doi:10.12068/j.issn.1005－3026.2026.20249063

变因素下基于多模态融合算法的煤矸识别

夏 蕊[1]，王 响[1,2]，李 博[1,2]，王学文[1,2]
（1.太原理工大学 机械工程学院，山西 太原 030024；2.智能综采装备技术山西省重点实验室，山西 太原 030024）

摘 要：为应对环境因素对煤矸识别任务的影响，本文提出了融合光谱与图像特征的新型网络架构.该架构通过整合双模态信息，挖掘丰富的特征信息，从而解决单一模态在恶劣环境下出现的特征退化问题.为验证模型性能，在实验室条件下构建了多因素数据集和工况数据集，并分别搭建光谱特征提取（1DCNN+CBAM）与图像特征提取（ResViT）网络，其中ResViT网络首次将残差机制与视觉Transformer结合用于煤矸分选任务，能够高效提取模态间的特异性特征.通过设计对比与消融试验确定了各分支搭配的合理性，并在两种数据集下验证了模型的稳定性，准确率最高达到97.92%.由此证明，提出的多模态融合算法是一种高效的分类模型，可为井下煤矸分选提供有力理论支持.

关 键 词：煤矸识别；近红外光谱；图像；多信息融合；多因素影响
中图分类号：TD 528 **文献标志码**：A **文章编号**：1005-3026(2026)04-0095-08

Coal Gangue Identification Based on Multimodal Fusion Algorithm Under Variable Factors

XIA Rui[1], *WANG Xiang*[1,2], *LI Bo*[1,2], *WANG Xue-wen*[1,2]
(1. College of Mechanical Engineering, Taiyuan University of Technology, Taiyuan 030024, China; 2. Shanxi Key Laboratory of Intelligent Fully Mechanized Mining Equipment Technology, Taiyuan 030024, China. Corresponding author: LI Bo, E-mail: libo@tyut.edu.cn.)

Abstract: To cope with the impact of environmental factors on the coal gangue identification task, a novel network architecture that fused spectral and image features was proposed. The architecture mined rich feature information by integrating the information of two modalities, so as to solve the feature degradation problem of a single modality in harsh environments. In order to verify the model performance, multi-factor dataset and working condition dataset were constructed under laboratory conditions, and spectral feature extraction (1DCNN+CBAM) and image feature extraction (ResViT) networks were built respectively, among which the ResViT network combined the residual mechanism with the visual Transformer for the first time for the task of coal and gangue sorting, which was able to efficiently extract the inter-modality specific features. The reasonableness of each branch collocation was determined by design comparison and ablation test, and the stability of the model was verified under two datasets, with an accuracy rate of up to 97.92%. It is proven that the proposed multimodal fusion algorithm is an efficient classification model, which can provide strong theoretical support for underground coal and gangue sorting.

Key words: coal gangue identification; near-infrared spectroscopy; image; multi-information fusion; multi-factor effect

在煤炭开采过程中，矸石作为与煤炭共生的废弃物，其处理问题日益凸显.矸石不仅可燃性弱，且在燃烧过程中会释放大量有害物质，对环境造成污染，因此，有效的矸石分离是亟待解决

收稿日期：2024-12-24
基金项目：山西省重点研发计划项目(202502020101022)；山西省回国科教创新资助项目(2025-104)；山西省回国留学人员科研资助项目(2024-049).
作者简介：夏 蕊(1988—)，女，陕西渭南人，太原理工大学副教授.
通信作者：李 博，E-mail:libo@tyut.edu.cn.

的关键问题.随着人工智能的发展,更为高效的煤矸分选技术应运而生,常用的方法有光谱识别、射线识别和图像识别等[1].近红外光谱和图像分析技术因其具有无损、快速等优点而被广泛应用.图像处理技术与深度网络的结合,克服了传统图像手工提取特征的繁琐.Gao等[2]将CNN与SVM结合,提高了煤矸识别的准确率,且实现了更快的识别速度.Pu等[3]使用VGG16并结合迁移学习的方法,实现了煤矸石的正确分类,验证集准确率达到了82.5%.光谱技术与卷积网络相结合,运用了卷积网络的强大特征提取能力,克服了传统分析方法无法处理干扰信息的局限性.Xiao等[4]使用深度置信网络提取光谱数据特征,并基于正则化双层极限学习机,有效预测出了煤的成分信息.Le等[5]使用卷积神经网络提取光谱特征,并构建了极限学习机算法成功预测出了煤的组分信息.

近年来,有学者将近红外光谱与图像技术融合进行样品的分类研究.Chen等[6]将近红外光谱与图像技术融合实现梨的疾病检测,结果证明融合模型在疾病检测中有着较好的性能.Song等[7]联合近红外光谱技术与计算机视觉对茶叶品质进行分级,结果证明卷积网络提取特征信息的方法对融合模型分类效果最好.

在煤矸石实际分选过程中,其表面会被煤粉、水分沾染,且设备光源的不稳定也会影响分类模型的性能.曹现刚等[8]针对光线不足的情况,提出了一种融合多个条件下的新型图像获取方案,结果表明该方法可以显著提高识别率.Luan等[9]针对真实采矿环境下光照不足的情况,提出了一种将CNN与SWIN变换器相结合的方法,并验证了该模型的性能.

