

基于注意力机制的岩石铸体薄片轻量化分割

杜睿山^{1,2}, 宋健辉¹, 孟令东²

(1. 东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318;

2. 油气藏及地下储库完整性评价黑龙江省重点实验室, 黑龙江 大庆 163318)

摘要: 储集层岩石铸体薄片的微观结构特征对研究储层的储渗能力、流体分布、采收率的大小、水驱油效率等均具有重要作用。岩石铸体薄片的分割是研究岩石微观结构特征的前提, 目前传统方法是通过染色剂对孔隙区域染色, 然后用阈值或连通域等方法进行分割, 这种方法准确率低且成本昂贵。基于深度学习的语义分割网络在不同的分割场景下都取得了很大进展。该文采用 DeepLabV3+ 网络作为模型框架, 首先, 针对语义分割网络参数多且在恢复空间细节方面表现欠佳等问题, 引入了轻量化特征提取网络, 优化原模型 Xception 特征提取网络的参数量; 其次, 优化残差结构, 减少参数计算量, 降低模型训练耗时; 最后, 为了弥补参数优化带来的精度损失, 在模型的高层特征图提取部分引入注意力机制 CBAM 模块, 以提高模型准确率。在岩石铸体薄片数据集上, 此方法与原模型相比准确率提高了 3.7 个百分点, 识别帧率提高了 106 个百分点。

关键词: 深度学习; 语义分割; 岩石铸体薄片; 轻量化网络; 注意力机制

中图分类号: TP391.1

文献标识码: A

文章编号: 1673-629X(2023)10-0128-07

doi:10.3969/j.issn.1673-629X.2023.10.020

Lightweight Segmentation of Rock Casting Sheet Based on Attention Mechanism

DU Rui-shan^{1,2}, SONG Jian-hui¹, MENG Ling-dong²

(1. School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China;

2. Key Laboratory of Oil and Gas Reservoir and Underground Gas Storage Integrity Evaluation, Daqing 163318, China)

Abstract: The microstructure characteristics of rock casting sheet of reservoir rocks play an important role in studying the reservoir permeability, fluid distribution, oil recovery and water displacement efficiency. The segmentation of rock casting sheet is the premise of studying the characteristics of rock microstructure. At present, the traditional method is to dye the pore area with dye, and then use threshold or connected domain methods to segment. This method has low accuracy and high cost. The semantic segmentation network based on deep learning has made great progress in different segmentation scenarios. We use DeepLabV3+ network as the model framework. Firstly, aiming at the problems of large number of parameters in the semantic segmentation network and poor performance in restoring spatial details, a lightweight feature extraction network is introduced to optimize the parameters of the original model's Xception feature extraction network. Secondly, the residual structure is optimized to reduce the amount of parameter calculation and model training time. Finally, in order to compensate for the accuracy loss caused by parameter optimization, the attention mechanism CBAM module is introduced in the high-level feature map extraction part of the model to improve the accuracy of the model. Compared with the original model, the accuracy of the proposed method is improved by 3.7 percentage points and the recognition frame rate is improved by 106 percentage points on the thin section data set of rock casting sheets.

Key words: deep learning; semantic segmentation; rock casting sheet; lightweight network; attention mechanism

0 引言

孔隙是流体赋存于岩石中的基本储集空间, 喉道则是连通孔隙的细小通道, 控制了流体在岩石中的流

通。储集层岩石的微观结构特征是影响储集层流体(油气水)的储集能力和开采油气资源的主控因素^[1], 对岩石微观结构特征的分析将为油气微观渗流机理的

收稿日期: 2022-11-05

修回日期: 2023-03-16

基金项目: 国家自然科学基金青年科学基金(41702156); 东北石油大学引导性创新基金(2020YDL-04)

作者简介: 杜睿山(1977-), 男, 硕士, 副教授, CCF 会员(51501M), 通讯作者, 研究方向为人工智能、机器学习; 宋健辉(1997-), 男, 硕士研究生, 研究方向为机器学习。

