

# 激光与光电子学进展

## 小样本太赫兹光谱识别

崔向伟<sup>1,2</sup>, 沈韬<sup>1,2\*</sup>, 刘英莉<sup>1,2</sup>, 朱艳<sup>1,2</sup>, 朱荣盛<sup>1,2</sup>

<sup>1</sup>昆明理工大学信息工程与自动化学院, 云南 昆明 650504;

<sup>2</sup>昆明理工大学云南省计算机技术应用重点实验室, 云南 昆明 650504

**摘要** 物质的太赫兹光谱具有独特的“指纹谱”特性, 可以利用该特性对物质进行识别。随着人工智能技术的发展, 深度学习算法在太赫兹光谱识别领域得到了越来越广泛的应用。然而在实际应用中, 受实验设备、实验条件以及实验环境等因素的影响, 所获取的太赫兹光谱数据并不总是大规模的, 无法满足深度学习算法所需的数据量要求。为了解决这一问题, 本文提出了一种基于生成对抗网络(GAN)的太赫兹光谱识别方法。首先利用 S-G 滤波器和三次样条插值法对物质的太赫兹光谱数据进行预处理, 然后通过 GAN 生成具有真实太赫兹光谱数据分布的仿真数据, 最后将生成的数据以及真实光谱数据作为训练样本对深层神经网络进行训练, 从而得出物质的识别结果。实验结果表明: GAN 模型生成的太赫兹光谱数据可以有效地模拟真实太赫兹光谱数据的总体特征, 扩充太赫兹光谱数据样本, 极大地提高了光谱的识别精度。

**关键词** 光谱学; 太赫兹光谱; 深度学习; 深层神经网络; 生成对抗网络

中图分类号 O433.5

文献标志码 A

doi: 10.3788/LOP202158.0130001

## Recognition of Small-Sample Terahertz Spectrum

Cui Xiangwei<sup>1,2</sup>, Shen Tao<sup>1,2\*</sup>, Liu Yingli<sup>1,2</sup>, Zhu Yan<sup>1,2</sup>, Zhu Rongsheng<sup>1,2</sup>

<sup>1</sup>Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan 650504, China;

<sup>2</sup>Computer Technology Application Key Lab of Yunnan Province, Kunming University of Science and Technology, Kunming, Yunnan 650504, China

**Abstract** Due to the unique “fingerprint spectrum” characteristic, terahertz (THz) spectrum can be used to recognize the materials. With the development of artificial intelligence, deep learning is widely used in the field of THz spectrum recognition. However, the acquired THz spectral data are not always on a large scale due to the influence of experimental equipment, conditions and environment, which cannot meet the data size requirements of the deep learning algorithm. In order to solve this problem, we proposed a method of THz spectrum recognition based on generative adversarial networks (GAN) in this paper. Firstly, an S-G filter and a cubic spline interpolation method were employed to pre-process the THz spectral data. Secondly, the simulation data with the distribution of real THz spectral data were generated by the GAN. Finally, the generated data and real spectral data were taken as the training samples to train the deep neural networks (DNN), thus obtaining the recognition results of the materials. The experimental results show that the THz spectral data generated by the GAN model can effectively simulate the overall characteristics of real THz spectral data and expand the THz spectral data samples, greatly elevating the spectral recognition accuracy.

收稿日期: 2020-04-13; 修回日期: 2020-04-23; 录用日期: 2020-06-12

基金项目: 国家自然科学基金(61971208、61671225)、云南省应用基础研究计划项目重点项目(2018FA034)、云南省万人计划青年拔尖人才项目(云南省人社厅 201873)

