
“拨云见日”重塑化学图像分辨率

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/28004.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

“拨云见日”重塑化学图像分辨率。拉曼光谱预处理

拉曼光谱学以无损无标记的方式原位探测样品化学组分分子键信息（概念解释>>>），在生物和临床医学研究中逐渐发展成一种新兴的光谱组学技术，并有望促进精准医疗的变革。目前拉曼光谱在生物学中的检测对象包括体液、外泌体、细胞/微生物和组织等，其中生物组织拉曼高光谱成像可提供化学键三维空间分布信息，有望成为生物和疾病发生发展分子机制筛查和研究工具。

然而拉曼散射信号微弱，因此容易受仪器、环境噪声和背景信号（非拉曼信号/基线）等干扰。这些噪声和基线信号的叠加严重影响拉曼光谱本征信号（拉曼峰）的分离和解析，限制了拉曼光谱的应用和普及，特别是跨仪器、跨样本应用。生物样本本身组分的复杂性和较强的荧光信号干扰使得拉曼光谱生物学应用举步维艰。同时具备高效光谱噪声去除和高保真基线校正能力的光谱预处理方法是高质量拉曼光谱应用的前提和挑战。

现状与挑战

传统的数值分析光谱预处理方法需要多次手工参数调整以达到较好的降噪和基线校正效果，不适用于临床诊断及高光谱图像处理等高通量应用场景。

现有深度学习预处理方法具有无需人工调参的优势，然而其噪声去除和基线校正光谱保真度仍有待提高，以实现跨仪器和跨样品的普适性应用。

常规监督深度学习光谱预处理方法的性能依赖高质量的标签数据训练。这种标签数据不含噪声和基线，其中噪声可以通过多次采集光谱取平均后消减；但是基线与拉曼信号同频率产生，难以通过常规仪器采集消除。另外，此方法训练的深度学习模型在跨仪器和跨样本使用时需要重新采集数据重新训练，泛化能力不足。

若采用数学模拟数据集进行训练，虽然可以跨仪器和跨样本使用，但由于实际仪器噪声和基线与数值模拟有差异，导致预处理模型光谱保真度仍然不足，无法实现生物医学样品拉曼高光谱图像的高保真降噪和基线校正；无法消除背景信号产生的虚假图像（直观体现在光谱生物静默区出现伪图像），从而释放光谱指纹区真实拉曼峰信号。

