

# 基于拉曼光谱和 DCGAN 数据增强的珍珠粉掺伪检测研究

谈爱玲<sup>1</sup>, 楚振原<sup>1</sup>, 王晓斯<sup>1</sup>, 赵勇<sup>2\*</sup>

1. 燕山大学信息科学与工程学院, 河北省特种光纤与光纤传感重点实验室, 河北 秦皇岛 066004

2. 燕山大学电气工程学院, 河北省测试计量技术及仪器重点实验室, 河北 秦皇岛 066004

**摘要** 珍珠粉和珍珠层粉化学成分相似, 但是珍珠层粉的药用价值远低于珍珠粉, 并且珍珠层粉制备容易, 成本低, 常被不法商家用于冒充或掺入珍珠粉中流入市场, 谋取利益。因此, 对珍珠粉掺伪鉴别和纯度检测具有重要的意义。采用激光拉曼光谱结合深度学习研究珍珠粉掺伪快速鉴别和纯度分析。将纯珍珠粉和珍珠层粉按一定比例混合, 制成珍珠粉质量百分数分别为 0%, 25%, 50%, 75%, 80%, 85%, 90%, 95% 与 100% 共 9 种纯度 270 个模拟掺伪珍珠粉样本。然后对样本进行拉曼光谱采集, 参数设置如下: 分辨率为  $4.5 \text{ cm}^{-1}$ , 积分时间为 3 000 ms, 激光功率为 20 mW。搭建了深度卷积生成式对抗神经网络(DCGAN)模型, 对样本拉曼光谱进行数据增强; 在此基础上, 结合 K 近邻(K-nearest neighbor)、随机森林(random forest)、决策树(decision tree)、一维卷积神经网络(1D-CNN)4 种分类器, 对纯度为 85%, 90%, 95% 与 100% 的小比例掺伪样本进行真伪鉴别分析; 同时, 结合一维卷积神经网络对 9 种纯度的珍珠粉掺伪样本建立纯度预测的定量模型。结果表明: 基于 DCGAN 数据增强方法所生成的拉曼光谱, 与原始光谱相比, 在峰值信噪比和结构相似度两个评价指标上均明显优于传统数据增强方法; 在珍珠粉掺伪定性鉴别方面, DCGAN 增强后的数据分别送入 4 种分类器, 对 4 种小比例掺杂样本的真伪鉴别正确率均达到 100%; 在对 9 种掺伪纯度样本纯度检测方面, 对测试集样本, DCGAN-1DCNN 方法所建纯度定量预测模型性能最优, 其决定系数  $R^2$  为 0.988 4, 预测均方根误差 RMSEP 为 0.034 8, 一维卷积神经网络的损失值 Loss 为 0.001 2, 定量模型拟合最好。拉曼光谱结合 DCGAN 算法为珍珠粉掺伪鉴别及纯度检测提供一种快速简便的方法。深度卷积生成式对抗网络的数据增强方法在光谱分析技术领域具有重要的研究意义和应用价值。

**关键词** 生成式对抗网络; 拉曼光谱; 深度学习; 珍珠粉掺伪; 一维卷积神经网络

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)03-0769-07

## 引言

珍珠粉是将双壳贝类动物受刺激所产的珍珠加工而制成的粉状物。珍珠粉作为一种名贵中药材, 在药用和美容方面都具有极高的价值, 在中国已有两千余年药用历史, 具有治疗溃疡、养肝明目、辅助降压、补充微量元素、润肤祛斑之功效<sup>[1]</sup>。珍珠层粉则是珍珠母贝内层经粉碎磨细而成的粉状物。两者成分较相似, 但因其内部有机质成分不同, 因此药用功效不同; 此外, 珍珠层粉制备比较容易, 成本较低, 常被不法商家用于冒充珍珠粉或掺入珍珠粉中, 谋取利益。目前, 珍珠粉和珍珠层粉的质量鉴别多采用 X 射线衍射法、显微红外光谱、电感耦合等离子体等分析方法<sup>[2-4]</sup>, 这些方法通

常需要进行样品热处理、操作繁琐、分析时间长。因此, 对珍珠粉进行快速、便捷的真伪鉴别和质量评价具有重要的意义。

激光拉曼光谱是基于拉曼散射效应的分子结构表征技术, 其谱线位置、谱带强度等可反映物质成分等信息, 无需对样品进行前处理, 具有操作简便、测定时间短、灵敏度高等优点, 已经广泛应用于诸多行业<sup>[5-7]</sup>。在中草药产品分析和检测领域, 明晶等采用拉曼光谱技术对鱼脑石、鹅管石、花蕊石、南寒水石 4 种中药材进行定性鉴别<sup>[8]</sup>。另外, 拉曼光谱技术也可以实现中药材的定量分析, 王清蓉等采用拉曼光谱法对丹参根鲜样的不同植物组织进行原位检测, 发现丹参酮类成分的分布与周皮细胞颜色分布之间呈正相关性<sup>[9]</sup>。