研究奠定关键基础^[2]。岩石铸体薄片的研究,也为面孔率、平均孔隙直径、平均表面比和平均孔喉比的研究提供了途径^[3]。

岩石铸体薄片是通过将有色液体压入岩石孔隙形成岩石样本,传统的图像分割识别方法主要借助染色剂染色,根据颜色的特性,通过 RGB 进行分割,A. Amankwah 和 Aldrich. C^[4]使用带形状标记的分水岭算法实现了岩石图像分割;Siebra、Hélio 等人^[5]使用模糊聚类的彩色纹理直方图实现了细岩的部分分割;魏雨等人^[6]通过阈值的方法,从二值图像中对孔隙连通进行快速的连通标记,进而对岩石图像信息进行分割和提取。但是由于岩石铸体薄片图像具有分辨率高、颗粒形状复杂、种类繁多、颜色相近处岩石颗粒与孔隙难以清晰划分的特性,这些传统的根据颜色特性分割的方法不仅在准确度上较低,而且无法对岩石薄片准确定量计算。并且不管是全局阈值分割还是局部阈值分割,这种分割的方式都没有考虑图像本身所反映的重要岩石物理信息,在处理图像时只考虑图像的灰度特征,比如在多种不同颜色岩石矿物分布的情况下进行划分,目标与背景的灰度有较大的重叠时,不能准确地将目标与背景分开。这将导致分割结果在理论上是最优的,但仍远低于实际。随着深度学习的发展,语义分割方法在岩石图像分割领域表现出巨大优势,韩巧玲等人^[7]通过 Fully Convolutional Networks (FCN) 在断层扫描图像中实现了孔隙分割,解决了因结构复杂多变而导致的准确率低下等分割问题,为岩石分割方面带来了新的思路。尚福华等人^[8]通过 Unet 完成了对岩石铸体图像的孔隙提取。虽然语义分割模型开始逐渐在岩石铸体图像分割中体现优势,但是现有网络存在模型参数庞大、内存依赖大的问题,使得运行结果缓慢,计算效率低下;并且由于岩石铸体薄片颗粒小的特性,编码器网络中容易丢失部分细节,从而导致分割边界损坏,结果较差。针对现有模型存在的问题,提出一种融合轻量化和注意力机制的岩石铸体薄片图像分割模型,以达到模型轻量化且实现对细小颗粒分割的问题,提升模型精度和速度。

1 岩石铸体薄片分割模型

语义分割是一种像素级的分割,大多数用于语义分割的深度网络是基于编码器-解码器结构搭建的。编码器网络利用卷积层和池化层提取有意义的特征,解码器网络利用反卷积层恢复原始分辨率。该文选用具有代表性且识别精度和性能相对较高的的语义分割模型 DeepLabV3+。DeepLabV3+ 模型的主干网络 Xception 提取网络参数庞大且运算量大,为了提高卷积运算的效率,该文采用在数学运算量方面具有良好

性能的轻量级网络 MobileNet V2,以优化参数运算量从而提高效率。其次,原有的经典分割技术在细节处的分割性能较低,特别是对于岩石铸体薄片中的细小颗粒,由于低分辨率和语义信息不足等问题,下采样会导致空间信息丢失,从而导致分割不准确。为此,该文在编码特征提取阶段引入自注意力机制,通过对特征提取阶段的权重分配来提高细小颗粒特征提取的敏感程度,以提高网络识别精度。

1.1 基础 DeepLabV3+网络

结合岩石铸体薄片小颗粒较多的问题,该文采用在 VOC 公开数据集模型分割能力相对较好的 DeepLabV3+ 模型,DeepLabV3+ 模型^[9]是由 FCN 模型^[10]发展而来的语义分割深度学习模型。基本 FCN 模型的应用存在两个技术障碍:下采样和空间不变性。下采样方法会降低分辨率,特别是在高层。空间不变性意味着从分类器获得以对象为中心的决策需要空间转换的不变性,这必然限制了空间的准确性。DeepLabV1 采用卷积算法和条件随机场 (CRF) 分别处理下采样和空间不变性。DeepLabV2 使用 atrous 空间金字塔池^[11] (ASPP) 在多个尺度上稳健地分割对象。为了对多尺度信息进行编码,DeepLabV3 提出了级联模块和改进的 ASPP 模块。DeepLabV3+ 通过添加一个简单而有效的解码器模块来优化分割结果,特别是沿着对象边界。一方面,它通过 ASPP 替代深层池化层保留了更多的细节,且不增加训练参数,提高了模型的预测性能。通过多尺度信息采样,获得信息量不同的目标样本,增强了模型的鲁棒性;另一方面,利用空洞卷积可以提取不同尺度的特征,在不损失信息的情况下,加大了感受野,并能有效扩展每个卷积输出的信息,其提取的特征信息更加丰富且蕴含更多的全局特征^[12]。DeepLabV3+ 的网络结构如图 1 所示。

