

基于选择性坐标注意力的SAR图像舰船目标检测

严春满^{1,2}, 王 钺¹

(1. 西北师范大学物理与电子工程学院, 甘肃兰州 730070; 2. 甘肃省智能信息技术与应用工程研究中心, 甘肃兰州 730070)

摘要: 针对SAR(Synthetic Aperture Radar)图像舰船目标检测结果虚警率和漏检率较高的问题, 本文提出一种基于选择性坐标注意力机制的舰船目标检测算法. 该算法以新的选择性坐标注意力机制为基础, 首先通过不同卷积核的特征提取分支对舰船目标进行特征提取; 然后融合所有分支的特征, 并沿融合后特征的不同空间方向进行编码形成两个一维特征向量, 以捕获空间方向上特征的位置信息; 最后利用这一对方向和位置敏感的特征向量编码形成“门”机制, 对各分支不同大小感受野提取的特征选择性地加权融合, 以增强舰船目标的特征表示. 本文以SSD(Single Shot MultiBox Detector)作为基础检测算法首先在SSDD(SAR Ship Detection Dataset)数据集上进行实验, 实验结果表明, 选择性坐标注意力机制相较于其他注意力机制能有效提升网络模型对舰船目标的检测能力, 同时, 基于选择性坐标注意力机制改进的SSD舰船目标检测算法平均检测精度达到了94.20%, 较原SSD算法提升了4.45%. 此外, 通过在其他两个舰船数据集上的进一步测试, 反映改进算法具有较好的泛化性, 其综合性能优于其他对比目标检测算法.

关键词: 合成孔径雷达; 舰船目标检测; 卷积神经网络; 选择性坐标注意力; 特征提取

基金项目: 国家自然科学基金(No.61961037); 甘肃省教育厅产业支撑计划项目(No.2021CYZC-30)

中图分类号: TN957.52

文献标识码: A

文章编号: 0372-2112(2023)09-2481-11

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB. 20211416

Ship Target Detection in SAR Image Based on Selective Coordinate Attention

YAN Chun-man^{1,2}, WANG Cheng¹

(1. College of Physics and Electronic Engineering, Northwest Normal University, Lanzhou, Gansu 730070, China;

2. Engineering Research Center of Gansu Province for Intelligent Information Technology and Application, Lanzhou, Gansu 730070, China)

Abstract: Aiming at the high false alarm rate and missed detection rate of ship target detection results in SAR (Synthetic Aperture Radar) images, a ship target detection algorithm based on selective coordinate attention mechanism is proposed in this paper. The algorithm is based on a new selective coordinate attention mechanism. Firstly, the feature of ship target is extracted by the feature extraction branches of different convolution kernels. Then, the features of all branches are fused, and in order to capture the position information of the features in the spatial direction, the features are encoded along different spatial directions of the fused features to form two one-dimensional feature vector codes. Finally, this direction and position sensitive feature vector coding is used to form a “gate” mechanism to get the weighted fusion of the features extracted from receptive fields with different sizes of each branch, so as to enhance the feature representation of ship targets. In this paper, the SSD (Single Shot MultiBox Detector) algorithm is used as a benchmark to test the detection results of ship targets on the SSDD (SAR Ship Detection Dataset) data set. The experimental results show that, compared with other attention mechanisms, the selective coordinate attention mechanism improves the ship detection ability of the network model more effectively. At the same time, the average detection accuracy of the SSD algorithm based on selective coordinate attention mechanism is improved to 94.20%, which is 4.45% higher than the original SSD algorithm. In addition, further tests on the other two ship data sets show that the improved algorithm has good generalization and its comprehensive performance is better than the comparison algorithms.

Key words: synthetic aperture radar; ship target detection; convolutional neural network; selective coordinate attention; feature extraction

Foundation Item(s): National Natural Science Foundation of China (No.61961037); Industrial Support Plan of Edu-

cation Department of Gansu Province (No.2021CYZC-30)

1 引言

合成孔径雷达(Synthetic Aperture Radar, SAR)是一种应用合成孔径原理的主动式微波成像雷达,具有主动发射电磁波、多波段、多极化散射特性和干涉测量成像等特点.与红外遥感等方式相比,SAR能够在能见度极差的气象条件下,全天时、全天候地获得高分辨率遥感图像,可以更广泛地应用在民用和军事各个领域^[1].舰船作为一种重要的军事装备和海洋运输载体,是一种需重点监测的高价值目标.近年来,随着SAR成像技术的发展,使用SAR图像进行海面舰船目标检测已经成为世界各国的研究热点^[2].通过SAR图像对海洋舰船目标实现实时连续检测对海运交通管理、国家海洋权益保护都具有重要意义.

传统的SAR图像舰船目标检测算法主要分为四类:基于分布模型的检测方法、基于变换的检测方法、基于模板的检测方法和基于阈值的检测方法^[3],其中研究和应用最广泛的是基于分布模型的恒虚警(Constant False Alarm Rate, CFAR)检测算法^[4].CFAR算法通过先验知识对海杂波分布建模,根据预先设定的虚警率自适应地选择阈值,从海洋背景中检测舰船目标.该算法的关键点在于分布模型函数的选择,最初通常选择高斯分布作为统计分布模型,但是对海杂波分布描述并不理想.直到Jakeman等人引入K分布概念,由于大多数海洋情况服从K分布模型,所以对海杂波具有很好的统计分布特性,但是海杂波一般并不严格服从某一种分布模型,在复杂的海面情况下检测效果并不好,并且易丢失目标形状信息.因此研究人员为适应不同海面状况,提出了众多分布模型,如伽玛分布、G0分布等.基于变换的检测方法是通过小波变换、Radon变换等方法在变换域进行目标检测.基于模板的检测方法以目标的典型模板作为基准进行检测.基于阈值的检测方法利用目标与背景之间的差异进行阈值分割以实现目标检测.这些方法在实践中获得了一定的应用,但多数存在依赖人工设计特征、步骤复杂、抗噪能力差等缺点.

