



中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

通过人工智能分析 同步辐射衍散射图像数据

汇报人：李庆梦 博士后

合作导师：赵丽娜 研究员

部门/课题组：多学科研究中心/AI分子组

2024.10.17



1. 研究背景-研究蛋白质结构

应用需求

理解生物机理^[2]

药物设计^[2]

疾病治疗^[3]

实验方法发展与挑战

X射线晶体学

- 优点: 解析高分辨率结构, 具有高灵敏度
- 缺点: 需要蛋白质结晶, 无法捕捉动态结构, 不代表生理状态

冷冻电镜

- 优点: 无需获得蛋白质晶体, 提供结构动态信息
- 缺点: 分辨率相对较低, 样品制备复杂, 需冷冻处理

小角X射线散射

- 优点: 溶液中蛋白质结构, 提供结构动态信息, 样品制备简便
- 缺点: 分辨率相对较低, 解析数据较为复杂

数据分析方法发展与挑战

- 单一数据处理
- 耗时8~246分钟

DAMMIN和DAMMIF软件生成模型^[13]

- 单一参数预测
- 秒级响应

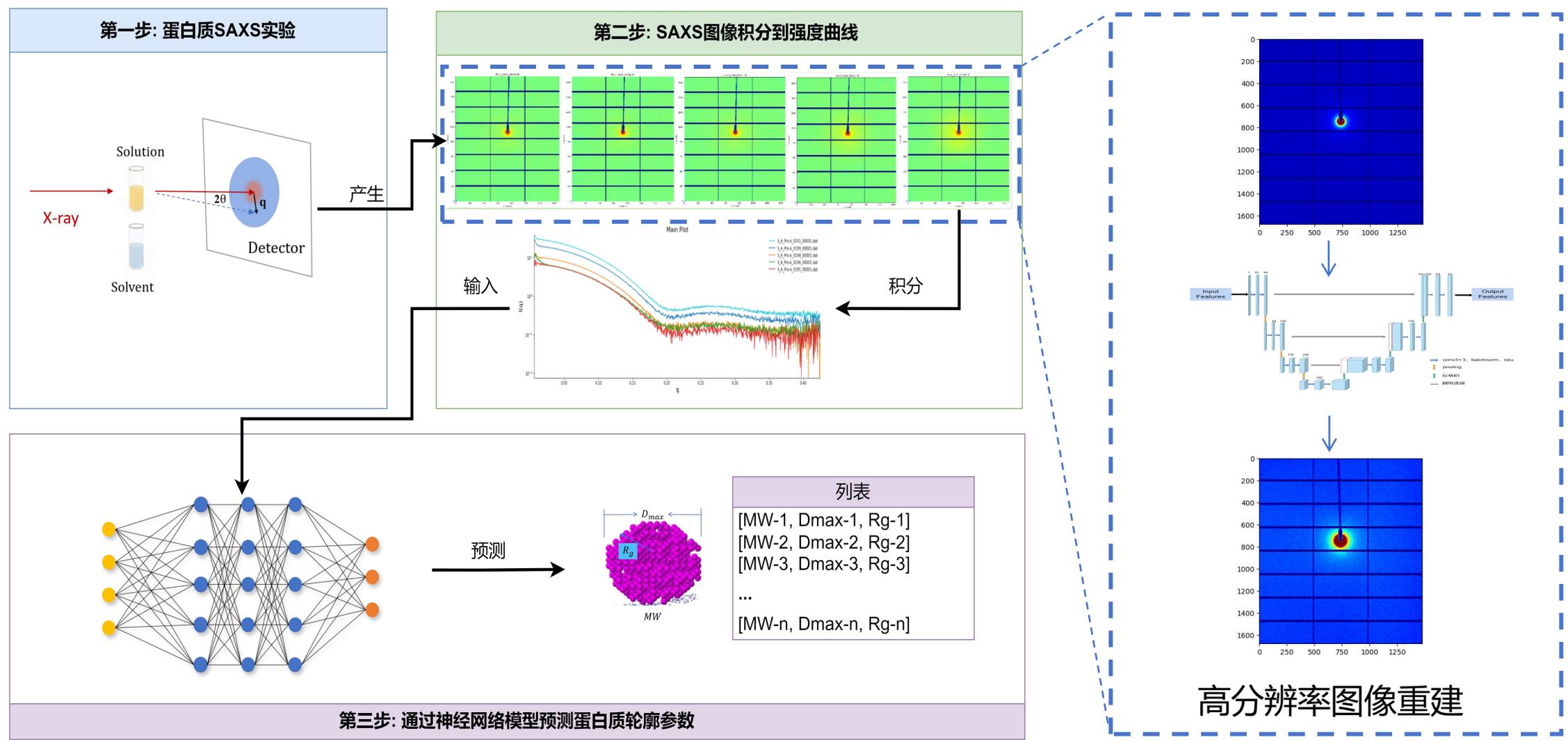
神经网络预测单一轮廓参数^[14]

