

激光与光电子学进展

基于深度学习的石油污染物三维荧光光谱识别技术研究

王忠东¹, 张云刚², 张亮敬^{1*}, 吴柳强¹¹广西科技师范学院数学与计算机科学学院, 广西 来宾 546199;²燕山大学电气工程学院, 河北 秦皇岛 066004

摘要 石油油品在一定的激发光照射下可产生相当强度的三维荧光光谱,是鉴别和分析石油污染物的重要依据。由于石油油品的荧光光谱特征复杂、数据庞大,不宜直接用数学模型描述,也不宜简单依靠人工观察分析。因此,根据深度学习的卷积神经网络(CNN)理论提出了一种直接利用石油油品原始荧光数据进行CNN建模的方法,利用其强大的非线性运算能力、自适应表示学习能力,自动隐性地从训练数据中进行特征学习,实现水环境中石油污染物种类识别。通过大量的荧光实验构建了石油油品(汽油、机油、柴油)的训练和验证光谱数据集,基于Python深度学习框架Keras建立了CNN模型,并对CNN模型在光谱数据集上进行了训练、验证与测试实验,实现了被测油品的种类判别。实验结果表明:该CNN模型对3种油品的训练集与验证集三维荧光光谱的分类准确率都达到了99.76%,综合测试分类准确率达到82.65%,对单物质分类准确率为100%,验证了三维荧光技术结合深度学习算法能够实现对石油油品准确可靠的判别分类,也为进一步研究基于深度学习的水环境污染智能识别系统提供了技术支持,为环境检测提供了一种新思路与新方法。

关键词 三维荧光光谱; 深度学习; 卷积神经网络; 分类识别; 石油污染物

中图分类号 O657.3

文献标志码 A

DOI: 10.3788/LOP221854

Research on Three-Dimensional Fluorescence Spectrum Identification Technology of Petroleum Pollutants Based on Deep Learning

Wang Zhongdong¹, Zhang Yungang², Zhang Liangjing^{1*}, Wu liuqiang¹¹Department of Mathematics and Computer Science, Guangxi Science & Technology Normal University, Laibin 546199, Guangxi, China;²College of Electrical Engineering, Yanshan University, Qinhuangdao 066004, Hebei, China

Abstract Petroleum oil products can produce three-dimensional fluorescence spectra with considerable intensity under certain excitation light, allowing the identification and analysis of petroleum pollutants. Because of the complex characteristics and huge data of the fluorescence spectrum of petroleum oil products, it is not easy to be described with a simple mathematical model, nor to rely on artificial observation and analysis. This paper presents a convolutional neural network (CNN) model constructed using raw fluorescence data of three petroleum products (gasoline, oil, diesel), which automatically learns features from training data and classifies petroleum pollutants in water, using its nonlinear computing ability and adaptive representation learning ability. Through various fluorescence spectrum experiments, training and validation spectral datasets of petroleum products are constructed, and the CNN model is established based on the deep learning framework Keras for Python. The CNN model is trained, validated, and tested on the spectral dataset, to classify and discriminate measured oil products. The experimental results show that the classification accuracy of the CNN fluorescence model on the training and validation sets of the three petroleum products is 99.76%, the classification accuracy in a comprehensive test is 82.65%, and the classification accuracy for a single substance is 100%. Additionally, the experimental results confirm the feasibility of combing three-dimensional fluorescence technology with a deep learning algorithm, to distinguish and classify petroleum products accurately and reliably. These results provide technical guidance

收稿日期: 2022-06-16; 修回日期: 2022-07-11; 录用日期: 2022-07-20; 网络首发日期: 2022-07-30

基金项目: 国家自然科学基金(62175208)、中央引导地方科技发展基金(216Z1701G)、广西创新驱动发展专项基金(桂科AA21077018-2)

通信作者: *1585550358@qq.com

for further research on creating intelligent identification models for environmental pollutants in water, as well as a new direction for environmental detection methods.

Key words three-dimensional fluorescence spectrum; deep learning; convolutional neural network; classification and identification; petroleum pollutant

1 引言

石油的不当使用和意外泄漏给环境带来了不同程度的污染,特别是对水的污染最为严重^[1]。因此,精准高效检测水中石油污染物对防止污染和保护环境具有重要意义。

石油是一种烃类混合物,受一定光的激发会发出荧光。可以通过荧光测量仪器获得荧光的二维或三维光谱。由于石油油品的化学分子结构不同,不同油品将发出独特的荧光光谱,这为鉴别分析提供重要依据。然而,由于油品种类较多,分子结构复杂,荧光光谱特征复杂、信息丰富、数据庞大,很难直接用某种数学模型简单描述,依靠人工分析通常效率低下、易受人的主观因素影响,且需专门分析人员及专业技能,当样品成分复杂时,光谱将产生严重重叠,仅凭人工判别是不可能完成的^[2]。因此,有必要找到一种新的方法实现对复杂石油油品的三维荧光光谱高效准确检测识别,从而提高荧光检测技术的质量与水平。

随着深度学习广泛而深入的研究,在图像分类、语音识别、自动驾驶等领域取得了重大突破。最具代表性的卷积神经网络(CNN)算法也随着深度学习理论的进步,在图像识别领域得到了快速发展^[3]。这为研究新的荧光光谱分析方法提供了新思路。利用深度学习算法对有机物质的荧光光谱进行识别分析正在成为当今研究热点。来文豪等^[3]提出一种基于激光诱导荧光技术获取待测白酒的荧光光谱,利用深度学习算法辨识白酒的种类与度数,识别率达 98.44%。赵志蕾^[4]基于荧光光谱结合深度学习的循环神经网络(RNN)、CNN 算法实现了对真假芝麻油的鉴别和定量分析,鉴别准确率为 100%,定量分析预测误差小于 3%。丛琳^[5]基于 CNN 提出了一种适用于蜂蜜三维荧光光谱

分析的网络模型且实现了蜂蜜掺假的快速检测,鉴别准确率为 97.09%,并具有良好的泛化能力。张学成等^[6]基于 CNN 和全卷积网络等深度学习算法对花粉等气溶胶物质的二维荧光光谱进行了识别分类,分类准确率达到 98.05%。Xu 等^[7]提出了一种基于深度学习方法的快速荧光识别网络(FFI-Net)模型,其由荧光成分数量和荧光成分图谱 2 个 CNN 模型构成,实现了对三维荧光光谱的快速预测,检测精度达 95.6%,最小平均绝对误差为 8.9×10^{-4} 。本文根据深度学习在图像识别领域的应用经验以及石油物质的荧光光谱特性,设计了面向汽油、机油、柴油的三维荧光光谱分类识别的 CNN 模型,实现了水环境中石油污染物的分类鉴别。

