doi:10.12068/j.issn.1005-3026.2026.20240240

一种提高推理速度的遥感图像实例分割算法

刘 军[1]，沈展鹏[1]，路国佳[1]，刘向军[2]
（1.东北大学 计算机科学与工程学院，辽宁 沈阳 110169；2.中软信息系统工程有限公司，北京 100081）

摘 要：针对遥感图像处理领域内实例分割算法推理速度慢的问题，选择SOLOv2-50作为基准算法，为提升算法的推理速度FPS，提出使用MobileNetV2作为骨干网络的实例分割算法M-SOLOv2和采用模型压缩及图像大小调整的改进算法L-SOLOv2，并在L-SOLOv2基础上融合无参注意力机制SimAM，提升模型分割精度，得到SL-SOLOv2模型.实验结果表明：在SSDD数据集上，相较于基准模型，改进算法L-SOLOv2在增加平均精度的前提下，推理速度大幅提升，并且相较于M-SOLOv2兼顾了分割性能和速率；而SL-SOLOv2算法相比于基准模型，平均精度均值（mAP）提升了11.3%，达到60.3%，推理速度也显著加快.

关 键 词：深度学习；遥感图像；实例分割；注意力机制；SSDD数据集
中图分类号：TP 391.41 **文献标志码**：A **文章编号**：1005-3026(2026)04-0024-09

A Remote Sensing Image Instance Segmentation Algorithm for Improving Inference Speed

LIU Jun[1], *SHEN Zhan-peng*[1], *LU Guo-jia*[1], *LIU Xiang-jun*[2]
(1. School of Computer Science & Engineering, Northeastern University, Shenyang 110169, China; 2. CS&S Information System Engineering Co., Ltd., Beijing 100081, China. Corresponding author: LIU Jun, E-mail: liujun@cse.neu.edu.cn)

Abstract: To address the challenge of slow inference speeds in instance segmentation algorithms during remote sensing image processing, SOLOv2-50 was selected as the baseline algorithm. To enhance the inference speed of this algorithm, an instance segmentation algorithm called M-SOLOv2 was proposed, utilizing MobileNetV2 as the backbone network. Additionally, an improved algorithm called L-SOLOv2 was introduced, which incorporated model compression and image size adjustment. Furthermore, the parameter-free attention mechanism, namely SimAM, was integrated into L-SOLOv2 to enhance the model's average precision (AP) in segmentation accuracy, resulting in the SL-SOLOv2 model. Experimental results show that on the SSDD dataset, the improved algorithm L-SOLOv2 significantly increases the FPS while enhancing AP compared to the baseline model. Moreover, compared to M-SOLOv2, it also takes into account both segmentation performance and speed. SL-SOLOv2 achieves an improvement of 11.3% in mean average precision (mAP) compared to the baseline model, reaching 60.3%, while also significantly accelerating inference speed.

Key words: deep learning; remote sensing image; instance segmentation; attention mechanism; SSDD dataset

随着深度学习技术的发展，遥感图像处理领域已经不再满足于单纯标记图像中的特定目标（目标检测[1]）或是只对一类目标进行像素级别的分类（语义分割[2]），而是更希望能准确快速识别出图像中的每一个实例.实例分割[3]是在目标检测方法和语义分割方法的基础上，对遥感图像进行逐像素的分割，将每一元素进行分类的同时，保留了分割目标的具体细节信息，实现了实例级别的目标预测与分割.当前的实例分割模型通常采用深度神经网络框架，常见的架构包括Mask

收稿日期：2024-12-30
基金项目：国家自然科学基金青年科学基金资助项目（62501133）.
作者简介：刘 军（1969—），男，辽宁沈阳人，东北大学教授，博士生导师.
通信作者：刘 军，E-mail：liujun@cse.neu.edu.cn.

R-CNN[4],YOLACT[5],SOLOv2[6]等.这些模型通常由骨干网络(如ResNet[7],EfficientNet[8])、区域提议网络[9]、特征金字塔网络[10]和分割头部组成.骨干网络用于提取特征,区域提议网络用于生成感兴趣的区域,特征金字塔网络用于处理不同尺度的特征,分割头部则用于预测每个像素的类别和对象边界框.

以Mask R-CNN为代表的基于锚框的双阶段实例分割算法,采用"先检测,再分割"的方法,首先通过检测算法获得预测框,保证足够的精确率与召回率,之后在预测框内进行语义分割,获得独立的实例.因此,双阶段实例分割算法通常会具有较高的分割精度,但分割速度较慢,难以实现实时分割.YOLACT是一种基于锚框的单阶段实例分割模型.它参考了Mask R-CNN的改进思想,同样增加了一条掩码分支来实现实例分割任务.而以SOLOv2为代表的单阶段无锚框实例分割算法,在具有较高分割速度的同时,分割精度较低.如今,越来越多的领域要求实现实时推理,但实时的高精度实例分割算法仍是一个挑战.

双阶段算法在实例分割任务上具有较好的平均精度(average precision,AP),然而在实时场景下,双阶段算法难以实现目标的实时推理,且其改进算法同样无法有效降低模型推理时间.因此,本文选择单阶段实例分割算法SOLOv2作为基准模型进行改进,研究如何在保证算法的分割精度的基础上,提高SOLOv2推理速度.具体工作如下:

1)提出了一个实时实例分割算法L-SOLOv2,通过压缩模型规模及重调输入图像大小,减少模型推理时间.