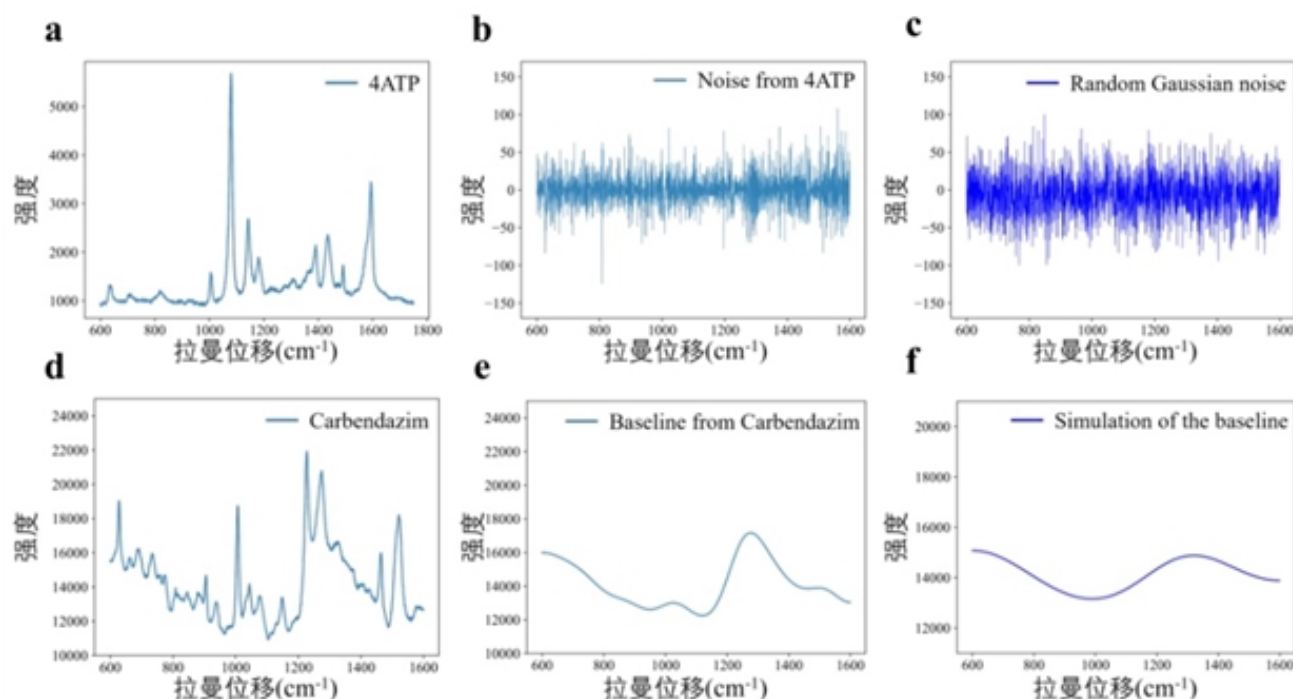


图1 实验采集光谱噪声(b)和基线(e)与仿真数据(c,f)的差异

具高保真高泛化能力的拉曼光谱预处理算法

近日，来自南方科技大学电子与电气工程系、光纤光缆先进制造与应用技术全国重点实验室、广东省集成光电子智感重点实验室沈平教授团队及其合作者提出一种自监督学习二步法拉曼光谱预处理策略（RSPSSL），实现跨仪器、跨样品和跨光谱类型的高保真光谱去噪和基线校正，并促进临床组织样本拉曼高光谱图像化学分辨率可视化。

文章以RSPSSL: A Novel High-fidelity Raman Spectral Preprocessing Scheme to Enhance Biomedical Applications and Chemical Resolution Visualization为题目，发表在Light：Science Applications，南方科技大学胡嘉祺博士生为论文第一作者、陈金娜副研究员（共一）与沈平教授为通讯作者。

该方案的第一步根据拉曼峰、噪声和基线之间物理关系的相互独立性建立自监督模型，对无标签训练光谱进行自我分解、重排及重构、并构建生成对抗网络获取大量带标签高仿真拉曼光谱，解决真实拉曼光谱无标签的问题。为了提高第二步光谱预处理模型RSBPCNN的泛化能力，无标签训练光谱采用来自多个实验室跨仪器、跨样本和跨光谱类型的多样化数据，以获取噪声和基线的多样性。

其次，为适应真实光谱数据的复杂性，预处理模型通过多个子模块端到端连接增强对复杂信号的拟合能力。通过第一步产生的多样化数据优化训练，该预处理模型RSBPCNN#可以用于来自任意仪器、样本类型和光谱类型的拉曼光谱预处理，无需人工干预或者再次训练。