在井下生产环境中,必定存在多种复杂条件交互作用的情况,此时基于单模态的分类方法可能无法达到理想的效果.本文在深度学习方法的基础上引进光谱和图像信息融合技术,构建双模态信息融合网络,通过消融试验与对比试验,验证各模块搭建的合理性,并将融合模型与单模态模型进行对比,分析融合模型相较于单模态模型的性能优劣.最后针对实际环境可能遇到的影响因素,模拟3种环境因素的混合条件并设计了3因素3水平正交表,探究环境对模型性能的影响.本研究可为多模态融合技术用于煤矸井下分选提供理论基础.

1 材料和方法

1.1 图谱数据采集

为验证模型性能,本文构建了多因素数据集、工况数据集两类具有不同特征的数据源.多因素数据集综合考量多种潜在因素的影响,工况数据集则聚焦于模拟实际工作条件下的数据特征,二者可为模型性能的全面评估提供坚实的数据支撑.试验样本均来自山西省太原市西铭煤矿的焦煤和黑色矸石,尺寸在30~50 mm.光谱仪采集的光谱数据波长在370~1 049 nm,共计特征2 048个;面阵相机采集的图像分辨率为2 048×2 448.数据采集过程中,将煤矸样本随机置于运动状态下的传送带上,传送带有效输送长度为4 m,宽度为0.8 m,运行速度可在0~2 m/s范围内调节.不同数据集对应的煤和矸石样本图像如图1所示.

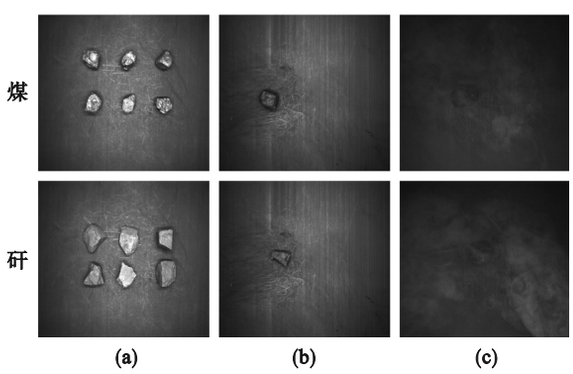


图1 煤和矸图像对比

Fig. 1 Comparison of coal and gangue images

(a)—理想数据集;(b)—多因素数据集;(c)—工况数据集.

1)多因素数据集.构建多因素数据集时,采用正交试验[10]的方法对不同照度、表面煤粉沾染量以及表面不同水分等情况展开工况模拟.多因素数据集使用光谱与图像样本数各240,并将其按照8∶2的比例划分为训练集和测试集,其中训练集数量为192,测试集数量为48.随后,将训练集部分按照3折交叉验证进行划分训练,训练集数量为128,验证集数量为64.表1为正交试验方案.

为确保试验的严谨性,在模拟多因素条件前,选择质地均匀的煤和矸石颗粒作为测试样本.然后将样品在暗室中存放48 h,以确保每个样品表面足够清洁.在照度模拟试验中,使用可调光源将照度分为2 000,3 000和4 000 lx 3个梯

度.在煤粉沾染试验中,洁净和未受污染的煤矸石样本被均匀涂上煤粉,以模拟生产活动中的污染.所使用的煤粉来自使用破碎机粉碎的原煤.同时,在均匀涂抹过程中,将煤粉的质量分别设定为0.05,0.10和0.15 g,并使用电子仪器测量其质量.在表面湿度试验中,设置0.08,0.16 和0.24 mL 3个水梯度,每次加湿试验均通过喷雾对其表面进行均匀加湿,喷洒量使用重量法进行控制.

表 1　正交试验设计

Table 1　Orthogonal experiment design

试验环境	1	2	3	4	5	6	7	8	9
照度/lx	2 000	2 000	2 000	3 000	3 000	3 000	4 000	4 000	4 000
粉尘沾染量/g	0. 05	0. 10	0. 15	0. 05	0. 10	0. 15	0. 05	0. 10	0. 15
水分/mL	0. 08	0. 16	0. 24	0. 16	0. 24	0. 08	0. 24	0. 08	0. 16

2) 工况数据集.为探究模型在实际生产中的应用效果,本文模拟实际生产环境制备了工况数据集.该数据集采集光谱与图像样本数各122,并将其按照8:2的比例划分为训练集和测试集,在训练集基础上进行交叉验证训练,并使用独立测试集对每一轮训练结果进行测试.借助此数据集,能够全方位评估模型在实际生产条件中的稳定性与可靠性.为模拟井下工作环境,采集1 460 lx卤素灯照度下的光谱,并将面阵光源的强度调整为1 000 lx来获取图像,以模拟低照度条件.利用烟雾发生器产生的烟雾,模拟实际生产中的粉尘环境,总烟雾质量浓度保持在0.2~0.5 mg/m^3.此外,样品表面均涂抹100 g/L的煤水混合物,模拟降尘系统喷水后与其他粉状杂质混合的效果.

