

DOI: 10.3969/j.issn.1671-024x.2025.01.008

基于卷积和 Transformer 的矿物拉曼光谱分类方法

耿磊^{1,2}, 仇怀志^{2,3}, 肖志涛^{1,2}, 张芳^{1,2}, 吴骏^{2,3}

(1. 天津工业大学 生命科学学院, 天津 300387; 2. 天津工业大学 天津市光电检测技术与系统重点实验室, 天津 300387; 3. 天津工业大学 电子与信息工程学院, 天津 300387)

摘要: 针对矿物类别众多、存在环境杂质等干扰信息以及部分拉曼光谱存在相似性等问题, 结合拉曼光谱时域和频域上多尺度特征信息, 提出一种基于卷积结构和自注意力结构的双分支分类网络 RT-Net (Residual-Transformer Net)。该网络利用卷积块搭建局部特征提取模块, 引入通道注意力增强局部特征提取能力; 利用自注意力结构学习拉曼光谱频域中的双向依赖关系来提取全局特征信息, 由注意力融合模块进行多尺度特征融合用以分类。实验结果表明: RT-Net 实现了对于 1 321 类矿物拉曼光谱快速准确的分类, 分类准确率达到 90.31%; 此外, 在精准率、召回率和 F1 得分 3 个评估指标上分别达到了 0.878 1、0.906 6 和 0.897 2, 进一步验证了 RT-Net 的有效性。

关键词: 矿物分类; 拉曼光谱; 频域; 注意力机制; 多尺度融合

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1671-024X(2025)01-0053-09

Classification method of Raman spectra of mineral based on convolution and Transformer

GENG Lei^{1,2}, QIU Huaizhi^{2,3}, XIAO Zhitao^{1,2}, ZHANG Fang^{1,2}, WU Jun^{2,3}

(1. School of Life Sciences, Tiangong University, Tianjin 300387, China; 2. Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems, Tiangong University, Tianjin 300387, China; 3. School of Electronics and Information Engineering, Tiangong University, Tianjin 300387, China)

Abstract: To address the problems of many mineral categories, the existence of environmental impurities and other interference information, and the similarity of some Raman spectra, a two-branch classification network RT-Net (Residual-Transformer Net) based on convolutional structure and self-attentive structure is proposed by combining the multi-scale feature information in the time and frequency domains of Raman spectra. The network uses convolutional blocks to build a local feature extraction module and introduces channel attention to enhance the local feature extraction ability; the self-attentive structure is used to learn the bidirectional dependencies in the frequency domain of Raman spectra to extract global feature information, and the attentional fusion module is used for multi-scale feature fusion for classification. The experimental results show that RT-Net achieves fast and accurate classification of 1 321 types of mineral Raman spectra with an accuracy of 90.31%. In addition, the three evaluation indexes of precision, recall and F1 score reached 0.878 1, 0.906 6 and 0.897 2, respectively, further validating the effectiveness of RT-Net.

Key words: mineral classification; Raman spectra; frequency domain; attentional mechanisms; multi-scale integration

矿物识别在地质分析、资源勘探以及许多工程应用中都是极为关键的一环。拉曼光谱能够提供丰富的物质分子信息, 具有反映物质分子“指纹”特征的特性, 能够获取被测物质的分子振动、转动能级信息并

进行分析, 已被广泛应用于各场景下的矿物质分析和识别任务中^[1]。然而, 自然界中矿物的种类繁多, 其拉曼光谱具有一定的相似性, 且因其成长环境不同, 成分结构复杂不统一, 同类矿物的拉曼光谱也存在一定

收稿日期: 2023-03-01

基金项目: 天津市科技计划项目(20YDTPJC00110)

通信作者: 耿磊(1982—), 男, 博士, 教授, 主要研究方向为机器视觉、深度学习等。E-mail: genglei@tiangong.edu.cn

差异性。这些问题都对矿物的分类造成许多困难,因此建立一个快速准确的矿物拉曼光谱分类模型有着重要的价值。

近年来,深度学习技术不断发展,被广泛应用于图像分类^[2]、语音识别^[3]、医学信号处理^[4]等领域。深度神经网络能够将原始数据的特征从低维空间映射到高维空间,具有优异的特征学习能力。对于多分类任务,深度神经网络能够很好地学习到样本间差异区分度较小的特征。已有学者将该技术应用于矿物拉曼光谱的识别任务上,并取得了一定效果。Liu 等^[5]采用包括特征提取的金字塔形卷积层和用于分类的 2 个全连接层的 LetNet 变体的深度卷积神经网络对拉曼光谱数据分类的方法,在 RRUFF(矿物拉曼光谱数据库)上取得很好的分类效果。郭志琦^[6]针对 192 类的矿物拉曼光谱多分类问题,提出了一种轻量级神经网络模型 FMCNN(factorization machine combined neural network),达到了 98.8% 的分类精度。Sang 等^[7]基于 VGG19 的网络结构构建了深层卷积神经网络(CNN)模型 1DCNN,在 RRUFF 数据库上构建的类别数为 192 和 1 332 的 2 个数据集上,准确率分别达到了 97.72% 和 87.91%。由于 CNN 架构的步长特征限制了感受野的范围,难以覆盖全部拉曼光谱,导致部分特征信息丢失。

矿物拉曼光谱的光谱信息在波段之间是高度相关的,关注大范围的波段关系有助于拉曼光谱特征的提取。Transformer 架构利用多头注意力机制提取到时序向量间的隐含关系,很好地解决了长距离依赖关系问题。Behinaein 等^[8]通过级联卷积结构与自注意力结构的方式对心电信号进行分类,其特点是通过卷积操作提取心电信号中的特征信息,从而提供给自注意力结构进行后续分类。孙嘉瑶^[9]提出一种双分支网络结构,各分支结构都使用卷积与自注意力层级联的方式,分别从时间维度和通道维度挖掘信号中的特征。除了级联卷积与自注意力的方式,葛君伟等^[10]提出将卷积结构和自注意力结构进行并行连接,从 2 种不同的角度捕获特征信息。