2 基本原理

2.1 卷积神经网络

CNN 是一个由多层“卷积层”和“采样层”等非全连接层组成的神经网络。CNN 的工作原理与反向传播(BP)神经网络相同,即在正向传播中逐层提取输入数据的特征,反向传播时利用梯度下降算法最小化损失误差调整训练参数,完成权值更新^[3-4,8]。CNN 算法本质上是一种输入到输出的非线性映射,但不需要精确的数学表达式,只要用简单模型对网络加以训练,就可完成复杂的分类等学习任务。CNN 具有很强的“特征学习”能力,其避免了由人类专家设计取样特征的缺点,能够全自动地、隐式地从训练数据中进行学习,非常适用于对具有复杂信息特征的对象进行分类识别^[3,5,8-9]。因此,选用 CNN 算法进行荧光光谱的识别分类。

CNN 通常由输入层、卷积层和池化层以及全连接层堆叠组成,如图 1 所示。网络的输入层为训练数据入口,输入可以是一维向量、也可以是二维或三维向量。

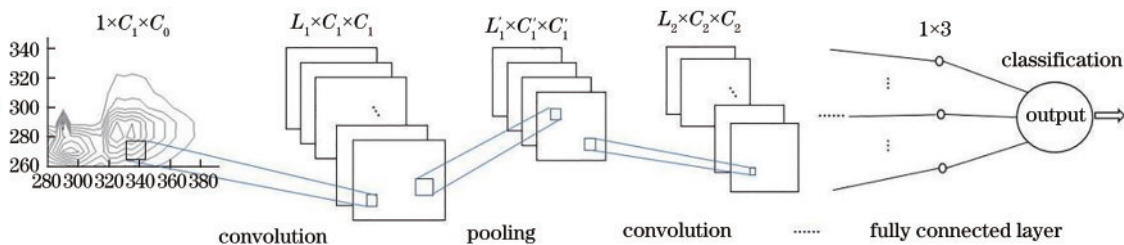


图 1 CNN 模型结构原理

Fig. 1 Structure principle of CNN

卷积层为特征映射层,每个特征映射是由多个神经元构成的“平面”,通过卷积滤波器(卷积核)提取输入的特征,卷积^[3,8-9]可表示为

$$Y_p^k = f \left[\sum_{m \in K_p} (X_p^{k-1} \times K_{mp}^k + B_p^k) \right], \quad (1)$$

式中: Y_p^k 、 X_p^{k-1} 分别为 k 层、 $k-1$ 层的第 p 个特征映射

的值; K_{mp}^k 为卷积滤波器权重值, m 为卷积滤波器序号; B_p^k 为 k 层的第 p 个特征映射的偏置; R_p 为卷积滤波器数量; f 为激活函数, 目的是增强模型的非线性, 在深层网络中一般选用 ReLU 函数^[3-4,9]。

在卷积层后, 都会连接一个实现局部平均和子抽样的计算层, 称池化层, 进行特征选择和信息过滤, 以降低特征映射的维度^[3-5]。池化层的池化函数可表示为

$$y_p^k = P(N_{ij}^k * Y_p^{k-1}), \quad (2)$$

式中: N_{ij}^k 为池化核; Y_p^{k-1} 为 $k-1$ 层特征值; y_p^k 为池化后特征输出; P 为池化函数, 一般采取平均值法和最大值法; $*$ 为卷积符号。

最后是全连接层, 一般设为 1~3 层。其功能是将前面经过卷积和池化后输出的分布式特征映射转换成一维特征映射再映射到样本标签空间, 起到“分类”的作用。对于二分类问题, 全连接层会连接 Sigmoid 分类函数, 多分类问题则会连接 Softmax 分类函数^[3-4,8-9]。

2.2 实验分析

2.2.1 实验方法

通过荧光光谱仪测量一个油品样本能够很容易得到一个 $I \times J$ 的激发-发射荧光光谱矩阵, 相当于一张二维图像, I 和 J 分别为激发波长数和发射波长数。测量 K 个样本就可以得到一个 $K \times I \times J$ 的三维荧光光谱图矩阵。将所有样本的荧光光谱去除背景干扰噪声后进行标准化处理, 得到深度学习建模用的荧光数据集。然后, 利用深度 CNN 强大的建模能力, 即使不清

楚样本数据明确的相互作用, 也可以直接利用 CNN 工具将荧光光谱与待测物质种类建立映射, 实现对样本种类的预测。

2.2.2 实验仪器与材料

实验仪器采用法国 HORIBA 荧光光谱仪。实验材料采用中石油 92# 汽油、10# 柴油、普通机油、纯净去离子水。

根据被测汽油、柴油、机油样品的分子荧光特性, 激发波长范围设置为 250~500 nm、激发光间隔为 5 nm, 发射波长范围为 260~700 nm, 发射间隔为 5 nm, 狭缝宽度为 2 mm, 采样积分时间为 200 ms。

2.2.3 实验分析

2.2.3.1 油品荧光实验

1) 标准溶液配制

分别称取适量的 92# 汽油、10# 柴油、普通机油与纯净去离子水充分搅拌混合, 将 3 种石油样品分别配制成 500 $\mu\text{g}/\text{L}$ 的标准液, 将样品溶液保存在 4 $^\circ\text{C}$ 以下, 溶液制备与保存需要避光。

2) 样本制作

用高精度移液器移取标准溶液, 在容量瓶中用纯净去离子水稀释, 定容配制 3 种油品不同比例、不同质量浓度的 100 组训练样品 ($C_1 \sim C_{100}$)。为了方便质量控制与分析, 配置样品的质量浓度梯度变化相同。样品溶液质量浓度范围都为 5~500 $\mu\text{g}/\text{L}$ 。同时制备 3 个纯净去离子水作为空白样本 ($K_1 \sim K_3$)。表 1 为部分样品质量浓度数据, 误差为 $\pm(2\% \sim 3\%)$ 。

表 1 实验样品质量浓度配置

Table 1 Mass concentration configuration of experimental samples

unit: $\mu\text{g} \cdot \text{L}^{-1}$

| Sample | 10# diesel oil (S_1) | 92# gasoline (S_2) | Machine oil (S_3) |
|----------------|--------------------------|------------------------|-----------------------|
| C_1 | 5.0 | 5.0 | 5.0 |
| C_2 | 5.5 | 5.5 | 5.5 |
| C_3 | 6.0 | 6.0 | 6.0 |
| C_4 | 6.5 | 6.5 | 6.5 |
| C_5 | 7.0 | 7.0 | 7.0 |
| C_6 | 7.5 | 7.5 | 7.5 |
| C_7 | 8.0 | 8.0 | 8.0 |
| C_8 | 8.5 | 8.5 | 8.5 |
| C_9 | 9.0 | 9.0 | 9.0 |
| C_{10} | 10.0 | 10.0 | 10.0 |
| C_{98} | 490.0 | 490.0 | 490.0 |
| C_{99} | 495.0 | 495.0 | 495.0 |
| C_{100} | 500.0 | 500.0 | 500.0 |
| $K_1 \sim K_3$ | 0 | 0 | 0 |