随着人工智能技术的快速发展,基于卷积神经网络(Convolutional Neural Network, CNN)的图像处理技术在许多视觉任务上取得了比传统方法更好的效果,如语义分割、图像分类及目标检测等^[5].在目标检测领域,基于CNN的检测模型主要分为两类:(1)以R-CNN系列为代表的两阶段检测模型^[6,7];(2)以YOLO(You Only Look Once)和SSD(Single Shot MultiBox Detector)等为代表的单阶段检测模型^[8-12].具体来说,两阶段检测模型在第一阶段生成可能包含目标的候选区域,在

第二阶段对每个候选区域实现分类和位置回归.单阶段检测模型去掉了候选区域生成步骤,直接对目标进行分类和位置回归.比较而言,前者精度更高,而后者速度较快.为进一步提升神经网络的性能,研究人员根据大脑信号的处理机制提出了注意力机制的网络结构,通过注意力机制使模型获得需要重点关注的目标特征并抑制其他冗余特征.注意力机制主要分为四类:(1)以SE(Squeeze and Excitation)、SK(Selective Kernel)等为代表的通道注意力机制^[13,14],通过对各通道的全局特征与目标的相关度进行表达生成一维权重掩码,以构建特征与通道的映射关系,然后通过权重掩码形成门机制,从原始通道中选择最有效的通道特征;(2)以CA(Coordinate Attention)为代表的空间注意力机制^[15],通过在特征的空间上构建目标位置权重掩码,增强目标所在的位置特征;(3)以CBAM(Convolutional Block Attention Module)为代表的混合注意力机制^[16],其依次沿着输入特征的通道维度和空间维度增强目标特征信息,弥补了单一注意力机制的缺陷;(4)以Transformer为代表^[17]的自注意力机制,可以捕获特征的全局依赖关系,但其内部计算量较大.

与传统方法相比,基于CNN的检测模型及注意力机制在SAR图像舰船目标检测上也展现出许多优势,但由于SAR图像成像机理及对象与自然场景图像不同,将上述目标检测算法简单应用到SAR图像舰船目标检测任务上并不能很好发挥其最佳性能.鉴于注意力机制的强大作用,研究人员针对SAR图像的特点采用注意力机制对目标检测模型进行改进.Wang等人^[18]将通道注意力机制和软阈值算法相结合,以抑制SAR图像中噪声与复杂背景对舰船目标检测的影响.Zhao等人^[19]提出一种级联耦合CNN引导空间注意力机制的检测方法,以充分利用舰船的空间位置信息,降低了检测的虚警率.Cui等人^[20]提出将混合注意力机制引入到金字塔特征结构中,在融合特征信息的同时进一步突出各尺度特征图下目标的显著特征,有效提高了各尺度舰船目标的检测能力.虽然注意力机制可以提高舰船检测效果,但是如何准确高效地增强舰船目标特征,进一步提升检测精度仍存在着很大的挑战.这主要有以下几个原因:(1)SAR图像舰船目标有多种长宽比,网络不能根据具体的舰船目标灵活地选择适合舰船目标的感受野,导致网络不能充分提取舰船的特征,虽然SK具备选择功能的门机制,但其感受野不适合SAR舰船目标;(2)SAR图像舰船目标通常分布在复杂的背景和噪声中,在通道注意力机制中,通常采用全局平均池化对空间特征信息压缩得出一维权重掩码,

这种方式会导致特征位置信息的丢失;(3)SAR图像舰船目标具有较少的特征信息,若不能将网络的注意力聚焦到舰船位置区域会导致漏检和误检率较高的问题.

因此,根据SAR图像舰船目标的特点,本文提出一种选择性坐标注意力机制(Selective Coordinate Attention, SCA). SCA由多特征提取、坐标编码和选择融合三个阶段组成. 多特征提取阶段根据SAR图像舰船目标多长宽比的特点,采用非对称卷积核形成具有不同大小和方向的矩形感受野,并构建多特征提取分支. 坐标编码阶段使用大小为 $(H, 1)$ 和 $(1, W)$ 的平均池化层分别沿着垂直和水平坐标方向构建包含所有分支特征位置信息的编码向量. 选择融合阶段依次将属于同一通道的编码向量输入softmax处理单元得到一维权重选择编码,并对不同分支的特征进行选择性地加权融合,以强化舰船目标特征. 为评估SCA的有效性,首先在SAR舰船数据集SSDD上进行实验,将其与其他注意力机制的检测性能进行了对比. 进一步,以SCA为基础对SSD算法进行了改进,并完成了改进算法与其他目标检测算法在SSDD数据集上的对比. 实验结果表明,选择

性坐标注意力机制可有效地提高SAR图像舰船目标的检测性能,改进的SSD算法其目标检测评价指标中的平均准确率可达94.20%,检测速度为31.9 FPS,优于其他对比算法,同时其参数量仅1.87 MB,可满足高精度、轻量化的实时目标检测任务需要. 另外,为测试改进算法的泛化性能,在HRSID以及SAR-Ship-Dataset数据集上与其他目标检测算法完成了对比实验,结果证明改进算法在SAR图像舰船目标检测中具有较好的性能且表现出一定的泛化能力.

2 算法原理

2.1 选择性坐标注意力机制

SCA通过更适合SAR图像舰船目标特点的非对称卷积核进行特征提取,并且将舰船特征的行列位置信息进行编码形成“门”机制,然后对不同卷积核所提取的各分支特征根据坐标编码选择性地强化特征融合. SCA结合了SKNet中的感受野动态选择机制以及CA中的空间坐标编码构建方式. SCA的结构如图1所示,主要分为多特征提取、坐标编码和选择融合三个步骤.

