

文章编号: 0253-2697(2024)12-1818-15 DOI:10.7623/syxb202412008

基于卷积神经网络与视觉变换器的微观剩余油分类方法

赵 娅 管 玉 贾 迪

(东北石油大学计算机与信息技术学院 黑龙江大庆 163318)

摘要:在石油开发领域,微观剩余油的精确识别和分类对提高油田开采效率和采收率至关重要。但传统的剩余油识别技术存在识别效率低、精度不高、资源消耗大等问题,限制了其在油田应用中的实际效果。因此,提出了一种基于卷积神经网络(CNN)与视觉变换器(ViT)的微观剩余油图像分类网络 LLGFormer,该网络架构通过融合局部与全局特征,显著提高了分类精度,同时改善了运行效率。首先,设计了边缘感知增强模块,增强了图像的边缘纹理信息;然后,通过 LLGFormer 数据块并行,提取剩余油的局部与全局特征。此外,引入贡献判别网络,指导 ViT 分支关注有效信息,并采用分步计算策略降低模型的计算量。在自制的微观剩余油数据集和公共数据集上实验验证了 LLGFormer 的有效性。研究结果表明,LLGFormer 在微观剩余油图像的处理速度和性能平衡方面具有显著优势,为石油行业中微观剩余油的自动化识别与分类提供了新的技术途径。

关键词:微观剩余油;图像识别;边缘感知增强;卷积神经网络;视觉变换器

中图分类号:TE19

文献标识码:A

Classification methods for microscopic remaining oil based on convolutional neural network and vision transformer

Zhao Ya Guan Yu Jia Di

(School of Computer and Information Technology, Northeast Petroleum University, Heilongjiang Daqing 163318, China)

Abstract:In the field of oil development, accurate identification and classification of microscopic remaining oil is crucial for improving the oilfield exploitation efficiency and oil recovery. However, the traditional remaining oil identification technique is confronted with the problems such as low identification efficiency, low accuracy and high resource consumption, thus restricting its practical effectiveness in oilfield applications. Therefore, LLGFormer, a microscopic remaining oil image classification network based on convolutional neural network (CNN) and vision transformer (ViT), is proposed, which can not only significantly enhance the classification accuracy but also improve the operational efficiency by fusing local and global features. Firstly, an edge perception enhancement module is designed to enhance image edge texture information, and then the local and global features of remaining oil are extracted in parallel by LLGFormer block. In addition, a contribution discriminant network is introduced to guide the ViT branch to focus on effective information, and a step-by-step computation strategy is adopted to reduce the calculation amount of the model. Moreover, the validity of LLGFormer is verified by experiments on the homemade microscopic remaining oil dataset and the public dataset. This not only proves the significant advantages of LLGFormer in terms of the balance between the speed and performance of microscopic remaining oil image processing, but also provides a new technical path for the automated identification and classification of microscopic remaining oil in the petroleum industry.

Key words:microscopic remaining oil; image recognition; edge perception enhancement; convolutional neural network (CNN); vision transformer (ViT)

引用:赵娅,管玉,贾迪.基于卷积神经网络与视觉变换器的微观剩余油分类方法[J].石油学报,2024,45(12):1818-1832.

Cite: ZHAO Ya, GUAN Yu, JIA Di. Classification methods for microscopic remaining oil based on convolutional neural network and vision transformer[J]. Acta Petrolei Sinica, 2024, 45(12): 1818-1832.

中国多数中、高渗油藏已进入高含水、高采出程度的“双高”开发阶段,70%以上的原油产量是在“双高”期采出的,面临着产量递减、效益降低等现实问题。油田复杂的地质情况和地层中更加复杂的流体分布成为

制约油气采收率的主要因素。另外,剩余油分布零散、复杂,以及开发后期仍存在局部剩余油富集的现象导致剩余油的挖潜率有所降低^[1-3]。因此,掌握油层孔隙中剩余油的赋存形态对油田增加可采储量、提高采收

基金项目:国家自然科学基金项目(No. 61702093)和黑龙江省自然科学基金项目(LH2022F006)资助。

第一作者及通信作者:赵 娅,女,1980年1月生,2019年获东北石油大学博士学位,现为东北石油大学副教授,主要从事智能信息处理、机器学习、深度学习方面的研究工作。Email:261980020273@nepu.edu.cn

率以及科学高效挖潜剩余油具有重要意义。

近年来,针对微观剩余油赋存形态的识别方法主要有传统人工方法和基于机器学习方法两大类。传统的微观剩余油分类方法通常依赖于人工进行分类,如孙先达^[4]通过专业知识手动划分。这种方法效率低下,易受主观因素影响,尤其是在处理不同类别但形状相似剩余油簇的情况下。李忠诚等^[5]根据油、水、岩石的灰度差异对剩余油图像进行二值化处理,清晰展现出剩余油的形状和轮廓,但二值化处理极易丢失色彩对比度不明显的剩余油区域。

随着技术的发展,机器学习方法逐渐被引入到微观剩余油赋存形态的分类中。如陈浩等^[6]依据偏心率、形状因子、长宽比、联通孔隙系数等多种剩余油几何形状特征构建了多叉判决树,实现了剩余油形态的自动分类,减少了人工判别产生的误差;程小龙等^[7]采用 K 近邻(KNN)分类方法取得了较好的效果,但并未解决样本不均衡带来的问题;成璐璐^[8]通过构造加权随机森林对分割图像进行了分类识别,有效缓解了数据样本不平衡的问题,但由于训练数据不足,导致分类器模型未达到最佳状态。相较于人工识别剩余油类型的方法,基于机器学习的方法在识别精度上已经取得了显著的提升,但需要使用专门设计的机器学习分类算法进行分类识别,模型的泛化性能不高,自动化水平仍有待提升。无论是传统人工分类方法还是基于机器学习的分类方法,现有的微观剩余油识别流程通常涉及图像预处理、剩余油区域分割以及剩余油类型分类等多个步骤。如陈浩等^[6]和成璐璐^[8]采用图像分割技术分割出剩余油区域,再设计分类网络对剩余油形态进行分类;李望奇等^[9]改进了之前的方法,直接设计分类网络进行剩余油图像分类,但却耗费大量时间进行图像预处理。这一系列操作不仅复杂,而且影响分类效率和准确性。

针对上述问题,笔者提出一种融合局部与全局特征的剩余油分类网络 LLGFormer(Lightweight Local-Global Feature Fusion Transformer Network),旨在通过深度学习自动提取深层特征进行分类,简化剩余油图像的处理流程^[10]。该方法主要通过设计边缘感知增强模块,利用 Sobel 算子^[11]和边缘对比度插值算法,检测并增强剩余油边界与纹理细节特征,再使用空洞卷积提升模型对剩余油图像中局部相关性和结构信息的提取能力。结合卷积神经网络(Convolutional Neural Networks, CNN)局部表示学习和视觉变换器(Vision Transformer, ViT)^[12]全局建模的特性,设计双分支特征提取模块的 LLGFormer 数据块,充分提取局部与全局特征。同时,为了进一步提升模型的

类性能,采用幽灵(Ghost)卷积、贡献判别网络和分步计算等方案对模型进行轻量化设计,实现网络运行速度与性能间的平衡。

1 微观剩余油赋存形态分类

对于不同方法采集的剩余油图像,存在各种各样的分类方法。目前,石油行业对微观剩余油赋存形态的划分尚无明确规范。笔者基于冷冻制片荧光显微镜技术拍摄的剩余油图像,依据现有的文献数据^[1,13-15]对油藏高含水开发阶段常见的微观剩余油的赋存形态进行了划分。根据形成原理和形态特征,将微观剩余油赋存形态划分为孔表薄膜状、角隅状、颗粒吸附状、簇状、粒间吸附状和狭缝状 6 类(表 1)。

2 基于 CNN 与 ViT 的微观剩余油分类方法

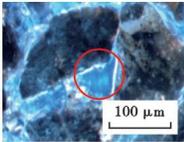
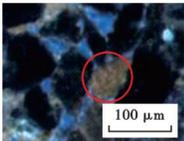
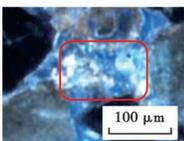
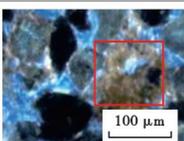
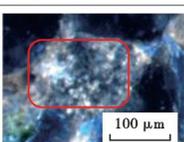
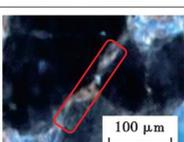
2.1 模型结构

针对人工分类方法和基于机器学习分类方法流程复杂、分类准确率低等问题^[16],笔者提出了一种新型的分类网络 LLGFormer(图 1)。该网络架构通过整合图像预处理与分类过程,简化剩余油图像的处理流程,实现深度学习自动提取图像的深层特征,避免了人为因素的影响。LLGFormer 省略了传统方法中的图像分割步骤,直接对原始图像进行深度学习分析,从而提高剩余油识别的速度和智能化水平。这种方法不仅简化了处理流程,还通过深度学习技术提升了分类的准确性和效率,为微观剩余油的快速、准确识别提供了一种有效的技术途径。

LLGFormer 整体采用多层级下采样^[17]的方式进行构造,分为 4 个阶段,每个阶段进行一次两倍下采样,用以处理不同层级的特征图像。为了平衡网络的复杂性与性能,对各阶段中的 LLGFormer 数据块设置不同的深度,堆叠深度为 $[L_1, L_2, L_3, L_4] = [2, 2, 5, 2]$ 。由于第 3 阶段具有较大的感受野,特征图的分辨率相对较低,因而增加网络层数不会导致计算量剧增,可以投入较大的计算资源。

基于 LLGFormer 法的具体分类流程为:对于输入的图像 $\mathbf{X} \in \mathbf{R}^{H \times W \times C}$,先通过卷积核大小为 3×3 、步长为 2 的卷积初步提取低维特征,随后输入第 1 阶段进行处理。具体而言,每个阶段由边缘感知增强模块和 L_n 个 LLGFormer 数据块组成。首先,将初步提取的低维特征图像输入到边缘感知增强模块中,进行边缘和纹理的细节增强,增强后的特征图经过两倍下采样、升维操作后传入 LLGFormer 数据块进行 L_n 次特征提取,输出融合局部与全局信息的特征图像。经过

