doi: 10.13413/j.cnki.jdxblxb.2023446

基于混合专家模型的岩石薄片图像分类

周程阳[1], 刘 伟[2], 吴天润[1], 李 骜[1], 韩霄松[1]
(1. 吉林大学 软件学院, 长春 130012; 2. 中国石油集团工程技术研究院, 北京 102206)

摘要: 以常见的 5 种岩石薄片作为研究对象构建数据集, 提出一种新的基于混合专家模型的岩石薄片图像分类模型. 该模型从薄片图像中学习到每种岩石图像的特征, 并对其进行分类. 首先, 使用多个基于卷积神经网络(CNN)和 Transformer 的图像分类模型(ResNet50, MobileNetV3, InceptionV3, DeiT 等)对数据进行训练; 其次, 选取效果较好的模型, 通过构建混合专家模型, 得到最终的预测结果, 其岩性识别准确率(ACC)和 AUC 在验证集上达到 85.33%和 96.69%, 在测试集上达到 87.16%和 96.75%; 最后, 通过混合专家模型结合多个模型, 综合各模型的优势, 平衡各模型间的贡献, 提高分类结果的准确性和鲁棒性, 使得到的分类结果更可靠、稳定.

关键词: 岩石薄片分类; 混合专家模型; 图像分类
中图分类号: TP391 **文献标志码:** A **文章编号:** 1671-5489(2024)04-0905-10

Classification of Rock Thin Section Images Based on Mixture of Expert Model

ZHOU Chengyang[1], LIU Wei[2], WU Tianrun[1], LI Ao[1], HAN Xiaosong[1]
(1. *College of Software, Jilin University, Changchun* 130012, *China*;
2. *CNPC Engineering Technology R&D Company Limited, Beijing* 102206, *China*)

Abstract: We proposed a new classification of rock thin section images based on mixture of expert model by using five common rock thin sections as the research object to construct a dataset. The model learned the characteristics of each rock image from the thin section images and classified them. Firstly, multiple image classification models based on convolutional neural network (CNN) and Transformer (such as ResNet50, MobileNetV3, InceptionV3, DeiT, etc.) were used to train the data. Secondly, models with better performance were selected, a mixture of experts model was built to obtain the final prediction result. The ACC and AUC of lithology recognition reached 85.33% and 96.69% on the validation set and 87.16% and 96.75% on the test set. Finally, by combining a mixture of experts model with multiple models, combining advantage of each model, balancing their contributions between each model, we improved the accuracy and robustness of classification results, making the obtained classification results more reliable and stable.

Keywords: classification of rock thin section; mixture of expert model; image classification

收稿日期: 2023-10-19.

第一作者简介: 周程阳(2000—), 男, 汉族, 硕士研究生, 从事智能算法的研究, E-mail: cyzhou23@mails.jlu.edu.cn. **通信作者简介:** 韩霄松(1984—), 男, 汉族, 博士, 副教授, 从事智能钻井的研究, E-mail: hanxiaosong@jlu.edu.cn.

基金项目: 国家重大科技专项基金(批准号: 2011ZX05044; 2011ZX05001)、国家自然科学基金(批准号: 62372494)、吉林省科技发展计划项目(批准号: 20220201145GX; 20220601112FG)和大学生创新创业训练计划项目(批准号: 202310183221).

岩石是地壳的主要组成部分，是由矿物和天然玻璃组成的具有固定外形的稳定集合体. 目前，自然界中已发现超过3 000种岩石，如砂岩、灰岩、板岩、花岗岩等. 岩性识别是地质学和资源勘探开发任务中的一项基础而重要的工作，它涉及对岩石的类型、成分、结构和成因特征进行判断和分类. 岩石薄片图像是岩性识别过程中的重要数据来源. 传统岩性识别过程依赖人工观察确定岩石的类别等信息，该过程耗时耗力且依赖于领域知识，受个人主观因素影响较大，识别准确率较低. 随着计算机科学技术的不断发展，岩石薄片的图像识别领域中引入了深度学习的方法.

目前，对岩石薄片分类的研究已有许多成果. 袁颖等[1]用主成分分析(PCA)方法提取评价指标的主成分，并通过遗传算法优化支持向量机(SVM)，对火成岩进行了分类训练；马陇飞等[2]用基于GBDT(gradient boosting decision tree)算法的识别模型识别多种泥岩和砂岩，准确率达92%；贺金鑫等[3]结合岩石光谱特征和4种机器学习分类模型进行投票，对辽宁省兴城地区的经典岩石样本进行分类，准确率达99.17%；Marmo等[4]通过图像处理技术和人工神经网络(ANN)对碳酸盐岩薄片进行识别，识别准确率为93.5%；Singh等[5]采用与文献[4]相同的方法，提取27维数值作为神经网络的输入，对玄武岩薄片图像的分类精度达92.22%；Chatterjee[6]提出了一种基于SVM的算法，从原始的189个特征中选择40个特征作为模型输入，对6种石灰岩类型进行识别，识别准确率达96.2%；Khorram等[7]提出了一个石灰岩分类模型，将SVM和Bayes技术用于分类，实现了在不同采矿阶段对岩性进行分类；Mlynarczuk等[8]利用偏振显微镜获取了9种岩石样本的薄片数字图像，使用4种模式识别方法自动识别岩石样本.

