

改进 BiSeNetV1 实时模型的岩屑图像识别^①



孙杰¹, 滕奇志¹, 罗崇兴¹, 何海波², 何小海¹

¹(四川大学 电子信息学院, 成都 610065)

²(成都西图科技有限公司, 成都 610041)

通信作者: 何小海, E-mail: hxh@scu.edu.cn

摘要: 在图像分割识别领域, 现有的深度学习方法大多使用高精度语义分割方法来实现, 存在着网络推理速度慢、计算量大、难以实际应用等问题. 借助于表现较好的 BiSeNetV1 实时网络模型, 通过扩展的空间路径卷积结构、空间金字塔注意力机制 (SPARM) 和简化的注意力特征融合模块 (S-iAFF) 等改进策略, 设计一种用于岩屑图像分割领域的 BiSeNet_SPARM_S-iAFF 实时网络. 扩展的空间路径卷积结构可以获取更丰富的岩屑图像空间特征, 上下文路径使用优化的空间金字塔注意力机制 (SPARM) 进一步细化高层语义特征提取, 在特征融合阶段使用简化注意力特征融合 (S-iAFF) 加强低层空间与高层语义特征的融合程度. 实验结果表明, BiSeNet_SPARM_S-iAFF 网络在 RockCuttings_Oil 岩屑数据集上的平均交并比 (mIoU) 为 64.91%, 相较于 BiSeNetV1 网络提高了 2.68%; 另外改进后的网络在精度上接近大部分高精度语义分割方法, 同时参数量大幅度减少、推理速度有着明显的提升.

关键词: 岩屑图像; 语义分割; BiSeNetV1 网络; 空间金字塔注意力; 迭代注意力特征融合; 深度学习; 卷积神经网络

引用格式: 孙杰, 滕奇志, 罗崇兴, 何海波, 何小海. 改进 BiSeNetV1 实时模型的岩屑图像识别. 计算机系统应用, 2023, 32(10): 45-53. <http://www.c-s-a.org.cn/1003-3254/9245.html>

Recognition of Cuttings Images Based on Improved BiSeNetV1 Real-time Model

SUN Jie¹, TENG Qi-Zhi¹, LUO Chong-Xing¹, HE Hai-Bo², HE Xiao-Hai¹

¹(College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)

²(Chengdu Xitu Technology Co. Ltd., Chengdu 610041, China)

Abstract: In the field of image segmentation and identification, the existing deep learning methods mostly perform tasks by high-precision semantic segmentation methods, which lead to a slow network inference speed, large amount of calculation, and difficult actual application. A real-time network model with better performance, namely BiSeNetV1 is used, and the extended spatial path convolution structure, spatial pyramid attention mechanism (SPARM), simplified iterative attention feature fusion (S-iAFF) module, and other optimization strategies are applied. As a result, a real-time BiSeNet_SPARM_S-iAFF network is designed for rock debris image segmentation. The extended spatial path convolution structure can obtain more abundant spatial features of rock debris images. The context path uses the optimized SPARM to further refine high-level semantic feature extraction. Finally, S-iAFF is used to enhance the fusion degree between low-level spatial and high-level semantic features in the feature fusion stage. The experimental results indicate that the mean intersection over union (mIoU) of the BiSeNet_SPARM_S-iAFF network on the RockCuttings_Oil dataset is 64.91%, which is 2.68% higher than that of the BiSeNetV1 network, and the precision of the improved network is close to that of the most high-precision semantic segmentation methods, while the number of parameters is greatly reduced, and the inference speed is significantly improved.

Key words: cuttings image; semantic segmentation; BiSeNetV1 network; spatial pyramid attention; iterative attention feature fusion; deep learning; convolutional neural network (CNN)

① 基金项目: 国家自然科学基金 (62071315)

收稿时间: 2023-03-13; 修改时间: 2023-04-20; 采用时间: 2023-04-27; csa 在线出版时间: 2023-07-21

CNKI 网络首发时间: 2023-07-24

近年来随着生活水平的提高,人类对能源的需求日益增加,其中石油作为国家经济发展的重要能源,对其进行开发和利用的相关技术也日益重要,石油勘探技术中,通过钻井作业对地质参数分析可以快速判断该地的石油含量,其中岩屑识别技术可作为判断地层岩性的重要手段^[1,2].工作人员从井口捞取初始岩屑样本,通过清洗等一系列处理过程得到利于识别的岩屑样本,最终通过图像采集得到岩屑颗粒图像用于后续的认识。

随着计算机技术的发展,深度学习逐渐代替传统图像算法应用到石油地质研究的分类识别上.特别是通过语义分割算法的应用大大提高了岩屑图像的分类效果,同时减少了人工识别岩屑的时间.语义分割是一种典型的计算机视觉研究方向,其基于像素级的分类方式能够很好地关联不同场景下每个像素的标签类别,最终通过不同标注结果实现像素级的分类任务,如FCN全卷积网络^[3]、轻量级ENet语义分割网络^[4]等均可实现对图像像素级的分类,它们广泛应用于视频监控、医学图像处理、岩屑图像处理等领域.语义分割的发展是逐步进行的,随着各种模型的涌现,在经典的训练场景下分割的精度也越来越高,然而大部分训练精度的提升是以参数量、训练时间的增加为代价的,特别是精度达到一定的高度后,模型精度的微小提升往往会导致训练成本(如计算量、时间、占用内存)成倍地增加.基于这种情况,近年来实时语义分割的发展逐渐得到人们的重视,在某些特定的应用场景下,实时语义分割的研究目标主要是在尽量保持分割精度的情况下简化训练模型的参数量、占用内存等,设计出低延迟、高效率、分割精度表现良好的模型^[5,6].