1.2　数据采集系统

光谱仪(Oceanview optics usb 2000+,Ocean Insight)连接光纤探头(1SMA1s-SI0.6-1.5 m,SL)实现数据采集.其配套的Oceanview 2.0.7作为采集显示软件,卤素灯(BBGD 12 V50 W,12 V75 W,12 V100 W MR16)作为光源.使用海康威视配有定焦镜头(MVL-MF0828M-8MP)的工业面阵相机(MV-VA050-12UC)采集图像数据,图像的固有分辨率为2 048×2 448,配备的光源为可调工业面阵光源(JS-LS18032),其中用于测量光照强度的照度计(DL 333204)量程为0~5 000 lx.设备的结构布局如图2所示.

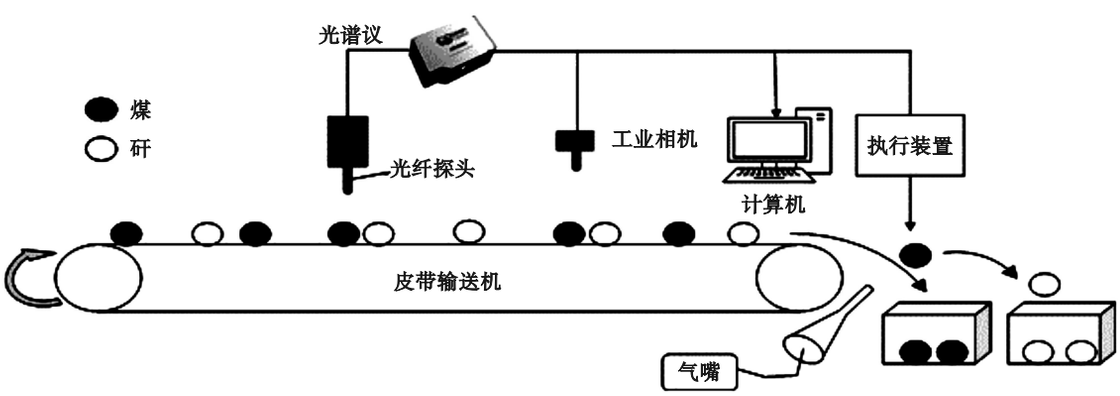


图2　设备示意图

Fig. 2　Schematic diagram of equipment

2　融合模型架构

2.1　光谱特征提取

在光谱分析过程中,选择有效的预处理方法可以极大程度减少非目标因素影响,提高数据质量.在本研究中,使用Savitzky-Golay平滑对采集的光谱数据进行预处理,该方法可以在保证样本特性的同时减少干扰[11].预处理后进行光谱特征提取操作,其网络结构如图3所示.在神经网络中,全连接层主要被用作整合前面层的输出[12],但在该网络中被用作改变数据维度固定输出.此外,为克服CNN在特征提取过程中出现的特征丢失问题[13],在网络中加入CBAM(convolutional block attention module)注意力机制模块[14],以实现性能的提升.CBAM是结合了通道注意力和空间注意力的注意力模块,其中通道注意力 $\boldsymbol{M}_c \in \mathbf{R}^{C\times1\times1}$ 能够实现通道选择.空间注意力模块 $\boldsymbol{M}_s \in$

$\mathbf{R}^{1\times H\times W}$能够选择注意区域.计算过程如下:

$$\boldsymbol{M}_{\mathrm{c}}(F)=\sigma(\mathrm{MLP}(\mathrm{AvgPool}(F))+\mathrm{MLP}(\mathrm{MaxPool}(F))), \tag{1}$$

$$\boldsymbol{M}_{\mathrm{c}}(F)=\sigma(\boldsymbol{W}_1(\boldsymbol{W}_0(F^C_{\mathrm{avg}}))+\boldsymbol{W}_1(\boldsymbol{W}_0(F^C_{\mathrm{max}}))). \tag{2}$$

式中:σ 为 Sigmoid 函数;$\boldsymbol{W}_0$ 和 $\boldsymbol{W}_1$ 为 MLP 的权重.

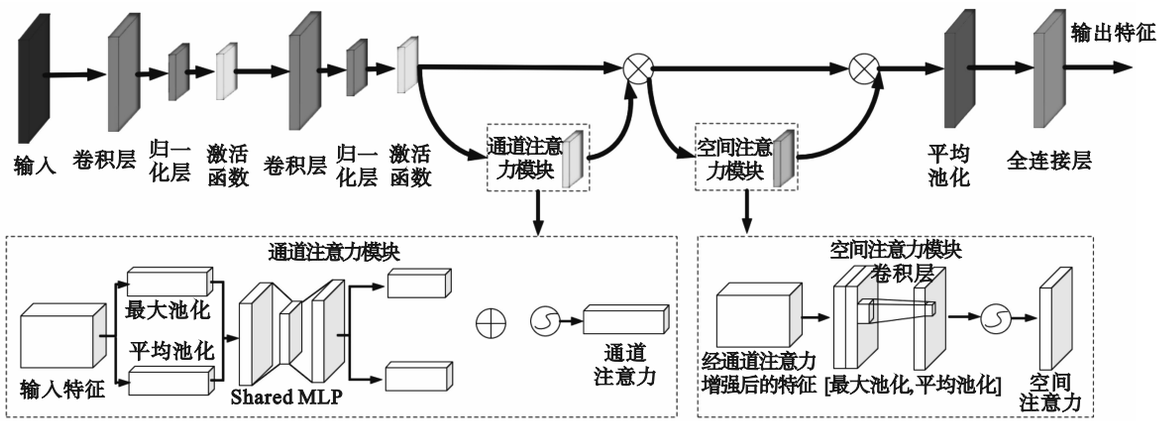


图3 光谱特征提取

Fig. 3 Spectral feature extraction

空间注意力的流程:首先将输入特征通过全局和平均池化进行通道压缩,并将结果进行拼接,最后将拼接结果进行卷积和 Sigmoid 函数处理.计算过程如下:

$$M_{\mathrm{s}}(F')=\sigma(f^{7*7}([\mathrm{AvgPool}(F');\mathrm{MaxPool}(F')])), \tag{3}$$

$$M_{\mathrm{s}}(F')=\sigma(f^{7*7}([(F')^S_{\mathrm{avg}};(F')^S_{\mathrm{max}}])). \tag{4}$$

