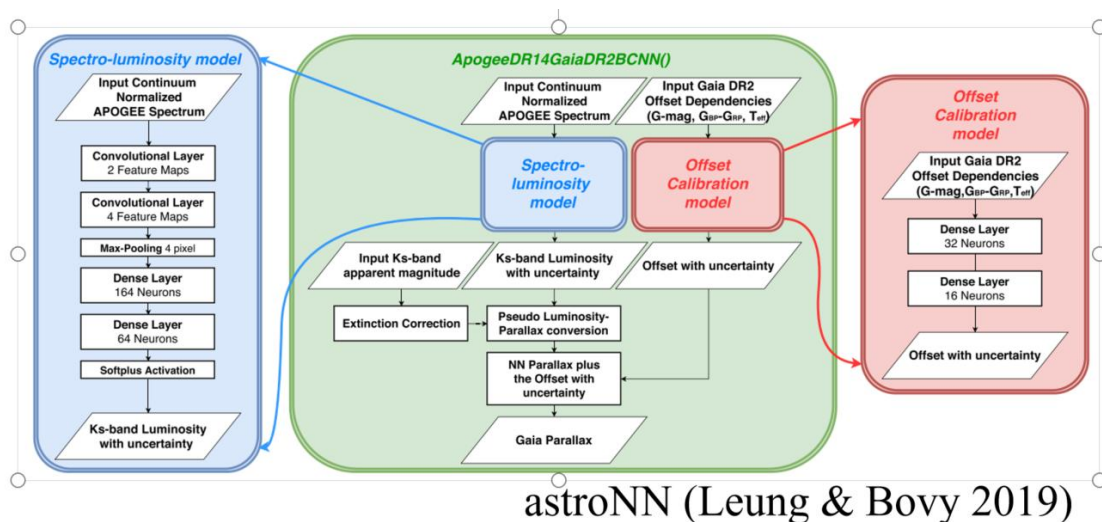
■ 恒星参数估计：有效温度 T_{eff} ，表面重力 $\log g$ ，金属丰度、自转速度……

巴尔默线轮廓拟合

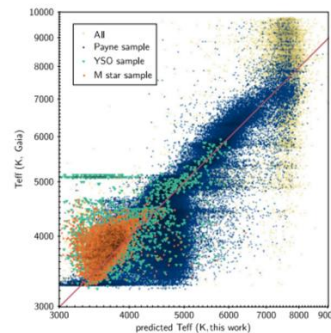
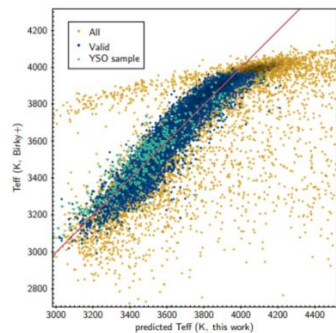
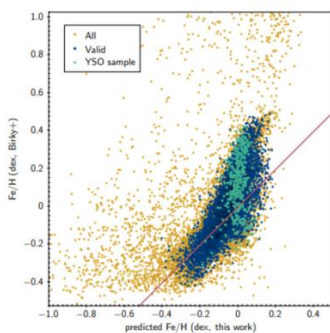
谱线比例方法

线指数方法

模板拟合方法



StarNet (Fabbro et al. 2018)



典型应用：天体光谱分析

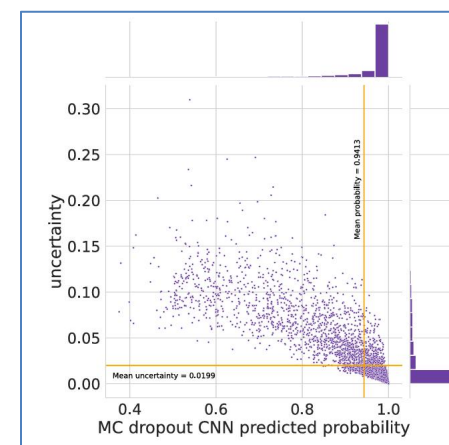
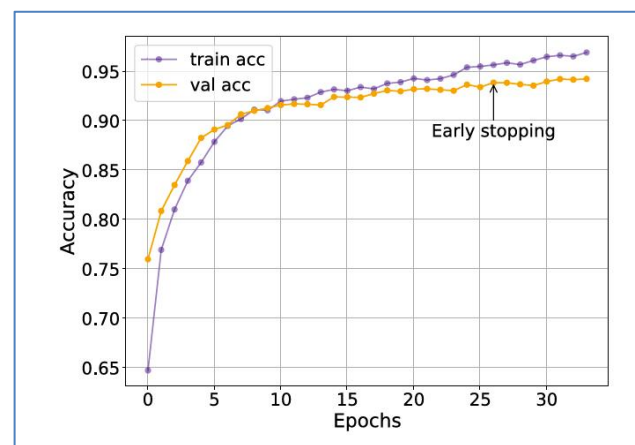
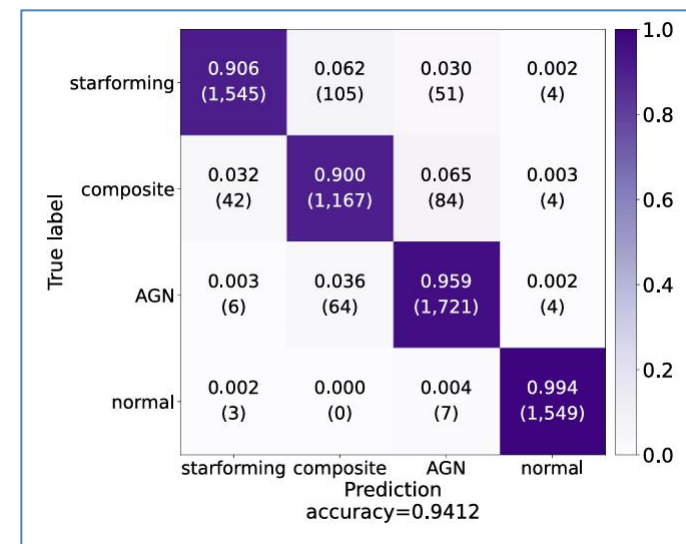
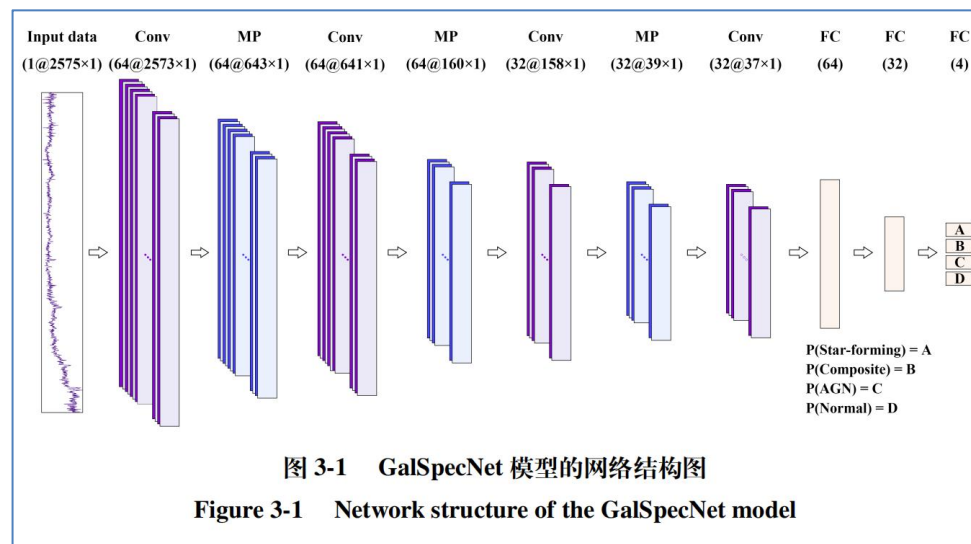
■ 星系光谱分类

