



计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331,CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目: 基于无锚框网络的航拍航道船舶检测算法
作者: 光睿智, 安博文, 潘胜达
网络首发日期: 2020-10-09
引用格式: 光睿智, 安博文, 潘胜达. 基于无锚框网络的航拍航道船舶检测算法. 计算机工程与应用.
<https://kns.cnki.net/kcms/detail/11.2127.TP.20201009.1052.002.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于无锚框网络的航拍航道船舶检测算法

光睿智, 安博文, 潘胜达

上海海事大学 信息工程学院, 上海 201306

摘要: 针对无人机航拍航道船舶影像中船舶目标较小、尺度变换大、背景复杂等问题, 提出了一种基于 FoveaBox 网络的单阶段无锚框的航道船舶检测算法 FoveaSDet。为提升小目标的检测精度, 该算法使用基于残差网络改进的 SEResNeXt-I 作为骨干网。为改善尺度变换问题, FoveaSDet 采用 Foveahead 实现无锚框目标检测。同时为提高复杂背景下检测框的定位精度, 使用完全交并比损失实现边框回归。经实验测试, FoveaSDet 算法在实景航拍数据集上的平均准确率(AP)和小目标准确率(AP_s)分别为 71.6%和 47.0%, 相较于原始的 FoveaBox 提高了 4.9%和 6.2%, 体现了更好的总体检测精度和小目标检测能力。

关键词: 深度学习; 目标检测; 特征提取; 无锚框; 损失函数

文献标志码:A 中图分类号:TP391.4 doi: 10.3778/j.issn.1002-8331.2005-0056

光睿智, 安博文, 潘胜达. 基于无锚框网络的航拍航道船舶检测算法. 计算机工程与应用

GUANG Ruizhi, AN Bowen, PAN Shengda. Channel ship detection algorithm for aerial image based on anchor-free network. Computer Engineering and Applications

Channel ship detection algorithm for aerial image based on anchor-free network

GUANG Ruizhi, AN Bowen, PAN Shengda

College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China

Abstract: In this paper, a one-stage anchor-free channel ship detection algorithm called FoveaSDet is proposed to address the problem of the small size ship, large scale transformation, and complex background in the image of drone aerial. First, FoveaSDet uses a backbone called SEResNeXt-I, which improved from ResNeXt, to increase the detection accuracy of small objects. Next, to solve the large scale transformation problem, the purpose method uses Foveahead to achieve anchor-free box object detection. Finally, the complete IOU loss has used to accomplish the bounding box regression so that the positioning accuracy of the detection box under a complex background could be raised. The experiment results show that the average

基金项目: 国家自然科学基金(No. 61302132, No.61504078, No.41701523)。

作者简介: 光睿智(1996-), 男, 硕士研究生, 主要研究方向为计算机视觉、目标检测, CCF 学生会员, E-mail: guangruizhi10@stu.shmtu.edu.cn; 安博文(1969-), 男, 博士, 教授, 主要研究方向为遥感图像处理; 潘胜达(1986-), 男, 通讯作者, 博士, 讲师, 主要研究方向为遥感图像处理。

precision and small objects' average precision of the FoveaSDet is 71.6% and 47.0% testing on the real aerial image dataset. It has increased by 4.9% and 6.2% compared with FoveaBox, reflecting better detection accuracy and small object detection ability.

Key words: deep learning; object detection; feature extraction; anchor-free; loss function

1 引言

伴随着国内航运事业的迅猛发展,内河航道愈发拥挤,给海事部门的监管带来许多困难。随着无人机技术的快速进步,采用无人机航拍影像进行航道船舶目标检测已经成为较有成效的辅助手段,其优势在于覆盖面广,可以实现目标航道无死角的监控。然而,无人机影像中船舶目标较小,尺度变换较大,具有易抖动、易模糊和航道两边背景特征复杂的问题,进行船舶目标检测的难度较高。传统的目标检测算法一般基于人工设计特征,如 HOG^[1]、SIFT^[2]、SURF^[3]等,因此在实际使用时检测效果不够理想,且泛化能力较弱。

现今计算机视觉领域中基于深度学习的目标检测算法已经在目标检测方向得到广泛的应用。这类算法若根据网络结构上的差异进行划分,依照是否有候选区域网络分为单阶段和两阶段两类。单阶段目标检测算法的检测精度普遍较之两阶段算法稍差,但其一般具有更快的检测速度,更适合工程应用。此外,依据不同的边框回归策略,基于深度学习的目标检测算法也分成基于锚框(anchor-base)和无锚框(anchor-free)两类。锚框是由数据集预处理产生的一种超参数,通过预先规范检测框的形状来提高网络召回率和精度。虽然使用锚框可以在一定程度上的优化算法性能,但是其存在并不符合人眼视觉特性,也不利于算法的泛化。因此,设计基于无锚框的算法实现目标检测已经成为该领域的重要方向。对于单阶段目标检测算法,基于锚框的算法有 YOLOV3^[4]、SSD^[5]、RetinaNet^[6]等,无锚框的算法也有 FoveaBox^[7]、FCOS^[8]等算法。

作为目标检测中的重要分支,基于深度学习的航拍图像船舶检测是当前目标检测领域的研究热

点之一。但航拍影像中目标普遍偏小、物体尺度变化较大,因此需要对这类算法进行一些有针对性的改进。在小目标检测精度提升方面,裴伟等^[9]使用基于 SSD 的方法,改进了算法的特征提取网络和检测网络的结构;史文旭等^[10]在 SSD 算法上增加特征增强模块,同时采用候选框匹配的方式改善检测框的定位问题。这两种方式均在目标数据集获得了较好效果,但受限于算法中特征提取网络的分类性能,其小目标检测精度仍有待提高。对于待测目标大范围尺度变换方面,谢学立等^[11]基于 RetinaNet 算法,使用网络旁路构造动态感受野改善尺度变换问题;李晓光等^[12]使用改进的 Faster R-CNN^[13]算法进行多尺度检测;Gui 等^[14]在 Faster R-CNN 中添加尺度变换模块,对残差网络的输出特征进行处理。这三者均在检测对象尺度变换较大的情况下实现算法精度提高,但其定位匹配方式均基于锚框,需要借助较多的超参数,限制了算法迁移。对于待测目标背景复杂方面,王健林等^[15]基于 R-FCN 网络通过在残差块中混合卷积尺度提高检测精度;于野等^[16]使用改进的 A-FPN 网络改善特征提取方式,提高了复杂背景下的算法性能。这类算法着重于提高船舶检测准确率,但算法复杂度较高。

