



(12) 发明专利申请

(10) 申请公布号 CN 113408549 A

(43) 申请公布日 2021.09.17

(21) 申请号 202110796466.7

(22) 申请日 2021.07.14

(71) 申请人 西安电子科技大学

地址 710071 陕西省西安市太白南路2号

(72) 发明人 李甫 张伟 谢玉圣 钱若浩

牛毅 李阳 谢雪梅 石光明

(74) 专利代理机构 陕西电子工业专利中心

61205

代理人 田文英 王品华

(51) Int.Cl.

G06K 9/46 (2006.01)

G06K 9/62 (2006.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

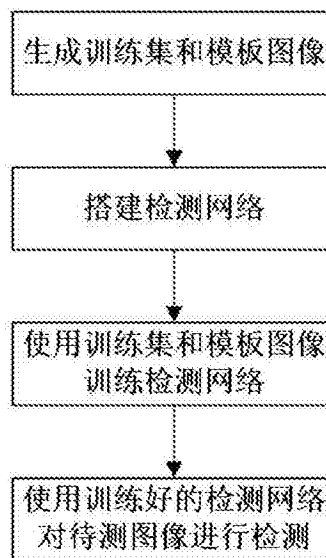
权利要求书2页 说明书6页 附图2页

(54) 发明名称

基于模板匹配和注意力机制的少样本弱小目标检测方法

(57) 摘要

一种基于模板匹配和注意力机制的少样本弱小目标检测方法,解决现有目标检测算法对于训练集样本数量需求大,在复杂场景下对小目标检测效果差的问题。本发明的步骤为:挑选出少量训练图片和一张模板图片;使用匹配卷积核生成网络生成匹配卷积核,使用匹配卷积核进行模板匹配;使用特征金字塔网络融合特征图;利用通道注意力模块和降噪模块串联处理特征图,增加特征图的有效通道响应;使用分类输出网络和坐标输出网络,得到检测网络预测结果,使用少量样本迭代更新网络权重。最后使用旋转矩形框对预测结果进行标注。本发明具有训练样本需求小,小目标检测效果好的优点。



1. 一种基于模板匹配和注意力机制的少样本弱小目标检测方法,其特征在于,利用匹配卷积核生成网络得到匹配卷积核,利用通道注意力模块和降噪模块来提升网络表示能力和降噪;该目标检测方法的步骤包括如下:

(1) 生成训练集和模板图像:

(1a) 选取至少10张图像,以256像素为裁剪步长将每张图像裁剪为 1024×1024 像素的大小且每张图像包含至少1个目标,将裁剪过程中被截断的目标标记为特殊目标,将此目标的标志位difficult置为1,将裁剪后的图像和标志位组成数据集;

(1b) 对数据集中每个目标的4个角点以旋转矩形框的形式进行标注,得到每个目标的标注坐标;

(1c) 将每个目标的标注坐标和数据集组成训练集;

(1d) 随机选取1张目标轮廓清晰的图像作为模板图像;

(2) 搭建匹配卷积核生成网络:

搭建一个由主干网络ResNet101和池化层组成的匹配卷积核生成网络,将池化层的核大小设置为 3×3 ,池化步长为2,池化层采用平均池化法;

(3) 搭建一个具有4个阶段、且每个阶段的输出通道数均为256的特征金字塔网络;

(4) 搭建通道注意力模块和降噪模块:

(4a) 搭建通道注意力模块,其结构依次为:全局平均池化层,第一全连接层,ReLU激活函数,第二全连接层,Sigmoid激活函数,其中,全局平均池化层的核大小设置为 3×3 ,池化步长为2,采用平均池化法,第一与第二全连接层的输入通道数分别设置为256和128,输出通道数分别设置为128和256;

(4b) 搭建降噪模块,其结构依次为:第一卷积层,ReLU激活函数,第二卷积层,ReLU激活函数,第三卷积层,Softmax函数,其中,第一、第二和第三卷积层的卷积核大小均设置为 3×3 ,步长均设置为1;