大部分的高精度方法在语义分割的结果上占据较大优势,如文献[7]的DANet模型在不同的应用场景中都有着优秀的表现,但从推理速度的角度来看,实时语义分割的优势较为明显.当处理任务用于移动设备上时更加关注的是模型的实时性,此时就需要在精度和实时性上进行衡量.近年来具有挑战性的实时语义分割任务出现许多实际的应用与研究,如文献[8]提出一种快速实现高质量分割的图像级联网络ICNet,文献[9]提出一种同时保持效率与准确性的ERFNet深层架构,文献[10]中的BiSeNetV1和文献[11]中的BiSeNetV2提出一种实时的双边分割网络架构,文献[12]提出一种基于高效空间金字塔的ESPNet实时分割模型,文献[13]提出一种轻量级上下文引导网络

CGNet,文献[14]提出一种基于高分辨率图像数据的FastSCNN超实时语义分割模型.不同网络对特定任务的适应性差异很大,某些应用场景下实时网络识别精度可能会接近甚至高于某些高精度方法,实时分割网络往往就是在这种情况下展现出巨大的优势。

1 BiSeNetV1 网络

BiSeNetV1^[10]是一种通过空间路径和上下文路径分别提取图像低级空间特征和高级上下文语义特征的基本架构,其优势在于同时计算两条路径来提高效率.首先空间路径使用3层步长为2的卷积编码丰富的空间信息,同时上下文路径利用预先训练好的Xception^[15]等轻量级模型提供了较大的接受域,该接受域编码高级语义上下文信息,上下文路径中使用注意力细化模块ARM细化每个阶段的特征,ARM使用全局平均池来捕获全局上下文,并计算注意力向量指导特征学习来轻松地集成全局上下文信息.由于两条路径的输出特征表示级别不同,最后使用FFM融合模块融合两条路径的输出特征用于实现像素级的分类。

2 改进的 BiSeNetV1 网络结构

通过初期实验表明BiSeNetV1双边分割网络在岩屑场景中表现较好,本文在BiSeNetV1网络架构的基础上,在保证网络推理速度基本不变的情况下,得到更适合岩屑图像识别的BiSeNet_SPARM_S-iAFF网络结构,其整体结构如图1所示。

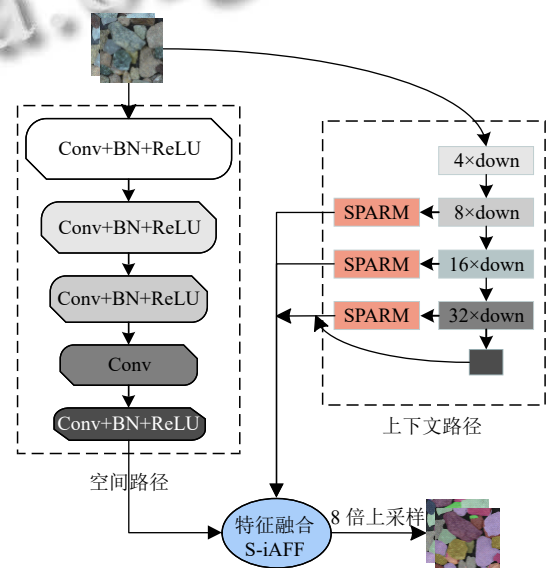


图1 BiSeNet_SPARM_S-iAFF网络整体结构

2.1 空间路径优化

为了捕获图像中足够多的空间信息,高精度方法往往通过增大卷积过程中的感受野,如池化操作或空洞卷积^[16]等方法,但增大感受野同时导致训练成本的增加.通过分析发现岩屑数据集中的图像特征与经典的城市场景数据集 Cityscapes^[17]特征不同,城市场景中的全局空间信息是非常重要的特征信息,因为城市场景中类别的形状具有多样性,同一类别内部也存在颜色和形状上的不同,不同种类的对象大小差距较大,需要足够的全局空间信息来辅助判断某区域的像素信息,而岩屑颗粒图像从整体上看相同种类的岩屑在形

状上和大小上的差别较小,同一类别的形状、大小、颜色、细节纹理较为相似,且图像中不同岩屑种类相互交叠的边缘信息较为复杂,需要细化卷积结构将空间路径的映射更多地关注到边缘等特征上.

基于上述情况,在保持原始输入图像空间大小的同时,将空间结构扩展至5层,同时控制各层通道数大小、卷积核大小、步长等减少参数数量的增加,卷积过程使用填充操作保持输出特征映射为原始图像的1/8,扩展的结构可以更多地提取原始图像的空间信息,图2显示了BiSeNetV1改进后的空间路径上的卷积结构.

