

CSST 科学数据处理系统月报

第 20 期

主办方：中国空间站巡天望远镜科学工作联合中心

时间：2022 年 7 月 11 日

简讯

CSST 科学数据处理系统软件系统设计说明 外部评审会召开

2022 年 7 月 7 日，中科院空间科学与应用总体部在北京组织召开了载人空间站工程巡天空间望远镜科学数据处理系统软件系统设计说明评审会。参加会议的有来自清华大学、北京师范大学、中科院空间应用中心、国家天文台、上海天文台的专家，会议成立评审组。

项目组成员首先做了《载人空间站工程巡天空间望远镜科学数据处理系统软件系统设计说明》的报告，主要内容包括整个科学数据系统的总体任务描述、工作模式、信息流描述、功能和技术要求描述；阐述了产品配套表和软件等级分析，并详细介绍了三个配置项的系统分析设计说明和内外部接口关系等。会上还报告了 CSST 科学数据系统《系统测试大纲》和《三个配置项研制任务书》。经过专家质询，与会专家认为整个系统设计说明完整可行，同意通过评审。目前，项目组根据专家意见正在进行闭环。

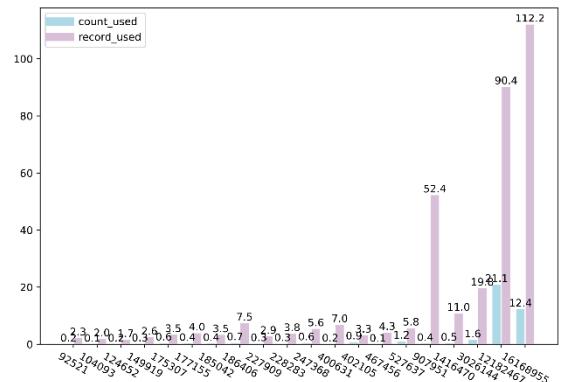
IFS 科学仿真及科学数据分析专家评审会 召开

2022 年 6 月 27 日，中科院上海天文台以线上形式组织召开了空间站数据系统项目外包项目“积分视场光谱仪的科学图像数据模拟”验收评审会。会议成立了由国内多家天文单位专家组成的评审组，听取了承研单位河北师范大学所作的积分视场光谱仪 IFS 的科学图像数据模拟工作汇报，经过质询和讨论，通过了该项目的验收，并给出了后续工作的具体建议。

突出进展

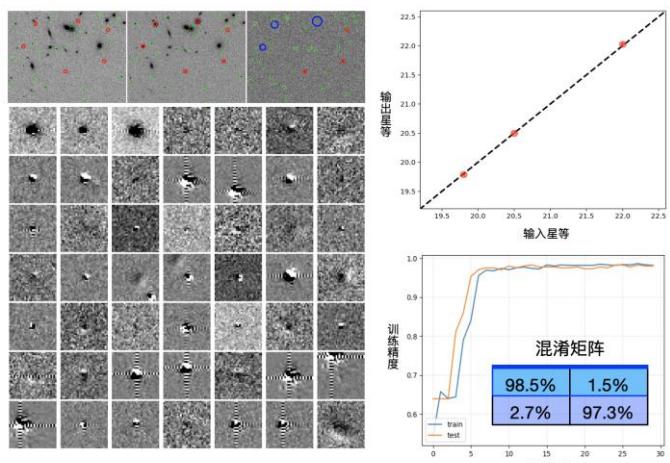
数据流管理模块研制工作进展

经过多次努力，数据流管理模块团队克服了设备与条件不足的问题，在分布式环境下高性能超大规模星表访问技术上取得突破。团队将 Gaia Edr3 放大 11 倍，在 10 节点，万兆以太网环境中模拟了 196 亿条星表数据，在密集天区数据检索测试中，半径为 1 度的情况下返回 16168955 条记录，检索与传输耗时共 112.2 秒，平均每秒高达 14 万记录以上。本项工作的突破，为 CSST 未来数据管理打下了坚实的基础。



多通道成像仪模块研制工作进展

MCI 图像相减/暂现源搜寻软件开发工作已经基本完成，目前正在对 Python 代码进行封装及测试。该软件主要应用于 MCI 暂现源探测识别以及科学研究工作。该模块分为两部分：1) 图像相减，包括参考图像和观测目标图像的位置、流量对齐以及点扩散函数的均一化，基于 ZOGY 算法的图像对减；2) 真假暂现源识别，包括对减图像中的源探测、测光以及基于 VGG 和 ResNet 深度学习算法的真假暂现源分类。我们利用 MCI 图像仿真软件产生 200 组包含暂现源的 gri 波段图像，以此进行算法的开发和训练样本的构建。经测试发现，该模块完成图像对减平均用时约 9 秒，暂现源的识别精度达到 98%。后续工作中，我们将利用理论以及观测的光变曲线模拟不同类型的暂现源图像，进而构建针对不同类型暂现源的分类程序。



