

第二轮 CSST 科学数据挑战大赛

评比报告

一、赛题名称：基于 CSST 无缝光谱图像的天体源和谱型分类

二、评比结果概述

参加本赛题的队伍总共有三支，CSST-DC2026-03、CSST-DC2026-04、CSST-DC2026-05。其中 CSST-DC2026-03 队得分为 96 分，CSST-DC2026-04 队得分为 90 分，CSST-DC2026-05 队得分为 87 分。经过评审，三支参赛队伍在完成本 CSST 科学数据挑战大赛题目的过程中都展现出了出色的物理、天文基础知识与编程能力，都有详细的报告文档、可复现的代码和优秀的科学分析解释。综合来看，三支队伍的分数也符合出题团队设置的优秀答题团队所需的分数线，因此本赛题将评选出 3 支优秀答题团队，即 CSST-DC2026-03、CSST-DC2026-04、CSST-DC2026-05。

三、评分规则

评分维度	评分内容	分值
图像建模质量	是否完成样本在 CSST 单块 CCD 上无缝光谱二维图像建模（包括 0 级和 1 级像）；图像是否符合 CSST 无缝光谱的观测模式	15 分
谱型分类算法	是否合理利用二维无缝光谱图像的特征（如谱线强度、吸收带宽度、连续谱形状等），方法是否具有物理动机或良好的数据适应性	20 分
分类质量与运行效率	评估分类准确率、召回率和完备度，是否评估算法在不同信噪比、不同星等和不同图像质量条件下的表现，给出能进行分类处理的最低要求信噪比，是否有误差分析	25 分
可复现性与代码质量	提交的代码是否结构清晰、文档完整、运行效率良好，是否易于部署和再现	20 分
结果可视化与科学解释	是否提供图示（如分类散点图、光谱特征分布等）；是	10 分

评分维度	评分内容	分值
创新性（加分项）	否结合天体物理知识解释分类边界与失败案例	10 分
	是否引入先进的分类算法（如机器学习、深度学习等），或提出具有推广潜力的新型特征参数	

四、评分结果

参赛队伍编号	图像建模质量得分	谱型分类算法得分	分类质量与运行效率得分	可复现性与代码质量得分	结果可视化与科学解释得分	创新性得分	总分
CSST-DC2026-03	15	18	24	20	10	9	96
CSST-DC2026-04	14	18	22	19	9	8	90
CSST-DC2026-05	13	16	22	18	9	9	87

队伍 CSST-DC2026-03 评审说明：

在图像建模质量上，团队成功实现了基于 CSST 真实光栅配置的单块物理 CCD 的 0 级像和 1 级谱图二维建模，流程图、理论过程与计算代码都很详实充沛。在完成图像正向建模的基础上设计了邻源光谱重叠污染评估模块。算法根据模拟源的位置和 grism 几何轨迹，定量计算了邻近天体 0 级和 1 级色散信号对目标谱迹的流量污染程度，导出了“污染分数（contamination score）”并划分了轻度、中度和重度污染区。这是该队在真实仪器效应建模上的最大优势，为去混叠与流量定标提供了极有力的物理基础。

在谱型分类算法上，团队从二维谱迹 cutout 中提取了连续谱形状、局部谱峰、空间轮廓等特征。实现了两阶段分类任务：恒星/星系源类型二分类；星系粗谱型三分类（将星系根据本征光谱的发射线强度与连续谱特征划分为 emission-line-like 发射线主导型、continuum-dominated 连续谱主导型和 ambiguous 模糊过渡型）。引入了多 Grism 感知策略（grism-blind, grism-aware 和 per-grism specialist）并进行了“留一波段交叉测试（leave-one-grism-out）”，这对多波段无缝光谱的联合分类与模型泛化力评估极具物理动机。但在星系粗谱型分类中，团队直接在训练和测试中剔除了过渡阶段的 ambiguous（模糊）样本。但在实际巡天中，处于恒星形成向宁静星系演化的过渡区天体非常常见。强行在实验中屏蔽这部分具有分类模糊性的源，人为降低了粗谱型分类的物理真实性与算法挑战度。此外，团队也避开了极其困难的恒星谱型细分类，缺少温度序列上的精细恒星大气物理特征区分。

