

船舶目标重叠下马赛克图像数据增强方法研究

曾广淼¹, 俞万能^{1,2}, 王荣杰^{1,2†}, 林安辉^{1,2}

(1. 集美大学 轮机工程学院, 福建 厦门 361021; 2. 福建省船舶与海洋工程重点实验室, 福建 厦门 361021)

摘要: 目标识别中的重叠遮挡问题一直以来是研究的难点, 船舶目标在狭窄水域发生相互遮挡的情况依然存在. 本文提出了一种改进的马赛克数据增强方法, 将训练图片拼接变换成3种不同的尺度, 并按照不同比例输入网络进行训练, 强化了检测算法对局部特征的学习能力, 在保持测试速度不变的情况下, 提高了对重叠目标的识别准确率, 降低了不同分辨率下识别能力的衰弱速度, 加强了算法的鲁棒性. 基于小型移动测试平台的实测实验证明, 相对于原始算法, 经过改进后的算法在重叠目标的识别准确率上提高了2.5%, 目标丢失时间减少了17%, 在不同视频分辨率下的识别稳定性上提高了27.01%.

关键词: 船舶识别; 目标重叠; 图像数据增强; Yolov4算法; 深度学习

引用格式: 曾广淼, 俞万能, 王荣杰, 等. 船舶目标重叠下马赛克图像数据增强方法研究. 控制理论与应用, 2022, 39(6): 1139 – 1148

DOI: 10.7641/CTA.2021.10329

Research on mosaic image data enhancement and detection method for overlapping ship targets

ZENG Guang-miao¹, YU Wan-neng^{1,2}, WANG Rong-jie^{1,2†}, LIN An-hui^{1,2}

(1. School of Marine Engineering, Jimei University, Xiamen Fujian 361021, China;

2. Fujian Provincial Key Laboratory of Naval Architecture and Ocean Engineering, Xiamen Fujian 361021, China)

Abstract: The problem of overlapping occlusion in target recognition has been a difficult research problem, and the situation of mutual occlusion of ship targets in narrow waters still exists. In this paper, we proposed an improved mosaic data enhancement method, which transforms the training image stitching into three different scales and inputs them into the network for training at different scales to strengthen the learning ability of the detection algorithm for local features. While keeping the test speed constant, the recognition accuracy of overlapping targets is improved, the rate of decay of recognition ability at different resolutions is reduced, and the robustness of the algorithm is enhanced. The experiments based on a small mobile testbed proved that, compared to the original algorithm, the improved algorithm improved the recognition accuracy of overlapping targets by 2.5%, reduced the target loss time by 17%, and improved the recognition stability at different video resolutions by 27.01%.

Key words: ship recognition; target overlap; image data enhancement; Yolov4 algorithm; deep learning

Citation: ZENG Guangmiao, YU Wanneng, WANG Rongjie, et al. Research on mosaic image data enhancement and detection method for overlapping ship targets. *Control Theory & Applications*, 2022, 39(6): 1139 – 1148

1 引言

目前随着计算机技术的迅速发展, 图像识别技术不只停留在实验室中, 在社会上的各个领域均得到了广泛的应用, 而亟待解决的海面目标识别问题就是其中之一, 作为船舶智能航行技术发展中的重要一环.

近年来, 对于海面目标的识别方法有很多, 大部分是通过合成孔径雷达得到的海面遥感图像, Guo等人对CenterNet的网络结构进行了改进, 提升了船舶小目标的识别能力^[1], Li等人提出了新的双分支回归网络, 提升了船舶的定位能力^[2], Fu等人利用特征平衡与细化

收稿日期: 2021-04-21; 录用日期: 2021-10-27.

†通信作者. E-mail: roger811207@163.com; Tel.: +86 13859925862.

本文责任编辑: 潘泉.

国家自然科学基金项目(51879118, 52171308), 福建省自然科学基金项目(2020J01688), 福建省科技重点项目(2021H0021), 福建省科技拥军项目(B19101), 交通运输行业高层次技术人才培养项目(2019-014), 集美大学青年拔尖人才项目(ZR2019006)资助.

Supported by the National Natural Science Foundation of China (51879118, 52171308), the National Natural Science Foundation of Fujian Province (2020J01688), the Key Projects of Science and Technology in Fujian Province (2021H0021), the Science and Technology Support Project of Fujian Province (B19101), the Transportation Industry High-Level Technical Talent Training Project (2019-014) and the Young Talent of Jimei University (ZR2019006).

网络的无锚框方法,提升了复杂场景的船舶检测能力^[3].上述文章都是以垂直视角对海面船舶目标进行识别,且需要借助飞机、卫星等航空器进行拍摄,而搭载在船舶上以水平视角对周围海面目标进行识别的相关研究较少.

