

基于改进 Unet 的矿石图像分割

曾中华¹ 曹东²

(1.河南理工大学物理与电子信息学院 焦作 454000;2.无锡东如科技有限公司工业智能研究院 无锡 214000)

摘要: 为了应对矿产领域的矿石识别困难、识别设备成本高等问题提出了改进的 Unet 矿石图像分割算法,首先对 EfficientNetV2 加以改造作为模型的主干网络提取矿石特征;其次,引入 MBconv 模块作为解码器中的基本模块,加强特征提取能力;采用 CA 注意力模块替换 SE 注意力模块,保留更多空间位置信息;最后使用滤波器响应归一化(FRN)层替换常用的批量归一化(BN)层,避免模型性能受批量大小的影响。基于 FeM 和 Cu 数据集实验结果表明,PA 分别为 96.58% 和 95.39%,Miou 分别为 92.8% 和 90.43%,F1_score 分别为 95.15% 和 93.47%,Efficient_unet 模型参数相比 Unet 减少了 60.4%,推理速度提升了 19.23%,可达 21.7 fps,所提模型在精度和速率方面均优于现有经典分割模型,同时表现出良好的泛化性能。

关键词: Unet;EfficientNetV2;矿石图像分割;注意力;滤波器响应归一化

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 510.4

Ore image segmentation based on improved Unet

Zeng Zhonghua¹ Cao Dong²(1. School of Physics and Electronic Information, Henan Polytechnic University, Jiaozuo 454000, China;
2. Wuxi Dongru Technology Co., Ltd., Institute of Industrial Intelligence, Wuxi 214000, China)

Abstract: Addressing the challenges of ore recognition in the mining industry and the high cost of recognition equipment, we propose an improved Unet ore image segmentation algorithm. Firstly, we modify EfficientNetV2 as the backbone network of the model to extract ore features. Secondly, we introduce the MBconv module as the decoder, enhancing the feature extraction capability. We then replace the SE attention module with the CA attention module to retain more spatial position information. Finally, we substitute the commonly used Batch Normalization (BN) layer with the Filter Response Normalization (FRN) layer to prevent model performance from being affected by batch size. Experimental results based on FeM and Cu datasets demonstrate that our proposed model achieves a PA of 96.58% and 95.39%, an MIoU of 92.8% and 90.43%, and an F1 score of 95.15% and 93.47%. Compared to Unet, the Efficient_Unet model parameters are reduced by 60.4%, and the inference speed is improved by 19.23%, reaching 21.7 frames per second. Our proposed model outperforms existing classical segmentation models in terms of accuracy and speed, exhibiting strong generalization performance.

Keywords: Unet;EfficientNetV2;ore image segmentation;attention;filter response normalization

0 引言

众所周知,矿石资源是工业生产的基本原料,矿石成分的区别是生产利用的基础。早期人们用肉眼根据外观和颜色来识别矿石,但效率较低,适用于常见且明显的特征和大型结晶矿物。制作矿石薄片样品^[1],用矿石显微镜观察矿石的微观结构来鉴别矿石变得越来越普遍。但其缺点是容易受主观因素影响,需要丰富的专业知识。由于该过程的高度可重复性,已经开发了许多自动识别系统^[2]。这些自

动化识别系统主要用扫描电子显微镜(scanning electron microscope,SEM)和 X 射线色散能谱仪(energy dispersive spectrometer,EDS)进行矿物识别。然而基于 SEM 的系统很难区分化学成分接近的矿物或主要元素相同的氧化物和氧羟基化合物^[3]。SEM 系统在背射式电子成像(back scattered electron imaging,BSE)图像中对赤铁矿和磁铁矿进行分割是可能的,但需要很强的图像对比度,这阻碍了对其他矿相的分割^[4]。同时由于 SEM 系统和 EDS 设备过于昂贵,成本高且限制于有限的矿物,仪器工作环境严苛,通

常采用反射光显微镜对矿石进行分析。对于具有相似化学成分矿物在 SEM 和 EDS 中无法区分,但在反射光显微镜上可以通过它们的颜色进行区分。人眼直接对矿石进行区分需要丰富经验且效率较低,经典的图像分析方法根据边界查找矿石颗粒与非矿石颗粒的间隙,分割精度受到边界轮廓清晰的限制^[5]。如何准确且迅速地对矿石图像进行分割,仍是矿产行业的一个技术难题。