由上述分析可知,面对航道船舶目标较小、尺度变换大、背景复杂等问题,目前存在的大部分单阶段算法的检测精度稍显不足,两阶段算法虽提高了准确率却使网络结构复杂化,同时这两类算法多数基于锚框实现回归策略。为此,本文提出了一种无锚框的单阶段目标检测算法 FoveaSDet。首先, FoveaSDet 算法为提高小目标和整体检测精度,在算法特征提取部分提出了基于残差网络^[17](Residual Network, ResNet)的骨干网 SEResNeXt-I,该网络在 ResNeXt^[18]的基础上改进了输入数据的特征提取结

构，同时也优化了网络特征层中数据的空间关联性；其次，为改善尺度变换对检测精度的影响，同时实现无锚框的边框回归，算法中引入了 FoveaBox 中的 Foveahead 模块实现目标分类和定位；最后，为增加复杂背景下的定位精度和检测框的交并比 (Intersection over Union, IOU)，使用完全交并比损失^[19](Complete-IOU, CIOU)作为边框回归损失。

2 FoveaSDet 目标检测算法

2.1 FoveaSDet 算法检测流程

本文所提出的 FoveaSDet 算法主要由三个模块构成：特征提取网络、无锚框检测模块和损失函数。

该算法使用了语义分割的思想设计正负样本区域划分，并与 FPN 相结合，用简洁的方法实现了无锚框的边框回归。

损失函数对算法的学习能力有决定性影响，可将其分为两类：分类损失函数及边框损失函数。针对本文要实现的船舶检测任务，选择分类损失函数 FocalLoss 实现分类，同时引入边框损失函数 CIOU。

FocalLoss 用以缓解数据集样本不平衡的问题，CIOU 旨在进一步挖掘图像特征信息与样本真值 (GroundTruth, GT)之间的联系，提升算法检测框的

特征提取网络，即骨干网，负责提取图像中包含的不同层次的特征信息。旨在提高 FoveaSDet 算法整体的检测精度及小目标检测性能，本文优化了基于 ResNeXt 的骨干网及网络块的内部结构两个部分，并称改进后的骨干网为 SEResNeXt-I。同时，为了适配多尺度检测，进一步提升算法的整体性能，在骨干网中增加了特征金字塔网络^[20](Feature Pyramid Network, FPN)结构。

无锚框检测模块旨在抛却传统算法基于锚框的边框回归的策略，使得算法的检测方式更符合人眼视觉，预设的超参数更少，并拥有更好的泛化能力。本文选择 Foveahead 实现该目标，其优势在于

定位精度。

FoveaSDet 算法实现目标检测的流程如图 1 所示。首先将待测图片输入特征提取网络，通过 SEResNeXt-I 及 FPN 完成对图片基本特征的提取，获得 5 层不同尺度的特征映射，每层的数据均需通过 Foveahead 分成两个子网络，分别完成目标的分类和检测框的预测。最后在各层完成检测之后按尺度融合，并对检测框进行软非极大值抑制^[21](Soft Non-Maximum Suppression, Soft-NMS)获得最后的检测结果。

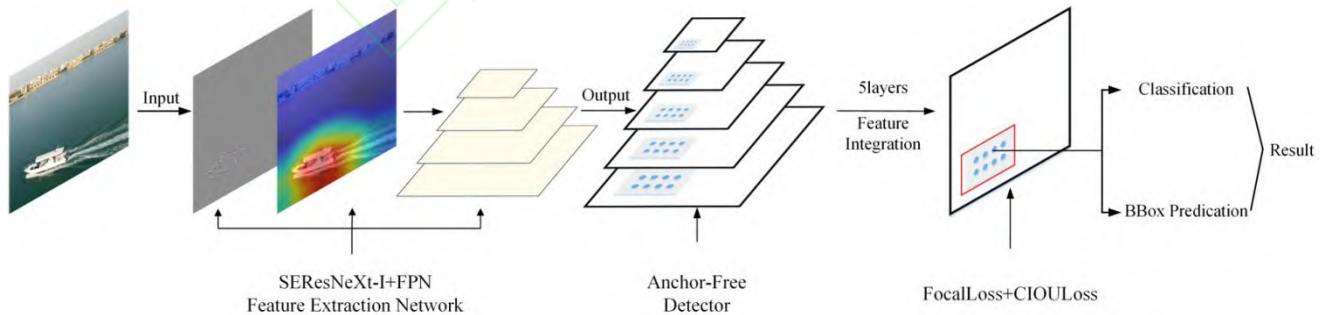


图 1 FoveaSDet 船舶检测流程

Fig.1 Process of FoveaSDet ship detection

2.2 SEResNeXt-I 骨干网

残差网络作为现今常用的特征提取网络，其特

点在于通过在网络块中构造恒等映射的方式，缓解了深层网络学习能力退化的问题。同时在残差块中增加批量归一化^[22](Batch Normalization, BN)层及