不仅如此,大部分研究都对小目标的识别问题进行了讨论,这考虑到了宽阔的水域中的视野相对较远的情况,但是,船舶还存在处于狭窄水域航行的情况,此时周围船舶较多,在观测视野中极易出现相互遮挡的现象,因此,快速准确地将不同重叠程度目标进行识别定位的问题就成为重点之一.对于遮挡目标的识别问题,国内外已有一定的研究成果,Wan等人通过生成对抗网络将局部面部特征进行恢复,从而实现遮挡情况下的人脸识别^[4],Chowdhury等人将渐进式扩展算法与图注意力网络相结合,提高街道拥挤状态下的车牌识别能力^[5],Liu等人利用耦合网络来提升小目标或被遮挡行人的识别准确率^[6].他们通过对神经网络进行优化,提升了对遮挡目标的识别能力,由于对网络结构进行了加强,或多或少地增加了计算复杂度,降低了目标识别的检测速度.

现阶段,目标检测方法主要分为两种,第1种为以R-CNN系列^[7-10]为代表的二阶段检测方法,其首先利用区域提议网络提取感兴趣区域,然后基于感兴趣区域的大小,从上到下,从左到右遍历测试图片进行识别与检测.第2种为以SSD^[11]、YOLO系列^[12-15]为代表的一阶段检测算法,其不依赖区域提议网络,而直接利用锚框,将测试图片分割成的多个方格区域,对每个区域分别进行识别与检测.相比之下,二阶段方法更准确但会产生一定量的冗余计算.一阶段方法在牺牲一部分准确率的情况下,大大提高了检测速度,有利于满足实时监测的需要.

Yolov4算法^[15]将作为实验的基本模型,虽然其提出的时间较短,但是已经在农业^[16]、建筑业^[17]、医学^[18]等领域得到了一定的应用.上述文献中的实验都是利用部署于陆上服务器进行训练与测试的,情况不同是,海上的通信信号相对不稳定,且由于海况复杂,需尽量减小识别系统的体积与能耗,因此需要考虑到检测系统处于离线使用的情况.因此在本文中选择了Yolov4-tiny轻量级算法作为主要的实验模型.

基于以上分析,本文提出了一种改进的马赛克(Mosaic)数据增强方法,主要贡献可以总结如下:

1) 对训练数据集中的图像进行数据增强,在不改变网络结构的情况下,加强目标识别算法对船舶重叠目标的识别能力.

2) 针对处于离线状态下的小型移动平台,选择合适的轻量级算法,在保持其识别速度的情况下提升识别准确率,降低不同分辨率的输入图像对算法识别性能的影响.

3) 通过在船舶数据集不同的海域进行实时检测测试,结果表明,经过改进Mosaic方法训练后的算法在船舶重叠目标的检测当中准确率更高,证明了改进方法的有效性和鲁棒性.

本文的剩余部分安排如下:第2节详细介绍了改进Mosaic方法与Yolov4-tiny算法的主要结构;第3节描述了仿真实验比较及实测实验结果;最后在第4节中得出结论.

2 方法论

Yolo^[12]算法自2016年首次提出以来受到了广泛的关注,而随后提出的Yolov2^[13]和Yolov3^[14]算法在其基础上对模型结构、数据预处理方式、损失函数计算方法进行了一系列的改进,极大提升了目标检测的速度与准确率.而Yolov4算法在Yolov3算法的计算方式上,增添了很多优化技巧,在识别速度不变的情况下提升了识别准确率.

2.1 Yolov4-tiny算法的概述

Yolov4-tiny网络是基于Yolov4网络进行简化,在牺牲一部分识别准确率的情况下,减少了10倍的参数量.从Yolov4网络中约6000万的参数降至Yolov4-tiny网络中约600万的参数.其网络结构图如图1所示.

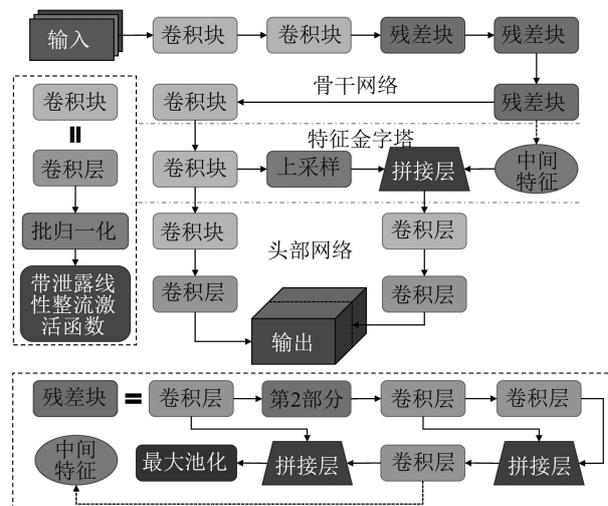


图1 Yolov4-tiny网络的特征结构

Fig. 1 Characteristic structure of Yolov4-tiny network

其中,骨干网络中的卷积块由卷积层、批归一化层^[19]、带泄露线性整流^[20]激活函数组成.而残差块是融合了跨阶段局部网络^[21]思想(CSPNet)的CSPDarknet53-Tiny,其结构如图2所示.