DeepLabV3+ 网络虽然分割效果相对较好,但仍存在网络参数庞大和模型在细节分割表现欠佳等问题,针对现有问题,该文在编码部分和深层提取部分做出调整,以实现轻量化和提高模型准确度。

1.2 改进编码部分网络

针对模型网络参数庞大的问题,为了在速度和准确性之间找到一个好的平衡点,提高对岩石铸体薄片分割的效率,满足实时分割需求,该文选用 MobileNet V2 作为提取特征的主干网络,以解决原 Xception 网络参数庞大的问题,同时最大限度地减少准确性的损失。它具有一个非常重要的特点就是使用了深度可分离卷积和 Inverted resblock,整个结构都由 Inverted resblock 组成。初始的网络首先利用 1×1 卷积进行扩张,增加特征层通道数,然后利用 3×3 深度可分离卷积对每个通道进行特征提取,并且降低运算量,最后再利用 1×1

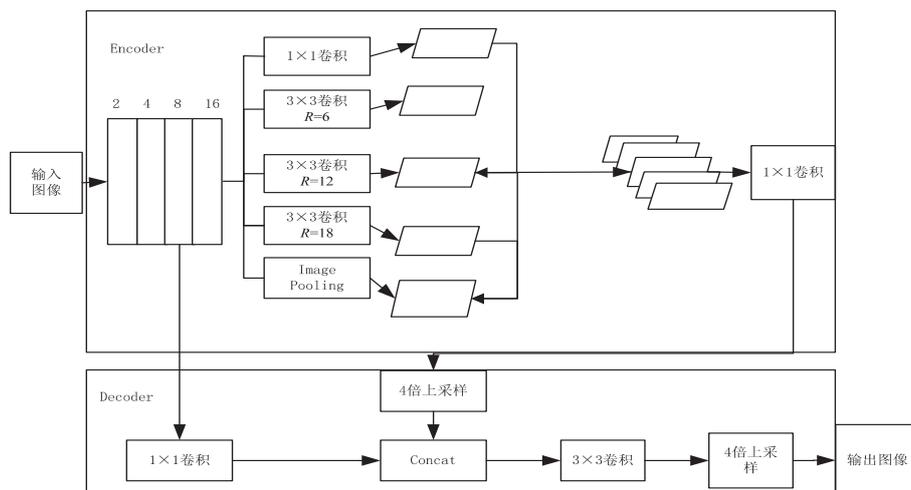


图 1 DeepLabV3+网络结构

卷积进行压缩,把高维特征映射到低维空间去,其中压缩后的激活函数采用 Linear 线性函数防止进一步破坏压缩后的特征,然后主干网络并联一个残差边,输入输出直接相接,原始的网络结构如图 2 所示。

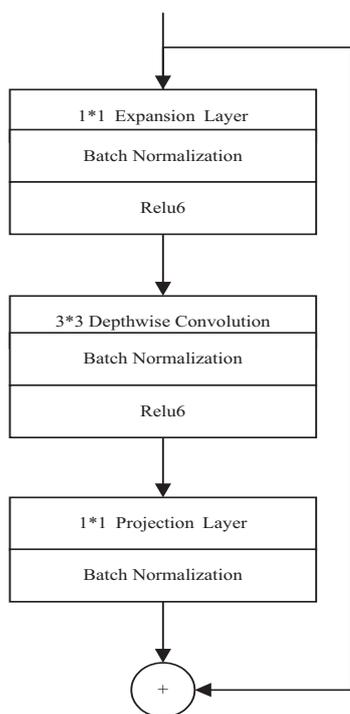


图 2 Bottleneck 结构

其中 BN 和激活层主要提升网络的非线性能力和网络的稳定性。然而,激活层的输入是残差网络中上一个残差结构的复合输出,它受此前残差网络的影响,已经具有较强的非线性能力,分布也比较稳定。所以第一个激活层对于一个残差结构分支来说几乎没有增强其线性能力的作用,尤其是对于深度较深的网络会降低训练速度。第二个激活层的输入是残差结构中第一个卷积层的输出,其分布会受到残差结构中第一个卷积层参数更新的直接影响。

为了简化网络参数和模型的训练耗时,该文还采

用了简化残差结构,其中图 3 为删除第二个激活层所得。与传统的残差结构相比,这种残差结构减少了残差分支中激活层的数量,节省了训练过程中的内存开支;还减少了残差分支中 BN 层对输入期望和方差的计算量,降低了模型训练耗时。其中 DeepLabV3+原模型参数数量为 41,034,394,最终经过优化原模型 Xception 和简化残差之后的模型参数数量为 22,130,628。

