基于SwinV2-EfficientNetV2的铜矿石品位分类方法研究*

黄源峰[1,2],何剑锋[1,2],张峰源[1,2],李卫东[1,2],聂逢君[1,3],汪雪元[1,2],瞿金辉[1]

(1.江西省核地学数据科学与系统工程技术研究中心, 江西 南昌 330013;
2.东华理工大学 信息工程学院, 江西 南昌 330013;
3.东华理工大学, 江西 南昌 330013)

摘要:针对现有铜矿石品位分类中所应用的卷积神经网络缺乏构建与归纳长距离特征关系不足的问题,提出了一种结合Swin Transformer和EfficientNet的集成模型的铜矿石品位分类方法。该方法充分利用了Swin Transformer V2-t架构对长距离特征关系的归纳能力,以及EfficientNet V2-s捕捉细微局部特征上的优势,通过增设线性层以整合两模型的输出结果,并根据单个模型自身输出动态调整线性层的权重,以优化映射关系,进而显著提升分类性能。试验验证表明,此融合模型在分类任务上的准确率达到92.891%,精确率达到93.095%,召回率达到92.654%。相较于未集成前的单一模型,集成后的综合模型在分类准确率上提升了1.3个百分点,精确率分别提升了1.9个百分点和2.186个百分点,召回率则分别提高了0.474个百分点和0.237个百分点。

关键词:铜矿石品位分类;图像分类;模型集成;视觉Transformer;卷积神经网络

中图分类号:TD92 **文献标识码**:A **文章编号**:1005-2763(2025)04-0220-09

Research on Copper Ore Grade Classification Method Based on SwinV2-EfficientNetV2

HUANG Yuanfeng[1,2], *HE Jianfeng*[1,2], *ZHANG Fengyuan*[1,2], *LI Weidong*[1,2],
NIE Fengjun[1,3], *WANG Xueyuan*[1,2], *QU Jinhui*[1]

(1.Jiangxi Engineering Technology Research Center of Nuclear Geoscience Data Science and System, Nanchang, Jiangxi 330013, China;
2.College of Information Engineering, East China University of Technology, Nanchang, Jiangxi 330013, China;
3.East China University of Technology, Nanchang, Jiangxi 330013, China)

Abstract: Aiming at the problem that the convolutional neural network applied in the existing copper ore grade classification lacked the relationship between the construction and the induction of long-distance features, a copper ore grade classification method based on the integrated model of Swin Transformer and EfficientNet was proposed. This method made full use of the inductive ability of Swin Transformer V2-t architecture for long-distance feature relations and the advantage of EfficientNet V2-s in capturing subtle local features. By adding a linear layer to integrate the output results of the two models, and dynamically adjusting the weight of the linear layer according to the output of a single model itself, the mapping relationship was optimized, and the classification performance was significantly improved. The experimental results show that the accuracy rate of the fusion model in the classification task is 92.891%, the accuracy rate is 93.095%, and the recall rate is 92.654%. Compared with the single model before integration, the integrated comprehensive model has improved the classification accuracy by 1.3 percentage points, the accuracy rate by 1.9 percentage points and 2.186 percentage points respectively, and the recall rate by 0.474 percentage points and 0.237 percentage points respectively.

Key words: Copper ore grade classification, Image classification, Model integration, Visual Transformer, Convolutional neural network

---

* **收稿日期**:2024-04-10

基金项目:国家自然科学基金资助项目(U2067202);江西省主要学科学术和技术带头人培养计划项目(20225BCJ22004);江西省重点研发计划项目(20203BBG73069)

作者简介:黄源峰(1999—),男,江西赣州人,硕士研究生,主要从事计算机视觉与自然语言处理等方面的研究工作。E-mail:yuanfenghuang767@gmail.com

通信作者:何剑锋(1977—),男,江西婺源人,博士,教授,主要从事智能核信息处理与嵌入式系统应用开发等方面的研究工作。E-mail:hjf_10@yeah.net

0　引言

铜矿是我国具有战略地位的矿产资源，但目前我国铜矿存在品位低、开发环境恶劣等不利条件[1]，因此对我国制铜工艺提出了更高的要求。而在制铜工艺中，提高矿石分选环节的效率能够帮助提升后续工业流程的制铜率。传统铜矿石选矿工艺一般采用重选法、浮选法、磁选法、人工手选或多种选矿工艺相结合的方式，在选矿效率、成本、环境友好等方面存在较大局限。有研究人员使用计算机视觉辅助区分矿石品位的方法来实现更高效、节能、环保的选矿。如 PATEL 等[2]使用支持向量机(SVM)算法对铁矿石进行品位层次上的分类，出错率仅为 0.27%，但该方法过于依赖人工设计算法进行矿石图像特征的提取。深度学习中的卷积神经网络也被运用到矿石分选当中，CHOROS 等[3]使用卷积神经网络对不同品位矿石的光谱图像进行识别，性能优异。王李管等[4]基于迁移学习的 Wu-VGG19(卷积神经网络)对黑钨矿石与围岩的图像进行分类识别，分辨率超过 90%。