式中,f^{7*7}表示卷积核大小为7*7的卷积运算.

2.2 图像特征提取

Vision Transformer(ViT)模型在视觉领域具有较大潜力,尤其在大数据集上与经典卷积网络相比有着较强的识别效果[15].然而,ViT 和 CNN 在数据处理方面关注度不同,CNN 关注数据的局部特征,ViT 模型则关注全局信息和上下文特征,对局部特征表现不明显.因此,本文将 CNN 与 ViT 相结合完成特征提取,实现网络互补.混合网络结构如图4所示.

在混合网络结构中,将 ResNet18 残差网络与 ViT 相结合,既保留了残差网络信息的完整性又结合了 ViT 的强大特征表示能力[16].ResNet 是经典的残差网络框架,在本文中将层4作为输出,该层是经过 ResNet18 所有残差块之后的输出,同时为了适应输出减小参数量使用自适应平均池化层对其尺寸进行缩减.由于 ViT 模型的输入是一种展平的图像块(patch),每个块都有固定的维度,为了适应输出格式将特征图从形状(b,c,h,w)转换为形状(b,n,c),其中:n是所有 patch 的数量$(n=h\times w)$,b为批次,c为通道数,h为高度,w为宽度.

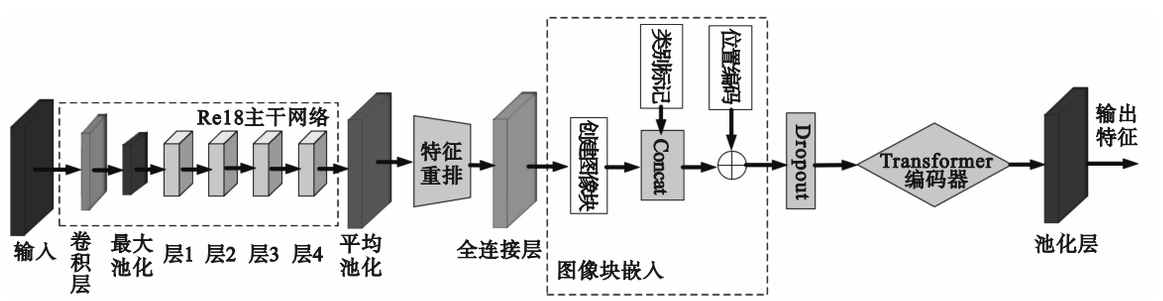


图4 混合网络结构

Fig. 4 Hybrid network structure

将输入序列进行线性映射得出 ViT 所需要的维度,并对输入的数据进行嵌入处理,在序列的最前面添加 class token 和位置嵌入,位置编码是为了使模型能够理解序列数据之间的相对位置关系,class token 作为一个全局表示,用于捕捉和整合目标中的小块信息.最后将 patch embedding 和 class token 拼接起来一同输入 Transformer 编码器中,Transformer 编码器模块结构如图5所示.

多头注意力是在自注意力的基础上进行的,其工作流程首先将输入序列 $X=(x_1,x_2,\cdots,x_n)$ 映射到更高的维度上得到 a_i,之后分别与 $\boldsymbol{W}^q$,$\boldsymbol{W}^k$,$\boldsymbol{W}^v$ 权重矩阵相乘生成对应的 q^i,k^i,v^i,将所有的

q，k 和 v 结合得到对应的 $\boldsymbol{Q}$（Query），$\boldsymbol{K}$（Key）和 $\boldsymbol{V}$（Value）. 紧接着通过另一个线性映射将注意力头合并，计算 q 和 k 的点积并进行缩放处理，最后对结果使用 Softmax 函数计算 v 上的注意力权重，得到最终的输出：

$$\text{Attention}=(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_{\mathrm{k}}}}\right)\boldsymbol{V}. \quad (5)$$

式中，d_{k} 为键向量维度.