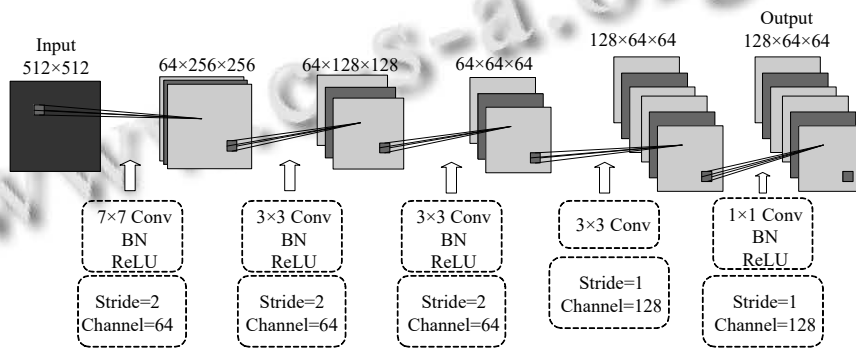


图2 BiSeNet_SPARM_S-iAFF网络空间路径结构

2.2 SPARM 空间金字塔注意力

注意力机制近年来被广泛应用于深度学习的诸多领域来提高神经网络的训练性能^[18,19]. BiSeNetV1中的ARM注意力细化模块类似于SENet^[20]的SE块,通过全局平均池化GAP^[21]聚合各个信道中的全局信息,其类似于结构正则化器的行为能够防止过度拟合,但处理结果更倾向于强调正则化的效果,导致中间特征图中结构信息的丢失,特别是对于比较大的中间特征图进行聚合平均值时,会导致特征表示能力的显著损失,这也是原始上下文路径只对16倍和32倍下采样的中间特征图使用ARM注意力细化模块的原因.

ARM对于图像中特征变化简单的稀疏区域有着较好的表现,但对于图像中特征变化复杂的密集区域,考虑到ARM使用单独全局平均池化,越密集的情况其特征变化越复杂,提取上下文语义特征效果越差,这导致在最终的分割效果中密集区域某一单独的岩屑颗粒受到周围区域岩屑颗粒较大的影响.

基于上述问题,本文借助SPA-Net^[22]中3个自适应全局平均池化的方法,使语义信息能够被更加充分

的利用,其中1x1自适应全局平均池化是具有结构正则化的传统全局平均池化,2x2自适应全局平均池化用来平衡结构信息和结构正则化之间的关系,4x4自适应全局平均池化可以捕获更多的特征表示和结构信息.相比于原始的单独全局平均池化,此池化方式可以兼顾特征图的正则化效果和原始特征的特征表示、结构信息,同时又继承全局平均池化的优点,SPA模块的具体结构如图3所示,其中AAP(n)表示 $n \times n$ 的自适应全局平均池化,Resize操作将经过自适应全局平均池化操作后的输出特征图调整为与输入特征图相同的大小.

注意力模块SPARM使用图3中的SPS空间金字塔池化结构代替单独自适应全局平均池化,并保留ARM注意力细化模块的连接路径,这种SPARM注意力模块能够更加全面的捕获包含特征正则化效果和结构信息的全局上下文信息,通过计算注意力向量来指导特征学习,达到细化上下文路径中下采样阶段的输出特征,图4显示了SPARM注意力模块的连接路径,其中 \odot 表示逐元素乘法(element-wise multiplication).

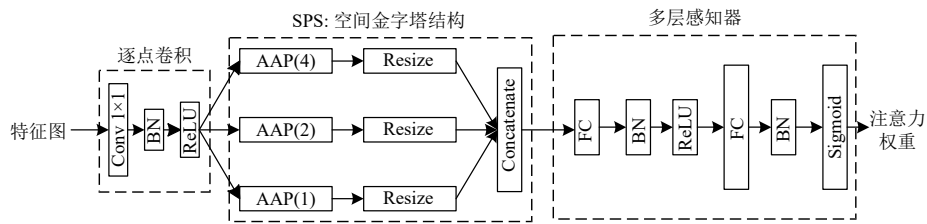


图3 SPA空间金字塔注意力模块的结构

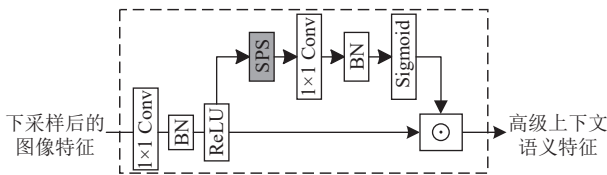


图4 SPARM空间金字塔注意力细化模块

另外使用空间金字塔的注意力模块能够克服对较大中间特征图特征提取能力差的问题,考虑到岩屑场景具有比较复杂的局部边缘信息,上下文路径的几个输出分支需要通过注意力模块捕获通道的依赖性,本文在上下文路径增加8倍下采样的注意力特征提取操作来加强网络对边缘信息的提取,且其输出权重设置相比于16倍和32倍的输出权重小(权重为0.2),使得上下文路径的输出更多地关注到深层语义特征。