综上所述，目前以计算机视觉辅助矿石分选的方法存在以下几点不足：(1)基于机器学习的方法较为依赖人工提取图像特征，针对矿石图像的预处理工作量大，应对不同品种矿石时模型可移植性较差；(2)受限于卷积神经网络本身不擅长挖掘长距离特征的缺陷[5]，基于卷积神经网络的矿石分选方法在应对铜矿石品位区分这一任务中，由于图像缺乏局部特征，模型可能难以区分其矿石品位。

针对上述问题，为了提高模型归纳长短距离特征的能力，利用卷积神经网络善于处理平移不变性和局部相关性的特性，以及视觉 Transformer[6]能够挖掘长距离的依赖并加以并行运算的特性，这两者相结合[7-8]组合成 EfficientNetV2-s-SwinV2-t 融合模型对铜矿石进行品位分辨。同时，在数据获取成本较高的情况下借助迁移学习(预训练数据为 ImageNet-1k)对模型进行预训练[9]，减少后续训练的数据需求量。因此，提出一种基于 Swin Transformer-EfficientNet 集成模型的铜矿石品位分类方法，应用双能 X 射线成像装置，将收集的不同品位的铜矿石双能 X 射线图像进行差异化处理，最后使用 EfficientNetV2-s-SwinV2-t 融合模型进行分类，从而达到铜矿石品位分辨的目的。

1　试验数据

1.1　数据集收集

矿石样本由某主营矿石分选相关领域的企业提供，矿石种类为黄铜矿[10]。将矿石样本送入自主搭建的图像采集平台进行图像采集，得到双能 X 射线矿石图像数据。平台设计如图 1 所示，由矿石粉碎系统、物料传输系统、工业 X 射线成像系统、计算机系统 4 部分组成。其中，矿石粉碎系统负责将矿石进行破碎，控制矿石颗粒直径范围为 7±2 mm，高品位铜矿石批次中平均铜质量占比约为 2.5%，低品位铜矿石批次中平均铜质量占比约为 0.1%[11]。经由传送带将矿石颗粒运送至成像系统进行 X 射线成像后，由人工使用计算机对影像进行图像分割、标签赋值、存储。

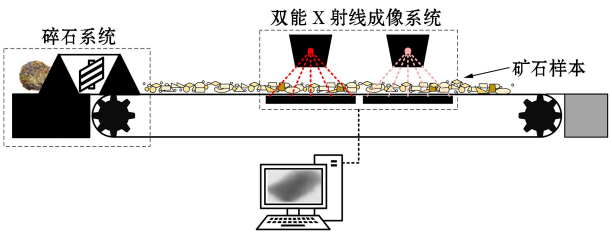


图 1　图像采集平台

Fig.1　Image capture platform

1.2　图像数据差异化处理

双能 X 射线成像设备对铜矿石进行图像采集时，图像当中金属部分与非金属部分区别较为明显，但由于 Cu 与 Fe 元素的原子序数接近，图像中富集该两种元素的部位之间未能显现出较大的阴影差距[12]，致使图像有效特征规模较小。为进一步扩大阴影差异，扩充图像特征规模，利用高能图像与低能图像的对位像素值相减从而得到差异图像，如图 2 所示。

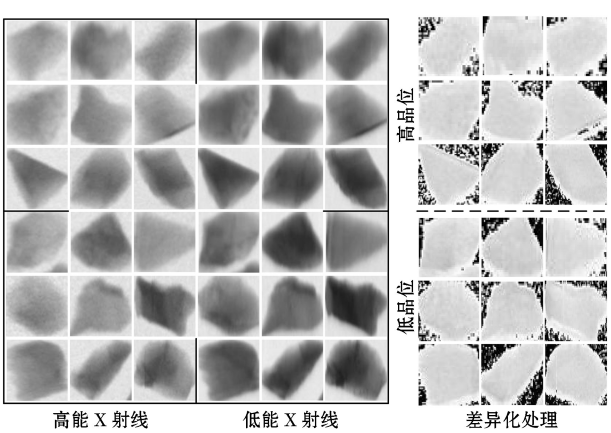


图 2　预处理矿石碎粒图像

Fig.2　Images of preprocessing ore fragments

试验图像数据包含高能 X 射线矿石碎粒图像与低能 X 射线矿石碎粒图像各 5 612 张,训练集、验证集、测试集的比例为 7 ∶ 1.5 ∶ 1.5,具体设置见表 1。

表 1 数据设置

Table 1 Data settings

数据集		所占比例/%	矿石碎粒图像数量/张	
			高能 X 射线下	低能 X 射线下
训练集	高品位铜/低品位铜	70	1 964/1 964	1 964/1 964
验证集	高品位铜/低品位铜	15	420/420	420/420
测试集	高品位铜/低品位铜	15	422/422	422/422

2 试验原理与方法

2.1 试验原理

2.1.1 EfficientNetV2-s 模型

EfficientNetV2 相对于 EfficientNet 摒弃了复合缩放尺寸规则,而是随着训练进行不断增加层(Layers)的数量,并采取限制缩放最大图片和最小图片的规则[13-14]。由于使用深度卷积不能很好地利用现代加速器(GPU),EfficientNetV2 提出了融合移动卷积模块并与移动卷积模块同时使用,融合移动卷积模块直接使用 3×3 的卷积替换 1×1 卷积加 3×3 卷积,并在颜色层扩充成 4 通道,加快了运行速度。EfficientNetV2 系列模型根据模型深度结合实际性能表现,主要研究对象是 EfficientNetV2-s,其具体结构如图 3 所示。