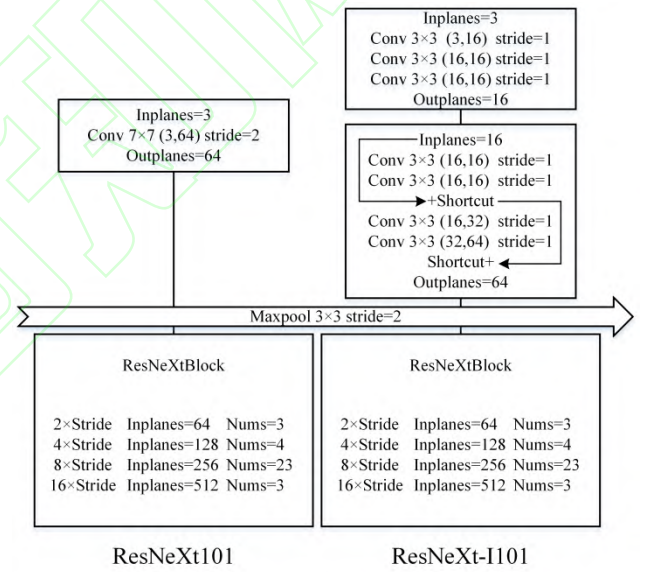
Relu 激活函数,降低了学习成本,并且改善了网络的训练环境。然而,即便残差网络具有很多良好的特性,其对于小目标的特征提取能力稍显逊色。为此,本文对骨干网的网络结构和残差块的内部结构两部分内容进行了改进。

首先,本文优化了骨干网用来处理输入数据的网络结构,旨在减少输入特征损失,保留更多特征信息,提高对小目标的检测精度,其改变如图 2(a)所示。左侧为主流的残差网络结构 ResNeXt101,在进入残差网络前使用步长为 2 的 7×7 卷积处理输入图片。虽然这种方式使用大小为 7 的卷积核来提高感受野,但是较大的步长仍使网络在处理输入数据时便损失了许多图片特征。为了减少特征损失,并且保留充足的感受野,将网络结构改进为图 2(a)右侧所示。将 7×7 卷积核拆分成 3 个 3×3 卷积核,同时步长变为 1,输入通道缩减为 16。这样既能保留网络的空间感知能力,又增加了输出特征数量。此外,通过降低网络的初始通道数,后续网络输入参数也随之减少。随后在网络结构中增加了 2 个跳跃连接结构,这使得网络中添加了更多的非线性映射,同时令不同层次的特征在块之间流动。本文将完成结构改进后的网络称为 ResNeXt-I。当设定其网络深度为 101 时,用 ResNeXt-I101 表示,文中的骨干网名称均采用此类方式命名。

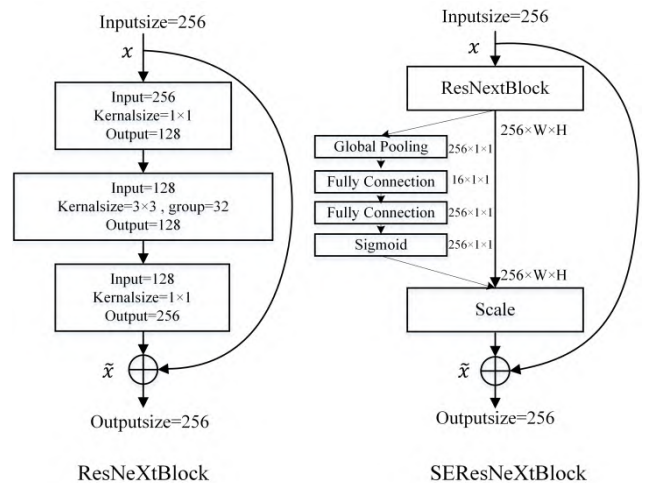
其次,为了提高网络层中不同特征的利用率,建立通道之间的依赖关系,同时改善网络全局的检测精度,本文对网络残差块的内部结构进行改进,将压缩激励^[23](Squeeze and Excitation, SE)模块引入残差块中,其样式如图 2(b)所示。SE 模块能使特征提取网络在学习过程中不断更新各个特征通道的重要程度,筛选出重要特征并抛弃掉无用的特征,这一步也剔除了冗余的参数,降低网络整体的计算开销。SE 模块可分为压缩和激励两个部分。首先为压缩部分,对于 ResNeXt 残差块输出的数据,使用全局平均池化完成对二维数据的空间压缩,将特征通道压缩成一个实数,该实数映射了对应特征通道的全局表达。令该值为 z_c ,通道的高为 h ,宽为 w ,二维平面的数据为 $x_c(i, j)$,则压缩部分的表达式为:

$$z_c = \frac{1}{h \times w} \sum_{i=1}^h \sum_{j=1}^w x_c(i, j) \quad (1)$$

完成压缩操作后,算法将要实现的是激励操作。对特征通道压缩后获得了当前通道的特征权重,为了平衡数据集整体的特征权重,使用全连接层建立不同通道的相关性。此外,为降低计算量,第一个全连接层将输入降低为原来 1/16,随之再经过一个全连接层用以实现特征通道在整体上的关联。这种方式也为输入输出增添了更多非线性元素,有助于更好的呈现特征。在完成激励操作后,通过 Sigmoid 激活函数完成归一化,并通过缩放操作将获得的权重加权到各个特征上,完成 SE 模块的整体流程。



(a) Improvement of the backbone structure



(b) Improvement of the bottleneck module

图2 骨干网结构和残差块结构改进

Fig.2 Improvement of backbone and residual block structure

最后,在特征提取网络的末端使用 FPN 结构。FPN 在原始的特征映射上进行步长为 2 的下采样,缩放出 4 个不同尺度的特征层。每个特征层不仅包含原始输入层的特征信息,还与经过上采样的下一层特征层相融合。这使得每层特征映射既包含了当前尺度的特征细节,又结合了来自上层的更抽象的语义信息。这 4 层特征与原始特征一起构成了 5 个不同尺度的特征映射,其输出数据将用于之后的无锚框模块,完成目标的分类和定位。相较而言,进行过两点改进并添加 FPN 结构的 SEResNeXt-I 网络比单纯的 ResNeXt 有更低的分类误差。

2.3 无锚框检测模块

传统的基于锚框的目标检测算法,在检测前首先通过聚类等手段计算与数据集相适应的 N 组锚框,随后在检测时根据锚框的大小进行边框回归。为实现无锚框边框回归策略,同时更好的处理船舶检测时尺度变换较大的问题,引入 FoveaBox 算法中 Foveahead 的实现无锚框检测,具体的检测流程如图 3 所示。Foveahead 首先对特征施行尺度划分。对于由 FPN 结构输出的 5 类尺寸的特征层, Foveahead 设定每层用于检测的基础大小为 S_l , $S_l \in \{32^2, 64^2, 128^2, 256^2, 512^2\}$ 。该检测面积可以通过常数 η 控制缩放,使得每层的尺寸范围为 $[S_l/\eta^2, S_l \cdot \eta^2]$ 。本文设定 $\eta=2$,这会使不同特征层的尺度范围存在相交的现象。因此 Foveahead 将根据 GT 的尺寸,在一个或多个符合尺度的特征层中完成检测并融合结果。

