

基于反向传播神经网络的拉曼光谱去噪方法

王 忠, 万冬冬, 单 闯, 李月娥, 周庆国*

兰州大学信息科学与工程学院, 甘肃 兰州 730000

摘 要 拉曼光谱技术作为一种典型的光学检测方法, 因其独特的非侵入性、快速、原位和极高的特异性, 在生物分析、疾病诊断及分子识别等众多领域得到广泛应用。拉曼光谱的指纹特性使其成为生物医学分析领域的重要工具, 但拉曼散射信号微弱, 数据处理分析大量依赖分析人员、自动化处理能力低等因素都会极大影响该技术在实际中的应用。实验设备、环境产生的噪声、待测生物样本的自发荧光等各种干扰因素使得高质量拉曼光谱数据的获得变得较为困难, 各种随机噪声会干扰拉曼光谱图中指纹谱峰信息的识别, 增大拉曼特征提取的难度, 因此, 噪声抑制在拉曼光谱预处理中显得十分重要。采用反向传播算法, 从数据本身的特征出发, 基于理想光谱片段和含噪光谱片段的线性差异, 分别构建噪声判定神经网络和光谱去噪神经网络模型。以随机生成的一系列洛伦兹峰数学模型叠加生成拉曼光谱, 对生成的拉曼光谱分别加入不同强度的噪声, 以此为实验数据, 对比新方法和经典滑动窗口均值法、Savitzky-Golay 滤波法、傅里叶变换法、小波阈值变换方法去噪的结果。对均方根误差和信号噪声比两个指标进行分析, 结果显示, 在低噪声干扰下所有去噪方法都能较好地完成任务, 但滑动窗口均值方法在光谱的边缘去噪效果会出现下降。随着噪声信号的增大, 滑动窗口均值方法、S-G 滤波方法、傅里叶变换方法去噪性能都出现明显下降。而基于反向传播神经网络的去噪方法要优于傅里叶变换法、滑动窗口均值法、S-G 滤波法, 同时该方法在避免复杂参数寻优设置的同时, 获得了和最优阈值小波变换方法近乎一致的去噪效果, 大大简化了参数设置, 更适合拉曼光谱去噪的自动化实现。

关键词 拉曼光谱; 神经网络; 光谱去噪

中图分类号: TN911.74 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)05-1553-08

引 言

拉曼光谱技术作为一种典型的光学检测方法, 因其独特的非侵入性、快速、原位和极高的特异性, 在生物分析、疾病诊断及分子识别等众多领域得到广泛应用^[1-5]。但拉曼散射信号微弱, 高质量拉曼光谱数据获取困难等因素都会极大影响该技术在实际中的应用。为了在后期的数据分析过程中尽量减少噪声、基线漂移等干扰因素对拉曼光谱数据分析的影响, 近年来, 科研人员作出了大量的努力和尝试。Tian 等考虑噪声出现的随机性, 将小波变换和 K-means 算法相结合去除拉曼光谱噪声, 大大提升了光谱去噪的自动化水平^[6]。Xi 等结合传统尺度相关性去噪法与软阈值法的优点提出了一种去噪改进算法, 使每个小波系数都有一个对应的阈值,

并将软阈值法的处理方式应用于各个小波系数, 从而实现小波系数的自适应处理^[7]。Leon-Bejarano 等基于温哥华算法 (vancouver Raman algorithm, VRA) 和经验模式分解 (empirical mode decomposition, EMD) 构建了基于生物样品的拉曼光谱去噪方法, 采用 EMD 代替均值滤波, 自动选择多项式阶数完成拟合^[8]。Zhao 等基于 Hilbert 振动分解方法, 将拉曼光谱分解为两个分量, 提出了一种基于特征提取的低信噪比拉曼光谱去噪方法^[9]。Sun 等提出了一种多尺度小波阈值去噪算法。拉曼信号被分解成不同的尺度, 每个尺度给出了拉曼信号中包含的不同频率相关信息^[10]。伴随着机器学习技术的蓬勃发展, Xie 等将线性稀疏编码去噪方法和机器学习模型相结合, 大大增强了算法的实用性^[11]。Xu 等提出了采用卷积神经网络获取图像细节, 反卷积网络恢复图片的去噪模型^[12]。Piriyatharawet 等提出了一种深度卷积多向长短

收稿日期: 2021-04-14, 修订日期: 2021-08-11

基金项目: 国家自然科学基金项目(61405083), 甘肃省自然科学基金项目(20JR5RA265), 兰州大学中央高校基本科研专项优秀青年教师创新项目(lzujbky-2018-129)资助

作者简介: 王 忠, 1978 年生, 兰州大学信息科学与工程学院副教授 e-mail: wangzhong@lzu.edu.cn

* 通讯作者 e-mail: zhouqg@lzu.edu.cn

期记忆网络模型,完成低光图像的噪声恢复^[13]。但这些适用于图像数据的神经网络模型不适合拉曼光谱数据,无法保护富含生物分子信息的拉曼谱峰信息。反向传播算法(back propagation, BP)的提出大大简化了人工神经网络的训练过程,不仅解决了非线性分类问题,而且在泛化能力、容错能力等方面都表现出了优良的性能,直到今天该方法仍然大量应用在研究和工程领域^[14-16]。针对拉曼光谱去噪问题,采用反向传播算法构建拉曼光谱判定和去噪网络,在提升去噪能力的同时,大大简化复杂参数的寻优设置,更适合自动化实现。