(5) 搭建分类输出网络和坐标输出网络:

(5a) 搭建一个由第一全连接层和第二全连接层串联的分类输出网络,其中,第一全连接层输入大小设置为49,输出大小设置为1024,第二全连接层输入大小设置为1024,输出大小设置为2;

(5b) 搭建一个由第一全连接层和第二全连接层串联的坐标输出网络,其中,第一全连接层输入大小设置为49,输出大小设置为1024,第二全连接层输入大小设置为1024,输出大小设置为5;

(6) 搭建检测网络:

(6a) 将匹配卷积核生成网络,特征金字塔网络,通道注意力模块,降噪模块,区域建议网络,RoIAlign网络,分类输出网络和坐标输出网络依次级联成检测网络;

(7) 训练检测网络:

(7a) 将训练集中的每一张图像分别与模板图像成一组,依次将每组数据输入到检测网络中,得到检测网络对每一组数据输出的预测框类别和预测框坐标;

(7b) 使用二值交叉熵,计算每组预测框类别与标签类别的类别损失值;使用平滑L1范数,计算每组预测框坐标与标签坐标的坐标损失值,将每组的类别损失值和坐标损失值相加,得到该组的损失值;

(7c) 利用反向传播法,用每组损失值迭代更新检测网络权重,直到检测网络损失值不再降低,得到训练好的检测网络;

(8) 对待检测图像进行检测:

(8a) 以256像素为裁剪步长,将一张待检测图像裁剪为多张 1024×1024 像素大小的子图像;

(8b) 分别将每一张子图像输入到训练好的检测网络中,得到每一张子图像的预测框坐标和预测框类别;

(8c) 将每张子图像的预测框坐标和预测框类别,按该子图像相对于原图位置映射到原始待检测图像上;

(8d) 使用非极大值抑制法,对待检测图像中的重叠预测框进行过滤,得到最终检测结果。

基于模板匹配和注意力机制的少样本弱小目标检测方法

技术领域

[0001] 本发明属于图像处理技术领域,更进一步涉及一种图像目标检测技术领域中的一种基于模板匹配和注意力机制的少样本弱小目标检测方法。本发明可用于对合成孔径雷达SAR图像、多光谱图像和遥感图像中的弱小目标进行检测。

背景技术

[0002] 目标检测技术通过对比当前信息与存储信息(记忆中的信息)实现影像内容的感知。深度学习在目标识别领域表现出无可比拟的优越性:通过逐层非线性变化,提取出从低到高的层级抽象特征,从而实现目标准确识别。然而在遥感图像的目标检测中,由于遥感图像的超高分辨率、广视场、样本数据缺乏以及目标所处背景干扰强和信噪比低的特点,使得现有目标检测的方法难以直接应用在此类场景的检测,即现有目标检测算法以大数据、海量清晰样本为驱动,无法直接处理少样本条件下的弱小目标检测问题。

[0003] 东南大学在其申请的专利文献“基于多层级特征选择卷积神经网络的遥感目标检测方法”(专利申请号:CN202110090408.2,申请公布号:CN112766184A)中公开了一种基于多层级特征选择卷积神经网络的遥感目标检测方法。该方法的步骤是:首先搭建卷积神经网络模型,并对搭建的卷积神经网络结构参数进行设定和训练参数初始化;再对训练图像进行预处理和标签格式转化,然后预处理和标签格式转化的训练图像进行数据增强;进行卷积神经网络模型训练,得到网络权重和偏置;将测试图像输入训练好的神经网络模型,得到定位和分类结果;该方法还使用了旋转框来检测物体,大大提高了定位精度。但是,该方法仍然存在的不足之处有两个:其一,该方法利用主干网络为ResNet50提取特征,由于此网络的深度较浅,对于弱小目标提取特征能力较弱,并且该方法在主干网络中使用了最大池化层,这很大程度上降低了主干网络提取出弱小目标特征图的质量;其二,该方法提出的模型的检测性能依赖于大量的训练数据作为支持,对于遥感图像这种数据量小的数据集来说适用性大大降低。