■ Star-forming

■ composite

■ AGN

■ Normal

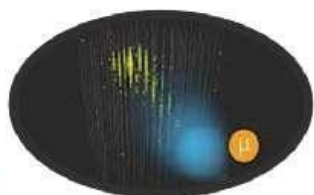


典型应用：时域天文学

Transients & Variability



High redshift GRB
Early universe

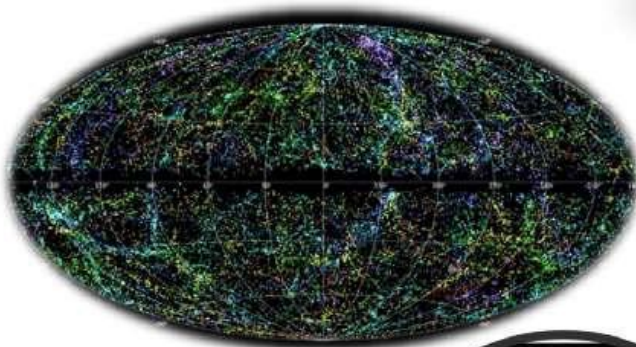


X-ray flares associated
with neutrino events
particle acceleration



Quasi-periodic eruption
EMRI (GW source?)
BH accretion?

SN shock breakout
SN physics
progenitors



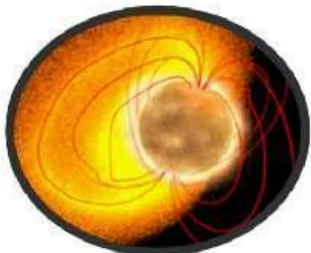
Star X-ray flares
Magnetic fields
Corona activity



Active galactic nuclei
Extreme gravity, BH
accretion/jets/growth



Intermediate-mass BH
BH physics/accretion



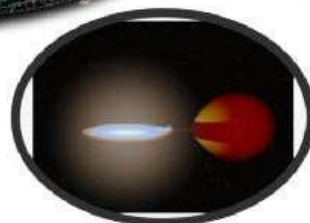
magnetar
extreme
magnetic field



Fast Radio Burst
X-ray sources



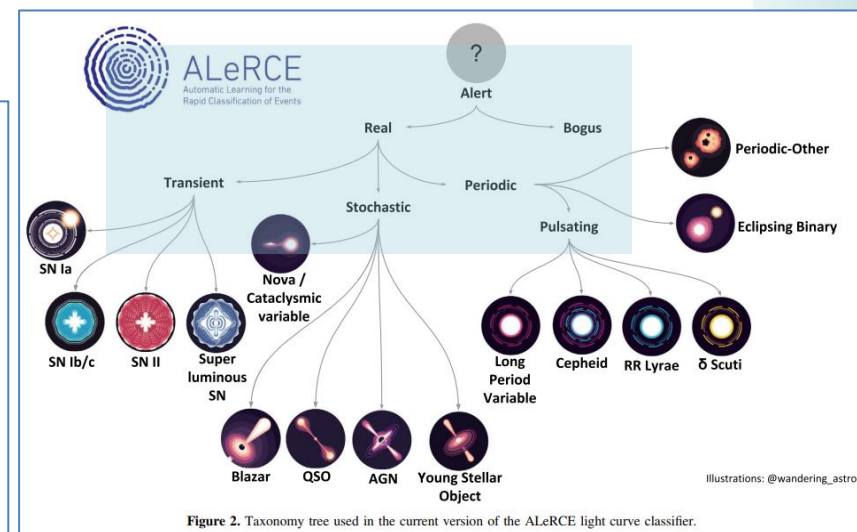
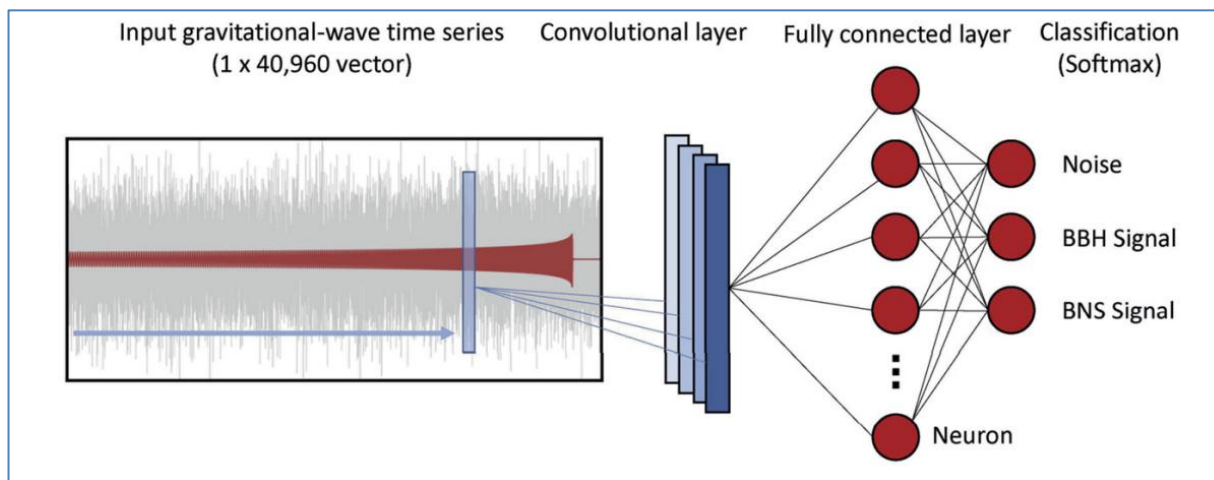
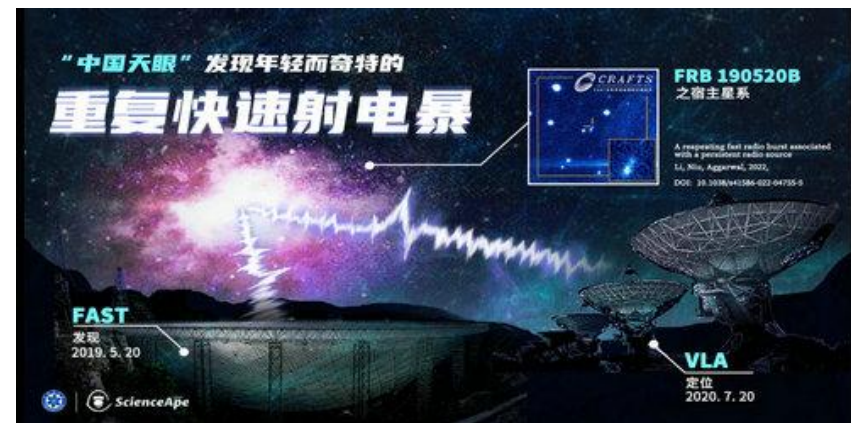
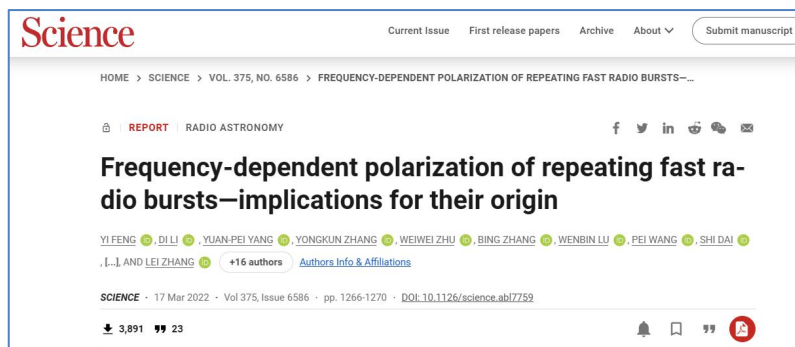
Thermal nuclear burst
Neutron stars physics



BH X-ray binary
Extreme gravity
BH physics/accretion

典型应用：时域天文学

- 快速射电暴
- 引力波探测
- 脉冲星探测
- 变源/暂现源证认及分类
- ...