1 算法思路

拉曼光谱的谱图是由拉曼采集实验设备中电耦合器件数千像元采样点获得的原始数据绘制而来的谱线,由于各种干扰噪声的存在,使得原本微弱的拉曼信号在噪声的干扰下很难得到高质量的拉曼指纹谱图。理想光谱片段中所有点拥有一致的线性特征,这里基于理想光谱片段和含噪片段之间的差异构建基于 BP 的神经网络模型。

1.1 算法步骤

Step1: 确定去噪神经网络模型。

神经网络采用 $n-k-1$ 结构,即第一个参数为输入层节点数,将切分后的每一小段 X^p 作为输入数据,用段长 n 表示输入层节点个数;第二个参数为隐含层,长度为 k ;第三个参数为输出层,用一个节点表示该片段的线性预测值。隐含层使用式(1)中的 tansig 函数作为激活函数,输出层采用式(2)中的 purelin 函数作为激活函数,训练方法采用反向传播算法,进行 5 000 次训练。

$$\text{tansig}(x) = \frac{2}{1 + \exp(-2x)} - 1 \quad (1)$$

$$f(x) = x \quad (2)$$

Step2: 确定分类神经网络模型。

网络结构采用 $n-k-1$ 结构,前两个参数与去噪神经网络模型相同,第一个参数代表输入层节点数 n ,第二个参数为隐含层节点数 k 。第三层为输出层,这里关注的是分类结果含噪或者不含噪,所以输出层用一个节点表示该片段光谱是否含噪。隐含层使用 tansig 函数作为激活函数,输出层 logsig 函数作为激活函数,即

$$\text{logsig}(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

采用 5 000 次反向传播算法训练。

Step3: 数据预处理。

将长度为 L 的光谱数据切分成 m 段(每段 n 个数据点),剩余 d 个数据,则 $mn+d=L$ 。将每段含噪数据记为 X_j^p ,理想光谱数据记为 S_j^p , p 表示光谱是整个训练集的第 p 条光谱, i 表示段号, j 表示段内数据点。

本文对输入数据,先计算数据片段的均值,根据噪声的随机性,将数据整体减去该值,并且将均值记录保存。对于标签数据,计算出每一小段的线性差值。两种预处理方法描述如下:

① 预处理 1: 将片段含噪数据减去该段所对应的均值,并保存每个片段的均值,即

$$Y_j^p = X_j^p - \frac{1}{n} \sum_{j=1}^n X_j^p \quad (4)$$

② 预处理 2: 理想光谱片段首尾两点相减后除以 $n-1$,得到理想片段对应的线性差值 k^p ,即

$$k^p = S_n^p - S_1^p / n - 1 \quad (5)$$

Step4: 神经网络训练。

去噪神经网络训练用 Step3 中的预处理①对含噪数据进行预处理,得到含噪数据 Y_j^p ,用 Step3 中的预处理②对理想光谱数据进行预处理,得到理想光谱数据 k^p 。将 Y_j^p 作为输入, k^p 作为标签进行训练得到所需要的去噪神经网络。

分类神经网络训练用 Step3 中预处理①对含噪数据和理想光谱分别进行预处理,得到含噪片段 X_j^p ,理想光谱片段 S_j^p 。然后分别生成标签数据,将数据和标签放入 BP 神经网络进行训练,得到分类神经网络。

Step5: 预测和数据还原。

将待去噪光谱 S' 经过预处理分成 m 段(段长为 n)和剩余 d 个数据点($d < n$),得到分段数据 Y_j^i ,和每一段的均值 mean_i 。先将预处理后产生的第 i 段含噪光谱 Y_j^i 放入 Step4 所得的分类神经网络,判断该段是否需要去噪,如果不需要,则不对该段进行处理。否则,则将 Y_j^i 放入去噪神经网络进行预测,得到预测值 k_i 。再根据预测值进行谱线数据恢复,即

$$S'_{(i-1) \times n + j} = k_i \times \left(j - \frac{n+1}{2} \right) + \text{mean}_i \quad (6)$$

式(6)中, i 为段数, j 为段中每个点的下标, $j=1 \cdots n$, $S'_{(i-1) \times n + j}$ 为还原后光谱的第 $(i-1) \times n + j$ 个点, k_i 为第 i 段的预测值, mean_i 为第 i 段的均值。

1.2 性能指标

这里采用如下两个通用评价指标来对比分析各种去噪方法的性能。

① 均方根误差(root mean squared error, RMSE),定义为

$$\text{RMSE} = \sqrt{\sum_{n=1}^L [s_{\text{ideal}}(n) - s'(n)]^2 / L} \quad (7)$$

这里为了适配实际实验设备的采样指标,选择光谱采样数据点为 1 650 个,即 $L=1\ 650$, $s_{\text{ideal}}(n)$ 是未含噪的原始理想光谱, $s'(n)$ 是去噪后得到的光谱,该指标用来表征去噪后得到的光谱逼近原始理想光谱的程度,该值越小说明去噪效果越好,越接近真实光谱。

② 信号噪声比(signal-noise ratio, SNR),该指标可以表征去噪后光谱有效信息保留程度,定义为

$$\text{SNR} = 10 \log_{10} (E_{\text{ideal}} / E_{\text{noise}})$$

$$E_{\text{ideal}} = \sum_{n=1}^L (S_{\text{ideal}}(n) - \overline{S_{\text{ideal}}})^2 \quad (8)$$

$$E_{\text{noise}} = \sum_{n=1}^L (S_{\text{ideal}}(n) - S'(n))^2$$

其中, E_{ideal} 为信号能量; E_{noise} 为噪声能量; S_{ideal} 为理想光谱; S' 为去噪处理后的拉曼光谱; n 为拉曼光谱横坐标; L 为一