在确定各个特征层的尺度划分后, Foveahead 将特征映射划分为正、负样本两部分,辅助完成目标的分类和回归。Foveahead 先通过样本的 GT 计算出在第 l 个尺度的特征映射中,位于目标中心的位置 (c'_x, c'_y) ,其公式如式(2)所示:

$$\begin{aligned} x'_1 &= \frac{x_1}{2^l}, y'_1 = \frac{y_1}{2^l}, x'_2 = \frac{x_2}{2^l}, y'_2 = \frac{y_2}{2^l} \\ c'_x &= x'_1 + 0.5(x'_2 - x'_1), c'_y = y'_1 + 0.5(y'_2 - y'_1) \end{aligned} \quad (2)$$

其中 (x'_1, y'_1) , (x'_2, y'_2) 分别对应第 l 个特征层中 GT 的左上角和右下角。计算出目标区域中心点后,通过参数 σ_1 调节区域边界的高和宽,使得区域整体向样本中心收缩。该目标区域内部的点将在训练中作为正样本,随后再使用参数 σ_2 使目标边界向外扩张,使得扩张后边界的外部区域成为负样本。最后去除正负样本之间的区域,实现了正负样本的差异化判别。下列为正样本区域的计算公式,负样本与之类似,其中缩放区域的左上角和右下角坐标为

(x''_1, y''_1) , (x''_2, y''_2) :

$$\begin{aligned} x''_1 &= c'_x - 0.5(x'_2 - x'_1)\sigma_1, y''_1 = c'_x - 0.5(y'_2 - y'_1)\sigma_1 \\ x''_2 &= c'_x + 0.5(x'_2 - x'_1)\sigma_1, y''_2 = c'_x + 0.5(y'_2 - y'_1)\sigma_1 \end{aligned} \quad (3)$$

样本划分完成后,对于正样本区域上每一个坐标, Foveahead 都将回归其分类结果和检测框。因各个特征层上坐标点位于不同尺度的特征映射,需要将坐标变换到原始图像。随后引入损失函数计算分类结果及检测框的定位结果,并衡量与 GT 的差异。最后通过大小为 0.05 的置信度阈值筛选出排名前 1000 的检测框,并使用阈值为 0.5 的 Soft-NMS 获得最终的检测结果。

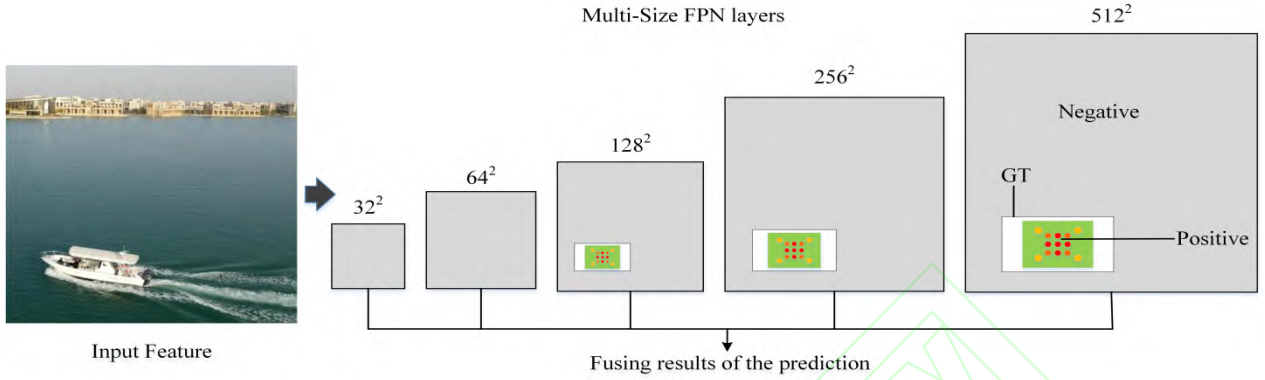


图3 Foveahead 检测流程

Fig.3 Process of Foveahead prediction

2.4 损失函数

对于 FoveaSDet 算法的分类损失，本文采用的是 FocalLoss，旨在缓解正负样本区域的不平衡，降低分类误差，其公式为：

$$L_{fl} = \begin{cases} -\alpha(1-y')^\gamma \ln y', & y=1 \\ -(1-\alpha)y'^\gamma \ln(1-y'), & y=0 \end{cases} \quad (4)$$

式中 y 为 GT 的标签，代表样本正负， y' 为检测分类置信值。 α 用于区分样本的重要性， γ 用来分辨样本的难易度，此两者共同约束了训练时获得的损失，本文设定 $\alpha=0.5$ ， $\gamma=4$ 。对于容易训练的样本，其损失 L_{fl} 相对较小，反之难于训练的样本的分类损失将会放大。

FoveaSDet 算法的边框回归损失选取 CIOU，该损失函数属于 IOU 损失函数中的一类。CIOU 损失继承了 IOU 损失具有的尺度不变性、非负性等优点，同时也克服了 IOU 损失无法反映检测框之间状态的缺点，该损失函数的优势为：