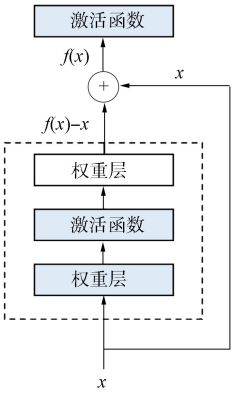
上述方法大部分将机器学习方法用在岩石类型分类中，显示出一定优势. 但这些方法严重依赖于研究人员提取的数字特征质量，直接决定了模型的最终性能. 近年来，随着深度学习算法的发展，许多研究人员将其应用于岩石薄片识别并取得了重大突破. 谭永健等[9]改进了InceptionV3网络中卷积操作，引入残差连接和迁移学习的思想，对10类岩石样本进行分类，准确率达86%；许振浩等[10]结合监督目标检测网络以及ResNet-101网络对岩石图像进行迁移学习，准确率达90.21%；程国建等[11]利用轻量卷积神经网络SqueezeNet训练岩石薄片图像分类，在验证集上的准确率达90.88%；张野等[12]对InceptionV3模型进行迁移学习，在岩石岩性分类识别上准确率超过90%；Li等[13]使用改进的TradaBoost算法对不同区域采集的微观砂岩图像进行识别；Polat等[14]利用两个神经网络对6种类型的火山岩进行自动分类，并对4种不同优化器的分类效果进行了评价；Dos Anjos等[15]提出了4种卷积神经网络(CNN)模型，利用微观Thin数学剖面图像识别了3种巴西盐下湖相碳酸盐岩.

目前，关于岩石薄片识别的研究已取得了很多成果. 混合专家(mixture of experts，MoE)通过将多个专家模型组合形成一个整体模型，以利用每个专家模型的优势. 每个专家模型可以专注于解决特定的子问题，而整体模型则能在复杂任务中获得更好的性能. 本文通过搭建多个基于CNN和Transformer的分类模型，如ResNet50，MobileNetV3，InceptionV3，DeiT，使用这些模型在较大数据集上的预训练参数，再对每个岩石薄片图像进行特征学习，并用于分类任务. 在单模型分类的基础上，引入MoE的方法将各模型视为不同专家，动态决定哪个专家模型应该被激活以生成最佳预测，进一步提高岩石薄片图像分类模型的性能.

1 深度学习方法

1.1 ResNet-50

ResNet-50是一种深度残差神经网络结构[16]，在图像处理任务中常用于特征提取，旨在解决传统神经网络中随网络层数增加，由于梯度消失或梯度爆炸导致的网络性能饱和或退化问题. 通过引入残差模块的概念，可有效解决上述问题. 在残差模块中，每个输入有两个主要路径：主路径和残差路径. 主路径通过若干卷积层学习特征，而残差路径则直接连接主路径的输入和输出，完成跳跃连接，将输入直接添加到主路径的输出中，使信息可直接传到更深层，促进信号在向前和向后路径的传播. 图1(A)为一个残差模块，其中主路径由两个卷积层组成；图1(B)为Resnet-50的网络结构，其中有50个卷积层.



(A) ResNet残差模块示例

stage	output	ResNet-50
conv1	112×112	7×7, 64, stride 2
conv2	56×56	3×3最大池化, stride 2 $\begin{pmatrix}1\times1,\ 64\\3\times3,\ 64\\1\times1,\ 256\end{pmatrix}\times3$
conv3	28×28	$\begin{pmatrix}1\times1,\ 128\\3\times3,\ 128\\1\times1,\ 512\end{pmatrix}\times4$
conv4	14×14	$\begin{pmatrix}1\times1,\ 256\\3\times3,\ 256\\1\times1,\ 1\,024\end{pmatrix}\times6$
conv5	7×7	$\begin{pmatrix}1\times1,\ 512\\3\times3,\ 512\\1\times1,\ 2\,048\end{pmatrix}\times3$
	1×1	全局池化 1 000-d fc, Softmax
#params.		25.5×10^6
FLOPs		4.1×10^9

(B) ResNet-50网络结构

图 1　ResNet 残差模块示例（A）及 ResNet-50 网络结构（B）

Fig. 1　Example of ResNet residual module (A) and ResNet-50 network structure (B)

1.2　MobileNetV3

MobileNet[17]是一种专为移动设备和嵌入式设备设计的轻量级卷积神经网络模型，其目标是在保持较高准确率的同时，显著减少网络参数量和计算量. MobileNet 在设计中提出了一种深度可分离卷积代替传统卷积的方案[18]，如图 2 所示.

深度卷积阶段在每个输入通道上分别应用一个小型的卷积核，可极大减少计算量. 逐点卷积阶段使用 1×1 的卷积核对深度卷积的结果进行线性组合，从而实现通道间的交互和特征融合. 这种深度可分离卷积的设计使 MobileNet 在减少参数量的同时，仍能保持较高的感受野和表达能力.

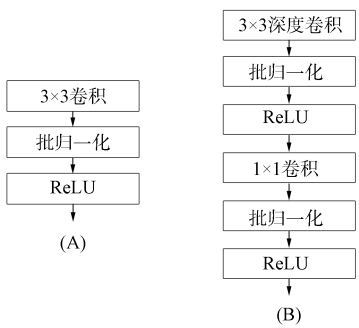


图 2　MobileNet 中传统卷积（A）和深度可分离卷积（B）

Fig. 2　Traditional convolution (A) and depth-separable convolution (B) in MobileNet

MobileNetV3 是对 MobileNet 的改进版本，网络结构列于表 1，参数量为 5.4 M，其中引入了几个关键的改进：h-swish 激活函数、SE 通道注意力机制和网络架构搜索.