条拉曼光谱长度； $\overline{S_{ideal}}$ 为理想拉曼光谱均值。

2 实验部分

2.1 实验分析

通常实验中获得的拉曼光谱由三部分组成：拉曼指纹峰 y_r ，荧光背景 b 和随机噪声 n 构成，因此拉曼光谱可以表示为 $y = y_r + b + n$ 。其中拉曼光谱数据中的拉曼谱峰通常由一系列洛伦兹峰叠加产生，即

$$y_r = \sum_{j=1}^N \frac{2A_j}{p} \frac{\omega_{lj}}{4(\omega - \omega_{oj})^2 + \omega_{lj}^2} \quad (9)$$

式(9)中， ω 是拉曼位移， j 为第 j 个拉曼峰， A_j 为第 j 个拉曼峰面积， ω_o 是拉曼峰中心位置， ω_{lj} 是第 j 个拉曼峰宽， N 是拉曼峰数量。

2.2 实验数据

为了保证数据具有一定的代表性，接下来的实验数据中，每条光谱拉曼峰的数量在 1—6 之间选择，峰宽选择在 10—100 之间，峰面积选择在 2 000—20 000 之间，具体数值随机产生。这里含噪判定神经网络和去噪神经网络均采用 11-15-1 的 BP 神经网络模型。先用预处理①对 1 000 条含噪数据进行处理，得到 150 000×11 的输入矩阵。再用预处理②对 1 000 条理想光谱进行处理，得到 150 000×1 的标签矩阵。将输入矩阵和标签矩阵送入 11-15-1 的 BP 神经网络进行

训练，得到去噪神经网络。用预处理①对 1 000 条含噪数据和 1 000 条理想光谱进行处理，得到 300 000×11 的输入矩阵，再建立 300 000×1 的标签矩阵。0 代表含噪，1 代表不含噪。将输入矩阵和标签矩阵送入 11-15-1 的 BP 神经网络进行训练，得到噪声判定神经网络。

3 结果与讨论

3.1 单光谱对比

按照前述方法随机生成 1 条拉曼光谱，分别加入 20 和 30 db 高斯噪声后，分别用神经网络、滑动窗口平均值、S-G 滤波法、傅里叶变换、小波阈值进行去噪，各类方法在不同噪声状态下相关参数均已设置为最优，如表 1 所示。在小波阈值中 tptr 为阈值选择标准，sorh 为阈值函数选择方式，scal 为阈值处理随噪声水平的变化， n 为分解层数，wname 为选取的小波名称。

各种方法在不同噪声下的处理结果如图 1 和图 2 所示。可以得出在低噪声干扰下，所有去噪方法都能较好的完成任务，但随着噪声信号的增大，滑动窗口均值方法、S-G 滤波法、傅里叶变换方法去噪性能都出现明显下降，但本文提出的神经网络去噪方法和小波阈值变换方法保持了较稳定的去噪效果。

表 1 各种方法参数设置

Table 1 Parameter settings of various methods

	滑动窗口平均值		S-G 滤波法			小波阈值			
	窗口大小	窗口大小	阶数	tptr	sorh	scal	n	wname	
20 db	15	21	3	heursure	s	sln	8	sym7	
30 db	30	21	2	sqtwolog	h	mln	6	sym5	