表1 微观剩余油赋存形态
Table 1 Microscopic remaining oil occurrence patterns

类别	图像	赋存特征描述	形成原理
孔表薄膜状		孔表薄膜状剩余油主要通过薄膜方式附着在岩石表面的矿物粒子上,仅当岩石表面具有亲油性时,这种形态的剩余油才会形成	岩石壁面亲油,原油吸附在岩石表面不易被剥离
角隅状		角隅状剩余油存在于由复杂孔隙结构构成的“U”形结构内部,一边是与储层的夹角凹槽相接触,另一边则暴露于开放空间	流线难以驱扫到的孔隙死角处
颗粒吸附状		颗粒吸附状剩余油主要以平铺和浸染的形式附着在矿物颗粒表面	
簇状		簇状剩余油位于孔隙和喉道的空间内,表现为分散的簇状、油滴状或团块状分布。在水驱过程中,这些剩余油通常位于未被水波及的区域,与周围多个孔隙相连	孔隙非均质性导致驱替流体绕流后富集
粒间吸附状		粒间吸附状剩余油主要散布在富含泥质和黏土的区域,与黏土矿物碎片混合在一起,呈不规则网状	
狭缝状		狭缝状剩余油存在于细长且狭窄的裂缝中,这些裂缝通常由层状矿物的溶蚀孔或颗粒裂缝以及微粒中含油的层状纹理裂缝构成	因毛细管力束缚,使剩余油卡断在喉道处

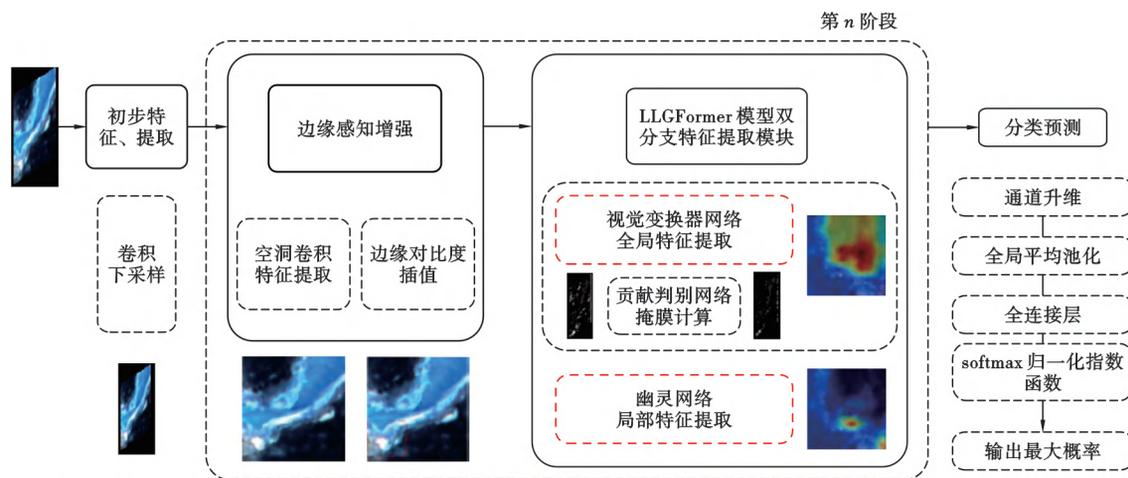


图1 LLGFormer 的整体结构

Fig. 1 Overall structure of LLGFormer

4个阶段的处理后,最后输出的特征图的通道数往往不会过大,故网络最后使用一个 1×1 的卷积来提高输出通道的维度,从而增强模型的表达能力,再利用全局平均池化层和一个全连接层进行分类操作,输出最终分类结果。

2.2 边缘感知增强模块

在剩余油图像分类任务中,细微的纹理和边缘特征对精准分类至关重要。但现有的剩余油处理方法中通常只涉及对剩余油色彩、饱和度等进行增强。如成璐璐^[8]通过饱和度增强和双边滤波实现剩余油图像的

去噪;李望奇等^[9]和 Zhao 等^[18]则采用基于阈值的方法凸显剩余油区域。这些方法忽视了边缘、纹理等关键信息,从而导致分类精度不足,特别是在边界和纹理复杂度较高的情况下,模型难以充分捕捉和利用这些信息。这种局限性可能导致模型在识别不同类型剩余油时表现不佳。为此,笔者设计了一个边缘感知增强(EPE)模块,通过结合空洞卷积和边缘对比度引导的图像插值(Contrast-guided Image Interpolation, CGI),对图像的边缘和纹理细节进行增强(图 2)。

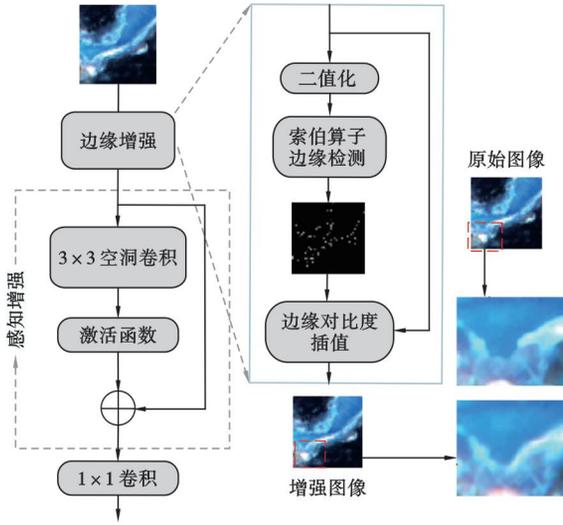


图 2 边缘感知增强模块

Fig. 2 Edge perception enhancement module

首先,对输入的 RGB 色彩模式图像 $F(x, y)$ (其中, $x \in \mathbf{R}^W, y \in \mathbf{R}^H$) 进行灰度化处理,输出灰度图像 $f(x, y)$,以降低边缘检测的计算量;然后,使用边缘检测算子提取图像轮廓信息,鉴于 Sobel 算子^[11]定位准确且可有效抑制噪声的特点,笔者选择该算子计算出 $f(x, y)$ 在水平方向和垂直方向的梯度 G_x 和 G_y ,并根据梯度计算每个像素的近似梯度 G_{xy} :

$$\begin{cases} G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \times f(x, y) \\ G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & 2 & -1 \end{bmatrix} \times f(x, y) \end{cases} \quad (1)$$

$$G_{xy} = \sqrt{G_x + G_y} \quad (2)$$

为了提高计算效率,将式(2)进行简化,对 G_x 和 G_y 的绝对值相加,来计算近似梯度值:

$$G_{xy} = |G_x| + |G_y| \quad (3)$$

将像素的灰度值与预设阈值进行比较,保留灰度值大于阈值的像素点,即图像的边缘点,从而得到边缘特征图像 Z 。接着,依据边缘特征图像 Z 中的像素点,

对 RGB 色彩模式图像应用 CGI 算法^[19]进行插值。CGI 算法依据边缘两侧像素的灰度值差异进行插值,特征图 Z 中保留的像素点即为差异度较大的点,差异度越大,扩散的边缘越宽。同时,该插值算法对边缘像素进行“保护”,即考虑距离边缘一定范围内的非边缘像素,并对其进行与边缘像素相同的插值处理,从而增强边缘细节信息,插值后的图像效果如图 2 中右下角所示。

在边缘增强处理后,笔者结合 3×3 空洞卷积和线性整流函数(ReLU)设计了感知增强模块,并添加一个残差分支进行连接,以增强网络对图像局部相关性和结构信息的提取能力:

$$\mathbf{X}_i = \text{ReLU}[\text{DConv}(\mathbf{X}_{in})] + \mathbf{X}_{in} \quad (4)$$

感知增强后的特征图经过 1×1 卷积调整通道数,并进行下采样操作后输出。

2.3 LLGFormer 数据块设计

2.3.1 LLGFormer 数据块

ViT 在捕捉图像全局特征方面具有优势,但缺乏 CNN 的局部感受能力,对图像中的细节特征和局部结构的捕捉较弱。CNN 通过卷积操作能够有效地提取图像中的局部特征,具有较好的平移不变性,尤其在边缘、纹理等细节特征提取方面有明显优势。因此,笔者提出了 LLGFormer 数据块(图 3),采用局部特征提取和全局特征提取双分支结构,并分别设计轻量级 CNN 和 ViT 分支。

2.3.2 ViT 原理

ViT 是一种将自然语言处理任务中的 Transformer 编码器应用于图像分类的模型。ViT 分类流程如图 4 所示。处理步骤为:①将图像分成若干“块”(patch);②每块被映射为一个“令牌”(token);③将这些图像块平铺成的“令牌”序列输入 Transformer 编码器中进行特征提取;④通过多层感知机(MLP)对这些特征进行分类^[8]。

自注意力模块(self-attention)是 Transformer 编码器中的关键组成部分,该模块通过计算输入图像序列中不同位置之间的关系,将这些位置关联起来,以生成序列的上下文表示。这种机制使得模型能够有效地捕捉全局依赖性,实现对输入序列的更准确表征。自注意力模块计算原理为:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}(\mathbf{Q} \cdot \mathbf{K}^T / \sqrt{d_k}) \times \mathbf{V} \quad (5)$$

首先将块序列投影为 \mathbf{Q} 、 \mathbf{K} 和 \mathbf{V} 矩阵,计算注意力分数(每个元素之间的相关性 $\mathbf{Q} \cdot \mathbf{K}^T$ 和 d_k);再通过 Softmax 函数(归一化指数函数)将注意力分数转为权重,归一化为概率分布。