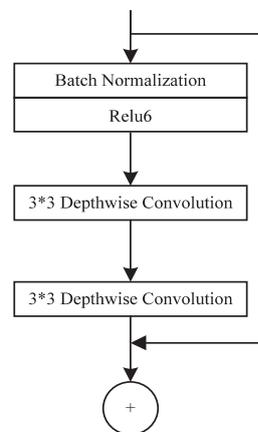


图 3 改进 Bottleneck 结构

1.3 注意力机制模块

减少网络参数,虽然会提高模型速度,但势必会影响模型分割的准确程度,所以为了弥补模型优化带来的精度损失,引入注意力机制模块提高岩石铸体薄片分割的准确率。注意力机制是 Treisman 和 Gelade 提出的一种信号处理机制^[13],在视觉任务中,注意力机制首先计算代表特征重要程度的注意权重,然后利用权重值从输入的特征映射中提取出信息量更大的特征,它如今在 CV 领域被广泛采用,并在各种基于深度学习的计算机视觉应用^[14-16]表现出显著的性能。由于引入了轻量级网络,减少了模型的参数量,会导致模型准确率下降,所以引入注意力机制来提高模型准确率。

Convolutional Block Attention Module (CBAM)关

注网络中重要的特征,抑制网络中不重要的特征,它不仅通道中含有丰富的注意力信息,特征图的像素点也包含大量注意力信息^[17],有效地提高了神经网络的性能。为了提高岩石铸体薄片分割模型的准确性,现将 CBAM 加入到编码器的特征提取网络中。CBAM

为两个模块,包括通道注意模块和空间注意模块,前者关注点为输入的特征,后者重点关注要识别的相关区域。两个子模块相互串联形成,分别提高了网络的通道和空间的重要性,并且容易嵌入到各个网络中,模型如图 4 所示。

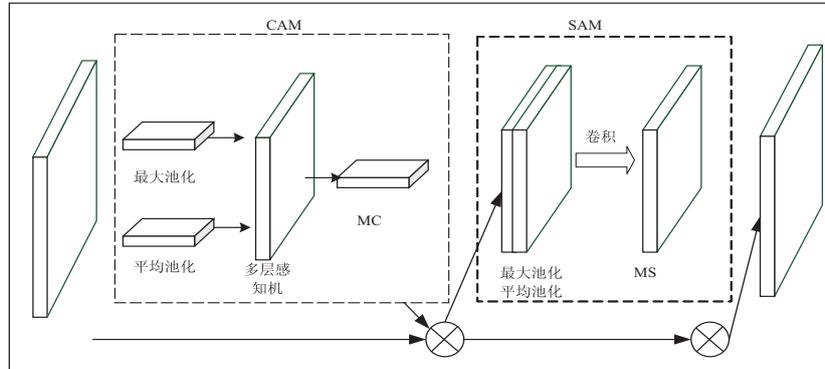


图 4 CBAM 模型结构

通道注意力模块分别采用 Ave-pooling 和 Max-pooling 来压缩特征图的空间维度得到全局平均池化层特征和全局最大池化特征,最大池化只考虑最大的元素,忽略其他元素,并保留图像的更多纹理信息,平均池化计算池化区域内所有元素的平均值,以保留更多的图像背景信息,最大池化和平均池化的过程如图 5 所示,通道注意力层表示为 F_{avg}^c 和 F_{max}^c 。然后,这两个特征传到一个共享网络中,以产生通道注意力图 M_c 。共享网络由一个多层感知器 (MLP) 组成,其中有一个隐含层。为减少参数开销,隐藏层的激活大小设为 $R/C = r \times 1 \times 1$,其中 R 为下降率。将输出的特征向量使用求和进行合并,过程如公式(1)所示,其中 σ 表示 sigmoid 函数。

$$M_c(F) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \quad (1)$$

空间注意力模块的关注点是对通道注意力的补充。它更加侧重于关注空间位置信息^[18],为了计算空间注意力,将串联的通道注意力得到的输出特征作为通道的输入,沿通道维度分别采用平均池化和最大池化,分别表示为 F_{max}^s 和 F_{avg}^s 。然后对这两个结果进行堆叠,此时获得了输入特征层每一个特征点的位于 0~1 之间的权值。在获得这个权值后,将这个权值乘上原输入特征层即可,过程如公式(2)所示:

$$M_s(F) = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \quad (2)$$

其中, F_{avg}^s 和 F_{max}^s 的大小均为 $1 \times H \times W$ 。 σ 表示 sigmoid 函数, $f^{7 \times 7}$ 表示一个滤波器大小为 7×7 的卷积运算。