在分类质量与运行效率方面，团队使用了选择性分类控制（Selective Classification）：在低信号强度下（测试缩放因子 0.01），模型不单给出硬标签，还利用验证校准置信度过滤低置信预测，使高置信子集上的分类精度获得显著提升。还利用污染感知低置信控制与掩膜处理，定量证明了中度/重度光谱污染会导致星系被误判为恒星，并探索了基于掩膜的去污染处理。提供了非常完美的性能与资源（RSS 峰值内存、磁盘消耗）的一键分析报告。但报告中二分类绝对准确度（Accuracy）偏低：在信号强度缩放因子为 0.02（低信号端）时，其总体准确率仅为 0.8895。在极低信号 0.01 下退化更加严重。相比于 04 队利用 CNN 在官方测试集上取得的 0.9923 精度，03 队的绝对分类表现略逊一筹。此外，低信噪比区间恒星样本真空依然存在，由于测试集中极低信噪比端恒星样本的缺失，模型对极暗弱天体的泛化评估在恒星端仍不完全。

在可复现性与代码质量上，该团队在可复现性架构上展现了出色的软件工程专业组织清晰，封装良好，设计了三种清晰的复现模式：figures（仅画图）、report（计算实验中间表并画图）和 full（从头前向仿真并全流程计算）。资源 profiling 极度扎实，详尽记录了各模块的峰值内存、耗时和生成路径。唯一的遗憾是复现的计算与存储开销极大：完整 full 模式需要 250 GB 的本地空闲磁盘，且特征提取和仿真中间产物（outputs/ 目录）规模高达 220 GB，限制了全流程重现的便利性。

在结果可视化与科学解释方面，提供了极其详细的报告文档和庞大的科学汇总表和图像，章节图表生成器清晰展现了光谱污染分箱与指标的关系、多 grism 的特征融合与 leave-one-grism-out 的评估。科学解释十分符合物理直觉，精辟阐述了“无缝光谱重叠的非对称特征漂移”，指出重叠光谱产生的额外通量不仅会使信噪比失真，还会改变空间形态的展宽核，因而加剧了星系向恒星分类边界的虚警概率。

在创新性方面，创造性地引入了光谱重叠污染分数的量化，并设计了多 grism 感知（grism-aware）和形态联合训练机制，这些设计与 CSST 宽视场无缝光谱巡天的真实科学数据处理管线贴合度极高，具有非常大的实用推广潜力。。

队伍 CSST-DC2026-04 评审说明：

在图像建模质量上，团队成功仿真了符合 CSST 仪器参数的单块物理 CCD 无缝光谱图像。该仿真完全契合 CSST 观测模式，不仅包含 1 级色散光谱（A 光束），还完整建模了 0 级像（B 光束）以及 +2 级（C）、-1 级（D）、-2 级（E）等 5 个衍射级次。图像生成机制符合天体物理本征性质：星系形态基于 Sersic 轮廓并与高斯 PSF 卷积生成二维展源，恒星作为点源处理，并将视场内数百个源的 0 级像和各级谱迹在一张大 CCD 画布上进行逐源累加重叠，考虑了真实的背景天光、暗流、Poisson 散粒噪声和读出噪声。团队还在配置文件中保留了坏列标记、平场非均匀性、宇宙线等探测器效应接口。

在谱型算法上，团队基于 PyTorch 设计了专用的三支卷积网络 ThreeBandClassifier。模型包含三个结构相同但参数不共享的单波段编码器 BandEncoder，可处理不同波段因色散率差异产生的尺寸不一的切片输入（GU/GV/GI）。引入了 GroupNorm 归一化层，避免了 BatchNorm 对大 batch 统计的依赖，非常适应多通道、强度动态跨度大的天体谱图输入。此外还实现了恒星谱型的多分类算法（B/A/F/G/K/M 六分类），并引入了“平方根反频率加权”

交叉熵损失函数，用以缓解冷热星不平衡的问题；同时设计了光谱数据增强（Spectral Augmentation，色散方向随机平移与噪声抖动），有效提高了对抽取定位不确定性的稳健性。但算法缺失。黑盒模型的局限性与物理约束缺失：CNN 网络在二维图像上滑动提取特征时，倾向于粗糙地捕捉二维形态与连续谱斜率，对于决定恒星光谱型的精细原子吸收线（如 Ca II H&K、G 带、Na I D、Mg b 等）或测光颜色（如 g-r），模型未提供任何物理约束或先验指导，使其在信噪比退化时缺乏物理解析力。