Xiaoxiong Zuo,^{1,2,3} Yihan Tao,^{1,3} Yuan Liu,¹ Chenzhou Cui,^{1,3} Yunfei Xu,^{1,3} Wenda Zhang,¹ Haiwu Pan,¹ Zhen Zhang^{1,2}

1. National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100101, China

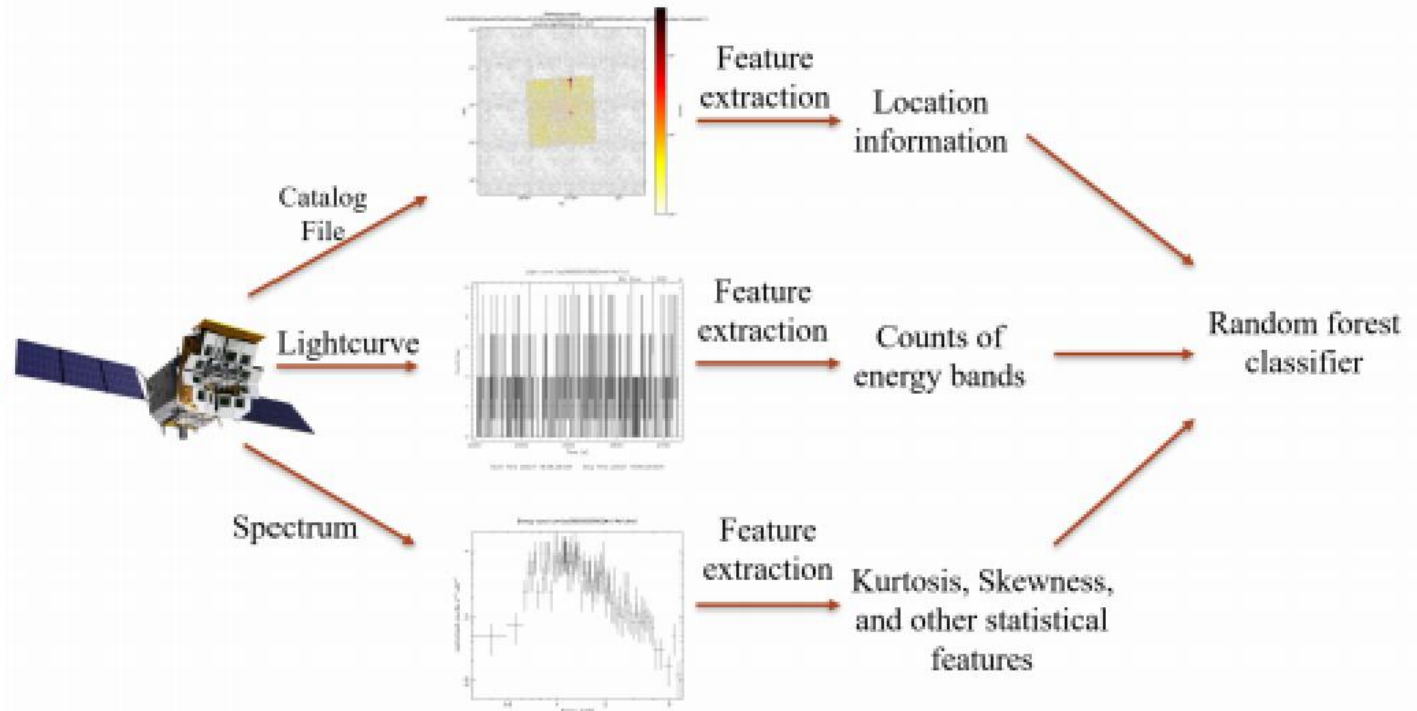
2. University of Chinese Academy of Sciences, Beijing 100049, China

3. National Astronomical Data Center, 20A Datun Road, Beijing 100101, China

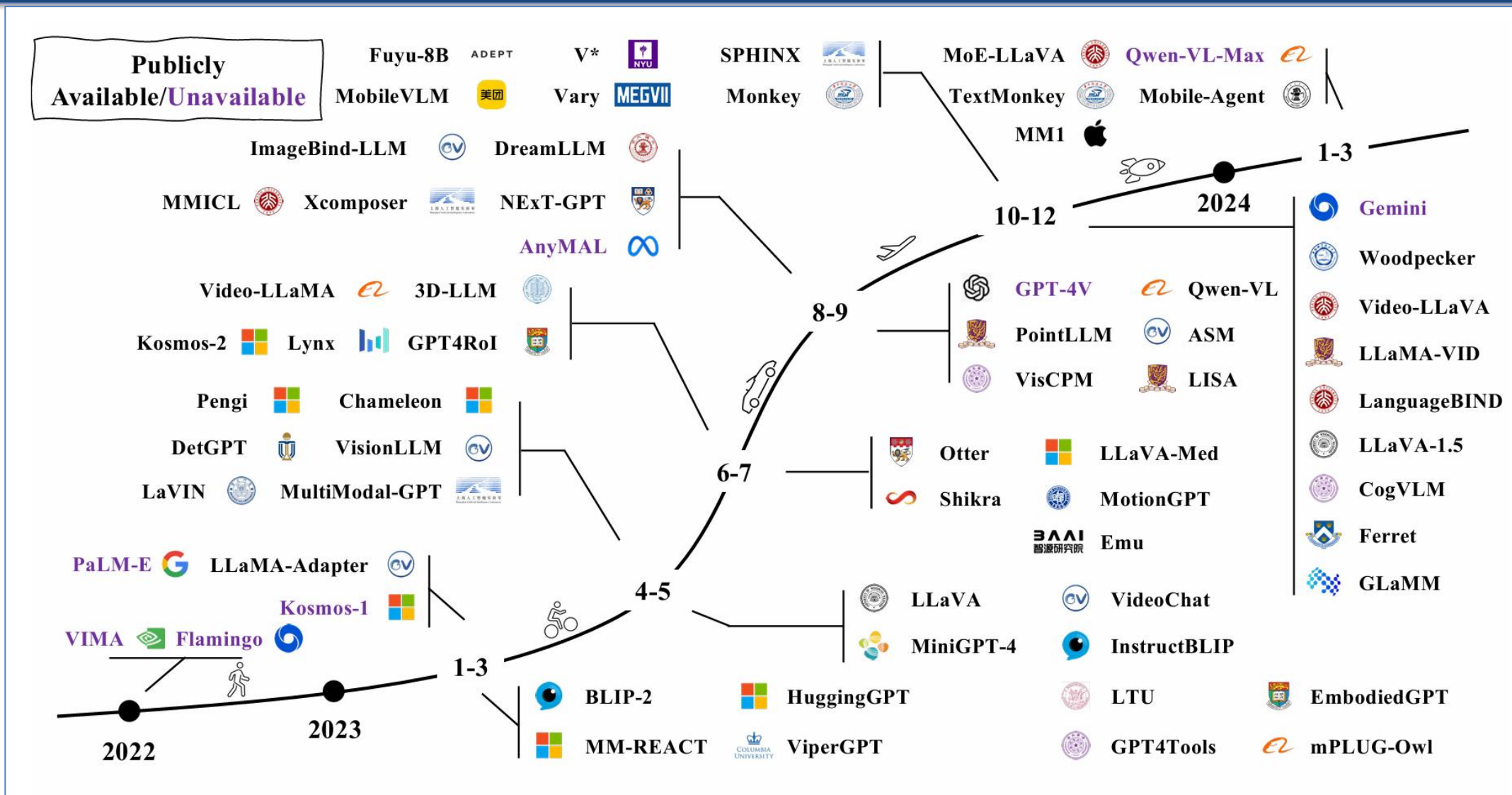
First author: zuoxx@bao.ac.cn Corresponding author: y.tao@nao.cas.cn