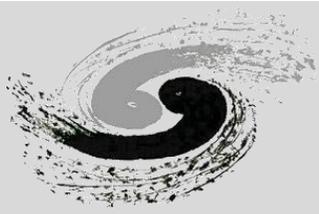
挑战

- 多参数预测
- 秒级响应, 实时在线处理
- 高通量数据处理



2. 研究目标--SAXS高分辨率图像重建与蛋白质轮廓参数预测



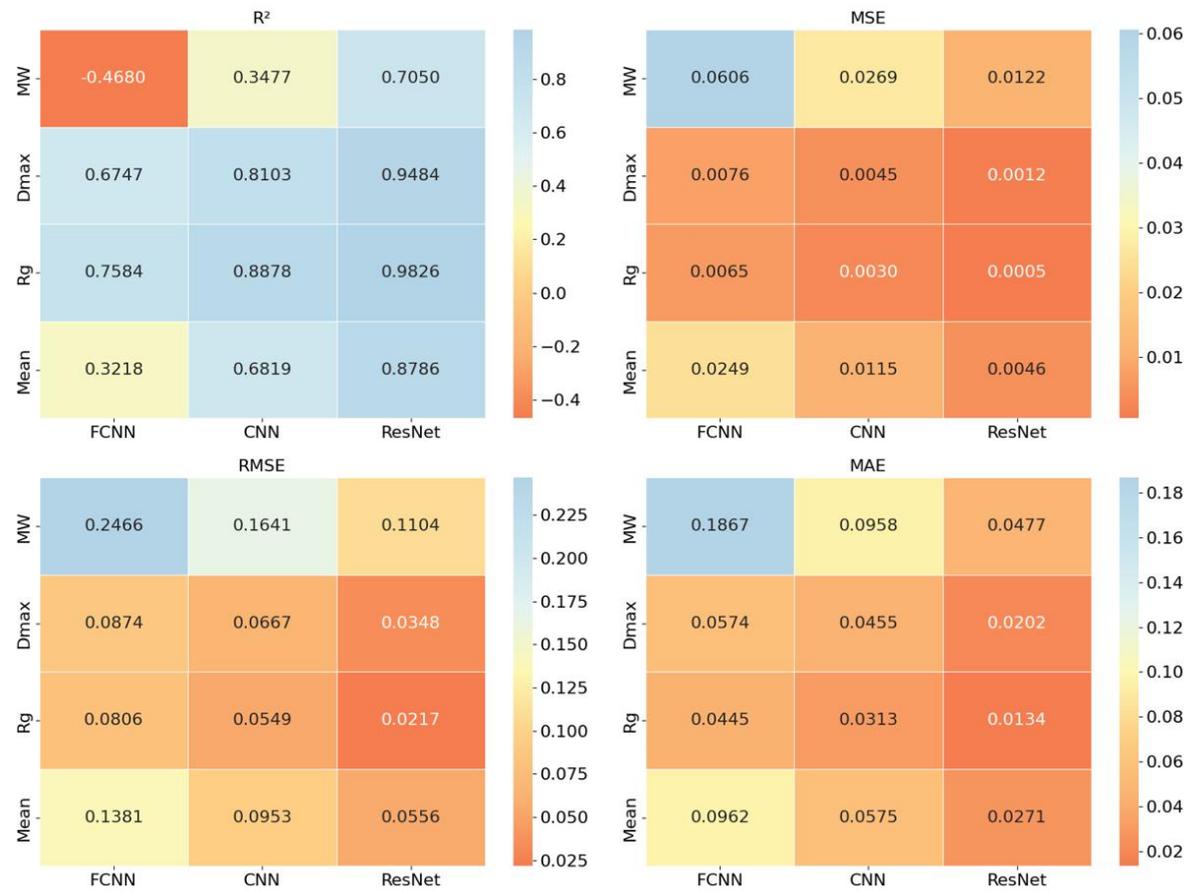
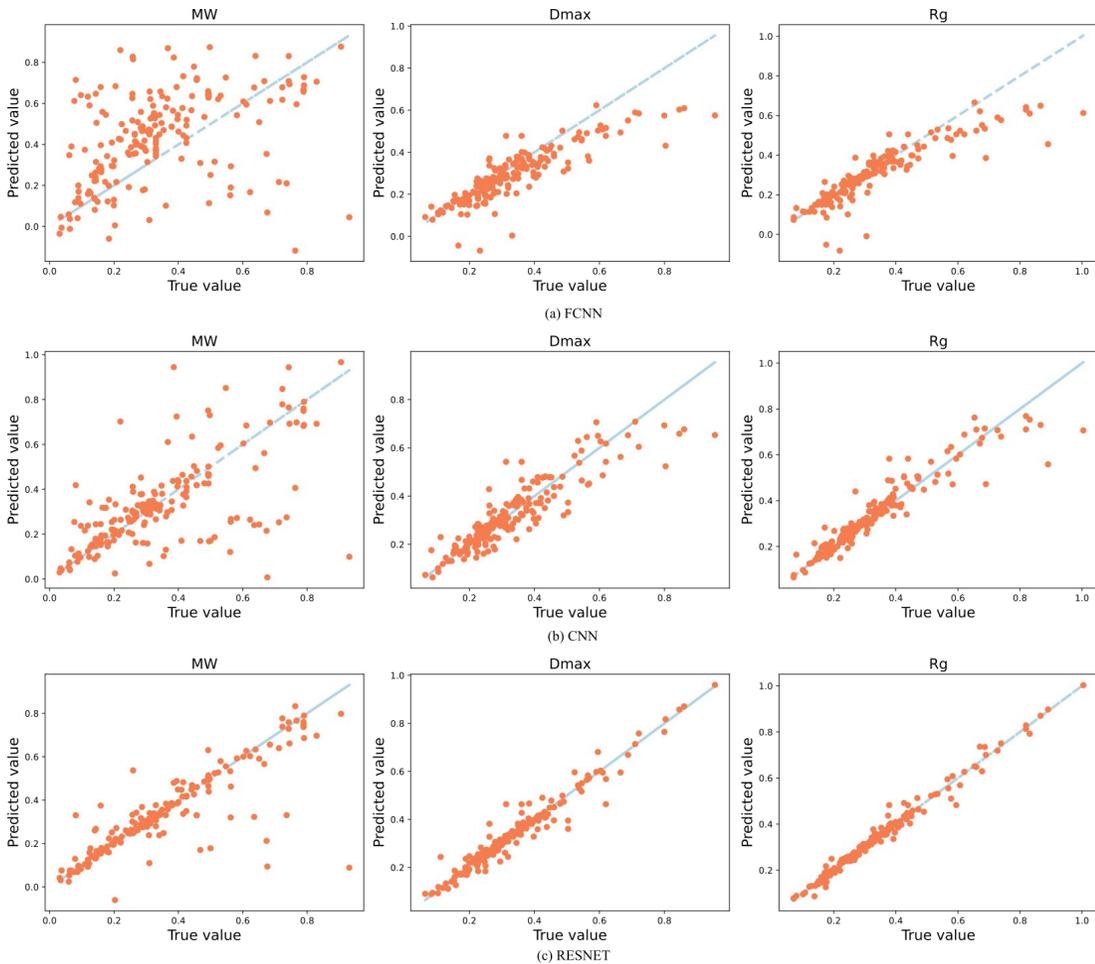