1) 衡量了 IOU=0 时的情况，使不相交的目标之间也存在梯度损失进行计算。

2) 考虑了边框重合时的情况，通过计算中心点距离使检测框更好的趋向 GT 中心移动。

3) 引入宽高比概念，使得检测框和 GT 有更相

近的矩形框形状。

若令检测框为 P ，GT 框为 T ，则 IOU 的公式为式(5)所示：

$$IOU = \frac{|P \cap T|}{|P \cup T|}, \quad L_{iou} = 1 - IOU \quad (5)$$

为评估在框的集合 S 中不相交的框 P 和 T ，设 C 为 S 集合中包含 P 、 T 的最小框，则广义交并比损失函数^[24](Generalized-IOU, GIOU)的公式为：

$$GIOU = IOU - \left| \frac{C(P \cup T)}{C} \right|, \quad (6)$$

$$-1 \leq GIOU \leq 1, \quad L_{giou} = 1 - GIOU$$

该公式计算出 C 中 P 、 T 以外的部分与 C 的比值，修正了 IOU 损失的计算值，可以更好反映检测框相交的情况，可将其视作 IOU 的下界。

在 GIOU 基础上，为了提升损失函数的整体性能，CIOU 引入了对检测框与 GT 框的中心点距离和宽高比的计算，在相交面积和形状比例两方面实现对检测框的约束。设 ρ 为中心点的距离， b, b^g 为两框的中心点， c 为最小包围框 C 的对角线距离， α 为比例平衡系数， ν 为一致性系数，则 CIOU 损失的表示式为：

$$\begin{aligned}
CIOU &= IOU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v, \\
L_{GIOU} &= 1 - CIOU, \\
\alpha &= \frac{v}{1 - IOU + v}, \\
v &= \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2
\end{aligned} \tag{7}$$

该算式首先将 GIOU 的设定的区域比值转化为距离比值, 这能更好的反映两框之间的关系。其次, 该方法降低了训练所需的计算量, 使得函数收敛的更快, 边框能较快的移动到目标位置。最后, 为更好的获得边框形状, 进一步提高检测结果, CIOU 增加参数 v 来评估两框一致性, 同时使用 α 控制尺度及梯度移动方向, 规范边框的收敛样式。

3 实验及结果分析

3.1 实验环境

本文采用的数据集源自于大疆M100无人机在黄浦江上空的实景拍摄, 共5743张图片, 分辨率均大于1280×720, 包含船舶和背景两类, 其样本示例如图4所示。相较于目前的许多公开数据集而言, 本文数据集的检测难度较高, 目标的尺度会随无人机的航迹发生变化, 船舶目标较小且样本数量偏少。通过考虑上述因素, 选择扩大训练集的比例, 将数据集的训练集、验证集和测试集划分比例为7:1:2, 同时将数据集标注为COCO格式方便进行更细致的结果展示。

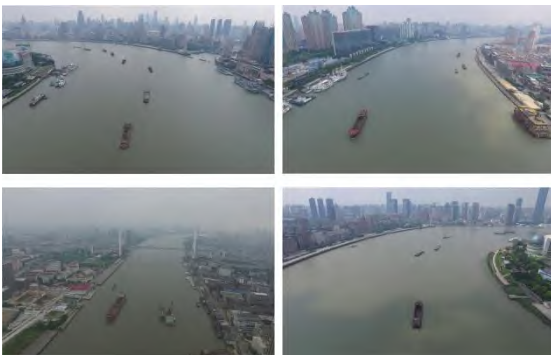


图4 数据集样本示例

Fig.4 Samples of the dataset

本文实验在Ubuntu16.04操作系统下完成, 算法部署于Pytorch1.1框架, 使用RTX2080Ti进行训练及测试, 并基于CUDA10.0及CUDNN7.3完成计算加速和算法优化。

本文在训练时采取了一系列数据增强^[25]手段, 旨在改善部分特征的展现形式, 提高算法最终的检测精度。首先, 通过对样本进行上下、左右翻折, 图像随机旋转和抖动, 图像扭曲模糊及亮度调整等操作, 完成了数据集的增广。这些手段也能使训练出的权重更加鲁棒, 抗干扰失真的能力更强。其次, 在训练中增加了多尺度训练、随机剪裁等技巧。多尺度训练通过在训练时缩放输入数据改变算法输入, 既提高了算法的训练速度, 又增强了鲁棒性。输入图像的随机剪裁变相扩大了数据集的样本数, 也使样本展现的形式更加丰富。最后, 使用预热学习率(warm-up)的方式设定学习率, 提高了算法学习的上限。

此外, 为减少超参数及训练样本数据差异对算法的影响, 在对各类算法进行测试时均未使用预训练权重, 旨在单纯测试算法在设定环境下的学习能力。本文算法选取随机梯度下降法完成梯度计算, 初始学习率为0.001, 动量为0.9, 衰减系数为0.0001, 迭代轮数为24轮。

3.2 评价标准

本文使用平均准确率(Average Precision, AP)、召回率(Recall)、交并比(IOU)及 PR 曲线来评估算法检测结果。平均准确率展示算法的检测精度; 召回率表征样本被正确检测的概率; 交并比作为衡量检测精度的门限, 影响边框回归与检测精度的关系; PR 曲线展示不同 IOU 阈值下准确率和召回率之间的比例关系, 该曲线闭合区域的面积数值即为对应算法的平均准确率。准确率 P 、召回率 R 和平均准确率 AP 的公式如式(8)、(9)所示:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad R = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{8}$$

$$AP = \int_0^1 P(R) dR \tag{9}$$

其中 N_{TP} 表示算法检测正确的正样本数, N_{FP}

表示算法检测错误的正样本数，两者相和为总的正样本数， N_{FN} 指代检测错误的负样本数。

3.3 消融实验

将 FoveaSDet 的骨干网和损失函数部分进行消融实验，旨在测试算法对小目标的检测优势及其边框定位性能。首先对骨干网的性能进行测试，主要

比较 SEResNeXt-I 和残差网络对小目标检测精度 AP_s 、网络中所含参量及 Flops。当骨干网深度为 101，分组数为 32 且宽度为 4 时，SEResNeXt-I 与各网络的对比结果展示在表 1 中。

表 1 骨干网对算法性能影响

Table 1 Influence of backbone on algorithm performance

Backbone	AP_s	Paramters	Flops
ResNet101/FPN	38.5	57.83 M	253.99
ResNeXt101-32×4d/FPN	40.6	56.29 M	252.39
ResNeXt-I101-32×4d/FPN	41.3	61.10 M	275.22
SEResNeXt-I101-32×4d/FPN	47.0	40.89 M	187.30