Abstract

The Einstein Probe (EP) is an astronomical satellite operating in the soft X-ray band, designed for the frontier of time-domain astronomy and high-energy astrophysics. In this study, we designed and implemented a machine learning model for the classification of observed sources. The model was trained using data acquired from the EP-WXT Pathfinder Lobster Eye Imager for Astronomy (LEIA) and the EP simulation data, which includes lightcurves, spectra, and location information. Statistical features extracted from the lightcurves and energy distributions within the spectra were used to build a Random Forest classifier. The final model proposed in this study achieved an accuracy of 98% on EP simulation data and 97.3% on LEIA data.



典型应用：天文大模型

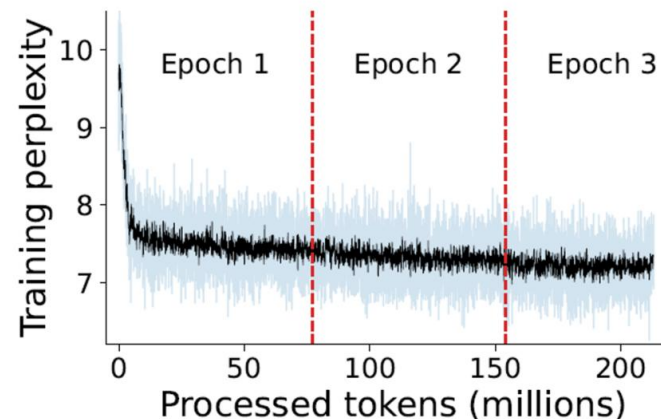


典型应用：天文大模型

AstroLLaMa:

- 为了增强通用大语言模型在天文领域的表现，丁源森团队训练了 AstroLLaMa;
- 基于LLaMA-2-70B (70亿参数)，该模型是在30万篇天文学论文摘要微调而来;
- 与通用模型相比，困惑度降低30%，具有强大的天文领域理解能力;

为天文问答、摘要生成、假说生成等语言、文本应用奠定基础



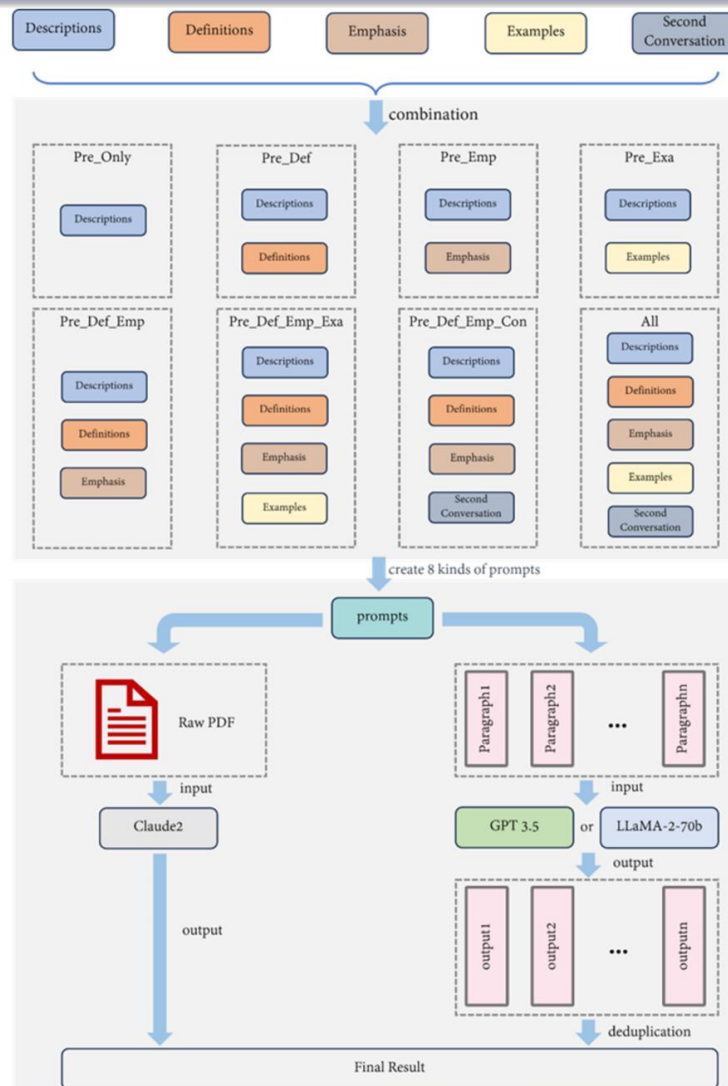
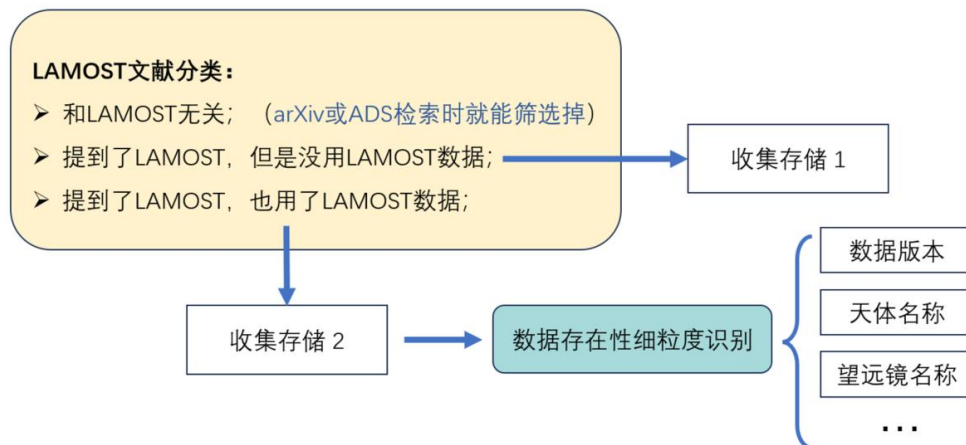
Original abstract	Completed by GPT-4
<p>The Magellanic Stream (MS) - an enormous ribbon of gas spanning 140° of the southern sky trailing the Magellanic Clouds - has been exquisitely mapped in the five decades since its discovery. However, despite concerted efforts, no stellar counterpart to the MS has been conclusively identified. This stellar stream would reveal the distance and 6D kinematics of the MS, constraining its formation and the past orbital history of the Clouds. We have been conducting a spectroscopic survey of the most distant and luminous red giant stars in the Galactic outskirts. From this dataset, we have discovered a prominent population of 13 stars matching the extreme angular momentum of the Clouds, spanning up to 10° along the MS at distances of 60-120 kpc. Furthermore, these kinematically-selected stars lie along a $[\alpha/\text{Fe}]$-deficient track in chemical space from -2.5 to -0.5, consistent with their formation in the Clouds themselves. We identify these stars as high-confidence members of the Magellanic Stellar Stream. Half of these stars are metal-rich and closely follow the gaseous MS, whereas the other half are more scattered and metal-poor. We argue that the metal-rich stream is the recently-formed tidal counterpart to the MS, and speculate that the metal-poor population was thrown out of the SMC outskirts during an earlier interaction between the Clouds. The Magellanic Stellar Stream provides a strong set of constraints - distances, 6D kinematics, and birth locations - that will guide future simulations towards unveiling the detailed history of the Clouds.</p>	<p>The Magellanic Stream (MS) - an enormous ribbon of gas spanning 140° of the southern sky trailing the Magellanic Clouds - has been exquisitely mapped in the five decades since its discovery. However, despite concerted efforts, no stellar counterpart to the MS has been conclusively identified. This stellar stream would reveal the distance and 6D kinematics of the MS, constraining its formation and the past orbital history of the Clouds. We propose a novel approach to search for the elusive stellar counterpart of the Magellanic Stream. By employing advanced photometric and spectroscopic techniques, combined with state-of-the-art data mining and machine learning algorithms, we aim to systematically analyze and cross-reference multi-wavelength survey data to identify potential candidate stars belonging to the hypothesized stellar stream. Our methodology involves leveraging the distinctive chemical signatures and proper motions associated with stars originating from the Magellanic Clouds, allowing us to differentiate them from the foreground and background contaminants. If successful, this discovery would not only provide essential insights into the origin and evolution of the Magellanic Stream, but also enhance our understanding of the interaction between the Magellanic Clouds and the Milky Way. Furthermore, the proposed approach establishes a powerful framework for investigating other unresolved galactic structures and tracing their underlying dynamical histories.</p>
<p>Completed by LLaMA-2</p> <p>The Magellanic Stream (MS) - an enormous ribbon of gas spanning 140° of the southern sky trailing the Magellanic Clouds - has been exquisitely mapped in the five decades since its discovery. However, despite concerted efforts, no stellar counterpart to the MS has been conclusively identified. This stellar stream would reveal the distance and 6D kinematics of the MS, constraining its formation and the past orbital history of the Clouds. We find a stellar stream in the MS, with a 6D velocity of 100 km s^{-1}, and a distance of 100 kpc. The stream is 100 pc wide and 1000 pc long, and is consistent with the MS. The stream is 1000 times more massive than the Magellanic Clouds, and is likely to be a tidal stream from the Large Magellanic Cloud.</p>	<p>Completed by AstroLLaMa</p> <p>The Magellanic Stream (MS) - an enormous ribbon of gas spanning 140° of the southern sky trailing the Magellanic Clouds - has been exquisitely mapped in the five decades since its discovery. However, despite concerted efforts, no stellar counterpart to the MS has been conclusively identified. This stellar stream would reveal the distance and 6D kinematics of the MS, constraining its formation and the past orbital history of the Clouds. We 1) present a new method to select MS candidate stars from wide field optical imaging, 2) apply this method to the MS, 3) compare the results to existing catalogs, and 4) identify the most promising MS candidate stars. We apply our method to 2547 stars in the Gaia EDR3 catalog that have Gaia proper motions consistent with the MS and use these as a pilot sample to optimize our method. We then apply our method to the entire Gaia EDR3 catalog to identify 46,000 stars that are candidates for the MS, and we use the Gaia-ESO Survey to identify 13,000 of these stars that are likely to be MS candidates. We identify 266 stars with (G-K) colours and [Fe/H] that are consistent with previous estimates of the MS. These stars have proper motions and radial velocities that are consistent with the MS, and we place them on a 3D map of the MS. We find a previously unknown stellar stream, which we call the Southeastern Stream (SES), in the direction of the Clouds and MS. The SES has an M31 distance of $32.2^{+0.5} \text{ kpc}$ and a 6D velocity of 136^{+10} km/s. The SES is about 20 times brighter in stars than the MS and about 40 times brighter in stars than the SMC. The SES is also 2.7 times more metal-rich than the MS, which suggests that it is a separate population.</p>