[0004] 中国科学院空天信息创新研究院在其申请的专利文献“一种基于迁移学习的小样本遥感目标检测方法及系统”(专利申请号:CN202010643231.X,申请公布号:CN111860236A)中公开了一种基于迁移学习的小样本遥感目标检测方法。该方法的步骤是:首先将所述待测遥感影像输入到预先训练好的双阶段目标检测模型中,得到待测遥感图像类别和水平框回归,其中,所述双阶段目标检测模型由源数据集训练得到所述双阶段目标检测模型的参数;然后在参数固定后由目标数据集微调双阶段目标检测模型中的迁移参数构建而成。该方法利用迁移学习的方法对已经训练好的模型在少样本遥感数据集上进行微调,来解决样本不足时模型的过拟合。该方法存在两个不足之处:其一,由于迁移学习使用时要求源数据集的特征和目标数据集特征不能相差过大,因而原模型需要在大量相似数据集上进行预训练,这大大增大了模型训练过程的难度,并且迁移学习使用不当会引入大量误检;其二,该方法对于遥感目标进行检测使用的是水平框,对于方向多变的遥感目标来说,水平框不能精确反映出遥感目标的方向而且水平框中会包含多余背景。

[0005] Adam Van Etten在其发表的论文“You Only Look Twice:Rapid Multi-Scale Object Detection In Satellite Imagery”(arXiv:1805.09512,24May 2018)中公开了一种基于YOLOV2目标检测网络的扩展结构用于遥感目标的检测方法。该方法的具体步骤如下,首先,先使用滑窗的方法对原始高分辨遥感图片进行切割,并保证一定重叠将图片切割为长宽均为416的子图;然后将子图送入检测网络进行训练,最终得到检测模型,在检测流程中使用水平框对遥感目标进行检测。该方法为了避免小目标的漏检,将主干网络中的步长从32减少到16,这样增大了该网络最后一层输出的特征图尺寸,以此来缓解小目标的丢失。但是该方法仍然存在两个不足之处:其一,该方法将主干网络的步长从32直接减少到16,虽然一定程度缓和了小目标丢失的情况,但是也很大程度降低了主干网络的提取特征的能力;其二,该方法对于遥感目标仍然使用水平框进行检测,不能精确反映出遥感目标的姿态以及会引入多余的背景。

发明内容

[0006] 本发明的目的是针对上述现有技术的不足,提出一种针对高分辨率图像中少样本弱小目标检测方法,旨在解决在上述图像中进行目标检测时使用迁移学习需要保证源数据集的特征和目标数据集特征不能相差过大,且数据集需要大量训练样本的问题,以及检测网络对小目标特征提取能力弱和水平框不能精确反映出遥感目标姿态的问题。

[0007] 实现本发明目的的技术思路是,使用主干网络ResNet101对图像特征进行提取,因为其比起主干网络ResNet50具有更深的网络结构,所以增大了检测网络的特征提取能力,解决了传统方法中使用较浅的主干网络ResNet50造成特征提取能力弱的问题。利用匹配卷积核生成网络生成匹配卷积核来进行模板匹配,由于此过程只需要一张模板图像和少量训练数据,因此解决了传统深度学习目标检测网络要求训练集具有大量训练样本,以及使用迁移学习会引入大量误检的问题。利用通道注意力模块和降噪模块串联来处理特征图,因为通道注意力模块可以增加特征图中重要通道的响应,降噪模块可以进一步提高特征图的信噪比,以此解决了传统检测网络对于小目标检测能力弱的问题。使用旋转矩形框对目标进行标注和检测,因为旋转矩形框相比于水平框具有更准确的标注精度,以此解决了水平框不能精确反映出遥感目标的姿态以及会引入多余的背景的问题。

[0008] 本发明的具体步骤如下:

[0009] (1) 生成训练集和模板图像:

[0010] (1a) 选取至少10张图像,以256像素为裁剪步长将每张图像裁剪为 1024×1024 像素的大小且每张图像包含至少1个目标,将裁剪过程中被截断的目标标记为特殊目标,将此目标的标志位difficult置为1,将裁剪后的图像和标志位组成数据集;

[0011] (1b) 对数据集中每个目标的4个角点以旋转矩形框的形式进行标注,得到每个目标的标注坐标;

[0012] (1c) 将每个目标的标注坐标和数据集组成训练集;

[0013] (1d) 随机选取1张目标轮廓清晰的图像作为模板图像;

[0014] (2) 搭建匹配卷积核生成网络:

[0015] 搭建一个由主干网络ResNet101和池化层组成的匹配卷积核生成网络,将池化层的核大小设置为 3×3 ,池化步长为2,池化层采用平均池化法;

- [0016] (3) 搭建一个具有4个阶段、且每个阶段的输出通道数均为256的特征金字塔网络；
- [0017] (4) 搭建通道注意力模块和降噪模块；
- [0018] (4a) 搭建通道注意力模块，其结构依次为：全局平均池化层，第一全连接层，ReLU激活函数，第二全连接层，Sigmoid激活函数，其中，全局平均池化层的核大小设置为 3×3 ，池化步长为2，采用平均池化法，第一与第二全连接层的输入通道数分别设置为256和128，输出通道数分别设置为128和256；
- [0019] (4b) 搭建降噪模块，其结构依次为：第一卷积层，ReLU激活函数，第二卷积层，ReLU激活函数，第三卷积层，Softmax函数，其中，第一、第二和第三卷积层的卷积核大小均设置为 3×3 ，步长均设置为1；
- [0020] (5) 搭建分类输出网络和坐标输出网络；
- [0021] (5a) 搭建一个由第一全连接层和第二全连接层串联的分类输出网络，其中，第一全连接层输入大小设置为49，输出大小设置为1024，第二全连接层输入大小设置为1024，输出大小设置为2；
- [0022] (5b) 搭建一个由第一全连接层和第二全连接层串联的坐标输出网络，其中，第一全连接层输入大小设置为49，输出大小设置为1024，第二全连接层输入大小设置为1024，输出大小设置为5；
- [0023] (6) 搭建检测网络；
- [0024] (6a) 将匹配卷积核生成网络，特征金字塔网络，通道注意力模块，降噪模块，区域建议网络，RoIAlign网络，分类输出网络和坐标输出网络依次级联成检测网络；
- [0025] (7) 训练检测网络；
- [0026] (7a) 将训练集中的每一张图像分别与模板图像成一组，依次将每组数据输入到检测网络中，得到检测网络对每一组数据输出的预测框类别和预测框坐标；
- [0027] (7b) 使用二值交叉熵，计算每组预测框类别与标签类别的类别损失值；使用平滑L1范数，计算每组预测框坐标与标签坐标的坐标损失值，将每组的类别损失值和坐标损失值相加，得到该组的损失值；
- [0028] (7c) 利用反向传播法，用每组损失值迭代更新检测网络权重，直到检测网络损失值不再降低，得到训练好的检测网络；
- [0029] (8) 对待检测图像进行检测；
- [0030] (8a) 以256像素为裁剪步长，将一张待检测图像裁剪为多张 1024×1024 像素大小的子图像；
- [0031] (8b) 分别将每一张子图像输入到训练好的检测网络中，得到每一张子图像的预测框坐标和预测框类别；
- [0032] (8c) 将每张子图像的预测框坐标和预测框类别，按该子图像相对于原图位置映射到原始待检测图像上；
- [0033] (8d) 使用非极大值抑制法，对待检测图像中的重叠预测框进行过滤，得到最终检测结果。
- [0034] 本发明与现有技术相比具有以下优点：
- [0035] 第一，由于本发明采用主干网络ResNet101对图像进行特征提取，通过使用主干网络ResNet101能够提取出图像中更加丰富的语义信息，有效的增加了检测网络的特征提取