表 1　MobileNetV3 网络结构

Table 1　Network structure of MobileNetV3

输入	操作	膨胀系数	#out	SE	NL	s
$224^2\times3$	conv2d	—	16	—	HS	2
$112^2\times16$	bneck，3×3	16	16	—	RE	1
$112^2\times16$	bneck，3×3	64	24	—	RE	2
$56^2\times24$	bneck，3×3	72	24	—	RE	1
$56^2\times24$	bneck，5×5	72	40	√	RE	2
$28^2\times40$	bneck，5×5	120	40	√	RE	1
$28^2\times40$	bneck，5×5	120	40	√	RE	1
$28^2\times40$	bneck，3×3	240	80	—	HS	2
$14^2\times80$	bneck，3×3	200	80	—	HS	1
$14^2\times80$	bneck，3×3	184	80	—	HS	1
$14^2\times80$	bneck，3×3	184	80	—	HS	1
$14^2\times80$	bneck，3×3	480	112	√	HS	1

续表 1

Continued to table 1

输入	操作	膨胀系数	#out	SE	NL	s
$14^2\times112$	bneck，3×3	672	112	√	HS	1
$14^2\times112$	bneck，5×5	672	160	√	HS	2
$7^2\times160$	bneck，5×5	960	160	√	HS	1
$7^2\times160$	bneck，5×5	960	160	√	HS	1
$7^2\times160$	conv2d，1×1	—	960	—	HS	1
$7^2\times960$	pool，7×7	—	—	—	—	1
$1^2\times960$	conv2d 1×1，NBN	—	1 280	—	HS	1
$1^2\times1\ 280$	conv2d 1×1，NBN	—	k	—	—	1

h-swish 是一种针对 MobileNetV3 特别设计的激活函数，它结合了线性和非线性特性，具有高计算效率和模型可训练性. h-swish 通过引入剪切的 ReLU 函数，提供了良好的非线性变换和梯度传播，增强了网络的表达能力和训练效果. SE 通道注意力机制先通过全局平均池化层得到一个通道数较小的特征向量，再用全连接层和激活函数得到一个注意力向量，最后将注意力向量和原始特征图相乘，得到一个经过注意力调整的特征图. 网络架构搜索利用强化学习和进化算法等技术，搜索算法可以在给定的计算约束下，自动搜索最优的网络结构和超参数，其提供了一种高效的方式设计出性能更好的模型.

1.3 InceptionV3

Inception[19]是一种深度卷积神经网络，又称 GoogLeNet，其主要特点是引入了 Inception 模块. Inception 模块包含了 1×1,3×3 和 5×5 的卷积层以及 3×3 的最大池化层. 这些操作并行进行，并在最后通过连接进行组合，如图 3 所示. 为减少计算量，Inception 在 3×3 和 5×5 的卷积操作前使用 1×1 的卷积作为瓶颈层进行降维，有效减少了网络的参数数量和计算复杂度. Inception 模块的这种设计使网络在每个层级都能自动学习并适应不同尺寸的特征. InceptionV3[20]对于 Inception 模块设计上更复杂，引入了更多的卷积和池化操作，在更小的卷积操作中引入了非线性激活，以增加模型的表达能力；并将大卷积核拆分成多个小卷积核，如将 5×5 的卷积核拆解成 3×3 的卷积核，在保持相同感受野的同时缩小模型数量，如图 3(B)所示. 同时，InceptionV3 使用了平滑标签技术.

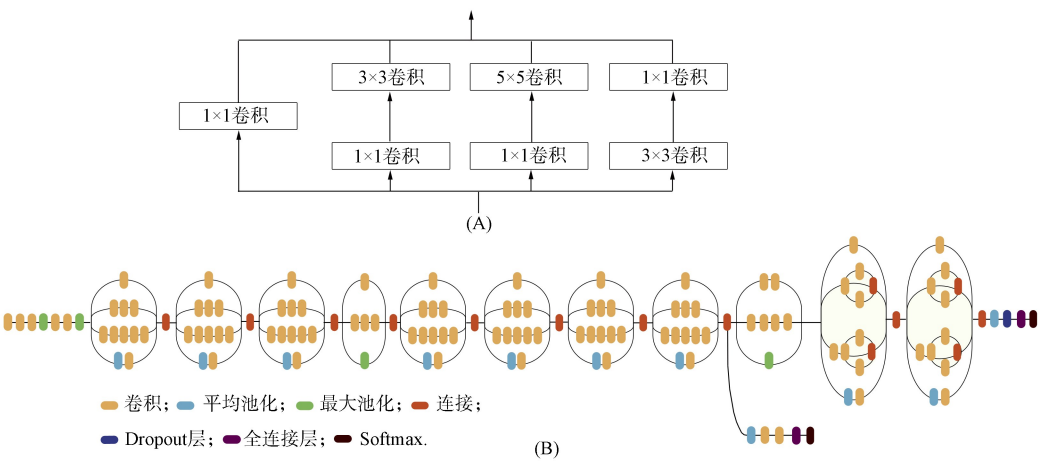


图 3 GoogLeNet 中引入瓶颈层的 Inception 块(A)和 InceptionV3 网络结构(B)

Fig. 3 Inception block that introduces bottleneck layer in GoogLeNet (A) and InceptionV3 network structure (B)

1.4 DeiT

DeiT(data-efficient image transformer)[21]是一种基于 Transformer 的视觉图像分类模型，DeiT 给出了一种新的训练方式，称为知识蒸馏(knowledge distillation)，使模型在数据较少的情况下也有很好的性能. 知识蒸馏是一种模型训练技术，旨在通过传递一个大型教师模型的知识训练一个小型学生模型. 该方法的目标是使学生模型能获得与教师模型相似的性能，同时减少学生模型的复杂性和计算