它由稠密层和连接层组成,首先对上一卷积层的输出 x_{U-1} 进行卷积操作,生成新的卷积层,将其输出 $x_0 = [x_{0'}, x_{0''}]$ 分成前后两个部分 $x_{0'}$ 和 $x_{0''}$ 进行前项计算,在Yolov4-tiny的网络结构中,先取第2部分进行前项传播,再将第1部分与第2部分一起直接连接到阶段的末尾,跳过稠密层,等到第2部分的 $x_{0''}$ 完成前项计算之后在过渡层中与 x_0 进行特征拼接,得到输出

x_T , 其经过最大池化后产生残差块的输出 x_U . 残差块前向计算与反向传播的过程如式(1)–(2)所示:

$$\begin{cases} x_T = w_T \cdot [x_{0''}, x_1], \\ x_U = w_U \cdot [x_0, x_1], \end{cases} \quad (1)$$

$$\begin{cases} \omega'_T = f_T(\omega_T \cdot \{g''_0, g_1\}), \\ \omega'_U = f_U(\omega_U \cdot \{g_0, g_T\}), \end{cases} \quad (2)$$

其中: ω_i 和 ω'_i 是前向计算与反向传播时的权重, f_i 表示权重更新的函数, g_i 表示传播到第*i*层的梯度, *i*等

于*T*或*U*, 分别代表着连接层或残差块的输出. 所以, 利用跨阶段局部网络的结构在反向传播时, 可以将不同通道上的梯度分别进行积分, 比如, 梯度信息经过稠密层时, 只会改变 $x_{0''}$ 通道上的权重却不会影响到 x_0 . 从而在保留不同深度的特征值的同时, 减少了过多的重复梯度信息, 在不影响网络特征提取效果的情况下, 减小了内存的开销, 提高了网络的运算速度.

在骨干网络之后, 利用特征金字塔结构^[22]对网络特征进行优化, Yolov4–tiny网络中小型特征金字塔的实现方式如图3所示.

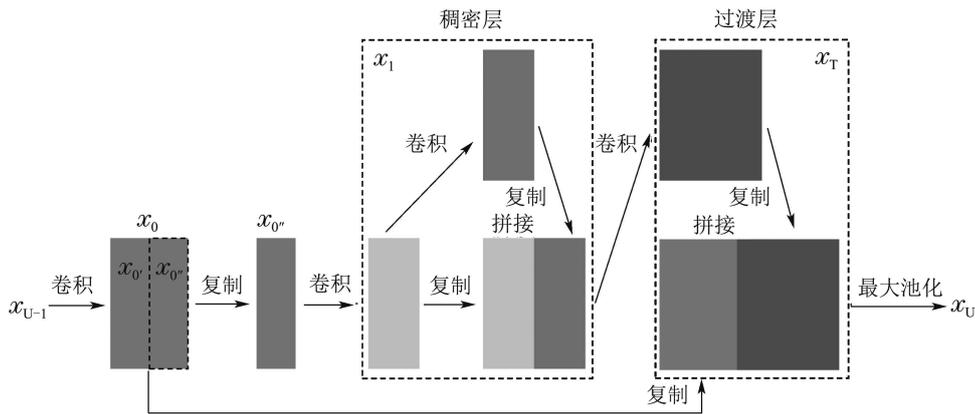


图 2 CSPDarknet53-Tiny网络中残差块的结构示意图

Fig. 2 Structure of residual blocks in CSPDarknet53-Tiny networks

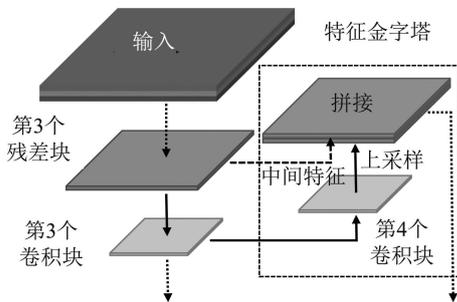


图 3 小型特征金字塔的结构示意图

Fig. 3 Structure of a small feature pyramid

中间特征是骨干网络的第3个残差块中第4个卷积层的输出, 与网络中第4个卷积块的输出经过上采样后的特征进行拼接. 在图1可以看出, 骨干网络只包含前3个CBL层, 因此图3中进行拼接操作生成的输出代表了浅层网络特征和深层网络特征的融合. 由于经过了多层卷积自上而下的特征提取, 深层网络会保留绝大部分大目标的特征值, 小目标的特征值被保存下来的很少甚至为零. 因此采用特征金字塔结构提取多个不同层级网络的特征, 通过上采样放大之后, 将其自下而上的拼接在一起, 实现了多个层级的特征融合, 提高了网络对多种分辨率下不同大小目标的识别能力.

之后将小型特征金字塔的两部分输出接入头部网

络进行计算, 生成两组包含不同感受野的图像, 分别对自身包含的先验框进行调整, 利用非极大值抑制的方法, 针对原图中大小不同的目标进行识别与检测, 提高神经网络对多尺度目标的整体检测能力.

2.2 数据增强方法

Mosaic方法是在切割混合^[23]方法上进行拓展, 生成的一个新的数据增强算法, 不同于切割混合方法的两张图片覆盖融合, 它是利用4张图片进行剪裁拼接, 形成一张新的图片. 这种方法可以更好地丰富目标物的背景, 防止由于训练集背景相似而导致的网络泛化能力降低.

Yolov4–tiny算法的输出图像中包含有两种不同的感受野, 而Yolov4算法的输出图像中则有3种不同的感受野, 因此Yolov4–tiny算法对于多尺度目标的识别能力就会相对弱一些. 所以, 对数据增强方法进行改进, 增强网络的泛化能力就显得尤为重要.