\*E-mail: shentao@kust.edu.cn

**Key words** spectroscopy; terahertz spectrum; deep learning; deep neural network; generative adversarial networks

**OCIS codes** 300.6495; 040.2235; 300.6300

## 1 引言

太赫兹 (THz) 波是指频率范围为 0.1~10 THz 的电磁波, 其波段能够覆盖等离子体、半导体、有机体和生物大分子等物质的特征谱<sup>[1-3]</sup>。太赫兹技术作为一门新兴的学科, 近年来发展十分迅速。作为“改变未来世界的十大技术”之一, 太赫兹技术已被广泛应用于雷达、遥感、高速数据通信与传输、医学成像等领域<sup>[4-6]</sup>。许多复合晶体的共振频率都在太赫兹波段, 因此太赫兹光谱可以在一定程度上反映晶体中声子的振动模式, 这使得太赫兹技术具有识别物质和分析物质成分的潜力<sup>[7-9]</sup>。随着太赫兹技术的应用越来越广泛, 国内外许多研究人员采用机器学习甚至是深度学习的方法来分析 and 处理太赫兹光谱数据, 如: 张文涛等<sup>[10]</sup>采用主成分分析和支持向量机完成了转基因大豆油光谱的检测; Liu 等<sup>[11]</sup>采用连续投影算法 (SPA) 和偏最小二乘算法 (PLS) 对转基因食用油进行了鉴别; 胡晓华等<sup>[12]</sup>将主成分分析和基于粒子群参数寻优的支持向量机分析方法应用于不同产地咖啡豆的鉴别, 他们先采用主成分分析对太赫兹光谱进行分析, 最后构造了一个粒子群参数寻优的支持向量机识别模型, 该模型对三种不同产地咖啡豆的综合识别准确率可达到 95% 以上。但当物质的太赫兹光谱曲线整体上非常相似时, 机器学习的方法就难以提取有效的光谱特征。针对这一问题, 马帅等<sup>[13]</sup>提出了一种基于深层信念网络 (DBN) 的太赫兹光谱识别方法。他们首先利用 DBN 自动提取太赫兹光谱数据的特征, 然后通过实验证明了利用 DBN 自动提取的特征可以在特征不明显的光谱数据中取得良好的分类效果。虞浩跃等<sup>[14]</sup>提出了一种基于双向长短期记忆网络 (BLSTM-RNN) 的太赫兹光谱识别方法, 该方法利用深度学习算法自动提取光谱特征。他们采用该方法对 15 种化合物进行了分类识别, 实验结果表明, 基于双向长短期记忆网络的算法可以自动提取光谱特征, 同时又能保证识别的准确率。然而, 深度学习算法需要大规模标注训练数据集, 当数据量较小时, 训练效果将大打折扣。目前, 在小样本数据量训练领域, 人们主要采用基于有监督学习、迁移学习以及数据增强的方法进行研究。生成对抗网络 (GAN) 是数据增强的一

种, 它通过生成器和判别器的不断博弈可以生成任意数据分布的仿真数据, 从而起到扩充训练样本的作用。针对太赫兹光谱训练样本不足的问题, 本文研究了基于 GAN 的太赫兹光谱识别方法。本文首先通过 GAN 模型生成数据, 以扩充太赫兹光谱数据集, 然后使用扩充后的数据集训练深层神经网络 (DNN) 模型, 得到了太赫兹光谱的识别结果。

## 2 理论部分

### 2.1 算法的主要思想

GAN 是 Goodfellow 等<sup>[15]</sup>在 2014 年提出的网络模型。最基本的 GAN 由生成器  $G$  和判别器  $D$  两个模型组成。生成器  $G$  的作用是给定一个随机噪声  $z$  生成样本  $G(z)$ , 而判别器  $D$  实质上是一个判别网络, 它的输出表示生成样本  $G(z)$  与真实样本相比较是否为真。通过二者不断地相互博弈学习, GAN 模型可以生成尽可能模拟真实数据分布, 又能与真实数据区分开的“数据样本”。本文中的真实样本指的是真实的太赫兹光谱数据。GAN 模型可以生成能模拟真实太赫兹光谱数据分布的仿真数据。DNN 又被称为多层感知机, 其最基本的网络结构由输入层、输出层以及隐藏层组成。将待识别物质的太赫兹光谱数据通过输入层输入至 DNN 网络, DNN 就会自动提取太赫兹光谱数据的特征并用于识别, 最终将物质的识别运算结果通过输出层输出。与直接输入原始训练数据集不同, 通过 GAN 模型生成的仿真光谱数据扩充了训练样本且增加了数据的多样性, 能有效提高光谱识别的准确率。

### 2.2 模型搭建

针对太赫兹光谱数据样本不足情况下导致识别效果较差这一问题, 本文提出了一种通过 GAN 生成太赫兹光谱数据以扩充数据集, 然后利用 DNN 对其进行识别的方法。GAN 的网络结构如图 1 所示。

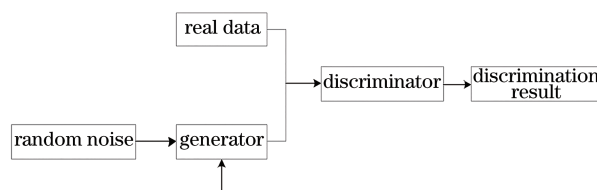


图 1 GAN 网络结构

Fig. 1 Network structure of GAN

GAN 通过判别器 D 和生成器 G 的相互博弈产生“以假乱真”的数据样本<sup>[16]</sup>。GAN 模型的构建过程如下：

判别器 D 是一个判别式网络，判定接收到的样本是否来自真实数据  $p_{\text{data}}(x)$ ，因此有

$$E_{x \sim p_{\text{data}}(x)} \{ \lg [D(x)] \}, \quad (1)$$

式中： $E$  指代期望，可根据正类（即判别出  $x$  是否属于真实数据）的对数函数构建。

生成器 G 生成尽可能服从真实数据分布的样本  $G(z)$ ，即生成器 G 通过训练不断提高欺骗判别器的概率。根据负类的对数函数构建

$$E_{z \sim p_z(z)} \{ \lg \{ 1 - D[G(z)] \} \}, \quad (2)$$

式中： $z$  为随机噪声； $p_z(z)$  为随机噪声的数据分布；

$D[G(z)]$  为判别器输出值。

GAN 训练过程的实质是训练判别器 D，使判别正确率最大化；而训练生成器 G，则是生成尽可能真实的数据，以使判别正确率最小，从而达到以假乱真的目的。目标函数为

$$\min_D \max_G V(D, G) = E_{x \sim p_{\text{data}}(x)} [ \lg D(x) ] + E_{z \sim p_z(z)} \{ \lg \{ 1 - D[G(z)] \} \}. \quad (3)$$