成本，通过传递教师模型的知识，可以在小型模型上实现接近教师模型性能的效果．DeiT 不仅要使学生模型在分类任务上的输出接近教师模型的输出，而且还要使学生模型在注意力分布上接近教师模型．通过模仿教师模型的注意力分布，学生模型可学习到如何聚焦于图像中的重要部分．DeiT 模型结构如图 4 所示．

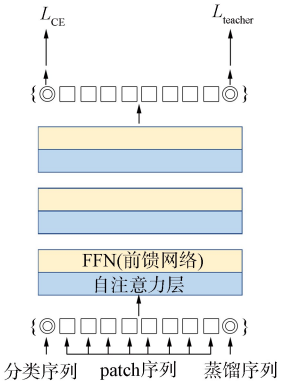


图 4　DeiT 模型结构

Fig. 4　DeiT model structure

DeiT 还引入了一种新的位置编码方式，称为 2D 相对位置编码．这种位置编码方式考虑了每个 patch 与其他所有 patches 在 2D 空间中的相对位置，使模型能更好地理解图像的空间结构．此外，为进一步提高模型的性能，DeiT 还引入了 token mixing 策略．在训练过程中，DeiT 会随机选择一些图像区域，并对它们进行局部调整，从而使模型更好地处理输入中的局部变化．这种随机性的引入可增加模型的鲁棒性和泛化能力．

2　岩石薄片图像分类模型

岩石薄片图像分类旨在从众多的岩石薄片图像中学习到每种岩石图像的特征，并对其进行分类．本文引入混合专家模型的思想，首先使用多个基于 CNN 和 Transformer 的图像分类模型对数据进行训练，通过将效果较好的多个单模型进行组合，综合各模型的预测结果，从而得到更准确和具有鲁棒性的分类结果，模型结构如图 5 所示．

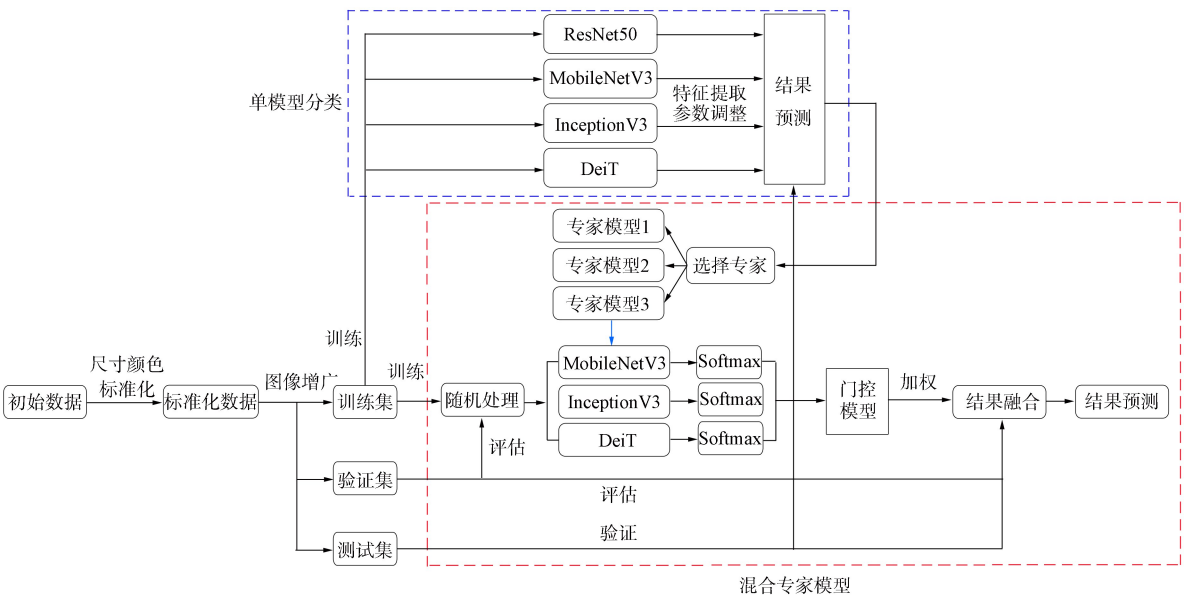


图 5　岩石薄片图像分类模型结构

Fig. 5　Structure of rock thin section image classification model

图 5 中蓝线框内使用多个基于 CNN 和 Transformer 的图像分类模型对数据进行训练，用这些模型在较大数据集上的预训练参数，再对每个岩石薄片图像进行特征学习，并用于分类任务．在验证集上评估各单模型的性能，通过对结果预测的准确性比较各模型，选取性能优秀的模型后续使用．红线框内选择性能最佳的模型，先用每个模型对图像样本进行分类预测，再通过 Softmax 输出各类别的概率结果．然后通过混合专家模型为选中的每个专家模型分配一个权重，通过模型在验证集上的预测结果得到各专家模型在分类任务中的贡献进行结果融合，寻找混合专家模型中各专家模型的最优权重集合，最终进行结果预测．