收稿日期: 2021-02-07, 修订日期: 2021-04-19

基金项目: 国家重点研发计划项目(2019YFC1407904), 河北省自然科学基金项目(2020203010)和国家自然科学基金项目(61971373)资助

作者简介: 谈爱玲, 1978 年生, 燕山大学信息科学与工程学院副教授 e-mail: tanailing@ysu.edu.cn

\* 通讯作者 e-mail: zhaoyong@ysu.edu.cn

本文基于拉曼光谱结合深度学习对掺伪珍珠粉进行检测。深度学习神经网络需要大量数据训练,通常情况下需要对原始数据进行扩充<sup>[10]</sup>。Esben 等针对一维光谱数据提出了左右平移、叠加噪声和平移噪声组合三种方法进行增强<sup>[11]</sup>。Frédéric 等提出相邻像素法进行拉曼光谱数据增强<sup>[12]</sup>。赵勇等提出线性叠加方法扩充拉曼光谱数据<sup>[13]</sup>。目前的方法没有充分利用原始样本的内部特征,导致增强的数据模型具有局限性。针对这一问题,本文提出基于深度卷积生成式对抗网络(deep convolutional generative adversarial networks, DCGAN)的拉曼光谱数据增强方法,得到大量与原始光谱数据集独立同分布的样本,并在此基础上结合一维卷积神经网络算法,建立掺伪珍珠粉真假鉴别的判别模型和纯度定量分析模型,为拉曼光谱技术结合深度学习在掺伪珍珠粉检测中的应用提供科学依据。

## 1 实验部分

实验样品珍珠粉(淡水)和珍珠层粉(淡水)均选购于北京同仁堂(亳州)饮片有限责任公司。将纯珍珠粉与纯珍珠层粉按一定比例混合以模拟掺杂珍珠层粉的掺伪珍珠粉,珍珠粉

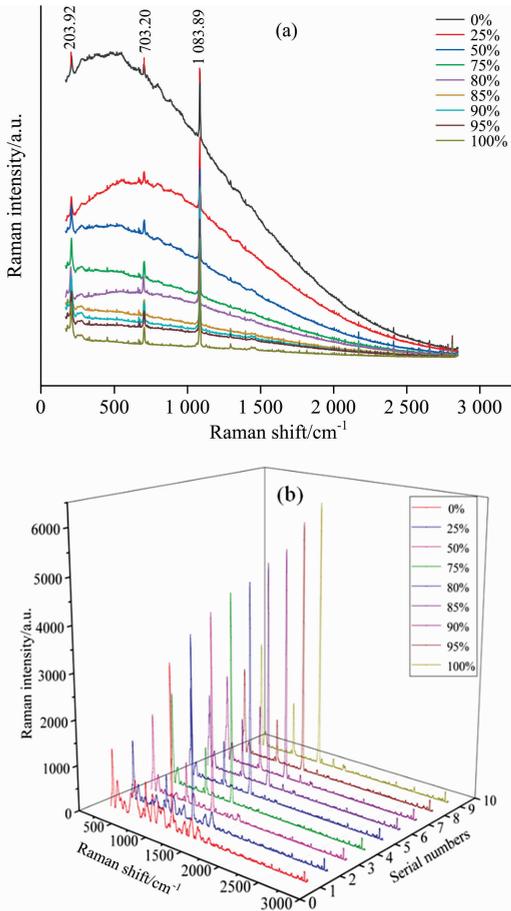


图 1 样本的平均拉曼光谱图

(a): 原始光谱; (b): 预处理后的光谱

Fig. 1 Mean Raman spectra of the samples

(a): Original spectra; (b): Spectra with pretreatment

在掺伪样本中的质量分数定义为珍珠粉纯度。同时,将纯珍珠粉纯度定义为 100%,纯珍珠层粉纯度定义为 0%,掺伪样本纯度由掺杂比例计算得到,控制掺杂比例使掺伪样本纯度在 25%~95% 范围内按 5% 的纯度梯度分布,因此共制得纯度为 0%, 25%, 50%, 75%, 80%, 85%, 90%, 95% 和 100% 共 9 种掺伪珍珠粉样本。

实验所用拉曼光谱仪型号为 i-RAMAN<sup>®</sup> plus 785S,激发光源波长为 785 nm,光谱范围为 166~2 851  $\text{cm}^{-1}$ ,分辨率为 4.5  $\text{cm}^{-1}$ ,最大功率为 300 mW。拉曼光谱采集激光功率设定为 20 mW,积分时间设为 3 000 ms,为避免杂散光影响,全部测量在暗室内进行。每种纯度配置 30 个样本,共采集 270 个拉曼光谱,每个样本采集 3 次光谱取平均值。9 种纯度样本原始平均拉曼光谱如图 1(a)所示。光谱信号的预处理可消除环境噪声、荧光背景等多种干扰因素的影响,提高定性分析和定量模型结果的稳定性。本文采用基线校正与平滑滤波对所有样本的拉曼光谱进行预处理,预处理后拉曼光谱如图 1(b)所示。

