

DOI: 10.11992/tis.202210039

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20230609.1351.002>

基于注意力机制的 SOLOA 船舶实例分割算法

孙雨鑫¹, 苏丽^{1,2}, 陈禹升¹, 苑守正¹, 孟浩^{1,2}

(1. 哈尔滨工程大学 智能科学与工程学院, 黑龙江 哈尔滨 150001; 2. 哈尔滨工程大学 船舶装备智能化技术与应用教育部重点实验室, 黑龙江 哈尔滨 150001)

摘要: 目前, 可见光船舶图像的实例分割仍然是较有挑战性的工作。由于船舶图像场景复杂多变, 多数实例分割算法无法对复杂场景下的船舶图像进行有效分割。本文提出了基于注意力机制的依靠位置分割目标 (attention based segmenting objects by locations, SOLOA) 船舶实例分割算法, 针对网络特征图里实例信息不完善、背景干扰较多的问题, 使用空间注意力机制来充分利用分类特征中的实例信息, 建模图像实例间的相互关系并与分割特征相融合。实验结果表明, 在新构建的船舶图像数据集上进行训练和测试, 改进的网络模型能有效地增强网络特征中的实例信息、减弱背景的干扰。SOLOA 算法的船舶实例分割准确率高于其他算法, 可以很好地适应复杂场景下的船舶分割需求。

关键词: 船舶目标; 实例分割; 复杂海上场景; 深度学习; 卷积神经网络; 注意力机制; 单阶段实例分割; 可见光图像
中图分类号: TP183 **文献标志码:** A **文章编号:** 1673-4785(2023)06-1197-08

中文引用格式: 孙雨鑫, 苏丽, 陈禹升, 等. 基于注意力机制的 SOLOA 船舶实例分割算法 [J]. 智能系统学报, 2023, 18(6): 1197-1204.

英文引用格式: SUN Yuxin, SU Li, CHEN Yusheng, et al. SOLOA ship instance segmentation algorithm based on attention[J]. CAAI transactions on intelligent systems, 2023, 18(6): 1197-1204.

SOLOA ship instance segmentation algorithm based on attention

SUN Yuxin¹, SU Li^{1,2}, CHEN Yusheng¹, YUAN Shouzheng¹, MENG Hao^{1,2}

(1. College of Intelligent Science and Engineering, Harbin Engineering University, Harbin 150001, China; 2. Key Laboratory of Ministry of Education on Intelligent Technology and Application of Marine Equipment, Harbin Engineering University, Harbin 150001, China)

Abstract: At present, the instance segmentation of visible ship images remains a highly challenging task. Most instance segmentation algorithms cannot effectively segment ship images in complex scenes due to the intricate and variable nature of ship images. A segmenting objects by locations based on attention (SOLOA) algorithm for ship instance segmentation, which utilizes the spatial attention mechanism to maximize the instance information in the classification features, is proposed in this paper. Here, the interrelationships between the image instances are modeled and fused with segmentation features. Training and testing results of the newly constructed ship image dataset show that the improved network model can effectively enhance the instance information in the network features and reduce the background interferences. The accuracy of ship instance segmentation by the SOLOA algorithm is higher than that of other algorithms; hence, the proposed algorithm can be effectively adapted to meet the demands of ship segmentation in complex scenes.

Keywords: ship object; instance segmentation; complex scene; deep learning; convolution neural network; attention; one-stage instance segmentation; visible light image

近年来, 芯片的生产能力以及计算机技术不断提升, 包括计算机视觉在内的多个人工智能领

域蓬勃发展, 人类已步入数字化、智能化时代。随着船舶数量、载运量不断增加, 船舶交通日益稠密, 人们对高性能的船舶智能监控设备的需求越来越高。船舶实例分割作为智能监控的重要基础, 直接影响了监控的质量, 因此寻求稳定且高效的船舶实例分割算法显得尤为重要。

然而目前多数船舶检测和分割算法大都基于合成孔径雷达 (synthetic aperture radar, SAR) 图

收稿日期: 2022-10-31. 网络出版日期: 2023-06-12.

基金项目: 国家重点研发计划项目 (2019YFE0105400); 船舶态势智能感知系统研制项目 (MC-201920-X01); 中央高校基本科研业务费专项资金-博士研究生创新基金项目 (3072022GIP0403).

通信作者: 苏丽. E-mail: suli406@hrbeu.edu.cn.