2.2.3.2 光谱实验

利用荧光光谱仪对配置好的实验样本进行荧光激发与采集, 获取全部样本的三维荧光光谱数据。

1) 空白样本 ($K_1 \sim K_3$) 测试

首先测试空白水样本的三维荧光光谱。获取水的

三维荧光光谱和等高线图, 如图 2 所示。实验表明, 水不发荧光, 但具有很强的瑞利散射, 在等高线图上呈现出密集带状信号, 位置在发射波长-激发波长的主对角线上。瑞利散射会将有用信息掩盖。由于各样本溶剂均为相同的水, 其本底的瑞利散射相同, 有利于去除背景干扰。

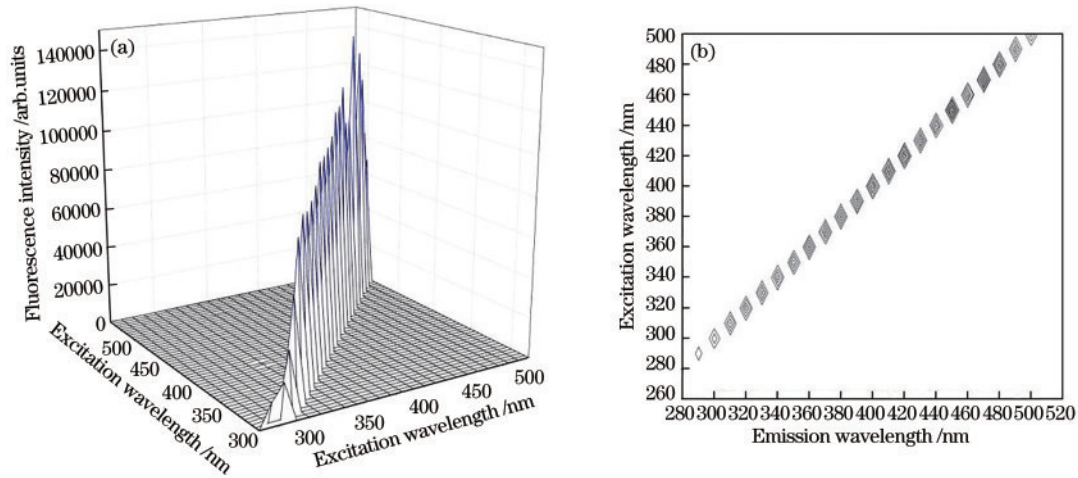


图 2 空白水。(a)三维荧光光谱图;(b)等高线图

Fig. 2 Blank water. (a) Three-dimensional fluorescence spectra; (b) contour maps

2) 训练样本测试

分别对 100 个训练样本 ($C_1 \sim C_{100}$) 进行荧光激发, 获取各个样本的三维荧光光谱。以训练样本表 1 中的 C_{10} 为例, 92# 汽油原始三维荧光光谱图和等高

线图如图 3 所示, 其消除背景干扰后三维荧光光谱图及等高线图如图 4 所示。机油与 10# 柴油经过消除背景干扰后的三维荧光光谱图和等高线图如图 5 所示。

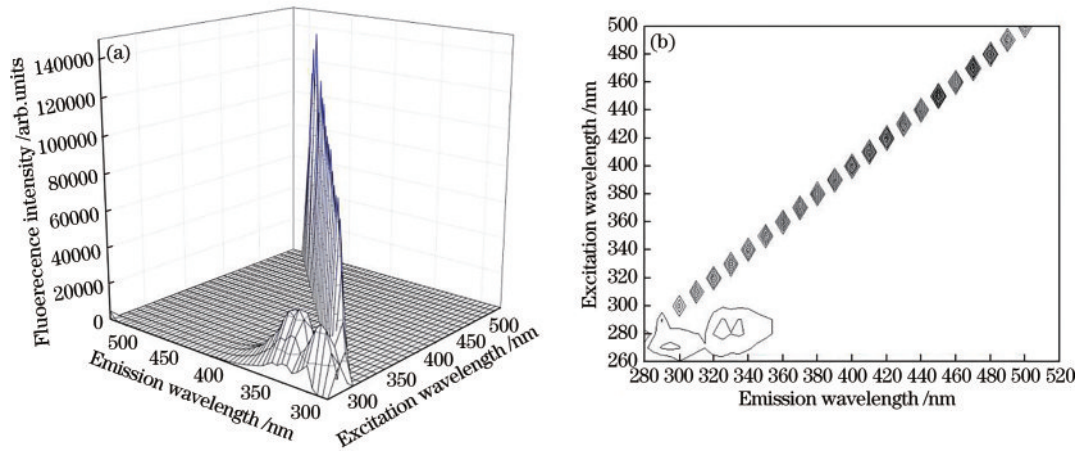


图 3 原始 C_{10} -92# 汽油。(a)三维荧光光谱图;(b)等高线图

Fig. 3 Original C_{10} -92# gasoline. (a) Three-dimensional fluorescence spectra; (b) contour maps

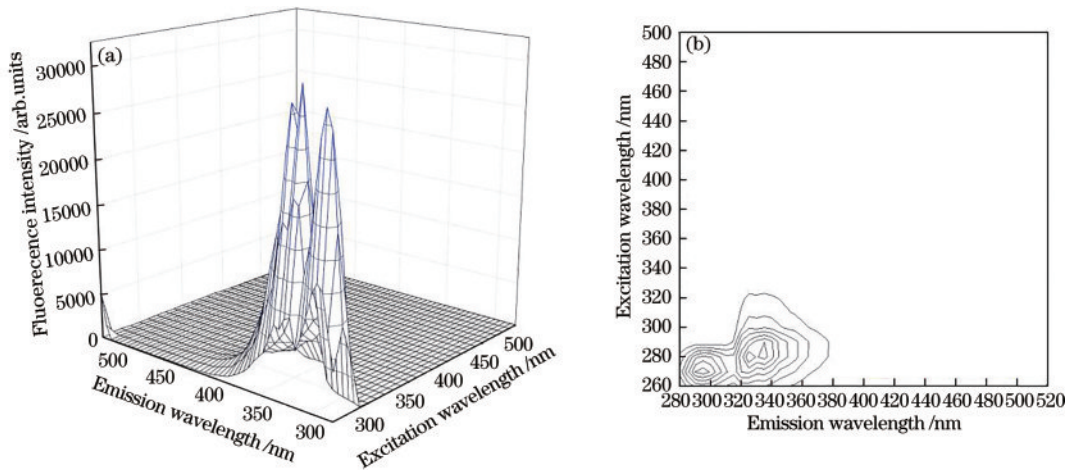


图 4 消除瑞利散射后 C_{10} -92# 汽油。(a)三维荧光光谱图;(b)等高线图

Fig. 4 C_{10} -92# gasoline after elimination of Rayleigh scattering. (a) Three-dimensional fluorescence spectra; (b) contour maps