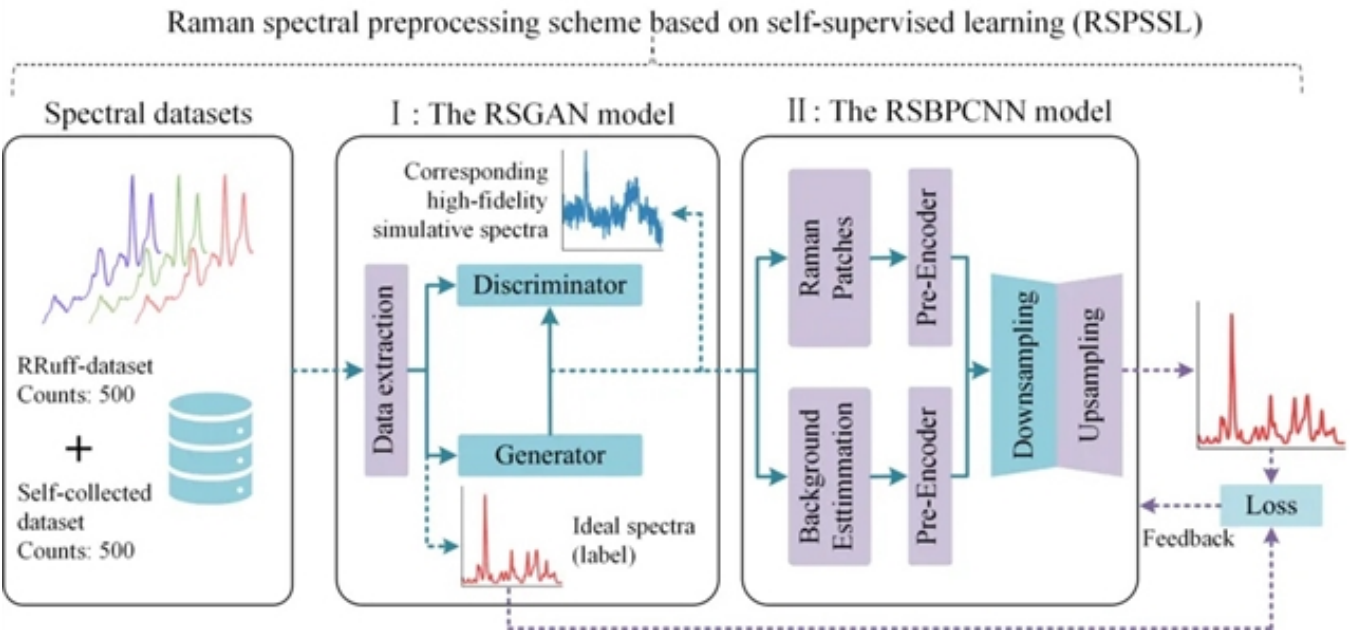


图2 基于自监督学习的拉曼光谱预处理策略（RSPSSL）

该预处理模型RSBPCNN#具有优异的噪声去除和基线校正能力，并且处理后的光谱保真度高。这种对不同信噪比微弱信号提取能力可减少采样时间，提高下游应用效果。

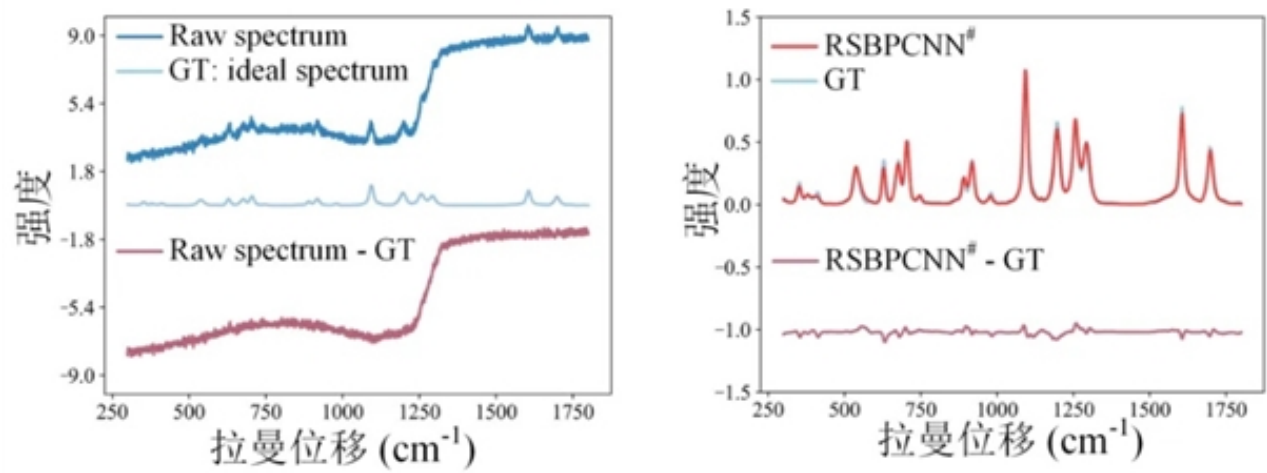


图3 高保真光谱预处理

高普适性拉曼光谱预处理

本研究通过多组实验验证该模型的普适性。在无需任何改动情况下，将该预处理模型RSBPCNN#直接应用于癌症诊断、除草剂浓度预测和高光谱成像中，显著提升其分类、定量准确性和图像质量。本方法在小样本情况下显著提高癌症诊断准确率和除草剂浓度预测精度，进一步验证了光谱预处理保真度。同时这些实验数据来自不同仪器及不同实验室，验证其跨仪器适应能力。