2)提出基于SimAM[11-12]注意力机制的骨干网实时实例分割算法SL-SOLOv2,专注于重要特征信息的提取,在避免AP损失的前提下,提高算法的推理速度.

1　SOLOv2算法简介

SOLOv2改进自Wang等[13]提出的基于位置分类的单阶段无锚框算法SOLO.实例分割任务可以解耦为两个像素级别的分类子任务,将待分割图像输入到SOLO中,在一个网络结构中同时实现对目标的分类和分割.SOLO算法是一个简单、直接、快速的分割模型,不需要规范化,但其分割性能不高.主要是因为掩码表示和学习的效率低下、分辨率不够高,不能实现更准确的掩码预测、掩码NMS较慢等.Wang等对SOLO进行改进,并按照实例分割模型的基本原理推导出了一些SOLO变体[14](Vanilla SOLO,Decoupled SOLO,Dynamic SOLO).

如图1所示,SOLOv2相对于SOLO的改进体现在两方面:一是将生成目标对象预测掩码的过程解耦成并行的两个分支,采用掩码核预测分支(mask kernel branch)生成卷积核,通过掩码特征学习分支(mask feature branch)实现对对应卷积的特征映射.图1中,$S\times S$表示图像被划分的特征网格尺寸;$H\times W$表示掩码特征图的空间分辨率(高×宽);C表示语义类别数或特征通道数;E表示中间层的特征通道数;D表示预测的动态卷积核参数量.通过各个分支维度的并行映射与动态卷积融合,最终生成实例分割掩码.二是改进设计了新的非极大值抑制算法matrix NMS,有效减少了模型的推理开销.

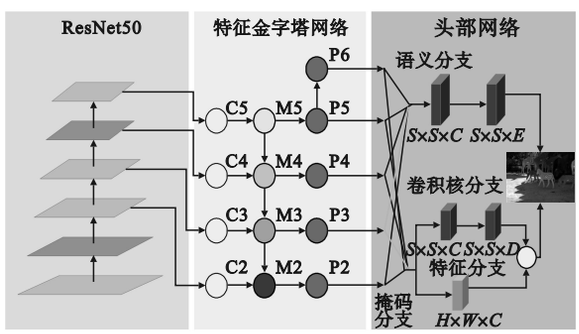


图1　SOLOv2实例分割模型

Fig.1　Instance segmentation model of SOLOv2

2　SOLOv2算法改进

针对实例分割模型推理速度低的问题,本文提出两种解决方案:一是采用轻量化模型作为实时模型的骨干网络,提出基于MobileNetV2[15]骨干网络的改进算法M-SOLOv2;二是通过裁剪或调整输入图像大小来限制输入尺寸的方法和基于修剪的模型压缩方法相结合,提出基于模型压缩和图像大小调整的改进算法L-SOLOv2.

针对模型分割精度不高的问题,本文提出基于SimAM注意力机制的骨干网络SL-SOLOv2,提取特征图中的关键信息,提高实例分割算法性能.

2.1　基于MobileNetV2骨干网络的改进算法

MobileNetV2是Google提出的改进的轻量级卷积神经网络模型,其内部采用倒残差结构

(inverted residuals)[15]来减少模型的运算量以及参数量.传统的残差模块(residual block)[16]采取先进行“降维”,再卷积的方式提取特征,导致逐深度卷积层获得的特征数目不足.因此MobileNetV2采取与传统的残差结构相反的操作,一开始对通道进行“升维”操作.即残差模块采取先“降维”再“卷积提特征”最后“升维”的流程,而MobileNetV2就变成了先“升维”再“卷积提特征”最后“降维”的流程,因此称为inverted residuals.这种设计在降低计算成本的同时,保持了特征的表达能力.

针对算法实时性,提高推理速度的问题,采用的第一种解决方案是选择MobileNetV2轻量级网络作为算法的骨干网络,并对其进行一些改进,得到改进的实时算法M-SOLOv2.将原来网络的最后两层,即平均池化层和1×1卷积层,从网络中移除,同时将激活函数从ReLU6变为Leaky ReLU[17],以保证模型的分割质量.之后选取1,2,4,6层作为输出特征层,通过特征金字塔网络FPN对不同尺寸的输入特征进行融合,其中输入通道数为[24,32,96,320].改进后的网络核心模块如表1所示,h和w分别为输入特征图的高和宽,k和k'分别为输入与最终输出的特征通道数,t为通道扩展因子,s为卷积步长.

表1 改进的MobileNetV2核心模块

Table 1 Core module for improved MobileNetV2

输入	操作	输出
$h\times w\times k$	1×1 conv2d,Leaky ReLU	$h\times w\times tk$
$h\times w\times tk$	3×3 dwise s=s,Leaky ReLU	$\frac{h}{s}\times\frac{w}{s}\times tk$
$\frac{h}{s}\times\frac{w}{s}\times tk$	Linear 1×1 conv2d	$\frac{h}{s}\times\frac{w}{s}\times k'$

平均池化(average pooling)操作在特征图上使用滑动窗口进行卷积操作,将感受野内的特征值取平均作为输出特征,实现下采样,缩小特征图尺寸.由于其采用平均值作为结果,因此可以实现提取整体特征而去除局部特征,但实例分割任务关注每个对象的边缘部分,聚焦于背景到目标的跨越,关注特征的局部特征,需要对位置进行准确定位,因此将平均池化从网络中移除.