针对矿物拉曼光谱存在的类间相似性和存在干扰性信息等问题,本文提出了一种同时提取时频域特征的多尺度信息融合双分支网络 RT-Net。主要研究内容有以下几个方面:

(1) 为提取矿物拉曼光谱中强度较小的次特征峰,解决杂质、环境等干扰性信息问题,设计了一种利用扩张卷积块搭建的局部特征提取模块,对小特征峰进行特征提取,同时通过将通道注意力模块嵌入卷积

网络的不同阶段,抑制低级语义特征。

(2) 针对部分矿物的拉曼光谱较为相似的问题,结合时频域特征信息,设计了一种由卷积和 Transformer 并行连接的双分支网络,学习光谱多尺度的特征信息。利用卷积网络提取光谱时域上的局部特征信息,捕获益于分类的拉曼特征峰信息;通过 Transformer 学习光谱频域上的全局依赖关系,融合局部特征信息用以分类。

(3) 利用注意力融合模块,将提取的局部特征信息和全局信息在保留各自最大信息完整性的同时进行充分融合。

(4) 通过实验验证了本文方法的有效性,并与传统光谱分类算法进行比较。

1 基于 RT-Net 网络的矿物拉曼光谱分类法

本文提出的 RT-Net 网络整体结构如图 1 所示。

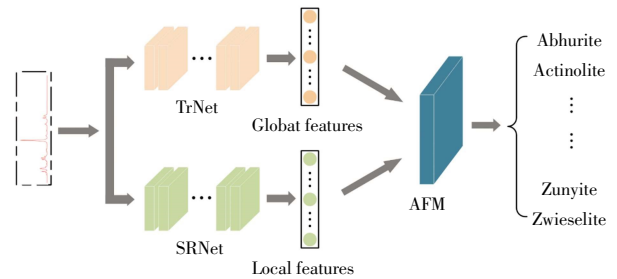


图 1 RT-Net 网络整体结构示意图

Fig.1 Overall structure diagram of RT-Net network

RT-Net 是由 SRNet 分支和 TrNet 分支并行的双分支网络,能够精确分类具有 1 321 类的矿物拉曼图谱。矿物的拉曼光谱有许多波峰和波谷,包含许多特征信息,因此在 SRNet 分支,利用其善于捕捉卷积感受野范围内的特征的优势,提取拉曼光谱的局部特征信息用以分类。由于某些矿物的拉曼光谱有相似的特征峰,依靠局部特征难以区分,因此 TrNet 分支借由 Transformer 的自注意力机制对全局依赖关系的学习能力,学习拉曼光谱的全局特征信息。网络从局部和全局对矿物的拉曼光谱进行特征提取,通过注意力融合模块(AFM),将各分支的特征信息进行融合分类。下面将对各分支网络和注意力融合模块进行更为详尽的叙述。

1.1 SRNet 分支网络

SRNet 分支是由 1 个 1×7 的一维卷积层和 1 个 1×3 的最大池化层,对拉曼光谱时域数据进行下采样,降低维数并保证数据基本特征的完整性。共 8 个特征提取模块分为 4 个阶段对拉曼数据进行时域特征的

细分, 结构如图 2 所示。

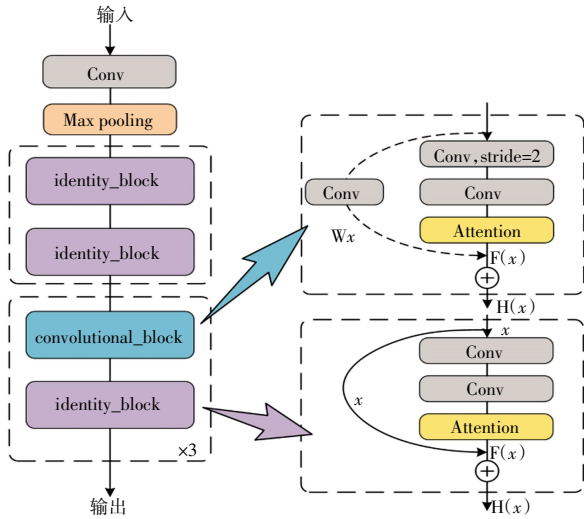


图 2 SRNet 网络结构示意图

Fig.2 Structure schematic of SRNet network

每个特征块有 2 个相同输出通道数的 3×3 卷积层, 每个卷积层后接 1 个批量规范化层和 ReLU 激活函数。然后通过跳过 2 个卷积层, 将输入直接加在 ReLU 激活函数前。假设输入为 x , 通过 2 个卷积层后的输出为 $f(x)$, convolutional_block 模块由于第 1 个卷积层步长为 2, 跳跃连接的输入 x 需要经过 1 个卷积层调整通道维度与 $f(x)$ 相匹配, 输出 $H_c(x)$ 的计算公式为:

$$H_c(x) = f(x) + Wx \quad (1)$$

式中: W 为卷积矩阵。identity_block 模块是 2 个步长为 1 的卷积层, 因此跳跃连接的输入 x 通道维度与 $f(x)$ 相一致, 输出 $H_i(x)$ 的计算公式为:

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

每个特征提取模块中都加入了 Squeeze-and-Excitation (SE) 模块^[11], 通过提取特征图通道之间的相

关性信息, 自适应的动态学习并加以调整, 对每个通道施加不同的权重, 选择性的突出利于识别的特征并抑制无用特征。

1.2 TrNet 分支网络

Transformer 网络主要应用于自然语言处理领域, 由于矿物拉曼光谱数据符合时间序列数据特点, 故使用 Transformer 模块作为矿物拉曼光谱全局特征的主要特征提取器, 通过自注意力机制学习其长距离双向依赖关系。同时, Transformer 网络含有 Encoder 编码器和 Decoder 解码器 2 部分, 分别负责对输入模型信号进行编码和解码, 而矿物拉曼光谱分类任务并不涉及 Decoder 解码操作, 故将 Transformer 中的 Encoder 结构作为分支网络 TrNet 的基础, 用于提取矿物拉曼光谱全局特征。由于自注意力模型在对当前位置的数据信息进行编码时, 易过度的集中注意力在自身的位置而忽视了其他的位置, 因此在 Transformer 中引入了多头注意力机制 (Multi-Head Attention) 来解决该问题, 其完整计算过程如公式 (3) 所示:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) \mathbf{W}^O$$

$$\text{where } \text{head}_i = \text{Attention}(\mathbf{C}\mathbf{W}_i^Q, \mathbf{C}\mathbf{W}_i^K, \mathbf{C}\mathbf{W}_i^V) \quad (3)$$

