

简 讯

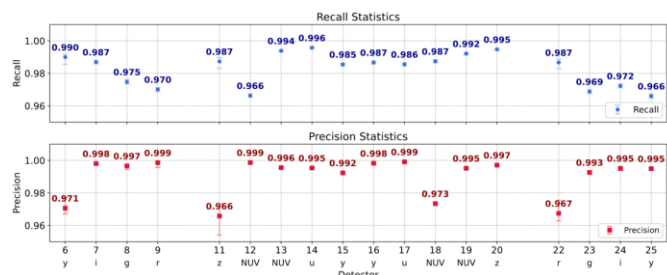
中国空间站工程巡天空间望远镜数据科学中心举行
揭牌仪式

2025 年 9 月 24 日，中国空间站工程巡天空间望远镜（简称 CSST）数据科学中心在之江实验室举行揭牌仪式。数据科学中心旨在探索实践新型科研模式和基于人工智能的科学研究新范式，促进国际一流科学成果产出。揭牌仪式后，中心随即召开“CSST 数据科学与人工智能”学术研讨会，邀请与会专家围绕中心的建设规划、运行管理、科学研究以及合作交流等议题展开深入交流，并就如何发挥之江实验室的优势、推进 CSST 数据科学与人工智能深度融合方面进行了专门讨论，与会专家提出了很多富有建设性的宝贵建议。

数据系统研制进展

多色成像数据处理模块研制进展

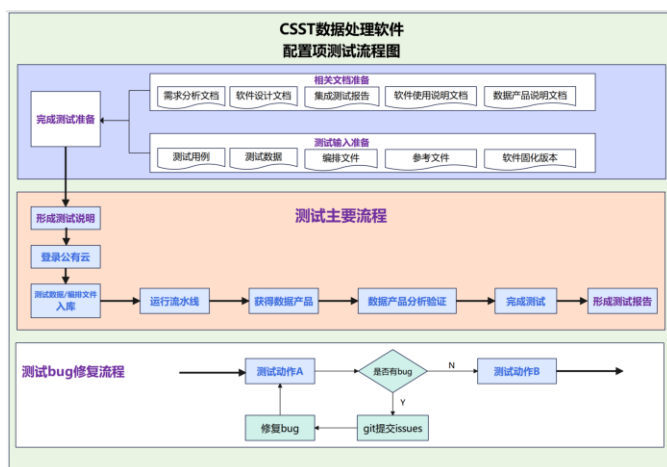
由于 CSST 主巡天相机的特殊性，无法使用传统的多次曝光叠加图像的方法进行宇宙线探测和去除，因此多色成像采用 Lacosmic、deepCR 等多种机器学习方法对单次曝光图像中的宇宙线进行探测。团队使用机器学习 deepCR 方法和基于新版训练得到的数据集，在一级流水线中对此版本的 25 平方度仿真数据完成宇宙线探测。25 平方度仿真数据包括有 5 个不同天球指向的 1221 次曝光，多色成像数据包括约 22000 幅图像。团队对这些图像数据的宇宙射线检测性能进行了系统性分析。以宇宙线识别召回率(recall)和精确率(precision)作为评价指标，总结了 18 块探测器的宇宙线探测性能。该宇宙射线检测功能在这批仿真数据集上表现出优异的整体性能。宇宙线识别召回率达 $97.90 \pm 0.18\%$ （均值 \pm 标准差），精确率达 $98.67 \pm 0.05\%$ 。下图为 18 个探测器的宇宙射线检测召回率（上图）与精确率（下图）。数据点及数值代表各探



测器对应的所有图像的中位数。在 1 级数据产品中，对这些被宇宙线污染的像元进行了标注，有效提高了后续对真实天体的探测效率。

CSST 数据处理软件配置项测试进展

CSST 科学数据处理系统数据处理软件是 CSST 科学数据处理系统三大配置项中任务最重，最为核心的部分，涵盖主巡天、多通道成像仪、积分视场光谱仪、星冕仪和太赫兹五个后端模块。其中主巡天包括多色成像，无缝光谱和在轨定标，共涉及 8 个数据处理软件。CSST 数据处理软件主要承担从 0 级到 1 级（1 级流水线：L1）和 1 级到 2 级（2 级流水线：L2）的数据处理任务。为了保障测试工作的顺利推进，数据处理系统组建了专门的配置项测试团队，确定了各模块的测试负责人，并配备了测试系统管理员、测试数据仿真人员，测试运行人员以及测试数据产品分析人员等，职责清晰、分工明确。自 2025 年 1 月测试启动会之后，团队首先开展了配置项测试专项培训，并建立每周例会机制持续推进测试进度。测试工作在公有云环境开展，截止目前，测试团队已完整走通全流程，包括公有云环境部署、软件安装、测试方案分析、测试用例设计、测试数据仿真、测试运行以及最后的测试数据产品分析和 issue 提交问题和闭环（流程图如下）。团队首先测试软件 L1 部分。经过多次线上线下的沟通与协作，团队已顺利完成公有云环境下的 L1 部分的软件部署和安装、测试方案分析和测试用例设计，目前正在进一步完善《数据处理软件配置项测试说明》文档，同时全面启动配置项测试工作。这些进展为 CSST 科学数据处理系统数据处理软件的高质量运行提供了坚实的保障。