2.3 S-iAFF 注意力特征融合模块

为了更好地融合提取到的空间和上下文语义信息,借助文献[23]中的迭代注意特征融合模块来代替BiSe-NetV1网络中的FFM特征融合模块,可以很好地解决融合不同尺度特征出现的问题,另外FFM融合方法结果更倾向于强调分布在全局的大型对象,融合过程会忽略中小对象大部分图像信息,中小对象的检测是目前先进网络的一大技术瓶颈^[24]。

FFM模块首先使用级联操作连接空间路径和上下文路径的输出特征,利用更新的权重值重新加权级联后的特征。假设空间路径输出特征为 X ,上下文路径输出特征为 Y , X 和 Y 通过Concatenation级联操作后的特征表示为 $Cat(X, Y)$,最终融合后的输出特征表示为 F ,可以得到FFM特征融合模块的计算方法如下:

$$F = G(Cat(X, Y)) \odot Cat(X, Y) + Cat(X, Y) \quad (1)$$

其中, G 表示经过全局平均池化GAP、批标准化BN、激活函数后的权重向量。

为了解决上述问题,引入了多尺度通道注意力模

块(MS-CAM)^[23],将多尺度的特征在注意力模块内聚合来缓解尺度变化和中小对象引起的问题。MS-CAM是在全局上下文的基础上添加局部上下文分支,选择逐点卷积(Point-wise Conv)作为局部分支的聚合器来淡化全局平均池化操作,使聚合更多地关注到局部上下文信息,最终得到的权重向量通过重新加权平衡局部和全局上下文信息。另外通过取消MS-CAM中的通道缩减策略降低复杂度。图5(a)显示了MS-CAM的模块结构,其中 r 表示通道缩减比,图5(b)显示了去除通道缩减后的结构,将其称为Simplified MS-CAM,简称SMS-CAM。

假设SMS-CAM模块中通过全局分支的输出为 $G(X_{in})$,通过局部分支的输出为 $L(X_{in})$,首先空间路径的输出 X 通过全局平均池化后得到 $g(X)$,其计算结果如下:

$$g(X) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_{[i,j]} \quad (2)$$

其中, $H \times W$ 表示特征图的尺寸大小, $[i, j]$ 表示输入张量 X 中所有批次、在高 i 和宽 j 处的切片位置。

全局 $G(X_{in})$ 和局部上下文 $L(X_{in})$ 输出计算如下:

$$G(X_{in}) = \delta(B(pwconv(g(X)))) \quad (3)$$

$$L(X_{in}) = \delta(B(pwconv(X))) \quad (4)$$

其中, δ 表示线性整流函数ReLU, B 表示批标准化(BN), $pwconv$ 表示逐点卷积。给定全局上下文和局部上下文,通过SMS-CAM模块后可获得如下结果:

$$X_{out} = X \odot SMS(X) = X \odot \sigma(G(X_{in}) \oplus L(X_{in})) \quad (5)$$

其中, σ 表示Sigmoid函数, SMS 为通过SMS-CAM模块的输出, \oplus 表示加法(broadcasting addition)。

考虑到空间与上下文路径的特征存在较大语义上的差别,采用两阶段SMS-CAM注意力融合方式,称为S-iAFF。由于一个阶段的SMS-CAM不能很好地抵消初始融合的影响,两阶段SMS-CAM的S-iAFF融合方法很好地降低初始融合对最终的融合权重的影响程

度, 输入特征 X 和 Y 通过 S-iAFF 的计算过程如下:

$$X\Phi Y = SMS(X\oplus Y)\odot X\oplus(1-SMS(X\oplus Y))\odot Y \quad (6)$$

$$F = SMS(X\Phi Y)\odot X\oplus(1-SMS(X\Phi Y))\odot Y \quad (7)$$

其中, $X\Phi Y$ 表示经过一阶段 SMS-CAM 的输出, F 表示经过两阶段 SMS-CAM 后的输出, 图 6 显示了 S-iAFF 模块的结构。

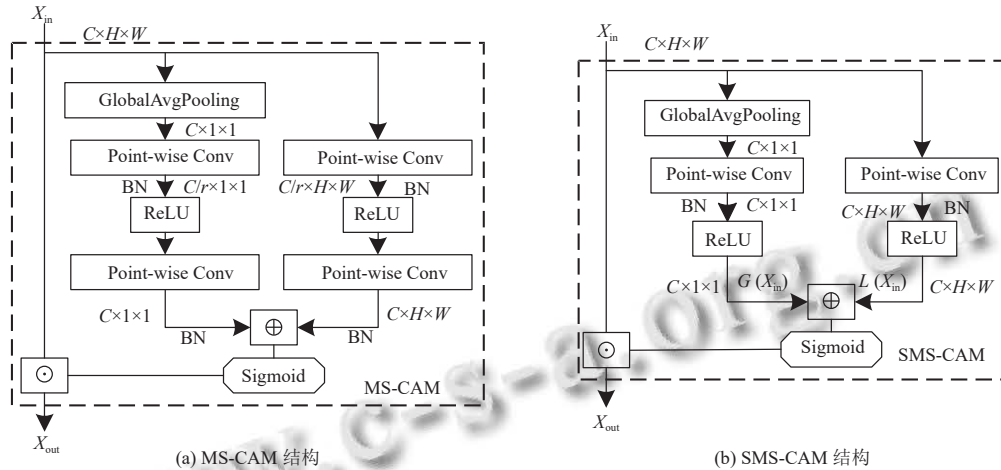


图 5 多尺度通道注意力模块

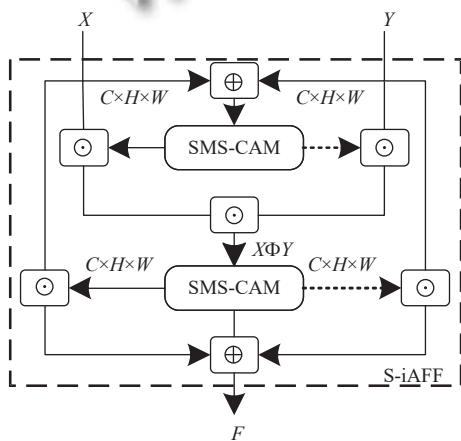


图 6 S-iAFF 注意力特征融合模块

3 实验结果与分析

3.1 实验配置

本文基于 Linux 操作系统的 PyTorch 框架进行实验, 使用 4 张 GPU 为 NVIDIA GeForce RTX 2080Ti 深度学习服务器, 代码基于 Python 3.8 实现。

3.2 实验数据集

本文数据集的岩屑初始样本由合作的石油公司提供, 是现场采集并经过清洗处理的干净样本, 得到的岩屑种类有 15 种, 分别为含油斑粉砂岩、含砾粗砂岩、灰色粉砂岩、灰绿色凝灰岩、石灰岩、灰绿色砂岩、紫灰色泥质粉砂岩、褐色粉砂岩、方解石、页岩、灰

绿色安山岩、深灰色安山岩、紫红色安山岩、褐色流纹岩和石英, 训练测试过程中将背景也作为一类. 通过图像采集系统^[25] 获取 100 张分辨率为 4000×3750 的岩屑原始图像, 由于高分辨率图像在训练过程中占用 GPU 内存大, 采集到的高分辨率图像需要按顺序裁剪, 岩屑图像裁剪后并不影响最终的分效果, 裁剪的图像分辨率统一为 512×512, 接着通过随机翻转、缩放、锐化等数据增强操作得到 8 540 张分辨率为 512×512 子图像, 数据集中 8 210 张图像用于训练, 330 张图像用于测试, 并将此数据集称为 RockCuttings_Oil 岩屑图像数据集^[26]. 图 7 展示了数据集中的图像样本, 图 7(a) 为通过采集系统得到的分辨率为 4000×3750 的岩屑图像, 图 7(b) 为由高分辨图像裁剪为分辨率为 512×512 岩屑子图像, 网络训练时使用的是裁剪后的图像, 图 7(c) 为对应于图 7(b) 的 RGB 彩色标签图。

3.3 网络训练细节

训练过程使用随机梯度下降 (SGD), 动量为 0.9, 权值衰减为 0.000 5, 应用类似文献^[27] 的 poly 学习策略, 初始学习率为 1E-4, 训练网络设置 3 个阶段的特征图输出, 主损失函数监控整个网络的输出, 同时使用两个辅助损失函数监控上下文路径的输出, 三者使用的交叉熵损失函数计算公式如下:

$$l = - \sum_{i=1}^N y_i \log(p_i) \quad (8)$$

其中, y 和 p 表示期望和真实预测, N 表示类别数. 3 个损失函数的损失权重均设置为 1, 那么融合损失函数的计算如下:

$$L(X, W) = l_m(X, W) + \sum_{i=1}^2 l_i(X_i, W) \quad (9)$$

其中, l_m 表示主损失函数的损耗, l_i 表示辅助损失函数的损耗, X_i 表示第 i 阶段的输出特征, W 表示模型的训练



图7 RockCuttings_Oil 岩屑数据集样本

首先对 BiSeNetV1 网络, 模块的其他部分均保持原始不变, 只改变上下文路径的网络模型, 表 1 展示了使用不同上下文路径基础模型的表现情况.

表 1 BiSeNetV1 下使用不同上下文基础模型的实验比较

网络模型	上下文基础模型	FLOPs (G)	Parms (M)	mIoU (%)
BiSeNetV1	ResNet18	14.83	13.43	62.23
BiSeNetV1	ResNet50	98.95	59.24	62.38
BiSeNetV1	ResNeXt50	98.95	59.24	62.54
BiSeNetV1	ResNeSt50	105.67	61.17	62.89

表 1 中 FLOPs 表示模型每秒的浮点运算次数, Parms 表示模型训练的参数量. 通过表 1 发现 BiSeNetV1 网络使用轻量化模型 ResNet18^[28] 的参数量较低, 虽然使用 ResNeXt50^[29] 和 ResNeSt50^[30] 后训练精度有所提升, 但使用该模型后的参数量甚至超过部分高精度方法, 考虑到轻量化的因素, 采用 ResNet18 作为上下文路径中的基础模型.