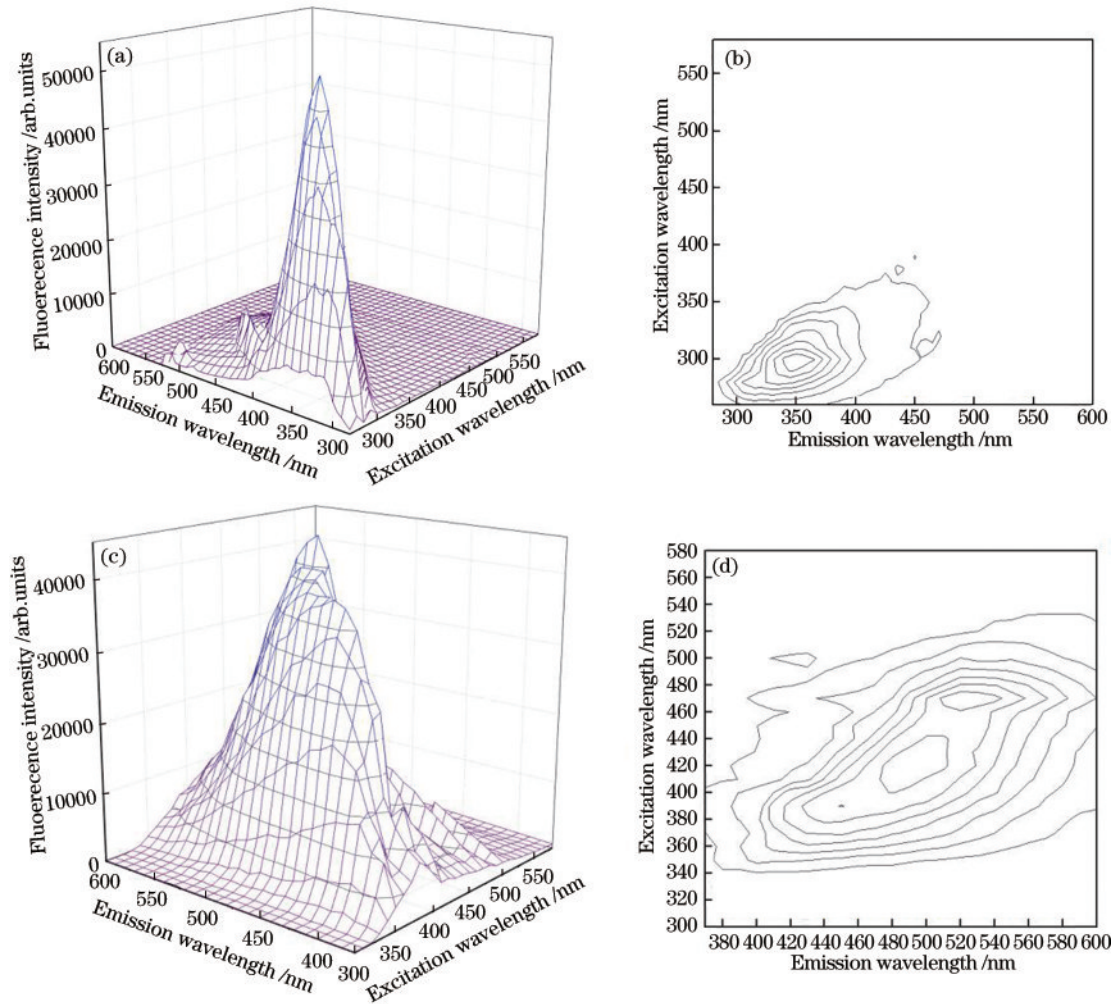


图5 消除瑞利散射后的C₁₀-机油和C₁₀-10#柴油。(a) C₁₀-机油的三维荧光光谱图;(b) C₁₀-机油的等高线图;(c) C₁₀-10#柴油的三维荧光光谱图;(d) C₁₀-10#柴油的等高线图

Fig. 5 C₁₀-machine oil and C₁₀-10# diesel oil after elimination of Rayleigh scattering. (a) Three-dimensional fluorescence spectra of C₁₀-machine oil; (b) contour maps of C₁₀-machine oil; (c) three-dimensional fluorescence spectra of C₁₀-10# diesel oil; (d) contour maps of C₁₀-10# diesel oil

由图4、图5可知,不同油品的三维荧光光谱特征是不同的,这与其分子结构有关^[2]。

由图3可知,92#汽油原始三维荧光光谱图中有很强的光谱峰,与空白水相似,可以断定为瑞利散射,而汽油的荧光强度相对较弱,几乎被淹没,无法有效辨识光谱信息。因此,原始光谱需要进行光谱校正,消除瑞利散射背景干扰后才能进行分析。消除瑞利散射后,如图4所示,样本的三维荧光光谱及等高线图中散射区域荧光数据拟合性好,光谱还原度较高,三维荧光光谱特征信息保存完好。汽油的荧光光谱图及等高线图突显出来,具有非常高的分辨率和荧光强度。由图4(a)、图4(b)可知,92#汽油有2个最佳激发波长(220 nm、235 nm),最佳发射波长为280 nm、330 nm、340 nm,光谱具有显著的三峰特征。图5为表1中C₁₀样品的机油和10#柴油经过消除瑞利散射背景干扰后的荧光光谱图和等高线图,其都具有良好的光谱特征和高分辨率,非常适合进行光谱分析。机油在290 nm

处有个最佳激发波长,相应的最佳发射波长为350 nm,光谱具有显著特征峰。而柴油的最佳激发波长为465 nm,最佳发射波长为530 nm,柴油只有一个单峰,荧光特征明显、分辨率高、易于鉴别。通过分析可知,92#汽油的荧光发射波长与机油的荧光发射波长相近,混合物的荧光发射可能存在较大重合,不易直接区分,而柴油的荧光峰波长与二者相差较远,不会产生明显影响。

3) 混合样本的荧光测试

按照表1配置好的3类样本的样品溶液分别按一定的体积比例进行混合,混合方式如表2所示,配置油品混合物的训练样本,然后对混合物样本进行荧光测量,获取各样本的荧光光谱数据,用于深度学习模型的训练与验证。

以表2中第1种情况为例,实验测得荧光光谱图如图6~图9所示。由图6可知,机油与汽油混合物的荧光光谱较图4(a)、图5(a)所示的单组分时光谱图发生

表2 石油样品混合方式

Table 2 Oil sample mixing method

| Machine oil+ gasoline | Machine oil+ diesel oil | Gasoline+ diesel oil | Machine oil+ gasoline+diesel oil |
|-----------------------|-------------------------|----------------------|----------------------------------|
| 1:1 | 1:1 | 1:1 | 1:1:1 |
| 1:2 | 1:2 | 1:2 | 1:2:1 |
| 1:5 | 1:5 | 1:5 | 1:5:1 |
| 2:1 | 2:1 | 2:1 | 1:1:2 |
| 5:1 | 5:1 | 5:1 | 1:1:5 |