数据系统研制进展

序号	模块名称	月度进展
1	多色成像	支撑配置项测试团队开展多色成像一级流水线配置项测试；利用 25 平方度仿真数据处理结果，完成多色成像一级流水线集成测试报告、产品说明、流水线和软件功能说明等文档；协同仿真团队完成 CTI 效应相关算法测试。
2	无缝光谱	基于新版 COSMOS 天区观测仿真数据，完成了无缝光谱 mosaic 图像的处理与生成；持续推进光谱提取流程的优化；有序开展二级流水线各项功能的整合与测试工作；积极配合系统团队推进整体流水线的集成与配置项测试，初步完成软件功能说明书、数据产品说明等关键文档的编制工作。
3	数据流管理	完成流水线相关命令行操作工具 csst-pipeline-cli 的 v0.0.1 开发。
4	流水线运行管理	完成 grpc 接口向 restful 接口的转换，提升了接口的通用性和易用性；新增了根据 dag_run 字段获取 Scalebox 中 DAG 实例状态和日志的接口功能，进一步增强了流水线的可观测性与运维便利性；配合进行之江环境的流水线测试工作。
5	多通道成像仪	确认了新版本编排策略可行性，即 C1、C2、C3 三通道拆分观测，同时对编排软件定义和软件进行了相应更新；开展配置项测试工作，按照要求编写配置项测试所用各类文档；开展二级流水线模块内部单元测试，接下来计划按照标准流程查漏补缺。
6	积分视场光谱仪	完成一级流水线软件集成测试对应的功能说明，数据产品说明相关文档以及推进二级流水线单元测试和集成；配合科学团队开展 IFS 相关早期观测提案的优化工作。
7	星冕仪	完成 CPIC 集成测试对应的测试总结、流水线软件功能说明、数据产品说明等文档，对配置项测试方案进行优化；对定标工具包进行了调试和优化；开展星冕仪观测编排软件的升级工作。
8	太赫兹	进一步开展太赫兹模块科学数据处理二级流水线单元测试工作，发现并修复 bug 并通过单元测试；完成集成测试相关文档初稿的提交；开展数据处理软件配置项太赫兹模块测试相关说明文件的编写；在公有云环境下，持续推进太赫兹模块科学数据处理流水线配置项测试相关工作。
9	天体测量	完成了天体测量二级流水线的模块内部对接以及各个模块内部的优化。
10	观测数据仿真软件	持续改正并优化仿真软件，更新了无缝光谱模块的光栅参数文件；完成 50 平方度深场的多色成像仿真任务，生成仿真数据 43TB；完成了 CTI 改正测试的仿真数据；完成了 1 平方度的强透镜图像星表。
11	数据可用性标记	完成天区标记第二版的单元测试工作；按照 CCDS 要求完成模型上传工作。
12	在轨定标	依据定标方案梳理内部定标的具体观测策略和可能的编排方案，与编排团队讨论后续将把内部定标观测策略格式化为标准的输入接口提供给编排软件；实现了将定标源置于探测器任意位置处的望远镜中心指向的计算，依据定标方案开展所有定标星场/定标星定标观测的多指向中心坐标的计算，其结果将提供给编排团队，从而测试与巡天穿插编排时对定标方案的影响及优化；实现了巡天模块 CTI 参考文件接入 CCDS，并提供给流水线使用。
13	观测需求编排	持续调整并形成《科学观测需求编排文件结构定义说明》，并依据此文档同步修改参数格式检查工具；正在开展在线系统多个主要页面样式的修改工作，优化交互式定制提案流程的 UI，解决部分显示 BUG。
14	数据处理软件集成	对齐 DFS1&2 的元数据关键字；实现 Scalebox 状态与日志的查询；基本完成数据处理命令行工具 csst-pipeline-cli 的设计与开发。
15	存储与计算	优化元数据服务的性能及代码。
16	天文算子	DeCTI2.0 是对“设备日期预测”及“CTI 效应改正”两个子任务联合建模，通过多任务学习提升模型性能，目前“CTI 效应改正”子任务已在 DeCTI1.0 中得到验证，“设备日期预测”子任务也完成网络设计及模型训练；针对目前结果，计划尝试多手段寻找部分样本精度差的原因；利用物理神经网络 PINN 的思想，将数理方程与深度学习相结合，替代传统的数值仿真方法，实现高维的恒星结构和演化模型。
17	CSST 计算工作台	完成 DFS_L2 V1.0 的发布上线；新增批量上传 fits 文件接口，元数据查询接口增加分页功能、过滤参数支持传入数组；完成流水线重处理及日志查看功能 PRD 评审工作，正在设计前后端接口。
18	数据融合	针对极暗矮星系分析的过程互操作等需求，完成需求、开发文档更新与前后端接口服务开发；增加应用于原始图的高斯滤波去噪功能模块；对无缝光谱多模态分析模型开展模型顶层解冻、投影层深度加深、学习率调整等实验。

数据系统近期节点和计划

时间节点	计划任务安排	时间节点	计划任务安排
7 月 31 日	完成在轨定标方案评审出所评审（已完成）	11 月 30 日	完成 1 级流水线的全部开发和版本固化
8 月 31 日	基本完成 1000 平方度仿真（已完成）	11 月 30 日	完成数据处理软件新功能开发
9 月 30 日	完成多色成像 50 平方度深场仿真（已完成）	12 月 15 日	完成 2 级流水线的全部开发和版本固化
9 月 30 日	完成定标产品生成流水线和生成工具包固化（已完成）	12 月 31 日	完成数据处理软件 2 级流水线配置项测试
10 月 31 日	完成《软件设计说明》评审	12 月 31 日	完成编排软件配置项测试
11 月 30 日	完成数据处理软件 1 级流水线配置项测试	26 年 1 月	实现公有云环境下的部署，开展系统测试
11 月 30 日	完成无缝光谱 50 平方度宽场仿真		