本文在上下文路径模型 ResNet18 的基础上对 BiSeNetV1 进行改进, 主要策略为空间路径使用 5 层扩展的卷积结构, 上下文路径使用 SPARM 代替 ARM 注意力模块, 空间和上下文路径的特征融合使用 S-iAFF 代替 FFM 模块. 表 2 展示了使用不同模块的比较, 其中 SP 表示原始空间路径, SP(5) 表示 5 层卷积空间路径, CP 表示原始上下文路径, CP(ARM) 表示采用 ARM 模块的上下文路径, CP(SPARM) 表示采用 SPARM 模块的上下文路径.

权重参数. 其中第 1 阶段使用主损失函数监控训练损耗, 第 2、3 阶段辅助损失函数来监控训练损耗.

3.4 对比实验与分析

在 RockCuttings_Oil 岩屑数据集上评估相同模型的不同上下文主干网络、不同注意力模块、不同特征融合模块以及不同模型下的分割效果, 验证改进方法中相应组件的有效性.

在实际岩屑分割场景中, 分割效果和实时性是同样重要的评判标准, 两种指标需要达到一定的平衡性. 通过表 2 实验对比发现对 BiSeNetV1 网络不同模块进行细节优化, 在参数量较小增加的情况下, 参数量由 13.43M 增加到 14.17M, 平均交并比 (mIoU) 由 62.23% 提升至 64.91%, 整体提升了 2.68%.

表 2 改进策略性能比较

网络训练策略	FLOPs (G)	Parms (M)	mIoU (%)
SP+CP(ARM)+FFM(BiSeNetV1)	14.83	13.43	62.23
SP(5)+CP(ARM)+FFM	16.17	13.68	63.47
SP(5)+CP(SPARM)+FFM	17.21	13.89	64.52
SP(5)+CP(SPARM)+iAFF	21.02	14.23	64.90
SP(5)+CP(SPARM)+S-iAFF (ours)	20.33	14.17	64.91

3.5 分割效果对比

对于测试集下的岩屑图像分割效果, 图 8 展示了 3 组岩屑测试集图片分别在 BiSeNetV2 网络、BiSeNetV1 网络、BiSeNet_SPARM_S-iAFF 网络和高精度方法 PSPNet、DANet 网络的分割效果对比.

对于岩屑分割效果, 一张岩屑图像包含的图像信息非常多. 通过图 8 分析发现: 从整体表现对比来看 DANet 分割效果最平滑, PSPNet 和 BiSeNet_SPARM_S-iAFF 次之, 表现较差的是 BiSeNetV1 和 BiSeNetV2. 从细节对比来看, 第 1 排岩屑分割图矩形框标注的区域, 通过原图可以发现该岩屑内部局部细节差异很大, 这种情况会导致此类石头的识别难度加大, 对于此区

域的分割结果上来看,其他很多网络的识别结果存在很明显的错误,相比较而言 BiSeNet_SPARM_S-iAFF 识别结果比较统一,且更具有参考性;第2排矩形框标注的区域内,BiSeNet_SPARM_S-iAFF 网络相比于 BiSeNetV1、BiSeNetV2 在密集区域对细节的处理更

好,误分类区域更少,但相较于高精度方法还存在一定的差距;第3排矩形框标注的区域内,BiSeNet_SPARM_S-iAFF 网络在右上侧区域分割效果略优于 DANet 网络,而下侧和左侧区域相比于 DANet 网络有所差距,但分割效果相比较 BiSeNetV1、BiSeNetV2 有明显提升.

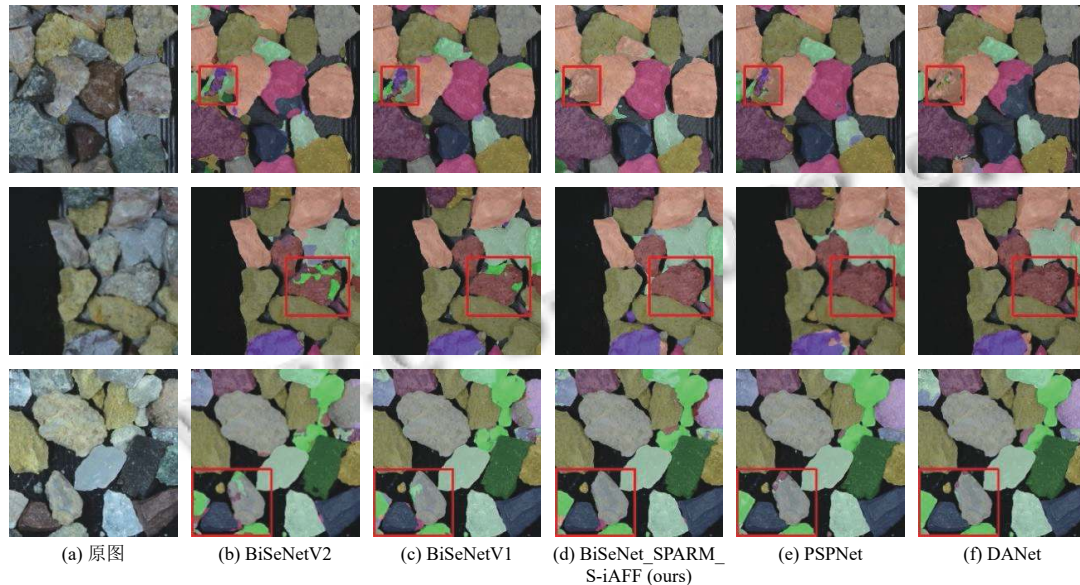


图8 不同模型的分割效果对比

由于最初采集到的岩屑图像的分辨率为 4000×3750 ,而训练过程的输入图像为裁剪后分辨率为 512×512 的子图集合,训练后最终输出对应输入图像的 512×512 子效果图,为了更直观地展示岩屑分割的效果,需要将神经网络输出的效果图通过拼接的方式合成原始分辨率大小为 4000×3750 的完整分割效果图,图9展示了相应的原始图像及对应输出子效果图拼接后的完整分割效果.