对得到的两个注意力模块串联后应用在编码器的深层特征提取部分,增强网络泛化能力,强化特征图通道和位置信息的特征提取,其最终改进的网络模型如图 6 所示。

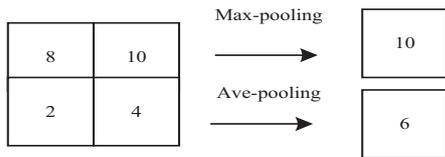


图 5 最大池化、平均池化

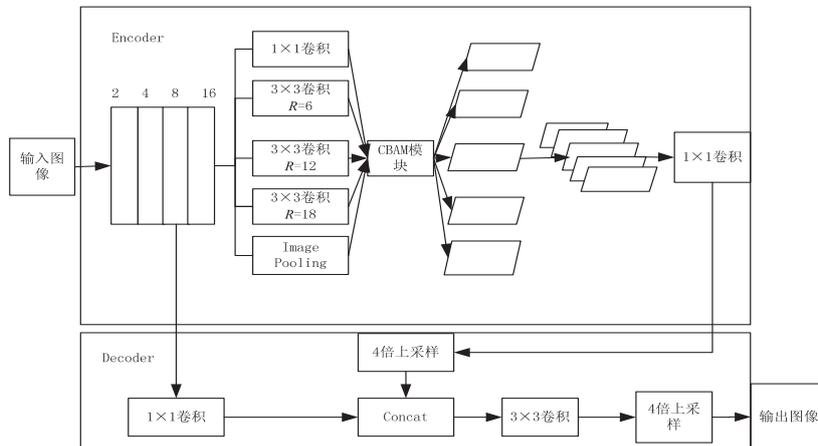


图 6 岩石铸体薄片分割模型

2 实验结果与分析

2.1 数据集

数据集包括原始图像与标签图像,本数据集来自东北石油大学非常规油气研究院的岩石铸体薄片数据,将原始图像在高清显微镜下采集薄片样本,由于岩石铸体薄片图像分辨率较高,整个模型的计算量会溢出。本研究采用图像分块等技术来降低计算量,通过分块将一张大的岩石铸体薄片图像分割成 $256 * 256$ 的小数据图像样本送入网络,然后对结果进行拼接。并且岩石铸体薄片数据集获取困难且成本昂贵,通过偏光、旋转、拼接切割等方法扩充数据,在节省成本的同时还能提高泛化能力。最终得到 3 400 张图像数据,然后通过图像处理技术对铸体薄片图像锐化,通过锐化技术使图像的一些细节信息以及边缘得到增益,目标物体的边缘鲜明,以便于提取特征。最后通过 Labelme 进行标注获得标签图像,构建出一套岩石铸体薄片训练数据集,如图 7 所示,分别为原图和标签图。