像^[1-3]以及红外图像^[4],较少有文献研究可见光船舶图像的实例分割任务。此外,红外图像存在背景和目标的对比度较低、信噪比较低以及细节特征不全等问题。而SAR图像含有大量伽玛分布的斑点噪声,且存在图像对比度较低、目标细微边缘不易区分^[5]等问题。若将SAR图像、红外图像应用到船舶实例分割任务,往往还需要图像降噪等预处理操作,难以直接应用。而谈到可见光图像,其在信噪比、图像对比度以及图像细节特征上均优于红外和SAR图像,它具有丰富的颜色、纹理信息和较高的分辨率,便于检查和观测。除此之外,可见光图像易于获取,使得基于可见光图像的船舶实例分割算法拥有广泛的应用空间。同时相较于SAR和红外图像的远景实例分割任务,可见光的实例分割可以有效应对近距离大目标,从而弥补对海上环境感知精度灵敏度较低的缺陷^[6]。基于可见光图像的实例分割,在通用物体分割领域已经涌现了许多优秀的算法^[7-19]并且在多分类的复杂任务中取得了卓越的性能。两阶段算法掩码分割网络^[7]在通用分割领域性能优越,常被应用到诸如汽车、鸟类实例分割等其他特定领域。两阶段算法的优势是其采用先检测后分割的范式来处理图像中的目标,具有一定的鲁棒性,不易被图像中过多的背景信息所干扰。而近期许多以速度见长的单阶段实例分割算法也不断涌现,其中性能和速度名列前茅的SOLOv2

算法^[20]具有巨大的应用价值。然而,在使用自建的船舶实例分割数据集对SOLOv2算法进行训练及测试时发现,对于普通海天背景下SOLOv2算法具有可观的性能,但在船舶密集、背景复杂的场景下,由于背景干扰较多,SOLOv2难以对船舶目标进行精确有效地分割。

本文根据上述问题,采用空间注意力机制来建模分类特征中的相互关系,并将其融合到分割特征中辅助船舶图像的分割,有效提高了复杂海上背景下的船舶实例分割性能。

1 SOLOv2 算法的原理

给定一个任意的图像输入,实例分割算法需要判断图片中是否存在实例,若存在则算法返回所有实例的分割掩码。SOLOv2算法把实例分割问题转化为语义类别预测问题和实例掩码预测问题。主要思路是将图像划分为 $S \times S$ 个网格,从而以网格为样本点来预测不同实例的类别和掩码。原理如图1所示,输入图像中的2个船舶实例的中心落在2个不同的网格中,则对应的2个网格分别负责这2个实例的掩码和类别预测,掩码和语义类别的预测分别由2个独立的分支网络进行预测。通过采用特征金字塔网络(feature pyramid networks, FPN)^[21]还可以将不同大小的实例有效地进行区分从而避免同一位置出现密集实例难以分割的情况。

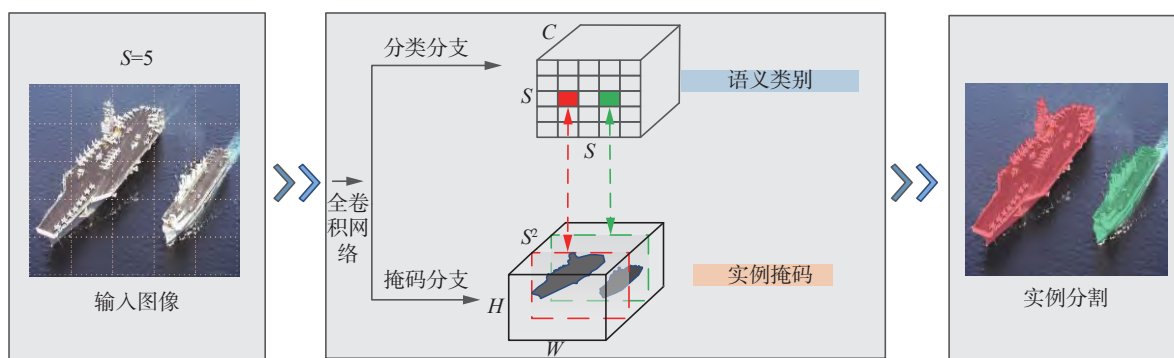


图1 SOLOv2 算法的原理

Fig. 1 Theory of SOLOv2 algorithm

1.1 语义分类

如图1的分类分支主要负责预测所有网格不同实例类别的概率(包含背景类的预测),对于 C 个类别的数据集,如果我们将输入的图片划分为 $S \times S$ ($S=5$)个网格,该预测网络的输出是 $S \times S \times C$ 维度的张量,通道维度包含了不同类别实例的概率。在这种设计下每个网格都有一个语义类别的预测来表示该网格位置存在不同类别的可能性。