由图 1(a)可以看出,所有样本在 203.92, 703.20 与 1 083.89  $\text{cm}^{-1}$  处具有强度较高的特征峰,包含丰富的所含成分信息。纯珍珠粉拉曼光谱的荧光背景最低,纯珍珠层粉拉曼光谱荧光背景最高,随着掺伪比例的增加,拉曼光谱的荧光背景逐渐增高。由图 1(b)可以看出,纯珍珠粉的拉曼信号强度最大,随着掺伪比例的上升,样本信号强度总体呈下降趋势。另外,纯珍珠层粉在更多拉曼位移处有特征峰,是珍珠层粉与珍珠粉有机质成分不同所导致。

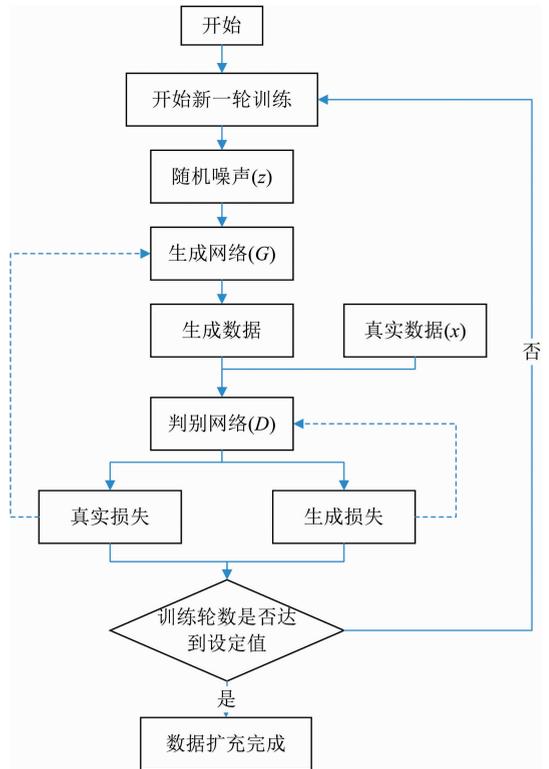


图 2 生成式对抗网络流程图

Fig. 2 Generative adversarial network flow chart

## 2 深度卷积生成式对抗神经网络

### 2.1 生成式对抗网络基本原理

蒙特利尔大学的 Goodfellow 等在 2014 年提出的生成式对抗网络 (generative adversarial net, GAN) 是生成效果非常好的一种网络模型<sup>[14]</sup>。GAN 主要由生成器与判别器组成, 具体执行流程如图 2 所示。

首先, 设置一个输入噪声  $z$ , 噪声符合随机分布, 然后通过生成器生成新的数据。生成器  $G$  的计算公式如式(1)所示

$$\min_G V(D, G) = E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

式(1)中,  $E$  为数学期望,  $P_z(z)$  为高斯噪声分布,  $D(G(z))$  为判别器对于输入伪造样本的输出概率。训练的目的是让  $G$  的值尽可能的小。生成器生成的新数据与原始数据  $P_{data(x)}$  输送到判别器  $D$  当中, 由判别器分辨生成数据的真伪, 判别器  $D$  的计算公式如式(2)所示

$$\max_D V(D, G) = E_{x \sim P_{data(x)}} [\log(D(x))] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (2)$$

式(2)中,  $D(x)$  为判别器对于输入真实样本的输出概率。判

别器  $D$  的值越大, 说明生成的光谱数据越“逼真”。GAN 的训练方式是使生成器生成的数据尽可能地“欺骗”过判别器, 如果判别器不能区别原始数据与生成数据, 则证明生成的数据具有“真实性”。同时, 判别器也会调整自身的判别能力, 与生成器形成一种对抗博弈关系。其理论公式如式(3)所示

$$\min_G \max_D V(D, G) = E_{x \sim P_{data(x)}} [\log(D(x))] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (3)$$

由式(3)可以看出, 对于判别器来说, 它需要使得  $V(D, G)$  获得最大值, 那么就需要最大化  $D(x)$ , 并且最小化  $D(G(z))$ ; 对于生成器来说, 它需要使得  $V(D, G)$  获得最小值, 那么它只与等式右边第二项有关, 即其只需最大化  $D(G(z))$ 。

### 2.2 深度卷积生成式对抗网络模型的搭建

本文在生成式对抗网络模型基础上, 增加 1 维卷积核、LeakyReLU 激活函数和最大池化层等组成部分, 构建深度卷积生成式对抗神经网络。深度卷积生成式对抗神经网络主要由随机噪声、生成网络、判别网络、全连接层与激活函数构成, 具体模型结构如图 3 所示。