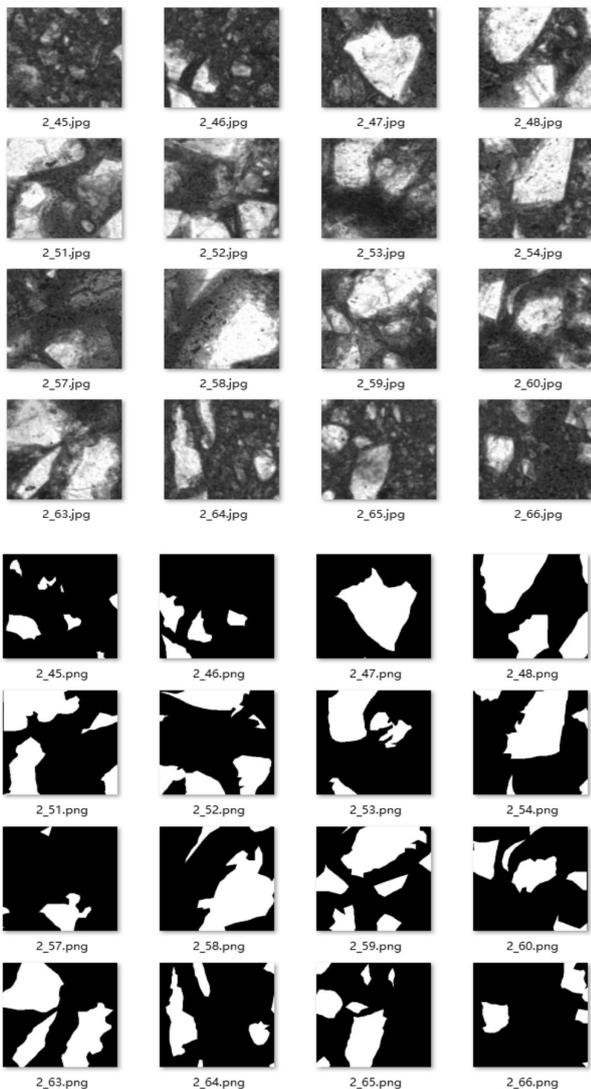


图 7 数据集构建效果

2.2 实验环境

实验在 Windows 环境下进行,具体配置和训练环境如表 1 所示。

表 1 实验环境配置

名称	参数
操作系统	Windows 10
深度学习框架	Tensorflow
GPU	GTX1030
算法语言	Python
内存	8 GB

2.3 评价指标

为了评价模型结果的好坏,采用 PA (Pixel Accuracy)、IOU (Intersection over Union) 和 FPS (Frames Per Second) 作为该模型的评价指标。

FPS 表示图像每秒传输的帧数,通过 FPS 指标来衡量模型的速度。PA 是分别计算每个类别分类正确的像素数占该类别像素总数的比值。IOU 表示交并比,指模型对某一类别预测结果和真实值的交集与并集的比值,背景部分在图像上覆盖了所有像素的很大比例,反映了分类模型的不平衡。因此选择对正负样本不平衡效果好的 Dice 函数来验证模型的准确性,能够更好地应对二分类预测问题,dice 函数由准确率和召回率决定,如式(3)~(6)所示:

$$L_{\text{precision}} = \frac{TP}{TP + FP} \quad (3)$$

$$L_{\text{recall}} = \frac{TP}{TP + FN} \quad (4)$$

$$IOU = \frac{TP}{TP + FP + FN} \times 100\% \quad (5)$$

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (6)$$

其中,TP 表示被正确分类为岩石颗粒的像素数;FP 表示背景被错误分类为预测样本的像素数;FN 表示岩石颗粒像素被错误地分类为背景像素的数量。那么 Dice 和 DiceLoss 差异度公式如下:

$$\text{dice} = \frac{2TP}{2TP + FP + FN} \quad (7)$$

$$P_{\text{diceLoss}} = 1 - \frac{2L_{\text{precision}} * L_{\text{recall}}}{L_{\text{precision}} + L_{\text{recall}}} \quad (8)$$

2.4 实验分析

将训练集和验证集划分为 9 : 1,下采样因子 downsample_factor 设置为 8,进行三次下采样,由于只分割岩石和背景,将 num_classes 设置为 2,Batchsize 设置为 4,设置学习率为 5×10^{-4} ,学习率衰减率为 0.1,momentum 超参数为 0.9,损失函数为 DiceLoss。输入薄片数据后进行训练,进行 80 次 epoch 迭代训练,采用梯度下降法对模型训练,随着训练次数增加损

失变化曲线逐渐趋于稳定状态,损失最终收敛至最小值,如图 8 所示。

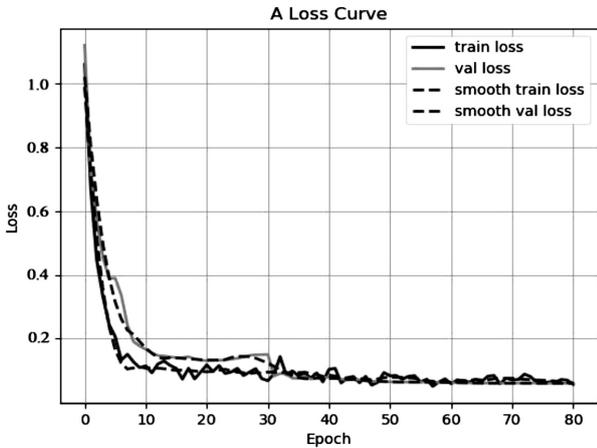


图 8 训练 Loss 图

对比实验是评价模型质量的关键因素,为了比较模型的实际效果,将岩石铸体薄片测试图像输入到该文训练的模型中,并与传统的阈值分割方法、PSPNet、经典神经网络模型 Unet 和未改进的 DeepLabV3+模型进行比较。与传统方法相比,分割精度有了很大的提高,特别是在背景信息丰富复杂的区域和岩石目标主体区域部分。与未优化的深度学习分割方法相比,此模型更加快速,也提高了岩石细节的分割精度。综合实验结果表明,该方法在多个方面优于其他模型,在提高模型速度的同时也提高了结果的准确性,所以该方法具有良好的推广性,评价指标的结果如表 2 所示。