1.2 实例掩码预测

针对之前划分的 $S \times S$ 个网格,掩码分支负责

预测每个网格对应 S^2 个实例的掩码,从而产生 $H \times W \times S^2$ 维度的张量来表示全图每个网格实例的掩码预测。 H 和 W 分别表示长和宽。如图1所示,分类分支的类别预测和分割分支的掩码预测是一一对应的,若设网格坐标为 (i, j) ,则其对应分割分支中第 $k = i \times S + j$ 个通道的输出。由于实例的掩码预测和不同位置的网格是相互关联的,我们需要不同的通道激活不同位置的实例从而产生具有位置敏感性的特征。然而在一般产生预测掩码的全卷积网络(fully convolutional networks,

FCN)^[22]中,由于卷积网络自身存在的空间不变性,产生的特征并不满足该网络的要求。因此,SOLOv2算法引入了坐标卷积^[23]运算,在图像输入时产生2个输入图像等大小的张量,并将其数据归一化到 $[-1,1]$ 用来表示 x 和 y 坐标,最后将该张量并到图像输入的通道维,也就是说对于 $H \times W \times D$ 维度的输入图像(H 和 W 分别表示长和宽, D 为通道数目),经过坐标卷积运算后转化为 $H \times W \times (D+2)$ 维度输入到后面的卷积网络中。

1.3 可视化分析

为了更直观地分析SOLOv2算法在船舶实例分割任务上存在的问题,本文采用ResNet101作为特征提取网络的SOLOv2算法,在构建的船舶实例分割数据集上进行36个周期迭代的训练以获得效果最好的SOLOv2模型,并在1044张图片的测试集上进行图像的可视化分析。通过可视化分析图2发现SOLOv2对于单一背景的图像具有良好的分割性能,而对船舶实例密集、船舶实例特征相似和背景复杂的图像的分割效果较差,出现了误分割等一系列问题。



图2 SOLOv2 船舶实例分割可视化

Fig. 2 SOLOv2 for ship instance segmentation visualization graph

从SOLOv2的实例掩码预测的网络结构可以看出,SOLOv2算法针对不同实例的掩码是由一组对位置敏感的卷积核与融合的分割特征进行卷积运算得出的,因此实例分割算法的性能受到两方面影响:一方面卷积核需要对全图像的特征进行逐一地分类判断,在没有区域筛选限制的情况

下,背景中的干扰极易影响到卷积核对实例的区分;另一方面融合的分割特征不能有效地对不同位置的实例产生不同程度的响应从而影响卷积核对不同实例的区分。

由于SOLOv2算法主要从卷积核优化的角度出发,利用坐标卷积增加位置信息来提升算法对实例位置的敏感性,而对于分割特征还缺少一定的措施来保证其对实例的区分性。因此本文考虑对融合的分割特征进行改进,寻求一定的方法使分割特征中的实例信息更加明显,让学习到的动态卷积核能更有效地区分不同的实例。通过建模网格间实例的相对信息,构建基于注意力机制的依靠位置分割目标(attention based segmenting objects by locations, SOLOA)算法。

2 SOLOA 船舶实例分割算法

针对于复杂海天背景下SOLOv2算法出现的误分割问题,本文通过分析SOLOv2的原理,构建对于复杂场景更稳定的基于注意力机制的依靠位置分割目标船舶实例分割算法。

2.1 网络结构

本文在SOLOv2原有的网络结构上进行改进,设计了将分类特征引入分割特征的注意力(Attention)模块,充分的利用分类网络的信息,建模网格间实例的相对信息,构建SOLOA算法。

2.1.1 分类网络

如图3所示,分类网络依托于特征金字塔网络的多尺度输入,根据特征金字塔网络输出尺度的不同选取合适的网格大小 S 值,后续将图像进行双线性插值到 $S \times S$ 大小。后面网络包括语义分类网络和卷积核预测网络,2个网络都采用4个卷积层作为特征提取。同时2个网络的权重在不同特征层级上是共享的。卷积核预测网络预测 D 维度的卷积核参数与 S^2 个网格对应,对于 1×1 卷积来说 D 等于卷积输入通道数,而对于 3×3 卷积, D 则等于输入通道数目的9倍。语义分类网络经过特征提取后通过最后的 1×1 卷积来产生 C 个通道的类别预测输出。

2.1.2 注意力模块

与SOLOv2算法不同的是,分析发现通过简单的叠加融合不同层的分割特征再进行上采样后的融合特征中具有许多干扰信息,使其在复杂的场景下无法进行有效分割。而分类网络将输入的特征插值到 $S \times S$ 的网格特征时也包含大量实例的信息,可以用其辅助分割产生更具有对比度、干扰较少的融合特征。另一方面,输入含有船舶的图像,分类特征中包含船舶实例的网格必然会获得较高的激活响应,这样根据分类特征的激活情

况就可以获得全图中实例的分布,从而利用分类网络中的实例分布信息来辅助 SOLOv2 算法的分割。

为了达到这一目的,引入了 Attention 模块来建模分类特征间的关系并叠加到分割特征中,来增强分割特征中存在实例信息而减弱背景的干扰,获取全图视角的实例分布信息。具体的 Attention