相比于传统卷积神经网络,融合移动卷积模块与移动卷积模块能够将原本的图像数据进一步卷积和直接输送到模块尾部,能够在提取局部特征的同时兼顾图像的全局数据,这一特性与试验所使用的铜矿石中铜部位较为发散的特点相适应。

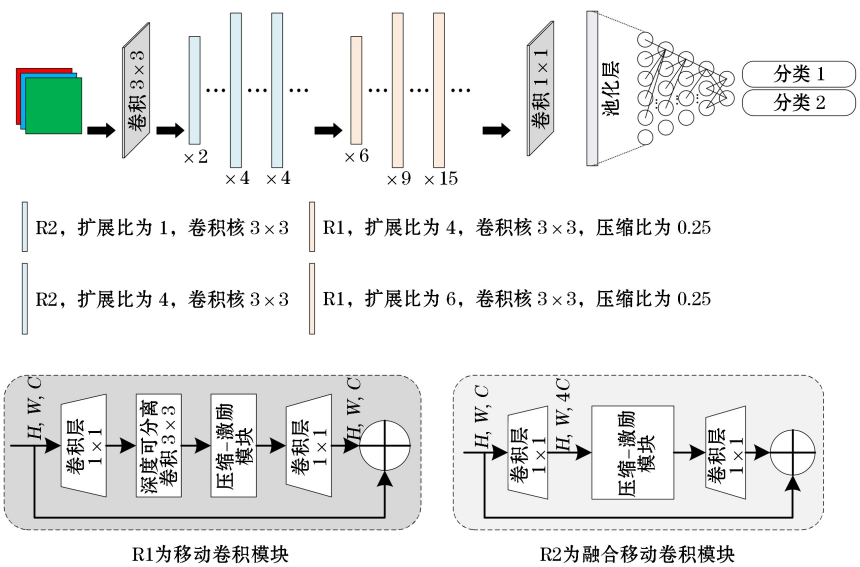


图 3 EfficientNetV2-s 网络结构

Fig.3 EfficientNetV2-s network structure

2.1.2 SwinV2-t 模型

相较于 WANG 提出的 Vision Transformer (ViT),Swin Transformer 将自注意力机制的计算限制在一个个滑动窗口内而并非全局运算[15],减少了计算复杂度和显存占用。同时,将每个窗口内的特征图沿着不同的方向循环移动,从而增加了窗口间的信息交互,在处理铜矿石样本时能够收集不同部位的特征信息,并加强关联性。Swin Transformer 的窗口注意力机制的数学表达式如下:

$$\mathrm{WindowAttention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}+\boldsymbol{B}\right)\boldsymbol{V} \tag{1}$$

式中:$\boldsymbol{Q}$ 为查询矩阵,表示在输入数据中搜索相关信息的模型部分;$\boldsymbol{K}$ 为键矩阵,用于存储与输入序列相关的信息,衡量元素当前位置与其他位置的相关性;$\boldsymbol{V}$ 为值矩阵,包含了注意力机制找到相关键后需要检索的实际信息内容;$\boldsymbol{B}$ 为相对位置编码矩阵,用以引入关于窗口内元素相对位置的信息,其尺寸匹

配 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 点积之后的形状；d 为特征维度。针对不同分辨率的图像(见图 4)，只进行单个红框(颜色标识见电子版)内的自注意计算，再进行红框之间的自注意力计算，相较全局自注意力运算，大幅减小了运算复杂度。

Swin TransformerV2 使用 res-post-norm 取代之前的 pre-norm 配置[15-16]；使用缩放的余弦注意力来替代原来的点积注意力；使用对数间隔的连续相对位置偏差方法来替代之前的参数化方法。SwinV2-t 的结构如图 5 所示。

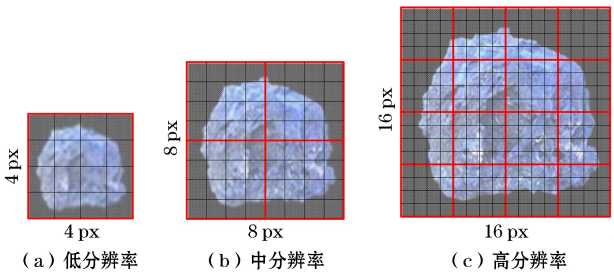


图 4　不同分辨率图片下的自注意力处理机制

Fig.4　Self-attention processing mechanism for dealing with images of different resolutions