随着近些年人工智能技术的发展,深度学习也逐渐应用到矿产方面。2019 年彭伟航等^[6]采用 Inceptionv3 模型对 16 种矿物晶体进行识别,平均准确率 86%。郭艳军等^[1]训练基于 ResNet18 网络模型来识别 5 种常见矿物,分类准确率为 89%,Filippo 等^[7]提出 Deeplabv3+ 变体模型对环氧树脂中的矿石颗粒进行分割,总体准确率达到 90% 以上。环氧树脂通常被用作矿石薄片的封装材料,而石英是矿石常见的成分,石英和树脂具有十分相近的反射率,人眼虽然可以区分树脂和石英,但当边界不清晰时无法准确描绘出轮廓,对此 Iglesias 采用 CNN 模型对反射光显微镜图像中的石英和树脂进行区分^[8],准确率达 95%。张娟娟等^[9]提出一种改进的 Soft-Mask RCNN 模型对碳纤维 SEM 图像分割,交并比达 90.6%。Unet 模型最早用于医学界神经细胞分割^[10],采用编码器解码器架构,在只有少量数据的情况下取得了很好的分割效果。矿石分割数据集制作困难,可用数据少,因此 Unet^[11]相比其他网络模型有更好的分割效果,本文借鉴原始 Unet 网络模型,提出了分割效果更好,参数量更少,推理速度更快的改进模型来分割树脂中矿石,为基于反射光显微镜的自动化分析系统提供可能性。

1 本文算法模型

本文算法在 Unet 网络模型基础上进行改进,主要工作体现在以下几个方面:

1) 引入特征提取能力更强的 EfficientNetV2 替代原始 Unet 模型的类似于 VGG16 结构作为主干特征提取网络,降低了模型参数量。

2) 使用坐标注意力 (coordinate attention, CA) 注意力模块对移动翻转瓶颈卷积 (mobile inverted bottleneck convolution, MBconv) 进行改进,命名为 CAMBconv 模块,保留更多空间信息。

3) 采用滤波器响应归一化 (filter response normalization, FRN) 替代常用的批量归一化 (batch normalization, BN) 归一化,减小对批量大小的依赖。

1.1 EfficientNetV2 主干网络

EfficientNetV2^[12] 是一种相比一般网络训练速度快、参数效率更高的网络,由谷歌实验室提出。EfficientNetV2 主要由 MBConv 和融合移动翻转瓶颈卷积 (fused-mobile inverted bottleneck convolution, Fused-MBConv) 组合而成。两个基本模块都采用了倒残差结构,MBConv 首先使用 1×1 的 pointwise 卷积对输入通道维度进行扩展,之后

使用 3×3 的 depthwise 卷积,中间加了个压缩和激励 (squeeze and excitation, SE) 注意力模块^[13],通道压缩比例为 0.25,然后再用 1×1 的 pointwise 卷积压缩通道维,需要注意的是前面的激活函数用的是 SiLu,最后的 1×1 pointwise 卷积激活函数为 linear,防止信息丢失,整个过程是“扩张-卷积-压缩”。由于 depthwise 卷积不能改变通道数,因此特征提取受限于输入的通道数,所以将通道数先提升上去,使得可以提取更多特征。如果卷积步长都为 1,则首尾使用捷径分支相连,实现恒等映射,防止网络加深造成的梯度消失。

EfficientNetV2 使用 Fused-MBConv 取代 EfficientNet^[14] 中的原始 MBConv。发现当在早期阶段时, Fused-MBConv 可以提高训练速度,同时在参数和 FLOP 上开销很小。但是,如果所有块都使用融合 MBConv,那么它会显著增加参数和 FLOP,同时也减慢训练速度。因此在前期使用 MBConv,后期使用 Fused-MBConv,使得推理速度和参数量两者兼顾,虽然增加了参数量,但是如果能够组合好 Fused-MBConv 和 MBConv,那么其训练速度反而可以提升。

Fused-MBConv 将 MBConv 中的深度卷积 3×3 和扩展卷积 1×1 融合为单个常规的 3×3 。同时由于应用在前期层删掉了 SE 注意力模块。MBConv 模块与 Fused-MBConv 模块对比结构如图 1 所示。