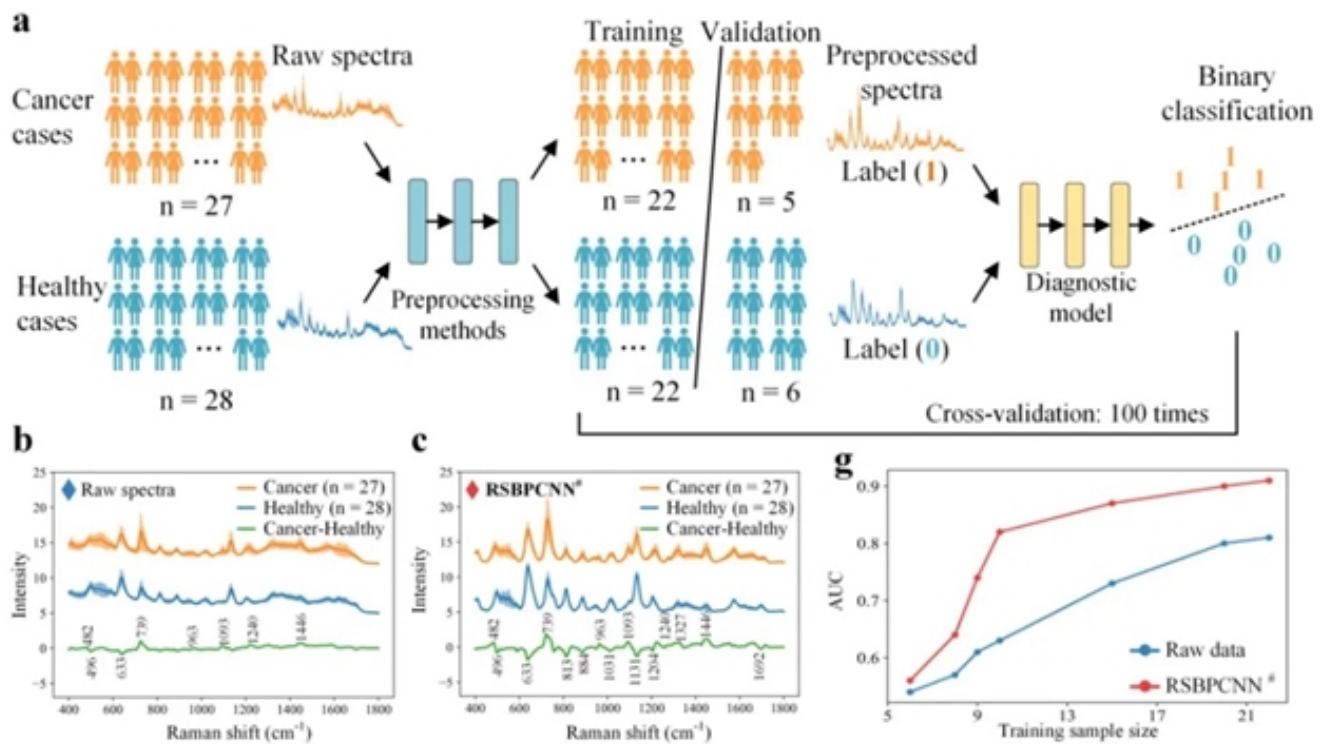


图4 预处理模型有效提升血清拉曼光谱诊断准确率

高光谱图像质量提升

拉曼高光谱图像最独特的能力是无标记化学成像。然而，微弱的生物成像信号由于叠加了基线信号而不可见。通过应用该预处理模型RSBPCNN#，拉曼峰强度被还原，重建出具有化学特异性的图像。同时该方法还可以显著提高信噪比，减少数十倍的采样时间。