在分类质量与运行效率方面，该工作分类精度评估扎实：在 62,389 个二分类样本上，实现了精度 0.9923、恒星召回率 0.9390。六分类恒星分类在 1,000 个官方样本上 accuracy 达 0.7940。运行效率也极高，给出了清晰的吞吐量量化指标。模型参数量仅 30 万，二分类推理速度达到 2413.5 samples/s，恒星多分类推理速度为 632.9 samples/s，非常易于处理巡天大级次的数据规模。鲁棒分箱合理，诚实地指出，若进行跨全谱型的六分类恒星谱型细分，由于中间谱型的谱线细节退化，最低可用信噪比推荐为 $SNR \geq 100$ 。但是，该工作在恒星候选体纯度较差（假阳性严重），虽然星系被大量正确划归，但由于测试集中星系与恒星存在 61:1 的极端比例，仅 0.68% 的星系假阳性（417 个）就污染了 939 个恒星真阳性，导致恒星的精确率（Star Precision）仅为 0.6925。对于实际科学观测而言，这意味着模型推荐的恒星候选体中有超过 30% 是星系假信号，在实用星表构建中面临极高的污染风险。此外，中段恒星多分类性能较差，模型在 F 型和 G 型谱型上表现极差，F 型星的召回率（Recall）仅为 20.62%，G 型仅为 46.28%。F 型星绝大多数被误分类为 A 型或 G 型，说明在没有物理吸收线约束下，CNN 极易混淆温度序列中段的变化，算法在部分恒星多分类中基本处于不可用状态。

在可复现性与代码质量方面，团队提供了非常规范的模块化 Python 库 csst_pipeline，并编写了直接重现评估的 Shell 脚 run_all_reports.sh，结构符合标准开源库的开发规范。同时提供了完整编译后的学术报告。

在结果可视化与科学解释方面，报告中的科学解释极为透彻，指出二分类由空间形态信号支撑，恒星在垂直色散维上的横向轮廓（受 PSF 点源主导）极窄，而星系通过 Sersic 二维延展形态会在图像横向上产生较宽展宽。这合理解释了为什么在低 SNR 下二分类仍能保持可用，且误分类主要集中在极暗弱、形态紧致（低面亮度）的边缘星系处。对恒星谱型误分类进行了温度序列漂移的深入剖析，并绘制了颜色分箱与分类纯度关系图。虽然提供了天体物理颜色-星等图，但对于三支卷积神经网络本身在二维图像上的关注点，模型没有进行任何特征图可视化（Feature Map）或类激活图（CAM/Grad-CAM）展示。

在创新性方面，该工作提出了一种利用多分支单波段 CNN 汇聚融合提取特征以实现不同尺寸输入二维光谱切片的架构，并在网络层级应用了 GroupNorm。该模型依然建立在成熟的二维卷积网络（CNN）和通用的平方根反频率权重加权损失函数之上，在算法底层和网络拓扑设计上缺乏具有行业颠覆性或高物理动机的原创性设计。

队伍 CSST-DC2026-05 评审说明：

在图像建模质量上，团队成功调用了一维无缝光谱仿真软件 SpecGenerator 类，实现了星系与恒星在 GI 和 GV 波段的二维无缝光谱图像仿真。但缺失 0 级

像与 1 级像的复合混叠建模,仅通过 beam='A' 单独仿真了理想无污染的 1 级像条带,脱离了真实巡天中 0/1 级像混叠的杂散光场景。此外 PSF 与消光模型过于简化,未考虑空间变动 PSF 以及波段内色差导致的色散 PSF 畸变,消光仅采用了常数项相乘,忽略了色散方向的消光倾斜。

在谱型分类算法上,团队巧妙地引入了拓扑数据分析 (Topological Data Analysis, TDA) 与持续同调 (Persistent Homology), 利用二维光谱图像的灰度值构建立方复形。物理动机阐述极为深刻: 设计了上水平集过滤 (对强发射线敏感, 捕捉发射线形成的“孤岛”连通分支 H0 与下水平集过滤 (反图像, 对吸收线敏感, 捕捉吸收谱线和连续谱断裂形成的“洼地”洞穴 H1, 将双波段 (GI + GV) 正反图像拓扑指纹压缩融合成一个 656 维特征向量。该方法不需要进行红移改正和消光改正, 直接提取其尺度和形变不变量, 在处理低信噪比、变消光和未知红移的天体源分类时极具物理适应性。然而, 对于天体精细谱型分类 (例如区分宽线 AGN、恒星形成星系、宁静星系或不同光谱型恒星), 谱线在波长维上的精确相对位置 (对应分子带或离子吸收边) 以及谱线宽度 (反映气体的运动学速度弥散) 是必不可少的物理诊断依据。TDA 特征在这里失去了对这些精细物理尺度的敏感性。