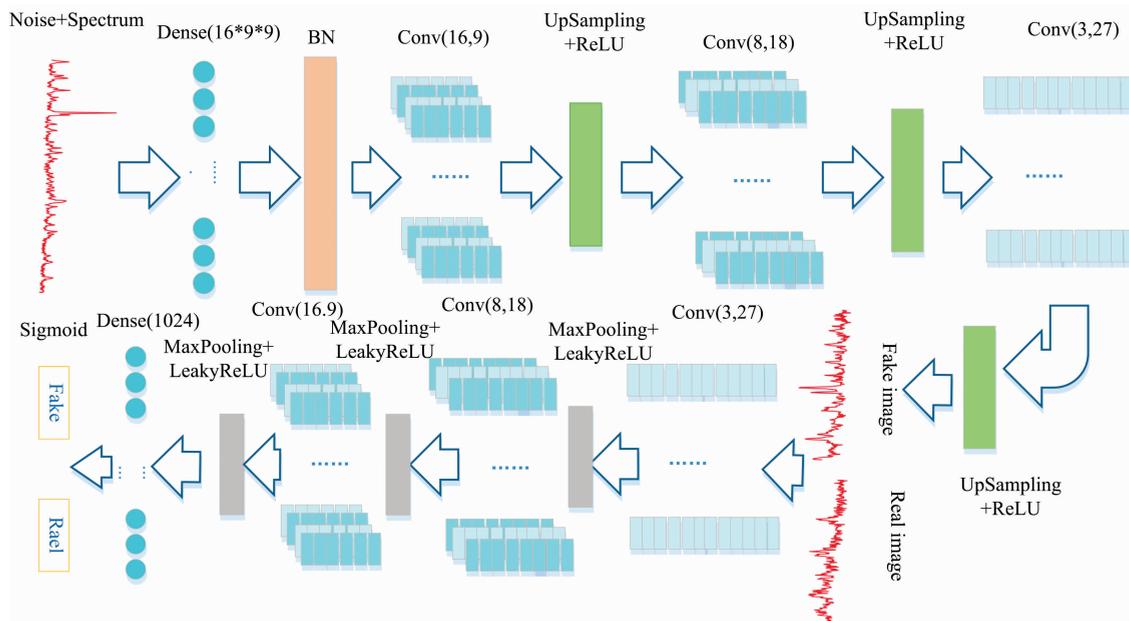


图 3 深度卷积生成式对抗网络结构图

Fig. 3 DCGAN structure diagram

首先, 在生成网络模型中输入叠加 100 维随机噪声的原始数据, 该数据经过全连接层, 并经标准化, 目的是更容易正确地收敛数据从而得到最优解。然后, 将标准化的数据输入给第一层的卷积核、ReLU 激活函数与上采样, 得到的数据再次输入第二层、第三层, 最终传递给 tanh 激活函数。模型具体参数配置如表 1 所示。

判别器具有判别生成数据与原始数据的作用, 主要由 3 个卷积层、3 个 LeakyReLU 激活函数, 3 个最大池化层组成, 最终通过 sigmoid 激活函数判别生成数据与原始数据的差异。具体参数设置如表 2 所示。

表 1 生成网络参数

Table 1 Parameters of generate network

网络层	1 维卷积核	步长	Padding	激活函数
Conv1	(16, 9)	3	same	ReLU
Conv2	(8, 18)	3	same	ReLU
Conv3	(3, 27)	3	same	ReLU

本文采用 Adam 优化器, 学习率设为 0.000 2, Batch Size 大小设为 50, 迭代训练次数设为 200。针对 9 种不同纯度的样本基于深度卷积生成式对抗网络进行光谱数据增强,

每种纯度的样本扩充到 1 000 个。

表 2 判别网络参数

Table 2 Parameters of discriminating network

网络层	1 维卷积核	步长	Padding	激活函数
Conv1	(3, 27)	3	same	LeakyReLU
Conv2	(8, 18)	3	same	LeakyReLU
Conv3	(16, 9)	3	same	LeakyReLU