改进的马赛克数据增强方法如图4所示. 原始Mosaic方法是以图4中的上方与中间两条通道进行的特征增强, 而改进Mosaic方法在原先的基础上增加了下方的通道, 采用3条通道进行特征增强. 第3条下方通道的输出, 相较于上面两种, 是对每行每列排列的图像数进行了增加而得到的. 为了方便说明, 将以 3×3 的规格新生成九合一的图片称之为*m9*, 将以 2×2 的规格生成的四合一图片称之为*m4*, 将不经过合并

以 1×1 规格生成的图片称之为 $m1, m1, m4$ 和 $m9$ 的比例为 $o : p : q$. 这样的组合方式, 一定程度上使得训练数据集的尺度变化特性更加多样, 从而进一步丰富了数据集, 增加了背景的复杂程度. 因此, 网络可以在复杂背景的干扰下, 更专注于对目标物特征的提取, 增强了网络的鲁棒性.

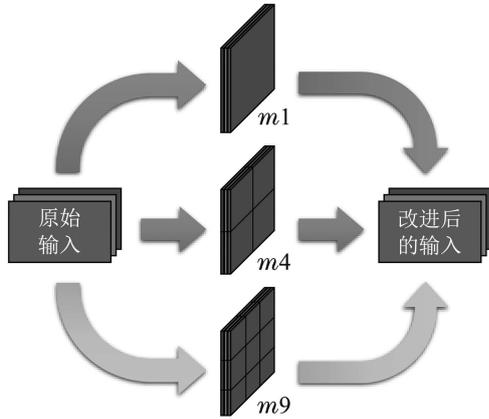


图4 改进的马赛克数据增强方法

Fig. 4 The improved Mosaic data enhancement method

九合一图像 $m9$ 的生成方式如图5所示, 主要分为A, B, C三个阶段. 在A阶段, 以输入图像的宽和高(W, H)作为边界值, 首先对图像进行缩放, X 轴和 Y 轴的缩放倍率为 t_x 和 t_y , 如式(3)–(4)所示:

$$t_x = f_{\text{rand}}(t_w, t_w + \Delta t_w), \quad (3)$$

$$t_y = f_{\text{rand}}(t_h, t_h + \Delta t_h), \quad (4)$$

其中: t_w 和 t_h 分别为宽和高缩放倍率的最小值, Δt_w 和 Δt_h 分别为宽和高缩放倍率随机区间的长度, 均为超参数. $f_{\text{rand}}()$ 表示随机值函数.

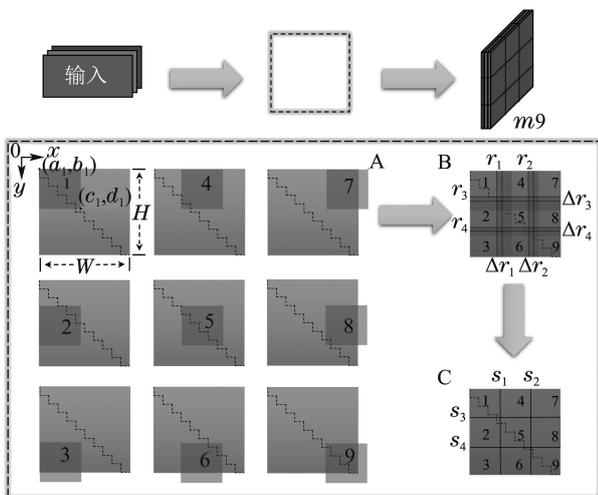


图5 $m9$ 图像的生成流程图

Fig. 5 Flowchart of $m9$ image generation

图像缩放后的左上角和右下角的坐标为 $[(a_i, b_i),$

$(c_i, d_i)]$, 由式(5)到式(8)可得

$$a_i = \begin{cases} 0, & i = 1, 2, 3, \\ W \cdot r_1, & i = 4, 5, 6, \\ W \cdot r_2, & i = 7, 8, 9, \end{cases} \quad (5)$$

$$b_i = \begin{cases} 0, & i = 1, 4, 7, \\ H \cdot r_3, & i = 2, 5, 8, \\ H \cdot r_4, & i = 3, 6, 9, \end{cases} \quad (6)$$

$$c_i = a_i + W \cdot t_w, \quad (7)$$

$$d_i = b_i + H \cdot t_h, \quad (8)$$

其中: r_1 和 r_2 分别为 X 轴的除0点以外的两组图像左上角坐标点与0点的距离占总宽度的比值, r_3 和 r_4 分别为 Y 轴的除0点以外的两组图像左上角坐标点与0点的距离占总高度的比值, 同样也均为超参数. 而灰色区域中的黑色短划线为比例尺, 每一小段代表宽或高的十分之一, 利用比例尺可以看出, 第2张至第9张的图像和第1张的缩放比例一致, 宽和高都是原来的 t_w 和 t_h 倍.