3. 研究进展--SAXS高分辨率图像重建

		1	2	3	4	5
曝光时间0.3s (low-resolution)	图像					
	PSNR	28.2605	29.0994	26.7884	27.0140	19.8979
	SSIM	0.3027	0.3275	0.3016	0.3849	0.3137
曝光时间1.0s (Target)	图像					
	Reconstruction					
Reconstruction	PSNR	29.2348	29.7376	29.4501	29.7965	29.0789
	SSIM	0.4318	0.3925	0.4921	0.5288	0.5150
三合一积分曲线 (扣背底)						



4. 研究进展--蛋白质轮廓参数预测

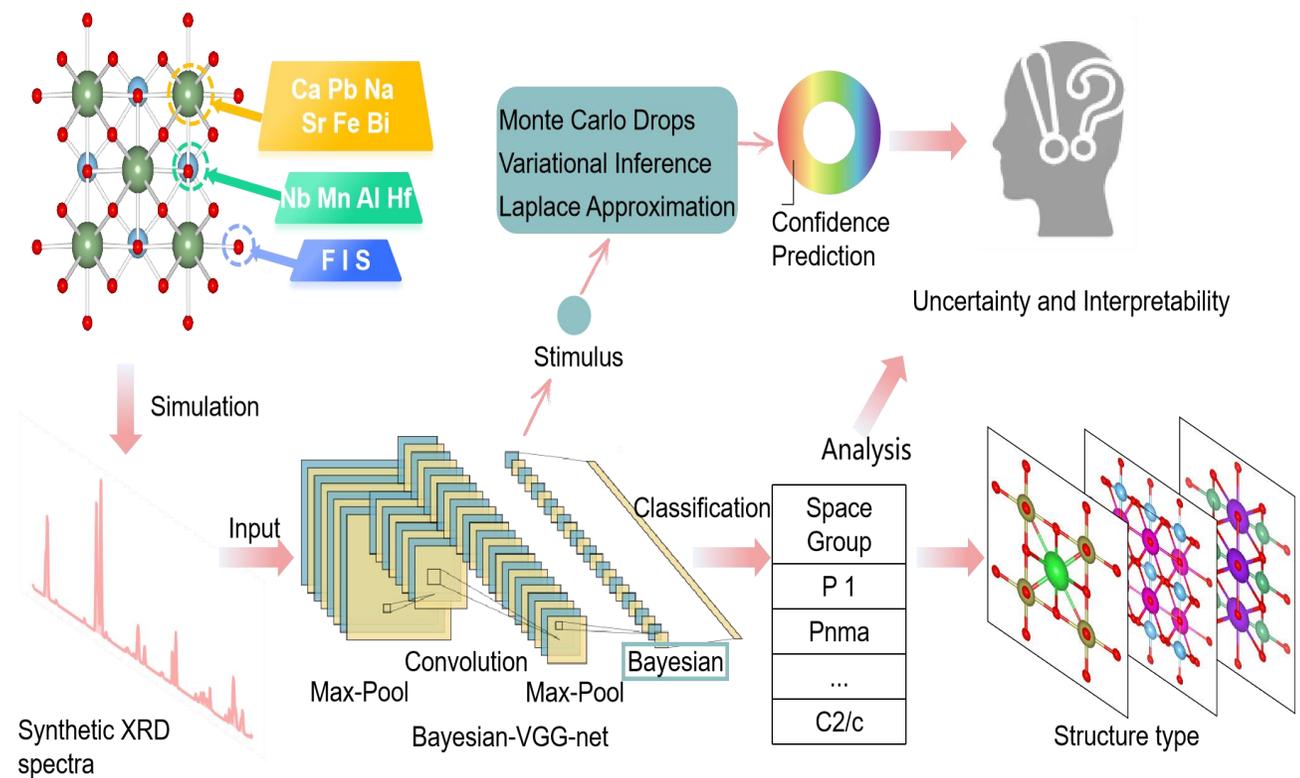


在实验数据上比较了三种深度学习模型预测和真实参数之间的差异。

误差值与准确性评估

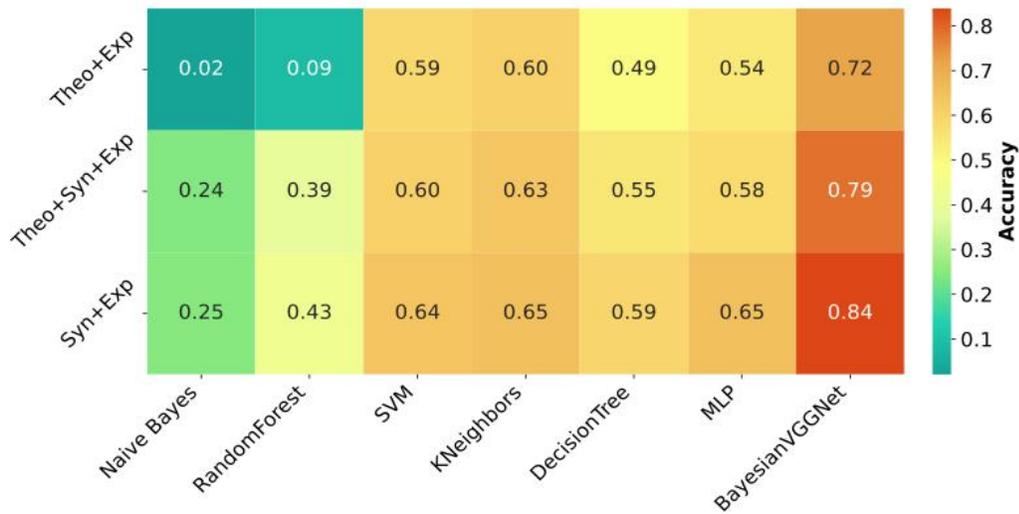


5.深度学习模型识别 X 射线衍射光谱



研究目标