式中: $\mathbf{Q}, \mathbf{K}, \mathbf{V} \in \mathbb{R}^{M \times d}$ 分别表示 query, key, value 矩阵; $\mathbf{W}_i^Q \in \mathbb{R}^{N \times d_q}$, $\mathbf{W}_i^K \in \mathbb{R}^{N \times d_k}$, $\mathbf{W}_i^V \in \mathbb{R}^{N \times d_v}$ 和 $\mathbf{W}_i^O \in \mathbb{R}^{hd \times N}$ 分别表示对应参数矩阵; h 表示多头注意力机制中 head 的个数; d_q, d_k, d_v 的值设置为 128。

由于矿物种类数量较多, 部分矿物类别的拉曼光谱存在相似性, 特征峰在时域上面难以区分, 频域则可以更好地从全局上表征矿石的特征。因此, 通过将输入进行快速傅里叶变换 (FFT) 转换到频域, 进而利用 Transformer 来学习全局依赖关系, 其结构如图 3 所示。

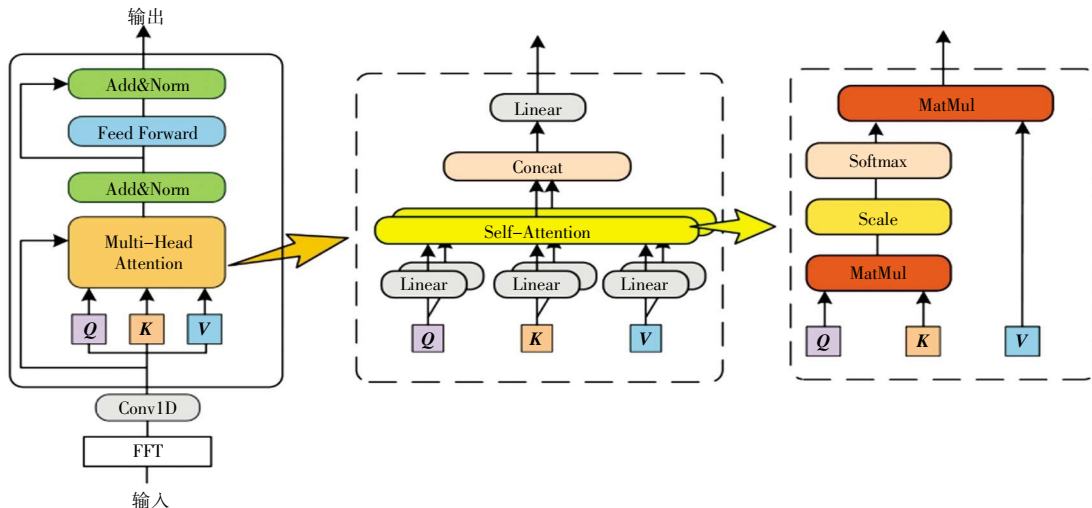


图 3 TrNet 网络结构示意图

Fig.3 Structure diagram of TrNet network

首先对 TrNet 的输入进行时频域转换,通过 1 个卷积核大小为 1×1 且通道数为 128 的一维卷积层,改变输入的维度大小以及提供矿物拉曼光谱的空间信息。然后将编码后的矿物拉曼光谱数据送入 Transformer 的 Encoder 结构中,在多头注意力层后接入 1 个前馈神经网络为模型增加非线性表达的能力,其计算公式如公式(4)所示:

$$\text{FFN}(x) = \text{ReLU}(0, xW_1 + b_1)W_2 + b_2 \quad (4)$$

式中: x 表示经过多头注意力层后的输出特征。

同时,在多头注意力层和前馈网络层周围都使用了残差连接(Residual Connection)的方式并进行层归一化(Layer Normalization, LN)操作。残差连接能够帮助网络底层特征信息有效的传递到网络的高层,增加了信息的传递效率并且增强网络的表达能力。层归一化操作是为了在激活函数之前,将输入转换为均值为 0 且方差为 1 的数据,避免出现梯度消失或者梯度爆炸。

1.3 注意力融合模块

为了将 SRNet 和 TrNet 在拉曼光谱数据中提取的局部特征与全局特征在保持各自信息完整度最大化的同时能够充分融合,利用多头注意力机制搭建了 1 个注意力融合模块(AFM),其结构如图 4 所示。

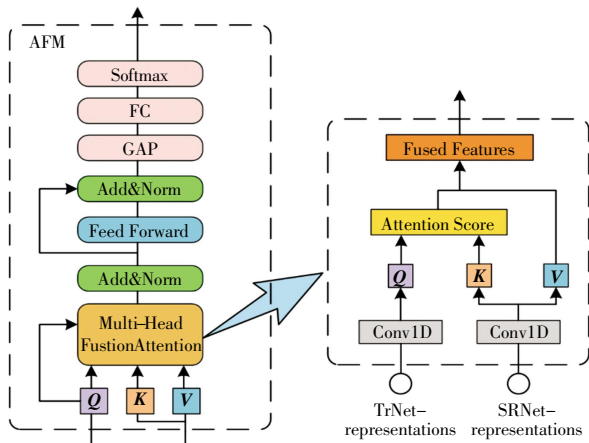


图 4 注意力融合模块结构示意图

Fig.4 Schematic diagram of attention fusion module

由于 2 个分支输出特征的维度不一致,因此在特征融合前要先进行维度统一。SRNet 分支网络输出的特征尺寸为 $T_c \times C$, 其中 T_c 表示矿物拉曼光谱通过 SRNet 的卷积操作之后的输出长度, C 表示最后 1 个特征提取模块中卷积操作的通道数。TrNet 分支网络输出的特征尺寸为 $T \times C$,其中 T 表示原始输入序列的长度, N 代表输入的矿物拉曼光谱的空间维度。超参数 C 和 N 的值统一设置为 128,保证了 2 个分支网络输出的特征通道维度相同,然后分别通过 2 个卷积核大小为 1×1 的卷积层,将 SRNet 分支卷积操作后的输出