(Nguyen et al. 2023)

典型应用：天文大模型

天文文献分类与命名实体识别：

- 为了更加精细化和准确地检索天文文献、构建天文知识图谱、实现数据库关联天文细粒度知识，国家天文科学数据中心团队利用大语言模型对天文文献进行精细化分类与命名实体识别；



典型应用：天文大模型

■ 天文大语文模型：

科研助手

LLM+知识库

■ 天文图文大模型：

自动识图



LAMOST数据文档智能检索问答助手

使用示例1

- 简单事实类问题：例如，数据发布包含的星表



LAMOST中分辨巡天数据发布包括哪些星表？



LAMOST中分辨巡天数据发布包括以下五个星表：

- "LAMOST中分辨总星表"
- "LAMOST中分辨率参数星表"
- "LAMOST中分辨重复观测源表"
- "LAMOST中分辨观测天区星表"
- "LAMOST中分辨输入星表"

✓ 知识库匹配结果

出处 [1] [LAMOST中分辨率巡天第五次国际数据发布说明文档1.pdf](#)

moon_flg 布尔型 每个目标到月亮的角距离 subproject 字符型 LAMOST 中分辨光谱巡天项目名称 二、LAMOST 中分辨参数星表 本节将介绍“LAMOST 中分辨率参数星表”。该表提供了 LASP 软件计算的大气参数及误差，有效温度在[4500 K, 8500 K]范围或者有效温度在[4000 K, 4500 K]范围内且表面重力小于 3.5 dex 的大气参数是使用蓝端合并光谱测量的结果，有效温度在[3300 K, 4000 K]范围或者在[4000 K, 4500 K]范围内且表面重

出处 [2] [LAMOST中分辨率巡天第五次国际数据发布说明文档1.pdf](#)

勒天区巡天”、“双星巡天”、“银河系巡天”、“TESS 天区后续巡天”、“恒星形成区巡天”、“疏散星团巡天”、“银河系星云巡天”和“四个 K2 天区双星测试观测”。这里提供了两种格式的“LAMOST 中分辨总星表”，一种是.fits 格式，另一种是.csv 格式，可以从 <http://www.lamost.org/dr11/v1.0/catalogue> 网站下载它们。表 7: LAMOST 中分辨总星表 字段 (单位) 数据类型 说明 mobsid 长整型 每次曝光每条中分辨光谱唯一的编号 obsid 长整型

出处 [3] [LAMOST中分辨率巡天第五次国际数据发布说明文档1.pdf](#)

星表 本节将介绍此次发布的五个 LAMOST 中分辨星表，用户可以从 <http://www.lamost.org/dr11/v1.0/catalogue> 网站上下下载它们。这五个星表分别是：“LAMOST 中分辨总星表”、“LAMOST 中分辨率参数星表”、“LAMOST 中分辨重复观测源表”、“LAMOST 中分辨观测天区星表”、以及“LAMOST 中分辨输入星表”。此次，到月亮的角距离小于 10 度的目标都没有观测，“LAMOST 中分辨总星表”和“LAMOST 中分辨率参数星表”中有三个与到月亮的角距离相关的字段，

LAMOST数据文档智能检索问答助手 使用示例1

- GPT-4o对比



你

LAMOST中分辨巡天数据发布包括哪些星表？



GPT-4o

LAMOST (Large Sky Area Multi-Object Fiber Spectroscopic Telescope, 大天区面积多目标光纤光谱望远镜) 巡天数据发布通常包括以下几类星表：