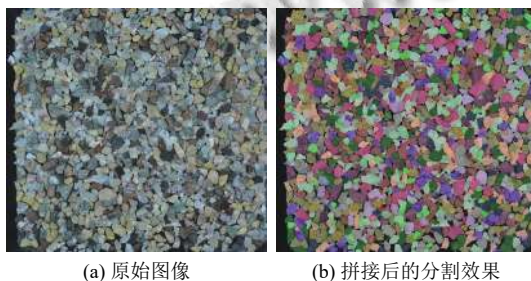


图9 完整分割效果展示

为了验证 BiSeNet_SPARM_S-iAFF 网络分割的效果,分别在高精度方法和实时分割方法中选取较为经典的网络进行对比实验,高精度方法如 FCN^[3]、DANet^[7]、

PSPNet^[31]、DeepLabV3+^[32] 和 UNet^[33],实时语义分割方法如 ICNet^[8]、ERFNet^[9]、BiSeNetV1^[10]、BiSeNetV2^[11]、CGNet^[13] 和 FastSCNN^[14],表3展示了不同模型在 RockCuttings_Oil 岩屑数据集、迭代次数为 80000 情况下的精度、参数量、单张图像处理时间的比较.

表3 不同模型的精度、参数量、单张图像处理时间比较

网络模型	mIoU (%)	Params (M)	时间 (s)
FCN	62.85	49.82	4.90
PSPNet	65.27	48.97	4.78
DeepLabV3+	64.12	43.58	5.60
UNet	62.78	29.06	3.94
DANet	69.97	49.86	4.98
ICNet	48.61	25.71	0.97
ERFNet	43.24	2.08	1.01
CGNet	18.8	0.489	0.92
FastSCNN	34.29	1.46	1.78
BiSeNetV2	57.84	14.8	1.67
BiSeNetV1	62.23	13.43	1.13
BiSeNet_SPARM_S-iAFF	64.91	13.77	1.19

通过对表3观察分析可以发现改进后的 BiSeNet_SPARM_S-iAFF 网络的训练精度与大部分高精度方法

的相近,且参数量和处理时间大大减小,另外相较于其他经典的实时网络精度表现最好。

4 结束语

本文基于 BiSeNetV1 改进设计一种用于岩屑图像分割任务的 BiSeNet_SPARM_S-iAFF 网络模型,改进后的网络依旧通过空间路径和上下文路径来提出特征信息,空间路径首先通过 5 层卷积结构提取足够多的空间信息,上下文路径的使用 SPARM 模块代替 ARM 模块来提取更多关注原始图像密集区域的语义信息,最后使用 S-iAFF 代替 FFM 融合模块来加强空间和语义特征的融合程度。整体实验结果表明 BiSeNet_SPARM_S-iAFF 网络相比于 BiSeNetV1 网络,在几乎没有增加网络参数量和推理速度的情况下,实现了更好的分割精度,此分割效果相比于其他实时分割网络具有巨大的优势,甚至优于部分高精度方法的分割效果。另外在中小对象特征信息和图像密集区域及特征融合阶段分别使用了优化策略,并通过相关实验得以验证这些优化策略具有应用上的可行性,但通过分析可知最终对岩屑图像密集区域的分割效果仍存在不足之处,需要后续进一步细化分割模型来进行改善。