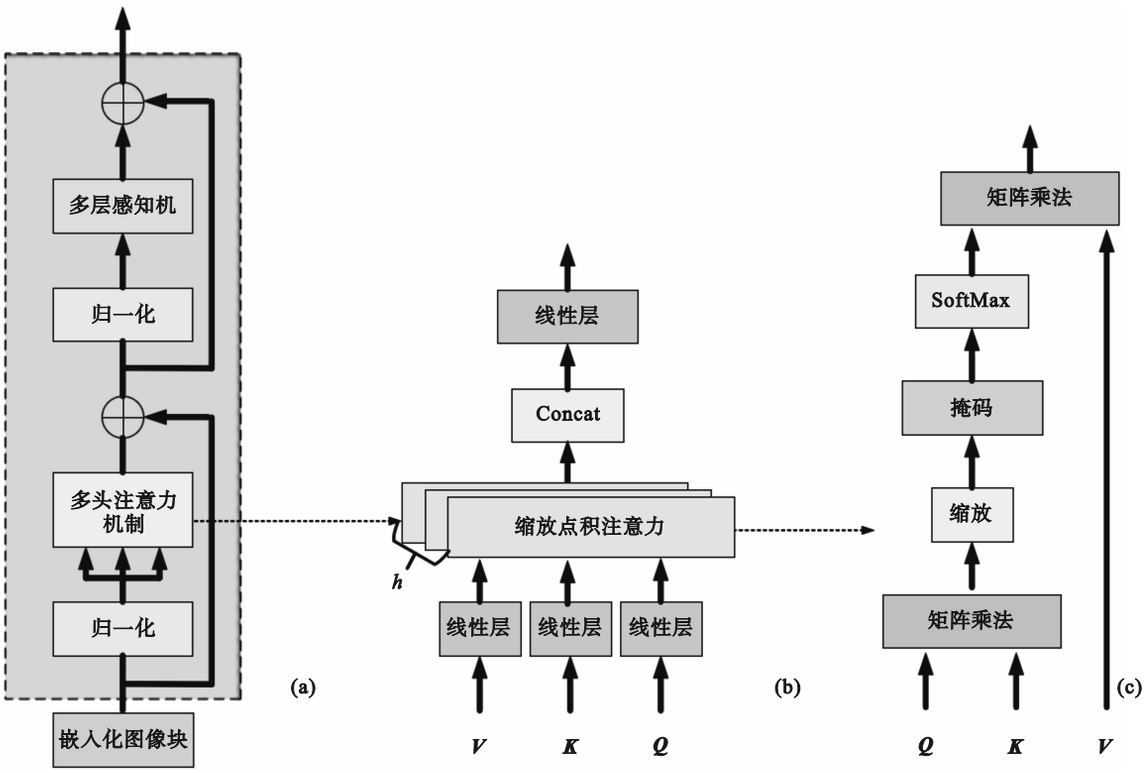


图 5　Transformer 编码器模块网络结构示意图

Fig. 5　Schematic diagram of network structure of Transformer encoder module

（a）—Transformer 编码器；（b）—多头注意力；（c）—缩放点积注意力.

将 $\boldsymbol{Q}$，$\boldsymbol{K}$，$\boldsymbol{V}$ 投影到一个低维度空间，经过 h 次不同的线性投影操作，形成多个不同的表示（注意力头），接着将每个头独立进行注意力计算：

$$\text{Head}_i=\text{Attention}(\boldsymbol{Q}\boldsymbol{W}_i^{Q},\boldsymbol{K}\boldsymbol{W}_i^{K},\boldsymbol{V}\boldsymbol{W}_i^{V}). \quad (6)$$

式中：$\boldsymbol{W}_i^{Q}\in\mathbf{R}^{d_{\text{model}}\times d_{\mathrm{k}}}$；$\boldsymbol{W}_i^{K}\in\mathbf{R}^{d_{\text{model}}\times d_{\mathrm{k}}}$；$\boldsymbol{W}_i^{V}\in\mathbf{R}^{d_{\text{model}}\times d_{\mathrm{v}}}$；$d_{\mathrm{v}}$ 为值向量维度.

将每个头的计算结果按维度进行拼接，并通过线性变换得到多头注意力输出：

$$\text{MultiHead}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{Concat}(\text{head}_1,\text{head}_2,\cdots,\text{head}_h)\boldsymbol{W}^{O}. \quad (7)$$

式中，$\boldsymbol{W}^{O}\in\mathbf{R}^{d_{\text{model}}\times hd_{\mathrm{v}}}$.

2.3　融合网络

融合领域中，特征级融合方法能够提取不同类传感器的信息特征，并将特征信息进行拼接，形成丰富的特征组合，从而减少信息丢失现象[17]. 在特征级融合中，使用 2.1 节的方法提取光谱特征，以 2.2 节的方法提取图像的表面特征. 将来自不同源的信息通过 Concat 进行拼接，充分挖掘序列与图像特征[18]. 将拼接后的特征向量传入 Dropout 层，随机丢弃一些神经元的输出，降低过拟合风险. 为了提高分类性能，使用全连接层，使融合向量能够进行线性变换并学习特征之间的关系，最后添加非线性激活函数 Softmax 将输出转换成概率.

3　结果与讨论

3.1　评价指标

为了评估融合模型的性能，使用准确率（Accuracy）、精确率（Precision）、召回率（Recall）以及 $F1$ 分数从多个角度了解模型性能，如式（8）~（11）所示. 为了直观全面显示模型的性能，引入混淆矩阵，清晰展示每个类被正确或错误分类的数量.

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}, \quad (8)$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}}, \quad (9)$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}},\tag{10}$$

$$F1=\frac{2\text{TP}}{2\text{TP}+\text{FP}+\text{FN}}.\tag{11}$$

其中:TP为正确预测为正类别的样本数量;TN为正确预测为负类别的样本数量;FP为错误预测为正类别的样本数量;FN为错误预测为负类别的样本数量.