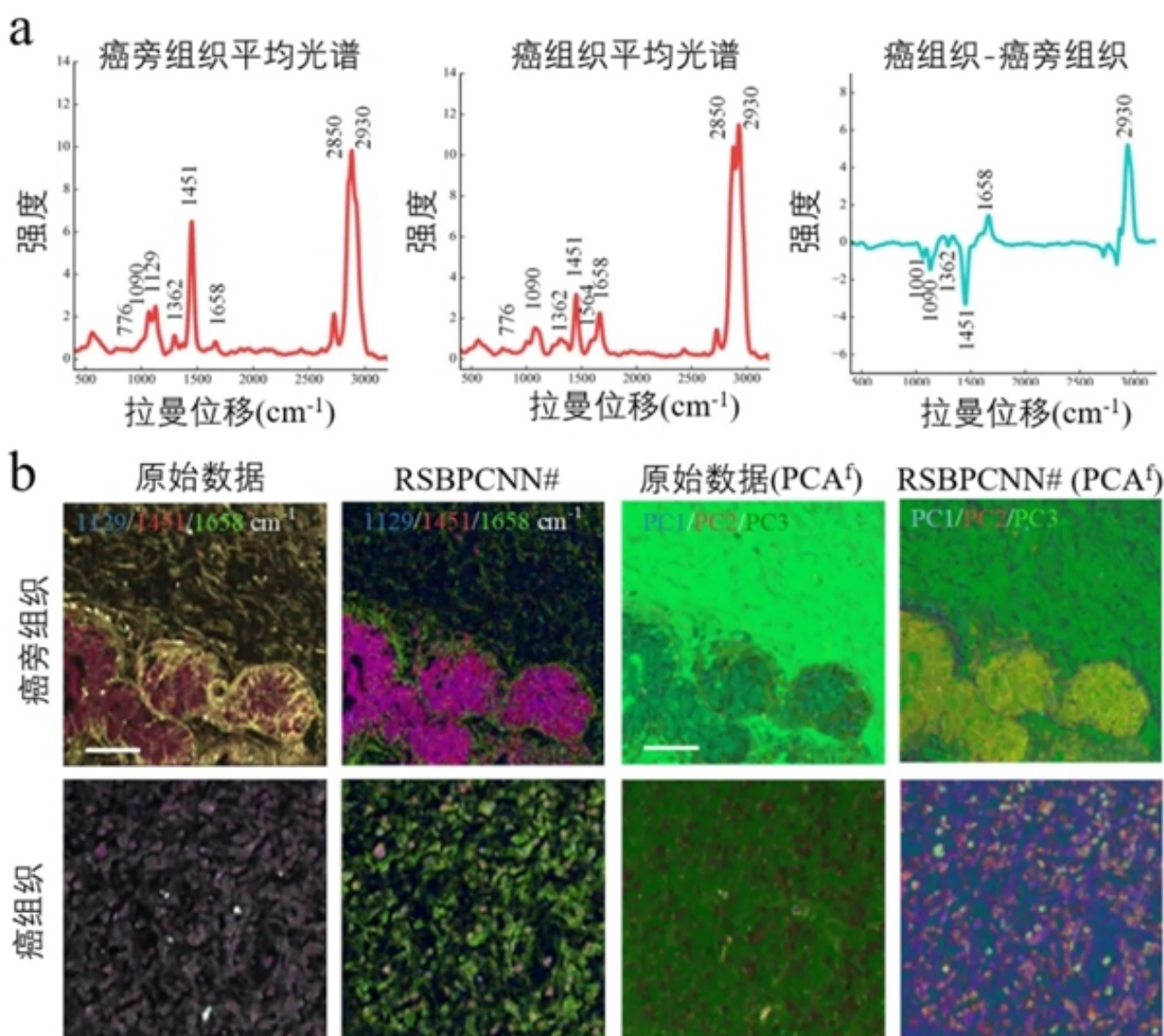


图5 临床组织高光谱图像预处理

总结与展望

本研究提出的自监督拉曼光谱预处理策略RSPSSL，通过精巧的算法模型设计，分离和重构多样化光谱特征，获取生成大量带标签高仿真光谱数据集，以此数据集训练优化高拟合能力的主预处理模型，得到通用型高鲁棒性的RSBPCNN#模型。该模型可以实现任意拉曼光谱的高通量高效噪声消除和基线校正。因其具有光谱高保真特性，在实验中显著提高癌症诊断和溶液浓度预测准确率，实现高光谱图像全谱质量提升，促进光谱指纹区化学图像分辨率可视化，同时体现出跨仪器、跨样本和跨光谱类型的广谱适用性。未来结合光谱空间分布关系可进一步提升高光谱图像分辨率，促进生物医学应用。

应用共享

本方法已集成到实验室共享平台，供科研使用。研究人员可通过批量载入拉曼光谱数据，实现快速光谱预处理（1900光谱/秒）。网址链接: <https://github.com/oilab-sustech/RSPSSL> .



图6 拉曼光谱预处理网页

该研究成果以RSPSSL: A novel high-fidelity Raman spectral preprocessing scheme to enhance biomedical applications and chemical resolution visualization为题在线发表在Light: Science Applications。（来源：LightScienceApplications微信公众号）

相关论文信息：<https://doi.org/10.1038/s41377-024-01394-5>

作者：沈平等 来源：《光：科学与应用》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发