当采用倒残差结构时,非线性激活函数ReLU6会破坏特征,出现较大的信息丢失.这是由激活函数ReLU6的自身特性所造成的,它把所有的负值都变为零.当压缩后的特征经过激活函数ReLU6后,会再损失一部分特征信息,模型检测精度出现严重下降.因此这里不采用ReLU6,以减少由其导致的信息损失,使特征信息更完整地保存在网络中.而所采用的Leaky ReLU激活函数则是在输入为负时,为其梯度赋一个小的正值,减少特征信息的丢失.

2.2 基于模型压缩和图像大小调整的改进算法

本文采用的第二种提高推理速度的方式是模型压缩和重调输入特征图的尺寸,实现模型针对实例的实时分割,得到模型L-SOLOv2.

如图2所示,在实现遥感图像实时分割的方法中,对SOLOv2模型进行压缩,并调整图像大小.首先压缩SOLOv2算法中的掩码头结构,将卷积层的数目从4缩小到2,在语义分支和掩码核预测分支分别进行Conv1和Conv2操作,再通过3×3卷积操作获得预测结果.之后分别将掩码特征头的特征通道和掩码特征学习分支的输出通道数设为原来的二分之一,实现对模型的压缩.最后,重新调整多尺度输入训练方法中的图像尺度,并使用“value”模式对输入特征图的大小进行调整,即从多个图像尺度(img_scale)中随机抽取一个元组,之后通过单尺度计算:先取出图像尺度中的大值和小值,再分别进行大值除以长边、小值除以短边的计算,选出比值最小的那个作为缩放因子,最后将缩放因子与真实图片大小相乘,获得调整后的输入图像.

2.3 基于SimAM注意力机制的骨干网络

卷积神经网络通过卷积核实现图像的特征提取,然而,卷积核在特征图像中不同位置获得的特征信息,其重要程度是不同的,获得图像中最重要的特征信息是模型分割精度提高的关键.传统的卷积神经网络通过不断增加网络的深度,来获得不同层次的特征信息,层次越高,特征图像中的语义信息越丰富,但是,仅仅采用增加网络深度的方法来获得特征信息会出现梯度爆炸的现象,最终无法从网络中获取特征信息.

研究表明,注意力机制可以解决由于模型实时化导致的分割精度不足的问题.传统的注意力机制通常沿着通道维度或空间维度进行操作,从而形成一维或二维权重.如图3所示,与现有的注意力机制不同,SimAM注意力机制在不增加参数量的前提下,推断特征图的特征映射三维注意力权重,降低模型的计算量和参数量,提高模型的推理速度.同时由于注意力机制只关注重要特征而忽略无关信息,有助于减少特征信息冗余问题,提高分割精度.