模块结构如图 4 所示,以 $S \times S$ 大小的分类特征 (Categories Feature) 为输入,通过 1×1 卷积产生 Q 、 K 、 V 等 3 个向量,并通过 Q 和 K 向量相乘获得特征间相互关系矩阵进行归一化 (softmax) 操作运算,最后与 V 向量相乘获取注意力矩阵并叠加到输入的网络特征中获得最终的注意力特征。

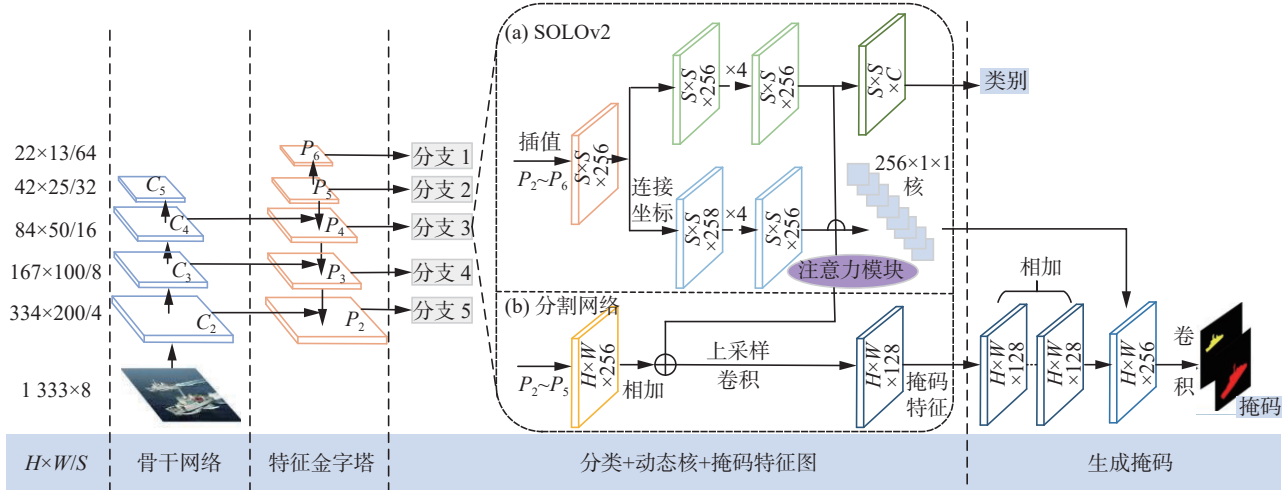


图 3 SOLOA 框架

Fig. 3 SOLOA framework

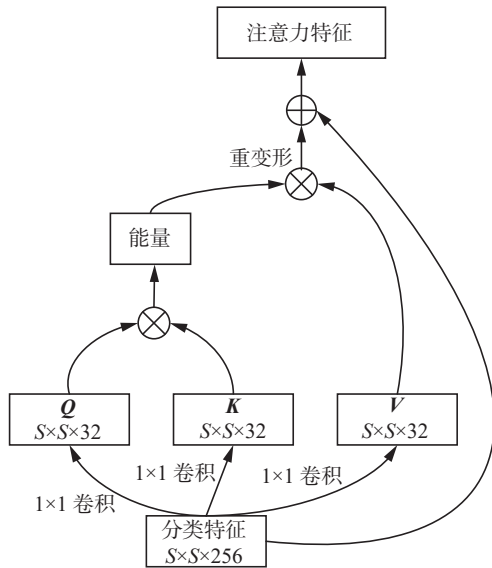


图 4 Attention 模块

Fig. 4 Attention module

2.1.3 分割网络

分割网络用来产生对实例敏感掩码预测。首先为了得到能包含多尺度特征的融合特征图,网络利用特征金字塔的 $\{P_2, P_3, P_4, P_5, P_6\}$ 层的特征图合并叠加本文引入的空间注意力模块的输出产生多尺度的实例敏感特征图,然后根据不同尺度的输入重复引入 3×3 卷积、组归一化^[24]、线性修正和 2 倍双线性插值上采样,使得 4 层网络都变为原图大小的 $1/4$, 然后进行叠加并通过 1×1 卷积网络调节通道融合特征信息,最终利用分类网络学

习到的动态卷积核来产生最终的预测,细节如图 3 所示,实例敏感特征图对于产生精确的分割掩码意义重大。

2.2 损失函数

采用常用的多任务损失函数来进行训练:

$$L = L_{\text{cate}} + \lambda L_{\text{mask}} \quad (1)$$

式中:分类损失 L_{cate} 采用的是 Focal Loss^[25],掩码预测损失 L_{mask} 的计算式为

$$L_{\text{mask}} = \frac{1}{N_{\text{pos}}} \sum_k \mathbb{I}_{\{p_{i,j}^* > 0\}} d_{\text{mask}}(m_k, m_k^*) \quad (2)$$