能力,克服了现有技术中由于使用较浅主干网络造成特征提取能力弱的问题,使得本发明具有满足对复杂背景下具有良好特征提取能力。

[0036] 第二,由于本发明采用了匹配卷积核生成网络来生成匹配卷积核,通过匹配卷积核对图像中的目标进行模板匹配,有效降低了检测网络对于训练集中样本数目的需求量,克服了现有技术中需要大量训练样本的问题,以及避免了使用迁移学习引入大量误检的问题,使得本发明满足在少样本场景下能够具有良好的检测性能。

[0037] 第三,由于本发明采用了通道注意力模块和降噪模块,通过通道注意力模块和降噪模块串联处理特征图,有效增加了特征图中有效通道的响应,并且进一步提高特征图的信噪比,克服了现有技术在面对目标较小以及目标特征不明显时无法正确检测出目标的问题,使得本发明满足在小目标以及目标特征不明显场景下能够具有良好的检测性能。

[0038] 第四,由于本发明采用旋转矩形框对目标进行标注和预测,有效提高了待检测目标预测框的表示精度,大大减少了预测框中包含的背景比例,克服了现有技术使用水平框对目标进行标注和检测时,不能准确反映目标姿态以及会引入多余背景的问题,使得本发明满足在目标姿态多变的场景下具有能够准确表示目标姿态的能力。

附图说明

[0039] 图1是本发明的流程图;

[0040] 图2是本发明的仿真图。

具体实施方式

[0041] 下面结合附图和实施例对本发明作进一步的描述。

[0042] 参照图1,对本发明的具体步骤作进一步的描述。

[0043] 步骤1,生成训练集和模板图像。

[0044] 选取至少10张图像,以256像素为裁剪步长将每张图像裁剪为 1024×1024 像素的大小且每张图像包含至少1个目标,将裁剪过程中被截断的目标标记为特殊目标,将此目标的标志位difficult置为1,将裁剪后的图像和标志位组成数据集。

[0045] 对数据集中每个目标的4个角点以旋转矩形框的形式进行标注,得到每个目标的标注坐标,将每个目标的标注坐标和数据集组成训练集。

[0046] 随机选取1张目标轮廓清晰的图像作为模板图像。

[0047] 步骤2,搭建检测网络。

[0048] 搭建一个由主干网络ResNet101和池化层组成的匹配卷积核生成网络,将池化层的核大小设置为 3×3 ,池化步长为2,池化层采用平均池化法。

[0049] 搭建一个具有4个阶段、且每个阶段的输出通道数均为256的特征金字塔网络。

[0050] 搭建一个通道注意力模块,其结构依次为:全局平均池化层,第一全连接层,ReLU激活函数,第二全连接层,Sigmoid激活函数,其中,全局平均池化层的核大小设置为 3×3 ,池化步长为2,采用平均池化法,第一与第二全连接层的输入通道数分别设置为256和128,输出通道数分别设置为128和256。

[0051] 搭建一个降噪模块,其结构依次为:第一卷积层,ReLU激活函数,第二卷积层,ReLU激活函数,第三卷积层,Softmax函数,其中,第一、第二和第三卷积层的卷积核大小均设置

为 3×3 ,步长均设置为1。

[0052] 搭建一个由第一全连接层和第二全连接层串联的分类输出网络,其中,第一全连接层输入大小设置为49,输出大小设置为1024,第二全连接层输入大小设置为1024,输出大小设置为2。

[0053] 搭建一个由第一全连接层和第二全连接层串联的坐标输出网络,其中,第一全连接层输入大小设置为49,输出大小设置为1024,第二全连接层输入大小设置为1024,输出大小设置为5。

[0054] 将匹配卷积核生成网络,特征金字塔网络,通道注意力模块,降噪模块,区域建议网络,RoIAlign网络,分类输出网络和坐标输出网络依次级联成检测网络。

[0055] 步骤3,使用训练集和模板图像训练检测网络。

[0056] 将训练集中的每一张图像分别与模板图像成一组,依次将每组数据输入到检测网络中,得到检测网络对每一组数据输出的预测框类别和预测框坐标。

[0057] 使用二值交叉熵,计算每组预测框类别与标签类别的类别损失值;使用平滑L1范数,计算每组预测框坐标与标签坐标的坐标损失值,将每组的类别损失值和坐标损失值相加,得到该组的损失值。