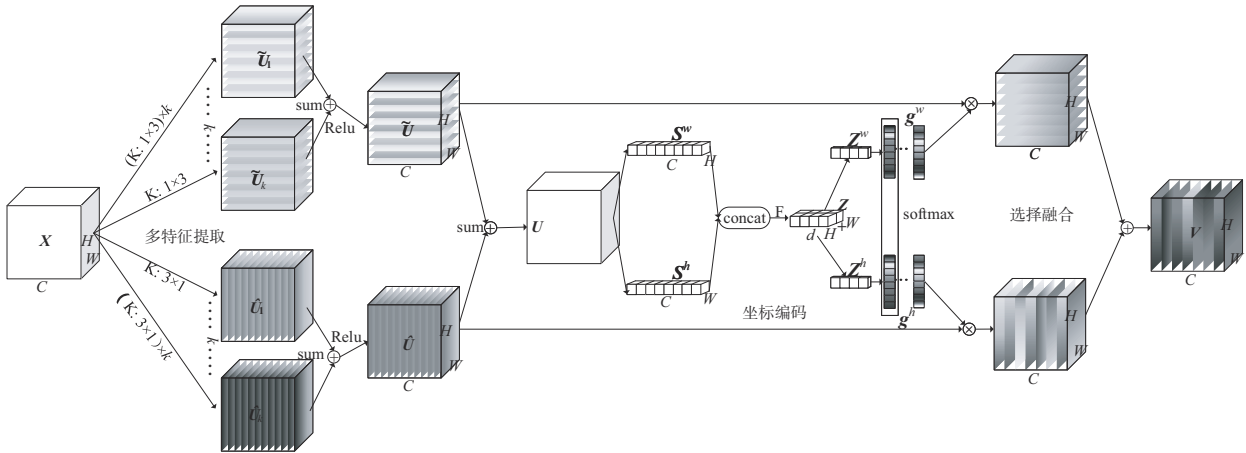


图1 SCA结构图

多特征提取:为减少参数量,采用非对称卷积核堆叠代替原卷积网络中的大卷积核. 对于任意给定的高为 H ,宽为 W ,通道数为 C 的输入 $X \in \mathbf{R}^{H \times W \times C}$,按 H 、 W 两个不同方向,堆叠大小为3的非对称卷积核形成特征提取分支,分别获取这两个方向的特征信息. 对于 W 方向,通过 1×3 横向卷积核的特征提取分支得到该方向特征信息,即 $\tilde{F}: X \rightarrow \tilde{U}_1, \dots, \tilde{U}_k \in \mathbf{R}^{H \times W \times C}$;对于 H 方向通过 3×1 纵向卷积核的提取分支得到该方向特征信息,即 $\hat{F}: X \rightarrow \hat{U}_1, \dots, \hat{U}_k \in \mathbf{R}^{H \times W \times C}$. 在此阶段引入超参数 k 以决定每个方向特征提取分支的数量以及各分支感受野的大小. 卷积核堆叠的数目与当前分支序号相同,如 W 方向第3个特征提取分支为3个 1×3 大小的非对称卷积

核堆叠,形成与 1×9 卷积核等效的感受野. 对提取到 W, H 方向的特征信息分别通过sum操作进行融合并通过Relu激活函数后得到两个特征输出 \tilde{U}, \hat{U} ,具体如式(1):

$$\begin{cases} \tilde{U} = \text{Relu}(\text{sum}(\tilde{U}_1, \dots, \tilde{U}_k)) \\ \hat{U} = \text{Relu}(\text{sum}(\hat{U}_1, \dots, \hat{U}_k)) \end{cases} \quad (1)$$

坐标编码:首先将上一阶段的两个特征输出采用sum操作进行总体融合得到输出 U ,如式(2):

$$U = \text{sum}(\tilde{U}, \hat{U}) \quad (2)$$

在SK、SE等注意力机制中一般采用全局平均池化操作,这种方式不利于提取目标的位置信息. 对舰船目

标识而言,空间位置信息的获取非常关键.为了提取到输出特征的空间坐标信息,本文采用CA中的编码方式,将全局编码分解为具有舰船目标位置信息的一对特征向量编码.具体来说,给定输入 U ,分别沿着水平坐标 w 和垂直坐标 h 对其每个通道进行编码,高度为 h 时的第 c 个通道的编码为:

$$S_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} U_c(h, i) \quad (3)$$

同样,宽度为 w 时的第 c 个通道的编码:

$$S_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} U_c(j, w) \quad (4)$$

上述两种操作分别沿两个空间方向聚合特征,得到一对具有位置感知能力的特征编码矩阵 S^h, S^w ,这种特征提取方式同自注意力机制具有类似的效果,可提取特征远程依赖关系,有助于网络更准确地定位目标区域.

在SE等通道注意力机制的研究中已经证明获取通道间的相互关系是非常重要的.所以,首先将两个方向的特征编码在空间维度进行拼接,然后通过共享参数的 1×1 卷积层 F 进行通道压缩,实现通道间特征信息的交流,最后得到水平编码与垂直编码的中间特征映射层 $Z \in \mathbf{R}^{d \times (H+W)}$.特征压缩如式(5):

$$Z = F(S) = \delta(\beta(W[S^h, S^w])) \quad (5)$$

其中, $[\cdot, \cdot]$ 表示沿通道维度的拼接操作; $W \in \mathbf{R}^{d \times (H+W)}$,其作用与SK中 1×1 中间映射层一致,以压缩特征, d 为压缩后的维度; β 为批归一化操作; δ 为非线性激活函数.为了研究 d 对模型性能的影响,采用与SK中类似的缩减比参数 r 以控制其值的选取,具体为:

$$d = \max(C/r, L) \quad (6)$$

其中, L 为 d 的最小值, $L=8$ 为实验中设置的默认值;参数 r 的值将在实验部分给出.最后将 Z 沿空间维度分解为两个紧凑的特征编码 $Z^h \in \mathbf{R}^{d \times H}$, $Z^w \in \mathbf{R}^{d \times W}$.

选择融合:选择性坐标注意力机制通过包含位置信息的特征编码生成软注意力权值,自适应选择不同空间的信息以强化特征.对属于同一通道的 H 和 W 方向的压缩特征编码 Z^h, Z^w ,采用softmax算子得到与输入 X 同维度的软注意力权值,具体表达如下:

$$g_c^h = \frac{e^{A_c Z^h}}{e^{A_c Z^h} + e^{B_c Z^w}} \quad (7)$$

$$g_c^w = \frac{e^{B_c Z^w}}{e^{A_c Z^h} + e^{B_c Z^w}} \quad (8)$$

其中, g_c^h 为 g^h 中第 c 个通道 H 方向的软注意力向量, $g^h \in \mathbf{R}^{C \times H \times 1}$ 为特征 \hat{U} 对应的软注意力矩阵; $A_c \in \mathbf{R}^{1 \times d}$ 为 A 的第 c 个通道, $A \in \mathbf{R}^{C \times d}$ 为中间映射层,对应 1×1 卷积层;同理, g_c^w 为 g^w 中第 c 个通道 W 方向的软注意力向

量, $g^w \in \mathbf{R}^{C \times 1 \times W}$ 为特征 \hat{U} 的软注意力矩阵; $B_c \in \mathbf{R}^{1 \times d}$ 为 B 的第 c 个通道, B 为中间映射层, $B \in \mathbf{R}^{C \times d}$; $g_c^w + g_c^h = 1$.

最后,通过两个空间方向的软注意力向量作为“门”机制对各分支的特征进行加权融合,输出最终的特征 V ,则 V 的第 c 个通道的第 (i, j) 特征值表示为:

$$V_c(i, j) = \tilde{U}_c(i, j) \times g_c^w(i) + \hat{U}_c(i, j) \times g_c^h(j) \quad (9)$$

其中, $0 < i < W, 0 < j < H$. SCA通过选择性特征融合,可以使网络更准确的定位目标的区域位置,加强对舰船目标的特征提取,提升目标检测模型对舰船的检测能力.