长度 T_c 与 TrNet 分支原始输入序列的长度 T 进行统一。

将 SRNet 与 TrNet 输出的特征分别对应矩阵 Q 和矩阵 K 来计算注意力得分,计算得到的注意力得分包含了局部特征与全局特征融合后的对应关系,注意力得分越高表明该位置的特征与总体特征的匹配程度越高,使用融合了全局特征与局部特征的对应关系的注意力得分标定 SRNet 的局部特征,计算过程如公式(5)所示:

$$\text{Fusion}(X_{\text{Trans}}, X_{\text{CNN}}) = \text{softmax}\left(\frac{Q_{\text{Trans}} K_{\text{CNN}}^T}{\sqrt{d_k}}\right) V_{\text{CNN}} \quad (5)$$

式中: X_{Trans} 和 X_{CNN} 分别代表 SRNet 的输出和 TrNet 的输出。

2 实验结果分析

2.1 实验数据

本次实验使用的数据集是 RRUFF 数据库中的矿物拉曼光谱数据集,该数据集包含 5 267 组光谱数据和 1 687 种矿石物质。由于训练模型时至少要划分 1 个样本用作测试集,因此筛除掉只有 1 个样本的矿石类别后剩余 1 321 类。

2.1.1 样条插值

RRUFF 数据库中收集的光谱数据因采集条件不同,导致拉曼位移及频率不同,光谱数据点的数量各不相同。因此,需要对光谱数据进行插值处理。

采用一阶样条插值法对各个光谱数据进行处理。通过估算插值点相邻的 2 个数据点,仅对 1 个小区间进行插值,插值结果如图 5 所示。经过插值处理后的光谱数据,可以得到均匀分布在 $0 \sim 1700 \text{ cm}^{-1}$ 拉曼位移上的 1 024 个数据点及对应强度,对某些类矿石拉曼位移未覆盖的区间用 0 进行填充。

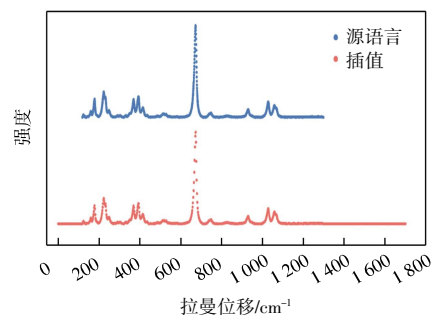


图 5 一阶样条插值结果

Fig.5 The results of first-order spline interpolation

2.1.2 数据增强

大部分矿石类别在选用的 RRUFF 数据集中样本数量比较少,而数据量少对模型的训练会带来过拟

合等一系列不利影响,因此采用数据增强的方式减轻对模型训练的不利影响。

在每一类矿石数据中随机抽选1个拉曼光谱样本放入测试集,剩余数据按照3:1的比例划分训练集和验证集。划分后训练集所含数据量为3580,测试集所含数据量为1321。然后对训练集中的数据采用以下操作进行数据增强:

(1) 平移操作:固定拉曼强度序列不变,将拉曼位移在 $0\sim 0.4\text{ cm}^{-1}$ 之间左右平移,对每一类矿石进行不同次数的操作,使每类矿石数据量在10个样本左右。

(2) 曲线加噪声:对平移后的拉曼光谱曲线在位移和强度序列中加入高斯噪声。每一条拉曼曲线中的每一个点都在 X 和 Y 轴方向上加入1个随机噪声,该噪声符合均值为0、标准差为0.5的正态分布,处理后每类矿石的数据量在20个样本左右。

2.1.3 归一化

同一矿物在不同激发光功率下会有不同的峰强,为了避免在模型训练时强度大小对拉曼曲线的影响,使模型只关注光谱图上特征峰之间的相对强度,对所有样本数据的强度序列分别进行了Min-Max归一化处理,计算方式如公式(6)所示:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (6)$$

式中: x 表示输入的强度值。

2.2 评价指标

本次实验使用4种评价指标判断RT-Net网络模型能否对矿物拉曼光谱进行1321类的精确分类。准确率是对模型效果最直观的评价指标,即测试集中正确样本占总样本的比例。此外,对于矿物拉曼数据的多分类问题,增加精确率、召回率和F1分数,从而尽可能的客观和全面的评估模型的性能。本文使用P和N表示正样本和负样本,用T和F表示正确预测和错误预测,4项评价指标的公式如下所示:

$$\text{准确率} = \frac{TP + TN}{TP + FP + TP + FN} \quad (7)$$

$$\text{精确率} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{召回率} = \frac{TP}{TP + FN} \quad (9)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (10)$$

式中:TP、FP、TN和FN分别表示真正、假正、真负和假负。

2.3 实验环境及超参数设置

硬件配置为:GPU,2个NVIDIA 2080Ti显卡。深度学习框架为:TensorFlow 2.0+python3.7。网络在训练

时采用Adam优化器训练模型,权重衰减系数为 $1e^{-4}$,初始学习率为0.0001, batch size为64。由于模型训练时其收敛速度会随着迭代轮次的增加而减缓,出现学习停滞的现象,通过降低学习率能有效改善此现象,因此设置训练的迭代轮次为200,当模型训练时的迭代轮次达到20且验证集准确率并未提升时,学习率将被调整为上轮迭代学习率的1/2。

2.4 实验结果与分析

2.4.1 各分支网络有效性验证

为了分析网络模型对矿物拉曼光谱的学习能力,通过将网络深层的特征集合与池化层获得的权重系数矩阵相乘,最终映射得到一个能够表征特征关注度的热力图。

SRNet在残差结构中通过堆叠扩张卷积块,能够有效地对矿物拉曼光谱中拉曼峰的特征信息进行提取并用以分类。本文随机选取了3类矿物的拉曼光谱曲线,通过SRNet对其特征提取并进行了特征热力图可视化,结果如图6所示。