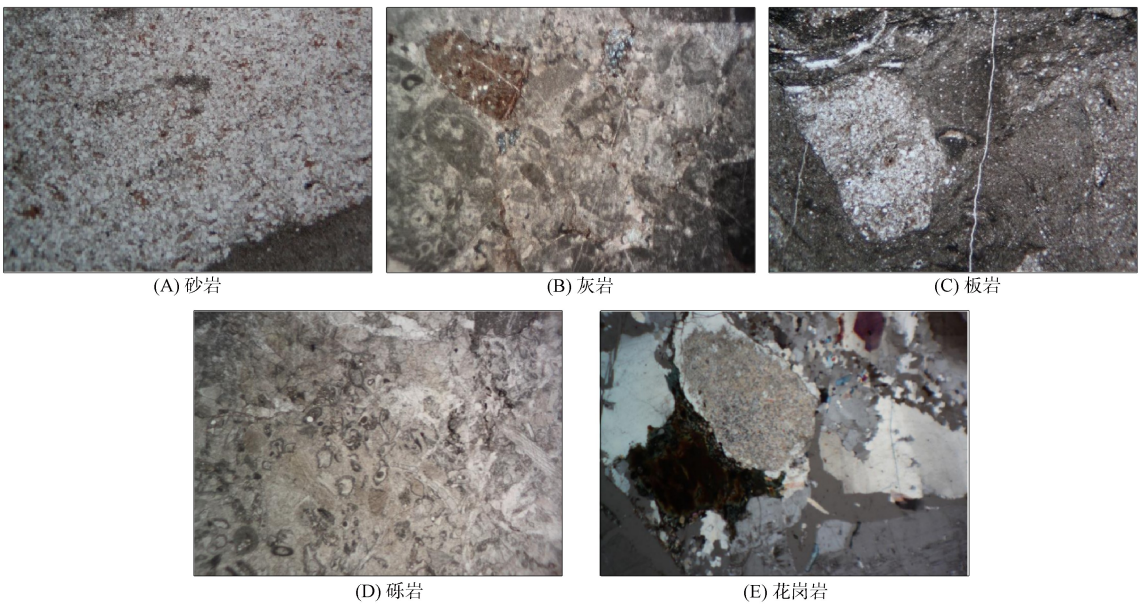
2.1 岩石薄片图像收集与处理

本文岩石薄片图像数据来自中国实物地质资料信息网的实物地质资料图像数据库(https://www.cgsi.cn/). 首先对岩石薄片图像进行分类和统计. 在数据库中选择3个较大的岩石薄片图像产地，共得到1 539组岩石薄片图像样本，其中包含11 370张岩石薄片图像. 对这些图像按岩石的名称进行分类，统计每个类别中的图像数量，并按数量进行排序，结果列于表2. 岩石薄片图像如图6所示.

表2 岩石薄片图像类别对应数量

Table 2 Corresponding number of rock thin section image categories

类别	砂岩	灰岩	板岩	砾岩	花岗岩
数量	4 737	3 913	1 306	1 001	773



(A) 砂岩 (B) 灰岩 (C) 板岩

(D) 砾岩 (E) 花岗岩

图6 岩石薄片图像示例

Fig. 6 Examples of rock thin section images

得到图像后对图像进行尺寸调整，使其大小一致，并进行颜色标准化操作，消除由于不同图像采集设备、光照条件和色彩校准等因素导致的颜色偏差，以提高模型对图像特征的识别能力. 为评估分类模型的性能，将数据集按14∶3∶3的比例划分为训练集、验证集和测试集. 为增加模型的泛化能力，对训练集和验证集中的每张图片进行随机裁剪、旋转、翻转及颜色抖动等处理，以生成更多样的训练样本，扩展数据集的多样性，并使模型对图像的变化有更好的适应能力.

2.2 单模型岩石薄片图像分类

搭建多个基于CNN和Transformer的分类模型，如ResNet50,MobileNetV3,InceptionV3,DeiT. 使用这些模型在较大数据集上的预训练参数，再对每个岩石薄片图像进行特征学习，并用于分类任务. 对每个模型，采用端到端的训练策略. 首先，将岩石薄片图像样本输入到模型中，对图像进行特征提取和表示学习；其次，通过反向传播算法和优化方法，对模型的参数进行调整和更新，使模型能更准确地预测岩石薄片图像的类别；最后，采用多个模型，并对不同模型采用不同的网络结构和参数设置，以考察不同模型的特征表示能力和学习能力.

为评估模型性能，使用交叉熵作为损失函数. 在多类别分类任务中，交叉熵损失可衡量模型的预测结果与真实标签之间的差异程度. 通过最小化损失函数，模型逐渐调整参数，提高对岩石薄片图像的分类准确性. 交叉熵损失函数的计算公式为

$$L^t = \sum_{i=1}^{N} y_i^t \log(p_i^t), \tag{1}$$

其中y表示图像的真实标签，N表示图像的总数量，L为损失函数的值.

通过训练多个单模型，并计算每个模型对图像样本的预测结果，最终获得多个独立的分类模型.

这些模型具有不同的架构和参数设置，能对岩石薄片图像的不同特征进行学习和表示.

2.3　混合专家模型岩石薄片图像分类

在单模型分类的基础上，引入混合专家模型进一步提高岩石薄片图像分类的性能. 通过将效果较好的多个单模型进行组合，综合各模型的预测结果，从而得到更准确和具有鲁棒性的分类结果. 混合专家模型的过程主要由一组专家模型和一个门控模型组成. 基本理念是将输入数据根据任务类型分割成多个区域，并将每个区域的数据分配一个或多个专家模型. 每个专家模型可专注于处理输入这部分数据，从而提高模型的整体性能.

门控模型主要由稀疏门网络组成，它接收单个数据元素作为输入，然后输出一个权重，这些权重用于表示不同专家模型在处理输入数据时做出的贡献，一般采用 Softmax 门控函数通过专家对概率分布进行建模. 例如，如果模型有 3 个专家，输出的概率可能为 0.6，0.3，0.1，则表示第一个专家对处理此数据的贡献为 60%，第二个专家为 30%，第三个专家为 10%.