2.2 网络体系结构

2.2.1 基本单元

SCA模块的输出维度与输入维度一致,故可以在不改变原网络情况下灵活嵌入到目前的网络体系结构中.本文将SCA模块嵌入到ResNet^[21]的基本单元Resblock中,形成网络基本单元SCA-Resblock.图2为经典的SE block、CA block结构与本文SCA-Resblock结构的对比.图中Avg Pool(Average Pooling)、GAP(Global Average Pooling)及FC(Fully Connected Layers)分别表示平均池化、全局平均池化和全连接层.SCA的嵌入方式与SE等注意力模块嵌入方式相同,通过在卷积层后插入SCA单元构建基于注意力机制的模型结构,提升对目标特征的提取能力.

2.2.2 模型结构

本文基于典型的单阶段目标检测算法SSD,通过嵌入SCA,构建针对复杂场景的SAR图像舰船目标检测算法.改进后的SSD算法框架如图3所示,命名为SSR-SSD(SCA Simple ResNet18-SSD).

由于SAR图像与自然场景图像相比具有较少的特征信息且为单通道图像,为了减少冗余特征的通道,SSR-SSD将conv1卷积层的通道数减少至1/4,之后的卷积层通道数沿用ResNet18的通道增加方式,这使得ResNet显著轻量化,最终形成了Simple-ResNet18(SResNet18).SCA-SResNet18在Simple-ResNet18骨干网络中以SCA-Resblock作为基本单元,其缩减比参数 r 的取值对模型结构有较大影响.表1列出了ResNet18、SResNet18以及SCA-SResNet18的骨干网络体系结构,模型中 r 默认设置为16.依据ResNet18结构,SCA-SResNet18使用了 $\{2, 2, 2, 2\}$ 个SCA-Resblock.若需得到不同的网络架构,仅需改变每个阶段的SCA-Resblock的个数.将conv4与conv5的第一个block卷积层的卷积核步长设置为1,使特征图保持原来大小.在骨干网络后添加额外的深度可分离卷积层进行下采样,使网络输出的特征图空间维度与原SSD算法中一致,形成最终特征提取网络.

SSR-SSD算法采用 320×320 大小的输入图像,经SCA-SResNet18骨干网络提取形成 40×40 大小的特征

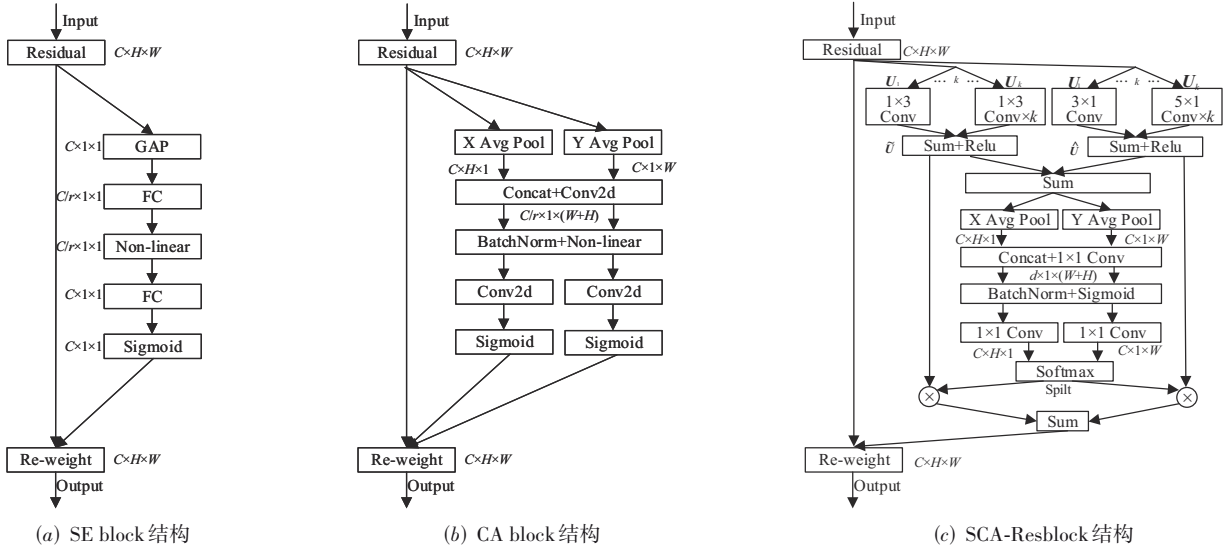


图2 三种注意力基本单元结构对比

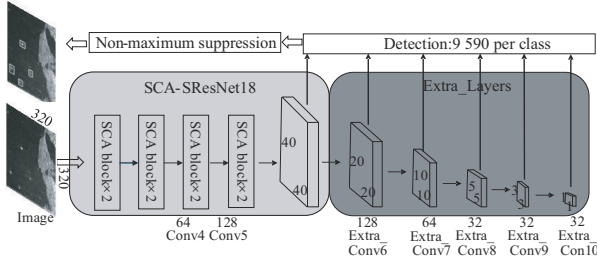


图3 SSR-SSD算法框架

图,然后经过额外的卷积层进一步提取特征,形成与原SSD模型输出大小一致的特征图,分别为 $20 \times 20, 10 \times 10, 5 \times 5, 3 \times 3, 1 \times 1$. 由于SAR为单通道图像,特征图的通道数分别为128, 64, 32, 32, 32. 然后在6个特征层生成与原SSD算法等数量相同宽高比的先验框. 在训练阶段,对先验框与真实框计算IoU以进行正负样本的匹配,并采用难例样本挖掘技术保证正负样本比例接近1:3. 最后通过分类损失和回归损失的加权构建损失函数来进行训练,具体如式(10):

$$L(x, c, l, g) = \frac{1}{N} [L_{\text{conf}}(x, c) + aL_{\text{loc}}(x, l, g)] \quad (10)$$

其中, N 为候选框与真实框匹配的数量, $L_{\text{conf}}(x, c)$ 为分类损失, $L_{\text{loc}}(x, l, g)$ 为位置损失, a 为平衡权重(默认值取1), x 为输入图像, c 为标签类别, l 为预测框, g 为真实框. 在预测阶段,由于预测框数量较多,故通过非极大值抑制算法(Non Maximum Suppression, NMS)对预测框进行筛选以消除重复框和无类别框,最后输出检测结果.