3.2 模型性能验证

融合模型由图像特征提取网络ResViT与光谱特征提取网络CNN-CBAM组成,在不改变图像特征提取网络的基础上,为验证CNN-CBAM网络在融合后模型中的表现能力,本文将经典神经网络VGG16、DenseNet以及Resnet18网络替换CNN-CBAM分支与ResViT分支完成融合任务[19-20].为减少融合后不必要的计算复杂度,在对比试验中不使用过深的网络结构,本着在有限资源的前提下,使用最简单的方法最大程度地完成融合,这也是本文的融合原则.

在对比试验中,将多因素数据集按照6∶4划分为训练集和测试集,得到训练集144个,测试集96个.在训练过程中保证所有网络的参数一致.如图6所示,从图中可以看出CNN-CBAM与ResViT融合后的效果最好,测试集准确率为98.96%,精确率为98.98%,召回率为98.96%,$F1$分数为98.96%.ResNet网络融合后的模型效果次之,DenseNet网络融合的效果较差.3个经典网络的融合效果没有CNN-CBAM好,但四者数值相差不大,均达到95%以上,可以看出多源信息融合能够很好地完成分类任务.

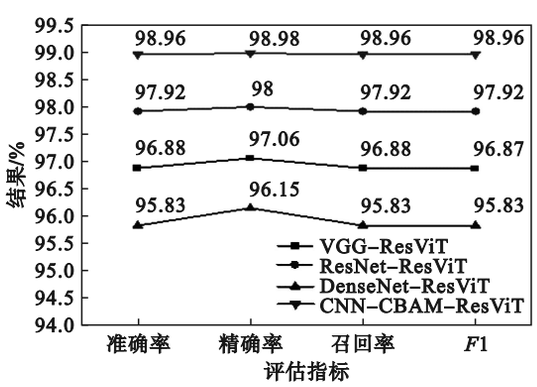


图6 对比试验结果

Fig. 6 Comparison experiments results

为了进一步说明本文融合模型的效果,在图7中给出4种模型的混淆矩阵,主对角线上是每个类别的正确预测,可以看出本文融合模型只有1个样本被错误识别,将矸识别成了煤,其次是ResNet网络,DenseNet网络效果最差,证明本文网络模型融合效果较好.

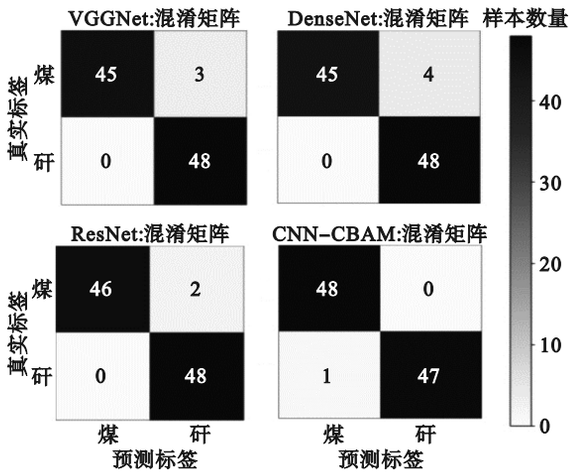


图7 4种融合模型的混淆矩阵

Fig. 7 Confusion matrix for four fusion models

3.3 消融试验

对于图像特征提取分支,本文使用ResNet18与ViT的混合模型完成特征提取任务,为验证ResNet18模块存在的必要性,借助消融试验的方法证明,消去ResNet18模块直接使用ViT模型进行特征提取并完成融合分类,结果如图8a所示.从图中可以看出,消去ResNet模块前后,4个指标均有所降低,尤其是召回率指标,同时也证明了两个网络结合的优势.为了验证融合模型性能,将单双模态进行了比较.如图8b所示,融合模型4种指标均高于单模态模型.准确率、精确率、召回率以及$F1$分数分别为97.92%,100%,95.83%和97.84%.从图中可以看出,光谱分支的召回率与融合模型的数值一致,而召回率为正确识别的正例数与实际正例总数的比值,其可能原因是光谱分支模型已经捕捉到了正例的关键特征,导致融合模型召回率没有提升.

4 模型在多因素数据集中的性能评估

4.1 参数设置

多因素试验和工况试验均在Python 3.8,环境版本PyTorch 1.9下完成,计算机为Intel Core i5-13500HX 2.50 GHz,配备Nvidia RTX 4050显卡.模型使用了Adam优化器,训练批次为12,正则化参数为0.001,学习率0.000 08.训练过程中采用交叉验证的方法进行训练,测试集独立于训练、验证过程,在训练过程中每次选择2个子集作

为训练集,剩下一个子集作为验证集,此过程重复3次,并将每次得到的评估结果用测试集进行测试,最后取3次结果的平均值.

4.2　模型评估

融合模型在9种不同条件下的训练结果如表2所示.结果表明融合模型无论在何种条件因素下均能达到较好的性能,准确率、精确率、召回率以及$F1$分数均在90%左右,最低值出现在试验的第3组中,准确率、精确率、召回率以及$F1$分数分别为90.28%,94.44%,86.11%和89.22%.其可能原因是在正交试验设计中第3组实验煤粉沾染量和表面湿度达到了梯度水平的最大值,进而影响模型性能,同时也说明煤粉和水分在煤矸分类任务中会对模型产生影响.