混合专家模型还包含一组专家模型，在训练过程中，门控模型将输入的数据分配到不同的专家模型中进行处理，不同的专家被分配到处理不同种类的输入数据；在推理过程中，被门控选择的专家会针对输入的数据产生相应的输出. 这些输出最后会与每个专家模型处理该特征的能力分配的权重进行加权组合，形成最终的预测结果. 混合专家模型在训练过程中通过“因材施教”的思想构建门控模型，进而在推理过程中实现各专家模型的优势. 在混合专家模型中，处理输入数据时只有少数专家模型被激活或使用，而大部分专家模型处于未被激活状态，这种状态即为稀疏. 稀疏性是混合专家模型的主要优点，也是提升模型训练和推理过程效率的关键.

图 7 为混合专家模型的原理，一个门控模型用于分配每个专家模型的输出权重. 对一个输入样本 c，第 i 个专家模型的输出为 o_c^i，真实数据为 d^c，则其损失函数可表示为

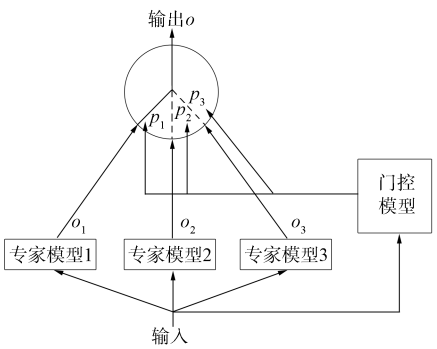


图 7　MoE 原理

Fig. 7　Principle of MoE

$$E^c = \sum_i p_i^c (d^c - o_i^c)^2, \tag{2}$$

其中 p_i^c 表示门控模型分配给每个专家模型的权重，将 p_i^c 放于括号外部使每个专家模型都能单独计算损失函数，鼓励不同的专家模型进行竞争.

每个专家模型都会对图像样本进行分类预测，并输出各类别的概率结果. 通过为每个专家模型分配的权重表示该模型对最终分类结果的重要性. 通过验证集将门控网络计算的权重进行加权聚合，每个专家模型的输出乘以相应的权重，并将这些加权的输出求和，从而得到最终模型输出的预测结果. 混合专家模型得到的结果为

$$\hat{y} = \frac{1}{N}\sum_{t=1}^{N} p_t y_t, \qquad \sum_{t=1}^{N} p_t = 1, \tag{3}$$

其中 p 表示每个专家模型对应的权重，y 表示每个模型预测各类别概率的结果，$\hat{y}$ 表示多个结果的加权和，N 表示进行加权的模型数量. 要求所有模型的权重之和等于 1，并使最终概率和仍等于 1.

在混合专家模型学习的过程中，通过合理设置权重平衡各模型之间的贡献. 性能更好的模型会被分配更高的权重，而性能较差的模型则会被分配较低的权重. 通过调整权重，可有效结合各模型的优势，提高模型整体的分类性能和鲁棒性. 通过引入混合专家模型，可充分利用多个单模型的优势，其加权组合机制使模型能在不同的输入下自适应地选择哪个专家模型能更有利于当前输入，从而弥补单个模型的不足，提高岩石薄片图像分类的准确性和可靠性.

3 实验与结果分析

3.1 实验环境与参数设计

实验以 Python3.8.10 作为编程语言，以 PyTorch 作为学习框架，版本为 1.11.0，GPU 选择 RTX 3090，显存为 24 GB，CPU 是 Intel Xeon Gold 6330，内存为 25 GB. 采用分批次训练方法，批次大小为 64，所有训练图像在模型中完成一次计算即为一次迭代. 网络初始学习率为 0.001，衰减步长为 10，衰减因子为 0.9.

3.2 评价指标

实验采用准确率(ACC)和 AUC 作为评价指标. ACC 是衡量模型在所有预测中正确分类的比例，假设 TP,TN,FP,FN 分别为真正例、真反例、假正例、假反例的数量，则 ACC 的计算公式为

$$\mathrm{ACC}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}. \tag{4}$$

AUC 是基于 ROC 曲线(receiver operating characteristic curve)计算得出的一个指标，用于评估模型在不同阈值下的分类性能. 二分类的 AUC 的计算方式是通过以假正例率(FPR)为横坐标、真正例率(TPR)为纵坐标绘制出的 ROC 曲线下的面积得出. 其中，FPR 和 TPR 的计算公式分别为

$$\mathrm{FPR}=\frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}}, \tag{5}$$

$$\mathrm{TPR}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}. \tag{6}$$

多分类的 AUC 通过计算多个种类的二分类 AUC 的值平均得出，计算公式为

$$\mathrm{AUC}=\mathrm{mean}(\mathrm{AUC}_1,\mathrm{AUC}_2,\cdots,\mathrm{AUC}_n). \tag{7}$$

3.3 单模型收敛性实验

选择 ResNet50,MobileNetV3,InceptionV3,DeiT 4 个模型进行训练，4 个模型在 200 次迭代中的损失值、ACC 值、AUC 值变化曲线如图 8 所示. 由图 8 可见，DeiT 的效果最好，在随机变换后验证集上的准确率达 80%以上，ResNet-50 效果最差.