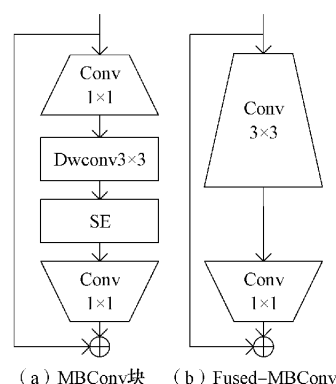


图 1 MBConv 与 Fused-MBConv 结构对比

由于 EfficientNetV2 是在图像分类任务上对网络进行优化的,因此对 efficientnetv2_base 进行了部分网络参数的改动,使之更适合本次分割任务。其具体网络结构如表 1 所示,减少了网络的深度,加宽了通道维度。表中每个 stage 描述的均为第 1 个步距,例如 Stage = Block4, layer = 4 的层,除了第 1 层 stride = 2 其他的默认都是 1,重复 4 层。

1.2 CAMBconv 模块

SE 注意力通过对各通道赋予不同的权重来形成注意力,分为两步:压缩,激励。首先通过 1 个自适应全局平均池化的操作获取全局信息。将 1 个 c 通道, $h \times w$ 的特征图,压缩成 c 通道 1×1 。再通过两个全连接层组成的瓶颈结构,第 1 个全连接层降维,接 SiLu 激活函数,第 2 个全连

表 1 编码器网络结构

阶段 (stage)	算子 (op)	步长 (stride)	通道 (out_c)	扩张 (expand)	层数 (layer)
Stem	Conv3×3	2	32	1	1
Block1	Fused-MB	2	64	4	2
Block2	Fused-MB	2	128	4	2
Block3	MBCConv, SE0.25	2	192	4	2
Block4	MBCConv, SE0.25	2	288	6	4

接层恢复到原来维度,后接 Sigmoid 激活,生成 0~1 的权值,代表每个通道的重要性,将权值逐通道相乘,对通道进行加权,形成了通道上的注意力。其网络结构如图 2 所示。由于对通道的特征图采用平均池化,丢失了空间信息。对通道压缩控制模块的参数量,不能很好地捕获通道间的信息。

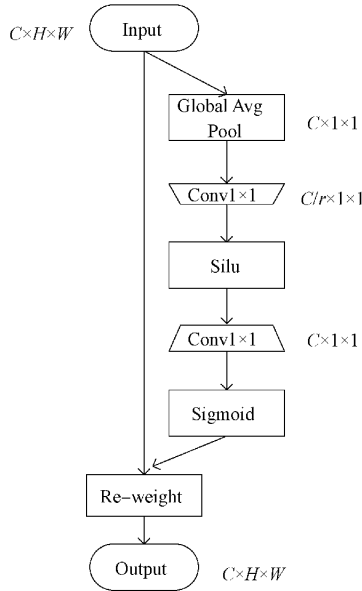


图 2 SE 注意力结构

CA^[15]将位置信息加入到通道注意力,在轻量级网络上取得了比较大的成功,它既能捕获通道之间的依赖也能很好地建模位置信息和长程依赖,实验表明其在图像识别、目标检测和语义分割任务上都有不错的改进,由于位置信息的加入,在这种依赖位置信息的密集预测任务效果更好。

通道注意力通常用 avgpool 对空间信息应用全局编码,并将其存储到通道维度中,位置信息被丢弃,然而对于视觉任务位置信息十分重要。CA 注意力对特征图分别沿横坐标和纵坐标两个方向进行特征编码,保留精确的位置信息。网络结构如图 3 所示。输入 $c \times h \times w$ 的特征图,分别在横坐标和纵坐标两个方向做平均池化,经转置后拼接为 $c \times (h+w) \times 1$ 的特征图,经过全连接层,压缩维度,接归一化层和 SiLu 激活函数层,再分拆为 $c \times h \times 1$ 和 $c \times w \times 1$ 特征图,经过 sigmoid 函数,生成 0~1 的权值,再扩展为 $c \times h \times w$ 的特征图,与输入相乘,形成带注意力的特