在B阶段, 需要将上一阶段裁剪好的9张图片进行拼接, 并裁减掉溢出边界框的部分, 可以看到合并后的图像存在一定程度的重叠, 因此需要对每个小区域进行划分. 从A阶段的示意图中可知, 当缩放后的图片按照坐标在指定位置进行放置时, 会存在溢出边框的情况. 此时需对溢出的部分进行裁剪, 如式(9)–(10)所示:

$$c'_i = \begin{cases} c_i, & c_1 < W, \\ W, & c_1 \geq W, \end{cases} \quad (9)$$

$$d'_i = \begin{cases} d_i, & d_1 < H, \\ H, & d_1 \geq H. \end{cases} \quad (10)$$

在边缘裁剪之后, 利用8条两两平行的圆点线围成的4个方形区域, 将其作为分割线的随机区间. 其中, $r_i = (r_1, r_2, r_3, r_4)$ 的值等于分割线坐标与0点的距离占边界长度的比值, Δr_i 为分割线随机区间的长度.

在C阶段, 将对内部重叠部分进行第2次裁剪, 其分割线坐标 s_i 可由式(11)得

$$s_i = f_{\text{rand}}(r_i, r_i + \Delta r_i), \quad i = 1, 2, 3, 4. \quad (11)$$

经过裁剪以后, 得到了拼接完成的 $m9$ 图像. 由于原图在缩放拼接过程中存在部分缺失, 对于处在原图边缘的目标有可能在操作过程中被截取部分或者完全截掉, 因此, 同样需要对这些目标对应的真实框进行裁剪甚至剔除, 以满足目标检测的需要.

生成 $m4$ 图像的方法与生成 $m9$ 类似, 如图6所示. 左上角坐标 (a_i^*, b_i^*) 与分割线坐标 s_i^* 由式(12)–(14)所

示:

$$a_i^* = \begin{cases} 0, & i = 1, 2, \\ W \cdot r_1^*, & i = 3, 4, \end{cases} \quad (12)$$

$$b_i^* = \begin{cases} 0, & i = 1, 3, \\ H \cdot r_2^*, & i = 2, 4, \end{cases} \quad (13)$$

$$s_i^* = f_{\text{rand}}(r_i^*, r_i^* + \Delta r_i^*), \quad i = 1, 2, \quad (14)$$

其中以*作为 m_4 图像与 m_9 图像的区别符号. 缩放倍率 t_x^* , t_y^* , Δt_w^* 和 Δt_h^* 及右下角坐标 (c_i^*, d_i^*) 与 m_9 图像的计算方式一致, 由于没有出现边框外溢出, 所以只需要对内部重合部分进行分割裁剪.

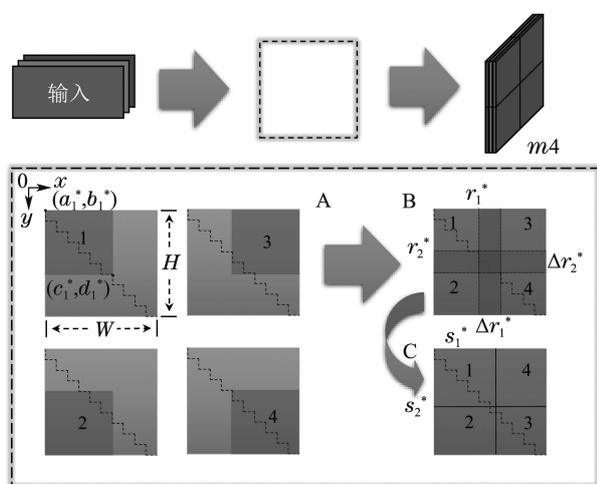


图6 m_4 图像的生成流程图

Fig. 6 Flowchart of m_4 image generation

m_1 图像由于不存在多张图片拼接的情况, 因此进入网络之前只需要通过翻转、色域变化等常规方法进行特征增强.

2.3 网络优化方法

为了更好地结合船舶数据集的特点, Yolov4-tiny算法在开始训练之前, 先利用 K 均值聚类算法将训练集内大小不一的真实框分成 m 类, 这样得到的锚框更适用于检测船舶目标. 在本文中 $m = 6$, 这些锚框将根据大小被分成2组, 每组3个框, 去检测不同尺度的目标物体. 此外, 在训练开始前, 先将网络模型在大型数据集中做训练, 当其具备提取基础特征及抽象特征的能力时, 再利用迁移学习方法进行微调, 将训练完的权重和偏差作为初始值迁移至新的任务中进行训练.

3 实验仿真及测试

3.1 数据集规划

目前广泛使用的数据集, 如: VOC数据集^[25]、COCO数据集^[26]等包含的船舶只被分为了1类, 存在图片尺寸随机且分辨率较低(不超过 640×360)的

点. 相比之下, 本文中用于训练及测试的船舶数据集^[27]一共包含7000张图片, 其分辨率为 1920×1080 , 它们截取自监控摄像机所拍摄的视频片段, 而这些摄像机是属于沿海岸线部署的海面监控系统, 其包括50个不同位置的156个摄像机. 数据集中包含6种不同类型的船舶, 其数目与类别如表1所示.