### 3 结果与讨论

#### 3.1 基于 DCGAN 的拉曼光谱增强结果分析

实际中,不法厂家通常在小比例范围内掺伪纯珍珠粉,

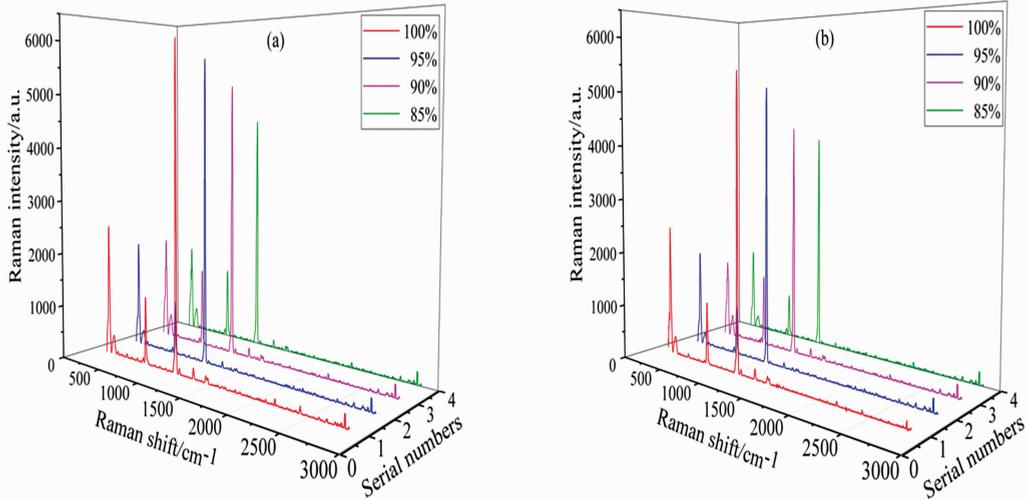


图 4 原始光谱与基于 DCGAN 增强的生成光谱

(a): 原始光谱; (b): 生成光谱

Fig. 4 Original spectrum and generated spectrum based on DCGAN

(a): Original spectra; (b): Generated spectra

PSNR 计算生成拉曼信号与原始拉曼信号的差异来反应二者失真程度,单位为 dB,其值越大,表示两种数据越接近,计算如式(4)所示

$$\text{PSNR} = 10 \log_{10} \left[ \frac{(I_{\max} - I_{\min})^2}{\text{MSE}} \right] \quad (4)$$

式(4)中,  $I_{\max}$  和  $I_{\min}$  分别是原始拉曼信号强度的最大值和最小值, MSE 指均方误差(mean square error, MSE),是生成光谱和原始光谱对应拉曼强度差平方和的均值,计算如式(5)所示

$$\text{MSE} = \frac{1}{H} \sum_{i=1}^H (x(i) - y(i))^2 \quad (5)$$

式(5)中,  $H$  为拉曼信号数量,  $x(i)$  和  $y(i)$  分别为第  $i$  个拉曼位移处原始信号和生成信号的强度值。

SSIM 是从结构来衡量原始和生成拉曼光谱信号  $x$  和  $y$  之间的结构相似度,计算公式如式(6)所示

$$\text{SSIM}(x, y) = \left( \frac{\sigma_{xy} + c}{\sigma_x \sigma_y + c} \right)^\gamma \quad (6)$$

式(6)中,  $\sigma_x$  和  $\sigma_y$  分别为原始光谱和生成光谱的方差,  $c$  为常数,  $\gamma$  是大于 0 的数值,通常取 1。该指标通常是介于 0 和

目的是欺瞒质检部门从而谋取非法利润。因此本文选取 100%, 95%, 90% 与 85% 的 4 种纯度的样本进行珍珠粉真伪鉴别定性分析。每种样本随机选取 20 条原始拉曼光谱与基于 DCGAN 增强生成的拉曼光谱,其平均值结果分别如图 4(a)和(b)所示。

由图 4 可知,基于 DCGAN 增强生成的拉曼光谱与原始的拉曼光谱并不相同,原因是生成式对抗神经网络在原始的归一化光谱中叠加了随机噪声,使得扩充后的光谱数据具备良好的多样性。本文采用峰值信噪比(peak signal to noise ratio, PSNR)和结构相似性(structural similarity, SSIM)指标来评价原始光谱与生成光谱数据的相似性,检验生成数据效果。

1 之间的一个数值,其值越趋近于 1,表示两种光谱在结构上越相似。

针对 100%, 95%, 90% 和 85% 这 4 种掺伪比例较小的珍珠粉样本,分别采用左右平移、叠加噪声及平移噪声三种数据增强方法和本文提出的 DCGAN 数据增强方法,采用 PSNR 和 SSIM 两个评价指标,相似性分析对比结果如表 3 所示。

由表 3 可知,对于四种纯度的掺伪珍珠粉样本拉曼光谱,基于 DCGAN 数据增强方法所生成的光谱,与原始光谱作比较,在峰值信噪比和结构相似度指标上均明显优于其他三种传统数据增强方法所得结果,PSNR 值均是最大,SSIM 值均最接近 1。结果表明深度卷积生成式对抗神经网络可以对掺伪珍珠粉样本拉曼光谱进行增强,同时增强后数据具有多样性。

#### 3.2 基于 DCGAN 光谱增强的掺伪珍珠粉真假鉴别

基于 DCGAN 方法对样本的拉曼光谱进行数据增强后,将 4 种纯度共 120 个掺伪珍珠粉样本数据集随机划分为训练集和测试集,其中训练集样本 72 个,测试集样本 48 个。然后,分别结合 K 近邻(K-nearest neighbor, KNN)、随机森林