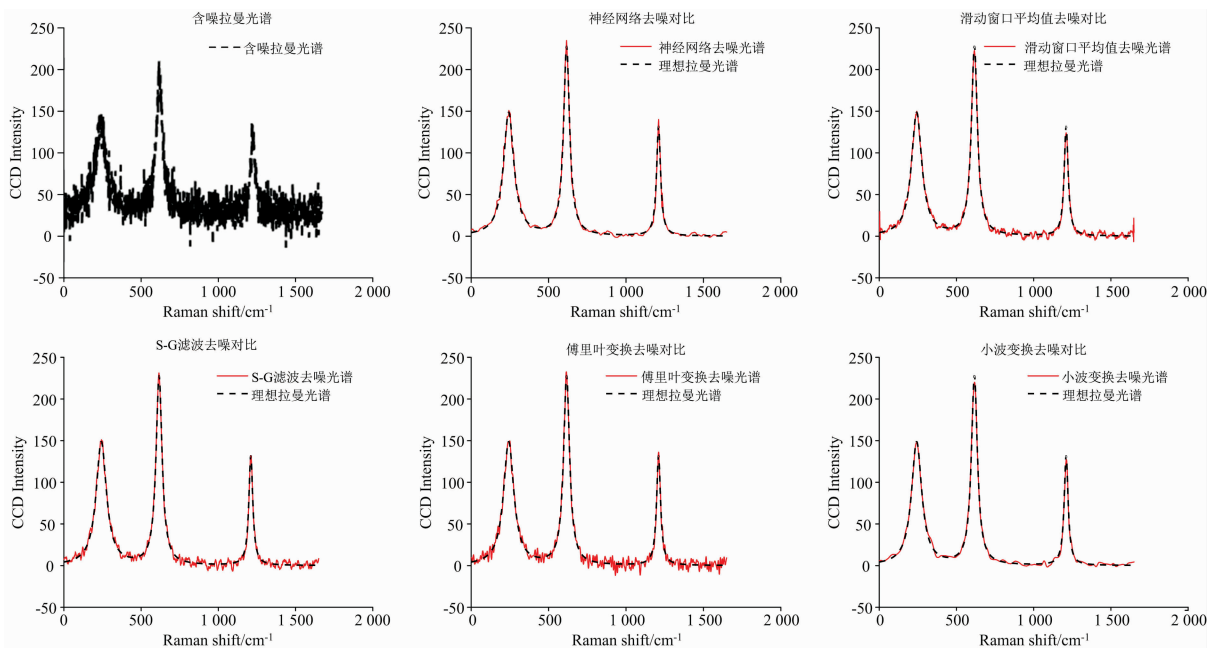


图 1 20 db 噪声下各种方法去噪效果对比

Fig. 1 Comparison of different denoising methods for 20 db noise

各种方法在不同噪声干扰下性能指标 RMSE 值和 SNR 值的变换如表 2 所示。通过比较可以得出，本文提出的神经

网络方法在 RMSE 指标和 SNR 指标方面，去噪后的得分都优于滑动窗口均值方法、S-G 滤波方法、傅里叶变换方法。

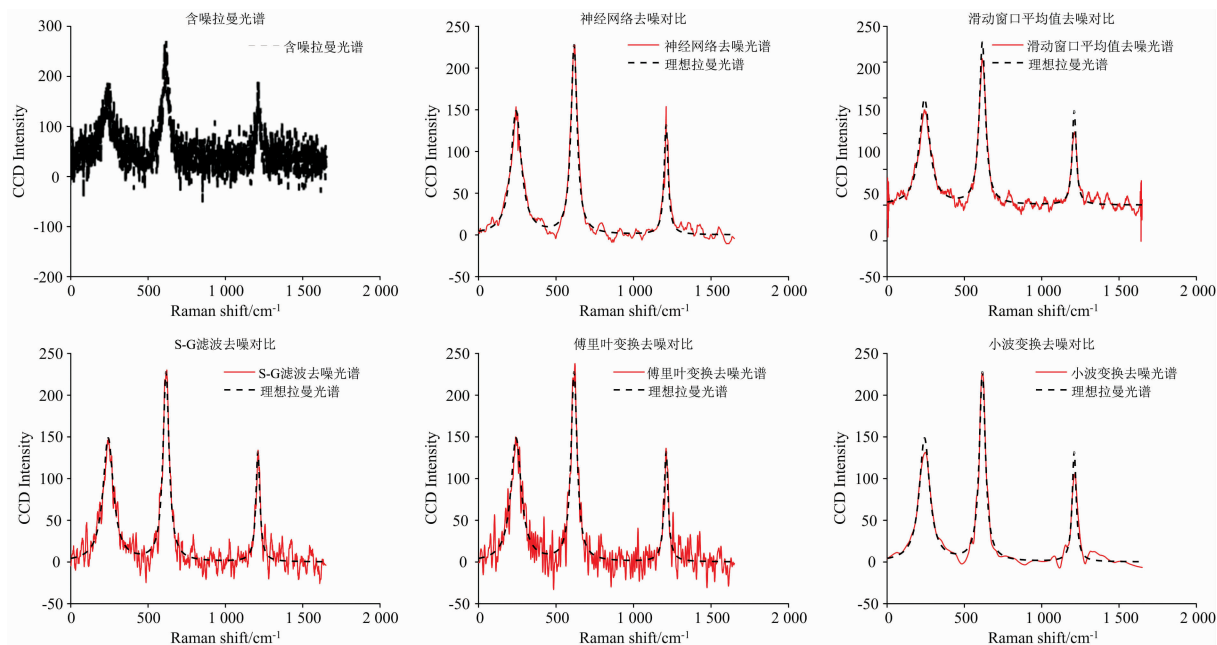


图 2 30 db 噪声下各种方法去噪效果对比

Fig. 2 Comparison of different denoising methods for 30 db noise

表 2 各种方法去噪性能指标 RMSE 和 SNR 对比

Table 2 Comparison of various methods for RMSE and SNR

		原始含噪	神经网络	滑动窗口均值	S-G 滤波	傅里叶变换	小波阈值
20 db	RMSE	9.719 0	2.616 4	2.990 8	3.372 4	4.379 9	2.317 7
	SNR	12.570 9	23.969 1	22.807 7	21.764 7	19.494 0	25.022 2
30 db	RMSE	31.605 3	6.772 7	8.127 8	11.054 5	13.855 8	6.183 7
	SNR	2.328 1	15.708 2	14.123 9	11.452 5	9.490 7	16.498 4