在分类质量与运行效率方面, 团队针对恒星与星系 1:61 的极端样本不平衡问题, 引入了 BorderlineSMOTE 算法, 在边界处对少数类 (恒星) 进行过采样, 并结合随机森林中 class_weight='balanced', 使模型宏平均 F1-score 达到 0.96。该工作还进行了细致的信噪比分箱估计发现在信噪比 1-3 之间恒星的召回率会降至 68.9% 且在低信噪比下性能恶化, 据此给出了该分类器的最低要求信噪比建议 $\text{SNR} \geq 3$ 。但在测试集的 $\text{SNR} < 1$ 分箱中, 总样本数为 16336, 但其中包含了 0 颗恒星。这意味着在最具挑战性的极低信噪比区间, 恒星的召回率与分类器完备度根本无法被有效测定。这个样本选择偏置导致该区间内的评估指标失真, 隐藏了低信噪比下更严重的恒星漏判问题。此外, 报告内也缺少量化的运行效率指标, 没有给出具体的代码运行时长和效率。

在可复现性与代码质量上, 提交的代码组织清晰, 封装良好, 提供了 train.slurm 调度文件, 非常有利于直接在高性能计算集群上提交作业。提供了用于拓扑动力学演化可视化的文件和一整套用于重现图表的绘图脚本, 可复现性极佳。但硬编码相对路径: 脚本中多处使用如 ../data/ 等硬编码的相对路径, 对不同环境下的灵活部署构成了阻碍。

在结果可视化与科学解释方面, 可视化图表极为完备, 提供了包括混淆矩阵、PCA 特征载荷贡献图、分类错误样本的 Betti 曲线比较图、t-SNE 降维聚类以及低信噪比退化过程的 GIF 动图。解释非常深刻且具有天体物理洞察力: 指出了恒星 (点源, 由本征 PSF 主导) 在低信噪比下, 其原本陡峭高耸的峰值信号会被背景噪声抹平, 使得 H0 连通分量的持久性缩短, 在几何上呈现为弥散的斑块, 从而漂移越过拓扑分类边界被误判为星系; 而星系是本征延展源, 即使在低信噪比下也具有更复杂的拓扑空腔 (由 H1 Betti 曲线反映), 所以极少反向误判。但未将结果与天体本征物理量 (星等、红移等) 关联, 虽然分析了 SNR 的影响, 但未给出分类精度或误分类样本在物理参数空间 (例如 Gaia 星等、星系 Sersic 指数、红移中的分布图示。这导致无法从天体物理学角度厘清到底是什么物理类型的恒星 (如极冷矮星 M 型还是明亮早型星) 或星系最容易与决策边界产生混淆。

在创新性方面，拓扑数据分析（TDA）在 CSST 二维光谱图像分类中的引入极具开创性。与目前流行的黑盒 CNN 神经网络相比，它对小样本（如测试集仅有 300 颗恒星）表现出了极强的鲁棒性，且提取的贝蒂曲线和持久熵等特征对于连续的灰度“拉伸”和“旋转”变形具备不变量特性。但正因为其“拓扑不变量”去除了红移和平移信息，该算法虽然在“星系/恒星”的粗粒度二分类上表现优异，但几乎无法推广应用到更精细的光谱分类任务（如测定恒星的光谱型 OBAFGKM、星系红移测定、或者区分活动星系核 AGN 与普通恒星形成星系），导致该创新性方法在外延应用上的局限性较大。

五、评比的结果

参赛队伍得分排序如下（由高到低降序排列）：

1. 参赛队伍 CSST-DC2026-03，总得分：96
2. 参赛队伍 CSST-DC2026-04，总得分：90
3. 参赛队伍 CSST-DC2026-05，总得分：87

根据各队伍得分情况，三支参赛队伍的报告都具有良好的理论建模与结果分析，均较好的完成了题目要求，且三支队伍的得分都满足出题团队的优秀答题团队标准，建议评选三支队伍（CSST-DC2026-03、CSST-DC2026-04、CSST-DC2026-05）为优秀答题团队。