**表 3 基于传统数据增强和 DCGAN 增强方法所生成光谱与原始光谱相似性评价**  
**Table 3 Similarity evaluation between the spectra generated by traditional data enhancement and DCGAN enhancement methods and the original spectra**

样本 纯度/%	左右平移		叠加噪声		平移+噪声		DCGAN	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
100	23.57	0.919 9	32.83	0.581 1	24.03	0.620 3	47.67	0.973 7
95	23.99	0.929 2	31.49	0.495 2	24.22	0.623 0	54.67	0.992 6
90	22.15	0.889 7	30.19	0.450 8	24.84	0.567 0	46.58	0.994 7
85	24.18	0.916 3	29.10	0.416 7	23.88	0.527 6	51.61	0.996 1

(random forest)、决策树(decision tree)和一维卷积神经网络(one-dimensional convolution neural network, 1DCNN)四种分类算法, 建立珍珠粉真伪鉴别的分类模型。同时, 为了验证本文提出的深度卷积生成式对抗神经网络的数据增强效果, 建立基于左右平移、叠加噪声以及平移与噪声组合的三种数据增强方法的定性分析结果。四种光谱数据增强方法结合四种分类算法的掺伪珍珠粉真假鉴别模型的综合正确率对比结果如表 4 所示。

**表 4 掺伪珍珠粉真假鉴别正确率结果对比**  
**Table 4 Comparison of the identification results of adulterated pearl powder**

分类算法	分类正确率/%			
	左右平移	叠加噪声	平移+噪声	DCGAN
KNN	98.03	97.38	97.63	100
random forest	94.50	92.75	73.38	100
decision tree	95.87	98.87	86.63	100
1DCNN	99.78	99.12	99.04	100

由表 4 可知, 基于 DCGAN 方法对掺伪珍珠粉拉曼光谱扩充的数据集, 在分别结合 KNN, random forest, decision tree 与 1DCNN 四种分类算法时, 所建模型的综合正确率均可以达到 100%, 优于基于左右平移、叠加噪声以及平移噪声组合数据增强方法结合四种分类算法的综合正确率。此外, 传统左右平移和叠加噪声等数据增强方法数据扩充所需时间较长, 相比而言, 深度卷积生成式对抗神经网络数据增强方法在建立模型的正确率和数据扩充效率上具有明显优势。

**3.3 基于 DCGAN 光谱增强的掺伪珍珠粉定量分析**

本文基于深度卷积生成式对抗神经网络光谱数据增强方法, 结合一维卷积神经网络算法, 建立了 9 种纯度的掺伪珍珠粉样本的纯度定量回归模型, 并与基于左右平移、叠加噪声以及平移+噪声三种传统数据增强方法所建定量模型结果进行对比。将 270 个掺伪珍珠粉样本数据集随机划分为训练集和测试集, 其中训练集样本 162 个, 测试集样本 108 个。采用测试集决定系数(coefficient of determination,  $R^2$ )和预测均方根误差(root mean square error of prediction, RMSEP)作为定量模型评估指标。 $R^2$  值越接近 1, 模型的拟合效果越好; RMSEP 值越小, 模型预测准确度越高。此外一维卷积神经网络模型的损失值反映样本预测纯度与真实纯度的差距, 其值越小, 表示预测纯度与真实纯度越接近。4 种数据

增强方法分别结合 1DCNN 算法所建定量模型指标对比结果如表 5 所示。

**表 5 不同数据增强方法结合 1DCNN 算法所建定量模型结果对比**

**Table 5 Comparison of quantitative models built by different data enhancement methods combined with 1DCNN**

数据增强方法	$R^2$	RMSEP	LOSS
左右平移	0.856 2	0.125 4	0.015 6
叠加噪声	0.943 8	0.078 2	0.006 1
平移+噪声	0.844 0	0.130 3	0.017 0
DCGAN	0.988 4	0.034 8	0.001 2

由表 5 可知, 基于 DCGAN-1DCNN 方法所建掺伪珍珠粉定量分析模型, 对于测试集样本, 其决定系数  $R^2$  为 0.988 4, 均方根误差 RMSEP 为 0.034 8, 一维卷积神经网络的损失值 Loss 为 0.001 2, 定量模型拟合度最高。模型性能明显优于其他三种数据增强方法结合一维卷积神经网络所建定量模型。

针对测试集样本, 基于 DCGAN、左右平移、叠加噪声以及平移+噪声四种数据增强方法结合 1DCNN 算法所建定量模型的预测纯度与真实纯度相关曲线分别如图 5(a), (b), (c)和(d)所示。从图 5 可以清晰地看出: 基于 DCGAN-1DCNN 方法所建掺伪珍珠粉定量模型的预测纯度与真实纯度非常接近, 预测精度最优。