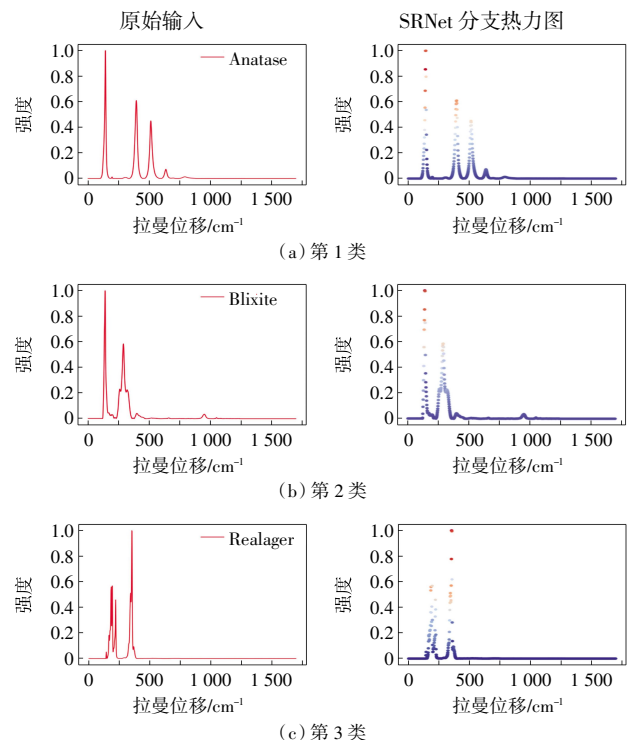


图6 SRNet特征可视化结果示意

Fig.6 Schematic diagram of SRNet feature visualization results

由图6可以看到,SRNet能够对每类矿物拉曼光谱中特征峰的特征信息进行提取,并且对最明显的主拉曼峰关注程度最高,对次拉曼峰的关注程度相对较低。由此可以证明,SRNet能够很好捕获到每一类矿物的局部特征信息,从而有效地对其进行分类。

TrNet 沿用 Transformer-Encoder 结构通过多头注意力结构和前馈神经网络级联的方式,利用自注意力机制对输入进行全局特征信息的特征提取。由于频域的变换结果是通过 FFT 得到的,其频域信息是时域波形对应频率下的幅度特征,通过对全局特征信息进行学习,并映射回原始拉曼光谱用于分类。选取了 3 类矿物进行了特征可视化分析,结果如图 7 所示。

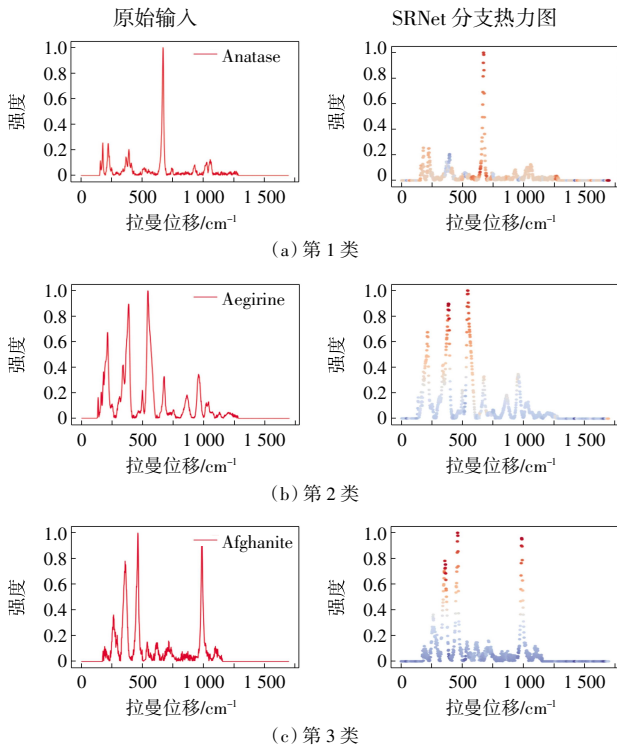


图 7 TrNet 特征可视化结果示意

Fig.7 Schematic diagram of TrNet feature visualization results

由图 7 可看出,TrNet 通过提取全局特征信息能够关注到拉曼频移在 500 ~1 200 cm^{-1} 之间的较小的拉曼峰,在主特征峰拉曼频移和形状极为相似的情况下,可以凭借其余较小的特征峰进行正确区分。然而某些矿物拉曼光谱在时域上较为相似,在频域则是具有差异性信息,考虑到时序数据可以从时域转换到频域获取频域信息,网络能够通过学习光谱频域上的特征信息映射回原始拉曼光谱,在提取主特征峰的前提下学习多尺度的特征信息用于分类。

为验证频域特征信息对于网络模型的性能影响,设置了消融实验来进行分析验证,实验结果如表 1 和表 2 所示。由表 1 可以看出,在时域上,SRNet 对矿物拉曼光谱上的局部特征提取能力较强从而有较高分类精度,而在频域上,TrNet 通过学习拉曼光谱全局特征信息进行分类效果较好。再与表 2 对比可知,RT-Net 的 2 个分支选择为 SRNet 提取时域特征信息和

TrNet 提取频域特征信息时,矿物拉曼光谱分类任务模型性能最优。

表 1 各分支网络在时频域性能比较

Tab.1 Comparison of performance of each branch network in time and frequency domain

时/频域	网络模型	准确率	精准率	召回率	F1 分数
时域	SRNet	0.876 5	0.839 4	0.878 4	0.847 6
	TrNet	0.848 2	0.763 9	0.820 2	0.794 9
	RT-Net	0.889 2	0.846 3	0.889 8	0.862 9
频域	SRNet	0.843 1	0.821 7	0.858 5	0.833 7
	TrNet	0.851 5	0.805 1	0.854 7	0.821 3
	RT-Net	0.867 2	0.817 4	0.869 1	0.823 7

表 2 RT-Net 网络模型在时频域性能比较

Tab.2 Comparison of performance of RT-Net network model in time and frequency domain

时域	频域	准确率	精准率	召回率	F1 分数
TrNet	SRNet	0.866 7	0.816 6	0.867 2	0.821 9
SRNet	TrNet	0.903 1	0.878 1	0.906 6	0.897 2