了明显变化,混合物的荧光光谱由于2种物质的共同作用变为双峰结构,激发光在波长为280 nm时,荧光最大峰值出现在335 nm,次峰出现在310 nm,荧光强度也较原来增大很多。但是由于2种混合物质的荧光

光谱重叠严重,已经无法辨别。图7为机油与柴油混合物光谱,由于柴油与机油的最佳激发波长、发射波长相差较远,荧光光谱没有明显重叠,光谱特征明显、分辨率高、易于鉴别。图8为汽油与柴油混合物光谱,与图7相似,由于柴油与汽油的最佳激发波长、发射波长也相差较远,荧光光谱没有明显重叠,易于鉴别。图9为机油、汽油、柴油3种物质混合的荧光光谱,虽然可区分为两部分,但与单成分时的光谱明显不同,3种物质的混合使光谱产生了较严重的重叠现象,很难分辨。在上述4种情况下,通过常规荧光光谱分析进行物质种类鉴别已经很难实现,因此,需采用新的原理与方法。后面将重点介绍基于深度学习的光谱识别方法实现混合物质的种类鉴别。

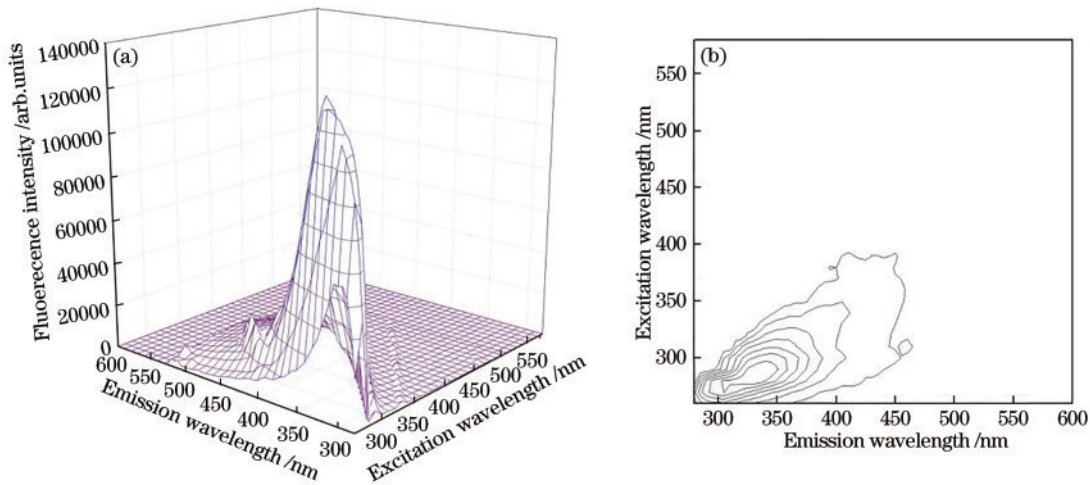


图6 机油与汽油混合。(a)三维荧光光谱图;(b)等高线图

Fig. 6 Machine oil+gasoline. (a) Three-dimensional fluorescence spectra; (b) contour maps

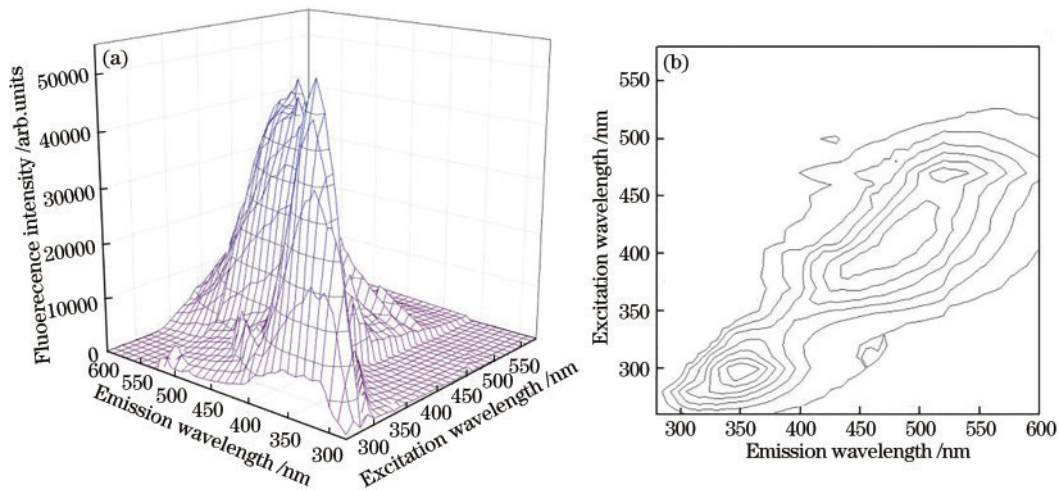


图7 机油与柴油混合。(a)三维荧光光谱图;(b)等高线图

Fig. 7 Machine oil+diesel oil. (a) Three-dimensional fluorescence spectra; (b) contour maps

2.2.3.3 荧光光谱处理

三维荧光光谱是一种激发-发射荧光光谱矩阵,可以将其看作一张二维数据图谱^[4-5]。但是,实验采集的原始荧光光谱中由于存在着大量的瑞利散射和杂质噪

声,样本数据的数量、规格等并不符合深度学习建模的要求。因此,还需将采集的荧光光谱数据进行处理,以满足深度学习模型的需要。

研究中将光谱数据处理分为光谱数据的预处理和

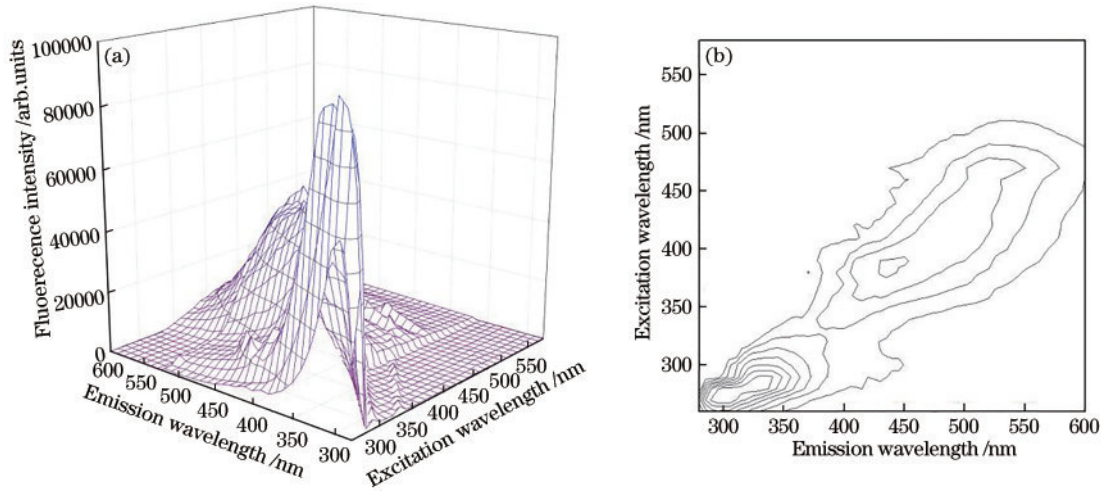


图8 汽油与柴油混合。(a)三维荧光光谱图;(b)等高线图

Fig. 8 Gasoline + diesel oil. (a) Three-dimensional fluorescence spectra; (b) contour maps