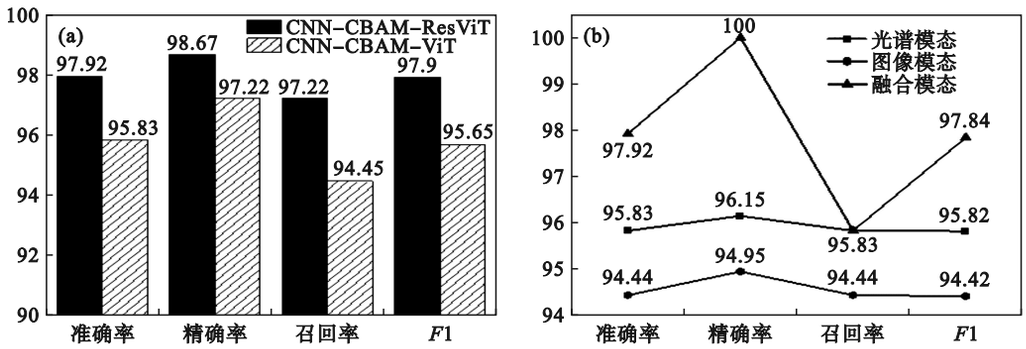


图8　试验结果

Fig. 8　Test results

(a)—消融试验结果;(b)—单双模态对比结果.

表2　测试集结果

Table 2　Test set results　　%

条件	评价指标			
	准确率	精确率	召回率	$F1$
因素1	97.92	100	95.84	97.87
因素2	97.22	97.22	97.22	97.13
因素3	90.28	94.44	86.11	89.22
因素4	96.53	97.22	95.83	96.39
因素5	95.83	94.45	97.22	95.65
因素6	91.67	96.96	86.11	91.14
因素7	97.22	94.45	100	97.10
因素8	95.14	95.83	94.45	94.88
因素9	96.53	98.61	94.45	96.39

5　模型在工况数据集中的性能评估

为验证融合模型性能,本研究使用准确率、精确率、召回率以及$F1$分数等4个评价指标对模型进行全方位评估,探讨融合模型在实际工况下的适应性.训练过程中使用3折交叉验证的方法,提高训练结果的准确性,减少随机误差.通过3次训练取其平均值,所得结果如图9所示.从图9可以看出,融合模型在工况数据集中表现良好,准确率、精确率、召回率以及$F1$分数分别达到了93.33%,94.44%,90.91%,95.24%.对比多因素数据集和工况数据集的指标差异可见,工况数据集中存在的粉尘、低照度环境会对模型产生不利影响,使模型的识别效果变差,同时从侧面可以说明所搭建融合模型相较于单模态模型有较高的稳定性和抗干扰能力.

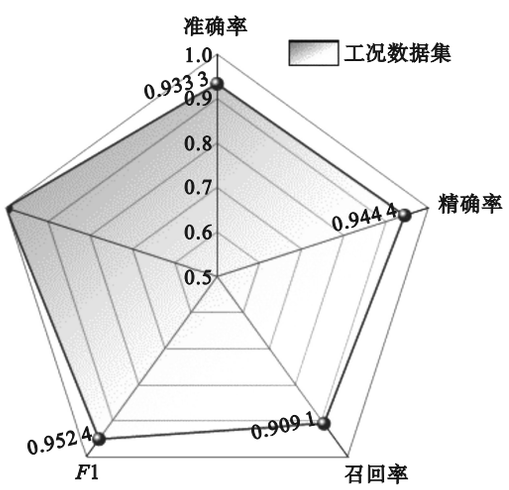


图9　工况数据集下融合模型性能雷达图

Fig. 9　Radar plot of performance of fusion model under working condition dataset

图10为交叉验证中各独立测试集的混淆矩阵结果,通过对比不同折的混淆矩阵,可以洞察模型在不同数据划分下的稳定性,为后续模型的优化和改进提供依据.

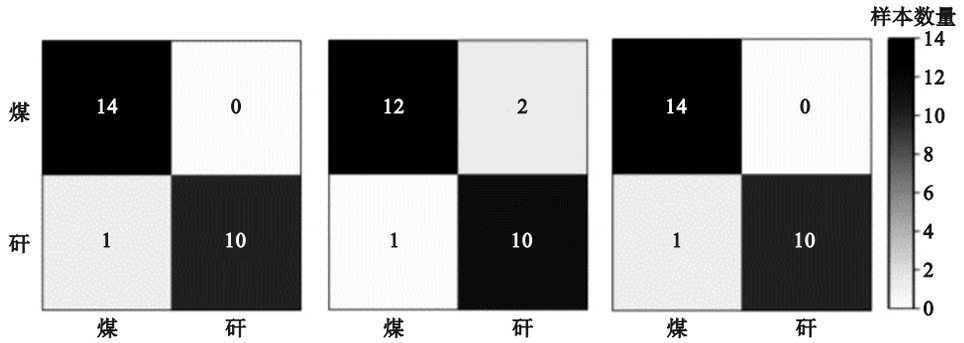


图 10 工况数据集下各折交叉验证对应的混淆矩阵

Fig. 10 Confusion matrix corresponding to cross-validation of each fold under working condition dataset

6 结 论

1) 本研究探索模型在不同环境影响下的适应性.采用了交叉验证取平均的方法对结果进行评估.测试集独立于训练过程,其余数据作为交叉验证的总数据集.结果表明,在多因素数据集和工况数据集下,模型均能达到理想的性能,测试集的准确率均在90%以上.