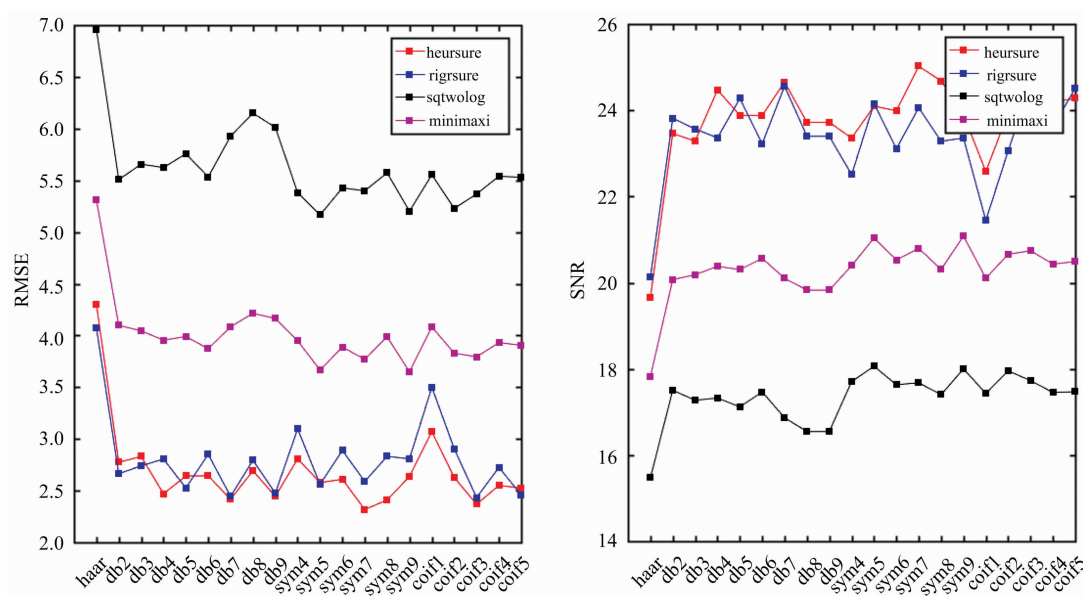


图 3 20 db 噪声下不同小波，不同阈值选择标准 RMSE 指标和 SNR 指标变化

Fig. 3 RMSE and SNR changes of different wavelets and different thresholds for 20 db noise

本文方法和小波阈值变换方法性能基本一致，但实验中小波阈值参数进行了最优选择，每次都需要重新寻找最优参数进行调整。如果选择不同的小波参数，产生的结果会有很大的差距，如图 3 和图 4 所示。当选取不同小波时，RMSE 和

SNR 指标都出现明显的变化。选取不同的阈值标准时，两种指标变化也很明显，这些都直接表明选择参数标准不同，去噪的准确性会受到很大的影响。

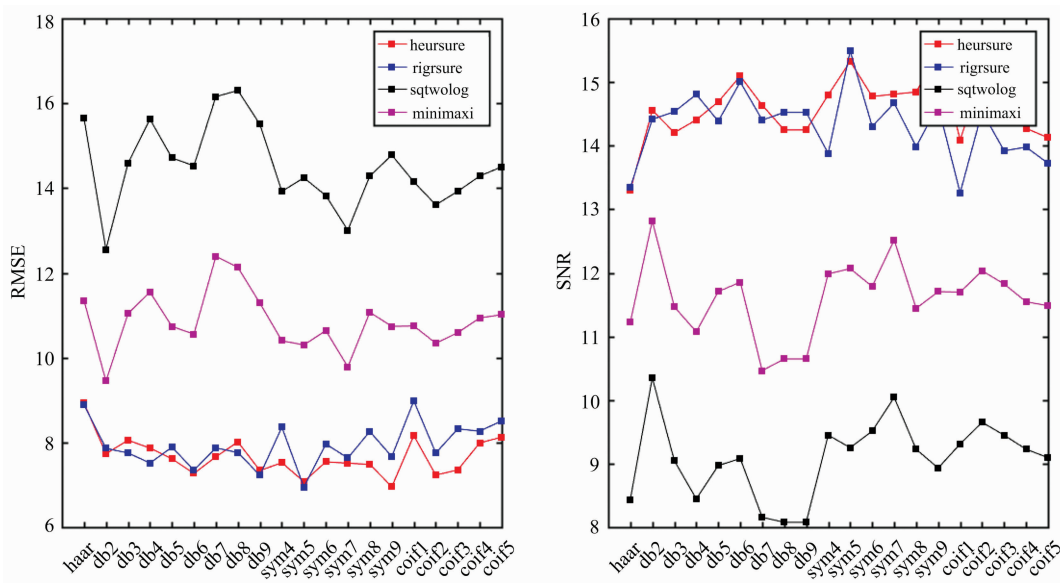


图 4 30 db 噪声下不同小波，不同阈值选择标准 RMSE 指标和 SNR 指标变化

Fig. 4 RMSE and SNR changes of different wavelets and different thresholds for 30 db noise

图 5 表明，对于不同噪声水平的光谱，最优分解层数也可能不同，层数过多会造成光谱信号中细节信息的丢失。小波函数、阈值选择标准，分解层数对于具有相同理想谱线，不同水平噪声的光谱数据的影响程度不同。换句话说，即使

知道该谱线在特定噪声水平下的理想小波参数，我们也很难预测一条谱线的全局最优理想小波参数是什么。因此在实际采集的光谱数据中，由于无法获得未含噪的对应理想光谱，所以想要选取最优小波参数是非常困难的事情。