征图。本文使用 CA 注意力对 MBCConv 中 SE 注意力进行替换,命名为 CAMBConv 模块。

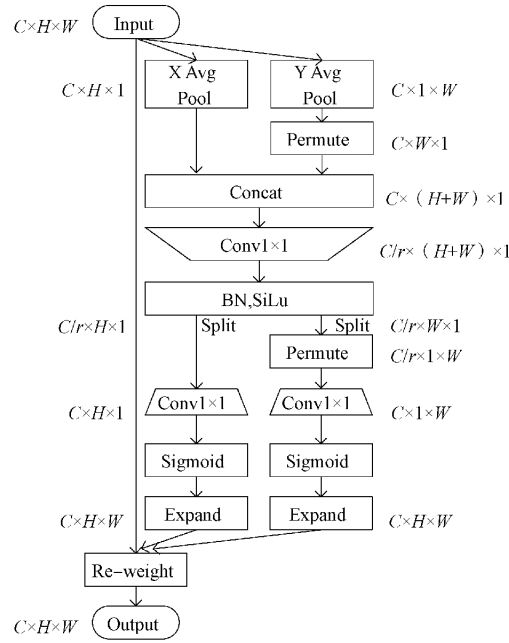


图 3 CA 注意力结构图

1.3 FRN 正则化层

在深度学习中,为使各层数据满足独立同分布,避免内部协变量偏移 (internal covariate shift, ICS) 问题^[16],归一化层的使用成为了神经网络的标配。BN 是最常用的一种归一化层。

训练时在每一 BN 层中,对 batch 中的每个样本的同一通道,计算并保存它们的均值和方差,利用计算出的均值和方差对数据进行归一化,运用白化的思想,如式(1)所示,使得归一化后的数据均值为 0,方差为 1,设置 γ 和 β 两个可学习参数,利用这两个参数对归一化的数据进行缩放和平移就可以重构数据,如式(2)所示。通过计算训练集所有小批量的均值和方差的期望值,以此来作为推理过程中该 BN 层的均值和方差。

$$\hat{x}^{(k)} = \frac{x^{(k)} - E[x^{(k)}]}{\sqrt{\text{Var}(x^{(k)})}} \quad (1)$$

$$y^{(k)} = \gamma^{(k)} \hat{x}^{(k)} + \beta^{(k)} \quad (2)$$

应用 BN 归一化层能够使用更大的学习率,可以在网络的任意层做归一化,不仅加快了收敛速度,而且能够避免梯度消失与梯度爆炸的问题,使得训练深层网络更稳定。

但由于现在模型和任务越来越复杂,尤其对于语义分割任务,通常没有足够显存去存放足够多的批量数据,而 BN 层对样本间的依赖性却是个问题:若是批量 batch 尺寸太小,或者样本间存在相关性,模型性能将会明显恶化。

与 BN 不同,FRN^[17]在通道层面进行归一化,对每个样例的每个 channel 单独进行归一化,解决了对批量大小的依赖问题,同时性能表现比 BN 归一化层更好。FRN 层

由两个组件构成,一个是 FRN (filter response normalization),一个是 TLU (thresholded linear unit),结构如图 4 所示。FRN 跟 IN(instance normalization)非常相似,也是基于单样本单通道,所不同的是 IN 减去了均值,再除以标准差,对于 $N=1$ 的情况归一化的结果是 0,但 FRN 可以避免这个问题。

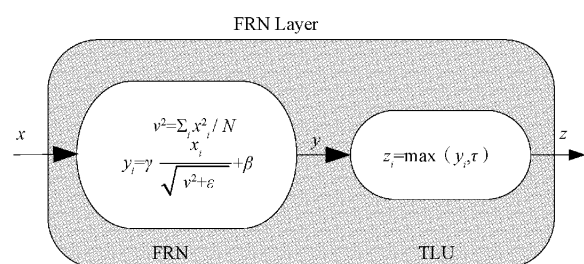


图 4 FRN 层结构

公式中的 v^2 是 x 的二次范数的平均值。这种归一化方式类似 BN 可以用来消除中间操作(卷积和非线性激活)带来的尺度问题,有助于模型训练。公式里的 ϵ 是 1 个很小的正值,以防止发生除 0 错误。