2.4.2 SRNet 中通道注意力机制有效性验证

为了验证通道注意力模块对残差网络的性能影响,通过消融实验和特征可视化进行分析。消融实验依照控制变量法的原则分别对未引入通道注意力模块和引入通道注意力后的残差网络进行训练和测试,实验结果如表 3 所示。由表 3 可见,引入通道注意力模块对于残差网络模型的整体性能有提升,并且 4 个评估指标都优于未引入通道注意力模块得网络模型。

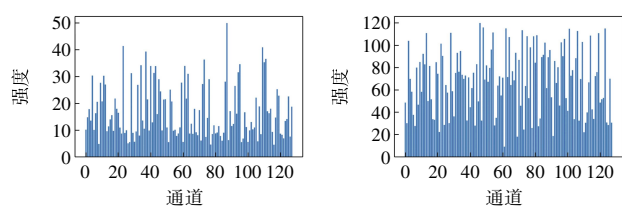
表 3 通道注意力结构加入前后网络性能比较

Tab.3 Comparison of network performance before and after adding channel attention structure

是否加入注意力结构	准确率	精确率	召回率	F1 分数
否	0.868 7	0.832 2	0.868 8	0.841 2
是	0.876 5	0.839 4	0.878 4	0.847 6

除了从评价指标能够直观的表现通道注意力模块对网络性能的提升,还可以通过引入通道注意力模块前后的特征图进行可视化,从而具体地分析通道注意力模块对残差网络带来的性能提升。将上述 2 种情况的 SRNet 网络中最后 1 个卷积层的特征进行可视化,将其输入展开为一维向量,该向量是网络模型从矿物拉曼光谱中提取的一组抽象特征,其可视化如图 8 所示。

图 8(a)将未引入通道注意力模块的 SRNet 的各通道输出特征进行可视化,其中各个通道的特征强度关系都相对独立。图 8(b)将通过引入通道注意力模块来增强 SRNet 的通道输出特征进行可视化,各通道的



(a) 未引入通道注意力模块特征图 (b) 引入通道注意力模块特征图

图 8 特征图可视化结果

Fig.8 Visualization result of characteristic map

特征强度变化是因为通道注意力结构通过学习通道间的依赖关系来判断通道的重要程度, 根据特征重要程度使网络模型更关注利于分类的有效特征。该实验验证了引入通道注意力模块能够很好的帮助 SRNet 对矿物拉曼光谱进行精准分类。

2.4.3 RT-Net 整体网络学习能力可视化分析

为了验证 RT-Net 对矿物拉曼光谱的特征学习能

力, 通过将深层神经网络的特征集合与全局平均池化层求得的权重系数矩阵做乘法运算, 在映射到拉曼光谱上得到 1 张能够表征特征关注度的热力图, 模拟网络模型在分类过程中对特征信息的关注度, 此方法能够直观的将网络模型的特征学习能力进行可视化。

选取任意 4 类矿物中各一条拉曼光谱曲线, 将 RT-Net 模型的特征学习能力进行了可视化分析。特征热力图是按照拉曼光谱曲线趋势的散点构成, 其中的散点的颜色反应了模型对该位置关注度的高低。对于网络模型判定类别的作用越大的取样点, 其对应的颜色越红, 从而反应 RT-Net 对矿物拉曼光谱分类时依靠的那些特征区域。而随着对应颜色由深红色逐渐到浅白色最后到深蓝色, 则表明模型认为该特征对于类别判断的作用是逐渐减小的甚至是无用的。RT-Net 对拉曼光谱特征学习能力可视化结果如图 9 所示。