1. **恒星光谱星表**：这些星表包括通过LAMOST观测所获得的大量恒星的光谱信息。包括恒星的标识号、位置、亮度、光谱类型、径向速度以及一些天文物理参数（如温度、表面引力和金属丰度等）。
2. **类星体 (Quasar) 星表**：LAMOST也观测到了一些类星体的数据，发布的类星体星表包含这些类星体的位置、红移、亮度和光谱特征等信息。
3. **活动星系核 (AGN) 星表**：类似类星体，LAMOST对一些活动星系核进行了观测，AGN星表包含这些活动星系的光谱数据及相关物理参数。
4. **候选恒星多重系统星表**：包括双星系统或多星系统的观测结果。
5. **候选星团成员星表**：一些特定星团内部或附近的恒星的光谱数据。
6. **特定天体及事件观测星表**：针对一些天文现象或特殊目标进行的观测结果，可能包括超新星、变星等。

以上是LAMOST数据发布中常见的几类星表，每次数据发布具体包括哪些星表，可以查阅当次的数据发布说明文档以获得详细信息。

China-VO

LAMOST数据文档智能检索问答助手 使用示例1

• GPT-4o对比

 你
LAMOST中分辨巡天数据发布包括哪些星表? 请给出参考文献

 GPT-4o
搜索了3个网站

根据查询到的多个数据发布文档, LAMOST中分辨巡天 (MRS) 数据发布包括以下主要星表:

1. LAMOST MRS General Catalog (LAMOST MRS 通用星表)

- 包含所有观测到的天体的光谱数据以及基本的物理参数。

2. LAMOST MRS A, F, G, K Stellar Parameters Catalog (LAMOST MRS AFGK 类型恒星参数星表)

- 专门针对恒星分类中的A型、F型、G型和K型恒星, 提供详细的光谱参数。

3. LAMOST MRS M giants Catalog (LAMOST MRS M型巨星星表)

- 涵盖所有观测到的M型巨星, 提供其光谱数据和物理参数。

4. LAMOST MRS Dwarfs and Subdwarfs Catalog (LAMOST MRS 矮星与亚矮星星表)

- 包括矮星和亚矮星的详细光谱数据及其物理参数。

5. LAMOST MRS Line Indices of A Type Stars Catalog (LAMOST MRS A型恒星光谱线指标星表)

- 专门记录A型恒星的光谱线指数。

6. LAMOST MRS White Dwarf Star Catalog (LAMOST MRS 白矮星星表)

- 收录白矮星的光谱数据和相关参数。

这些星表为研究恒星和其他天体的光谱特性提供了丰富的数据支持, 适用于各种天文学研究目的。

参考文献

- LAMOST Data Release Documents:
 - [Data Release Three of LAMOST Medium-Resolution Spectroscopic Survey \(MRS\)](#)
 - [Data Release One of LAMOST Medium-Resolution Spectroscopic Survey \(MRS\)](#)
 - [Home - LAMOST DR8](#)

LAMOST数据文档智能检索问答助手 使用示例2

- 简单事实类问题：例如，星表中包含的参数



LAMOST 低分辨类星体发射线特征星表包含哪些参数



LAMOST 低分辨类星体发射线特征星表包含 184 个参数。其中，前面 34 个参数（“sn_ratio_conti”字段之前）也包含在“LAMOST 低分辨总星表”。后面 147 个字段是测量类星体光谱中六组谱线过程的得到相关参数。这些参数包括赤经、赤纬、信噪比、Gaia 的源编号、Gaia 星等。

LAMOST数据文档智能检索问答助手

使用示例2

✓ 知识库匹配结果

出处 [1] [LAMOST低分辨巡天第十一次数据发布说明文档.pdf](#)

32 “LAMOST 低分辨类星体发射线特征星表”发布了 19,052 条类星体光谱的 184 个参数，前面 34 个参数 (“sn_ratio_conti”字段之前) 也包含在“LAMOST 低分辨总星表”。后面 147 个字段是测量类星体光谱中六组谱线过程的得到相关参数。测量类星体光谱参数之前，使用 SDSS 测光数据对 LAMOST 类星体光谱进行流量定标[19]，并将类星体发射线分为六组进行测量：(1) 第一组是 Lyman alpha (Lya) 的宽线成分和窄线成分；(2) 第二组是 C IV 的宽线成分和窄线成分；(3)



出处 [2] [LAMOST低分辨巡天第十一次数据发布说明文档.pdf](#)

“LAMOST 低分辨总星表”、“LAMOST 低分辨 A、F、G 和 K 型星参数星表”、“LAMOST 低分辨 A 型星线指数星表”、“LAMOST 低分辨 M 巨星、矮星和亚矮星星表”、“LAMOST 低分辨重复观测源星表”、“LAMOST 低分辨观测天区星表”、“LAMOST 低分辨输入星表”、“LAMOST 低分辨激变星星表”、“LAMOST 低分辨白矮星星表”、“LAMOST 低分辨类星体发射线特征星表”和“LAMOST 低分辨星系的星族合成星表”。这些星表中包含了很多参数，例如：赤经、赤纬、信噪比、Gaia 的源编号、Gaia 星等、恒星大气参数（有效温

出处 [3] [LAMOST低分辨巡天第十一次数据发布说明文档.pdf](#)

需要注意的是，“LAMOST 低分辨 A、F、G 和 K 型星星表”、“LAMOST 低分辨 A 型星线指数星表”、“LAMOST 低分辨 M 巨星、矮星和亚矮星星表”、“LAMOST 低分辨激变星星表”、“LAMOST 低分辨白矮星星表”、“LAMOST 低分辨星系的星族合成星表”和“LAMOST 低分辨类星体发射线特征星表”都是“LAMOST 低分辨总星表”的子集。表 6、7、8、9、10、11、12、13、14、15 和 16 分别给出了这 11 个星表中的所有字段，并且对每个字段进行详细说明。这些星表中大多数字段已经在第二章介绍过，下面只介绍前面章节没有解释过的字段。

LAMOST数据文档智能检索问答助手 使用示例2

- GPT-4o对比



你

LAMOST 低分辨类星体发射线特征星表包含哪些参数



GPT-4o

搜索了3个网站

LAMOST 低分辨类星体发射线特征星表包含以下参数:

1. 赤经 (RA)
2. 赤纬 (Dec)
3. 星等 (Magnitudes)
4. 信噪比 (Signal-to-Noise Ratio)
5. Gaia的源编号 (Gaia Source ID)
6. Gaia星等 (Gaia Magnitudes)
7. 恒星大气参数 (包括有效温度、表面重力和金属丰度)
8. 视向速度 (Radial velocity)
9. 元素丰度 (Element Abundances)
10. 光谱谱线指数 (Spectral Line Indices)

更多详情可以参考 [LAMOST DR8数据发布文档](#)¹.

China-VVO

LAMOST数据文档智能检索问答助手

使用示例3

- 简单事实类问题: 例如, 星表或光谱文件中参数的含义、计算方式等



LAMOST中分辨光谱大气参数怎么计算的?