**4 结 论**

将激光拉曼光谱和深度学习数据增强应用于珍珠粉掺伪的定性鉴别和纯度定量回归分析。针对拉曼光谱数据不足的问题, 提出一种基于深度卷积生成式对抗神经网络光谱数据增强方法, 并结合一维卷积神经网络建立珍珠粉掺伪的定性和定量分析模型。结果表明基于 DCGAN 数据增强方法扩充的光谱数据与原始数据相比, PSNR 和 SSIM 性能指标优于传统数据增强方法; DCGAN 结合四种分类器对小比例掺伪珍珠粉样本的分类正确率均达到 100%; DCGAN-1DCNN 方法所建立的掺伪珍珠粉纯度预测模型, 测试集决定系数  $R^2$  和均方根误差 RMSEP 均优于传统数据增强方法所建模型结果, 预测纯度与真实纯度非常接近。深度卷积对抗式神经网络数据增强方法具有精度高、抗干扰性好、简便快速等优点, 为深度学习在光谱数据分析领域应用提供了新思路。

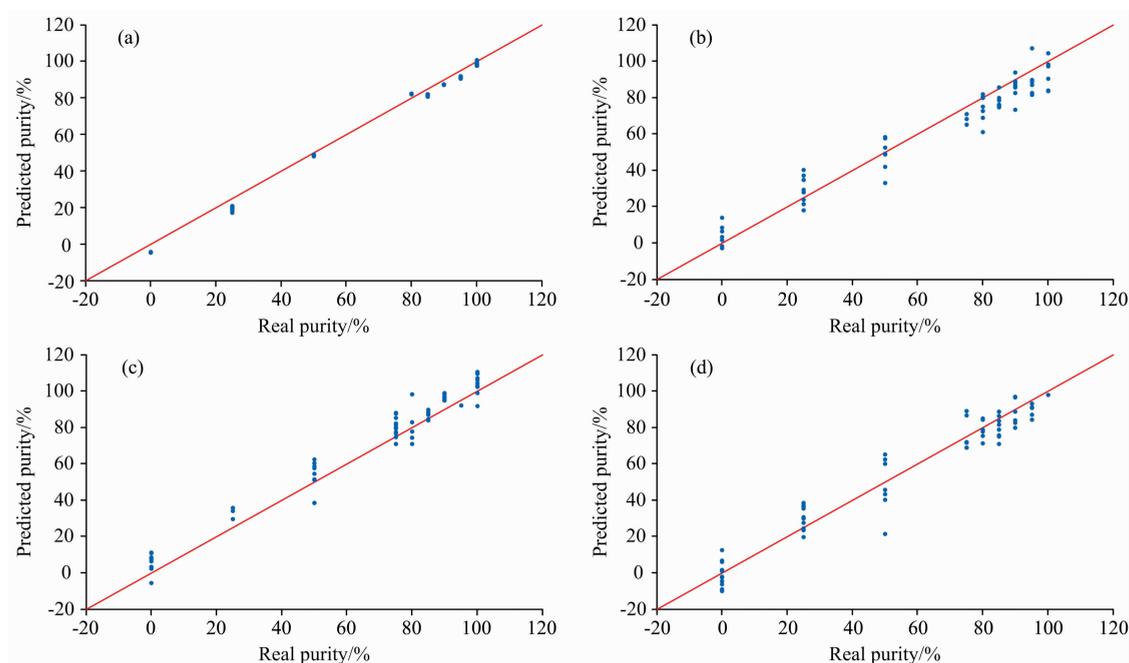


图 5 不同数据增强方法结合 1DCNN 算法所建定量模型真实纯度与预测纯度相关曲线

(a): DCGAN-1DCNN; (b): 叠加噪声-1DCNN; (c): 左右平移-1DCNN; (d): 噪声+平移-1DCNN

Fig. 5 Correlation curve between real and predicted purity of quantitative models built by different data enhancement methods combined with 1DCNN

(a): DCGAN-1DCNN; (b): Noise addition-1DCNN; (c): Translation-1DCNN; (d): Noise+Translation-1DCNN