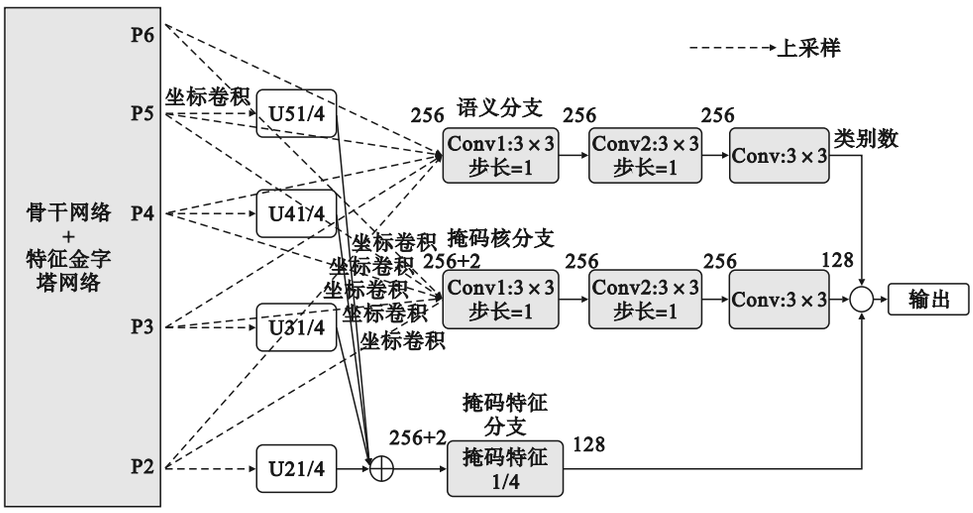


图2　压缩模型框架图

Fig. 2　Framework diagram of compressed model

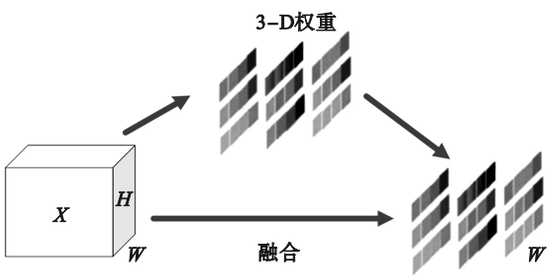


图3　SimAM注意力机制结构图

Fig. 3　Structure diagram of SimAM attention mechanism

本文采用注意力机制模块SimAM,通过重新配置权重来重点关注特征图中的关键信息,而忽略其他信息,在不增加网络深度的前提下提取了更多的局部特征信息,有效避免了传统局部特征提取(深层网络)方法下出现的参数量、计算量剧增及梯度爆炸等问题.如图4所示,将SimAM模块加入到残差模块的Conv2之后.因为SimAM是一种无参注意力机制,可以有效推断出特征图的三维注意力权重,而三维权重在信息搜索方面的能力优于一维和二维权重.

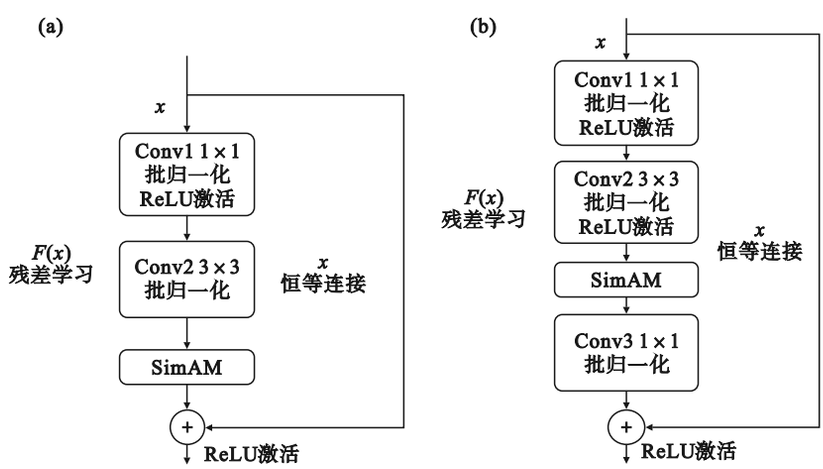


图4　基于注意力机制的残差结构

Fig. 4　Residual structure based on attention mechanism

(a)—基础残差块;(b)—瓶颈残差块.

模型通过在骨干网络ResNet-50的每一层(stage)后加入SimAM注意力模块,增加重点特征信息的权重值,避免特征信息在多次卷积后消失,并将通过无参数SimAM注意力机制后提取的特征图作为FPN特征金字塔网络的输入信息,对不同尺度的特征信息进行融合,提高算法分割精度,得到SL-SOLOv2模型.

2.4　改进的SOLOv2的实时实例分割算法

SL-SOLOv2算法总体架构图如图5所示.首先,采用基于SimAM的ResNet-50骨干网络提取输入遥感图像的特征信息.其次,多尺度特征融合网络FPN取后4层特征进行上采样融合,并对

最深输出层进行最大池化操作,得到5层输出.最后,SOLOv2头结构中的语义分割和掩码分支(掩码核预测分支和掩码特征学习分支)通过上采样(双线性插值)和坐标卷积(增加x和y方向上的特征图)操作分别提供精准的目标分类和实例分割掩码.