2) 为探究双模态融合网络性能优势,将融合模型与单模态模型进行了比较.数据表明,融合模型在测试集上的准确率达到97.92%,单模态模型的测试集准确率分别为95.83%和94.44%,可以看出,融合模型具有显著优势.为验证ResViT分支搭配的合理性,本文设计了消融试验.结果显示,残差结构的加入使模型性能提高了2%.

参考文献:

[1] 郭永存,何磊,刘普壮,等.煤矸双能X射线图像多维度分析识别方法[J].煤炭学报,2021,46(1):300-309.
(Guo Yong-cun, He Lei, Liu Pu-zhuang, et al. Multi-dimensional analysis and recognition method of coal and gangue dual-energy X-ray images[J]. *Coal Journal*, 2021, 46(1): 300-309.)

[2] Gao R X, Du Y B, Wang T F. Research on coal gangue classification recognition method based on the combination of CNN and SVM [J]. *Journal of Real-Time Image Processing*, 2023, 20(6): 110.

[3] Pu Y Y, Apel D B, Szmigiel A, et al. Image recognition of coal and coal gangue using a convolutional neural network and transfer learning[J]. *Energies*, 2019, 12(9): 1735.

[4] Xiao D, Le B T. Rapid analysis of coal characteristics based on deep learning and visible-infrared spectroscopy [J]. *Microchemical Journal*, 2020, 157: 104880.

[5] Le B T, Xiao D, Mao Y, et al. Coal analysis based on visible-infrared spectroscopy and a deep neural network[J]. *Infrared Physics & Technology*, 2018, 93: 34-40.

[6] Chen Y F, Liu L, Rao Y, et al. Identifying the "Dangshan" physiological disease of pear woolliness response via feature-level fusion of near-infrared spectroscopy and visual RGB image[J]. *Foods*, 2023, 12(6): 1178.

[7] Song Y, Wang X Z, Xie H L, et al. Quality evaluation of Keemun black tea by fusing data obtained from near-infrared reflectance spectroscopy and computer vision sensors [J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2021, 252: 119522.

[8] 曹现刚,郝朋英,王鹏,等.多因素光照条件下高质量煤矸图像获取方法研究[J].煤炭科学技术,2023,51(1):455-463.
(Cao Xian-gang, Hao Peng-ying, Wang Peng, et al. Research on method of acquiring high quality coal gangue images under multi-factor illumination condition [J]. *Coal Science and Technology*, 2023, 51(1): 455-463.)

[9] Luan H X, Xu H, Tang W, et al. Coal and gangue classification in actual environment of mines based on deep learning[J]. *Measurement*, 2023, 211: 112651.

[10] Zhang L L, Lu Y D, Kang J F, et al. Selection of optimum composition of alumino borosilicate glasses with excellent dielectric properties according to orthogonal experiment design [J]. *Journal of Materials Science: Materials in Electronics*, 2018, 29: 5746-5752.

[11] Zhou L N, Zhou L J Y, Wu H B, et al. Estimation of cadmium content in Lactuca sativa L. leaves using visible-near-infrared spectroscopy technology [J]. *Agronomy*, 2024, 14(4): 644.

[12] Jiao J Y, Zhao M, Lin J, et al. A multivariate encoder information based convolutional neural network for intelligent fault diagnosis of planetary gearboxes [J]. *Knowledge-Based Systems*, 2018, 160: 237-250.

[13] Peng C, Zhong L, Gao L L, et al. Implementation of near-infrared spectroscopy and convolutional neural networks for predicting particle size distribution in fluidized bed granulation [J]. *International Journal of Pharmaceutics*, 2024, 655: 124001.

[14] Hao Y, Zhang C X, Li X Y, et al. Establishment of online deep learning model for insect-affected pests in "Yali" pears based on visible-near-infrared spectroscopy[J]. *Frontiers in Nutrition*, 2022, 9: 1026730.

[15] 徐昊,郭黎,李润泽.基于紧凑型Vision transformer的细粒度视觉分类[J].控制与决策,2024,39(3):893-900.
(Xu Hao, Guo Li, Li Run-ze. Fine-grained visual classification based on compact vision transformer [J]. *Control and Decision*, 2024, 39(3): 893-900.)

[16] Guo C R, Li M X, Xu J F, et al. Ultrasonic characterization of small defects based on Res-ViT and unsupervised domain adaptation [J]. *Ultrasonics*, 2024, 137: 107194.

[17] Xu M, Wang J, Gu S. Rapid identification of tea quality by E-nose and computer vision combining with a synergetic data fusion strategy [J]. *Journal of Food Engineering*, 2019, 241: 10-17.

[18] Zhang H C, Ai G J. An image text sentiment analysis method using bimodal attention mechanism[J]. *Journal of Circuits, Systems and Computers*, 2024, 34: 2550036.

[19] Fayemiwo M A, Olowookere T A, Arekete S A, et al. Modeling a deep transfer learning framework for the classification of COVID-19 radiology dataset [J]. *PeerJ Computer Science*, 2021, 7: e614.