当值较小时,归一化相当于 1 个符号函数,这时候梯度

几乎为 0,严重影响模型训练;当值较大时,曲线变得更圆滑,此时的梯度利于模型学习。本文设为 1 个可学习的参量,初始值设为 1×10^{-6} 。

归一化之后同样需要进行缩放和平移变换,这里的 γ 和 β 也是可学习的参数,如式(3)所示。FRN 缺少去均值的操作,这可能使得归一化的结果任意地偏移 0,如果 FRN 之后是 ReLU 激活层,可能产生很多 0 值,这对于模型训练和性能是不利的。为了解决这个问题,FRN 之后采用的阈值化的 ReLU,即 TLU,如式(4)所示。

$$y = \gamma \hat{x} + \beta \quad (3)$$

$$z = \max(y, \tau) = \text{ReLU}(y - \tau) + \tau \quad (4)$$

1.4 总体框架

本文网络框架如图 5 所示,沿用 Unet 网络的编码器-解码器架构,输入图片首先经过左边的主干特征提取网络,编码部分是经过改造后的 EfficientNetV2 网络,该阶段下采样生成了 5 种不同尺寸大小的特征图,经过 concat 操作特征融合后送入解码网络,拼接后的特征图通过带坐标注意力的模块,增强网络的特征提取能力,经过 5 次双线性插值上采样还原成与输入同等尺寸的特征图,最终经过 1×1 卷积生成分割掩码图。

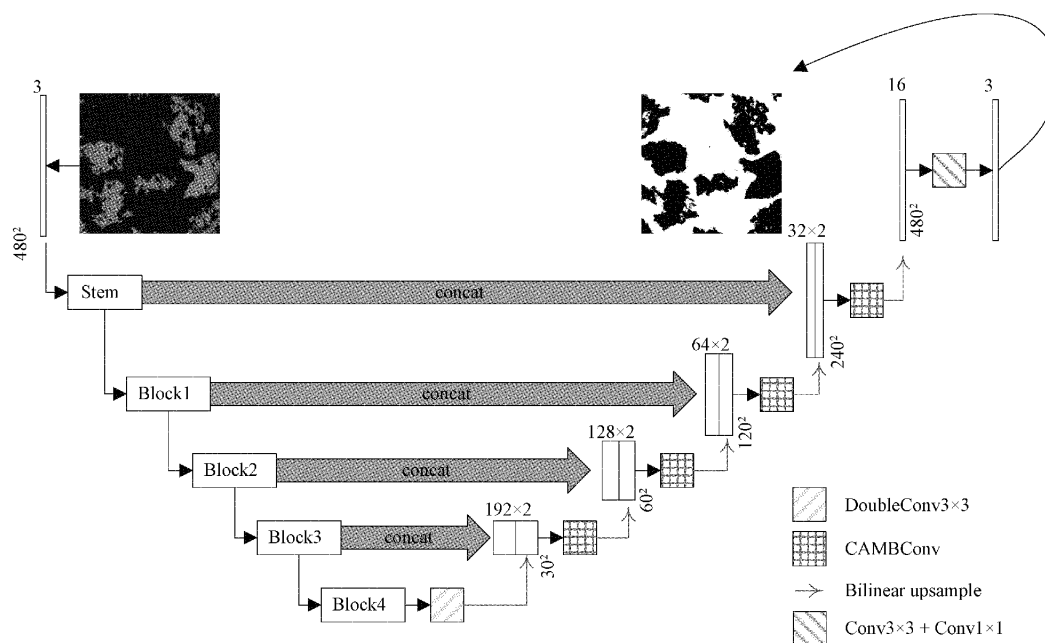


图 5 Efficient_Unet 网络结构

2 实验与结果分析

2.1 数据集介绍

FeM 数据集^[18]图像采集自巴西钛铁矿精矿,主要由赤铁矿和石英组成,极少量磁铁矿和针铁矿。矿石样品用环氧树脂冷安装密度,随后进行研磨和抛光。在反射光显微镜和 SEM 上总共 81 个场进行成像。依次进行配准,得

到 999×756 像素的图像,分辨率为 $1.05 \mu\text{m}/\text{像素}$ 。最后,对来自 SEM 的图像进行阈值化,以组成参考图像:将灰度级介于 $0 \sim 70$ 之间的像素分割为树脂,将灰度级高于 70 的像素设置为矿石颗粒。

Cu 数据集^[19]图像采集自秘鲁铜矿石,其矿物学复杂,主要由硫化物、氧化物、硅酸盐和天然铜组成。样品用环氧树脂冷安装,随后研磨和抛光。在反射光显微镜和 SEM

上共成像 121 个场。随后,对它们进行配准,得到 $1\ 017 \times 753$ 像素的图像,分辨率为 $0.53\ \mu\text{m}/\text{像素}$ 。最后,对来自 SEM 的图像进行阈值化以构成参考图像:将灰度级介于 $0 \sim 30$ 的像素分割为树脂,将灰度级高于 30 的像素设置为矿石颗粒。