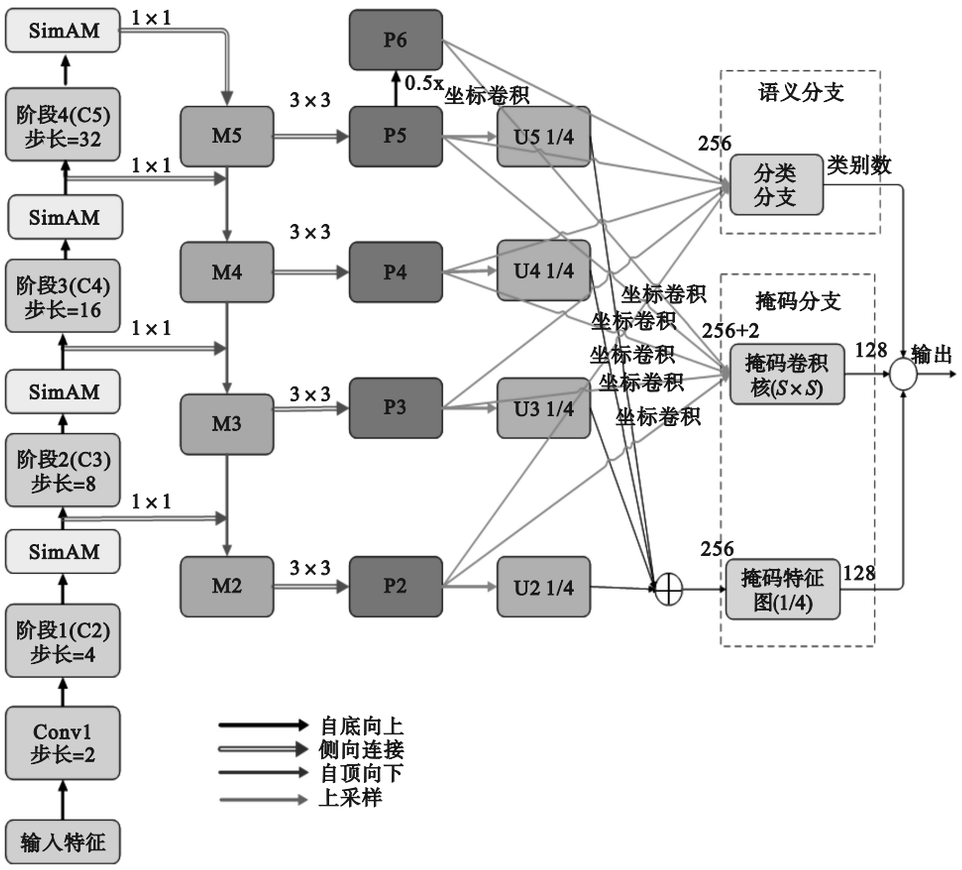


图5 SL−SOLOv2总体架构图

Fig. 5 Overall architecture diagram of SL−SOLOv2

3 实验及结果分析

3.1 实验环境与数据集

本实验在Intel® Xeon® Gold 5218R CPU@ 2.10 GHz(×80)处理器、125.4 GB内存的硬件平台上进行,操作系统为Ubuntu 18.04.2 LTS,GPU采用NVIDIA GeForce RTX 3090,磁盘容量为2.0 TB.实验软件环境为Python 3.9.12,CUDA 12.1以及PyTorch 1.11.0.

本文采用了Su等[18]基于合成孔径雷达的船舶检测数据集SSDD[19].分辨率在1~15 m的SSDD数据集共有1 160张图像.其中,数据集主要分为近岸舰船和远岸舰船图像,近岸图像背景较为复杂,且舰船分布不均匀,整个数据集中共有2 540艘船.在所有的消融和对比实验中,SSDD数据集被随机分为7份和3份,分别作为训练样本(812张图像)和验证样本(348张图像).

3.2 实验设计及训练参数

本文采用的基准模型是SOLOv2−50,学习率(learning rate)、动量(momentum)和权重衰减(weight decay)等模型参数的设置,分别是0.002 5,0.9以及0.000 1.共训练36个轮次,在第27和33个轮次后调整学习率为原来的0.1,并使用梯度下降算法来优化目标函数.

实例分割实验采用MS COCO格式的评估指标mAP,AP_{50},AP_{75},AP_S,AP_M,AP_L以及推理速度对实例分割模型的性能进行评价.AP_{50}与AP_{75}表示IoU阈值分别取0.50与0.75时的AP指标;mAP表示IoU阈值以步长0.05从0.50到0.95所对应AP的平均值;AP_S表示对小目标对象(面积小于32像素×32像素)进行实例分割后获得的AP;AP_M表示对中目标对象(面积大于32像素×32像素且小于96像素×96像素)进行实例分割后获得的AP;AP_L表示对大目标对象(面积大于96像素×96像素)进行实例分割后获得的AP.

IoU(intersection over union)是预测框和真实框面积的交并比.

$$\mathrm{IoU}=\frac{M_{\mathrm{p}}\cap M_{\mathrm{gt}}}{M_{\mathrm{p}}\cup M_{\mathrm{gt}}}. \tag{1}$$

其中:M_{p}和M_{gt}分别为实例分割掩码的预测值和地表真值.

召回率(R)是正确识别出的正样本样例在整个测试集正样本中的概率.

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}. \tag{2}$$

其中:TP表示真实为正且预测为正;FN表示真实为正但预测为负.

精确率(P)是正样本在检测结果中所占的比例.

$$P=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}. \tag{3}$$

其中FP表示真实为负但预测为正.

以精确率与召回率为坐标构成的曲线是PR曲线,AP即PR曲线与坐标轴围成的面积,预测结果的AP为

$$\mathrm{AP}=\int_0^1 P(R)\mathrm{d}R. \tag{4}$$

mAP是各类别的平均精度均值:

$$\mathrm{mAP}=\sum_{i=1}^{M}\frac{\mathrm{AP}_i}{M}. \tag{5}$$

其中:i表示分割样本中第i类对象;M为分割样本的总类别数;AP_i为类别i平均精度.

3.3　实验结果及分析

为了验证改进实例分割算法在SAR(合成孔径雷达)数据集SSDD上的分割性能,本文与几种经典实例分割算法(自顶而下的双阶段算法Mask R-CNN、单阶段实时实例分割算法YOLACT以及本文的基准模型SOLOv2)分别在网络深度为50和101的模型上依据MS COCO评估指标和模型推理速度进行对比分析.表2为各实例分割算法的性能结果.