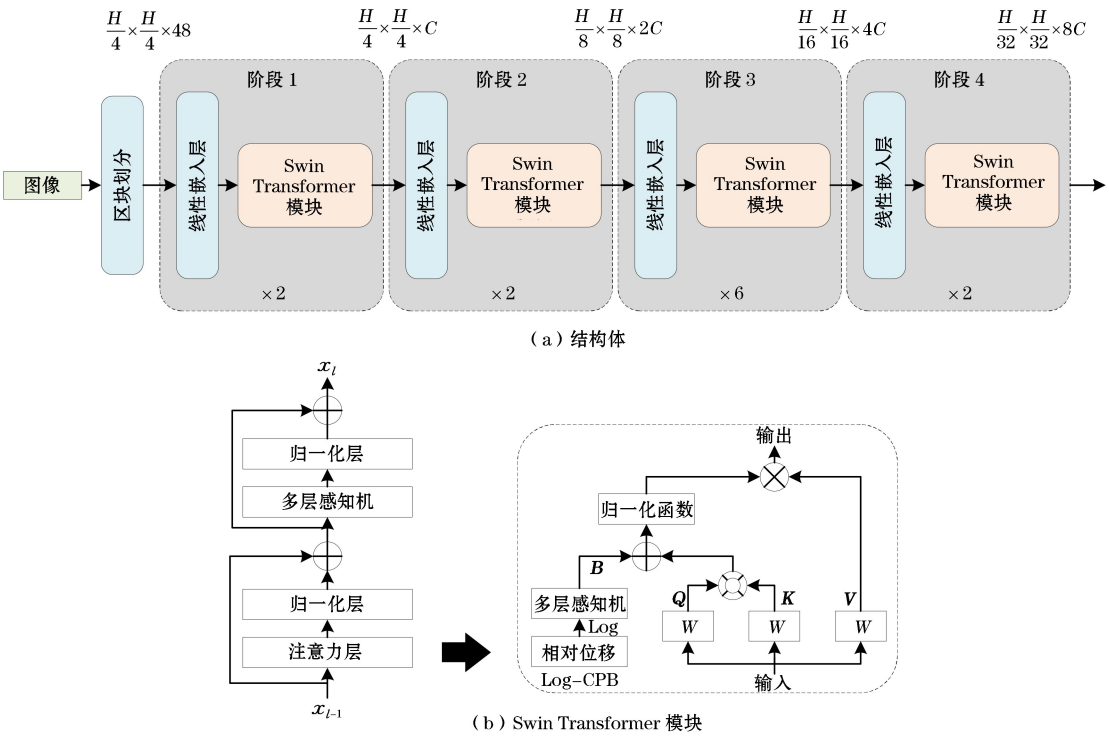


图 5　SwinV2-t 结构

Fig.5　SwinV2-t structure

2.1.3　双模型输出融合

集成学习可以提高预测性能和泛化能力，使得模型具有更强的稳定性和鲁棒性[17]。试验所选用的 EfficientNetV2 模型与 SwinV2-t 模型分别起源于 CNN 与视觉 Transformer，前者善于处理平移不变性与局部相关性，后者能够挖掘长距离的依赖并加以并行运算，同时两者结构与技术路线相差较大，在集成之后可能会表现出更加优异的性能[18]。

集成方式为使用线性层将两个并行运行的模型的全连接层的输出作为输入，线性层的输出为两个神经元(二分类)，如图 6 所示。

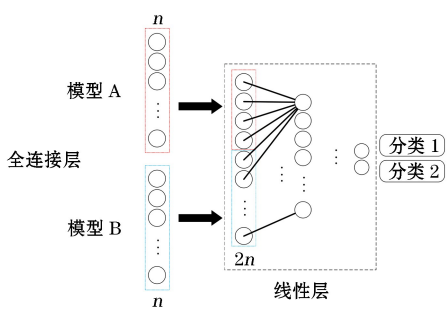


图 6　模型集成示意

Fig.6　Schematic diagram of model integration

假设设有训练数据集 $\{(x^{(1)},y^{(1)}),(x^{(2)},y^{(2)}),\cdots,(x^{(m)},y^{(m)})\}$，其中，$x^{(i)}$ 为第 i 个样本的

特征,$y^{(i)}$ 为对应的真实标签的独热编码。定义训练时线性方程模型为:

$$\hat{y}^{(i)}=\mathrm{Softmax}(\boldsymbol{w}\times x^{(i)}+\boldsymbol{b}) \tag{2}$$

式中:$\hat{y}^{(i)}$ 是模型对 i 个样本的预测概率分布;$\boldsymbol{w}$ 为权重矩阵;$\boldsymbol{b}$ 为偏置向量。

定义交叉熵损失函数:

$$J(\boldsymbol{w},\boldsymbol{b})=-\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{c}y_j^{(i)}\log(\hat{y}_j^{(i)}) \tag{3}$$

式中:$J(\boldsymbol{w},\boldsymbol{b})$ 是交叉熵损失;m,c 是类别数;$y_j^{(i)}$ 是样本 i 的真实标签的第 j 个元素。

最小化交叉熵损失值通过梯度下降来减小,使 $\boldsymbol{w}$ 和 $\boldsymbol{b}$ 趋近于最优值,梯度下降的迭代更新规则为:

$$\boldsymbol{w}=\boldsymbol{w}-\alpha\frac{\partial J}{\partial \boldsymbol{w}} \tag{4}$$

$$\boldsymbol{b}=\boldsymbol{b}-\alpha\frac{\partial J}{\partial \boldsymbol{b}} \tag{5}$$