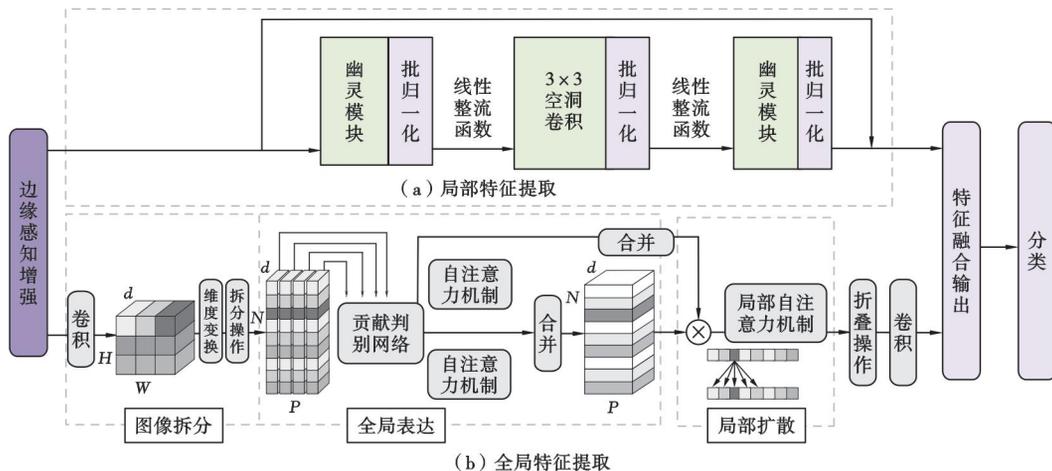


图3 LLGFormer数据块 Fig.3 LLGFormer block

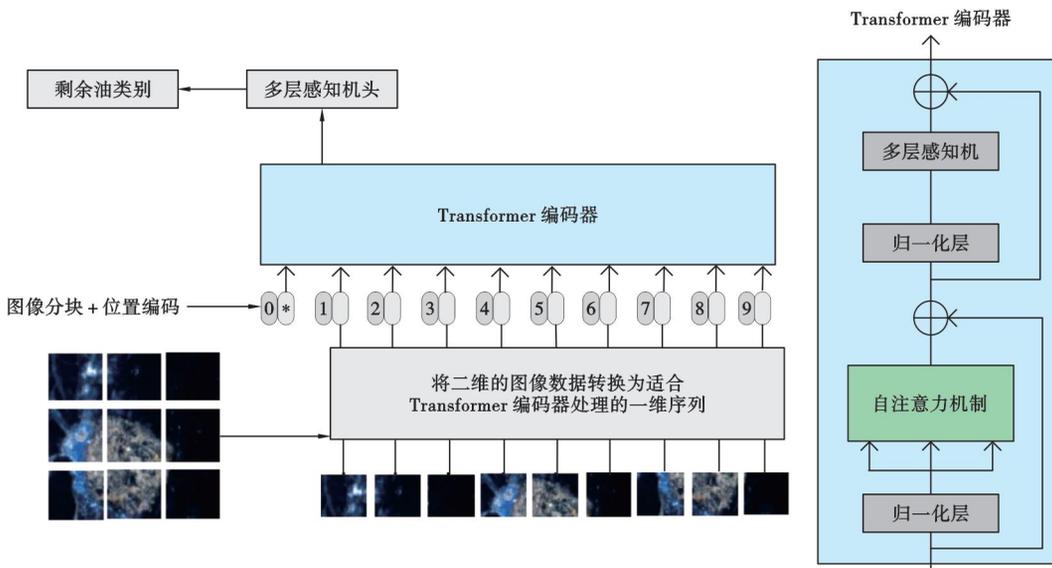


图4 ViT分类流程 Fig.4 ViT classification process

2.3.3 局部特征提取

在CNN提取特征图的过程中,通常存在大量冗余和相似的特征图。这些特征图对于模型的精度又是不可或缺的,都是由卷积操作得到,并输入到下一个卷积层进行运算,包含大量的网络参数,消耗了较多计算资源。为此,笔者结合Ghost卷积^[20]和空洞卷积,设计了一个轻量级的CNN局部特征提取模块,以更低的计算成本来获取这些冗余特征图。

Ghost卷积结构如图5所示。Ghost卷积首先使用少量卷积核对输入特征图进行初步特征提取,然后通过分组卷积操作对提取到的特征进行线性变换,生成“冗余”特征^[20]。这些冗余特征通过低计算成本的操作来补充网络所需的特征表达。最后,将这些特征与初始特征进行拼接,得到最终的输出特征图。

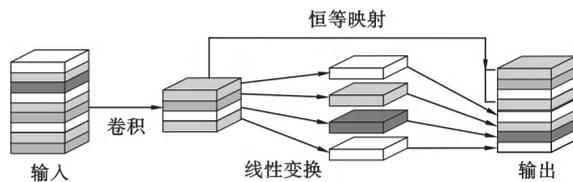


图5 Ghost卷积 Fig.5 Ghost convolution

局部特征提取模块[图3(a)]正是利用了Ghost卷积的这一特性。先使用少量卷积核进行初步特征提取,再通过分组卷积生成“冗余”特征,在减少计算量和参数的同时保持特征表达的完整性。该模块包含2个堆叠的Ghost模块和1个3x3的空洞卷积。原始输入特征分为两部分,一路进行残差连接,另一路经过2次Ghost卷积提取冗余特征,中间加入空洞卷积以扩大感受野,增强局部信息的提取能力。此外,模块中还

加入了批归一化以加速收敛。前半部分使用 ReLU 防止神经元不学习;后半部分则未使用激活函数,以保持数据分布的一致性,加快模型收敛速度。

2.3.4 全局特征提取

自注意力机制通过对输入序列中任意两个位置之间的关系进行建模来获取全局关系,然而,需要计算每个位置与其他位置之间的相似度,并根据相似度进行加权。当输入序列较长时,计算的复杂度呈平方增长,增加了模型的计算负担,此外,自注意力机制在计算相似度时无法明确指示模型关注哪些信息,可能导致模型过度关注一些无关的细节。

受 MobileViT 轻量级网络架构^[21]和稀疏自注意力机制启发,笔者针对上述问题设计了稀疏判别自注意力机制(SDSA),旨在让网络重点关注剩余油部位特征,并使用较少的参数对输入信息进行全局建模。

图 4(b)为 SDSA 模块。与传统的 ViT 模型不同的是,SDSA 模块通过 1×1 卷积进行升维,将输入特征图像切割成多个块,并将这些“块”重新映射排列为“令牌”序列,输入自注意力机制。然后,设计判别网络来评估每个“令牌”的贡献度,并通过掩码机制筛选出重要特征。最后,对筛选出的特征进行局部自注意力计算,通过将全局信息扩散至局部区域,增强模型对有效信息的关注。SDSA 模块具体设计如下:

(1) 特征图像拆分

对于输入的特征图 $X_i \in \mathbf{R}^{H \times W \times C}$,首先通过 1×1 卷积进行升维,增强模型的代表能力和特征提取能力,输出 $X_G \in \mathbf{R}^{H \times W \times d}$ 。然后,将 X_G 切成 N 个“块”,经过将高维度数据展开为多个“块”的变换后得到 $X_U \in \mathbf{R}^{P \times N \times d}$ 。然后,对 X_U 在“块”维度进行拆分(图 6),生成 P 个形状为 $[1, N, d]$ 的序列 $X_p \in \mathbf{R}^{1 \times N \times d}$,每个序列代表 N 个“块”中相同位置元素的集合。生成序列在与自注意力机制结合时,计算从原来的全局计算降为局部计算。

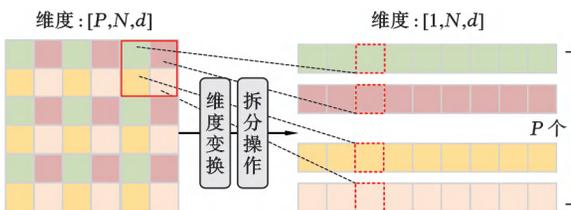


图 6 特征图像拆分示意

Fig. 6 Schematic diagram of feature image splitting

(2) 判别自注意力机制计算(全局表达)

为了进一步降低自注意力机制的计算负担,笔者设计了判别网络,对序列 X_p 进行贡献度得分计算(图 7),

并对贡献度得分较低的冗余特征进行掩码处理^[22]。对输入特征序列 X_p 在 0 维度进行挤压操作,去除空间维度,生成待判别的二维权重序列 $Y_p \in \mathbf{R}^{N \times d}$ 作为判别网络的输入。构建 MLP 网络,对输入序列进行一次全连接计算,隐藏层使用 Sigmoid 函数(神经网络的激活函数)进行非线性激活,再利用一次全连接计算输出 N 个节点的序列。最后,对 MLP 输出的 N 个序列利用 Softmax 函数进行归一化处理,生成每个“令牌”对类别信息的贡献得分 s_p^n ,进而得到权重得分序列 $S_p \in \mathbf{R}^{N \times 1}$ 。

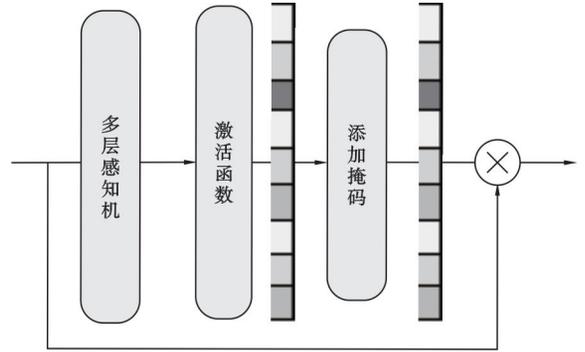


图 7 贡献判别网络

Fig. 7 Contribution discriminant network

贡献判别网络的计算为:

$$Y_p = [y_p^1, y_p^2, y_p^3, \dots, y_p^n] \quad (6)$$

$$E_p^n = b_2 + W_2 \text{Sigmoid}(W_1 Y_p + b_1) \quad (7)$$

$$s_p^n = \exp(E_p^n) / \sum_n \exp(E_p^n) \quad (8)$$

$$S_p = [s_p^1, s_p^2, s_p^3, \dots, s_p^n] \quad (9)$$

对权重得分序列 S_p 进行筛选,利用掩码的方式对无关令牌进行遮挡,设置超参数 z ,将贡献度得分后 $1/z$ 位置处的权重设置为负无穷,以便在计算注意力分数时将其抑制为接近零的概率。考虑在到越往后的阶段,特征图中每个像素所含的语义信息越多,需要保留更多的特征信息,因此对于 4 个阶段的超参数 z 分别设置为 2、2、4 和 4。经过掩码计算后得到的权重序列为:

$$S_{p_Mask} = [-\infty, s_p^2, s_p^3, -\infty, \dots, s_p^n] \quad (10)$$

最终,将 S_{p_Mask} 与原始序列 X_p 相乘,保留显著区域。笔者将第 1 阶段使用掩码运算和未使用掩码运算的单通道特征图像进行可视化(图 8),像素值的范围从 0 到 255,0 对应黑色,255 对应白色,因此暗色块的权重低于亮色块。由图 8 可以发现,在未经过贡献判别网络掩码运算的特征图中,网络耗费了大量的计算资源用在背景区域,而经过笔者提出的贡献判别网络掩码运算后,无关区域的权重明显降低,以此引导模型重点关注有效的信息,提高模型的准确性。