表1 船舶数据集目标物类别与数量

Table 1 Ship data set target object category and number

| 船舶类别 | 数量/个 |
|------|------|
| 矿砂船 | 2084 |
| 散货船 | 1811 |
| 杂货船 | 1426 |
| 集装箱船 | 898 |
| 渔船 | 1539 |
| 客船 | 455 |

这些图像中的船舶拥有不同的光照条件、观测角度、远近尺度以及重叠程度, 使数据集的复杂程度大大提升, 增加了目标检测算法识别的难度.

3.2 训练与测试结果

本文的算法是在开源的神经网络框架Pytorch (3.8.5)上进行实现的. 计算工作站配置包含1个GPU (GeForce RTX 3090), CPU (AMD Ryzen 9 3950x 16 Core/3.5 GHz/72 M), 以及128 G RAM. 小型移动测试平台是基于NVIDIA Jetson Xavier NX开发板进行搭建, 由1个1080p摄像头模组, 供电模组, 显示输出模组及控制模组组成.

在训练开始之前, 先将数据集进行分类. 选择6000张几乎无遮挡的船舶目标图片作为训练集, 1000张发生不同严重程度重叠的船舶图片作为测试集. 通过对目标识别算法的优化, 使目标船舶在发生不同程度的重叠与遮挡时, 能够更加快速且准确地捕捉到目标, 减少提高识别准确率. 实验中Mosaic方法的参数如表2所示, 网络优化方法中的参数如表3所示.

表2 马赛克数据增强方法中的实验参数

Table 2 Experimental parameters in Mosaic data enhancement methods

| 参数 | 数值 | 参数 | 数值 | 参数 | 数值 |
|--------------|------|--------------|------|----------------|-----|
| W | 608 | Δr_3 | 0.05 | Δr_1^* | 0.2 |
| H | 608 | Δr_4 | 0.05 | Δr_2^* | 0.2 |
| r_1 | 0.3 | t_w | 0.4 | t_w^* | 0.4 |
| r_2 | 0.65 | t_h | 0.4 | t_h^* | 0.4 |
| r_3 | 0.3 | Δt_w | 0.05 | Δt_w^* | 0.2 |
| r_4 | 0.65 | Δt_h | 0.05 | Δt_h^* | 0.2 |
| Δr_1 | 0.05 | r_1^* | 0.4 | | |
| Δr_2 | 0.05 | r_2^* | 0.4 | | |

表3 网络优化方法中的参数

Table 3 Parameters in network optimization methods

| 参数 | 数值 |
|---------------|--------------------|
| φ | 0.1 |
| γ_0 | 1×10^{-3} |
| ε | 0.95 |

训练数据集中用于训练和验证的图片以9比1的比例随机分割, 训练开始后, 网络经过100次迭代后停止, 图7表示Yolov4-tiny算法在不同迭代次数下的识别准确率, 图中图例的数值为 $o : p : q$ 的值, 准确率的高低由各类别平均精度的平均值(mean average precision, mAP)来表示.

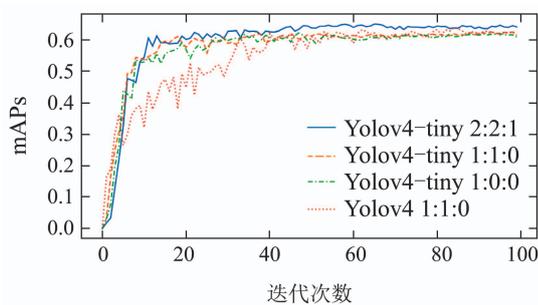


图7 Yolov4-tiny算法在不同迭代次数下的识别准确率
Fig. 7 mAPs of Yolov4-tiny algorithm with different number of iterations

从图7中可以看出, 在趋向于平稳的后二十次迭代过程中, 黄色曲线的值略高于绿色曲线的值, 使用Mosaic方法之后Yolov4-tiny算法的识别结果略有提升, 而蓝色曲线表示的改进Mosaic方法则大大提高了识别准确率, 甚至高于使用原始Mosaic方法的Yolov4算法的识别准确率. 所以改进Mosaic方法不仅提高了

Yolov4-tiny算法的识别准确率, 相对于Yolov4算法大大提高了船舶识别的检测速度. 运用 $m4$ 和 $m9$ 方法的数据增强实验过程如图8所示.

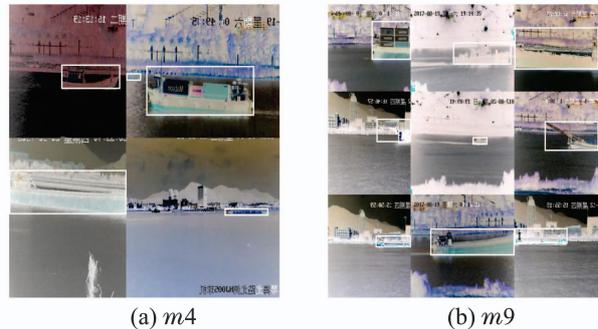


图8 $m4$ 与 $m9$ 方法的实验过程示意图

Fig. 8 Experimental procedure of $m4$ and $m9$ methods

图7中识别准确率最高的曲线是使用改进Mosaic方法进行训练的, 其使用 $m1$, $m4$ 和 $m9$ 方法作为输入的概率按照比例 $o : p : q = 2 : 2 : 1$ 来进行. 为了探究比例的不同对识别准确率的影响, 本文进行了多组实验比较, 结果如表4所示. 同时, 将改进Mosaic方法与3种同类代表性数据增强方法进行了比较, 结果如表5所示.