表2　SL-SOLOv2与各实例分割算法的性能结果

Table 2　Performance results of SL-SOLOv2 and various instance segmentation algorithms

模型	层数	mAP/%	AP_{50}/%	AP_{75}/%	AP_{S}/%	AP_{M}/%	AP_{L}/%	推理速度/(帧·s^{-1})
Mask R-CNN	50	67.9	96.5	86.3	67.9	68.5	—	27.7
	101	67.9	96.6	86.8	67.6	69.8	—	21.8
YOLACT	50	54.2	90.1	61.3	53.7	56.9	—	56.3
	101	54.3	90.8	60.9	53.3	59.5	—	42.0
SOLOv2	50	49.0	77.1	58.8	46.9	56.2	—	26.5
	101	49.2	78.0	60.9	46.4	58.9	—	20.8
L-SOLOv2	50	53.3	85.1	65.6	51.5	60.0	—	51.4
SL-SOLOv2	50	60.3	92.9	75.1	58.8	66.2	—	44.7

如表2所示,双阶段算法Mask R-CNN具有较高的分割精度和较低的推理速度,单阶段实时算法YOLACT具有中等的分割精度和较高的推理速度,而原始的SOLOv2算法在分割精度和推理速度方面都表现一般.改进的SL-SOLOv2实例分割算法推理速度达到了44.7帧/s,比Mask R-CNN提高了17.0帧/s,并且在MS COCO的某些评估指标上接近双阶段算法Mask R-CNN的分割精度.对比实时分割算法YOLACT,改进的SL-SOLOv2算法在mAP,AP_{50}和AP_{75}上分别高出了6.1%,2.8%和13.8%,并且在小目标和中等大小目标的分割精度上分别高出了5.1%和9.3%.对比原始SOLOv2实例分割算法,改进的SL-SOLOv2算法在获得了巨大的精度提升的同时,推理速度提高了18.2帧/s,mAP,AP_{50}和AP_{75}分别提高了11.3%,15.8%和16.3%,小目标和中等大小目标的分割精度也分别提高了11.9%和10.0%.这说明:改进的实例分割算法SL-SOLOv2在精度上不断接近双阶段模型,而在推理速度上大于双阶段模型,兼顾了实例分割的性能与效率,体现了改进算法的有效性.本文中AP类指标的提升/降低均表示绝对差值(AP points).

为了更好分析改进算法与对比算法的分割性能,对近岸分割图片与远岸分割图片的分割结果进行可视化分析.在近岸图像中选择编号分别为000022,000759和001034的3张图片作为分析对象,对应图6a~6c.在远岸图像中选择编号分别为000050,000065,001100的3张图片作为分析对象,对应图7a~7c.图6和图7自上而下分别是GT图(地面真值图),Mask R-CNN-50,YOLACT-50,SOLOv2-50以及SL-SOLOv2-50的实例分割可视化结果图.

从图6中可以发现,面对多尺度目标以及复杂环境背景下的分割任务,4种实例分割方法普遍出现了误检、漏检现象.Mask R-CNN-50在图6a,6c中出现了明显的误检、漏检,在分割时混淆了周围环境与船只.YOLACT-50在图6a中未分割出任何目标,出现了漏检现象,在中间图像中出现了误检现象,并且将图6c的大片海岸误检为船体,分割效果较差.SOLOv2-50在分割图6a时轮廓细节特征未能准确提取,轮廓线萎缩,导致分割结果中断,并且同样在图6c中出现了掩膜叠掩、漏检问题,边界细节提取不充分,分割精度不足.改进的SL-SOLOv2-50算法通过注意力机制重新分配权重,为重要特征信息分配了更大的权重值,而抑制了不相关的信息,增强了细节处理能力,因此图6a的分割轮廓更接近于GT图,然而在图6c中又出现了误检、漏检现象,误检数相较Mask R-CNN-50少,改进算法在复杂环境中分割小目标容易造成失误.

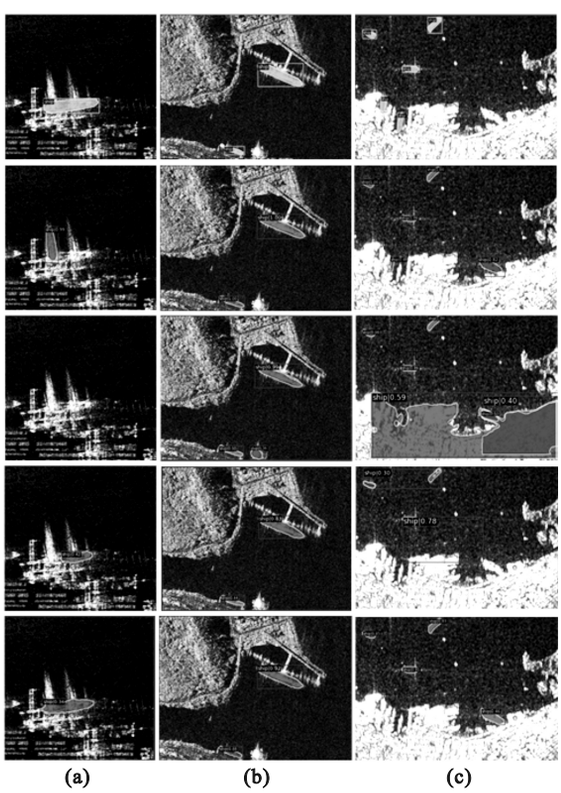


图6 近岸复杂地形分割对比图(每列自上而下分别是GT图、Mask R-CNN-50、YOLACT-50、SOLOv2-50及SL-SOLOv2-50分割结果)

Fig. 6 Near-shore complex terrain segmentation contrast map (from top to bottom in each column are GT, Mask R-CNN-50, YOLACT-50, SOLOv2-50, and SL-SOLOv2-50 segmentation results)

(a)—000022;(b)—000759;(c)—001034.