通过以上训练过程，GAN 模型的输出就是能模拟真实光谱数据分布的仿真数据。最终将生成的太赫兹光谱数据和原始太赫兹光谱数据共同输入到 DNN 分类器中。DNN 会在不断的迭代训练中自动学习太赫兹光谱数据特征，并将其用于识别，得出识别结果。DNN 的网络模型如图 2 所示。

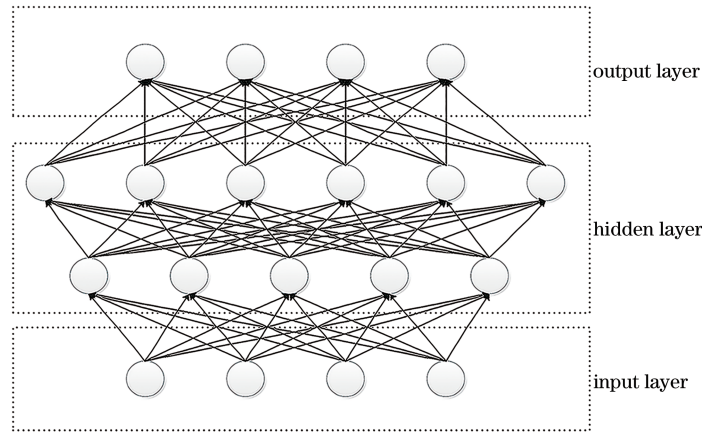


图 2 DNN 网络模型

Fig. 2 Network model of DNN

### 3 光谱数据处理

受测试环境和测试设备的影响，经离散傅里叶变换收集得到的太赫兹时域信号中通常包含着许多噪声信息。此外，不同的太赫兹光谱采集设备得到的光谱曲线中包含的数据点的数量也不相同，不利于太赫兹光谱的识别。因此，需要有效地过滤光谱数据以减少噪声，并且需要将数据转换为统一格式，以便于进一步对光谱进行识别。太赫兹光谱数据处理过程包括 S-G 滤波和三次样条插值。

#### 1) S-G 滤波

这里假设对某一物质的太赫兹光谱数据进行平滑去噪处理。以光谱数据上的任意一点为例进行说明，以该点为中心，前后分别取  $n$  个点，对这些点进行多项式拟合<sup>[17]</sup>。拟合公式为

$$Y_i = c_0 + c_1 i + c_2 i^2 + \dots + c_p i^p, \quad (4)$$

式中： $i$  为变量； $c_0, c_1, \dots, c_p$  为系数。则拟合上述多项式后的累积误差为

$$S = \sum_{j=-m}^m (Y_j - y_j)^2, \quad (5)$$

式中： $y_j$  为光谱数据的实际值； $Y_j$  为光谱数据的拟合值。

对各误差项求偏导并使其值为 0，即

$$\begin{cases} \frac{\partial S}{\partial c_0} = 2 \sum_{j=-m}^m (c_0 + c_1 i + c_2 i^2 + \dots + c_p i^p - y_i) = 0 \\ \frac{\partial S}{\partial c_1} = 2 \sum_{j=-m}^m j (c_0 + c_1 i + c_2 i^2 + \dots + c_p i^p - y_i) = 0 \\ \frac{\partial S}{\partial c_2} = 2 \sum_{j=-m}^m j^2 (c_0 + c_1 i + c_2 i^2 + \dots + c_p i^p - y_i) = 0, \\ \dots \\ \frac{\partial S}{\partial c_p} = 2 \sum_{j=-m}^m j^p (c_0 + c_1 i + c_2 i^2 + \dots + c_p i^p - y_i) = 0 \end{cases} \quad (6)$$

这样可使累计误差值最小。然后提取各项系数得到

$$\begin{cases} c_0 \sum_{j=-m}^m 1 + c_1 \sum_{j=-m}^m j + c_2 \sum_{j=-m}^m j^2 + \dots + c_p \sum_{j=-m}^m j^p = \sum_{j=-m}^m y_j \\ c_0 \sum_{j=-m}^m j + c_1 \sum_{j=-m}^m j^2 + c_2 \sum_{j=-m}^m j^3 + \dots + c_p \sum_{j=-m}^m j^{p+1} = \sum_{j=-m}^m j y_j \\ c_0 \sum_{j=-m}^m j^2 + c_1 \sum_{j=-m}^m j^3 + c_2 \sum_{j=-m}^m j^4 + \dots + c_p \sum_{j=-m}^m j^{p+2} = \sum_{j=-m}^m j^2 y_j \\ \dots \\ c_0 \sum_{j=-m}^m j^p + c_1 \sum_{j=-m}^m j^{p+1} + c_2 \sum_{j=-m}^m j^{p+2} + \dots + c_p \sum_{j=-m}^m j^{p+p} = \sum_{j=-m}^m j^p y_j \end{cases} \quad (7)$$