表 2 边框损失函数对算法性能影响

Table 2 Influence of bounding box loss function on algorithm performance

BoundingBox Loss	$AP IOU=0.75 $	$AR IOU=0.5:0.95 $
Smooth-L1	78.7	74.4
GIOU	80.8	75.6
CIOU	81.6	75.9

由表 1 中可知，将骨干网从 ResNeXt 改进为 ResNeXt-I 时，小目标检测精度提高了约 0.7%。增加 SE 模块后，SEResNeXt-I 对小目标精度相较原版 ResNeXt 提高 6.4%，其中 SE 模块贡献约 5.7% 的准确率增益。SEResNeXt-I 通过改变网络结构，并使用 SE 模块进行特征筛选，两者相结合降低了网络整体的参量。

随后测试边框损失 CIOU 损失对 FoveaSDet 算法目标检测性能的影响。为了衡量在较高 IOU 阈值下边框回归的检测精度，使用 IOU 阈值为 0.75 的

AP 及边框平均召回率进行评估。将 CIOU 损失与 $L1^{[26]}$ 损失、GIOU 损失比较，其结果如表 2 所示。经测试证明采用 CIOU 对算法的整体性能有一定提高，其检测精度提升了 2.9%，平均召回率上升了 1.5%。使用 CIOU 损失改善了网络的训练效率，同时也对算法检测精度提升做出了贡献。

消融实验的测试结果证明，本文改进所得的骨干网及边框损失两个模块均在原有基础上提高了算法的检测精度。骨干网 SEResNeXt-I 在提高小目标检测精度的同时降低了网络参量，损失函数

CIOU 为算法的边框定位提供了增益。

3.4 算法对比结果及分析

将 FoveaSDet 算法与基于锚框的算法 SSD、YOLOV3、RetinaNet，无锚框的算法 FCOS 及 FoveaBox 进行比较，其测试结果如表 3 所示。FoveaSDet 在使用网络深度为 101，分组数为 32 且宽度为 4 时的骨干网 SEResNeXt-I101-32×4d 时，其在整体 AP 上相较于使用 ResNeXt101-32×4d 的 FoveaBox 上升了 4.9%。对于高 IOU 判决下的检测精度，两算法的差异更为明显，FoveaSDet 算法的 AP_{75} 较后者提高了约 8.6%，同时对小目标的准确率 AP_s 也提高了约 6.2%，各项性能均实现了对 FoveaBox 的超越。随后比较 FoveaSDet 算法使用 L1、GIOU 和 CIOU 回归损失的性能差异。GIOU 相比 L1 损失对于算法整体 AP 均有提升，而算法采用 CIOU 相较于 GIOU 而言虽然各项数值提升较小，但改善了算法的训练及收敛情况，其测试结果受数据集及骨干网性能影响。最后分析 FoveaSDet 算法的计算效率。在骨干网相同的情况下，该算法相较于 FoveaBox 的帧率提高了 23%，在使用 SEResNeXt-I 骨干网后，其帧率相较 ResNeXt 版本降低了 0.5 至 1 帧。虽然 SEResNeXt-I 使用 SE 模块提高了检测性能，同时压缩了网络参数，但其较为复杂的数据处理方式仍对检测速率产生了影响。

将 FoveaSDet 与常用的一些单阶段检测算法，如 YOLOV3 和 SSD 算法进行比较，本文算法的 AP

较其中性能最好的 SSD512 提升了约 9.6%。与同样使用 FocalLoss 进行分类的 RetinaNet 相比，本文算法在整体精度上也有较大提升，同时拥有更快的检测速率。相较于另一种较好的无锚框单阶段检测算法 FCOS，其 AP 比之高了 2.2%，检测速率也与之基本持平。

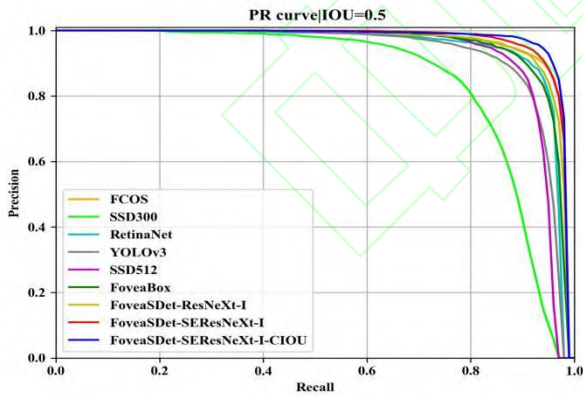
图 5 显示 FoveaSDet 算法在使用不同骨干网的情况下，与表 3 中其余算法在 IOU 阈值分别为 0.5 和 0.75 时的 PR 曲线，该图表明本文算法相较于其他算法在检测精度和召回率两部分都有一定优势。图 6 展示了 FoveaSDet 算法对测试集部分数据的检测结果。从图中可以看出，本文算法在航道两边的背景极为复杂，光线、阴影、视角及船舶尺度因素各不相同的情况下，仍能很好的实现船舶的目标检测，其对于小目标的检测效果较好。并且 FoveaSDet 算法在尺度变换较大并且背景较复杂的情况下，检测框的定位良好。

图 7 展示了在不同环境下，YOLOV3、SSD 与 FoveaSDet 三种算法的检测结果。图中错检漏检的部分用红框标注，检测框偏移等问题用黄框标出。对于不同背景分布、光照环境不同、船舶数量尺度不一的四类图片中，YOLOV3 和 SSD 因精度等问题出现了一定的错检漏检情况，同时伴有检测框堆叠的情况。相较而言，FoveaSDet 算法在相同场景下实现了更好的检测结果，其检测框的定位相对较好。对小目标检测精度和检测框定位精度的提升得益于算法对骨干网和边框回归损失的改进。