式中:α 是学习率;$\frac{\partial J}{\partial \boldsymbol{w}}$ 和 $\frac{\partial J}{\partial \boldsymbol{b}}$ 分别是损失函数关于 $\boldsymbol{w}$ 和 $\boldsymbol{b}$ 的偏导数。

2.2 试验方法

在实际矿物品位识别研究中,由于矿石样本采集困难且成本高昂,直接基于有限的样本量构建一个深度学习模型可能会导致过拟合、泛化能力低下等问题,尤其是在面对高维度图像数据和复杂的纹理特征时。因此,借鉴计算机视觉领域的先进实践,采用迁移学习的方法来有效利用已有的大规模图像数据集——ImageNet-1k 预训练的模型。ImageNet-1k 是一个包含 1 000 个类别、共有 1 281 167 张图像的数据集,是图像分类任务中最常用的基准数据集之一[19]。通过在如此庞大的数据集上预先训练的卷积神经网络(CNN),能够学习到丰富的底层特征表示,包括边缘检测、形状识别、颜色分布以及其他高级抽象概念等。这种预训练过程赋予了模型一种强大的先验知识,即对自然界各种物体的普适视觉理解能力。

为了适应特定的矿石分类任务,试验并非从零开始训练一个新的模型,而是基于在 ImageNet-1k 上预训练得到的网络架构,冻结部分或全部早期层以保留其捕捉通用特征的能力,同时微调(fine-tune)网络的后期层以适应特定矿石图像的特点。这样的策略允许试验在相对较少的专用矿石样本上进行高效训练,利用预训练模型已经学习到的特征表达能力,减少对大量标记样本的需求,进而克服因样本不足带来的训练难题。该方法可以在有限的试验数据集上实现较高的模型性能,并缩短训练时间,提高最终模型在新样本上的泛化性能,对于具有高成本采样特性的矿石品位分类研究至关重要。表 2 展示了所使用的基准模型在 ImageNet-1k 的表现。

表 2 试验所用模型的相关信息

Table 2 Information about the model used in the experiment

模型名称	发布时间/年份	参数量/M	前五命中率/%
EfficientNetV2-s	2022	21.5	96.878
EfficientNetV2-m	2022	54.1	97.156
EfficientNetV2-l	2022	118.5	97.788
ResNet-152	2015	60.2	94.046
SwinV2-t	2022	28.4	96.132
SwinV2-s	2022	49.7	96.816
SwinV2-b	2022	87.9	96.864
ViTb-32	2020	88.2	92.466

最终模型由单一模型与线性拟合网络组合而成(见图 7),模型训练基本流程为:

(1) 使用双能 X 射线成像仪对铜矿石进行成像,获取训练数据;

(2) 单模型载入预训练权重使其后续训练过程快速收敛;

(3) 使用铜矿图像对单个模型分别进行训练,使用验证集进行模型的调试,防止模型陷入过拟合以及局部最优解的状态;

(4) 将两个已调整至最佳权重的模型的全连接层的神经元作为线性拟合模型的输入,双模型并行运行,以动态输入的方式对线性拟合模型进行训练。训练集、验证集、测试集与基础模型所使用的试验数据相同;

(5) 使用测试集观察单个模型以及融合模型的分辨能力。

2.3 模型设置以及环境配置

试验环境设置见表 3,深度学习框架为 Pytroch2.1,Python 版本为 3.9,GPU 加速处理器为 RTX4090×2。图像为单通道图像,分类任务为二分类,模型输入设置为单通道,输出设置为 2 个单元。优化器使用 Nadam,自适应调节学习率,使模型权重可以加速收敛,迭代次数为 50 次。