构建一个深度学习模型，以实现XRD光谱的高准确性晶体结构分类，同时进行不确定性评估。此外，探讨模型的分类原理，提升可解释性，确保在实际应用中遵循基本物理原理。

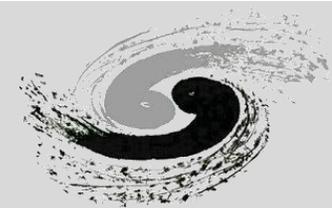


使用 B-VGGNet 进行 XRD 分类以及不确定性、可解释性分析框架

基于实验数据集验证的各种机器学习模型的分类准确率。

研究结果

1. 使用“模板元素替换”方法构建了全面的钙钛矿化学空间，93类空间群。
2. 所有经典机器学习模型在空间群分类任务中的准确率均小于70%。而B-VGGNet在所有数据集上的准确率至少提高了10%。



➤ 总结:

- I. 开发了SAXS高分辨率图像重建深度学习模型，提高了低曝光实验图像的信噪比；
- II. 开发了蛋白质轮廓参数预测机器学习模型，实现了快速准确的预测；
- III. 实现了高通量数据实时解析技术。

➤ 展望:

- 更多的实验数据，用来训练优化模型；
- 更多衍散射数据相关的科学问题研究。



中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

谢谢批评与指正!