从表4的前5行的对比可以看出, 如果训练集中的所有图像都使用 $m4$ 或 $m9$ 方法进行数据增强, 而抛弃原始图像的话, 相较于不使用数据增强方法的情况, 其识别准确率反而会降低, 而无论是 $m4$ 或 $m9$ 方法, 结合50%的原始图像同时训练, 其准确率都会提升. 因此, 即使使用数据增强方法, 也不可以丢弃原始图像, 这一点同时也可以通过表5中的前8行数据来印证.

表4 不同比例下Mosaic数据增强方法对目标识别准确率的影响

Table 4 Effect of Mosaic data enhancement methods on target recognition accuracy at different scales

| 算法类型 | $m1, m4, m9$ 的比例 | $m1$ 占比 | $m4$ 占比 | $m9$ 占比 | 100次迭代中的最高mAP | 后20次迭代中的平均mAP | 相对于原始Mosaic (1:1:0)的提升比例 |
|-------------|------------------|---------|---------|---------|---------------|---------------|--------------------------|
| Yolov4-tiny | 1:0:0 | 100% | 0% | 0% | 62.28% | 61.54% | 99.55% |
| | 0:1:0 | 0% | 100% | 0% | 60.17% | 55.98% | 96.18% |
| | 0:0:1 | 0% | 0% | 100% | 31.34% | 30.69% | 50.10% |
| | 1:1:0 | 50% | 50% | 0% | 62.56% | 61.99% | 100.00% |
| | 1:0:1 | 50% | 0% | 50% | 60.91% | 58.24% | 97.36% |
| | 1:1:1 | 33% | 33% | 33% | 62.39% | 60.53% | 99.73% |
| | 2:2:1 | 40% | 40% | 20% | 65.06% | 64.09% | 104.00% |
| | 2:1:1 | 50% | 25% | 25% | 63.11% | 61.44% | 100.88% |
| | 3:2:1 | 50% | 33% | 17% | 63.78% | 62.49% | 101.95% |
| | 4:2:1 | 57% | 29% | 14% | 63.31% | 62.86% | 101.20% |
| | 4:3:2 | 45% | 33% | 22% | 63.80% | 62.21% | 101.98% |
| 5:3:2 | 50% | 30% | 20% | 65.01% | 64.48% | 103.92% | |
| Yolov4 | 1:1:0 | 50% | 50% | 0% | 63.90% | 62.56% | 102.14% |

表5 不同数据增强方法的目标识别准确率

Table 5 Accuracy of target recognition with different data enhancement methods

| 算法类型 | 数据增强方法 | 训练集进行数据增强的图片比例 | 100次迭代中的最高mAP | 后20次迭代中的平均mAP | |
|-------------|------------------------|------------------------------|---------------|---------------|--------|
| Yolov4-tiny | Mixup ^[28] | 100% | 36.19% | 33.04% | |
| | | 50% | 40.16% | 38.19% | |
| | Cutout ^[29] | 100% | 61.17% | 60.24% | |
| | | 50% | 61.75% | 61.31% | |
| | Cutmix ^[30] | 100% | 22.69% | 20.18% | |
| | | 50% | 34.77% | 24.19% | |
| | | Mosaic | 100% | 60.17% | 55.98% |
| | | | 50% | 62.56% | 61.99% |
| | 改进的Mosaic | 60% (40% m_4 + 20% m_9) | 65.06% | 64.09% | |

而同比例使用 m_1 , m_4 及 m_9 三种方法进行训练时(当 $o:p:q=1:1:1$ 时),其准确率较原始Mosaic方法(当 $o:p:q=1:1:0$ 时)几乎一致,并没有得到预期的提升,考虑到可能是因为按照此方法训练时原始图片的比例过小(仅有30%),所以在表4的第7至12行进行了多种比例的实验,并且将 m_1 的占比调至40%以上,准确率较原始Mosaic方法都有了不同程度的提升,这说明数据集应当以 m_1 数据作为重点,让网络学习好目标的整体特征,在此基础上利用 m_4 与 m_9 分别对局部特征进行加强学习,提高网络的泛化能力.并且可以发现,当 m_4 与 m_9 的占比只差在10%~20%且 m_1 的占比在40%~50%时,网络的训练效果会更好,即当 $o:p:q=2:2:1$ 和 $o:p:q=5:3:2$ 时,mAP的最高值相对原始Mosaic方法提升了2.5%,并且高于使用原始Mosaic方法训练的Yolov4算法.

为了进一步测试网络对重叠目标的识别能力,利用小型移动测试平台对海面进行实时检测,测试地点位于鼓浪屿(中国,厦门).以一段在行进过程中发生重叠的两艘渔船的实时视频片段作为实验内容,计算网络在各个时刻的识别能力,视频片段分辨率为1080P,时长38 s,每秒24帧,共912帧,测试结果如图9所示.实测实验中选用的权重文件为100次迭代中最高mAP所对应的权重.