3 实验

3.1 实验环境

实验中采用公开的SAR图像舰船目标数据集

SSDD、HRSID和SAR-Ship-Dataset^[22,23]. SSDD数据来源包括卫星RadarSat-2、TerraSAR-X和Sentinel-1,采用HH、HV、VV和VH四种极化方式,包含1~15 m不同的分辨率,数据集标注了在不同场景下像素值大于3的舰船目标,共1 160幅图像,包含2 346艘舰船目标. 高分辨率HRSID数据来源包括Sentinel-1B、TerraSAR-X等卫星,成像分辨率0.5~3 m,总共包含5 604张 800×800 像素的高分辨率SAR图像,共有16 951个舰船目标. SAR-Ship-Dataset数据集以高分三号和Sentinel-1卫星SAR数据为主数据源,包含了多源、多模式的SAR图像,总共包含43 819张 256×256 像素的SAR图像.

实验使用的计算机硬件平台CPU为Intel(R) Core(TM) i5-6300HQ, GPU为NVIDIA GTX1060. 操作系统为Ubuntu16.04,采用Pytorch1.7.0深度学习框架,编程语言为Python. 实验采用随机初始化网络结构参数进行模型训练,批处理大小为8,初始学习率为0.01,衰减系数为0.1,总训练轮数为300,优化算法采用SGD,动量因子参数为0.9. 经实验验证,将2.1节中的参数 r 值设置为16.

3.2 评估指标

采用检测速率FPS(Frames Per Second)、平均准确率AP(Average Precision)、以及参数量作为模型性能的评价指标. FPS定义为每秒处理图像的帧数,衡量模型检测速度. 参数量为网络模型中的权重和偏置参数存储总量. AP表示查准率 P (Precision)与查全率 R (Recall)曲线围成的面积, P 与 R 如式(11)和式(12)所示:

$$P = TP / (TP + FP) \quad (11)$$

$$R = TP / (TP + FN) \quad (12)$$

其中,TP为正确检测为正例的个数,FP为错误检测为正例的个数, FN为错误检测为负例的个数, AP的计算如

表1 三种骨干网络的体系结构

Layer name	Output size	ResNet18	SResNet18	SCA-SResNet18
conv1	160×160	7×7, 64, stride 2	7×7, 16, stride 2	7×7, 16, stride 2
conv2	80×80	3×3 max pool, stride 2	3×3 max pool, stride 2	3×3 max pool, stride 2
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$[(3 \times 3, 16) \times 2] \times 2$	$\begin{bmatrix} (3 \times 3, 16) \times 2 \\ \text{SCA}[r=16] \end{bmatrix} \times 2$
conv3	40×40	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$[(3 \times 3, 32) \times 2] \times 2$	$\begin{bmatrix} (3 \times 3, 32) \times 2 \\ \text{SCA}[r=16] \end{bmatrix} \times 2$
conv4	40×40	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$[(3 \times 3, 64) \times 2] \times 2$	$\begin{bmatrix} (3 \times 3, 64) \times 2 \\ \text{SCA}[r=16] \end{bmatrix} \times 2$
conv5	40×40	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$[(3 \times 3, 128) \times 2] \times 2$	$\begin{bmatrix} (3 \times 3, 128) \times 2 \\ \text{SCA}[r=16] \end{bmatrix} \times 2$
extra_conv6	20×20	$[(1 \times 1, 128), (3 \times 3, 128)]$		
extra_conv7	10×10	$[(1 \times 1, 64), (3 \times 3, 64)]$		
extra_conv8	5×5	$[(1 \times 1, 32), (3 \times 3, 32)]$		
extra_conv9	3×3	$[(1 \times 1, 32), (3 \times 3, 32)]$		
extra_conv10	1×1	$[(1 \times 1, 32), (3 \times 3, 32)]$		

式(13)所示,在本文实验中求解AP时IoU设置为0.5.

$$AP = \int_0^1 P(r) dr \quad (13)$$

3.3 参数 k 的选择

在多特征提取阶段,将 k 作为超参数引入网络中可以改变特征提取分支数量以及各分支的感受野大小.为了研究 k 的取值对SAR舰船目标检测性能和计算成本的影响,以SSD为基准算法,采用不同 k 值的SCA-SResNet18在SSDD数据集上进行测试,对比结果如表2所示.

表2 k 值对网络性能的影响

超参数 k	AP/%	Params/MB
1	92.99	1.35
2	94.20	1.87
3	93.45	2.66
4	93.26	3.71

$k=1$ 时,每个特征提取方向仅包含卷积核大小为3的一条特征提取分支,由于较小的感受野不利于全局信息的获取,所以其AP仅为92.99%.当 $k=2$ 时,增加了一条由两个大小为3的非对称卷积核堆叠形成的特征提取分支,形成了更适合舰船目标的多尺度感受野,其矩形感受野大小为5,AP值提高到了94.20%.当 $k=3,4$ 时,特征提取分支数量进一步增加,矩形感受野也进一步增大,但大的感受野会降低网络对小目标特征的提取能力且会包含较多的噪声.实验结果反映 k 值继续增大并没有进一步提升SCA的性能,AP值随着 k 值的增大反而降低,但是依然高于 $k=1$ 时的AP,说明多尺度感受野对舰船目标检测十分重要.另外,随着 k 的增大

模型参数量相应提高,故在本实验中,对网络内所有的SCA模块无额外说明时, k 值默认为2.

3.4 SSD改进算法对比实验

为了验证本文提出的SCA对SAR图像舰船目标检测的有效性,首先在SSDD数据集上,将采用ResNet18、Simple-ResNet18以及SCA-SResNet18改进后的SSD算法与原SSD算法进行对比,并按照7:1:2的比例,将数据集划分成训练集、验证集和测试集,各算法均采用320×320大小的输入图像进行训练与测试,对比结果如表3所示.