对图像通过随机裁剪、旋转、镜像的方式进行数据增广,FeM、Cu 数据集分别扩充为 486,726 张图片,像素尺寸为 480×480 ,按照 6:2:2 的比例划分为训练集、验证集和测试集。

2.2 实验环境

本实验所用服务器搭载 Ubuntu16.04 64 位操作系统,显卡为 NVIDIA GeForce GTX 1070 8 G,搭载 i5-10400F CPU@2.90 GHz×12 处理器,在 pytorch 框架下实验,编程语言为 python3.8.13,使用 cuda11.3 加速库,利用 Adam 优化算法对模型参数优化,采用学习率预热策略(wramup)训练模型,1 个 epoch 后达到初始学习率 1×10^{-3} ,采用自适应学习率(Reduce lr on plateau)根据损失函数值的变化情况对学习率进行调整,当损失连续 6 个 epoch 不降时降低学习率,衰减比例为 0.6,使用 L2 正则化,防止过拟合,正则化参数为 3×10^{-4} ,选择交叉熵损失

函数作为损失函数。

2.3 评价指标

评价指标采用语义分割常用的指标像素准确率(pixel accuracy,PA),平均交并比(mean intersection over union, MIOU),F1 分数(F_1_score)。具体计算公式如下:

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$MIOU = \frac{\frac{TP}{TP + FP + FN} + \frac{TN}{TN + FN + FP}}{2} \quad (6)$$

$$F_1_score = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (7)$$

2.4 实验结果分析

本文在 FeM 数据集上设计了一系列的消融实验,表 2 展示了各项改进措施分割精度结果,“√”表示使用改进措施,“×”表示不采用。①是采用了 EfficientNetV2 作为主干网络,②是在 ①基础上使用 MBconv 替换解码器的普通卷积,③在 ②基础上使用 CA 注意力模块替换 MBconv 中的 SE 模块,④在 ③基础上使用 FRN 层替换常用的 BN 层,从精度指标的结果可以看出各项改进措施是有效的。

表 2 改进措施消融实验结果

模型	EfficientNetV2	MBConv	CAMB	FRN	PA/%	Miou	$F_1_score/\%$
①	√	×	×	×	94.63	89.13	92.71
②	√	√	×	×	96.18	91.98	94.56
③	√	√	√	×	96.24	92.11	94.66
④	√	√	√	√	96.58	92.8	95.15

之后将 ④作为最终确定的改进模型,与主流经典分割模型在两个数据集上进行了分割精度和效率方面的对比,分给精度结果如表 3、4 所示,图 6、7 展示了各模型分别在

表 4 各模型参数量与推理速度对比

模型	参数量/M	推理速度 FPS/(帧/s)
FCN-8s	134.27	15.6
PSPNet	65.57	8.1
Deeplabv3+	54.7	14.2
SegNet	39.79	11.2
Unet	28.94	18.2
Efficient_Unet	11.46	21.7

表 3 与主流经典分割模型精度对比

数据集	模型	PA/%	Miou	$F_1_score/\%$
FeM	FCN-8s ^[20]	93.96	87.67	91.45
	PSPnet ^[21]	95.91	91.48	94.24
	Deeplabv3+ ^[22]	95.97	91.62	94.35
	SegNet ^[23]	96.12	91.89	94.51
	Unet ^[11]	96.27	92.19	94.72
	Efficient_Unet	96.58	92.8	95.15
Cu	FCN-8s	92.57	85.06	89.51
	PSPNet	93.73	87.08	90.89
	Deeplabv3+	93.29	86.39	90.5
	Segnet	93.21	86.1	90.12
	Unet	94.49	88.65	92.15
	Efficient_Unet	95.39	90.43	93.47

FeM 和 Cu 数据集上的分割图对比。实验结果显示,对于 FeM 数据集上,Efficient_Unet 相比 Unet 模型在 PA, Miou, F_1_score 上分别提升了 0.31%,0.61%,0.43%,对于 Cu 数据集上,Efficient_Unet 相比 Unet 在 PA 上提升 0.9%,Miou 上提升 1.78%, F_1_score 上提升 1.32%。从图中也可以看出 Efficient_Unet 分割图最为接近金标准,Unet 表现次之,其他经典模型分割效果一般。在 Cu 数据集示例中表现更为明显,Efficient_Unet 表现最好,在边缘小细节分割更为精细。在两个数据集上都表现优异,体现了良好的泛化性能。