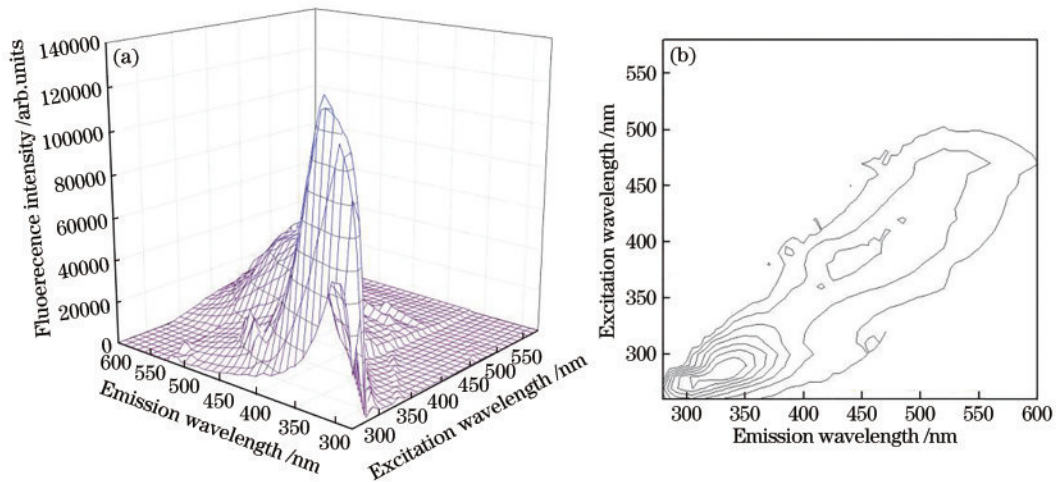


图9 机油、汽油与柴油混合。(a)三维荧光光谱图;(b)等高线图

Fig. 9 Machine oil + gasoline + diesel oil. (a) Three-dimensional fluorescence spectra; (b) contour maps

光谱数据的扩充。光谱数据预处理,首先采用插值拟合方法和空白扣除法^[10]消除荧光光谱中存在的瑞利散射以及杂质噪声等干扰,再通过光谱图裁减,或光谱图扩大以及数据归一化等操作,保证深度学习模型输入的光谱数据标准统一^[4,11]。实验中,由于每一个荧光光谱都是一个 $I \times J$ 大小的荧光强度矩阵,根据CNN建模的需要,将每个荧光光谱 I 与 J 等范围进行了扩大或裁减处理,构成了范围相同标准统一的光谱数据矩阵,并对光谱矩阵数据进行了归一化处理,把光谱强度变换为0~1之间的数值。根据卷积要求,在数据外层进行了补零层处理。

光谱数据扩充主要包括光谱扩充与光谱数据增强,目的是防止CNN模型因数据不足而产生过拟合现象^[4-5,11]。深度CNN模型依赖大量训练数据来进行表征学习的非线性函数。实际上,获得足够多的样本数据来训练深度模型是非常困难的,可以通过光谱数据扩充的办法解决此难题^[4-5]。

本研究根据有机分子荧光特性采用加和性原则对

所测得的荧光光谱样本数量进行了扩充^[2,5]。根据朗伯-比尔定律,如果各溶剂之间没有相互作用,溶液的总吸光度等于多组分系统中各组分的吸光度之和,满足荧光的加和性原则^[5]:

$$I = \sum_{k=1}^n a_k S_k, \quad (3)$$

式中: I 为混合溶液荧光强度; a_k 为混合溶液中 k 组分所占比例; S_k 为混合溶液中 k 组分的荧光强度; $k=1, 2, \dots, n$ 。该方法可以模拟更多样本的混合比例,减少实验室测量样品的数量。

荧光光谱数据增强通过深度学习的常用数据增强方法如翻转、缩放、平移、剪切和噪声注入^[4-5,11]等方法组合使用,对光谱数据进行扩充或增加光谱训练样本的数量。研究中采用Keras的图像生成器ImageDataGenerator类实现光谱数据增强,通过ImageDataGenerator生成每一个批次的训练光谱图,适时地进行数据增强处理,可以随机地增强训练数据,避免了样本数据稀少的问题^[12]。

通过数据扩充,用于深度模型训练的数据集包括 92#汽油、机油、10#柴油、汽油+机油+柴油混合物等样本的原始光谱分别由 100 个增加到 2000 个。由于 ImageDataGenerator 图像生成器的功能特性是在深度学习模型训练过程中,根据生成器的构造函数参数设置实时数据增强,生成张量光谱图数据批次,且可以循环迭代,因此,数据增强的样本数量是不确定的,可以理解为无限^[12]。模型训练时,从每种样本中随机选择 70% 作为训练集,剩下 30% 作为验证集。

3 分析与讨论

3.1 CCN 模型构建

三维荧光光谱实质上是一个以激发-发射波长构成的二维区域上分布的离散值,即激发波长和发射波长作为矩阵的行序和列序,对应位置的荧光强度为矩阵元素值^[3-4]。该矩阵反映了物质荧光强度的状态与变化趋势,如同一个图片的像素值反映了图片的内容。传统方法是通过荧光光谱形状特征进行物质的定性识别,利用深度学习全自动数据分析和表征学习能力,根据 CNN 方法基于荧光数据与待测物质的种类建立映射关系,实现对物质种类的鉴别,此方法与传统方法相比,能够更好地挖掘原始光谱的信息特征^[13-14]。

根据实验获取的油品荧光数据特点,在保证深度学习模型分类精度的同时,尽可能减少模型的训练时间。文中的 CNN 模型基于 Python 的深度学习框架 Keras 构建,如图 10 所示。网络含有 3 个卷积层,3 个池化层,2 个全连接层(FC layer),1 个多标签分类层(ML layer)。为了增强网络各层的非线性输出,避免网络节点输出零,造成梯度消失现象,激活函数选择 ReLU 函数,损失函数为二元交叉熵函数。网络采用 RMSProp 优化算法对模型进行更新训练, N 为迭代次数(200~1000),卷积核大小为 3×3 、步长为 1,池化层采用最大池化法,池化核为 2×2 、步长为 2、学习率为 0.0001、分类数为 3。