当  $\sum_{j=-m}^m j^r$  为奇数时, (7) 式为 0。进一步化简 (7) 式并对线性方程组进行求解, 可得系数序列  $(c_0, c_1, c_2, \dots, c_p)$ 。根据 (4) 式即可得到该光谱数据

点的拟合值, 然后用拟合值代替原来点的数值, 以达到去除高频噪声点的目的。以 riboflavin 的太赫兹透射光谱数据为例, 对其进行 S-G 滤波处理, 得到的光谱曲线如图 3 所示。

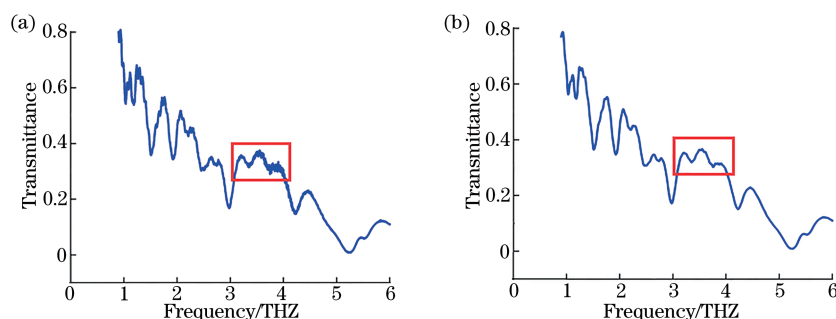


图 3 S-G 滤波前后 riboflavin 的太赫兹光谱曲线。(a) 滤波前; (b) 滤波后

Fig. 3 THz spectral curves of riboflavin before and after S-G filtering. (a) Before filtering; (b) after filtering

## 2) 三次样条插值

太赫兹光谱数据经 S-G 滤波处理后, 就得到了去除噪声的光谱。为了使光谱曲线更加平滑, 本文使用三次样条插值法对曲线进行拟合, 这样可以根据有限个点的的光谱数据得到一条平滑曲线。

给定太赫兹光谱数据  $(x_i, y_i), i = 0, 1, \dots, n$ , 则在这批数据中间有插值数据节点  $a = x_0 < x_1 < x_2 < \dots < x_n = b$ , 本文要在插值节点区间内寻找一个函数  $S(x)$ , 使其满足下列条件:

$$a) S(x) \in C^2[a, b];$$

b)  $S(x)$  在每个子区间  $[x_{i-1}, x_i]$  上寻找能满足插值条件  $s(x_i) = y_i$  的三次多项式的三次样条函数, 其中,  $i = 1, 2, \dots, n$ 。

利用以上插值函数代替原先的有限数据点便可拟合出光滑的太赫兹光谱曲线。

采用以上过程就可以完成太赫兹光谱数据的处理。下面以 maltotetraose、maltopentaose 和 malthexaose 在 0.9~6 THz 之间的太赫兹光谱数据为例具体说明: 首先使用 S-G 滤波器对太赫兹频谱

进行滤波处理, 然后用三次样条插值法进行拟合, 最终得到了曲线平滑的光谱数据。原始数据和预处理后数据的对比如图 4 所示。

由图 4 可以看出: 在光谱数据未进行处理前, 曲线整体比较粗糙, 且不同物质的光谱曲线的纵轴单位未统一; 经 S-G 滤波和三次样条插值法处理后, 得到了去除噪声、平滑且单位统一的太赫兹光谱曲线。

## 4 实验与结果分析

### 4.1 实验数据

本文以 anthraquinone、benomyl、carbazole、mannose、riboflavin、malthexaose、maltotetraose、maltopentaose、maltotriose 在 0.9~6 THz 频段范围的太赫兹透射光谱曲线 (每种 20 条, 共计 200 条) 作为实验数据, 这些数据均来自日本理化研究所太赫兹光谱数据库, 其网址为 [www.riken.jp/THzdatabase/](http://www.riken.jp/THzdatabase/)。对实验数据进行数据预处理后, 从每一种物质的太赫兹光谱数据集中随机挑选一条绘制成光谱曲线, 如图 5 所示。