图：左上从左至右分别为参考图像、目标图像和对减图像；左下为暂现源候选体，其中前三个为真实的暂现源；右上为暂现源的输入星等以及对减图像中的测量星等的比较；右下为深度学习网络分类精度。


数据系统研制进展

序号	模块名称	月度进展
1	多色成像	完成单次曝光流水线的全流量封装，实现 Cycle5 数据批量处理；完成复杂背景扣除，星表改造，天光梯度拟合等测试报告；开展使用 photutils 等 python 包替代 sextractor 部分功能的测试。
2	无缝光谱	基本完成无缝光谱 Cycle5 流水线原型代码集成、封装、及本地运行调试；初步完成 Cycle5 研制任务总结；制定了 Cycle6 研制任务详细计划。
3	数据流管理	完成 IFS 和 MCI 数据导入模块实现和测试；在百亿级星表构建与高性能访问上取得突破；完成星表导入中包含数组类型的导入，数据库中创建对应数组类型的数据测试导入；完成 Cycle5 工作任务总结。
4	流水线运行管理	完善星表合并流水线；形成一级、二级流水线运行支持方案，启动二级流水线管理模块的研发，支持对二级流水线业务模块的运行调度。
5	多通道成像仪	完成 Cycle5 任务总结报告；完成仪器效应公共模块测试报告；完成图像相减模块。
6	积分视场光谱仪	完成 Cycle5 工作任务总结；对 IFS 仿真和数据处理软件中的科学仿真及科学数据分析模块进行了专家评审。
7	星冕仪	完成 Cycle5 任务总结报告；初步完成在轨定标程序和数据处理流水线，实现了星冕仪数据处理的完整流程。
8	太赫兹	完成 Cycle5 研制任务总结并编写任务总结报告；完成 Cycle6 研制任务规划并编写研制任务计划书。
9	天体测量	完成太阳系小天体数据库的创建，完成了部分太阳系移动天体检测模块中的天文定位程序和太阳系移动天体识别程序，近地天体的轨道外推、轨道类型识别等工作正在有序推进；单帧图像天体测量搭建了包括参考星表构造及调用、参考星坐标转换、图像预处理、目标检测、天文定位等多个模块的数据处理 pipeline；目标检测完成 7 类检测算法的对比测试，各种算法具有不同优缺点，基于连通域识别算法因其高检出率、低虚警率，强可扩展性和具备自主产权等特色，在稀疏星场图像上表现较为突出，密集的星场图像可以考虑使用图像卷积增强目标信噪比的升级优化版算法。
10	观测数据仿真软件	添加了自定义的星等截断功能，完成共 7 组不同工况的 PSF 数据仿真，分析了杂散光对无缝光谱仿真的影响，对比确认了不同 SED 模型的正确性，完成了一次主程序开发版本的迭代更新，并开始推进代码版本的规范和注释添加等工作，同时调研不同探测器的仪器效应，讨论 Cycle6 阶段的任务规划，为下阶段仿真任务的开展做准备。
11	数据可用性标记	完成 Cycle5 阶段任务总结；部分修复了稳像小天区的代码；同时展示平台也恢复工作，正在加紧搭建过程。
12	在轨定标	与巡天观测编排方面进行内部定标及定标星场观测编排方面的讨论；部署最新版的仿真软件进行测试，基于巡天观测编排方面提供的指向信息，正在重新对 3 个典型定标星场进行仿真；考虑 Cycle6 阶段在定标研究方面的仿真需求情况；梳理定标参考文件类型与科学数据处理 pipeline 步骤之间的对应；完成 Bias 和 Dark 等的特性和变化统计程序，并在长光辰芯前照式 6k CMOS 测试中使用形成相应报告；开展 CRDS 相关研究讨论，研究定标产品生成 pipeline；完成 Cycle5 任务总结并考虑 Cycle6 工作计划。
13	观测需求编排	观测需求编排系统实现警报系统接入用户管理与主系统框架；对比多个 Web 可视化框架在实现 CSST 观测计划、运行状等功能的可行性、适用性；基于目前编排系统原型和流程设计讨论后续需求及任务；完成 Cycle5 模块任务总结。


近期节点和计划

2022 年下半年 Cycle6 节点：

时间节点	计划任务安排	时间节点	计划任务安排
22 年 7 月中旬	系统分析设计说明外审闭环	22 年 9 月	协调与数据中心的外部接口
22 年 7 月中旬	Cycle5 任务总结/Cycle6 计划	22 年 10 月	配置项需求规格说明书评审
22 年 7 月中旬	V1.5 仿真软件发布	22 年 10 月	协调与运控系统的外部接口
22 年 7 月-12 月	不同探测器地面测试	22 年 12 月	配置项软件设计评审
22 年 7 月-12 月	观测需求编排外部接口定义	22 年 12 月	实现所有模块统一程序框架并完成主要功能封装
22 年 7 月-12 月	开展算法科学验证	22 年 12 月	二级数据流水线应用接口发布
22 年 8 月底	一级数据定义	22 年 12 月	Cycle6 任务总结