[0058] 利用反向传播法,用每组损失值迭代更新检测网络权重,直到检测网络损失值不再降低,得到训练好的检测网络。

[0059] 步骤4,使用训练好的检测网络对待测图片进行检测。

[0060] 以256像素为裁剪步长,将一张待检测图像裁剪为多张 1024×1024 像素大小的子图像。

[0061] 分别将每一张子图像输入到训练好的检测网络中,得到每一张子图像的预测框坐标和预测框类别。

[0062] 将每张子图像的预测框坐标和预测框类别,按该子图像相对于原图位置映射到原始待检测图像上。

[0063] 使用非极大值抑制法,对待检测图像中的重叠预测框进行过滤,得到最终检测结果。

[0064] 以下结合仿真实验,对本发明技术效果进行进一步说明:

[0065] 1. 仿真条件。

[0066] 本发明在Intel (R) Core (TM) i7-10700K CPU@3.80GHz处理器,Nvidia (R) RTX3090显卡的电脑上,软件使用深度学习框架Pytorch完成仿真。

[0067] 2. 仿真内容与结果分析。

[0068] 仿真场景设置:为了验证本发明提出的基于模板匹配和注意力机制的少样本弱小目标检测方法能够在少样本条件下对弱小目标进行检测,本发明的仿真实验场景设置为从SAR图像中检测出舰船目标,可用数据集图像数量为10张。以256像素为裁剪步长将每张图像裁剪为 1024×1024 像素的大小且每张图像包含至少1个目标的子图像,从这些子图像中随机选取1张目标轮廓清晰的图像作为模板图片,剩余的图像使用旋转矩形框标注后作为训练集。

[0069] 搭建完检测网络后,设置最大训练轮数为36轮,将训练集中的每一张图像分别与模板图像成一组,依次输入网络迭代更新网络权重,最后得到训练完好的检测网络。

[0070] 为了证明仿真效果,将待检测图像设置为舰船目标较小和背景复杂的场景。使用训练好的检测网络对待检测图像进行检测,得到检测结果。图2(a)和附图2(b)是采用本发明方法在背景复杂场景下的检测结果,使用旋转矩形框对检测结果进行标注。图2(c)是采用本发明方法在小目标场景下的检测结果,使用旋转矩形框对检测结果进行标注。其中,复杂场景指的是图像中背景的面积大于目标面积的场景,小目标场景指的是图像中包含目标面积小于 32×32 像素大小的场景。

[0071] 图2(a)给出了本仿真实验在复杂场景下单一目标的检测结果图,由图2(a)可知,在复杂场景下,即使目标数目较少本发明方法也能取得良好的检测效果。附图2(b)给出了本仿真实验在复杂场景下多目标的检测结果图,由图2(b)可知,在复杂场景下,即使目标数目较多本发明方法也能取得良好的检测效果。附图2(c)给出了本仿真实验在多个小目标场景下的检测结果图,由图2(c)可知,在小目标场景下,即使目标数目较多本发明方法也能取得良好的检测效果。

[0072] 综上所述,从仿真效果图的分析可知,本发明提出的一种基于模板匹配和注意力机制的少样本弱小目标检测方法,实现了在少样本条件下,对小目标能够进行准确检测,目标检出数目准确,标注位置准确,同时由于采用了匹配卷积核生成网络生成匹配卷积核进行模板匹配,使得检测网络对于训练集样本数量需求量大大减少,并且通过使用通道注意力模块和降噪模块,使得检测网络在背景复杂场景下也能具有良好的检测性能,使该发明在现实工程应用中更有优势。