## References

- [ 1 ] Chinese Pharmacopoeia Commission(国家药典委员会). Pharmacopoeia of the People's Republic of China(Volume 1)(中华人民共和国药典(1部)). Beijing: Chemical Industry Press(北京: 化学工业出版社), 2018. 159.
- [ 2 ] PU Yue-hua, HE Jin-feng, GAO Zhen-sheng, et al(蒲月华, 何锦锋, 高振声, 等). Food Research and Development(食品研究与开发), 2016, 37(16): 125.
- [ 3 ] QIAO Yi-han, SUO Ya-ran, MENG Xue-dan, et al(乔艺涵, 索亚然, 孟雪丹, 等). Chinese Journal of Pharmaceutical Analysis(药物分析杂志), 2019, 39(5): 911.
- [ 4 ] Liu S Q, Wei W, Bai Z Y, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, 189(189): 265.
- [ 5 ] Weng S Z, Zhu W X, Zhang X Y, et al. Artificial Intelligence in Agriculture, 2019, 3(9): 1.
- [ 6 ] Meikun Fan, Gustavo F S. Andrade, et al. Analytica Chimica Acta, 2020, 1097: 1.
- [ 7 ] Wang Dongqi, He Peisheng, Wang Zijian, et al. Current Opinion in Biotechnology, 2020, 64: 218.
- [ 8 ] MING Jing, CHEN Long, CHEN Ke-li, et al(明晶, 陈龙, 陈科力, 等). China Journal of Traditional Chinese Medicine and Pharmacy(中华中医药杂志), 2017, 32(3): 1340.
- [ 9 ] WANG Qing-rong, LI Xiao-yan, SUN Ying, et al(王清蓉, 李晓艳, 孙莹, 等). The Journal of Light Scattering(光散射学报), 2018, 30(4): 351.
- [ 10 ] Félix Lussier, Vincent Thibault, Benjamin Charron, TrAC Trends in Analytical Chemistry, 2020, 124: 115796.
- [ 11 ] Esben Jannik Bjerrum, Mads Glahder, Thomas Skov. arXiv: 2017, 1710. 01927v1.
- [ 12 ] Frédéric Jamme, Ludovic Duponchel. Journal of Chemometrics, 2017, 31(5): e2882.
- [ 13 ] ZHAO Yong, RONG Kang, TAN Ai-ling(赵勇, 荣康, 谈爱玲). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2019, 39(12): 3755.
- [ 14 ] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Advances in Neural Information Processing Systems, 2014, 27: 2672.

# Detection of Pearl Powder Adulteration Based on Raman Spectroscopy and DCGAN Data Enhancement

TAN Ai-ling<sup>1</sup>, CHU Zhen-yuan<sup>1</sup>, WANG Xiao-si<sup>1</sup>, ZHAO Yong<sup>2\*</sup>

1. School of Information and Science Engineering, Yanshan University, the Key Laboratory for Special Fiber and Fiber Sensor of Hebei Province, Qinhuangdao 066004, China

2. School of Electrical Engineering, Yanshan University, the Key Laboratory of Measurement Technology and Instrumentation of Hebei Province, Qinhuangdao 066004, China

**Abstract** The chemical composition of pearl powder and nacre powder is similar, However, the medicinal value of nacre powder is far lower than that of nacre powder, and nacre powder is easier to prepare at low cost, which is often used by illegal businesses to fake or mix into nacre powder to enter the market and seek illegal interests. Therefore, identifying pearl powder adulteration and purity analysis is of great significance. This paper used Raman spectroscopy combined with deep learning technology to study the rapid identification and purity analysis of pearl powder adulteration. The pure pearl powder and nacre powder were mixed according to a certain proportion to make 270 samples with 9 kinds of purity of 0%, 25%, 50%, 75%, 80%, 85%, 90%, 95% and 100% respectively. Then the Raman spectrum of the samples was collected, and the parameters were set as follows: the resolution is  $4.5\text{ cm}^{-1}$ , the integration time is 3 000 ms, and the laser power is 20 mW. A Deep Convolutional Generative Adversarial Network (DCGAN) model was built to enhance the Raman spectra of the samples. Furthermore, K-nearest neighbor, Random forest, Decision tree, and one-dimensional convolution neural network (1DCNN) classifiers were used to identify the authenticity of a small proportion of adulterated samples with the purity of 85%, 90%, 95% and 100%. At the same time, a quantitative model for the purity prediction of 9 kinds of adulterated pearl powder samples was established by using a one-dimensional convolutional neural network. The results showed as follows: compared with the original spectral data, the Raman spectral data generated by the DCGAN data enhancement method was significantly better than the traditional data enhancement methods in the two evaluation indexes of peak signal-to-noise ratio and structural similarity. For the identification of pearl powder adulteration, the accuracies of the qualitative models established by DCGAN data enhancement method combing with four different classifiers have all reached 100%. For the quantitative detection of the purity of the pearl powders, the model established by DCGAN-1DCNN method has achieved the best performance. For the test set, the determination coefficient ( $R^2$ ) was 0.988 4 and the prediction root mean square error (RMSEP) was 0.034 8 as the loss value of 1DCNN network was 0.001 2. Raman spectroscopy combined with the DCGAN method provides a rapid and simple method for identifying and purity analysis of pearl powder adulteration. The data enhancement method of deep convolution generation countermeasure network has important research significance and application value in spectral analysis technology.

**Keywords** Generative adversarial network; Raman spectroscopy; Deep learning; Adulteration of pearl powder; One-dimensional convolutional neural networks

(Received Feb. 7, 2021; accepted Apr. 19, 2021)

\* Corresponding author