参考文献

- 1 陈晓明,杨涛,梁坤,等.不同勘探类别油气发现成本评价方法及应用.中国石油勘探,2021,26(3):88-94.[doi:10.3969/j.issn.1672-7703.2021.03.008]
- 2 潘柯宇,卢正东,唐明,等.四川盆地 SYS 构造录井岩屑识别方法探讨及应用.钻采工艺,2020,43(S1):24-26,34.
- 3 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 3431-3440. [doi: 10.1109/CVPR.2015.7298965]
- 4 Paszke A, Chaurasia A, Kim S, et al. ENet: A deep neural network architecture for real-time semantic segmentation. arXiv:1606.02147, 2016.
- 5 Lin PW, Sun P, Cheng GL, et al. Graph-guided architecture search for real-time semantic segmentation. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 4202-4211. [doi: 10.1109/cvpr42600.2020.00426]
- 6 Fan MY, Lai SQ, Huang JS, et al. Rethinking BiSeNet For real-time semantic segmentation. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 9711-9720. [doi: 10.1109/CVPR46437.2021.00959]
- 7 Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 3141-3149. [doi: 10.1109/CVPR.2019.00326]
- 8 Zhao HS, Qi XJ, Shen XY, et al. ICNet for real-time semantic segmentation on high-resolution images. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 418-434. [doi: 10.1007/978-3-030-01219-9_25]
- 9 Romera E, Álvarez JM, Bergasa LM, et al. ERFNet: Efficient residual factorized ConvNet for real-time semantic segmentation. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(1): 263-272. [doi: 10.1109/TITS.2017.2750080]
- 10 Yu CQ, Wang JB, Peng C, et al. BiSeNet: Bilateral segmentation network for real-time semantic segmentation. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 334-349. [doi: 10.1007/978-3-030-01261-8_20]
- 11 Yu CQ, Gao CX, Wang JB, et al. BiSeNet V2: Bilateral network with guided aggregation for real-time semantic segmentation. International Journal of Computer Vision, 2021, 129(11):3051-3068.[doi:10.1007/s11263-021-01515-2]
- 12 Mehta S, Rastegari M, Caspi A, et al. ESPNet: Efficient spatial pyramid of dilated convolutions for semantic segmentation. Proceedings of the 15th European Conference Computer Vision. Munich: Springer, 2018. 561-580. [doi: 10.1007/978-3-030-01249-6_34]
- 13 Wu TY, Tang S, Zhang R, et al. CGNet: A light-weight context guided network for semantic segmentation. IEEE Transactions on Image Processing, 2021, 30: 1169-1179. [doi: 10.1109/TIP.2020.3042065]
- 14 Poudel RPK, Liwicki S, Cipolla R. Fast-SCNN: Fast semantic segmentation network. Proceedings of the 30th British Machine Vision Conference. Cardiff: BMVA Press, 2019.
- 15 Chollet F. Xception: Deep learning with depthwise separable convolutions. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 1800-1807.
- 16 Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. Proceedings of the 4th International

- Conference on Learning Representations. San Juan: ICLR, 2016.
- 17 Cordts M, Omran M, Ramos S, *et al.* The Cityscapes dataset for semantic urban scene understanding. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 3213–3223.
- 18 Xu K, Ba J, Kiros R, *et al.* Show, attend and tell: Neural image caption generation with visual attention. Proceedings of the 32nd International Conference on Machine Learning. Lille: JMLR.org, 2015. 2048–2057.
- 19 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 20 Hu J, Shen L, Albanie S, *et al.* Squeeze-and-excitation networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011–2023. [doi: [10.1109/TPAMI.2019.2913372](https://doi.org/10.1109/TPAMI.2019.2913372)]
- 21 Lin M, Chen Q, Yan SC. Network in network. arXiv:1312.4400, 2014.
- 22 Guo JD, Ma X, Sansom A, *et al.* Spanet: Spatial pyramid attention network for enhanced image recognition. Proceedings of the 2020 IEEE International Conference on Multimedia and Expo. London: IEEE, 2020. 1–6. [doi: [10.1109/ICME46284.2020.9102906](https://doi.org/10.1109/ICME46284.2020.9102906)]
- 23 Dai YM, Gieseke F, Oehmcke S, *et al.* Attentional feature fusion. Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2020. 3559–3568. [doi: [10.1109/WACV48630.2021.00360](https://doi.org/10.1109/WACV48630.2021.00360)]
- 24 Singh B, Najibi M, Davis LS. SNIPER: Efficient multi-scale training. Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: Curran Associates Inc., 2018. 9333–9343.
- 25 滕工生, 张国良, 陈树秋, 等. 一种新型的 CIS 岩屑录井数字成像系统. 录井工程, 2015, 26(1): 53–57. [doi: [10.3969/j.issn.1672-9803.2015.01.013](https://doi.org/10.3969/j.issn.1672-9803.2015.01.013)]
- 26 王茜. 岩屑图像分类系统研究与设计 [硕士学位论文]. 成都: 四川大学, 2022.
- 27 Chen LC, Papandreou G, Kokkinos I, *et al.* DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834–848. [doi: [10.1109/TPAMI.2017.2699184](https://doi.org/10.1109/TPAMI.2017.2699184)]
- 28 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778. [doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90)]
- 29 Xie SN, Girshick R, Dollár P, *et al.* Aggregated residual transformations for deep neural networks. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 5987–5995. [doi: [10.1109/CVPR.2017.634](https://doi.org/10.1109/CVPR.2017.634)]
- 30 Zhang H, Wu CR, Zhang ZY, *et al.* ResNeSt: Split-attention networks. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans: IEEE, 2022. 2735–2745. [doi: [10.1109/CVPRW56347.2022.00309](https://doi.org/10.1109/CVPRW56347.2022.00309)]
- 31 Zhao HS, Shi JP, Qi XJ, *et al.* Pyramid scene parsing network. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6230–6239. [doi: [10.1109/CVPR.2017.660](https://doi.org/10.1109/CVPR.2017.660)]
- 32 Chen LC, Zhu YK, Papandreou G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 833–851. [doi: [10.1007/978-3-030-01234-2_49](https://doi.org/10.1007/978-3-030-01234-2_49)]
- 33 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015. 234–241. [doi: [10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28)]

(校对责编: 孙君艳)