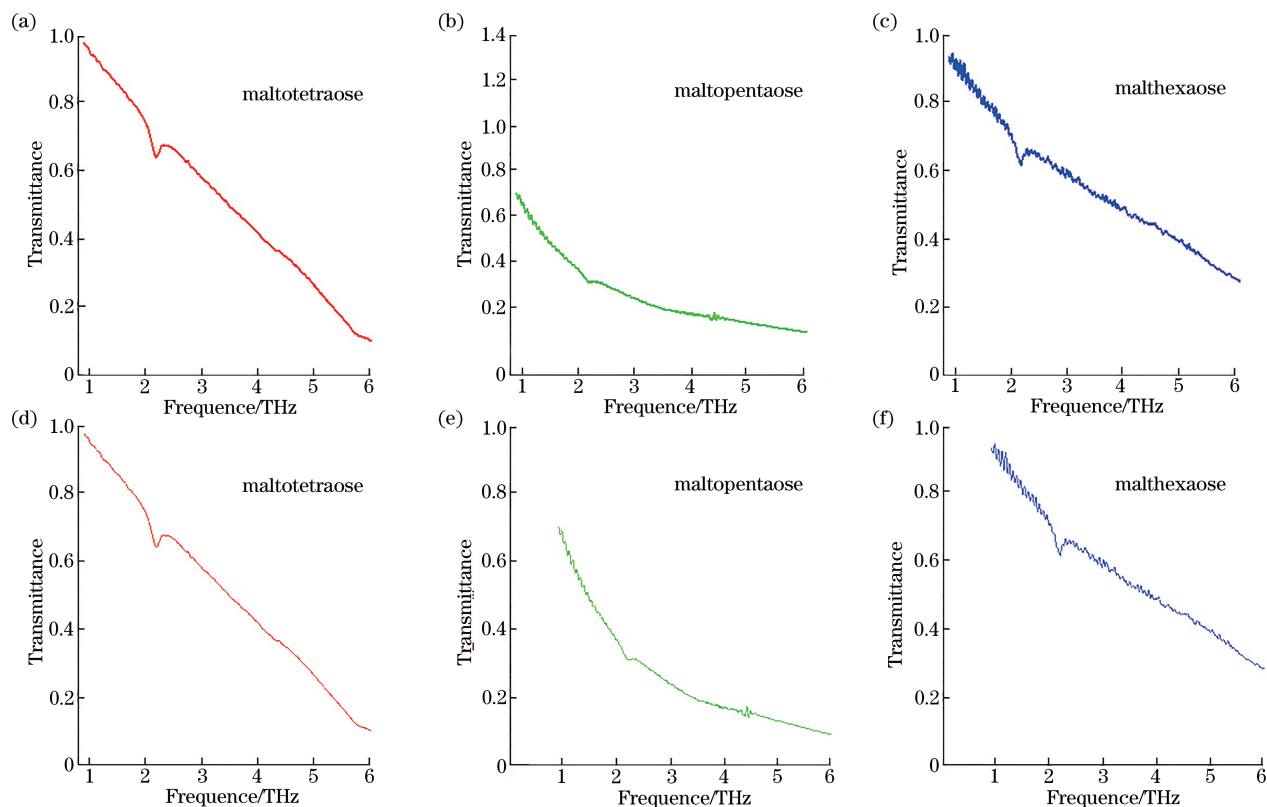


图 4 数据预处理前后的对比。(a)~(c)预处理前;(d)~(f)预处理后

Fig. 4 Comparison before and after data preprocessing. (a)~(c) Before preprocessing; (d)~(f) after preprocessing

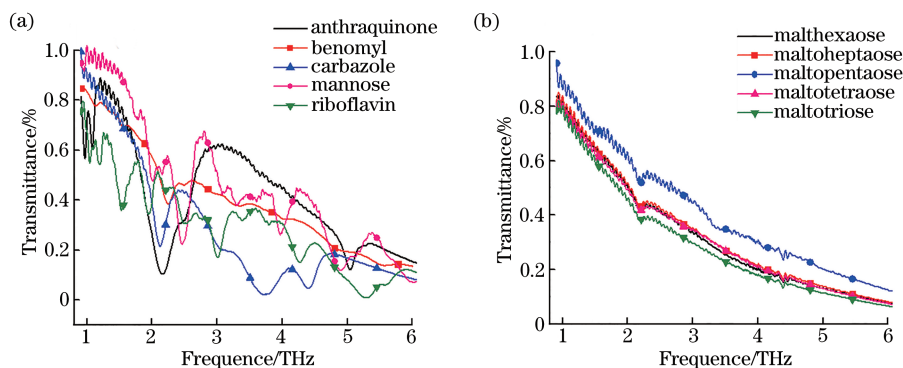


图 5 10 种物质的太赫兹光谱曲线

Fig. 5 THz spectra of 10 substances

#### 4.2 实验过程及结果分析

将 10 种物质的太赫兹透射光谱分为 Dataset-1、Dataset-2 和 Dataset-3 数据集。其中:Dataset-1 为 5 种物质 anthraquinone、benomyl、carbazole、mannose、riboflavin 的太赫兹透射光谱数据集,这些物质的太赫兹透射光谱均具有明显的峰值特征,易于人工定义特征;Dataset-2 为 5 种物质 malthexaose、maltoheptaose、maltopentaose、maltotetraose、maltotriose 的太赫兹透射光谱数据集,这些物质的太赫兹透射光谱都没有明显的峰值特征,且曲线整体较为相似,不便于