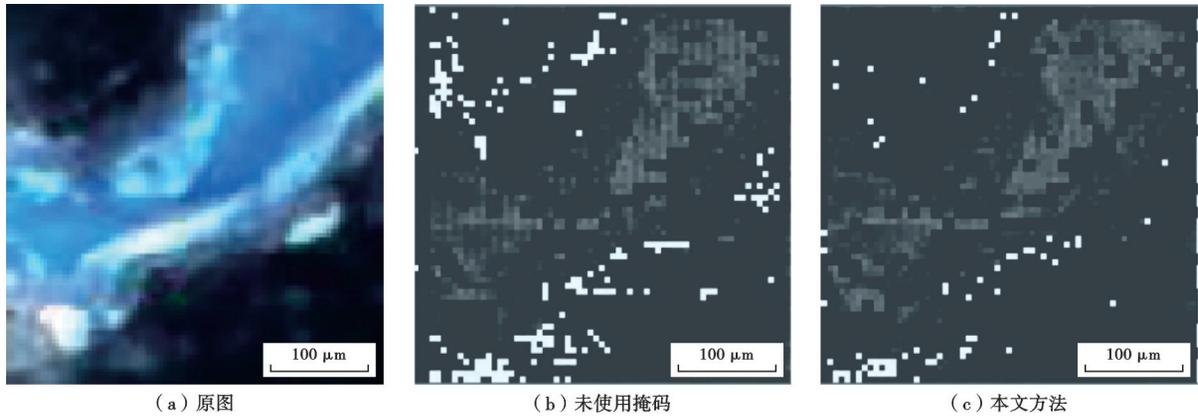


图8 第1阶段特征图可视化

Fig. 8 Visualization of stage 1 feature map

将经过掩码处理后的 P 个序列分别传入自注意力机制进行初步的全局表达运算,输出特征序列,进行合并操作后得到特征图像 $\mathbf{X}_g \in \mathbf{R}^{P \times N \times d}$ 。

(3) 局部自注意力机制计算(局部扩散)

特征图像 \mathbf{X}_g 经过全局表达后,已经将显著区域关联起来,但不同“块”之间的像素仅在对对应位置处存在联系,因而关联程度不高。显著区域“块”包含重要信息,对分类的精度影响较大,加之对经过筛选的特征图像进行计算的成本远小于对整个特征图进行自注意力机制计算的成本,因此,需要对 \mathbf{X}_g 使用稀疏注意力机制进行进一步特征提取。对全局表达输出的特征图像 \mathbf{X}_g 在 P 维度进行展开,将 \mathbf{S}_{p_Mask} 在 N 维度拼接后的权重序列与 \mathbf{X}_g 相乘,从而屏蔽非显著区域特征,输出 \mathbf{X}_G :

$$\mathbf{X}_G = \text{concat}(\mathbf{S}_{p_Mask}, N) \times \text{view}(P \times N, d, \mathbf{X}_g) \quad (11)$$

对 \mathbf{X}_G 中的每个元素进行局部自注意力机制(local self-attention)运算。设置超参数 k ,每个“令牌”只对周围 $1+2k$ 范围内的“令牌”计算相关性,以此“令牌”将全局关联扩散至局部,从而让 Transformer 编码器更好地关注有效的信息,实现对关键信息的进一步增强。基于该方法的像素感受野效果如图9所示。

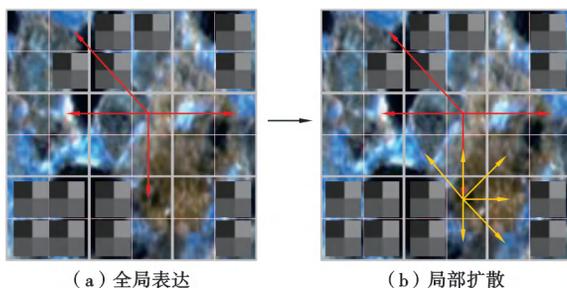


图9 全局特征感受野

Fig. 9 Global feature reception field

2.3.5 稀疏判别自注意力机制计算量分析

ViT 分支采用全局表达与局部扩散相结合的分步计算策略,自注意力模块的计算量可分为两部分。全局表达部分可表示为 $N^2 \times P \times d$,其中 $N \times P$ 代表原始图像大小。经过掩码运算后,最终自注意力模块的计算量可表示为:

$$N^2 \times P \times d = \frac{W \times H}{w \times h} \times \left(1 - \frac{1}{z}\right) \times W \times H \times \left(1 - \frac{1}{2}\right) \times d = \frac{\left[\left(1 - \frac{1}{2}\right) \times W \times H\right]^2}{w \times h} \times d \quad (z \neq 1) \quad (12)$$

对于局部自注意力机制,每个元素与周围 $1+2k$ 个元素相联系,因此整个序列 N_L 个元素总的计算量为 $(1-1/z) \times W \times H \times (1+2k) \times d$,ViT 分支中自注意力机制相关的计算量为 $N^2 \times P \times d + N_L \times (1+2k) \times d$,而传统的自注意力计算的复杂度为 $O(N_L^2 \times d)$,计算量为 $[W \times H \times (1-1/z)]^2 \times d$ 。为了更清晰地反映计算量大小,去除相同变量 d 并代入实验数据进行计算。取 $z=2, k=2, w \times h=4$,则 ViT 分支的计算量为 $0.0625(W \times H)^2 + 2.5W \times H$ 。在 $W \times H$ 较大时, $2.5W \times H$ 可以看作偏置项,而传统自注意力机制为 $0.25(W \times H)^2$,约是笔者方法计算量的4倍,因此笔者的方法更加轻量且高效。

3 实验应用

3.1 实验环境及参数设置

笔者使用 Python 语言,在 PyCharm2022 软件中进行编程,GPU 采用 12GB 显存容量的 RTX3080Ti 显卡。实验基于 Pytorch1.12 深度学习框架进行网络模型的搭建、训练与测试,使用 Cross Entropy Loss 作为训练时的损失函数,采用随机梯度下降(SGD)算法

优化网络模型,应用余弦退火算法来调整学习率,初始学习率设置为 0.02。模型迭代的次数设置为 200,批大小设置为 16。为了避免迭代过程中存在多个拥有相同数据的批大小,每次训练前对数据集进行随机打乱,可以有效加快模型收敛的速度。

3.2 数据集及评价指标

由于冷冻制片荧光显微镜技术成本昂贵以及采集的图片存在保密性等原因,导致微观剩余油相关的公共数据集十分稀缺,从网络中爬取的微观剩余油图像存在清晰度低、色彩昏暗、质量差等问题,与实际场景中微观剩余油分布情况存在较大差异,容易对剩余油的分类研究产生不利影响。因此,笔者利用荧光显微镜对剩余油的冷冻制片在不同视域和光照强度下进行拍摄,人工截取不同类型的剩余油图像,构建了一个新的数据集。剩余油数据集由 6 种共计 5 450 张微观剩余油图像组成(表 2),其中一部分来自筛选后的网络图像。首先将图像分辨率调整为 256×256 ,再按 7:2:1 的比例划分成训练集、验证集和测试集。

表 2 剩余油数据集组成

Table 2 Composition of remaining oil dataset

分组	类别	拍摄图像 截取量	网络图像 截取量
1	孔表薄膜状	823	90
2	角隅状	941	100
3	颗粒吸附状	678	80
4	簇状	857	50
5	狭缝状	864	90
6	粒间吸附状	827	30

在实验图像的分类任务中,常用 Top-1 准确率($P_{accuracy}$)、精确率($P_{precision}$)、召回率(P_{recall})和 F1 分数(S_{F1})作为评估模型性能的指标。此外,引入参数量 Params、浮点运算数 FLOPs 和画面数来辅

助评价模型的运行效率。 $P_{accuracy}$ 即正确预测样本数的比例:

$$P_{accuracy} = (n_{TP} + n_{TN}) / (n_{TP} + n_{FP} + n_{FN} + n_{TN}) \quad (13)$$

$P_{precision}$ 表示在模型预测的正类样本中,实际为正类的比例:

$$P_{accuracy} = n_{TP} / (n_{TP} + n_{FP}) \quad (14)$$

P_{recall} 是模型对实际正类样本的捕获能力,即实际正类样本中被正确预测为正类的比例:

$$P_{recall} = n_{TP} / (n_{TP} + n_{FN}) \quad (15)$$

S_{F1} 结合了精确率和召回率,通过二者的调和平均数来衡量模型在正类预测中的整体表现:

$$S_{F1} = 2P_{precision} \times P_{recall} / (P_{precision} + P_{recall}) \quad (16)$$

对于剩余油图像的多分类任务,笔者采用宏平均(macro-averaging)对模型的整体性能进行评估。宏平均方法通过分别计算每个类别的精确率、召回率和 F1 分数,然后对所有类别的结果进行算术平均,得到平均精确率、平均召回率和平均 F1 分数,可以有效避免类别不平衡的影响,更好地反映模型在不同类别上的表现。

3.3 实验结果及分析

3.3.1 主流模型对比实验

为验证 LLGFormer 的有效性,笔者将其与目前高效且轻量的网络模型进行了比较。算法类型分为基于 CNN 的方法、基于 ViT 的方法和基于 CNN 与 ViT 混合的方法,具体的网络架构为 EfficientNetV2-S^[23]、MobileNet-V2^[24]、MobileNetV3-S^[25]、ShuffleNetV2_1.5^[26]、ViT-B、CrossViT-S^[27]、Swin-T^[28]、BiFormer-S^[29]、CVT^[30]、MobileViT-S、FastViTS12^[31]以及笔者的 LLGFormer 法,12 种网络架构在微观剩余油数据集的分类结果见表 3。