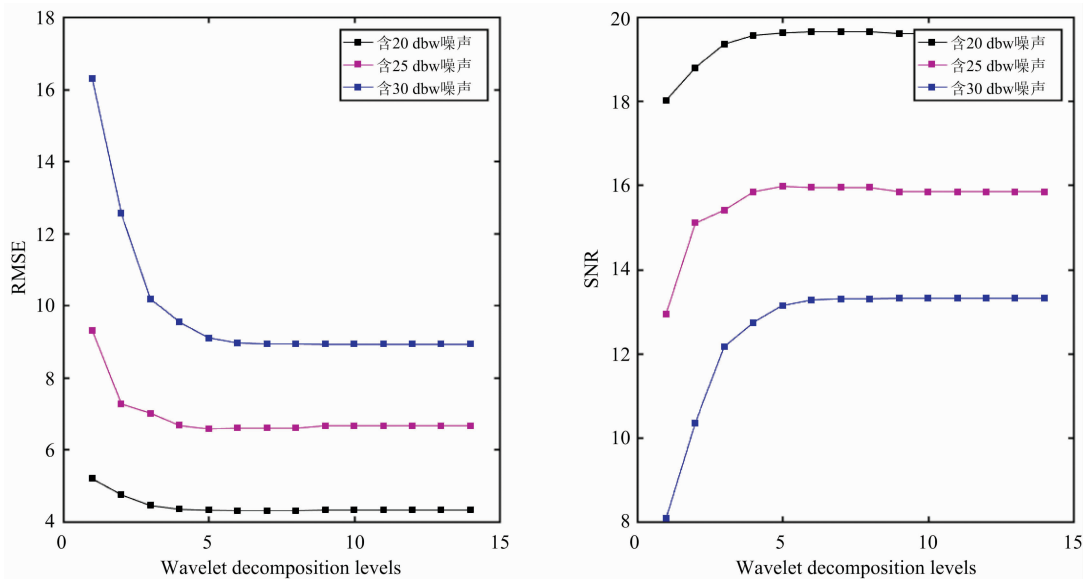


图 5 不同噪声水平下，不同小波分解层数 RMSE 指标和 SNR 指标变化

Fig. 5 RMSE and SNR changes of different wavelet decomposition levels for different noise