表3 骨干网络及SCA对模型性能的影响

Backbone	AP/%	FPS	Params/MB
VGG16	89.75	34.2	23.8
ResNet18	90.57	39.4	15.15
Simple-ResNet18	90.43	70.9	1.07
SCA-SResNet18($r=16$)	94.20	31.9	1.87

原SSD算法采用VGG16作为骨干网络,时间复杂度与空间复杂度较高,AP为89.75%,FPS为34.2,可以满足实时检测的需要.采用ResNet18作为骨干网络改进后,由于骨干网络深度的增加以及残差结构的优势,相比原SSD算法检测效果有所提升,参数量也显著下降.采用Simple-ResNet18作为骨干网络进行改进后,FPS提升了36.7,AP提升到了90.43%,且参数量降低了95.5%,虽然与ResNet18相比AP稍有下降,但参数量的大幅下降使模型显著轻量化.进一步,采用SCA-Simple-ResNet18改进后,由于SCA的多分支和动态选择机制会使其复杂度增加,FPS下降为31.9,但平均准

准确率继续提升. 与原SSD算法相比,改进算法检测能力的提升较为显著,AP达到了94.20%,为对比算法中的最高值,FPS与原SSD算法接近,而参数量仅为原算法的7.86%,获得了模型性能及轻量化两个方向的显著提升,可满足实时监测的需要.

3.5 注意力机制对比

为了进一步验证本文SCA对SAR图像舰船目标检测的有效性,依旧在SSDD数据集上,将其与目前典型的注意力机制模型在基于Simple-ResNet18改进的SSD算法上进行对比,测试结果如表4所示.

表4 典型注意力机制的性能对比

模型	AP/%	FPS	Params/MB
SSD	90.43	70.9	1.07
SSD +SK	91.90	36.5	1.20
SSD +CBAM	91.93	53.2	1.10
SSD +SE	92.11	62.7	1.08
SSD +CA	92.34	38.4	1.09
SSD +SCA	94.20	31.9	1.87

SK及SE为典型的通道注意力机制,采用多分支的特征提取以及动态选择机制,对模型目标检测性能的提升有一定的帮助,AP提升,但参数量上升,FPS下降.相比较而言,SE好于SK,AP提升了1.68%,FPS下降且参数量的提高不大.对于CA,采用坐标注意力机制有效提升了网络对目标的位置敏感性,尽管其FPS下降,但AP继续提升,说明特征位置信息对于SAR图像舰船目标检测具有更重要的作用.CBAM为混合注意力机制,其按照通道维度和空间维度增强目标特征信息,可有效提高AP,但由于模型复杂度的提高,FPS有所下降.SK、SE及CBAM采用全局平均池化或较大卷积核对空间信息进行编码,没有特别关注目标特征的位置信息.CA虽然采用了位置信息编码方式,但是其直接将权重编码乘到输入特征中,未构建两个方向位置编码的关系.对于本文提出的SCA,在多特征提取阶段采用与SAR图像中舰船目标相匹配的矩形感受野进一步对空间特征进行提取,通过 H, W 两个方向的一维平均池化操作生成两个方向位置信息特征编码,能有效保留目标特征的空间信息.SCA通过将同一特征的坐标编码输入softmax函数构建两个方向特征编码的关系,然后利用动态选择“门”机制选择性融合不同分支的特征信息,而不是直接对输入特征进行加权,同时相比CA增加的多特征提取阶段能进一步提取特征.最终,SCA相比其他注意力机制能有效的提升网络对SAR图像中舰船目标的检测能力,其AP达到了94.20%.由于SCA为多分支的特征提取以及编码结构,所以其参数量上升,FPS为31.9,但仍能满足实时检测的需要.

另外,通过Grad-CAM以及Guided Grad-CAM技

术^[24]对特征图进一步可视化,以分析不同的注意力机制对SAR图像中大、中、小舰船目标检测的影响,对比结果如表5所示.

在对大型舰船目标的检测中,从表5大型舰船目标特征可视化图像中可以看出各注意力机制的表现并不相同.由于通道注意力机制SE感受野较小,不利于获取目标的局部信息,所以对不包含舰船目标的海域有较大的关注度.空间注意力机制SK和混合注意力机制CBAM具有较大感受野,能获取目标的局部信息,所以对其他海域关注度较低.CA中虽然不具有大感受野的卷积核,但是通过两个空间方向的特征编码可以捕获目标的长依赖关系,所以对目标区域具有更高的关注度.SCA采用矩形感受野,其更适合舰船目标并可捕获长依赖关系的坐标编码,在Guided Grad-CAM图中呈现出更多的细粒度特征,对舰船目标有更好的关注度.在对中型舰船目标特征可视化中,SE、SK、CBAM及CA注意力机制对舰船目标的关注度均有所加强.SE与CA相对SK、CBAM的感受野较小,导致对陆地的关注度偏多,但对所有目标都具有更加均衡的关注度,而CA由于坐标注意力机制的细粒度关注能力,所以相比SE其感受野能更加贴合目标.SCA的选择性坐标注意力机制能有效提高目标位置的特征权重,抑制非目标区域的特征权重,相比其他注意力机制取得了更加均衡的注意力能力.在对小型舰船目标的特征可视化对比结果中,由于SE对有效特征分配了更大的权值,使得目标特征更加清晰.CA由于其更加细粒度的信息编码,使得对小目标的关注度加强,但是对SAR图像中相干斑噪声的敏感性也增强,对噪声区域出现较大的关注度.SCA不直接将细粒度的权重加权到原始特征上,所以对噪声敏感性较低,并且对其他注意力机制关注度较低的右下角小目标也得到较高的关注度.综上分析可见,选择性坐标注意力相比其他注意机制能更精确地关注SAR图像中的舰船目标所在的区域.

3.6 典型目标检测模型性能对比

为了更好的评估本文提出的基于SCA改进的SSD算法对SAR图像舰船目标的检测性能,将其与其他几种典型的多尺度目标检测算法在SSDD数据集上进行对比测试.为保证结果的公平性,均在本实验平台重新进行训练,除本文改进后的SSD算法(SSR-SSD)外,其他模型均采用加载预训练模型的迁移学习方式进行训练,并且均采用原始算法的数据增强方式,实验结果如表6所示.