表 3 算法在剩余油数据集的性能对比

Table 3 Comparison of algorithm performance in remaining oil datasets

算法类型	网络架构名称	Top-1 准确率/%	平均精确率/%	平均召回率/%	平均 F1 分数/%	参数量/ 10^6
CNN	EfficientNetV2-S	94.12	93.98	93.92	93.83	21.46
	MobileNetV2	93.57	93.68	93.92	93.70	3.50
	MobileNetV3-S	91.24	91.26	90.17	90.52	2.54
	ShuffleNetV2_1.5	92.37	92.78	92.78	91.85	3.50
ViT	ViT-B	91.09	91.22	89.98	90.38	86.42
	CrossViT-S	94.41	94.17	93.92	93.71	26.70
	Swin-T	93.39	93.15	92.25	92.37	28.27
	BiFormer-S	94.94	94.58	94.63	94.26	25.50
CNN 与 ViT 混合	CVT	92.57	92.40	92.06	92.14	20.10
	MobileViT-S	93.73	93.94	92.02	92.58	5.02
	FastViT-S12	94.78	94.89	94.10	94.39	9.46
	LLGFormer	95.07	95.37	94.21	94.54	4.12

与 CNN 相比,LLGFormer 的 Top-1 准确率较 EfficientNetV2-S、MobileNetV2、MobileNetV3-S 和 ShuffleNetV2_1.5 分别提升了 0.95%、1.50%、3.83% 和 2.70%,参数量略微高于目前轻量级的 CNN;相较于纯 ViT 网络架构,LLGFormer 由于引入多层次下采样结构,参数量大幅度下降,其 Top-1 准确率较 ViT-B、CrossViT、Swin-T 和 BiFormer-S 分别提升 3.98%、0.66%、1.68% 和 0.13%;对于同类型的混合网络,LLGFormer 的 Top-1 准确率较 CVT、MobileViT-S 和 FastViT-S12 分别提升 2.5%、1.34% 和 0.29%,参数量分别下降 15.98×10^6 、 0.90×10^6 和 5.34×10^6 ,在平均精确率、平均召回率和平均 F1 分数上均高于其他网络。实验结果表明,与其他主流分类网络相比较,LLGFormer 法能够通过较低的运行成本获得更高的分类准确率。

LLGFormer 法在测试集中的预测结果如图 10 所示。从图 10 可以看出:模型在颗粒吸附状方面的精确率表现突出,达到 98.68%,显著优于其他类型;在召回率方面,簇状剩余油的表现最佳,达到 98.12%;在 F1 分数指标中,对颗粒吸附状和簇状剩余油的分类效果依旧领先其他类型。模型对角隅状和粒间吸附状剩余油的分类效果相对较差。

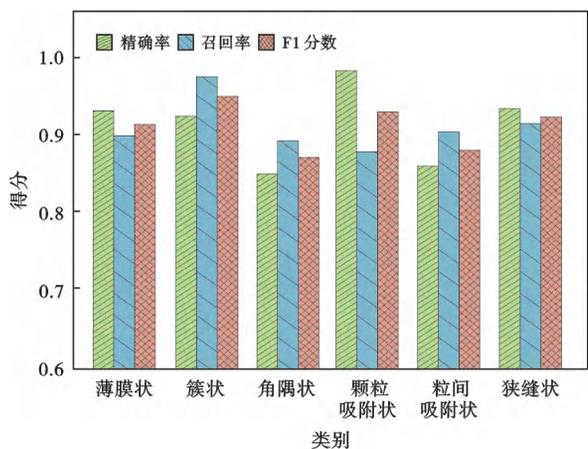


图 10 LLGFormer 法在每一类剩余油上的识别性能对比
Fig. 10 Comparison of LLGFormer recognition performance on each type of remaining oils

基于图 11 的混淆矩阵可以进一步分析分类结果。粒间吸附状剩余油较多地被误分为簇状,而角隅状剩余油则较多地被误分为狭缝状。结合表 1 的剩余油形态对比可以发现,粒间吸附状与簇状剩余油的结构较为相似,狭缝状与角隅状剩余油的颜色和形态也较为接近。因此,模型在分类时容易混淆这些相似结构的类别。此外,模型在预测过程中偏向数据量较大且易于划分的类别,对于数据量较少且结构复杂的类别,例

如粒间吸附状剩余油,模型的准确率受到数据类别不平衡的影响,导致分类效果较差。这说明在处理复杂且数据量较少的类别时,模型的泛化能力有所不足。

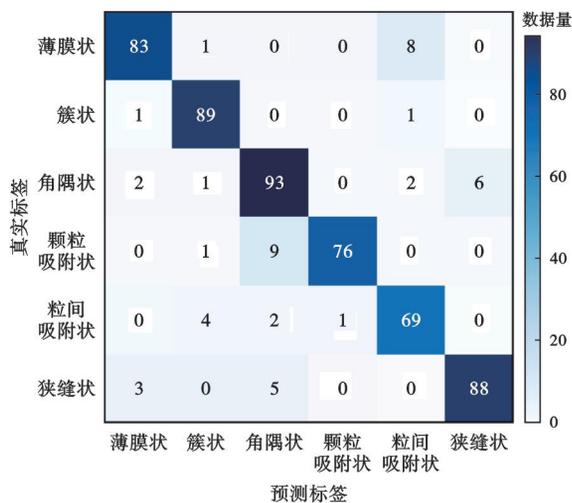


图 11 LLGFormer 预测结果的混淆矩阵
Fig. 11 Confusion matrix of LLGFormer prediction results

3.3.2 主流模型对比实验可视化分析

结合各模型在剩余油测试集上的分类性能对比(图 12)可以发现,相较于 CNN,LLGFormer 的 Top-1 准确率大幅提升,而对于纯 Transformer 编码器和 CNN 与 ViT 的混合架构,笔者的方法在显著降低参数量的情况下可提升分类准确率,表现出突出的综合性能。

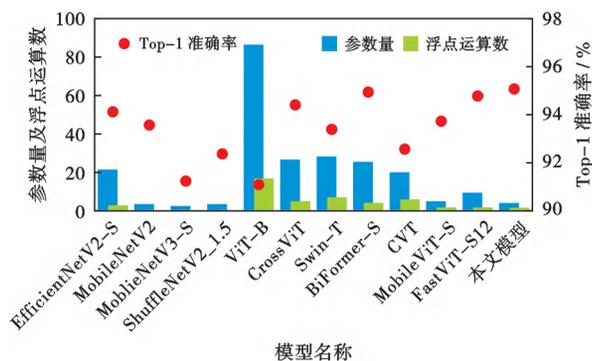


图 12 各类模型的性能对比结果
Fig. 12 Comparison of classification performance of various models

为进一步分析模型的特征提取能力,笔者采用梯度加权类激活映射(Grad-CAM)算法对网络最后一层的特征图进行了可视化实验。通过与主流模型的特征可视化图像(图 13)对比,得出以下结论:

(1) CNN 侧重于局部特征的提取,例如 EfficientNetV2 通过引入更深、更宽的网络结构和更多的特征提取层,对图像中的剩余油区域聚焦性更强。从图 13(b)可看出,EfficientNetV2 能够精准定位剩余油的显著

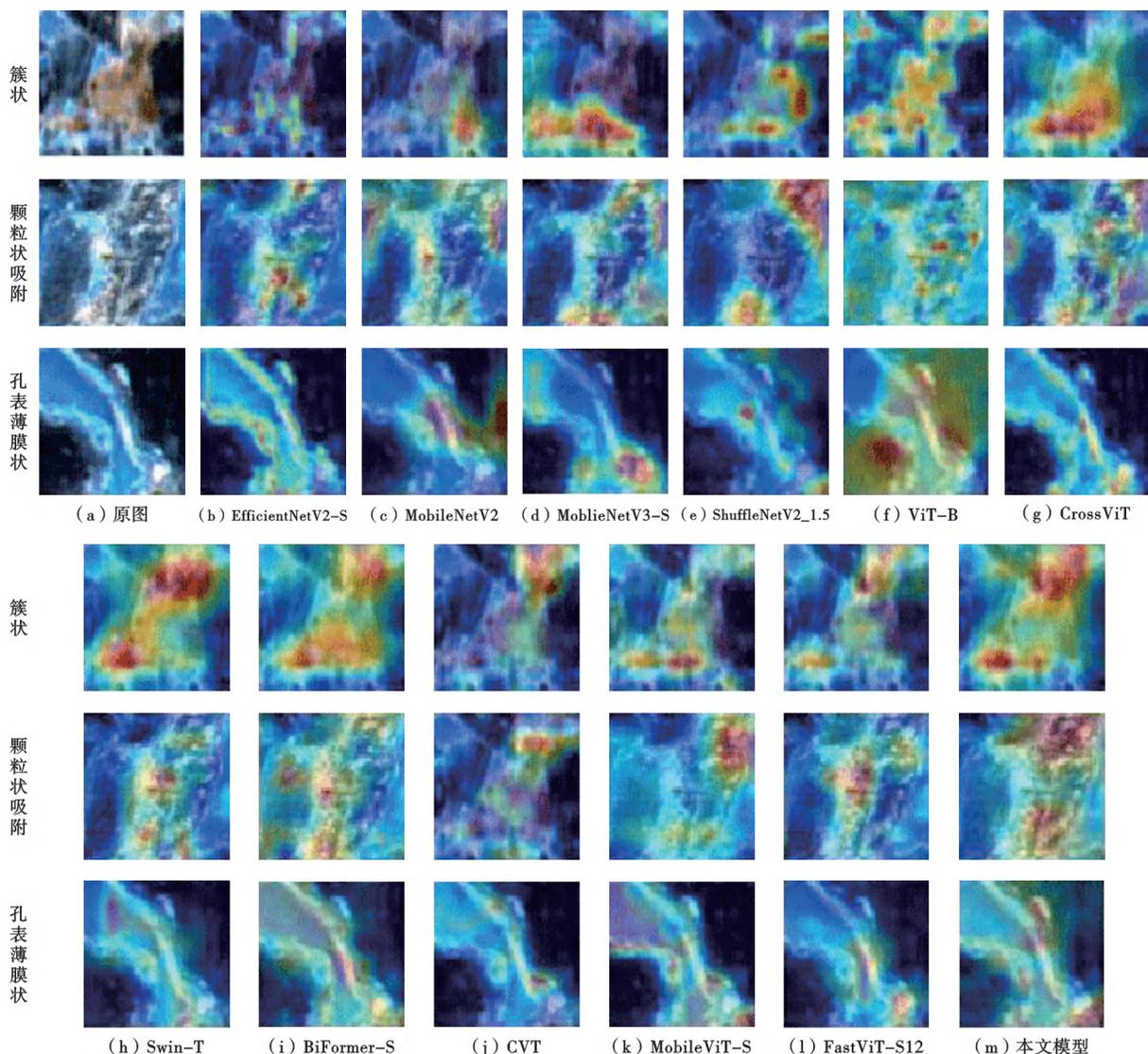


图 13 不同模型的注意力可视化结果

Fig. 13 Attention visualisation results for different models

区域,但由于感受野受限,忽视了鉴别图像类别的其他关键区域。

(2) 相较于 CNN, ViT 网络架构具有较强的全局特征学习能力。如 Swin-T 通过窗口注意力机制在保持局部依赖性的前提下,扩大了模型的感知范围,但对于需要全局上下文信息才能正确分类的图像,模型可能难以建立长距离的依赖关系,导致模型在处理复杂结构的图像时性能下降。在对颗粒吸附状剩余油分类时,由于剩余油呈零散分布, Swin-T 的窗口难以建立长距离关联,丢失大量全局信息。