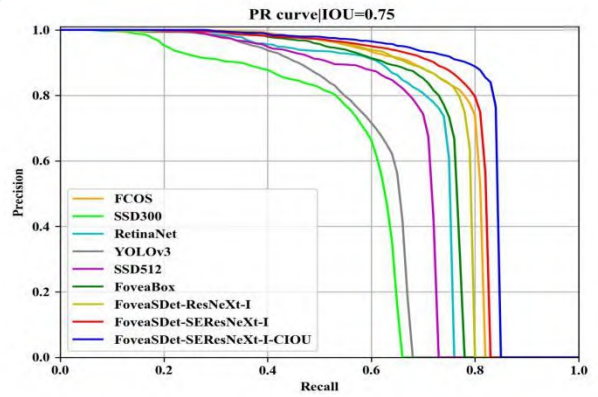
表 3 多种算法船舶检测精度及召回率

Table 3 Multiple algorithms for ship detection accuracy and recall rate

Mothed	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L	AR	FPS
<i>Anchor-Base</i>								
SSD300-VGG	53.0	84.1	56.3	26.5	58.9	63.3	63.3	61.6
SSD512-VGG	62.0	92.6	67.4	35.2	65.1	73.9	68.1	40.3
YOLOV3-Darknet53	58.1	92.8	60.6	28.5	61.4	68.6	63.6	55.0
RetinaNet-ResNeXt101-32×4d	64.5	94.5	70.8	28.2	67.1	77.8	69.7	11.2
<i>Anchor-Free</i>								
FCOS-ResNeXt101-32×4d	69.4	96.2	77.4	40.0	70.0	81.2	73.5	14.9
FoveaBox-ResNeXt101-32×4d	66.7	95.0	73.0	40.8	68.5	77.3	71.4	12.4
FoveaSDet-ResNeXt101-32×4d	68.6	95.2	76.3	40.6	70.3	79.7	73.4	15.2
FoveaSDet-ResNeXt-I101-32×4d	68.5	95.8	75.8	41.3	70.4	79.3	72.5	14.8
FoveaSDet-SEResNeXt-I101-32×4d-L1	70.2	96.6	78.7	45.0	72.0	80.2	74.4	14.5
FoveaSDet-SEResNeXt-I101-32×4d-GIOU	71.3	96.8	80.8	46.2	73.3	81.9	75.5	14.3
FoveaSDet-SEResNeXt-I101-32×4d-CIOU	71.6	96.9	81.6	47.0	73.5	81.6	75.9	14.7



(a) IOU=0.50 PR curve



(b) IOU=0.75 PR curve

图 5 两种阈值 IOU 下 PR 曲线比较

Fig.5 Comparison of PR curves under two thresholds IOU

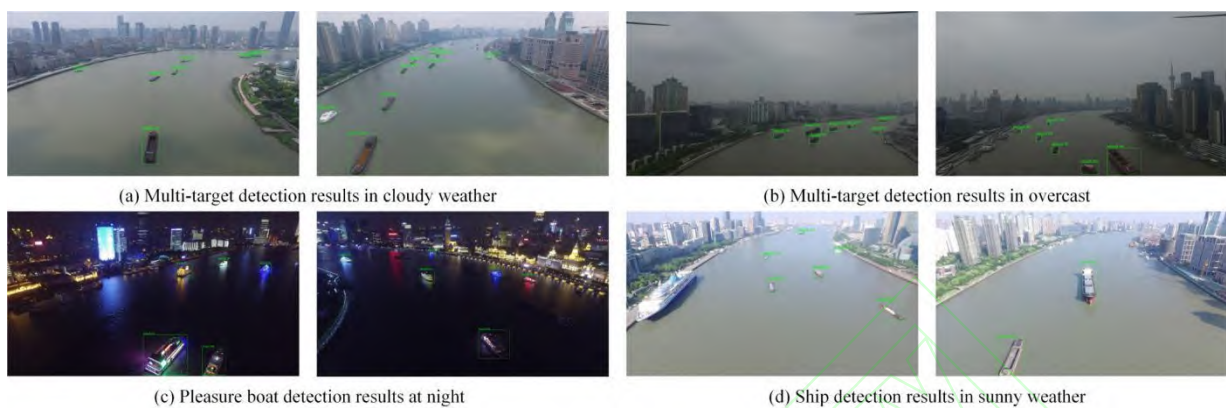


图 6 FoveaSDet 航道船舶目标检测样例

Fig.6 Samples of FoveaSDet channel target detection

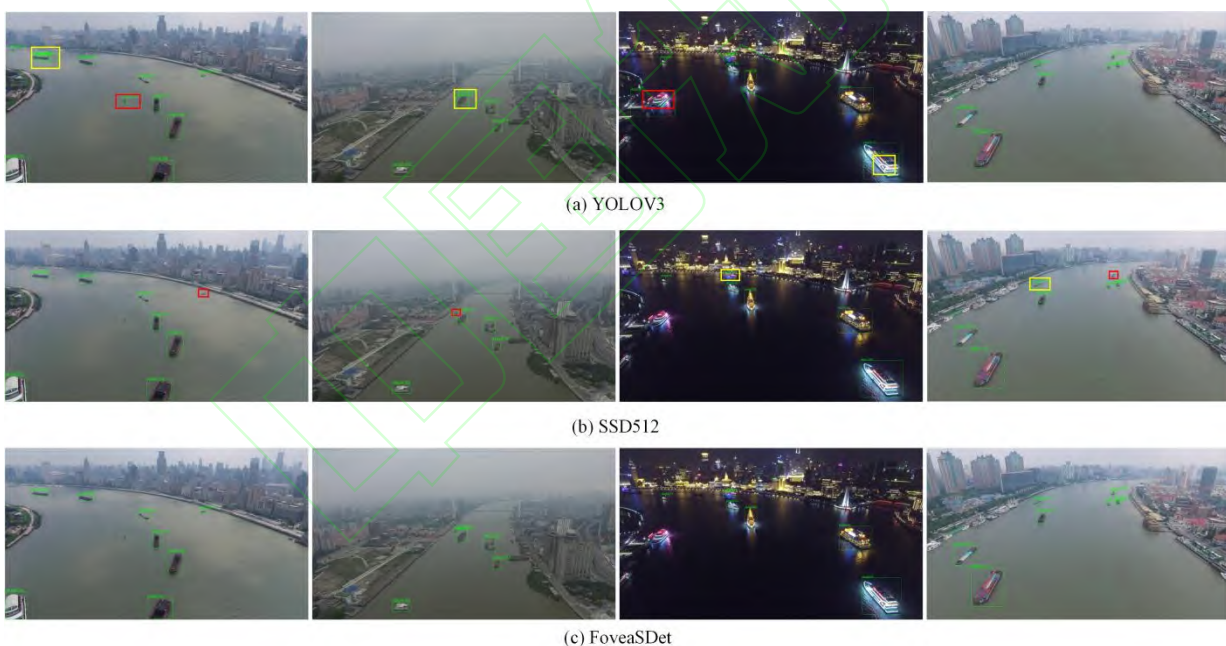


图 7 三种算法对不同环境场景检测结果比较

Fig.7 Comparison of prediction results of three algorithms for different environmental scenarios