3.2 模型训练与种类识别

实验时,利用搭建好的深度学习模型,将处理后的荧光光谱数据批次输入模型进行训练,模型训练的评估指标由分类精度和损失曲线表示,如图 11 所示。

由图 11(a)可知,本研究设计的深度学习模型在训练开始时收敛速度较快,大约经历 35 次迭代后,训练集分类精度可达到 92.5%,验证集精度则能达 95% 以上;此后缓慢上升,在训练迭代 150 次后,训练集和验证集分类精度均达 98%,变化更加缓慢;迭代 200 次时,训练集分类精度达 99.76%,模型变化趋于稳定。图 11(b)体现了 CNN 模型训练集和验证集 Loss 值的变化情况,其规律与分类精度曲线相反,从开始迭代至迭代 50 次时,Loss 值迅速下降,随着迭代次数的增加,Loss 值开始缓慢下降,直至迭代次数达到 125 次时,模型收敛接近最优。验证集 Loss 值的幅度变化较训练

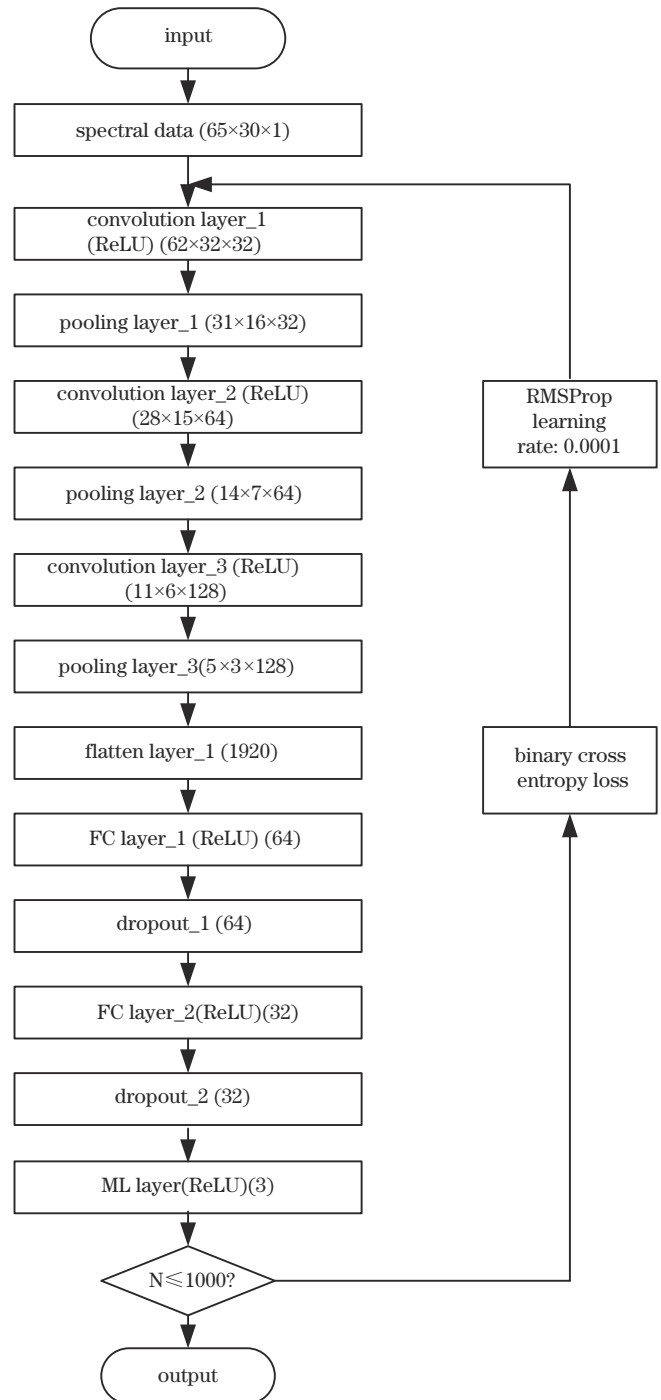


图 10 深度学习模型流程图

Fig. 10 Flow chart of deep learning model

集 Loss 值更加显著,随着迭代次数的增加,能够较快地收敛至最优。训练好的模型测试结果如表 3 所示,共测试 10 次,取其平均值。

由表 3 可知,深度学习训练好的模型对 3 种油品的综合识别率为 99.76%,这表明深度学习模型能够准确分类识别汽油、柴油、机油的荧光光谱,深度学习模型的性能良好。

3.3 深度学习模型性能测试

将训练好的 CNN 模型分别对单组分油品测试样本、双组分混合测试样本和 3 组分混合测试样本

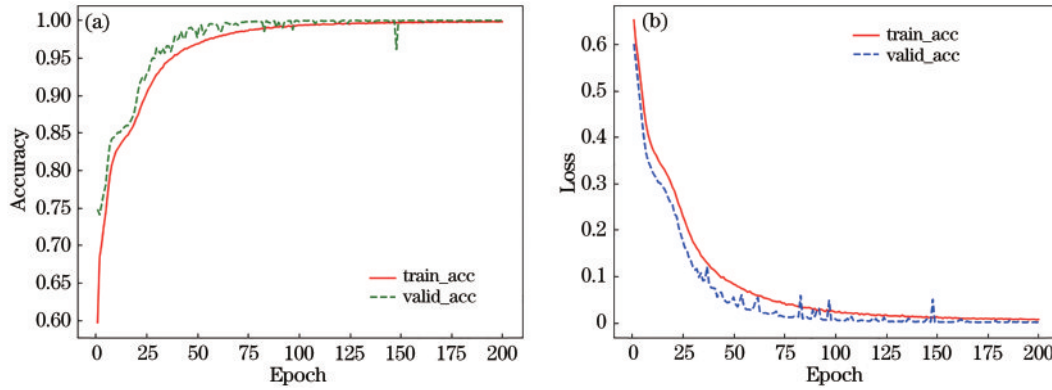


图 11 CNN 模型训练过程图。(a)分类精度;(b)损失函数

Fig. 11 Deep learning model training process curves. (a) Classification accuracy; (b) loss function

表 3 石油光谱的训练模型测试结果

Table 3 Test result of petroleum spectra

| Type | Iteration | Accuracy / % | Loss |
|----------------|-----------|--------------|-------|
| Training set | 200 | 99.76 | 0.015 |
| Validation set | 200 | 99.76 | 0.015 |