3.2 100 组光谱对比

为得出更加全面的比较结果，接下来随机生成 100 组光

谱进行整体性能分析，分别加入 20 和 30 db 的随机噪声。结果如图 6 和图 7 所示。横坐标表示每条光谱的编号，纵坐标

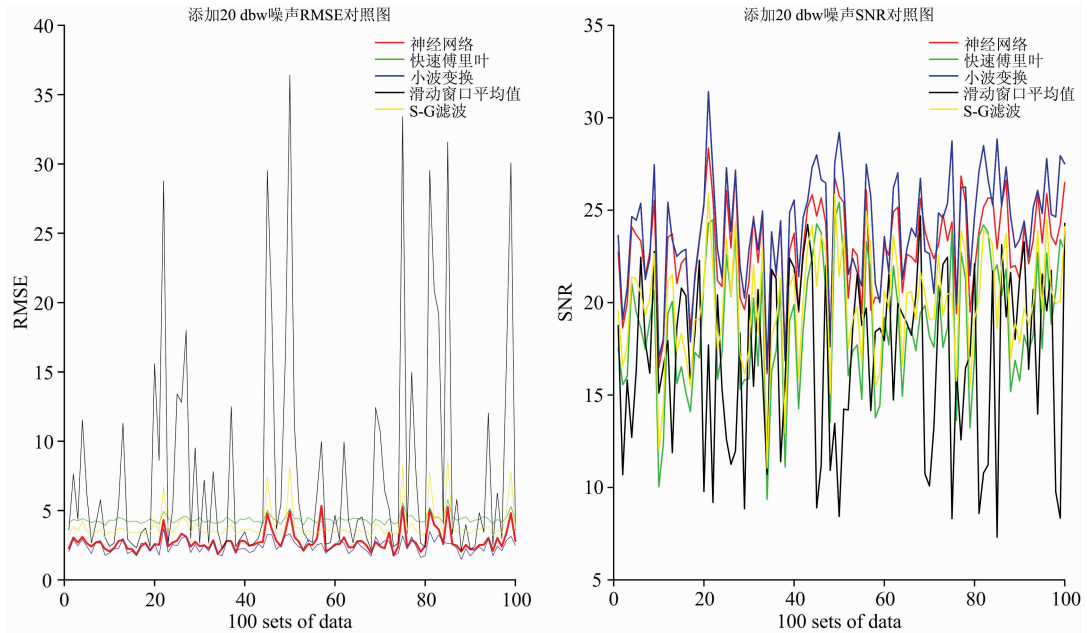


图 6 20 db 噪声中不同去噪方法的 RMSE 指标和 SNR 指标变化

Fig. 6 RMSE and SNR changes of different denoising methods for 20 db noise

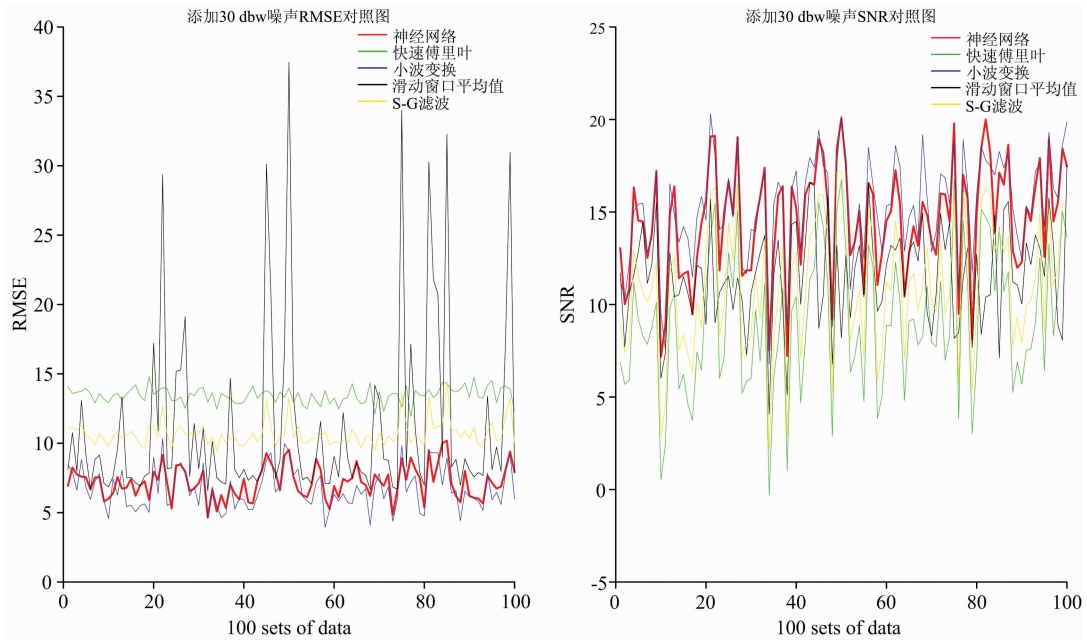


图 7 30 db 噪声中不同去噪方法的 RMSE 指标 SNR 指标变化

Fig. 7 RMSE and SNR changes of different denoising methods for 30 db noise

表示对应光谱在相应去噪算法下计算得到的 RMSE 和 SNR 评价指标值。

从上述图中,我们可以观察到在所有噪声水平下,神经网络去噪和小波阈值去噪的 RMSE 值小于其他三种方法而 SNR 值大于其他三种方法,显示了较好的去噪效果。表 3 给出了所有方法 RMSE 和 SNR 指标均值数据。表中 \overline{RMSE} 表示 100 条光谱的 RMSE 指标平均值, \overline{SNR} 表示 100 条光谱的 SNR 指标平均值, S 为神经网络去噪方法在 100 条光谱中

RMSE 指标和 SNR 指标得分优于其他方法的数量,包括指标值得分相同的情况。

在加入 20 db 噪声的数据中,共有 99 个点优于傅里叶变换去噪,21 个点优于小波阈值去噪,96 个点优于滑动窗口平均值法去噪,99 个点优于 S-G 滤波器法去噪。在加入 30 db 噪声的数据中,所有点都优于傅里叶变换去噪、S-G 滤波器法去噪,93 个点优于滑动窗口均值法去噪、27 个点优于小波阈值去噪。

表 3 各种方法 RMSE 和 SNR 指标对比
Table 3 Comparison of various methods for RMSE and SNR on different Noise

		神经网络	傅里叶变换	小波阈值	滑动窗口均值	S-G 滤波器
20 db	\overline{RMSE}	2.787 9	4.362 6	2.434 4	7.745 6	3.912 1
	\overline{SNR}	22.950 8	18.819 3	24.031 6	16.958 0	19.989 8
	S		99	21	96	99
30 db	\overline{RMSE}	7.197 4	13.460 8	6.692 8	11.268 4	10.688 7
	\overline{SNR}	14.565 0	9.020 3	15.285 0	11.543 0	11.043 6
	S		100	27	93	100

3.3 光谱去噪

为了进一步验证方法的有效性，对实验采集的癌变组织样品含噪拉曼光谱进行了去噪处理，实验仪器采用德国 WiTec 公司的共聚焦显微拉曼光谱系统，光斑直径大约在 1~2 μm 之间，采用面扫描采集样品区域 20 μm×20 μm，单

点采样积分时间 5 ms，区域采样点不少于 100。由于实验光谱本身包含噪声信息，无法获得理想未含噪光谱，这里只能和经验丰富的实验员人工处理后结果相比，统一进行基线校正后，随机抽取一组数据完成去噪实验，结果如图 8 所示，在提升处理效率的同时大大降低了人工处理的复杂度。

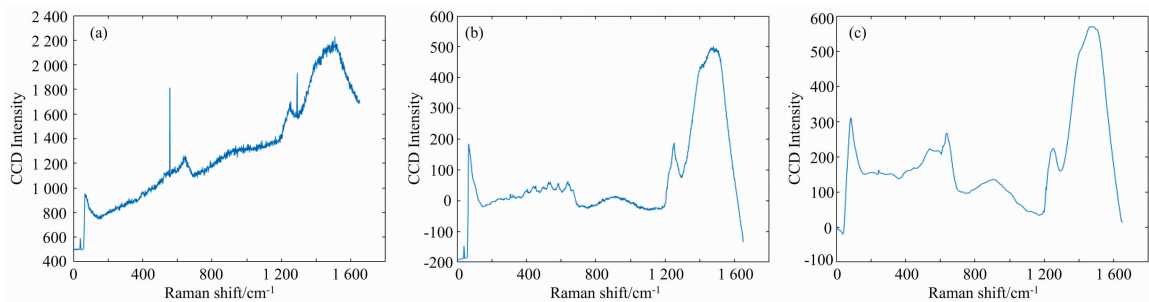


图 8 实验样品拉曼光谱

(a): 含噪光谱; (b): 人工去噪光谱; (c): BP 网络去噪光谱

Fig. 8 Raman Spectrum of experimental sample

(a): Noisy Raman spectrum; (b): Denoised spectrum by analyst; (c): Denoised spectrum by BP network

4 结 论

从整体情况来看，基于反向传播算法的神经网络去噪要优于傅里叶变换、滑动窗口均值法、S-G 滤波器法，但稍逊于最优参数的小波阈值去噪，但性能差距并不大。从 RMSE 均值可以看出，对于加入 20 db 噪声的数据，神经网络去噪相比小波阈值去噪相差 0.353 5；对于加入 30 db 噪声的数据，神经网络去噪相比小波阈值去噪相差 0.504 6。