人工定义特征;Dataset-3 为以上 10 种物质的太赫兹透射光谱数据集。

利用 GAN 模型对以上 10 种物质的太赫兹光谱数据进行扩充,每种物质均生成 980 条光谱数据。最终生成的太赫兹光谱数据与原始 Dataset-1、Dataset-2、Dataset-3 数据集一起形成扩充数据集 Augmented dataset-1、Augmented dataset-2 和 Augmented dataset-3。其中:Augmented dataset-1 是 Dataset-1 数据与生成数据的集合 (anthraquinone、benomyl、carbazole、mannose、riboflavin 各 1000 条,

共计 5000 条), Augmented dataset-2 是 Dataset-2 数据与生成数据的集合 (malthexaose、maltoheptaose、maltopentaose、maltotetraose、maltotriose 各 1000 条, 共计 5000 条), Augmented dataset-3 是 Augmented dataset-1 和 Augmented dataset-2 太赫兹光谱数据的集合。

下面以 maltotetraose 为例, 展示通过 GAN 模型

生成光谱数据的过程, 如图 6 所示。其中, 图 6(a) 为原始光谱数据图, 图 6(b) 为 GAN 模型迭代 1000 次生成的数据图, 图 6(c) 为 GAN 模型迭代 5000 次生成的数据图, 图 6(d) 为 GAN 模型迭代 10000 次生成的数据图, 图 6(e) 为 GAN 模型迭代 100000 次生成的数据图, 图 6(f) 为 GAN 模型迭代 200000 次生成的数据图。

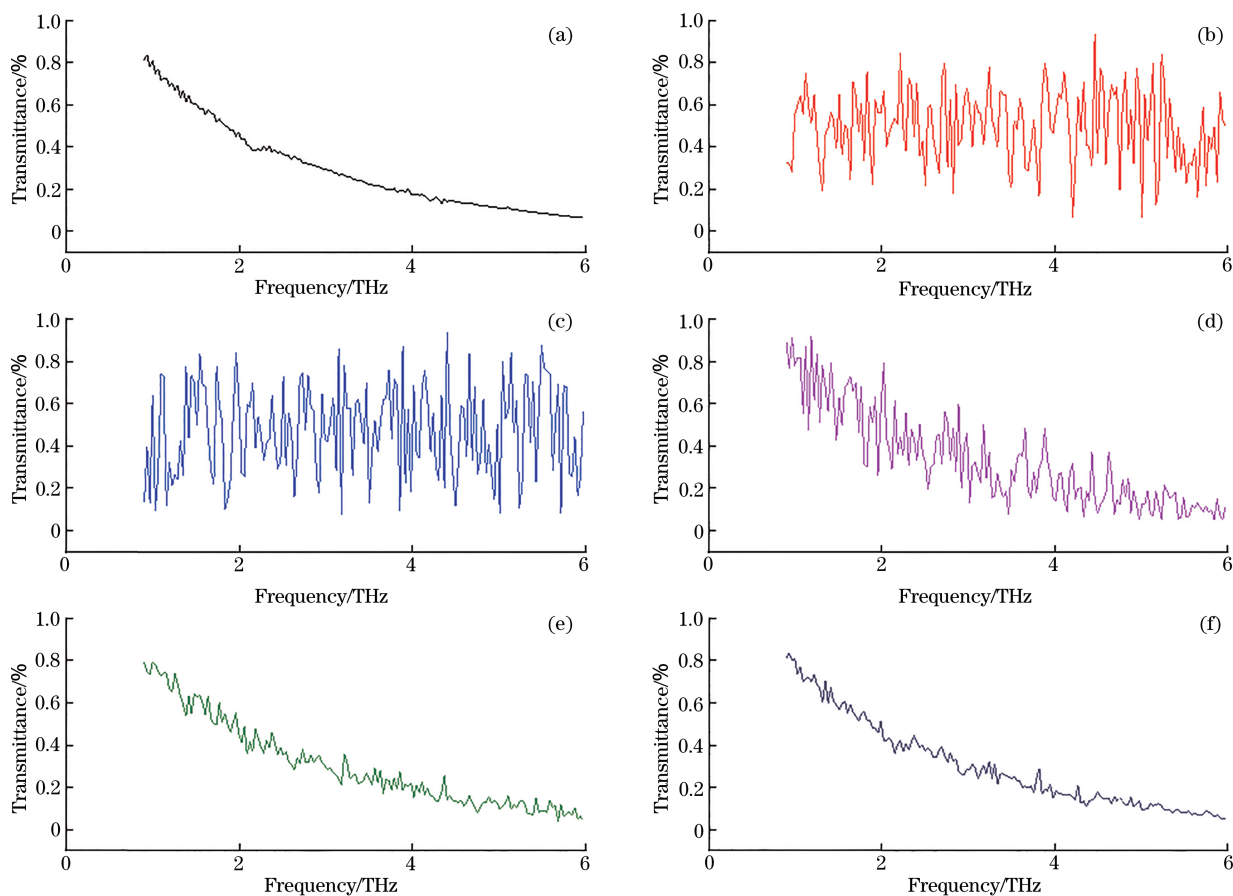


图 6 不同迭代次数下 GAN 生成的数据图。(a) 原始光谱数据图; (b) 迭代 1000 次生成的数据图; (c) 迭代 5000 次生成的数据图; (d) 迭代 10000 次生成的数据图; (e) 迭代 100000 次生成的数据图; (f) 迭代 200000 次生成的数据图