表 2 不同模型性能比较

分割方法	PA/%	IOU/%	FPS
传统阈值分割	84.2	67.8	42.2
PSPNet	86.2	79.8	24.2
Unet	89.4	81.4	26.4
DeepLabV3+	92.1	86.5	17.5
Ours	93.2	90.2	36.2

融合注意力和轻量化的 DeepLabV3+模型对于岩石铸体薄片图像的处理性能有了较大加强,与传统分割模型相比,测试集的 IOU 和 PA 均得到提升,相较于未改进的 DeepLabV3+模型,IOU 提升了 3.7 个百分点,像素精确度提高了 1.1 个百分点,由于引入了轻量化模型,相对原模型识别帧率提高了 106 百分点。

图 9 表示原始图像和不同模型的分割效果,以便可视化模型的图像分割结果。

通过对比不同模型方法,传统的阈值分割方法在分割效果方面最差,其中最主要的原因在于阈值分割对于颜色的依赖性太强,尤其在分割这种错综复杂的岩石铸体薄片图像方面,孔隙和岩石边缘颜色相近,导致分割效果差;PSPNet、Unet 网络和传统 DeepLabV3+

网络均出现细小岩石分割不连续且出现断点的问题,对细节处的处理效果也较差。文中网络在解决分割精度的同时改进了主干提取网络,使得模型速度更快、精确度更高。

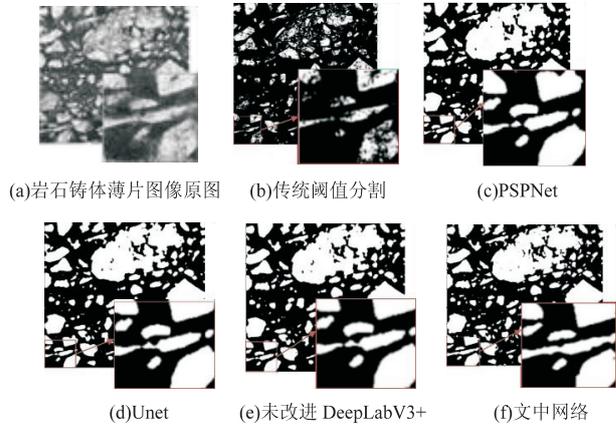


图 9 不同模型的分割效果

2.5 模型泛化性分析

为了验证文中算法的可行性,选取 VOC2007 公开数据集验证模型泛化能力。如图 10 所示,展示了不同算法与文中改进轻量化与添加 attention 后的效果对比。可以看出无论何种数据集,对模型好坏能力的评价更体现在对图像细节的识别能力上。在不同的场景下,该文提出的模型相对较好,如飞机尾翼的细节和在多目标场景下的摩托车驾驶员的脚部细节,相对来说有着较为准确的分割效果。