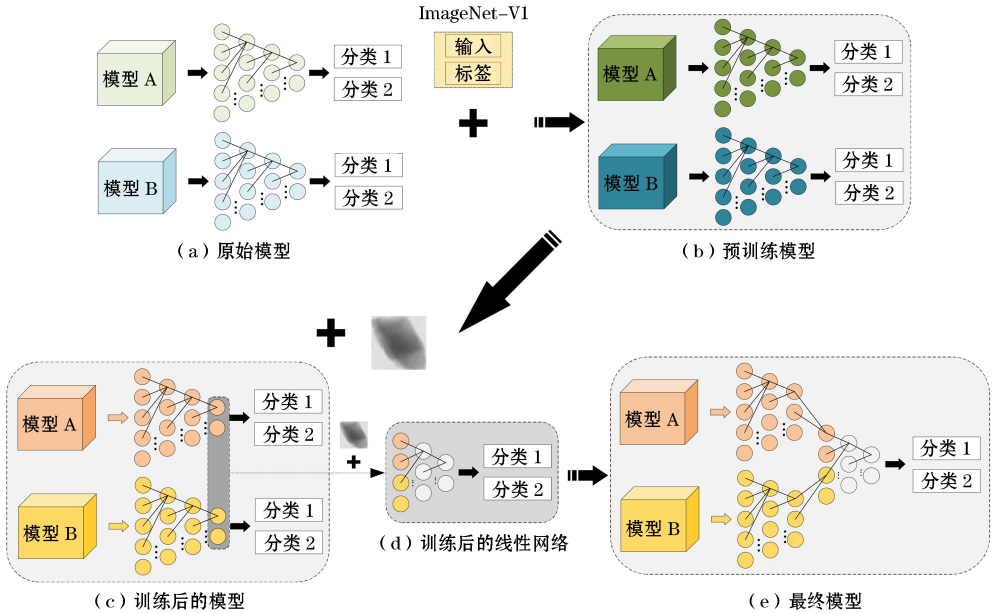


图 7　试验方法示意

Fig.7　Schematic diagrams of experimental method

表 3　模型设置及相关环境配置

Table 3　Model settings and related environment configuration

条件设置	参数名称	值
模型优化相关设置	优化器	NAdam
	学习率	1×10^{-4}
	损失函数	交叉熵损失函数
批次设置	批量大小	32
	轮次	50
运行环境设置	中央处理器	AMD EPYC 9654
	图形加速器	RTX4090×2 CUDA11.8
	深度学习框架	Pytorch2.1

3　试验结果分析

3.1　基础模型训练结果分析

在基础模型训练过程，使用验证集准确率作为性能指标对模型进行初步筛选。基础模型分为 A、B 两组，A 组为视觉 Transformer 类，B 组为卷积神经网络类，从两组当中分别挑选验证集准确率较高的模型组合，以寻求更高性能的融合模型。基础模型验证集准确率对比如图 8 所示，由图 8 可知，A 组中 SwinV2-t 表现最优，其准确率在训练过程中均高于其他 3 个模型。同理，B 组中 EfficientNetV2-s 表现最优。两者的验证集最高准确率都为 91.667%，即验证集 420 张图片，正确分辨了 385 张。

在 A 组中，相较于 SwinV2-b 与 SwinV2-s，SwinV2-t 在结构上有着更少的 Swin Transformer Block，这也意味着针对同样的数据，更深层次的网络不一定能够带来准确率的提升。相较于初代 VIT，Swin Transformer 有着更优异的性能。在 B 组当中，相较于 ResNet-152，EfficientNetV2-s 加入了类自注意力机制，性能相比更优。

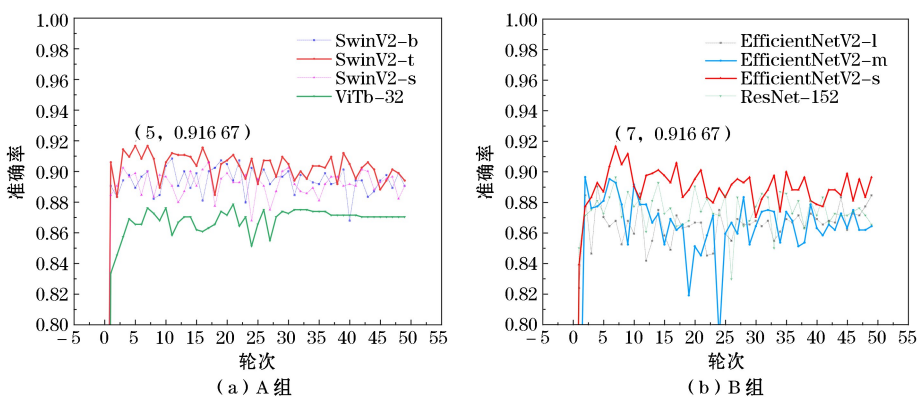


图 8　基础模型验证集准确率对比

Fig.8　Comparison of accuracy of basic model validation set

3.2 集成模型训练结果分析

冻结验证集准确率处于最高位时的 SwinV2-t 与 EfficientV2-s 的权重,使用随机权重的线性层对两者的全连接层的输出进行连接,形成集成模型。使用验证集损失和验证集准确率观察集成模型的训练程度(见图 9)。由图 9 可知,第 30 轮时验证集损失处于较低位置,其正确率最高,为 92.619%。相比于单模型,较 SwinV2-t 与 EfficientNetV2-s 高出 0.952 个百分点。

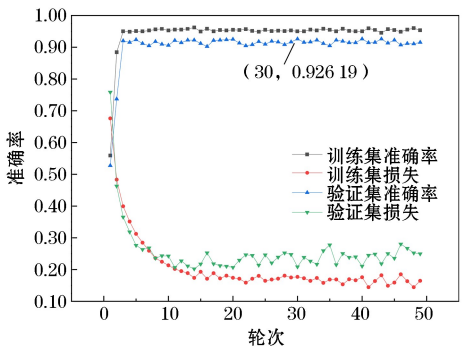


图 9 集成训练阶段指数展示

Fig.9 Integrated training phase index display

在测试阶段,试验将使用测试集对模型进行测试,相对于验证集,测试集能够反映模型最终在实际使用的性能表现。借助混淆矩阵对模型性能进行量化(见图 10),得出准确率(A)、精确率(P)、召回率(R)以及 $F1$ 指标[20](见图 11)。

$$A=\frac{LN+HP}{LN+HP+HN+LP} \tag{6}$$

$$P=\frac{HP}{HP+LP} \tag{7}$$

$$R=\frac{HP}{HP+HN} \tag{8}$$

$$F1=\frac{2PR}{P+R} \tag{9}$$

式中:LN 表示判定为低品位类的低品位样本数量;HP 表示判定为高品位类的高品位样本数量;HN 表示判定为低品位类的高品位样本;LP 表示判定为高品位类的低品位样本。其中 $F1$ 指标用于评估二分类和多标签分类问题中模型性能,为模型的整体表现提供单一的、均衡的评分。