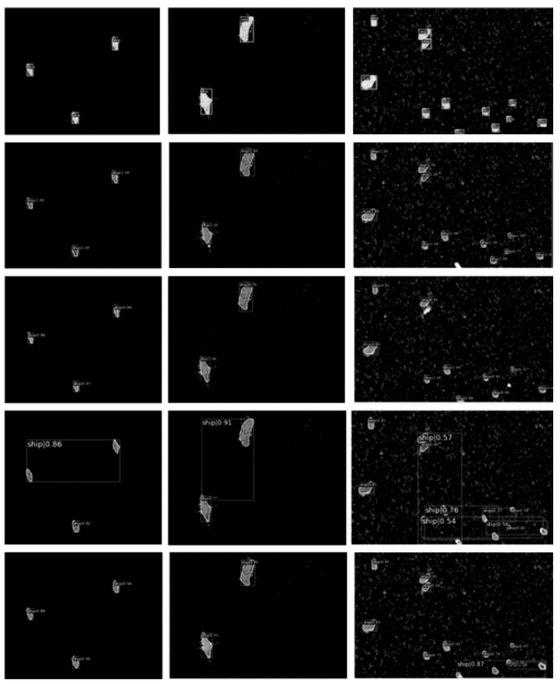


图7 远岸舰船分割对比图(每列自上而下分别是GT图、Mask R-CNN-50、YOLACT-50、SOLOv2-50及SL-SOLOv2-50分割结果)

Fig. 7 Far-shore ship segmentation contrast map (from top to bottom in each column are GT, Mask R-CNN-50,YOLACT-50, SOLOv2-50, and SL-SOLOv2-50 segmentation results)

(a)—000050;(b)—000065;(c)—001100.

仔细分析图7,可以明显看出SSDD数据集中的待分割呈现出明显的目标分布不均现象,且具有分割轮廓不规则以及尺度大小不一的特征.对于稀疏分布的图像(图7a,图7b),4种分割算法都能比较完整地实现分割任务,其中SL-SOLOv2算法在不规则的目标分割中具有较好的特征提取能力,相较于另外3种对比算法,分割轮廓较为准确,细节特征清晰.而对于密集分布的舰船目标(图7c),Mask R-CNN-50和YOLACT-50分别出现一个和两个目标漏检,改进的SL-SOLOv2-50算法相较于原始模型提高了分割的准确度,降低了掩膜叠掩出现的概率,分割轮廓更为明显.

3.4 消融实验

为了对比分析提出的两个实时性方案,选择最佳方案作为本文的实时实例分割算法,其中方案1使用轻量级网络MobileNetV2作为模型的骨干网络,方案2采用压缩模型与调整图像大小的操作提高训练速度.实验结果如表3所示.

为了分析SimAM注意力机制在改进的实例分割算法中的作用.在选定的实时实例分割算法

上添加注意力模块,评估注意力模块添加前后模型的性能.实验结果如表4所示.

表3　不同实时性改进方案的验证结果

Table 3　Verification results of different real-time improvement schemes

方案1	方案2	mAP/%	AP_{50}/%	AP_{75}/%	AP_S/%	AP_M/%	AP_L/%	推理速度/(帧·s^{-1})
		49.0	77.1	58.8	46.9	56.2	—	26.5
√		38.8	68.9	44.1	35.0	51.8	—	33.5
	√	**53.3**	**85.1**	**65.6**	**51.5**	**60.0**	—	**51.4**

表4　SimAM不同位置的验证结果

Table 4　Verification results of SimAM at different positions

策略一	策略二	mAP/%	AP_{50}/%	AP_{75}/%	AP_S/%	AP_M/%	AP_L/%	推理速度/(帧·s^{-1})
		53.3	85.1	65.6	51.5	60.0	—	**51.4**
√		54.9	86.0	67.1	53.4	60.5	—	44.9
√	√	**60.3**	**92.9**	**75.1**	**58.8**	**66.2**	—	44.7

如表3所示,SSDD数据集中的目标对象是小目标和中目标,因此MS COCO评估指标中的AP_L指标不存在.当采用"方案1"来改进算法时,改进算法的mAP相较于原始模型降低了10.2%,AP_{50},AP_{75}分别降低了8.2%,14.7%.同时,数据集中的小目标和中等大小目标的分割精度也有所下降,分别降低了11.9%和4.4%,而算法的推理速度只提高了7.0帧/s.显而易见,使用MobileNetV2轻量化模型作为模型的主干网络的方法有效提高了网络的推理速度,但是算法的分割精度却急剧下降,改进代价较大.

当采用方案2来改进算法时,算法的mAP相较于原始算法提高了4.3%,AP_{50},AP_{75}提高了8.0%和6.8%.同时,在对小目标和中等大小的目标进行分割时,改进算法的分割性能分别提高了4.6%和3.8%.在提高分割精度的同时,改进算法的推理速度提高了24.9帧/s.方案2不仅在分割精度方面相对于原始SOLOv2模型取得了良好的结果,同时还有效提高了算法的推理速度,增强了算法推理的实时性,很好兼顾了遥感图像实例分割的性能和效率.因此选择方案2作为改进的实时分割算法L-SOLOv2.