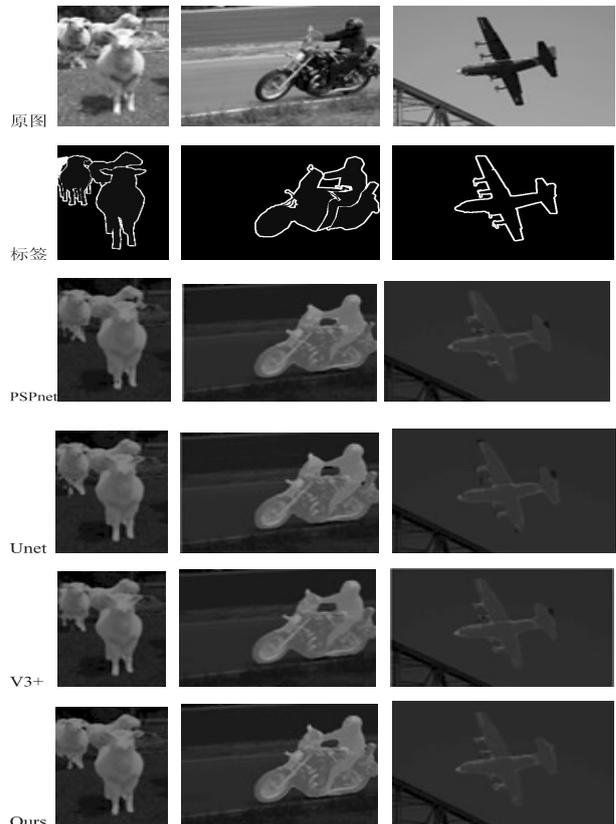


图 10 不同模型在 VOC 公开数据集上的效果展示

3 结束语

针对传统阈值方法分割准确率低且成本昂贵等问题,采用目前流行的语义分割技术。并针对这些传统模型网络参数量多、运行速度慢并且在恢复空间细节方面仍然表现欠佳等问题,采用分割精度较好的 DeepLabV3+模型作为基本分割模型框架。为了提高模型的速度和准确率,引入了轻量化特征提取网络,优化原模型的参数量,然后通过引入注意力机制,以提高模型准确率,弥补因减少参数而导致的精度下降问题。实验结果表明,该方法在原网络上参数大大减少,运行处理速度更快,并且与传统分割方法和网络相比有更高的准确率和边缘敏感度,为后续对岩石铸体薄片的分割研究提供了一种思路。

参考文献:

- [1] LI Xizhe, GUO Zhenhua, WAN Yujin, et al. Geological characteristics and development strategies for Cambrian Longwangmiao Formation gas reservoir in Anyue gas field, Sichuan Basin, SW China[J]. Petroleum Exploration and Development, 2017, 44(3): 428-436.
- [2] LIN W, LI X, YANG Z, et al. A new improved threshold segmentation method for scanning images of reservoir rocks considering pore fractal characteristics[J]. Fractals, 2018, 26(2): 1840003.
- [3] 赵明, 郭志强, 卿华, 等. 岩石铸体薄片鉴定与显微图像分析技术的应用[J]. 西部探矿工程, 2009, 21(3): 66-68.
- [4] AMANKWAH A, ALDRICH C. Rock image segmentation using watershed with shape markers[C]//2010 IEEE 39th applied imagery pattern recognition workshop (AIPR). Washington: IEEE, 2010: 1-7.
- [5] SIEBRA H, CARVALHO B M, GARDUÑO E. Fuzzy clustering of color textures using skew divergence and compact histograms; segmenting thin rock sections[J]. Journal of Physics: Conference Series, 2015, 574(1): 012116.
- [6] 魏雨. 融合遗传算法与 OSTU 的岩石薄片图像分割方法研究[J]. 现代信息科技, 2022, 6(16): 1-5.
- [7] 韩巧玲, 赵玥, 赵燕东, 等. 基于全卷积网络的土壤断层扫描图像中孔隙分割[J]. 农业工程学报, 2019, 35(2): 128-133.
- [8] 尚福华, 张瑞, 杜睿山, 等. 基于 U-Net 岩石薄片的孔隙提取方法[J]. 海南大学学报: 自然科学版, 2022, 40(2): 134-141.
- [9] CHEN L C, ZHU Y, PAPANDEOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European conference on computer vision (ECCV). Munich: Springer, 2018: 801-818.
- [10] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Boston: IEEE, 2015: 3431-3440.
- [11] CHEN L C, PAPANDEOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [12] 刘文斌, 崔学英, 程小霞. 基于多尺度残差网络构架的低剂量 CT 图像去噪[J]. 宁夏师范学院学报, 2020, 41(10): 61-68.
- [13] MNIH V, HEES N, GRAVES A. Recurrent models of visual attention[J]. Advances in Neural Information Processing Systems, 2014, 27(3): 2204-2212.
- [14] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seoul: IEEE, 2019: 3146-3154.
- [15] PANG Y, XIE J, KHAN M H, et al. Mask-guided attention network for occluded pedestrian detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. Seoul: IEEE, 2019: 4967-4975.
- [16] HUANG L, WANG W, CHEN J, et al. Attention on attention for image captioning[C]//Proceedings of the IEEE/CVF international conference on computer vision. Seoul: IEEE, 2019: 4634-4643.
- [17] 张宸嘉, 朱磊, 俞璐. 卷积神经网络中的注意力机制综述[J]. 计算机工程与应用, 2021, 57(20): 64-72.
- [18] 王囡, 侯志强, 蒲磊, 等. 空洞可分离卷积和注意力机制的实时语义分割[J]. 中国图象图形学报, 2022, 27(4): 1216-1225.