式中: $i = k/S$, $j = kS$, N_{pos} 代表正样本的数目, p^* 和 m^* 分别代表分类和掩码预测任务的真实标记值, \mathbb{I} 表示指示函数。

在比较不同的 Focal Loss、BCE Loss 以及 Dice Loss 等常用的分割预测损失函数后,选取 Dice Loss 作为算法的最终损失:

$$L_{\text{Dice}} = 1 - D(p, q) \quad (3)$$

D 为 Dice 系数,详细公式为

$$D(p, q) = \frac{2 \sum_{x,y} (p_{x,y} \times q_{x,y})}{\sum_{x,y} p_{x,y}^2 + \sum_{x,y} q_{x,y}^2} \quad (4)$$

式中 $p_{x,y}$ 和 $q_{x,y}$ 表示坐标为 (x,y) 的像素点在预测的掩码 p 和真实掩码 q 中的值。

2.3 推理阶段

推理阶段同 SOLOv2 一致。给定一张输入图像,

先将其导入特征提取和特征金字塔网络以及其后的分类和分割网络,产生每张图多达3600多个的掩码及类别预测,最后通过矩阵非极大值抑制算法筛选出预测效果最好的实例掩码进行输出。

3 实验与结果分析

3.1 算法平台

本次实验的硬件平台主要采用CPU为Xeon E5-2620,内置一块NVIDIA GeForce GTX 1080Ti GPU。软件平台选择以Linux Ubuntu 16.04、Pytorch 1.5、CUDA 10.1在商汤发布的MMDetection 1.0框架上进行训练,保证模型训练的速度和稳定性。

3.2 数据集构建

本文主要以各类典型的目标为主要对象,其中包括军舰、渔船、货船、快艇、帆船等多类典型的目标。由于数据集中具体实例情况直接影响模型的训练结果,因此,收集图片时充分考虑了实例的角度、位置、距离、数目以及环境等不同因素。主要的数据来源包括3部分:通过实船采集的海上图片、抽取通用标准数据集(VOC^[26]、COCO^[27])中含有船舶的图片、网络上下载的船舶图片。数据集图片总数为5225张,其中4181张作为训练集,1044张作为测试集。数据集图片示例如图5所示。

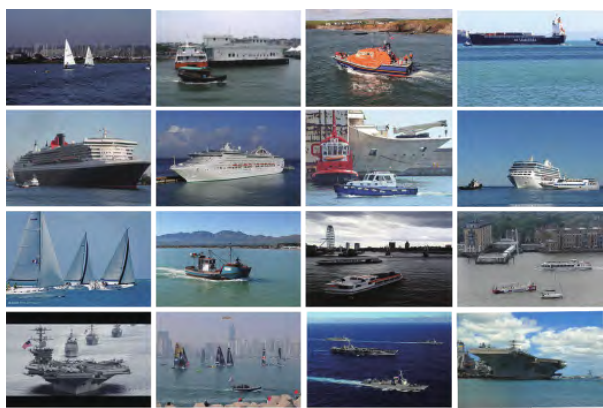


图5 船舶实例分割数据集

Fig. 5 Ship instance segmentation dataset

3.3 模型训练

模型训练使用ImageNet^[28]数据集预训练的ResNet模型^[29]进行初始化权重。初期设置了相对较大的学习率即0.01实现模型训练的初期损失函数的快速下降,然而在实验过程中出现了损失函数收敛较慢、振荡和爆炸等问题。经过多次测试,最后选取较小的学习率即0.00125进行训练,并采用warmup的形式保证训练初期的稳定性,最终采用70000次迭代,并为了保证模型平稳收敛在训练的第60000次迭代时将学习率缩小到0.00013左右。

模型的训练损失情况如图6所示,在训练的初期损失下降较快,由于引入了注意力模块在10000到20000次迭代时出现了一定程度的缓慢下降,但在30000次迭代后损失逐渐收敛,训练达到60000次迭代时由于调整学习率,以较小的步长进行学习,整体损失基本收敛。