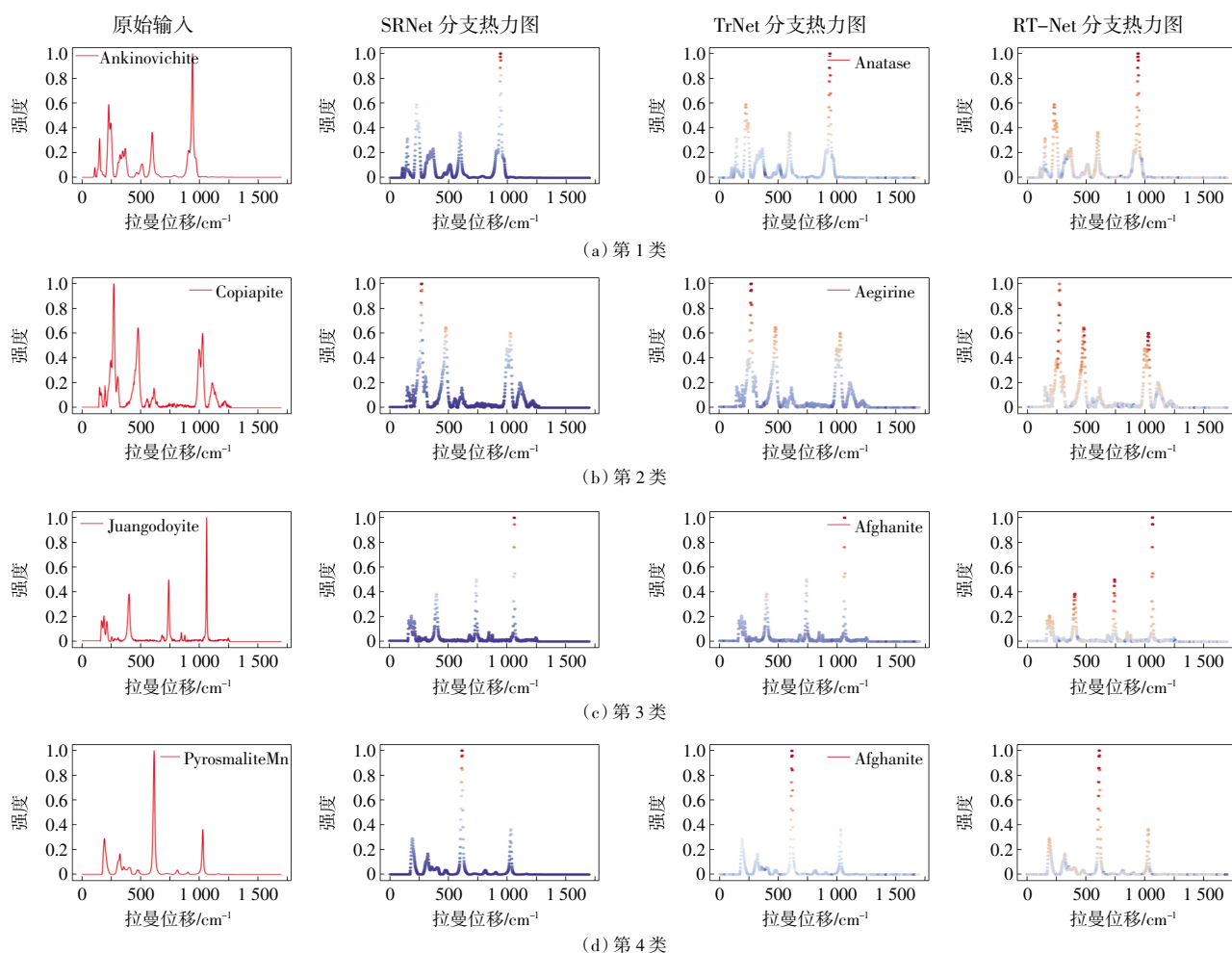


图 9 RT-Net 模型特征学习能力可视化结果示意

Fig.9 Schematic diagram of visualization results of feature learning ability of RT-Net model

图 9 中: 第 1 列分别是 4 种矿物 Ankinovichite、Copiapite、Juangodoyite 以及 PyrosmaliteMn 的拉曼光谱曲线; 第 2 列是 SRNet 在矿物拉曼光谱时域上对特征信

息关注程度的可视化结果, 观察发现, SRNet 通过扩张卷积能够学习到拉曼光谱中的局部特征峰; 第 3 列是 TrNet 在拉曼光谱频域上学习到的特征信息映射到原

始图谱上的可视化结果图,观察发现,TrNet 利用自注意力结构有效地学习到矿物拉曼光谱的全局信息;第 4 列是 RT-Net 对输入拉曼光谱的特征关注程度的可视化结果图,相比较于支路模型,RT-Net 通过注意力模块将双分支提取到的局部特征信息和全局信息有效融合后,能够更加全面的关注最有利于类别判断的特征信息。该实验也验证了设计 RT-Net 的初衷,通过卷积网络和自注意力机制并行的双分支网络结构对 1 321 类矿物拉曼光谱实现精准分类。

2.4.4 RT-Net 与其他分类算法性能对比

为验证 RT-Net 对矿物拉曼光谱分类的可行性,选用几类经典的机器学习算法进行对比试验分析。利用 PCA 算法将矿物拉曼光谱数据进行降维处理,将其高维空间特征映射到低维特征空间,然后将样本数据的主要特征分量进行提取,输入到分类器中用以分类。最终在各模型在测试集上的性能比较如表 4 所示。

表 4 不同传统分类算法在数据集中性能对比

Tab.4 Comparison of performance of different traditional classification algorithms in data set

实验方案	分类算法	准确率
Yan X. et al ^[12]	RBF SVM	0.743 4
Kramer O. et al ^[13]	Nearest Neighbors	0.774 1
Myles A. et al ^[14]	Decision Tree	0.736 4
Rish I. et al ^[15]	Naive Bayes	0.621 7
Tharwat A. et al ^[16]	QDA	0.815 5
Guo G. et al ^[17]	KNN	0.768 8
本文所提方法	RT-Net	0.903 1

由表 4 可知,RT-Net 准确率远高于其他经典机器学习分类方法。分类算法按照其各自的分类准对矿物拉曼光谱进行类别划分,但由于本文数据集中矿物拉曼光谱类别数量较多,且其特征信息极为相似,导致拉曼光谱数据在特征空间分布较为密集,传统的机器学习分类算法难以找到准确的决策面,因而难以精准分类。

本文将 RT-Net 与其他深度学习算法以及同类任务的分类网络进行了对比试验,在同一数据集上的实验结果如表 5 所示。

表 5 不同深度学习算法在数据集中性能对比

Tab.5 Comparison of performance of different depth learning algorithms in data set

实验方案	分类方法	准确率	精准率	召回率	F1 得分
Ney H. et al ^[18]	LSTM	0.701 6	0.642 5	0.611 9	0.624 1
Cho K. et al ^[19]	GRU	0.683 6	0.599 1	0.524 4	0.573 2
Lea C. et al ^[20]	TCN	0.788 4	0.736 5	0.799 3	0.789 6
Zha B. et al ^[21]	Transformer-Encoder	0.813 5	0.778 1	0.819 7	0.798 9
Liu J. et al ^[5]	CNN	0.866 4	0.841 5	0.862 4	0.860 3
Sang X. et al ^[7]	1DCNN	0.877 4	0.843 9	0.879 1	0.862 2
郭志琦 ^[6]	FMCNN	0.854 1	0.799 3	0.855 7	0.811 6
本文所提方法	RT-Net	0.903 1	0.878 1	0.906 6	0.897 2

从表 5 中可以看出,LSTM 和 GRU 两类算法的性能在数据集上表现并不理想,是因为对矿物拉曼光谱而言,局部特征峰是进行类别判断的首要特征信息,而 RNN 结构的深度学习算法善于学习中长距离序列的依赖关系,导致没能有效学习到有价值的特征信息。相较于单一的 RNN 结构,TCN 算法加入了卷积结构对矿物拉曼光谱的局部特征进行学习,能够对有利于类别判断的特征信息进行提取,实验也表明其性能要优于 RNN。原始的 Transformer-Encoder 结构在数据集上进行了训练测试,能够学习矿物拉曼光谱的全局特征信息用以分类,表现出较好的性能。对于矿物拉曼光谱分类这一特定任务领域的算法 CNN、1DCNN 和 FMCNN,在该数据集上表现的效果并不理想。由于上述 3 种算法都是基于 CNN 结构提取矿物拉曼光谱的局部特征信息进行分类的,而矿物拉曼光谱数据集中样本类别为 1 321 类,类别数量更多且存在许多同谱异物的样本数据,仅凭借 CNN 结构提取样本中的局部特征信息难以判别这些“困难”样本数据。但是 RT-Net 具备局部特征提取和全局特征提取的能力,并且引入了频域特征信息,在数据集上展现出强大的分类能力,各项评价指标均为最高。