为了进一步分析注意力机制在模型改进中的效用,本文针对注意力模块不同融合位置进行实验分析.如表4所示,在改进的实时算法L-SOLOv2的基础上对比分析不同的注意力机制连接位置对提升算法分割性能的作用.分别将注意力机制融合在残差模块内部的Conv2之后(策略一)和残差模块外部的Stage之后(策略二).

由表4可知,当同时采用策略一和策略二时,算法的mAP达到了60.3%,比实时算法L-SOLOv2提高了7.0%,比只采用策略一提高了5.4%.AP_{50},AP_{75}比实时模型L-SOLOv2分别提高了7.8%和9.5%,比只采用策略一提高了6.9%和8.0%.当对小目标和中等大小目标进行分割时,改进算法SL-SOLOv2的分割精度相较于实时模型L-SOLOv2提高了7.3%和6.2%,相较于只采用策略一提高了5.4%和5.7%,其中AP_{75}获得了最好的增益.算法的推理速度相较于实时算法L-SOLOv2降低了6.7帧/s,比只采用策略一降低了0.2帧/s.改进的实时实例分割算法SL-SOLOv2在推理速度上有所降低,但是提高的分割精度幅度更大,大致实现了实例分割算法在分割准确率和实时性上的平衡,验证了改进算法的有效性.

4　结　　语

针对实例分割模型的推理时间长和实例分割准确率低的问题,本文以单阶段实例分割模型SOLOv2作为基准,通过模型压缩和重调输入图像尺寸来降低算法计算量,提高算法推理速度,满足实时分割的需求.为了弥补实时需求导致的精度下降,在主干网络融入无参数SimAM注意力模块,在不增加额外参数的前提下,重新分配三维权重大小,提高算法的分割精度.实验表明:在SSDD数据集上,改进算法SL-SOLOv2的mAP达到60.3%,高于原始模型11.3%,整体分割精度大大提高.模型推理速度为44.7帧/s,高于原

始算法18.2帧/s,可以实现快速而精准的实例分割.

参考文献:

[1] Guo M H, Xu T X, Liu J J, et al. Attention mechanisms in computer vision: a survey [J]. *Computational Visual Media*, 2022, 8(3):331-368.

[2] Soylu B E, Guzel M S, Bostanci G E, et al. Deep-learning-based approaches for semantic segmentation of natural scene images: a review[J]. *Electronics*, 2023, 12(12): 2730.

[3] Tian D, Han Y, Wang B Y, et al. Review of object instance segmentation based on deep learning [J]. *Journal of Electronic Imaging*, 2022,31(4): 041205-041205.

[4] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, 2017: 2980-2988.

[5] Bolya D, Zhou C, Xiao F Y, et al. YOLACT: real-time instance segmentation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, 2019: 9157-9166.

[6] Wang X L, Zhang R F, Kong T, et al. SOLOv2: dynamic and fast instance segmentation [J]. *Advances in Neural Information Processing Systems*, 2020, 33: 17721-17732.

[7] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization [C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, 2017: 618-626.

[8] Tan M X, Pang R M, Le Q V. EfficientDet: scalable and efficient object detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 10781-10790.

[9] Song P H, Li P T, Dai L H, et al. Boosting R-CNN: reweighting r-cnn samples by RPN's error for underwater object detection[J]. *Neurocomputing*, 2023,530:150-164.

[10] Chen P Y, Hsieh J W, Wang C Y, et al. Parallel residual bi-fusion feature pyramid network for accurate single-shot object detection [J]. *IEEE Transactions on Image Processing*, 2019, 30: 9099-9111.

[11] Yang L X, Zhang R Y, Li L D, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks [C]// International Conference on Machine Learning. Virtual: PMLR, 2021: 11863-11874.

[12] Qin X Y, Li N, Weng C, et al. Simple attention module based speaker verification with iterative noisy label detection [C]// 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Singapore, 2022: 6722-6726.

[13] Wang X L, Kong T, Shen C H, et al. SOLO: segmenting objects by locations [C]// Computer Vision-ECCV 2020: 16th European Conference. Cham: Springer, 2020: 649-665.

[14] Wang X L, Zhang R F, Shen C H, et al. SOLO: a simple framework for instance segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(11): 8587-8601.

[15] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, 2018: 4510-4520.

[16] Gao S H, Cheng M M, Zhao K, et al. Res2Net: a new multi-scale backbone architecture[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 43(2): 652-662.

[17] Zhou L M, Li Y H, Rao X H, et al. Feature enhancement-based ship target detection method in optical remote sensing images[J]. *Electronics*,2022,11(4): 643.

[18] Su H, Wei S J, Liu S, et al. HQ-ISNet: high-quality instance segmentation for remote sensing imagery [J]. *Remote Sensing*, 2020, 12(6): 989.

[19] Zhang T W, Zhang X L, Li J W, et al. SAR ship detection dataset (SSDD): official release and comprehensive data analysis[J]. *Remote Sensing*, 2021, 13(18):3960.