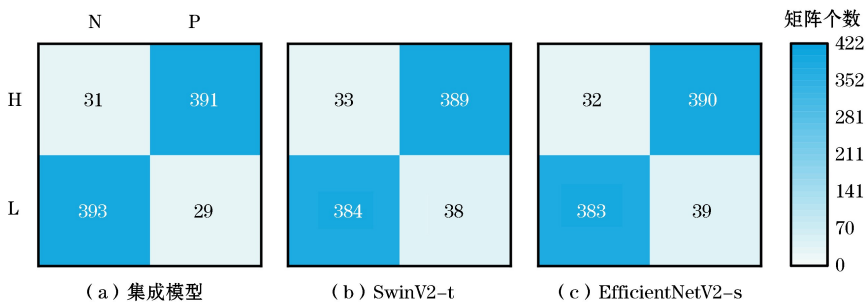


图 10 集成模型、SwinV2-t 与 EfficientNetV2-s 模型的混淆矩阵

Fig.10 Confusion matrix of Ensemble model, SwinV2-t and EfficientNetV2-s models

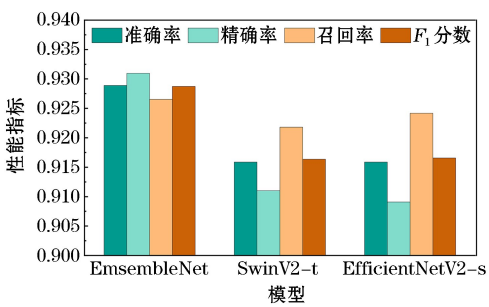


图 11 集成模型、SwinV2-t、EfficientNetV2-s 性能指标对比

Fig.11 Comparison of performance indicators of Ensemble model, SwinV2-t and EfficientNetV2-s

根据图 11 所展示的数据,集成模型的表现显示出更为均衡且优越的性能,准确率达到 92.891%,精确率达到 93.095%,召回率达到 92.654%。相较于单独的 Swinv2-t 和 EfficientNetv2-s 模型,整合后模型的准确率提升了 1.3 个百分点,并且在精确率方面分别提升了 1.9 个百分点和 2.186 个百分点。相对于 SwinV2-t 模型,其他模型各指标也有相应提高。此外,集成模型在召回率方面的提升也较为显著,相较于 SwinV2-t 和 EfficientNetV2-s 模型,集成模型的召回率分别增长了 0.474 个百分点和 0.237 个百分点。与此同时,F1 指标这一综合性评价指标亦有所进步,相较于 SwinV2-t 和 EfficientNetV2-s 模型,集成模型的 F1 指标分别提高了 1.237 个百分点和 1.217 个百分点。

结合模型性能表现以及应用环境，经过集成优化的模型展现出更强的识别效能。模型对于高品位矿石图像甄别性强，能够有效避免有价值的矿石被误弃，从而减少资源浪费。同时，对于筛选剔除废矿石、尾矿石的能力也得到了提升，从而帮助减少出现废矿、尾矿过加工的现象。理论层次上能够提升工业选矿效率。海外圣拉斐尔锡矿项目[21]引进同类型基于 X 射线透射的矿石分选系统，用于从边缘开发废弃物堆中排除废石，将大约 0.6%品位的锡矿石浓缩至 2.8%品位，提高了进入加工冶炼程序的铜含量，从而增强工序的有效性，使该企业全年总产量提升了 36%。

4 结论

在应用深度学习网络模型对矿石品位进行分选的过程中，传统的卷积神经网络往往难以妥善解决局部特征关联性和长距离依赖结构之间的平衡问题，为应对铜矿石图像特征分散、复杂的特点，提出了一种基于 Swin Transformer-EfficientNet 集成模型的铜矿石品位分类方法。该方法结合了卷积神经网络(如 EfficientNet)的局部特征提取优势与视觉 Transformer(如 Swin Transformer)捕获长距离依赖关系的能力，有效增强了模型在复杂纹理和全局理解上的性能。试验表明，集成模型在铜矿石双能 X 射线图像分类任务上表现出显著的优势。对比单一的 SwinV2-t 和 EfficientNetV2-s 模型，集成模型在各项评价指标上均有不同程度的提升。具体表现为：准确率提升了 1.3 个百分点，精确率分别比 SwinV2-t 和 EfficientNetV2-s 模型提高了 1.9 个百分点和 2.186 个百分点，召回率分别提高了 0.474 个百分点和 0.237 个百分点，而 F1 指标也分别提高了 1.237 个百分点和 1.217 个百分点。

总之，基于 Swin Transformer-EfficientNet 的集成模型的铜矿石品位分类方法充分利用了两种模型架构的优点，实现了对铜矿石双能 X 射线图像的高效分类，在数据规模方面降低了对大量标注数据的依赖，有效提高了品位分类的准确性和稳定性。

参考文献(References)：

[1] 应立娟，陈毓川，王登红，等.中国铜矿成矿规律概要[J].地质学报，2014，88(12)：2216-2226.
YING Lijuan, CHEN Yuchuan, WANG Denghong, et al. Metallogenic regularity of copper ore in China[J]. Acta Geologica Sinica, 2014, 88(12): 2216-2226.