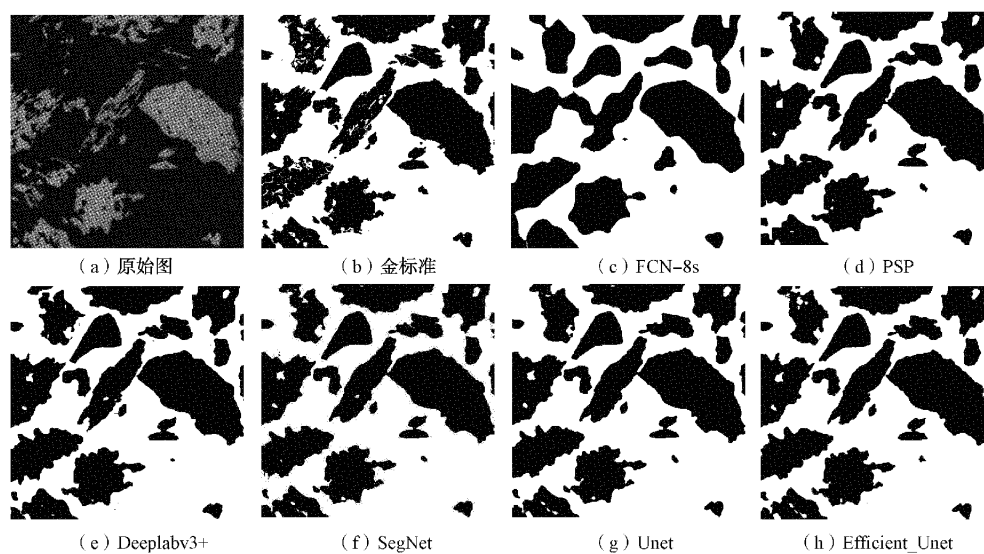


图 6 FeM 数据集分割示例

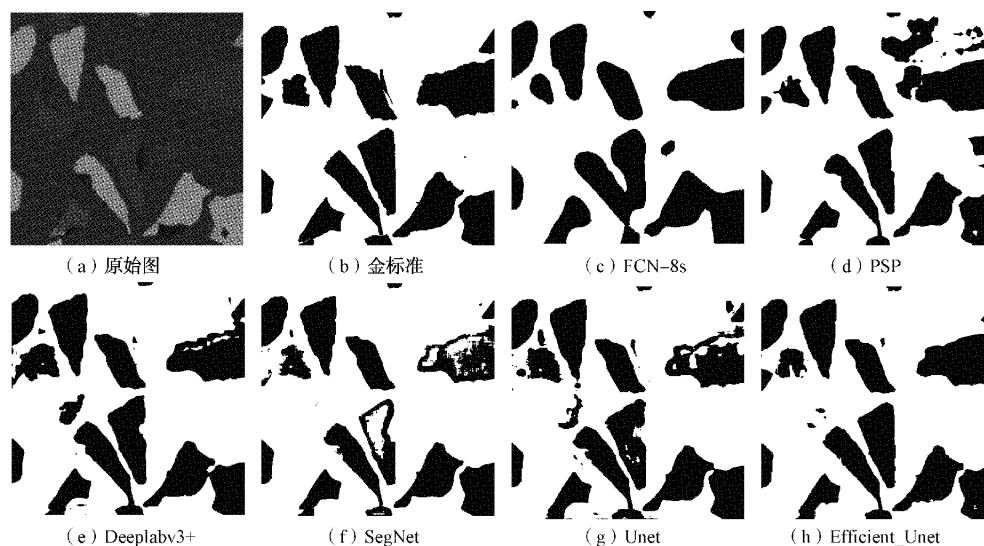


图 7 Cu 数据集分割示例

从表 4 可以看出 Efficient_Unet 模型参数相比 Unet 参数量减少了 60.4%，推理速度提升了 19.23%，对单张 480×480 图片进行预测，取推理时间的 100 次之和的平均，得到 FPS 推理速度可达 21.7 fps，同时相比其他经典模型更为优秀。