4 结束语

本文提出了一种单阶段的无锚框目标检测算法 FoveaSDet，旨在解决复杂背景下船舶目标小、尺度变换大、难以检测定位的问题，提高对航拍航道船舶小目标的检测精度。对于船舶目标较小的特

点，改善了特征提取网络的结构，并在残差块内增加 SE 模块实现对于小目标提升；针对目标尺度变换较大的问题，使用 Foveahead 无锚框检测模块实现多尺度检测；针对检测框难以定位的问题，使用了 CIOU 损失函数予以改善，使检测框的定位精度更高，提升算法的整体性能。经过实验证实，

FoveaSDet 算法能够很好的实现航道船舶检测功能, 算法的鲁棒性和泛化能力较强, 可以推广到更广泛的应用场景。下一阶段的任务为使 FoveaSDet 算法更为轻量化, 需要提高其检测速率, 使其能部署在边缘计算设备上, 更好的实现落地应用。

参考文献:

- [1] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. New York: IEEE, 2005, 8588935.
- [2] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
- [3] HERBERT B, TINNE T, LUC J. V G. SURF: Speeded Up Robust Features[C]//European Conference on Computer Vision, 2006.
- [4] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[C]//IEEE conference on Computer Vision and Pattern Recognition, Utah: arXiv preprint, 2018: 1804.0276
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]//European Conference on Computer Vision, 2016, 6:21-27.
- [6] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017(99): 2999-3007.
- [7] Kong T, Sun F, Liu H, et al. FoveaBox: Beyond Anchor-based Object Detection[J]. IEEE Transactions on Image Processing, 2020, PP(99):1-1.
- [8] Tian Z, Shen C, Chen H, et al. FCOS: Fully Convolutional One-Stage Object Detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2020.
- [9] 裴伟, 许晏铭, 朱永英, 等. 改进的 SSD 航拍目标检测方法[J]. 软件学报, 2019, 30(3): 738-758.
- PEI Wei, XU Yanming, ZHU Yongying, et al. The target detection method of aerial photography images with improved SSD[J]. Journal of Software, 2019, 30(3): 738-758.
- [10] 史文旭, 谭代伦, 鲍胜利. 特征增强 SSD 算法及其在遥感目标检测中的应用[J]. 光子学报, 2020, 49(1): 154-163.
- SHI Wenxu, TAN Dailun, BAO Shengli. Feature enhancement SSD Algorithm and Its Application in Remote Sensing Image Target Detection[J]. Acta Photonica Sinica, 2020, 49 (1): 154-163.
- [11] 谢学立, 李传祥, 杨小冈, 等. 基于动态感受野的航拍图像目标检测算法[J]. 光学学报, 2020, 40(4): 0415001.
- XIE Xueli, LI Chuanxiang, YANG Xiaogang, et al. Dynamic Receptive Field-Based Object Detection in Aerial Imaging[J]. Acta Optica Sinica, 2020, 40(4): 0415001
- [12] 李晓光, 付陈平, 李晓莉, 等. 面向多尺度目标检测的改进 Faster R-CNN 算法[J]. 计算机辅助设计与图形学学报, 2019, 31(7): 1095-1101.
- LI Xiaoguang, FU Chenping, LI Xiaoli, et al. Improved Faster R-CNN for Multi-Scale Object Detection[J]. Journal of Computer-Aided Design and Computer Graphics, 2019, 31 (7): 1095-1101.
- [13] REN S, HE K, 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.
- [14] GUI Y, LI X, XUE L, et al. A scale transfer convolution network for small ship detection in SAR images[C]//2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 2019: 1845-1849. 2019.8785805.
- [15] 王健林, 吕晓琪, 张明, 等. 基于改进 R-FCN 的遥感图像舰船检测[J]. 激光与光电子学进展, 2019, 56(16): 222-228.
- Wang Jianlin, Lü Xiaoqi, Zhang Ming, et al. Remote sensing image ship detection based on improved R-FCN [J]. Laser and Optoelectronics Progress, 2019, 56(16): 222-228

- [16] 于野,艾华,贺小军,等.A-FPN 算法及其在遥感图像船舶检测中的应用[J].遥感学报,2020,24(2):107-115.
- Yu Ye, Ai Hua, He Xiaojun, et al.Attention-based feature pyramid networks for shipdetection of optical remote sensing image[J]. Journal of Remote Sensing, 2020, 24(2):107-115.
- [17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 770-778.
- [18] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]// Computer Vision and Pattern Recognition, 2017: 5987-5995.
- [19] Zheng Z, Wang P, Liu W, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression[C]// AAAI Conference on Artificial Intelligence. 2020.
- [20] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2017:2117-2125.
- [21] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS: Improving object detection with one line of code[C]// 2017 IEEE International Conference on Computer Vision. ICCV.2017:593.
- [22] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International Conference on International Conference on Machine Learning, 2015:423-434.
- [23] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018:7132-7141
- [24] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society: Piscataway, NJ. 2019: 658-666.
- [25] PEREZ L, WANG J. The effectiveness of data augmentation in image classification using deep learning[J]. arXiv preprint arXiv: 1712. 04621, 2017.
- [26] SCHMIDT M, FUNG G, ROSALES R. Fast Optimization Methods for L1 Regularization: A Comparative Study and Two New Approaches[J]. Lnai, 2007, 4701:286-297.