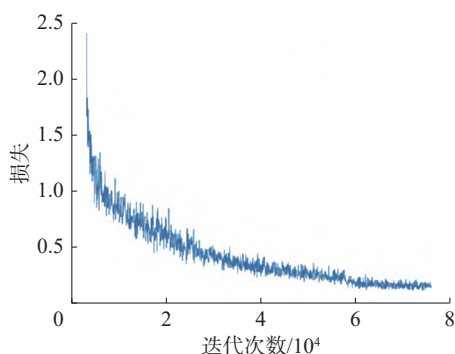


图6 SOLOA模型训练损失值

Fig. 6 Loss of the SOLOA model training

3.4 可视化实验

主要从改进前后的可视化分割效果和融合特征的可视化图2个方面来说明SOLOA算法的效果。

采用ResNet-101为特征提取网络,超参数使用0.00125的学习率和0.9的动量来训练SOLOv2及其SOLOA,最终的分割结果对比如图7所示。其中图7(a)为SOLOv2算法的分割结果图,图7(b)为改进后的SOLOA分割效果图,可见改进后实例的掩码质量得到有效的提高,同时一定程度降低了对背景的误分割。



(a) SOLOv2 分割结果

(b) SOLOA 分割结果

图7 实例分割结果可视化对比

Fig. 7 Visual comparison of instance segmentation results

为了更清晰地体现改进的作用机制,选取含有64通道融合特征的模型进行训练,并和同样训练参数的SOLOv2算法进行对比,其效果如图8所示。图8(a)为网络输入的原图,图8(b)为SOLOv2算法融合特征图的输出,通过观察可以发现许多通道同时激活了图像中的船舶和背景,如图中标记的2个通道,背景和实例之间的对比度较低,在后期的 1×1 卷积中引入额外的干扰信息,导致分割效果不好。在引入了空间注意力模块后的SOLOA算法的特征图可视化如图8(c),可见通道抑制了背景信息的干扰,同时背景和实例的对比度更高,且大多数通道都仅仅激活了图像中的船舶信息,减少背景的干扰使得最终 1×1 卷积核的掩码的预测输出效果更好。

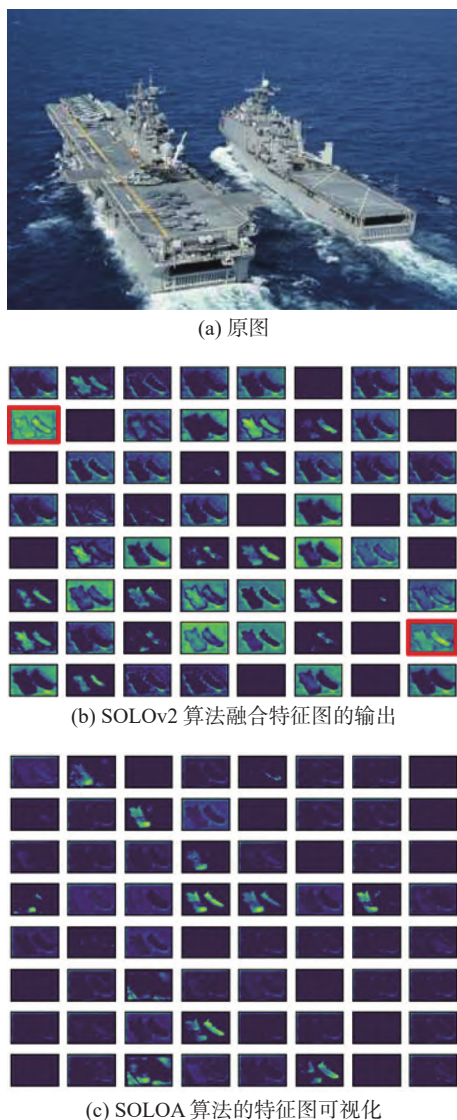


图8 特征图可视化

Fig. 8 Feature map visualization

3.5 消融实验

考虑到注意力模块引入方式的多样性,本文对于相加和相乘的引入方式进行了测试。实验选

择标准COCO数据集的平均精度(average precision, AP)指标来评估算法的性能,AP指标考虑了预测掩码和真实掩码的交并比(intersection over union, IoU)、准确率、召回率等指标,综合的反映了算法的整体性能。通过计算获得图像中每个预测掩码和真实掩码的IoU,通常评估时会选择一个合适的IoU阈值来判断预测结果是否正确。划分好正确预测和错误预测后,对照数据集中所有标注的实例预测就可以计算平均精度。

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

式中: P_{TP} 表示真实标记为正,预测也为正的情况; P_{FP} 表示真实标记为负,但预测为正的情况; P_{FN} 表示真实标记为正,但预测为负的情况; P_{TN} 表示真实标记为负,但预测为负的情况。具体主要包括指标 P_{AP} 、 P_{AP50} 、 P_{AP75} 来验证所有类别实例分割预测的综合情况和 P_{APS} 、 P_{APM} 、 P_{APL} 等3个指标来表示图像中的小目标、中目标以及大目标分割的性能。 P_{AP50} 和 P_{AP75} 分别表示以IoU阈值是0.5和0.75来评价预测样本的精度值。 P_{AP} 则为综合指标,表示使用IoU阈值(范围为0.5~0.95,步长为0.05)来评价预测样本的精度值,即 P_{AP50} 到 P_{AP95} 范围内的平均精度。最后的消融实验结果如表1所示。