3 结 论

为了应对矿产领域的矿石识别困难、识别设备成本高等问题，本文借鉴 Unet 医学分割网络，提出了 Efficient_Unet 模型用来分割树脂中的矿石，该模型主要通过引入 EfficientNetV2 并加以改造作为主干特征提取网络，有效地降低模型的体积；采用 MBconv 优化解码器结构，增强特征提取能力；使用 CA 注意力替换 SE 注意力重构 MBConv 模块，更好地保留目标位置信息；利用 FRN 层替代传统的

BN 归一化层，解决对批量样本大小的依赖问题，提高模型性能。实验结果显示，本文提出的 Efficient_Unet 模型相比其他经典分割网络分割精度更高，参数量更少，推理速度更快，具有实际应用的潜能。

参考文献

- [1] 郭艳军,周哲,林贺洵,等.基于深度学习的智能矿物识别方法研究[J].地学前缘,2020,27(5):39-47.
- [2] IGLESIAS J C Á, AUGUSTO K S, GOMES O D F M, et al. Automatic characterization of iron ore by digital microscopy and image analysis[J]. Journal of Materials Research and Technology,2018,7(3):376-380.
- [3] DONSKOI E, POLIAKOV A, HOLMES R, et al. Iron ore textural information is the key for prediction of downstream process performance [J]. Minerals

- Engineering, 2016, 86: 10-23.
- [4] GOMES O, PACIORNIK S. Iron ore quantitative characterisation through reflected light-scanning electron co-site microscopy[C]. International Congress on Applied Mineralogy, 2008: 699-702.
- [5] OTÁVIO D, FONSECA M G, SIDNEI P. Multimodal Microscopy for Ore Characterization[M]. Rijeka: IntechOpen, 2012.
- [6] 彭伟航, 白林, 商世为, 等. 基于改进 InceptionV3 模型的常见矿物智能识别[J]. 地质通报, 2019, 38(12): 2059-2066.
- [7] FILIPPO M P. Deep learning semantic segmentation of opaque and non-opaque minerals from epoxy resin in reflected light microscopy images [J]. Minerals Engineering, 2021, 170.
- [8] IGLESIAS J C Á, SANTOS R B M, PACIORNIK S. Deep learning discrimination of quartz and resin in optical microscopy images of minerals [J]. Minerals Engineering, 2019, 138: 79-85.
- [9] 张娟娟, 宋圭辰, 刘斌, 等. 改进的基于 Mask R-CNN 的碳纤维图像分割方法 [J]. 中国科技论文, 2021, 16(11): 1189-1194, 1208.
- [10] 于宁波, 刘嘉男, 高丽, 等. 基于深度学习的膝关节 MR 图像自动分割方法 [J]. 仪器仪表学报, 2020, 41(6): 140-149.
- [11] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]. International Conference on Medical Image Computing and Computer-assisted Intervention, 2015: 234-241.
- [12] TAN M, LE Q. Efficientnetv2: Smaller models and faster training [C]. International Conference on Machine Learning, 2021: 10096-10106.
- [13] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), DOI:10.48550/arXiv.1709.01507.
- [14] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C]. International Conference on Machine Learning, 2019: 6105-6114.
- [15] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [16] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]. International Conference on Machine Learning, 2015: 448-456.
- [17] SINGH S, KRISHNAN S. Filter response normalization layer: Eliminating batch dependence in the training of deep neural networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11237-11246.
- [18] GOMES O D F M, PACIORNIK S, FILIPPO M P, et al. FeM dataset-An iron ore labeled images dataset for segmentation training and testing [J]. Zenodo, 2021, DOI:10.5281/zenodo.5014700.
- [19] GOMES O D F M, PACIORNIK S, FILIPPO M P, et al. Cu dataset-A copper ore labeled images dataset for segmentation training and testing [J]. Zenodo, 2021, DOI:10.5281/zenodo.5020566.
- [20] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
- [21] LI H, XIONG P, AN J, et al. Pyramid attention network for semantic segmentation [J]. ArXiv Preprint, 2018, ArXiv:1805.10180.
- [22] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 801-818.
- [23] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

作者简介

曾中华, 硕士研究生, 主要研究方向为计算机视觉、图像分割。

E-mail: 212011020025@home.hpu.edu.cn

曹东(通信作者), 博士, 高级工程师, 主要研究方向为人工智能技术与计算机视觉算法、智能制造与工业智能化。

E-mail: cd@eastru.com