(3) 混合型网络架构很好地综合 CNN 和 ViT 的优势, MobileViT 和 Fastvit 依赖 CNN 提取局部信息的能力,使得 ViT 部分可以围绕 CNN 定位的区域进行全局特征提取,令剩余油显著区域被全局注意力覆盖。但当提取的局部信息较多时, ViT 关注的区域就

会遍布整个图像,导致网络学习到大量无关特征信息,造成注意力退化及计算量倍增等问题。

(4) 笔者提出的 LLGFormer 模型充分结合了 CNN 和 ViT 的优点,并在捕获整体特征的基础上,重点增强局部特征区域。同时,得益于判别网络去除无关背景的影响, LLGFormer 提取的全局特征紧紧围绕在剩余油区域附近,使网络得以集中注意力关注有效信息。

3.3.3 LLGFormer 法与现有分类方法对比

LLGFormer 法与现有的微观剩余油识别方法进行对比(表 4),在识别精度和智能化水平上取得了显著提升。传统方法依赖人工分类,不仅效率低下,还容易受到主观因素的干扰。基于机器学习的分类方法尽管可提高识别精度,但通常需要设计特定的分类算法,泛化能力较弱。现有的深度学习方法多采用先分割后

表4 LLGFormer法与现有方法的对比

Table 4 Comparison between LLGFormer method and existing methods

方法	划分依据	识别方法	准确率/%	参数量/ 10^6	来源
人工方法	形成机理以及最终存在的状态	人工判别			文献[2]
	形成机理以及最终存在的状态	根据油、水、岩石的灰度差异进行二值化处理			文献[5]
机器学习方法	圆形成度、凹度、长宽比、最小 Feret 直径、周长	KNN 算法	90.8		文献[7]
	偏心率、形状因子、长宽比、联通孔隙系数	多叉判决树	90		文献[6]
	荧光颜色、形状	边缘检测法、加权随机森林	78.2	>100	文献[8]
深度学习 方法	圆形成度、长宽比、矩形度、凹度、偏心率	BP 神经网络分类器	84.1		文献[16]
	油状类型	DeepLabV3+ 分割, MobileNetV2 分类	91.8	>60	文献[32]
	多维度多尺寸的特征信息	acgan 数据增强、VGG 分类	92	>143	文献[9]
	形成机理以及最终存在的状态	CNN 与 ViT 混合模型的 LLGFormer 法	95.07	4.12	本文

分类的双阶段识别方式,虽较机器学习方法有所改进,但仍存在计算复杂度高的问题。相比之下,LLGFormer 网络架构不仅进一步提升了分类精度,还通过轻量化设计实现了高效的特征提取与融合,显著减少了模型的计算量,满足了实际应用中轻量化部署的需求。

3.3.4 拓展对比实验

为了验证 LLGFormer 网络架构的适用性和有效性,笔者选用南京大学火成岩教学薄片照片数据集^[33]与 CUB-200-2011 数据集^[34]进行拓展实验。火成岩教学薄片数据集的图像背景环境与微观剩余油数据集具有一定相似性,其中包含 40 种不同类型的岩石,经过数据增强处理后,共有 5 832 张图像;CUB-200-2011 数据集含有 200 种鸟类,11 788 幅图像。对以上数据集按 7:2:1 的比例划分成训练集、验证集和测试集,采用多种评价指标,所有方法均在相同的实验环境下训练。

从算法在公开数据集上的性能对比结果(表 5)和评价指标得分曲线(图 14)可知:在分类精度方面,

LLGFormer 法在两个公共数据集数据上的 Top-1 准确率相较 CNN 和 ViT 类算法都有明显提升;与同类型中性能较为优异的 FastViT-S12 网络架构相比,LLGFormer 在两个数据集上的 Top-1 准确率分别提升了 0.67% 和 0.65%,平均精确度、平均召回率和平均 F1 分数均高于其他模型。在运行效率方面,LLGFormer 与最轻量的 CNN 类型的网络架构存在一定差距,但与当前的 ViT 类型和混合类型的网络架构相比,仍保持较高的效率;其浮点运算数仅为 1.68×10^9 ,画面数仅略低于 MobileViT-S 和 FastViT-S12,但远高于其他 ViT 类型和部分混合类型的网络架构,展现出了良好的运行性能和计算效率。

实验结果表明,LLGFormer 网络架构不仅在剩余油数据集上表现出色,在处理其他较高复杂性数据集时同样拥有强大的泛化能力和鲁棒性。

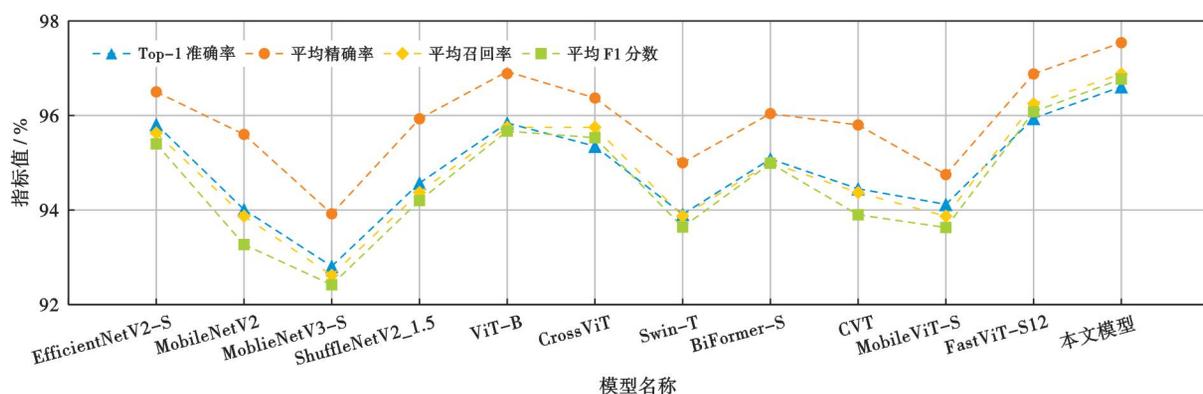
3.3.5 消融实验

为了验证 LLGFormer 网络架构中不同模块的有效性,笔者在自制剩余油数据集上进行了一系列的消融实验。消融实验的具体数据见表 6。

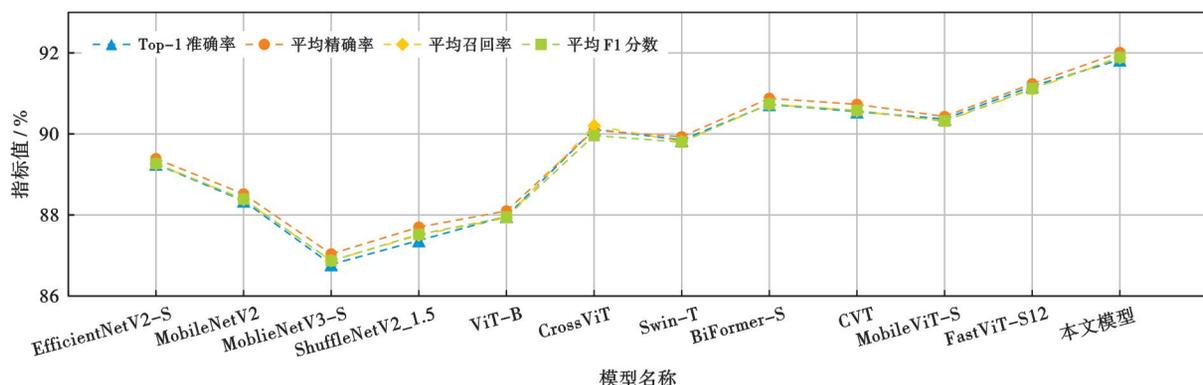
表5 算法在公开数据集性能对比

Table 5 Comparison of algorithm performance in public datasets

算法类型	网络架构	火成岩数据集				CUB-200-2011 数据集				运行效率	
		Top-1 准确率/%	平均精确率/%	平均召回率/%	平均 F1 分数/%	Top-1 准确率/%	平均精确率/%	平均召回率/%	平均 F1 分数/%	浮点运算数/ 10^9	画面数/(帧/s)
CNN	EfficientNetV2-S	95.81	96.50	95.63	95.40	89.26	89.39	89.26	89.27	2.89	57.32
	MobileNetV2	94.01	95.60	93.87	93.27	88.35	88.52	88.38	88.39	0.42	48.70
	MobileNetV3-S	92.81	93.92	92.62	92.42	86.78	87.04	86.87	86.87	0.08	120.76
	ShuffleNetV2_1.5	94.57	95.93	94.37	94.20	87.37	87.70	87.50	87.52	0.40	114.58
ViT	ViT-B	95.84	96.93	95.75	95.67	87.97	88.10	87.93	87.95	16.86	35.20
	CrossViT	95.35	96.37	95.75	95.53	90.12	90.09	90.21	89.96	5.08	52.02
	Swin-T	93.91	95.00	93.87	93.64	89.85	89.93	89.81	89.80	7.11	53.68
	BiFormer-S	95.08	96.04	95.00	94.99	90.73	90.88	90.73	90.74	4.20	35.40
CNN 与 ViT 混合	CVT	94.45	95.80	94.37	93.90	90.55	90.73	90.57	90.58	6.03	44.88
	MobileViT-S	94.12	94.75	93.87	93.63	90.37	90.43	90.31	90.32	1.82	69.36
	FastViT-S12	95.94	96.88	96.25	96.08	91.18	91.24	91.10	91.12	1.84	75.22
	LLGFormer	96.61	97.54	96.88	96.77	91.83	92.01	91.88	91.89	1.68	61.59