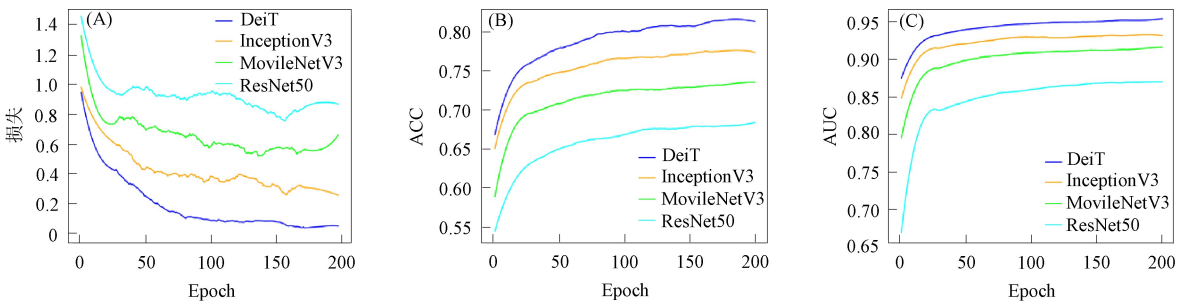


图 8 不同模型训练的损失值、ACC、AUC 变化曲线

Fig. 8 Loss values, ACC and AUC change curves of different model training

3.4 混合专家模型权重实验

构建混合专家模型，观察使用该方法后模型对岩石薄片的分类预测准确率对比单模型是否有提升. 由上述实验结果可知，ResNet50 模型的效果比其他模型相差较多，并且与其他模型结构的重复度较高，故选择剩余的 MobileNetV3,InceptionV3 和 DeiT 作为专家模型构建混合专家模型.

由于样本分布不均匀，最大类别与最小类别数据量约相差 6 倍，故选择 AUC 最大值作为专家模型的选取，选择专家模型后，分别对经过 Softmax 的结果进行加权求和，在验证集采用网格搜索的方式寻找混合专家模型中各专家模型最佳的权重集合，其中 Softmax 加权网格搜索的效果如图 9 所示. 由图 9 可见，当 DeiT,InceptionV3,MobileNetV3 的权重分别为 0.67,0.27,0.06 时模型效果最好. 对 3 个专家模型，构建混合专家模型在验证集上的最佳模型，混合专家模型和单模型在验证集和测试集上的测试效果列于表 3.

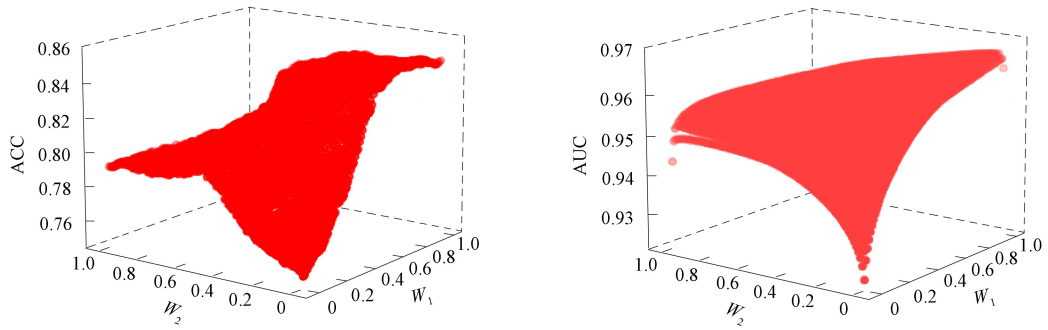


图 9　专家模型通过 Softmax 进行加权的网格搜索效果

Fig. 9　Weighted grid search effect of expert model through Softmax

表 3　模型验证集 ACC，AUC 最佳效果比较

Table 3　Comparison of the best effect of model validation sets ACC and AUC

模型	验证集		测试集	
	ACC	AUC	ACC	AUC
ResNet50	0.690 2	0.887 4	0.701 7	0.898 5
MobileNetV3	0.751 0	0.924 5	0.751 7	0.925 8
InceptionV3	0.791 4	0.943 3	0.817 6	0.946 0
DeiT	0.847 1	0.963 8	0.866 5	0.964 8
混合专家	0.853 3	0.966 9	0.871 6	0.967 5

由表 3 可见，通过最佳权重集合构建的混合专家模型在测试集和验证集上的性能均较好，对比最优秀的单一分类模型 DeiT，在验证集和测试集上的准确率分别高出 0.006 2 和 0.005 1. 实验结果表明，在验证集上得到的模型在测试集上性能也较好，可有效增加模型预测的准确率. 对最终的混合专家模型，其对测试集的图像预测热力图如图 10 所示，其中从左到右的类别分别为板岩、灰岩、砂岩、砾岩、花岗岩. 由图 10 可见，模型对各岩石类别的预测效果相对较好，但也存在一些问题. 当预测标签为砂岩一列时，模型可能会将其他类别的图像错误地预测为砂岩，这可能是因为砂岩类别在岩石分类中属于一个广泛的类别，包含了多种不同的样本，即使部分岩石的名称被定义为其他岩石，可能最终类别仍在砂岩类中，导致模型产生混淆.

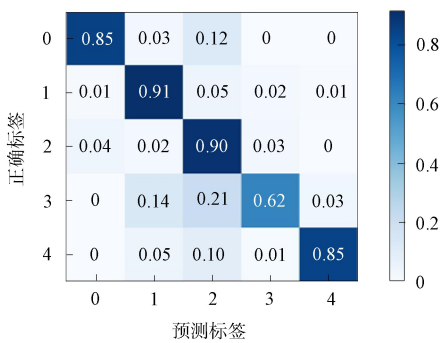


图 10　混合专家模型在测试集上的预测热力图

Fig. 10　Prediction heat map of mixture of expert model on test set

综上所述，本文基于多个 CNN 和 Transformer 的单模型分类器，通过端到端训练对岩石薄片图像进行特征提取和学习，以预测岩石的类别. 通过引入构建混合专家模型的思想合理设置权重，平衡各模型之间的贡献，提高了模型整体的分类性能和鲁棒性. 使用中国实物地质资料信息网的岩石薄片图像数据库构建数据集进行实验的结果表明，在单个模型的预测准确率达 86.65%，在 MoE 模型中达 87.16%.