由表6可知本文SSR-SSD在SAR图像舰船目标检测任务中的AP、FPS性能指标均优于其他检测算法,而且模型参数量较少.从检测精度上看,本文算法与相同输入图像大小的DSSD和YOLOv4网络相比AP分别提

表5 多种注意力机制对舰船检测影响的特征可视化对比结果

模型	大型舰船目标特征可视化	中型舰船目标特征可视化	小型舰船目标特征可视化
SSD			
SSD +SE			
SSD +SK			
SSD +CBAM			
SSD +CA			
SSD +SCA			

高了 5.37%, 2.7%, 与较大输入图像尺寸的 Faster RCNN、RetinaNet 和 RFBNet 相比 AP 分别提高了 5.57%, 9.02%, 6.67%。可见相对加深骨干网络以及增大输入图像尺寸并不能单调的提升目标的检测能力, 相比之下, 采用细粒度的特征编码指导网络的注意力位置能更加有效的提升 SAR 图像中舰船目标的检测能力。从检测速度上看, 本文算法的检测速度达到了 31.9FPS, 与其他单阶段目标检测算法相比依然具有较快的检测速度。在模型大小上, 本文算法仅为 1.87 MB 参数量, 在模型部署时占用较少的内存空间。特别对于本文提出的 SCA 模块来说, 作为一种即插即用的模块, 其相比增加的参数量而言带来的更多的是检测精度的提升。综上所述, 本文通过 SCA 改进的 SSD 算法相对于其他检测模型在 SAR 图像舰船目标检测中具有更好的综合性能。


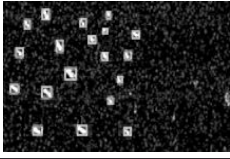
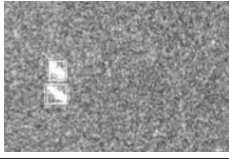
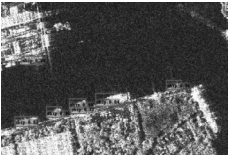
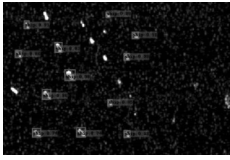
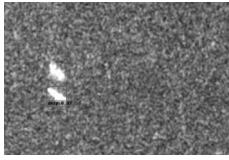
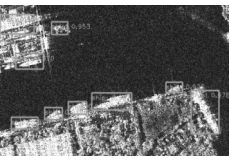
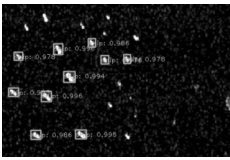
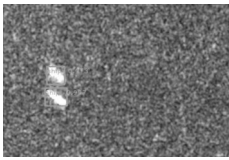
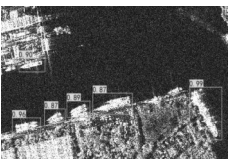
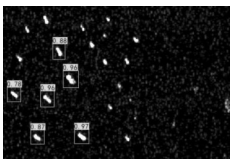
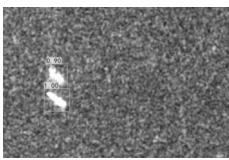

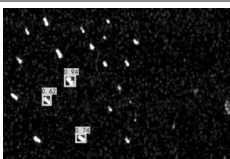
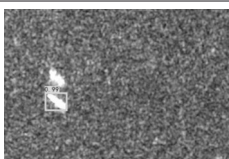
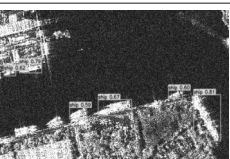
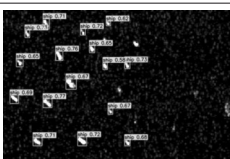
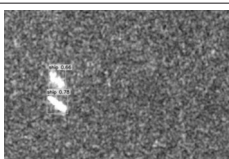
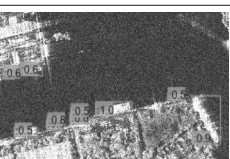
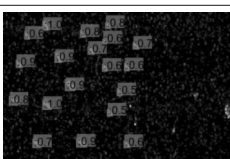
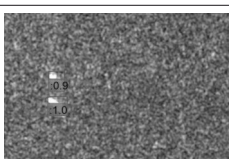
表 7 所示为 SSDD 数据集中三种典型的场景下, 不同目标检测模型对 SAR 图像舰船检测样例。在近岸多目标舰船的检测结果, 由于图中舰船目标密集且有多个尺度, 在 DSSD, Faster RCNN, RetinaNet, RFBNet, YOLOv4 中均出现了不同情况的漏检。虽然 Faster RCNN 对所有多尺度目标均正确检出, 但是对陆地区域出现较多虚警目标。本文算法具有较好的多尺度目标检测效果, 由于可捕获长距离感受的空间编码, 能

表6 本文方法与其他经典目标检测模型对比结果

模型	Backbone	Input size	AP/%	FPS	Params/MB
RetinaNet ^[25]	ResNet50	600×600	85.18	14.4	36.33
RFBNet ^[26]	VGG16	512×512	87.53	15.4	34.44
Faster RCNN ^[7]	VGG16	600×600	88.63	6.3	136.69
DSSD ^[27]	ResNet101	320×320	88.83	15.7	121.8
YOLOv4 ^[11]	CSPDarkNet53	320×320	91.50	30.9	63.93
SSR-SSD	SCA-ResNet18	320×320	94.20	31.9	1.87

很好的抑制虚警目标的产生。在远海密集小目标检测结果中, 除 YOLOv4 外, 其他目标检测模型均出现了较多的漏检情况。YOLOv4 对小目标采用 Mosaic 数据增强手段, 有效提高了小目标的检出能力。本文算法采用了选择性坐标注意力机制, 采用更适合舰船目标的感受野并且通过空间编码增强小目标位置的特征权重, 对远海密集小目标得到了更好的检测效果。在强干扰下的目标检测结果中, DSSD 和 RFBNet 出现漏检情况, 其他模型均正确检出舰船目标, 本文算法通过 SCA 机制使得其具有较好的抗干扰特性, 在复杂的背景下相比其他模型依然以较高的置信度检出所有舰船目标。

表7 多种场景下不同目标检测方法检测结果对比

	近岸多目标检测结果	远海多目标检测结果	强干扰下目标检测结果
真值			
SSD			
Faster RCNN			
RetinaNet			
RFBNet			
YOLOv4			
Our			

为了进一步验证本文提出的基于SCA改进的SSD算法的泛化能力,将其与其他目标检测模型在HRSID以及SAR-Ship-Dataset数据集上进行模型迁移测试对比实验.随机抽取数据集的80%为测试集,即分别抽取HRSID,SAR-Ship-Dataset内4 484及35 056张图片作为测试集.此外,为保证测试结果的可比性,模型的输入图像大小与之前保持一致.