(a) 模型在火成岩数据集上的性能对比



(b) 模型在 CUB200 数据集上的性能对比

图 14 不同模型在公共数据集上的性能指标对比

Fig. 14 Comparison of performance metrics of different models in public datasets

表 6 消融实验定量评估

Table 6 Quantitative assessments of ablation experiments

实验编号	方法	Top-1 准确率/ %	平均精确率/ %	平均召回率/ %	平均 F1 分数/ %	参数量/ 10 ⁶	浮点运算数/ 10 ⁹	画面数/ (帧/s)
实验 1	基准网络	89.92	90.79	89.75	89.93	4.484	1.701	73.31
实验 2	基准网络 + 局部特征提取	92.31	92.92	92.02	92.27	1.583	0.487	82.81
实验 3	基准网络 + 全局特征提取	93.45	93.33	92.25	92.56	4.171	1.584	70.55
实验 4	基准网络 + LLGFormer 数据块	94.63	94.29	93.92	94.01	4.650	1.772	65.55
实验 5	基准网络 + LLGFormer 数据块 + 边缘感知增强模块	95.07	95.37	94.21	94.54	4.717	1.784	61.59

实验 1 采用多层次下采样方式构造四阶段的基准网络,每个阶段采用纯 Transformer 编码器进行全局特征建模,Transformer 编码器的深度设置为[2,2,5,2]。实验 2 和实验 3 基于基准网络,分别将纯 Transformer 编码器替换为笔者提出的 LLGFormer 数据块的局部特征提取分支和全局特征提取分支,Top-1 准确率较实验 1 分别提升了 2.39%和 3.53%,参数量和浮点运算数均有所下降,表明 LLGFormer 数据块的两个分支均可有效识别微观剩余油类型。实验 4 在基准网络的基础上,将纯 Transformer 编码器替换为 LLGFormer 数据块,使用通道拼接再卷积降维的方法对局部特征提取分支和全局特征提取分支输出的特征图进行融合,Top-1 准确率较实验 1 提升了 4.71%,参数量和浮点运算数变动较小,但画面数(FPS)下降了近

8 帧/s,这归因于网络使用多分支结构并大量使用特征融合,使得模型将大量的时间用在从显存中读写数据,GPU 的算力未得到充分利用。尽管如此,笔者提出的方法依旧具有较快的识别速度,可以满足实时分类的要求。实验 5 在实验 4 的基础上,在模型的每个阶段引入边缘感知增强模块,在不影响模型计算和运行效率的情况下,精度提升了 0.34%。这表明,边缘感知增强模块能有效增强特征图的边缘细节信息并扩大感受野,提升模型对图像的局部特征提取能力。

笔者利用消融实验训练得到的权重文件来处理狭缝状、粒间吸附状和角隅状这 3 类剩余油图像,使用 Grad-CAM 算法输出特征热力图(图 15),颜色越深表明权重越大,关注度越高。

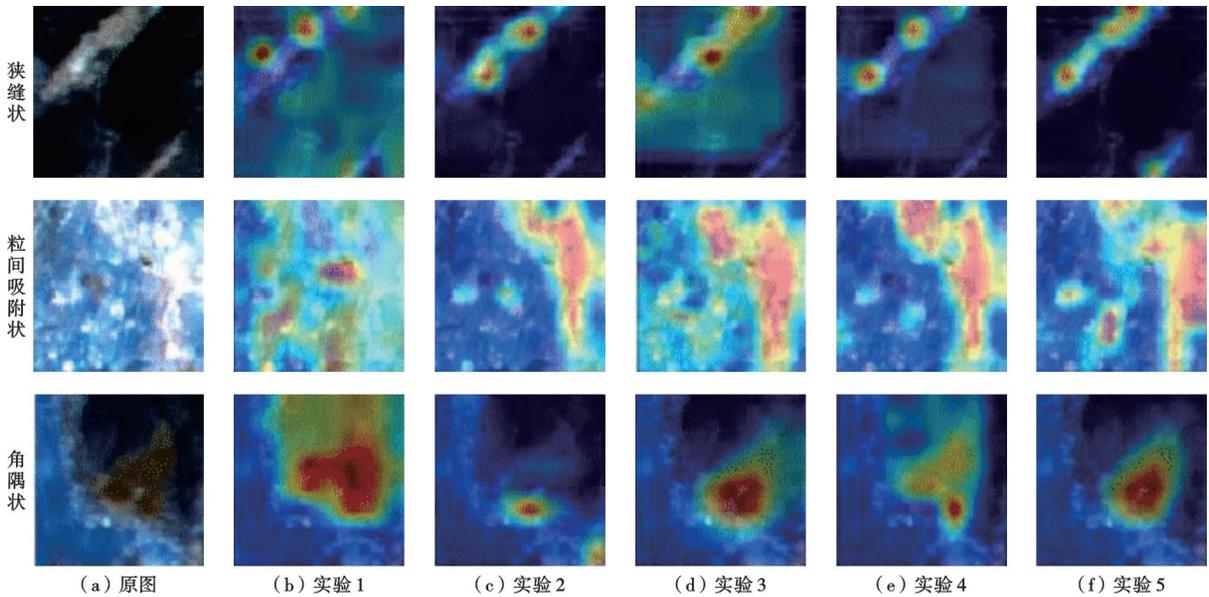


图 15 消融实验结果的注意力可视化

Fig. 15 Attention visualization of ablation experiment results

从消融实验不同阶段的可视化热力图(图 15)中可以看出,单独添加局部特征提取后,模型能够精准定位关键特征。当引入全局特征提取后,模型的关注区域明显扩大,相较于实验 1 减少了对无关区域的关注。当同时添加全局与局部特征提取模块时,模型在重点识别局部区域的基础上,增强了对周围关键特征的检测。在实验 5 中,当添加边缘感知增强模块后,狭缝状和粒间吸附状剩余油中的微小颗粒边缘特征更加清晰,表明局部特征提取模块显著提升了模型对复杂结构的感知能力。

4 结论

(1) 提出了基于 CNN 和 ViT 的微观剩余油分类方法,融合局部与全局特征,显著提高了分类精度。在自建的剩余油数据集上,Top-1 准确率达到 95.07%,较现有最优方法提升了 3.07%,且模型参数量和计算复杂度均大幅降低。

(2) 设计了边缘感知增强模块和贡献判别网络,增强了模型对图像边缘细节和关键特征的提取能力。通过全局表达和局部扩散策略,模型以较低的计算代价实现了关键特征的高效提取,画面数满足实时应用的需求,适合部署在资源受限的设备上。

(3) LLGFormer 存在一定局限性,需进一步优化。受数据集的规模和多样性限制,模型在其他领域或不同剩余油数据集上的泛化能力可能受限。此外,在处理数据量较少且结构复杂的类别时,性能有待提升,且与同类型中性能较为优异的 FastViT-S12 网络架构相比,处

理的画面数偏低。未来将针对模型压缩与加速、数据集扩展与多样化、处理类别不平衡等问题展开研究,以提升 LLGFormer 的能力,拓展其应用领域。

符号注释: n —任一阶段; L_n —第 n 个阶段中 LLGFormer 数据块的堆叠深度,堆叠深度 L_1 、 L_2 、 L_3 和 L_4 分别取值 2、2、5 和 2; \mathbf{X} —模型输入的原始图像; \mathbf{R} —实数域; H —图像的高度; W —图像的宽度; C —图像的通道数; $F(x, y)$ —输入的 RGB 色彩模式图像; $x \in \mathbf{R}^W$ 、 $y \in \mathbf{R}^H$ —图像的宽和高的坐标; $f(x, y)$ —灰度化后的图像; \mathbf{G}_x 、 \mathbf{G}_y —灰度图像 $f(x, y)$ 在水平方向和垂直方向的梯度; \mathbf{G}_{xy} —每个像素的近似梯度值; \mathbf{Z} —通过边缘检测后的边缘特征图像;DConv— 3×3 的空洞卷积;ReLU—线性整流函数; \mathbf{X}_{in} —边缘增强后的图像; \mathbf{X}_i —边缘增强模块最终输出的特征图像; \mathbf{Q} —查询向量,用来表示当前处理的输入序列中的某个位置的特征; \mathbf{K} —键向量,表示与查询相关联的特征或上下文信息,用来与 \mathbf{Q} 进行匹配和计算相似度; \mathbf{V} —值向量,表示与 \mathbf{K} 相关的实际信息;Softmax—激活函数; $\mathbf{Q} \cdot \mathbf{K}^T$ — \mathbf{Q} 和 \mathbf{K} 的点积,表示两个向量的相似度; d_k — \mathbf{Q} 和 \mathbf{K} 的维度;Attention($\mathbf{Q}, \mathbf{K}, \mathbf{V}$)—通过计算 \mathbf{Q} 与 \mathbf{K} 的相似度生成权重,并用这些权重加权得到最终的输出; \mathbf{X}_c —升维后输出的特征图像; d —通道数; \mathbf{X}_U —经过高维度数据展开变换操作后的特征图像序列; h —“块”的高; w —“块”的宽; P —每个“块”的大小, $P = h \times w$; N —“块”的数量, $N = H \times W / P$; \mathbf{X}_p —生成的序列,每个序列代表 N 个“块”中相同位置元素的集合; \mathbf{Y}_p —输入判别网络的二维权重序列,去除空间维

度; y_p^n —序列 Y_p 中的第 n 个分量; p —第 p 个序列的索引;Sigmoid—神经网络的激活函数; S_p —权重得分序列; s_p^n —第 p 个样本在不同位置的得分; W_1 、 W_2 —判别网络中的权重矩阵; E_p^n —第 p 个样本在第 n 个位置的输出值; b_1 、 b_2 —偏置项; z —掩码机制中的超参数; S_{p_Mask} —掩码计算后的贡献得分序列; X_g —经过全局自注意力机制处理后的特征图像;concat—拼接操作; X_G —经过拼接操作和重新排列后得到的特征图像;view($P \times N, d, X_g$)—对特征图 X_g 进行视图变换,将其展平为一个新的维度, $P \times N$ 代表原始图像大小; k —超参数; N_L —经过掩码后的序列长度; $O(N_L^2 \times d)$ —大 O 表示法,描述算法的计算复杂度; $P_{Accuracy}$ —准确率; $P_{Precision}$ —精确率; P_{Recall} —召回率; S_{F1} —F1 分数; n_{TP} —真正类且预测为正类的样本数量; n_{FN} —真正类但预测为负类的样本数量; n_{FP} —误将负类预测为正类的样本数量; n_{TN} —真实负类且预测为负类的样本数量。