Fig. 6 Data graphs generated by GAN under different iterations. (a) Original spectral data graph; (b) data graph generated by 1000 iterations; (c) data graph generated by 5000 iterations; (d) data graph generated by 10000 iterations; (e) data graph generated by 100000 iterations; (f) data graph generated by 200000 iterations

从图 6 可以看出: 当 GAN 模型迭代次数较少时, 模型的输出跟真实太赫兹光谱数据有很大差别; 随着迭代次数增加, GAN 模型的输出开始逐渐与原始光谱数据相似, 且整体曲线越来越相似。这表明, 通过 GAN 模型生成的太赫兹光谱数据可以有效映射原始光谱数据。

通过以上步骤, 将 GAN 模型迭代 200000 次后生成的太赫兹光谱数据保存下来, 以扩充训练数据集。原始数据集与扩充数据集中的样本数量如表 1 所示。

表 1 原始数据集与扩充数据集中的样本数量对比

Table 1 Comparison of sample size between original and extended data sets

Data set	Sample size
Dataset-1	100
Dataset-2	100
Dataset-3	200
Augmented dataset-1	5000
Augmented dataset-2	5000
Augmented dataset-3	10000

原始待识别物质的太赫兹光谱数据集只有 100~200 条,训练样本数量非常少,无法满足 DNN 所需的大规模数据集的要求,导致分类准确率不高。本文通过 GAN 网络模型生成了一部分太赫兹光谱数据以扩充数据集,从表 1 中可以看出,扩充后的训练数据集有 5000~10000 条。最后构造一个 DNN,并分别对原始数据集和扩充数据集进行分类识别。DNN 分类准确率为分类准确样本数量与总样本数量之比。

在实验中,为了验证 GAN 模型生成的太赫兹光谱数据的有效性,利用同一分类模型分别对扩充数据集和原始数据集进行训练并得出识别结果。同时,为了说明深度学习算法在太赫兹光谱识别领域的优越性,本文还对比了不同分类器下的识别准确率。表 2 列出了不同分类器的识别准确率。

表 2 不同分类器的识别准确率

Table 2 Recognition accuracy rate of different classifiers

Classifier	Accuracy / %		
	SVM	LSTM	DNN
Dataset-1	90.25	91.34	93.26
Augmented dataset-1	91.17	93.35	97.89
Dataset-2	85.15	86.16	91.12
Augmented dataset-2	89.12	91.64	94.35
Dataset-3	86.23	87.64	92.14
Augmented dataset-3	89.86	92.12	94.87

Notes: LSTM represents long short-term memory and SVM represents support vector machine.

从表 2 中可以看出:针对同一数据集,SVM 的识别准确率低于 LSTM 的识别正确率;DNN 相较于 SVM 和 LSTM 具有更好的识别效果;对于不同的数据集,在同一分类器下,扩充数据集取得了更为优秀的分类效果。实验结果表明,GAN 能有效扩充太赫兹光谱数据集,且 DNN 能更好地对太赫兹光谱进行识别。同时,通过对比数据集 Augmented dataset-1 和 Augmented dataset-2 的识别准确率可以发现,数据集 Augmented dataset-2 具有较低的识别准确率。这主要是因为数据集 Augmented dataset-2 的光谱曲线的整体相似度较高,导致模型识别困难,而对于数据集 Augmented dataset-1,其光谱曲线有明显的波峰、波谷,更容易识别。

## 5 结 论

深度学习算法在处理大规模数据集方面具有很大优势,但当某些物质的太赫兹光谱数据样本不

足时,深度学习的方法将产生过拟合问题,难以有效识别目标物体。因此,本文提出了一种基于 GAN 的太赫兹光谱识别方法。该方法利用 GAN 模型自动生成与原始数据相匹配的数据,以此来扩充训练样本数据集,然后利用 DNN 对扩充数据集进行训练,得出识别结果。实验结果表明,该方法可以极大地提高太赫兹光谱的识别准确率,从而验证了该方法的有效性,为小样本数据集下物质的太赫兹光谱识别提供了新途径。