[2] PATEL A K, CHATTERJEE S, Gorai A K. Development of a machine vision system using the support vector machine regression (SVR) algorithm for the online prediction of iron ore grades[J]. Earth Science Informatics, 2019, 12(2): 197-210.

[3] CHOROS K A, JOB A T, EDGAR M L, et al. Can hyperspectral imaging and neural network classification be used for ore grade discrimination at the point of excavation?[J]. Sensors, 2022, 22(7): 2687.

[4] 王李管，陈斯佳，贾明滔，等.基于深度学习的黑钨矿图像识别选矿方法[J].中国有色金属学报，2020，30(5)：1192-1201.
WANG Liguan, CHEN Sijia, JIA Mingtao, et al. Beneficiation method of wolframite image recognition based on deep learning[J]. The Chinese Journal of Nonferrous Metals, 2020, 30(5): 1192-1201.

[5] 周飞燕，金林鹏，董军.卷积神经网络研究综述[J].计算机学报，2017，40(6)：1229-1251.
ZHOU Feiyan, JIN Linpeng, DONG Jun. Review of convolutional neural network[J]. Chinese Journal of Computers, 2017, 40(6): 1229-1251.

[6] 孙振涵.基于单头自注意力轻量级视觉 Transformer 骨干网的研究[D].南京：南京邮电大学，2023.
SUN Zhenhan. Research on lightweight transformer backbone based on single head self attention[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2023.

[7] 袁媛，陈明惠，柯舒婷，等.基于集成卷积神经网络和 Vit 的眼底图像分类研究[J].中国激光，2022，49(20)：108-116.
YUAN Yuan, CHEN Minghui, KE Shuting, et al. Fundus image classificaiton research based on ensemble convolution neural network and vision transformer[J]. Chinese Journal of Lasers, 2022, 49(20): 108-116.

[8] JIANG Z C, DONG Z X, WANG L Y, et al. Method for diagnosis of acute lymphoblastic leukemia based on ViT-CNN ensemble model[J]. Computational Intelligence and Neuroscience, 2021: 7529893.

[9] PAN S J, YANG Q. A Survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.

[10] HURLBUT, CORNELIUS S. Manual of Mineralogy[M]. New York: Wiley, 1977.

[11] 中华人民共和国国土资源部.铜、铅、锌、银、镍、钼矿地质勘查规范：DZ/T 0214—2002[S].北京：中国标准出版社，2002.
Ministry of Land and Resources of the People's Republic of China. Specifications for copper, lead, zinc, silver, nickel and molybdenum mineral exploration: DZ/T 0214—2002[S]. Beijing: China Standards Press, 2002.

[12] REBUFFEL V, DINTEN J M. Dual-energy X-ray imaging: benefits and limits[J]. Insight-Non-Destructive Testing and Condition Monitoring,2007,49(10):589-594.

[13] TAN M X, LE V Q. EfficientNet: rethinking model scaling for convolutional neural networks[C]//36th International Conference on Machine Learning: ICML 2019, Long Beach, California, USA, 9-15 June 2019, Part 16 of 19.:Curran Associates, Inc., 2019:10691-10700.

[14] MINGXING TAN, QUOC V. LE. EfficientNetV2: Smaller Models and Faster Training[C]//International Conference on Machine Learning: ICML 2021, Online,18-24 July 2021, Part 13 of 16.:Curran Associates, Inc.,2022:10086-10096.

[15] SONG Huaxiang. FST-EfficientNetV2: exceptional image classification for remote sensing[J]. Computer Systems Science and Engineering,2023,46(3):3959-3978.

[16] LYU Ningning, YUAN Min, XIE Yufei, et al. Non-local sparse attention based swin transformer V2 for image super-resolution[J].Signal Processing,2024,222:109542.

[17] GANAIE M A, HU M H, MALIK A K, et al. Ensemble deep learning: a review[J]. Engineering Applications of Artificial Intelligence,2022,115:105151.

[18] JIANG Z C, DONG Z X, WANG L Y, et al. Method for diagnosis of acute lymphoblastic leukemia based on ViT-CNN ensemble model[J]. Computational Intelligence and Neuroscience,2021:7529893.

[19] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision,2015,115:211-252.

[20] HOSSIN M, SULAIMAN M N. A review on evaluation metrics for data classification evaluations[J]. International Journal of Data Mining & Knowledge Management Process, 2015,5(2):1-11.

[21] ROBBEN C, CONDORI P, PINTO A, et al. X-ray-transmission based ore sorting at the San Rafael tin mine[J]. Minerals Engineering,2020,145:105870.