参　考　文　献

[1] 袁颖，李绍康，周爱红. 基于 PCA-GA-SVM 的火成岩分类方法研究 [J]. 数学的实践与认识，2017，47(12)：121-128.（YUAN Y，LI S K，ZHOU A H. Research on Igneous Rock Classification Method Based on PCA-GA-SVM [J]. Mathematics Practice and Understanding，2017，47(12)：121-128.）

[2] 马陇飞，萧汉敏，陶敬伟，等. 基于梯度提升决策树算法的岩性智能分类方法 [J]. 油气地质与采收率，2022，29(1)：21-29.（MA L F，XIAO H M，TAO J W，et al. Lithology Intelligent Classification Method Based on Gradient Boosting Decision Tree Algorithm [J]. Oil and Gas Geology and Recovery Factors，2022，29(1)：21-29.）

[3] 贺金鑫，任小玉，陈圣波，等. 融合学习模型的岩石光谱特征自动分类 [J]. 光谱学与光谱分析，2021，41(1)：141-144.（HE J X，REN X Y，CHEN S B，et al. Automatic Classification of Rock Spectral Features Based on Fusion Learning Model [J]. Spectroscopy and Spectral Analysis，2021，41(1)：141-144.）

[4] MARMO R，AMODIO S，TAGLIAFERRI R，et al. Textural Identification of Carbonate Rocks by Image Processing and Neural Network：Methodology Proposal and Examples [J]. Computers & Geosciences，2005，31(5)：649-659.

[5] SINGH N，SINGH T N，TIWARY A，et al. Textural Identification of Basaltic Rock Mass Using Image Processing and Neural Network [J]. Computational Geosciences，2010，14(2)：301-310.

[6] CHATTERJEE S. Vision-Based Rock-Type Classification of Limestone Using Multi-class Support Vector Machine [J]. Applied Intelligence，2013，39(1)：14-27.

[7] KHORRAM F，MORSHEDY A H，MEMARIAN H，et al. Lithological Classification and Chemical Component Estimation Based on the Visual Features of Crushed Rock Samples [J]. Arabian Journal of Geosciences，2017，10(15)：324-1-324-9.

[8] MLYNARCZUK M，GÓRSZCZYK A，ŚLIPEK B. The Application of Pattern Recognition in the Automatic Classification of Microscopic Rock Images [J]. Computers & Geosciences，2013，60：126-133.

[9] 谭永健，田苗，徐德馨，等. 基于 Xception 网络的岩石图像分类识别研究 [J]. 地理与地理信息科学，2022，38(3)：17-22.（TAN Y J，TIAN M，XU D X，et al. Research on Rock Image Classification and Recognition Based on Xception Network [J]. Geography and Geo-Information Science，2022，38(3)：17-22.）

[10] 许振浩，马文，林鹏，等. 基于岩石图像迁移学习的岩性智能识别 [J]. 应用基础与工程科学学报，2021，29(5)：1075-1092.（XU Z H，MA W，LIN P，et al. Intelligent Identification of Lithology Based on Rock Image Transfer Learning [J]. Journal of Applied Basic and Engineering Sciences，2021，29(5)：1075-1092.）

[11] 程国建，李碧，万晓龙，等. 基于 SqueezeNet 卷积神经网络的岩石薄片图像分类研究 [J]. 矿物岩石，2021，41(4)：94-101.（CHENG G J，LI B，WAN X L，et al. Research on Rock Thin Section Image Classification Based on SqueezeNet Convolutional Neural Network [J]. Minerals and Rocks，2021，41(4)：94-101.）

[12] 张野，李明超，韩帅. 基于岩石图像深度学习的岩性自动识别与分类方法 [J]. 岩石学报，2018，34(2)：333-342.（ZHANG Y，LI M C，HAN S. Automatic Identification and Classification Method of Lithology Based on Deep Learning of Rock Images [J]. Acta Petrologica Sinica，2018，34(2)：333-342.）

[13] LI N，HAO H Z，GU Q，et al. A Transfer Learning Method for Automatic Identification of Sandstone Microscopic Images [J]. Computers & Geosciences，2017，103：111-121.

[14] POLAT O，POLAT A，EKICI T. Automatic Classification of Volcanic Rocks from Thin Section Images Using Transfer Learning Networks [J]. Neural Computing and Applications，2021，33(18)：11531-11540.

[15] DOS ANJOS C E M，AVILA M R V，VASCONCELOS A G P，et al. Deep Learning for Lithological Classification of Carbonate Rock Micro-CT Images [J]. Computational Geosciences，2021，25(3)：971-983.

[16] HE K M，ZHANG X Y，REN S Q，et al. Deep Residual Learning for Image Recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，NJ：IEEE，2016：770-778.

[17] HOWARD A G，ZHU M L，CHEN B，et al. MobileNets：Efficient Convolutional Neural Networks for Mobile Vision Applications [EB/OL]. (2017-04-17)[2023-01-10]. https://arxiv.org/abs/1704.04861.

[18] HOWARD A，SANDLER M，CHU G，et al. Searching for Mobilenetv3 [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway，NJ：IEEE，2019：1314-1324.

[19] SZEGEDY C，LIU W，JIA Y Q，et al. Going Deeper with Convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，NJ：IEEE，2015：1-9.

[20] SZEGEDY C，VANHOUCKE V，IOFFE S，et al. Rethinking the Inception Architecture for Computer Vision [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，NJ：IEEE，2016：2818-2826.

[21] TOUVRON H，CORD M，DOUZE M，et al. Training Data-Efficient Image Transformers & Distillation through Attention [C]//International Conference on Machine Learning. [S.l.]：PMLR，2021：10347-10357.

（责任编辑：韩 啸）