表1 注意力引入方式的对比

Table 1 Comparison of the ways of using attention %

方式	P_{AP}	P_{APS}	P_{APM}	P_{APL}
相加	49.8	26.1	46.9	71.5
相乘	49.0	23.0	46.9	71.2
无注意力	48.6	25.7	45.4	70.0

通过观察数据可以发现,相比未改进前无注意力的算法,采用相乘的形式将实例信息引入分割特征虽然对中、大目标有一定的提升效果,但会极大影响对小目标的分割效果。而采用相加的形式引入注意力特征并不会降低小目标的分割准确率同时会极大地提升中、大目标的分割效果。

3.6 主要实验结果

为了全面的评估改进后算法的性能,本文采用ResNet-50和ResNet-101这2种特征提取的骨干网络,分别进行训练SOLOA和一些其他主流的实例分割算法,包括YOLACT^[8]、SOLOv2^[20]和Polarmask^[30]其实例分割结果如表2所示。

相比SOLOv2,由于采用分类特征来辅助分割显著地提高了网络的表示能力,改进后的SOLOA算法在轻量级的ResNet-50网络中有显著的提高效果,相比改进前的网络平均精度提升3.6%。而采用更复杂的ResNet-101-FPN主干网络并采用3倍迭代训练的模型性能最好,AP可以达到49.8%,

相比 SOLOv2 算法 AP 共增长了 1.2%。并且改进后的 SOLOA 其小、中、大目标的分割效果都有了一定的提升。同时, 相比 YOLACT 和 Polarmask, SOLOA 算法性能也是最优的。此外, 为了验证改进后 SOLOv2 算法的运行速度, 本小节采用单张

1080Ti 显卡对模型进行速度测试和模型复杂度(参数量)比较, 实验结果如表 3 和表 4 所示。增加了额外的融合机制并没有损失过多的速度性能, 也没有增加过多的模型参数量, 总体来看改进后的算法满足船舶实例分割的实时性需求。

表 2 实例分割结果对比

Table 2 Comparison of the instance segmentation results

%

算法	骨干网络	P_{AP}	P_{AP50}	P_{AP75}	P_{APS}	P_{APM}	P_{APL}
YOLACT [8]	Res-101-FPN (3×)	46.2	77.2	47.1	24.9	44.1	65.7
Polarmask [21]	Res-101-FPN (1×)	39.7	77.3	34.8	24.4	39.4	52.7
SOLOv2 [11]	ResNet-50-FPN (1×)	42.1	71.2	43.1	18.8	37.9	64.9
SOLOv2 [11]	ResNet-101-FPN (3×)	48.6	77.3	50.4	25.7	45.4	70.0
SOLOA	ResNet-50-FPN (1×)	45.7	78.1	46.8	20.0	44.1	68.4
SOLOA	ResNet-101-FPN (3×)	49.8	78.6	52.1	26.1	46.9	71.5

表 3 改进前后运行速度对比

Table 3 Comparison of the running speed before and after improvement

f/s

骨干网络	SOLOv2	SOLOA
ResNet-50	11.6	10.4
ResNet-101	9.6	8.6

表 4 改进前后模型参数量对比

Table 4 Comparison of the number of model parameters before and after improvement

骨干网络	SOLOv2/10 ⁶	SOLOA/10 ⁶
ResNet-50	46.2	46.4
ResNet-101	65.5	65.8

4 结束语

本文针对 SOLOv2 算法在船舶实例分割领域存在相似实例和背景误分割的问题, 提出了一种利用分类特征中的实例分布信息进行注意力建模来融合分割特征的方法, 构建了 SOLOA 船舶实例分割算法。相比于 SOLOv2 算法, SOLOA 算法对密集实例能有效分割, 具有较好的位置敏感性和抗背景干扰能力, 对不同尺度的实例尤其是中大尺度的实例均有明显的精度提升, 在保证算法运行速度的前提下全面提升了分割精度, 可以满足船舶实例分割的需求。

参考文献:

- [1] HUANG Guoquan, WAN Zining, LIU Xinggao, et al. Ship detection based on squeeze excitation skip-connection path networks for optical remote sensing images[J]. *Neurocomputing*, 2019, 332: 215–223.
- [2] SU Hao, WEI Shunjun, LIU Shan, et al. HQ-ISNet: high-quality instance segmentation for remote sensing imagery[J]. *Remote sensing*, 2020, 12(6): 989.
- [3] WEI Shunjun, ZENG Xiangfeng, QU Qizhe, et al. HRSID: a high-resolution SAR images dataset for ship detection and instance segmentation[J]. *IEEE access*, 2020, 8: 120234–120254.
- [4] 胡欣, 郭庆昌. 基于红外图像的海上船舶分割算法研究[J]. *水雷战与舰船防护*, 2009, 17(1): 5–8.
- [5] HU Xin, GUO Qingchang. Research on ship segmentation algorithm based on IR image[J]. *Mine warfare & ship self-defence*, 2009, 17(1): 5–8.
- [6] SUN Yuxin, SU Li, LUO Yongkang, et al. Global mask R-CNN for marine ship instance segmentation[J]. *Neuro-computing*, 2022, 480: 257–270.
- [7] SUN Yuxin, SU Li, LUO Yongkang, et al. IRDCLNet: instance segmentation of ship images based on interference reduction and dynamic contour learning in foggy scenes[J]. *IEEE transactions on circuits and systems for video technology*, 2022, 32(9): 6029–6043.
- [8] HE Kaiming, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2980–2988.
- [9] BOLYA D, ZHOU Chong, XIAO Fanyi, et al. YOLACT: real-time instance segmentation[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2020: 9156–9165.
- [10] GAO Naiyu, SHAN Yanhu, WANG Yupei, et al. SSAP: single-shot instance segmentation with affinity pyramid[J]. *IEEE transactions on circuits and systems for video technology*, 2021, 31(2): 661–673.
- [11] BOLYA D, ZHOU Chong, XIAO Fanyi, et al. YOLACT++ better real-time instance segmentation[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2022, 44(2): 1108–1121.
- [12] SUN Yuxin, SU Li, CUI Haohao, et al. Ship instance segmentation in foggy scene[C]//2021 40th Chinese Control

- Conference. Shanghai: IEEE, 2021: 8340–8345.
- [12] LIU Shu, QI Lu, QIN Haifang, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8759–8768.
- [13] ZHANG Rufeng, TIAN Zhi, SHEN Chunhua, et al. Mask encoding for single shot instance segmentation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 10223–10232.
- [14] CAI Zhaowei, VASCONCELOS N. Cascade R-CNN: high quality object detection and instance segmentation[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2021, 43(5): 1483–1498.
- [15] CHEN L C, HERMANS A, PAPANDREOU G, et al. MaskLab: instance segmentation by refining object detection with semantic and direction features[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 4013–4022.
- [16] KIRILLOV A, WU Yuxin, HE Kaiming, et al. PointRend: image segmentation As rendering[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 9796–9805.
- [17] CHENG Tianheng, WANG Xinggang, HUANG Lichao, et al. Boundary-preserving mask R-CNN[C]//European Conference on Computer Vision. Cham: Springer, 2020: 660–676.
- [18] CHEN Kai, PANG Jiangmiao, WANG Jiaqi, et al. Hybrid task cascade for instance segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2020: 4969–4978.
- [19] WANG Shaoru, GONG Yongchao, XING Junliang, et al. RDSNet: a new deep architecture for reciprocal object detection and instance segmentation[C]//Proceedings of the AAAI conference on artificial intelligence. Washington, DC: AAAI press 2020, 34(7): 12208–12215.
- [20] WANG X, ZHANG R, KONG T, et al. SOLOv2: dynamic and fast instance segmentation[J]. *Advances in neural information processing systems*, 2020, 33: 17721–17732.
- [21] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 936–944.
- [22] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[C]//IEEE Transactions on Pattern Analysis and Machine Intelligence. Piscataway: IEEE, 2016: 640–651.
- [23] LIU R, LEHMAN J, MOLINO P, et al. An intriguing failing of convolutional neural networks and the CoordConv solution[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: ACM, 2018: 9628–9639.
- [24] WU Yuxin, HE Kaiming. Group normalization[C]//European Conference on Computer Vision. Cham: Springer, 2018: 3–19.
- [25] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2020, 42(2): 318–327.
- [26] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge[J]. *International journal of computer vision*, 2010, 88(2): 303–338.
- [27] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//European Conference on Computer Vision. Cham: Springer, 2014: 740–755.
- [28] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84–90.
- [29] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770–778.
- [30] XIE Enze, SUN Peize, SONG Xiaoge, et al. PolarMask: single shot instance segmentation with polar representation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 12190–12199.

作者简介:



孙雨鑫, 博士研究生, 主要研究方向为计算机视觉。



苏丽, 副教授, 博士生导师, 主要研究方向为环境感知与智能控制、智能船舶、机器视觉检测技术、多传感器信息融合、先进控制理论及应用、智能系统。作为项目负责人和主要研究人员承担了 4 项国家及省部级科研项目, 发表学术和教改论文 40 篇。



陈禹升, 硕士研究生, 主要研究方向为目标检测。