但是神经网络去噪方法避免了复杂的参数设置，使得它的适应力更强。在实验采集的癌变生物组织拉曼光谱图中，前述方法能较为稳定地处理干扰噪声，但在局部细节处和经

验丰富的实验员分析还存在着一些差异。这里主要体现在实验分析人员会根据待测样本可能存在的成分信息重点关注某些区域的细节，保留某些主观认定的重要区域，而算法程序处理完全从数据本身的特征出发来给出判定，没有预期的经验作为辅助。实际生物样品的数据采集集中，扫描区域往往包含数百上千的点扫描光谱，数据处理分析任务繁重，快速的去噪方法可以很好地帮助实验分析人员进行数据特征分析和判定，减少其工作负荷，进一步提高处理效率。由于噪声的存在，我们无法准确获得它的理想谱线，因而也就很难选取最优小波参数进行小波阈值设置。相反，神经网络去噪方法在拥有和最优小波阈值去噪一致性能的前提下，大大简化了参数设置，更适合拉曼光谱去噪的自动化实现。

References

[1] Cicchi R, Cosci A, Rossari S, et al. J. Biophotonics, 2014, 7(1-2): 86.
 [2] Lin K, Zheng W, Lim C M, et al. Biomed. Opt. Express, 2016, 7(9): 3705.
 [3] McKinney S M, Sieniek M, Godbole V, et al. Nature, 2020, 577(7788): 89.
 [4] Schleusener J, Gluszczynska P, Reble C, et al. Exp. Dermatol., 2015, 24(10): 767.
 [5] LIANG Hao-yue, CHEN Xue-lian, YANG Wan-zhu, et al(梁昊岳, 程雪莲, 杨晚竹, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2020, 40(12): 3670.
 [6] Tian Y, Burch K S. Appl. Spectrosc., 2016, 70(11): 1861.

- [7] Xi Y, Li Y, Duan Z, et al. *Appl. Spectrosc.*, 2018, 72(12): 1752.
- [8] Leon-Bejarano F, Mendez M O, Ramirez-Elias M G, et al. *Appl. Spectrosc.*, 2019, 73(12): 1436.
- [9] Zhao X Y, Liu G Y, Sui Y T, et al. *Spectrochim. Acta A: Mol. Biomol. Spectrosc.*, 2021, 250: 119374.
- [10] Sun P, Shi Y, Li L, et al. *International Conference on Optical Instruments and Technology-Optoelectronic Imaging/Spectroscopy and Signal Processing Technology*, Beijing, 2019.
- [11] Xie Junyuan, Chen Enhong. *Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012.
- [12] Xu L, Ren J S J, Liu C, et al. *Proceedings of 27th International Conference on Neural Information Processing Systems*, 2014. 1790.
- [13] Piriayatharawet T, Kumwilaisak W, Lasang P. *18th International Symposium on Communications and Information Technologies (ISCIT)*, 2018. 90.
- [14] Kosko B, Audhkhasi K, Osoba O. *Neural Networks*, 2020, 129: 359.
- [15] Kim J, Kwon D, Woo S Y, et al. *Neurocomputing*, 2021, 428: 153.
- [16] Chen C, Wang H, Yuan F, et al. *Neural Computing and Applications*, 2020, 32(14): 10435.

A Denoising Method Based on Back Propagation Neural Network for Raman Spectrum

WANG Zhong, WAN Dong-dong, SHAN Chuang, LI Yue-e, ZHOU Qing-guo*

School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China

Abstract As a typical optical detective method, Raman spectroscopy has been widely used in many fields such as biological analysis, disease diagnosis and molecular recognition due to its unique non-invasive, fast, in-situ and extremely high specificity. The fingerprint characteristics of Raman Spectroscopy make it an important tool in the biomedical field. However, many problems greatly influence the application of this technology, including weak Raman scattering signal, the large dependence on analysts for data processing and analysis, and the low capabilities of automated processing. It is usually difficult to obtain an effective and stable Raman spectrum of various interference factors, including the noise from experimental equipment or environment and spontaneous fluorescence of the biological sample. All kinds of random noises will interfere with identifying fingerprint peak information in Raman spectroscopy and increase the difficulty of Raman feature extraction, so denoising is an important task in the preprocessing of Raman Spectroscopy analysis. In this paper, using the backpropagation algorithm and considering the characteristics of the data, we built the noise judgment and denoising neural network model respectively, based on the linear difference between the ideal spectral segment and the noisy spectral segment. For the simulation experiments, the data of Raman spectra are composed of a series of randomly generated mathematical models of the Lorentz peak. Different intensities of noises were added to the randomly generated single Raman spectrum and 100 groups of Raman spectra, respectively. In addition, the new method has been compared with the classic sliding window average method, Savitzky-Golay filter method, Fourier transforms, and wavelet threshold transform method. Using the Root mean square error and Signal to noise ratio indicators for analysis, the results show that all methods can complete the denoising task under low noise, but the denoising effect of the sliding window average method is obviously reduced at the edge of the spectrum. With the increase of the noise signal, the denoising performance of the sliding window average method, S-G filtering method, and Fourier transform method have all decreased significantly. Overall, the denoising method based on a backpropagation neural network is better than the Fourier transform method, sliding average window method, and S-G filter method. This method avoids complex parameter optimization settings and at the same time, obtains the denoising effect that is almost the same as the optimal wavelet transform method. It greatly simplifies the parameter setting and is more suitable for the automated denoising of the Raman spectrum.

Keywords Raman spectroscopy; Neural network; Spectroscopy denoising

(Received Apr. 14, 2021; accepted Aug. 11, 2021)

* Corresponding author