实验结果如表8所示.分析实验结果可见,由于HRSID,SAR-Ship-Dataset与SSDD数据集的图像分辨率、成像模式及目标分布等方面都存在差异,导致各个模型的AP均有下降.另外,由于HRSID数据集小目标占比较多且目标分布背景复杂,模型的目标检测结果

整体低于在SAR-Ship-Dataset数据集上检测结果.相对而言,在HRSID数据集上,本文提出的SSR-SSD算法均达到了较好的检测结果,并且随着 k 值增大,即多特征提取分支数量的增加,进一步增加了对目标多尺度信息的提取能力,模型的AP也逐渐上升.在SAR-Ship-Dataset数据集上,当 $k=1$ 及 $k=2$ 时,SSR-SSD算法的AP高于其他对比模型.当 $k=3$ 及 $k=4$ 时,SSR-SSD算法的AP低于RFBNet.这是由于RFBNet模型相对SSR-SSD的输入图像分辨率较大,这有利于舰船小目标检测;另一方面,SAR-Ship-Dataset数据集内含有较多强噪声下的舰船目标,在 $k=2$ 时,SCA仅有两条多特征提取分支,感受野与RFBNet相似,其AP值也近似,但随着感受野的进一步增大,即 k 值的增大,使得噪声信息干扰舰船

目标特征,导致检测性能有所下降.综上所述,本文通过SCA改进的SSD算法相对于其他检测模型在SAR图像舰船目标检测中具有较好的性能且表现出一定的泛化能力.

表8 模型迁移测试对比结果(AP) 单位:%

模型	HRSID	SAR-Ship-Dataset
YOLOv4	44.00	56.40
Faster RCNN	52.60	55.40
RetinaNet	56.55	63.80
RFBNet	58.97	64.00
DSSD	59.26	63.01
SSR-SSD($k=1$)	59.32	64.40
SSR-SSD($k=2$)	59.51	64.14
SSR-SSD($k=3$)	59.75	63.65
SSR-SSD($k=4$)	59.86	63.60

4 总结

本文提出一种适用于SAR图像舰船目标检测的选择性坐标注意力机制,不同于常规的注意力机制方法,该方法采用了适用于SAR图像舰船目标的感受野进行特征提取,并采用可保留目标空间位置信息的编码方式构建“门机制”,最后通过空间和位置信息敏感的软注意力向量动态选择性地融合不同分支提取的舰船特征信息.通过在SSDD数据集上的实验验证,SCA相比其他注意力机制可提高网络对舰船目标所在区域的关注度,并且通过SCA改进的SSD算法有效的降低了虚警率和漏检率.在HRSID及SAR-Ship-Dataset数据集上的进一步测试也验证了改进算法对舰船目标的检测能力,表现出较好的泛化性能.改进模型大小仅为1.87 MB,对未来SAR图像舰船目标检测网络轻量化设计具有较好的实用价值.

参考文献

- [1] 彭书娟,曲长文,李健伟. K近邻优化估计的SAR图像建模与目标检测算法[J]. 控制与决策, 2020, 35(9): 2199-2206.
- [2] PENG S J, QU C W, LI J W. K nearest neighbors optimized estimation algorithm for SAR image statistical modeling and target detection[J]. Control and Decision, 2020, 35(9): 2199-2206. (in Chinese)
- [3] WACKERMAN C C, FRIEDMAN K S, PICHEL W G, et al. Automatic detection of ships in RADARSAT-1 SAR imagery[J]. Canadian Journal of Remote Sensing, 2001, 27(5): 568-577.
- [4] 唐沐恩,林挺强,文贡坚. 遥感图像中舰船检测方法综述[J]. 计算机应用研究, 2011, 28(1): 29-36.
- [5] TANG M E, LIN T Q, WEN G J. Overview of ship detection methods in remote sensing image[J]. Application Research of Computers, 2011, 28(1): 29-36. (in Chinese)
- [6] QIN X X, ZHOU S L, ZOU H X, et al. A CFAR detection algorithm for generalized gamma distributed background in high-resolution SAR images[J]. IEEE Geoscience and Remote Sensing Letters, 2013, 10(4): 806-810.
- [7] 程旭,宋晨,史金钢,等. 基于深度学习的通用目标检测研究综述[J]. 电子学报, 2021, 49(7): 1428-1438.
- [8] CHENG X, SONG C, SHI J G, et al. A survey of generic object detection methods based on deep learning[J]. Acta Electronica Sinica, 2021, 49(7): 1428-1438. (in Chinese)
- [9] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2016: 1440-1448.
- [10] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 779-788.
- [12] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 6517-6525.
- [13] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-08)[2021-09-18]. <https://arxiv.org/abs/1804.02767>.
- [14] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [EB/OL]. (2020-04-23)[2021-09-18]. <https://arxiv.org/abs/2004.10934>.
- [15] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector[C]//Computer Vision—ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
- [17] LI X, WANG W H, HU X L, et al. Selective kernel networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 510-519.

- [15] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 13708-13717.
- [16] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//Computer Vision—ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. (2017-06-12)[2021-09-18]. <https://arxiv.org/abs/1706.03762>.
- [18] WANG R, SHAO S H, AN M Y, et al. Soft thresholding attention network for adaptive feature denoising in SAR ship detection[J]. IEEE Access, 2021, 9: 29090-29105.
- [19] ZHAO J P, ZHANG Z H, YU W X, et al. A cascade coupled convolutional neural network guided visual attention method for ship detection from SAR images[J]. IEEE Access, 2018, 6: 50693-50708.
- [20] CUI Z Y, LI Q, CAO Z J, et al. Dense attention pyramid networks for multi-scale ship detection in SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(11): 8983-8997.
- [21] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 770-778.
- [22] WEI S J, ZENG X F, QU Q Z, et al. HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation[J]. IEEE Access, 2020, 8: 120234-120254.
- [23] WANG Y Y, WANG C, ZHANG H, et al. Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery[J]. Remote Sensing, 2019, 11(5): 531.
- [24] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 618-626.
- [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] LIU S T, HUANG D, WANG Y H. Receptive field block net for accurate and fast object detection[C]//Computer Vision—ECCV 2018. Cham: Springer International Publishing, 2018: 404-419.
- [27] FU C Y, LIU W, RANGA A, et al. DSSD: Deconvolutional single shot detector[EB/OL]. (2017-01-23)[2021-

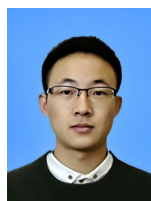
09-18].<https://arxiv.org/abs/1701.06659>.

作者简介



严春满 男,1970年出生,甘肃人.西北师范大学副教授,硕士生导师.主要研究领域为模式识别、神经网络、深度学习及其在计算机视觉中的应用.

E-mail: yanacha02@163.com



王 铖 男,1997年出生,山东人.西北师范大学物理与电子工程学院硕士研究生.主要研究方向为深度学习、嵌入式系统与应用等.

E-mail: 610487244@qq.com