LAMOST中分辨光谱大气参数的计算采用了LASP软件和卷积神经网络方法。LASP软件同时给出了大气参数误差, 这些误差不仅包含了算法引入的误差, 同时还考虑了光谱观测过程中引入的误差。LAMOST数据发布网站的“Release Note”文件中会展示中低分辨光谱大气参数性能相关的实验结果, 如不同信噪比范围大气参数误差的分布、LAMOST与SDSS同源目标大气参数对比、LAMOST与APOGEE同源目标大气参数对比以及LAMOST与APOGEE同源目标元素丰度对比图等。可靠性的用户可以参考“Release Note”。

LAMOST数据文档智能检索问答助手 使用示例3

- GPT-4o对比



你

LAMOST中分辨光谱大气参数是怎么计算的？



GPT-4o

LAMOST (Large Sky Area Multi-Object Fiber Spectroscopic Telescope, 即郭守敬望远镜) 是中国科学家建造的大型天文设备, 用于进行大规模光谱观测。对LAMOST获得的光谱进行高分辨率分析, 可以提取恒星大气的参数, 如有效温度、表面重力、金属丰度等。这些参数通常是通过光谱分析软件计算出来的。以下是一般步骤:

1. **光谱数据预处理**: 这一步包括数据的平场校正、波长标定和天光扣除等处理, 以确保光谱数据的质量。
2. **归一化**: 将光谱数据归一化, 使得光度数据在不同波长段之间具有可比较性。
3. **光谱拟合**: 使用标准星光谱或合成光谱库与观测光谱进行比较和拟合。常用的方法包括:
 - **模板匹配**: 将观测光谱与已知参数的标准星光谱进行比较, 以确定最佳匹配的模板。
 - **合成光谱匹配**: 利用理论模型生成合成光谱, 将其与观测光谱进行匹配以确定参数。
4. **最小二乘法拟合**: 对观测光谱和模型光谱之间的差异进行最小二乘法拟合, 以最小化差异并确定最佳参数。
5. **贝叶斯方法**: 使用贝叶斯统计方法, 通过计算后验概率分布来推导出大气参数。
6. **机器学习**: 近年来, 机器学习算法 (如神经网络、随机森林等) 也被引入到光谱分析中, 以更高效地提取大气参数。

总体来说, 计算大气参数涉及多个步骤和复杂的算法, 具体实现依赖于用于分析的软件和模型数据库。LAMOST项目中使用的软件工具一般包括Laspec (LAMOST Stellar Parameters Pipeline) 和WADS (Wide-field Analysis of Datasets System), 它们综合利用了上述方法来处理和分析光谱数据。

China-VO

LAMOST数据文档智能检索问答助手

使用示例3

✓ 知识库匹配结果

出处 [1] [用户常见问题及解答.pdf](#)

(2) 国内和国际数据发布时需要包括从 2012 年测试巡天以来的所有数据。4、LAMOST 中分辨光谱大气参数怎么计算的？分别使用 LASP 软件和卷积神经网络方法为中分辨光谱计算了大气参数，并同时发布它们的结果。LASP 软件同时给出大气参数误差，这些误差不仅仅包含了算法引入的误差，同时还考虑了光谱观测过程中引入的误差。每次数据发布时，数据发布网站“DOCUMENTS”菜单里面名为“Release Note”的文档会展示中低分辨光谱大气参数性能相关的实验结果，比如：不同信噪比范围大气参数误差的分布、LAMOST 与 SDSS 同源目标大气参数对比、LAMOST 与 APOGEE 同源目标大气参数对比、以及 LAMOST 与 APOGEE 同源目标元素丰度对比图等。想了解每次发布的大气参数等数据可靠性的用户可以参考“Release Note”。

出处 [2] [LAMOST中分辨率巡天第五次国际数据发布说明文档1.pdf](#)

moon_flg 布尔型 每个目标到月亮的角距离 subproject 字符型 LAMOST 中分辨光谱巡天子项目名称 二、LAMOST 中分辨参数星表 本节将介绍“LAMOST 中分辨率参数星表”。该表提供了 LASP 软件计算的大气参数及误差，有效温度在[4500 K, 8500 K]范围或者有效温度在[4000 K, 4500 K]范围内且表面重力小于 3.5 dex 的大气参数是使用蓝端合并光谱测量的结果，有效温度在[3300 K, 4000 K]范围或者在[4000 K, 4500 K]范围内且表面重



LAMOST数据文档智能检索问答助手

使用示例4

- 整合总结类问题: 例如, 如何检查光谱数据存在的问题, 需综合知识库中多处内容回答



如何检查光谱数据是否存在问题, 有什么问题?



根据已知信息, 可以检查光谱数据是否存在以下问题:

1. 检查“与标记位”和“或标记位”字段, 如果它们的二进制表示中某一位是1, 则说明该波长处存在问题。
2. 检查“band_b”和“band_r”字段, 如果它们为“1”, 则说明该波长处存在问题。
3. 检查“NODATA”字段, 如果为“1”, 则说明该行光谱数据没有好数据。

此外, 还可以检查光谱数据中是否存在“倒方差”为0或流量小于或等于0的情况, 以及检查“mobsid”字段中是否存在“lmjm”为0的情况。



LAMOST数据文档智能检索问答助手

使用示例4

✓ 知识库匹配结果

^

出处 [1] [LAMOST中分辨率巡天第五次国际数据发布说明文档1.pdf](#)

4 BRIGHTSKY 天光水平太高 5 BADCENTER 光纤轨迹超过 CCD 图像边界 6 NODATA 没有好数据 如果用户想检查一条光谱发生了表 5 所列的哪种情况, 可以首先把十进制的“与标记位”和“或标记位”转换为六位二进制数。如果二进制“与标记位”中某一位是 1, 则说明该波长处每次曝光肯定发生了表 5 中与这一二进制位对应的情况。同理, 如果二进制“或标记位”中某一位是 1, 则说明该波长处至少有一次曝光发生了表 5 中与这一二进制位对应的情况。

出处 [2] [LAMOST低分辨巡天第十一次数据发布说明文档.pdf](#)

NODATA 没有好数据 郭守敬望远镜(LAMOST)技术报告 第13页 编号: LAMOST-YX-TR-SCB-1C-11-v1.0 如果用户想检查一条光谱发生了表 5 所列的哪种情况, 可以首先把十进制的“与标记位”和“或标记位”转换为六位二进制数。如果二进制“与标记位”中某一位是 1, 则说明该波长处每次曝光肯定发生了表 5 中与这一二进制位对应的情况。同理, 如果二进制“或标记位”中某一位是 1, 则说明该波长处至少有一次曝光发生了表 5 中与这一二进制位对应的情况。

China

出处 [3] [用户常见问题及解答.pdf](#)

单次曝光光谱和合并光谱, “mobsid”的构成不同。单次曝光光谱, mobsid = obsid + lmjm + band; 合并光谱, mobsid = obsid + band。其中, “lmjm”是本地修正儒略日, “band”有两个值分别是 B 和 R, 分别标记是蓝端还是红端光谱。11、一般什么情况下, 光谱的流量可能不可靠? 一般来说, 出现下列情况, 光谱流量可能不可靠。因此, 数据处理过程中, 建议尽量去掉出现这几种情况对应的波长和流量。i. “倒方差”为 0 ii. 流量小于或者等于 0 iii. andmask 或 ormask 标记位不为 0 12、2018 年 9 月以后, 中分辨星表中天区名 planid 的命名规则是什么? (1) 第一类: 中分辨天区