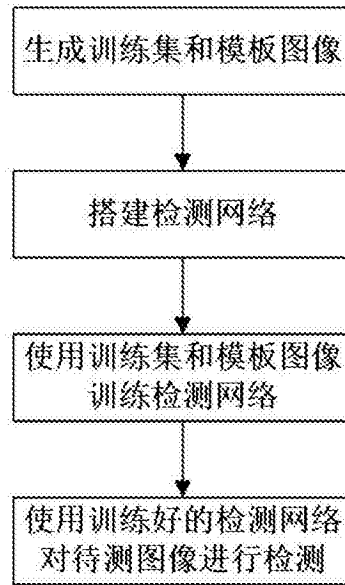


图1

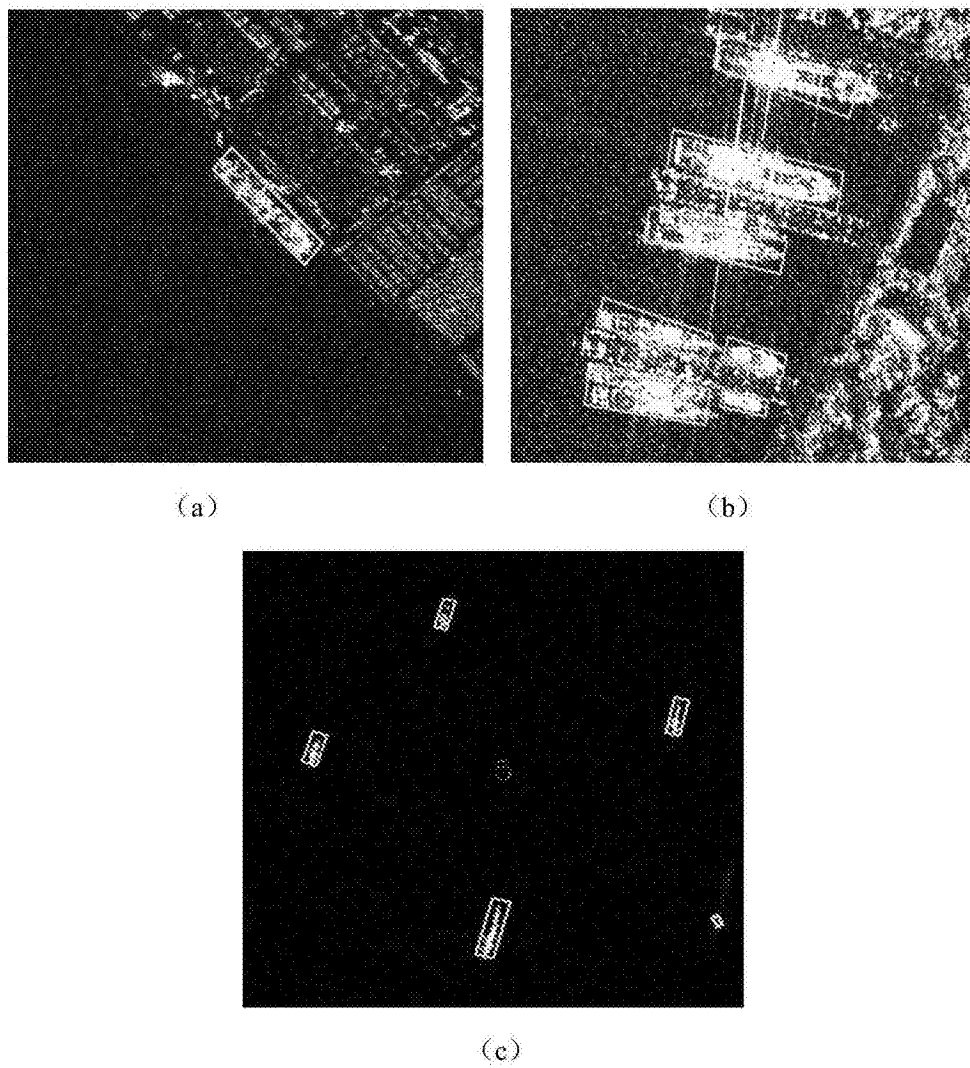


图2