## 参 考 文 献

- [1] Tonouchi M. Cutting-edge terahertz technology [J]. *Nature Photonics*, 2007, 1(2): 97-105.
- [2] Baxter J B, Guglietta G W. Terahertz spectroscopy [J]. *Analytical Chemistry*, 2011, 83(12): 4342-4368.
- [3] Qin J Y, Xie L J, Ying Y B. Feasibility of terahertz time-domain spectroscopy to detect tetracyclines hydrochloride in infant milk powder [J]. *Analytical Chemistry*, 2014, 86(23): 11750-11757.
- [4] Liang M Y, Ren Z Y, Zhang C L. Progress of terahertz space exploration technology [J]. *Laser & Optoelectronics Progress*, 2019, 56(18): 180004. 梁美彦,任竹云,张存林.太赫兹空间探测技术研究进展[J].*激光与光电子学进展*, 2019, 56(18): 180004.
- [5] Guillet J P, Recur B, Frederique L, et al. Review of terahertz tomography techniques [J]. *Journal of Infrared, Millimeter, and Terahertz Waves*, 2014, 35(4): 382-411.
- [6] Hérault E, Garet F, Coutaz J L. On the possibility of identifying substances by remote active THz spectroscopy [J]. *IEEE Transactions on Terahertz Science and Technology*, 2016, 6(1): 12-19.
- [7] Liu X Q, Yao J L, Huang F, et al. Study on detection of penicillin drugs based on terahertz time-domain spectroscopy [J]. *Acta Optica Sinica*, 2020, 40(6): 0630001. 刘晓庆,姚嘉丽,黄凡,等.基于太赫兹时域光谱的青霉素类药物检测研究[J].*光学学报*, 2020, 40(6): 0630001.
- [8] Amenabar I, Lopez F, Mendikute A. In introductory review to THz non-destructive testing of composite mater [J]. *Journal of Infrared, Millimeter, and Terahertz Waves*, 2013, 34(2): 152-169.
- [9] Liu J X, Du B, Deng Y Q, et al. Terahertz-spectral identification of organic compounds based on

- differential PCA-SVM method[J]. Chinese Journal of Lasers, 2019, 46(6): 0614039.
- 刘俊秀, 杜彬, 邓玉强, 等. 基于差分-主成分分析-支持向量机的有机化合物太赫兹吸收光谱识别方法[J]. 中国激光, 2019, 46(6): 0614039.
- [10] Zhang W T, Li Y W, Zhan P P, et al. Recognition of transgenic soybean oil based on terahertz time domain spectroscopy and PCA-SVM [J]. Infrared and Laser Engineering, 2017, 46(11): 1125004.
- 张文涛, 李跃文, 占平平, 等. 基于太赫兹时域光谱技术与 PCA-SVM 的转基因大豆油鉴别研究[J]. 红外与激光工程, 2017, 46(11): 1125004.
- [11] Liu J J, Fan L L, Liu Y M, et al. Application of terahertz spectroscopy and chemometrics for discrimination of transgenic camellia oil [J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 206: 165-169.
- [12] Hu X H, Liu W, Liu C H, et al. Rapid identification of producing area of coffee bean based on terahertz spectroscopy and support vector machine[J]. Transactions of the Chinese Society of Agricultural Engineering, 2017, 33(9): 302-307.
- 胡晓华, 刘伟, 刘长虹, 等. 基于太赫兹光谱和支持向量机快速鉴别咖啡豆产地[J]. 农业工程学报, 2017, 33(9): 302-307.
- [13] Ma S, Shen T, Wang R Q, et al. Terahertz spectroscopic identification with deep belief network [J]. Spectroscopy and Spectral Analysis, 2015, 35(12): 3325-3329.
- 马帅, 沈韬, 王瑞琦, 等. 基于深层信念网络的太赫兹光谱识别[J]. 光谱学与光谱分析, 2015, 35(12): 3325-3329.
- [14] Yu H Y, Shen T, Zhu Y, et al. Terahertz spectral recognition based on bidirectional long short-term memory recurrent neural network[J]. Spectroscopy and Spectral Analysis, 2019, 39(12): 3737-3742.
- 虞浩跃, 沈韬, 朱艳, 等. 基于双向长短期记忆网络的太赫兹光谱识别[J]. 光谱学与光谱分析, 2019, 39(12): 3737-3742.
- [15] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [C]// Proceedings of the 27th International Conference on Neural Information Processing Systems, Dec. 2-5, 2014, Montreal, Canada. [S. l.]: IEEE Press, 2014: 135-142.
- [16] Creswell A, White T, Dumoulin V, et al. Generative adversarial networks: an overview [J]. IEEE Signal Processing Magazine, 2018, 35(1): 53-65.
- [17] Lu Y B, Liu W Q, Zhang Y J, et al. An adaptive hierarchical savitzky-golay spectral filtering algorithm and its application [J]. Spectroscopy and Spectral Analysis, 2019, 39(9): 2657-2663.
- 鲁一冰, 刘文清, 张玉钧, 等. 一种自适应层进式 Savitzky-Golay 光谱滤波算法及其应用[J]. 光谱学与光谱分析, 2019, 39(9): 2657-2663.