LAMOST数据文档智能检索问答助手

使用示例4

• GPT-4o对比



你

如何检查LAMOST光谱数据存在的问题，有什么问题？



GPT-4o

LAMOST（大型光纤光谱天文望远镜，也称郭守敬望远镜）数据的质量检查是确保天文研究准确性的重要步骤。以下是一些常见的方法和可能存在的问题：

1. 数据质量检查方法

1. 信噪比（Signal-to-Noise Ratio, S/N）评估：

- 高信噪比意味着光谱数据具有较高的信号质量，噪声较低。
- 检查不同波长范围内的信噪比，确认是否均匀。

2. 平滑和归一化：

- 对光谱进行平滑处理，以减少噪声对具体信号的干扰。
- 将光谱归一化，以便于比较不同光谱之间的强度。

3. 光谱分辨率：

- 确认光谱的分辨率是否满足研究要求。
- 检查仪器分辨率和数据分辨率的一致性。

4. 背景噪声和仪器效应：

- 检查背景噪声水平，消除天空背景和仪器噪声的影响。
- 对比暗场（Dark Frame）和平场（Flat Field）校正数据。

5. 系统误差和恒星光谱标准：

- 对比恒星光谱标准，检测系统误差。
- 使用已知光谱标准星进行标定。

6. 光谱线识别和匹配：

- 检查特征光谱线的识别准确性（如氢线、钠线等）。
- 与理论模型或已知数据库进行匹配，确认光谱线识别的准确性。

7. 红移和速度信息：

- 计算红移，确认其与实际观测一致性。
- 检查星际介质引起的光谱偏移和吸收。

2. 可能存在的问题

- 低信噪比：**数据噪声较高，会影响光谱特征的辨识和测量的精度。
- 观测条件不佳：**云层、光污染等环境因素会导致数据质量下降。
- 仪器效应：**仪器老化、校准不准或故障会影响光谱数据的质量。
- 光谱线混淆：**光谱线可能被错识别或误判，影响结果的准确性。
- 数据校正不足：**如背景噪声、暗场和平场校正不充分，影响光谱数据的真实度。
- 大气干扰：**地球大气层吸收或散射光线，对天文光谱造成干扰。

LAMOST数据文档智能检索问答助手

使用示例5

推理类问题: 例如, 某个字段具体值代表的含义等, 需根据文档中表格参数进行推理



FIB_MASK字段的第1位的值为1说明什么



根据已知信息, 当FIB_MASK字段的第1位的值为1时, 代表光纤没有被分配 (NOALLotted) 。

LAMOST数据文档智能检索问答助手

使用示例5

✓ 知识库匹配结果

出处 [1] [LAMOST中分辨率巡天第五次国际数据发布说明文档1.pdf](#)

一位的值为 1，则代表光纤存在相应的问题，表 2 给出了九位二进制数各位所代表的光纤问题。表 1：“FIB_MASK 字段的 9 个二进制位 9 8 7 6 5 4 3 2 1 表 2: 光纤问题 位数 问题
注释 1 NOALLOTTED 没有被分配的光纤 2 BADTRACE “TRACECENTER”程序追迹结果错误的光

出处 [2] [LAMOST低分辨巡天第十一次数据发布说明文档.pdf](#)

FIB_MASK：整数，表示光纤存在的问题。如果想检查光纤存在的问题，首先需要把十进制的“FIB_MASK”转换成九位二进制数，如表 1 所示。如果某一位的值为 1，则代表光纤存在相应的问题，表 2 给出了九位二进制数各位所代表的光纤问题。表 1：
“FIB_MASK”字段的 9 个二进制位 9 8 7 6 5 4 3 2 1 表 2: 表 1 中 9 个二进制位分别对应的光纤问题 位数 问题

出处 [3] [The eleventh Data Release of LAMOST Low-Resolution Spectral \(LSR\) Survey.pdf](#)

should first convert the decimal value of “FIB_MASK” to a nine-bit binary number, which is shown in table 1. If the value of a bit is 1, it represents the fiber has associated problem, and the following table 2 lists the associated fiber problem of each bit. Table 1: Nine bits of the “FIB_MASK” field 9 8 7 6 5 4 3 2 1 Table 2: The fiber problems Bit Problem Comment 1
NOALLOTTED Fiber not allotted



LAMOST数据文档智能检索问答助手 使用示例5

- GPT-4o对比



你

LAMOST光谱fits文件中，FIB_MASK字段的第1位的值为1说明什么



GPT-4o

LAMOST (Large Sky Area Multi-Object Fiber Spectroscopic Telescope, 大天区面积多目标光纤光谱望远镜) 光谱FITS文件中的FIB_MASK字段用于记录每个光纤的状态信息。FIB_MASK字段是一个位掩码 (bitmask), 其中每一位 (bit) 代表一个特定状态或条件。当特定位上的值为1时, 表示该状态或条件成立。

具体到你问的第1位 (即最低位, bit 0) 值为1的情况, 这需要查阅LAMOST数据文档, 找到该位对应的状态信息。根据LAMOST的常见使用文档, 第1位值为1通常表示该光纤存在某种特定的问题。例如, 这一位可能指示该光纤被某种干扰或影响所掩盖。

对于精准的信息, 你可以参考以下资源:

1. [LAMOST 数据发布文档](#)
2. 数据版本的发布说明和技术文档通常会详细列出FIB_MASK字段各个位所对应的状态信息。

China-VO

LAMOST数据文档智能检索问答助手

使用示例5

推理类问题: 例如, 某个字段具体值代表的含义等, 需根据文档中表格参数进行推理



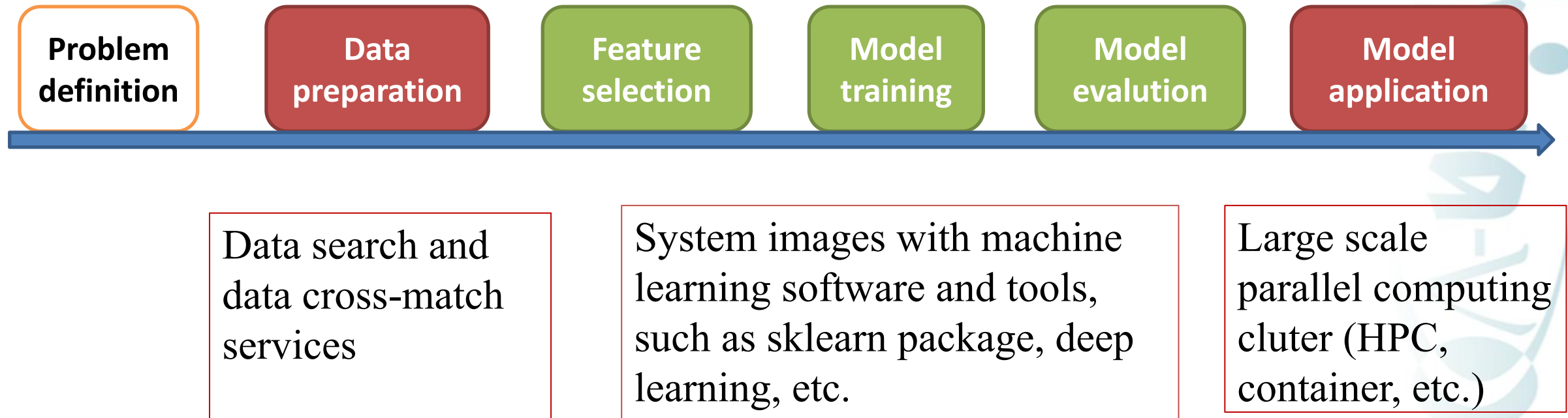
FIB_MASK字段的第4位值为1代表什么?



FIB_MASK字段的第4位值为1代表该光纤存在问题, 具体问题取决于该位的二进制数值。根据表2, 如果是一位二进制数中的值为1, 则代表光纤存在相应的问题。如果 FibBADTRACE 字段中的“TRACECENTER”程序追迹结果错误的光纤, 则该位代表的问题是“BADTRACE”。如果 FibNOPROFILE 字段中的“ORIENTATION” orient 不是“NORMAL”, 则该代表的问题是“BADPROFILE”。

基于AI的天文研究流程

□ The workflow of astronomical data analysis using AI/ML





Thanks!

<https://nadc.china-vo.org>

vo@nao.cas.cn