的光谱数据进行测试实验,平均测试结果如表 4 所示。

由表 4 测试结果可知,在油品光谱识别中,深度学习模型对单物质如汽油、柴油、机油的分类准确率达

表 4 石油光谱的分类测试结果

Table 4 Results of classification test of petroleum spectra

| Sample | Sample number | Right sample number | Accuracy / % |
|---------------------------------|---------------|---------------------|--------------|
| Gasoline | 12 | 12 | 100 |
| Diesel oil | 10 | 10 | 100 |
| Machine oil | 11 | 11 | 100 |
| Gasoline+diesel oil | 19 | 17 | 89.47 |
| Gasoline + machine oil | 21 | 17 | 80.95 |
| Machine oil+diesel oil | 24 | 22 | 91.67 |
| Gasoline+machine oil+diesel oil | 19 | 16 | 84.21 |
| Comprehensive test | | | 82.65 |

100%。对混合物质分类时,光谱没有明显重叠,光谱特征分辨清晰的分类精度相对较高,如机油与柴油的混合、汽油与柴油的混合,如图 7、图 8 所示,机柴油混合时的荧光光谱重叠少,因而其识别准确率高达 91.67%,汽柴油混合光谱部分重叠,分类准确率为 89.47%,而机油与汽油混合的光谱重叠严重,分辨率低,因而其分类测试准确率为 80.95%。3 种油混合时,由于光谱更加复杂,分类准确率为 84.21%。最终模型测试的综合分类准确率为 82.65%,模型分类识别性能较好。

4 结 论

石油类污染物组成复杂,传统荧光分析鉴别困难大。提出了一种直接通过荧光光谱数据进行物质种类识别的 CNN 方法。在荧光实验基础上获取了汽油、柴油、机油及其混合物的荧光光谱数据,建立了被测油品样本的荧光数据集,构建了基于 Python 深度学习框架 Keras 的 CNN 模型,并对 CNN 模型进行了训练、验证与测试实验,实现了被测油品的种类判别。实验表明,该模型对 3 种油品的训练验证的分类准确率都达到了 99.76%,对 3 种油品混合时的综合测试分类准确率达到 82.65%,对单物质的分类准确率达 100%。结果表明,三维荧光技术结合深度学习算法实现对石油油品的种类识别是可行的。该方法为进一步研究基于深度

学习的水环境污染物质智能识别系统提供了一种新思路与新方法。

参 考 文 献

- [1] 唐远河,刘青松,蒙磊,等. 荧光光谱智能监测水中矿物油浓度的研究[J]. 光谱学与光谱分析, 2015, 35(2): 424-428.
Tang Y H, Liu Q S, Meng L, et al. Study on the concentration of mineral oil in water by online intelligent detection based on fluorescence spectrum[J]. Spectroscopy and Spectral Analysis, 2015, 35(2): 424-428.
- [2] 左兆陆. 土壤石油烃类污染物荧光测量技术及应用研究[D]. 合肥: 中国科学技术大学, 2020: 16-40.
Zuo Z L. Application study on fluorescence measurement technology of petroleum hydrocarbon pollutants in soil

- [D]. Hefei: University of Science and Technology of China, 2020: 16-40.
- [3] 来文豪, 周孟然, 王亚, 等. 深度学习与激光诱导荧光在假酒识别中的应用[J]. 激光与光电子学进展, 2018, 55(4): 043001.
Lai W H, Zhou M R, Wang Y, et al. Application of counterfeit liquor recognition based on deep learning and laser induced fluorescence[J]. Laser & Optoelectronics Progress, 2018, 55(4): 043001.
- [4] 赵志蕾. 基于荧光光谱与深度学习的植物油品质鉴别研究[D]. 秦皇岛: 燕山大学, 2021: 23-26.
Zhao Z L. Research on identification of vegetable oil quality based on fluorescence spectroscopy and deep learning[D]. Qinhuangdao: Yanshan University, 2021: 23-26.
- [5] 丛琳. 基于三维荧光光谱的蜂蜜掺假智能检测技术研究[D]. 天津: 天津工业大学, 2019: 22-23.
Cong L. Research on intelligent detection technology of honey adulteration based on three-dimensional fluorescence spectrum[D]. Tianjin: Tianjin Polytechnic University, 2019: 22-23.
- [6] 张学成, 金尚忠, 赵天琦, 等. 基于深度学习的气溶胶荧光光谱识别应用研究[J]. 应用光学, 2022, 43(3): 466-471.
Zhang X C, Jin S Z, Zhao T Q, et al. Application of aerosol fluorescence spectrum recognition based on deep learning[J]. Journal of Applied Optics, 2022, 43(3): 466-471.
- [7] Xu R Z, Cao J S, Feng G Y, et al. Fast identification of fluorescent components in three-dimensional excitation-emission matrix fluorescence spectra via deep learning[J]. Chemical Engineering Journal, 2022, 430: 132893.
- [8] Gan J H, Zhou L H, Cui J, et al. Classification of blood species using fluorescence spectroscopy combined with deep learning method[J]. Journal of Applied Mathematics and Physics, 2019, 07(10): 2324-2332.
- [9] 刘金香, 班伟, 陈宇, 等. 融合多维度 CNN 的高光谱遥感图像分类算法[J]. 中国激光, 2021, 48(16): 1610003.
Liu J X, Ban W, Chen Y, et al. Multi-dimensional CNN fused algorithm for hyperspectral remote sensing image classification[J]. Chinese Journal of Lasers, 2021, 48(16): 1610003.
- [10] 孔德明, 李雨蒙, 崔耀耀, 等. 海面溢油三维荧光光谱消除瑞利散射方法的研究[J]. 光谱学与光谱分析, 2020, 40(9): 2791-2797.
Kong D M, Li Y M, Cui Y Y, et al. Correction methods of Rayleigh scattering of three-dimensional fluorescence spectra of spilled oil on sea[J]. Spectroscopy and Spectral Analysis, 2020, 40(9): 2791-2797.
- [11] Wu X J, Zhao Z L, Tian R L, et al. Exploration of total synchronous fluorescence spectroscopy combined with pre-trained convolutional neural network in the identification and quantification of vegetable oil[J]. Food Chemistry, 2021, 335: 127640.
- [12] 弗朗索瓦·肖莱. Python 深度学习[M]. 张亮, 译. 北京: 人民邮电出版社, 2018: 111-113.
Francois C. Deep learning with Python[M]. Zhang L, Transl. Beijing: Posts & Telecom Press, 2018: 111-113.
- [13] 杨国威, 周楠, 杨敏, 等. 融合卷积神经网络和相关滤波的焊缝自动跟踪[J]. 中国激光, 2021, 48(22): 2202011.
Yang G W, Zhou N, Yang M, et al. Automatic weld tracking based on convolution neural network and correlation filter[J]. Chinese Journal of Lasers, 2021, 48(22): 2202011.
- [14] 刘喆, 孟辉, 张永彬, 等. 基于 LLE 与 Gold-SA-SVM 的海藻三维荧光光谱门类识别[J]. 光学学报, 2022, 42(4): 0430001.
Liu Z, Meng H, Zhang Y B, et al. Category recognition of three-dimensional fluorescence spectra of algae based on LLE and gold-SA-SVM[J]. Acta Optica Sinica, 2022, 42(4): 0430001.