参 考 文 献

- [1] 高文彬,李宜强,何书梅,等.基于荧光薄片的剩余油赋存形态分类方法[J].石油学报,2020,41(11):1406-1415.
GAO Wenbin, LI Yiqiang, HE Shumei, et al. Classification method of occurrence mode of remaining oil based on fluorescence thin sections[J]. Acta Petrolei Sinica, 2020, 41(11): 1406-1415.
- [2] 李志军,肖阳,田建章,等.渤海湾盆地冀中坳陷新领域、新类型油气勘探潜力及有利方向[J].石油学报,2024,45(1):69-98.
LI Zhijun, XIAO Yang, TIAN Jianzhang, et al. Potentials and favorable directions for new fields, new types of oil-gas exploration in Jizhong depression, Bohai Bay Basin[J]. Acta Petrolei Sinica, 2024, 45(1): 69-98.
- [3] 李宜强,何书梅,赵子豪,等.基于剩余油动用规律的高含水油藏水驱扩大波及体积方式实验[J].石油学报,2023,44(3):500-509.
LI Yiqiang, HE Shumei, ZHAO Zihao, et al. Experiment on enlargement of swept volume by water flooding in high water cut reservoir based on the remaining oil displacement law[J]. Acta Petrolei Sinica, 2023, 44(3): 500-509.
- [4] 孙先达.储层微观剩余油分析技术开发与应用研究[D].长春:吉林大学,2011.
SUN Xianda. New technique invention and application of characterizing microscopic remnant oil in reservoirs[D]. Changchun: Jilin University, 2011.
- [5] 李忠诚,鲍志东,王洪学,等.基于高压汞灯荧光显微观测的剩余油定量分析方法[J].石油钻探技术,2024,52(3):112-117.
LI Zhongcheng, BAO Zhidong, WANG Hongxue, et al. Quantitative analysis method of remaining oil based on fluorescence microscopic observation of high-pressure mercury lamp[J]. Petroleum Drilling Techniques, 2024, 52(3): 112-117.
- [6] 陈浩,滕奇志,何小海,等.基于几何形状特征的剩余油形态识别[J].微型机与应用,2017,36(01):18-21,28.
CHEN Hao, TENG Qizhi, HE Xiaohai, et al. Shape recognition of residual oil based on geometry feature[J]. Microcomputer & Its Applications, 2017, 36(1): 18-21, 28.
- [7] 程小龙,王正勇,滕奇志.基于KNN的剩余油形态识别[J].信息技术与网络安全,2020,39(1):104-107.
CHENG Xiaolong, WANG Zhengyong, TENG Qizhi. K-nearest neighbor method for recognizing the shape of residual oil[J]. Information Technology and Network Security, 2020, 39(1): 104-107.
- [8] 成璐璐.基于机器学习的微观剩余油赋存形态分类识别研究[D].大庆:东北石油大学,2023.
CHENG Lulu. Research on the classification and identification of microscopic residual oil reservoir storage based on machine learning[D]. Daqing: Northeast Petroleum University, 2023.
- [9] 李望奇,滕奇志,何小海,等.基于深度学习的剩余油形态分类[J].计算机系统应用,2023,32(12):224-232.
LI Wangqi, TENG Qizhi, HE Xiaohai, et al. Morphological classification of remaining oil based on deep learning[J]. Computer Systems & Applications, 2023, 32(12): 224-232.
- [10] 刘合,李艳春,贾德利,等.人工智能在注水开发方案精细化调整中的应用现状及展望[J].石油学报,2023,44(9):1574-1586.
LIU He, LI Yanchun, JIA Deli, et al. Application status and prospects of artificial intelligence in the refinement of waterflooding development program[J]. Acta Petrolei Sinica, 2023, 44(9): 1574-1586.
- [11] ZOU Xiangxi, ZHANG Yonghui, ZHANG Shuaiyan, et al. FPGA implementation of edge detection for Sobel operator in eight directions [C]//Proceedings of 2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS). Chengdu: IEEE, 2018: 520-523.
- [12] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[C/OL]//Proceedings of the Ninth International Conference on Learning Representations. ICLR, 2021. <https://arxiv.org/abs/2010.11929>.
- [13] 王然,吴涛,尤新才,等.玛湖凹陷二叠系风城组页岩油储层岩相特征及定量评价[J].石油学报,2023,44(7):1085-1096.
WANG Ran, WU Tao, YOU Xincan, et al. Petrographic characteristics and quantitative comprehensive evaluation of shale oil reservoirs in Permian Fengcheng Formation, Mahu sag[J]. Acta Petrolei Sinica, 2023, 44(7): 1085-1096.
- [14] 孙歧峰,李克昊,段友祥,等.基于卷积神经网络与特征聚类的荧光薄片分析方法[J].石油学报,2024,45(3):548-558.
SUN Qifeng, LI Kehao, DUAN Youxiang, et al. Fluorescent thin section analysis method based on convolutional neural network and feature clustering[J]. Acta Petrolei Sinica, 2024, 45(3): 548-558.
- [15] 王立辉,夏惠芬,韩培慧,等.剩余油分布的微观特征及其可动用程度的定量表征[J].岩性油气藏,2021,33(2):147-154.
WANG Lihui, XIA Huifen, HAN Peihui, et al. Microscopic characteristics of remaining oil distribution and quantitative characterization of its producibility[J]. Lithologic Reservoirs, 2021, 33(2): 147-154.
- [16] 毛国庆,滕奇志,吴拥,等.基于BP神经网络的剩余油形态识别[J].太赫兹科学与电子信息学报,2014,12(6):858-864.
MAO Guoqing, TENG Qizhi, WU Yong, et al. Shape recognition of remained oil based on BP neural network[J]. Journal of Terahertz Science and Electronic Information Technology, 2014, 12(6): 858-864.

- [17] WANG Wenhai, XIE Enze, LI Xiang, et al. Pyramid vision transformer: a versatile backbone for dense prediction without convolutions[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, IEEE, 2021: 548-558.
- [18] ZHAO Lin, SUN Xianda, LIU Fang, et al. Study on morphological identification of tight oil reservoir residual oil after water flooding in secondary oil layers based on convolution neural network[J]. Energies, 2022, 15(15): 5367.
- [19] WEI Zhe, MA Kaikuang. Contrast-guided image interpolation[J]. IEEE Transactions on Image Processing, 2013, 22(11): 4271-4285.
- [20] HAN Kai, WANG Yunhe, TIAN Qi, et al. GhostNet: more features from cheap operations[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, IEEE, 2020: 1577-1586.
- [21] MEHTA S, RASTEGARI M. MobileViT: light-weight, general-purpose, and mobile-friendly vision transformer[C]//Proceedings of the Tenth International Conference on Learning Representations. ICLR, 2022.
- [22] 康宇, 郝晓丽. 联合判别区域特征的细粒度视觉分类方法[J/OL]. 计算机工程与应用, 1-8[2024-01-29]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20240129.1011.010.html>. KANG Yu, HAO Xiaoli. Fine grained visual classification method for combined discriminative region features[J/OL]. Computer Engineering and Applications, 1-8[2024-01-29]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20240129.1011.010.html>.
- [23] TAN Mingxing, LE Q V. EfficientNetv2: Smaller models and faster training[C]//Proceedings of the 38th International Conference on Machine Learning. New York, PMLR, 2021: 10096-10106.
- [24] SANDLER M, HOWARD A, ZHU Menglong, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake Cit; IEEE, 2018: 4510-4520
- [25] HOWARD A, SANDLER M, CHEN Bo, et al. Searching for MobileNetV3[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul; IEEE, 2019: 1314-1324.
- [26] MA Ningning, ZHANG Xiangyu, ZHENG Haitao, et al. ShuffleNetV2: Practical guidelines for efficient CNN architecture design [C]//Proceedings of the 15th European Conference on Computer Vision-ECCV 2018. Munich; Springer, 2018: 122-138.
- [27] CHEN C F R, FAN Q F, PANDA R. CrossViT: cross-attention multi-scale vision transformer for image classification[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada. IEEE, 2022: 347-356.
- [28] LIU Ze, LIN Yutong, CAO Yue, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal; IEEE, 2021: 9992-10002.
- [29] ZHU Lei, WANG Xinjiang, KE Zhanghan, et al. BiFormer: vision transformer with bi-level routing attention[C]//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada; IEEE, 2023: 10323-10333.
- [30] WU Haiping, XIAO Bin, CODELLA N, et al. CvT: introducing convolutions to vision transformers[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal; IEEE, 2021: 22-31.
- [31] VASU P K A, GABRIEL J, ZHU J, et al. FastViT: a fast hybrid vision transformer using structural reparameterization[C]//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris; IEEE, 2023: 5762-5772.
- [32] 赵玲. 基于数字化孔道的聚驱后微观剩余油定量描述研究[D]. 大庆: 东北石油大学, 20119. ZHAO Ling. Study on quantitative description of micro residual oil after polymer flooding based on digital pore and throat[D]. Daqing; Northeast Petroleum University, 2019.
- [33] 赖文, 蒋璟鑫, 邱检生, 等. 南京大学岩石教学薄片显微图像数据集[DS/OL]. 科学数据银行, 2020(2020-07-28)[2024-05-14]. <https://cstr.cn/31253.11.sciencedb.j00001.00097>. LAI Wen, JIANG Jingxin, QIU Jiansheng, et al. A photomicrograph dataset of rocks for petrology teaching at Nanjing University[DS/OL]. V1. Science Data Bank, 2020(2020-07-28)[2024-05-14]. <https://cstr.cn/31253.11.sciencedb.j00001.00097>.
- [34] WAH C, BRANSON S, WELINDER P, et al. CUB-200-2011 (1.0)[DS/OL]. CaltechDATA, 2012(2022-04-11)[2024-05-14]. <https://data.caltech.edu/records/65de6-vp158>.

(收稿日期 2024-06-25 改回日期 2024-10-17 编辑 雷永良)