3 结束语

本文提出了一种基于卷积和 Transformer 的双分支网络 RT-Net 拉曼光谱分类模型,实现了 1 321 类矿物拉曼光谱的精确分类。通过卷积网络分支在局部特征提取模块引入通道注意力机制来抑制低层语义信息,聚焦深层语义信息;利用 Transformer 使光谱特征建立全局依赖关系,增强全局前后文建模能力。与传统分类算法和深层卷积网络模型相比,RT-Net 能够全面地提取多尺度特征信息,利用注意力融合模块将时域特征和频域特征进行多尺度融合从而实现精准分类。实验结果表明,RT-Net 是一种性能优异的矿物拉曼光谱分类模型,相比于传统光谱分类算法无需进行预处理和特征提取等步骤,具有简单快速、分类精度高的优点,为更多类别的矿物拉曼光谱分类任务提供了新思路。

参考文献:

- [1] VENERANDA M, MANRIQUE J A, LOPEZ-REYES G, et al. Spectroscopic study of olivine-bearing rocks and its relevance to the ExoMars rover mission[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 223: 117360.

- [2] 余东行, 张保明, 赵传, 等. 联合卷积神经网络与集成学习的遥感影像场景分类[J]. 遥感学报, 2020, 24(6):717-727.
YU D X, ZHANG B M, ZHAO C, et al. Remote sensing image scene classification based on combined convolution neural network and ensemble learning[J]. Journal of Remote Sensing, 2020, 24(6):717-727(in Chinese).
- [3] JIA G M, CHENG F Y, YANG J F, et al. Intelligent checking model of Chinese radiotelephony read-backs in civil aviation air traffic control[J]. Chinese Journal of Aeronautics, 2018, 31(12):2280-2289.
- [4] YU M X, SUN Y C, ZHU B F, et al. Diverse frequency band-based convolutional neural networks for tonic cold pain assessment using EEG[J]. Neurocomputing, 2020, 378:270-282.
- [5] LIU J C, OSADCHY M, ASHTON L, et al. Deep convolutional neural networks for Raman spectrum recognition: A unified solution[J]. Analyst, 2017, 142(21):4067-4074.
- [6] 郭志琦. 基于深度学习的拉曼光谱分析与应用研究[D]. 乌鲁木齐:新疆大学, 2021.
GUO Z Q. Analysis and application of Raman spectroscopy based on deep learning[D]. Urumqi: Xinjiang University, 2021 (in Chinese).
- [7] SANG X C, ZHOU R G, LI Y C, et al. One-dimensional deep convolutional neural network for mineral classification from Raman spectroscopy[J]. Neural Processing Letters, 2022, 54(1):677-690.
- [8] BEHINAIEIN B, BHATTI A, RODENBURG D, et al. A transformer architecture for stress detection from ECG[C]//2021 International Symposium on Wearable Computers. Virtual, USA: ACM, 2021:132-134.
- [9] 孙嘉瑶. 深度学习注意力机制在脑电信号分类中的应用[D]. 深圳:中国科学院大学(中国科学院深圳先进技术研究院), 2021.
SUN J Y. Application of deep learning attention mechanism in EEG signal classification[D]. Shenzhen: University of Chinese Academy of Sciences (Chinese Academy of Sciences Shenzhen Advanced Technology Research Institute), 2021(in Chinese).
- [10] 葛君伟, 涂兆昊, 方义秋. 基于融合 CNN 和 Transformer 的分离结构机器翻译模型[J]. 计算机应用研究, 2022, 39(2):432-435.
GE J W, TU Z H, FANG Y Q. A machine translation model with separate structure based on fusion CNN and Transformer[J]. Application Research of Computers, 2022, 39(2):432-435 (in Chinese).
- [11] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 7132-7141.
- [12] YAN X, GE H, YAN Q. SVM with RBF kernel and its application research[J]. Computer Engineering and Design, 2006, 27(11): 1996-1997.
- [13] KRAMER O. K-Nearest Neighbors [C]//Dimensionality Reduction with Unsupervised Nearest Neighbors. Berlin, Heidelberg: Springer, 2013: 13-23.
- [14] MYLES A J, FEUDALE R N, LIU Y, et al. An introduction to decision tree modeling[J]. Journal of Chemometrics: A Journal of the Chemometrics Society, 2004, 18(6): 275-285.
- [15] RISH I. An empirical study of the naive Bayes classifier[C]//IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence. Seattle, WA, USA: IJCAI, 2001: 41-46.
- [16] THARWAT A. Linear vs. quadratic discriminant analysis classifier: A tutorial[J]. International Journal of Applied Pattern Recognition, 2016, 3(2): 145-180.
- [17] GUO G D, WANG H, BELL D, et al. KNN Model-based approach in classification [C]//MEERSMAN R, TARI Z, SCHMIDT D C. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003:986-996.
- [18] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [19] CHOK, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. ArXiv e-Prints, 2014: arXiv: 1406.1078.
- [20] ZHAO B D, LU H Z, CHEN S F, et al. Convolutional neural networks for time series classification[J]. Journal of Systems Engineering and Electronics, 2017, 28(1):162-169.
- [21] ZHA B, VANNI A, HASSAN Y, et al. Deep transformer networks for time series classification: The NPP safety case[J]. ArXiv preprint arXiv: 2021. 2104. 05448.

本文引文格式:

耿磊, 仇怀志, 肖志涛, 等. 基于卷积和Transformer的矿物拉曼光谱分类方法[J]. 天津工业大学学报, 2024, 44(1): 53-61.
GENG L, QIU H Z, XIAO Z T, et al. Classification method of Raman spectra of mineral based on convolution and Transformer[J]. Journal of Tiangong University, 2024, 44(1): 53-61 (in Chinese).