在图9中可以看出,第691帧两艘船已经发生严重重叠时,使用改进Mosaic方法的Yolov4-tiny算法依然可以将两艘船同时识别出来,而使用原始Mosaic方法

和未使用Mosaic方法的Yolov4-tiny算法只能识别其中的一艘船.在第756帧两艘船正在摆脱重叠的过程中,使用改进Mosaic方法和原始Mosaic方法的Yolov4-tiny算法已经可以将两艘船同时识别出来,而未使用Mosaic方法的Yolov4-tiny算法依然只能识别其中的一艘船.这进一步说明了改进Mosaic方法对提升重叠目标检测能力的有效性.

与此同时,为了探究网络对于目标物特征识别能力的影响,其梯度加权类激活映射图(Grad-CAM)^[30]如图10所示.

在图10中,通过对比在不同训练方式下的梯度加权类激活映射图,可以直观地观察到网络对于测试图片感兴趣区域的不同.第一,对于发生重叠且不完整的船舶大目标,当 $o:p:q=2:2:1$ 时,此时网络所关注的中心点有3个,其中两个位于船舶上层建筑的位置,可以更好地对船舶的个数进行判断,而且重点关注了图片的上半区域,相比于其他两种情况,对下半区域的背景特征有了更好的抗干扰能力.另外,相较于未使用Mosaic方法训练的情况,原始Mosaic方法的确增强了网络对于船舶的关注度.第二,对于发生重叠且完整的船舶小目标,当 $o:p:q=2:2:1$ 时,网络对于不同船舶位置的关注点较为准确且边界感较强,同样可以增加对船舶个数的判断能力,而未使用Mosaic方法训练下的网络,对于小目标的识别能力明显更弱.因此,对于不同尺度的船舶目标,改进的Mosaic方法在一定程度上可以提高网络的识别能力.

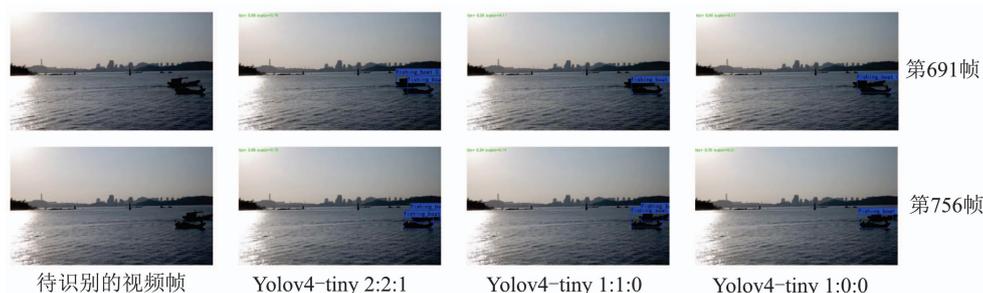


图9 重叠检测实验对比图

Fig. 9 Overlap detection experiment comparison chart

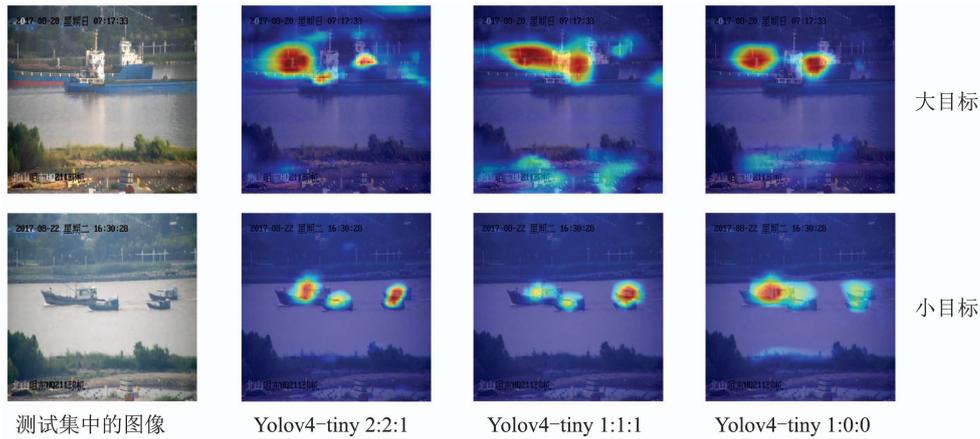


图 10 不同训练方式下梯度加权类激活映射图的对比

Fig. 10 Comparison of gradient-weighted class activation mapping maps under different training methods

由于小型移动设备的算力有限, 在1080p分辨率下识别速度较慢, 因此对不同分辨率视频中的船舶目标进行检测测试, 网络的识别速度由表6所示。

由表6可以看出, 由于Yolov4网络结构较复杂, 权重参数较多, 导致其识别时需要的算力较高, 因此在4种分辨率下的识别速度都很低, 而Yolov4-tiny算法网络结构较简单, 权重参数较少, 识别时需要的算力较低. Yolov4-tiny算法在360p视频分辨率下的识别速度为其在1080p视频分辨率下的179.50%, 为Yolov4算法在360p视频分辨率下的识别速度的462.60%。

由于不同分辨率下不同算法的识别速度不同, 因此改进Mosaic在不同分辨率下的识别能力的重要性

不言而喻. 于是, 通过对比在重叠开始至重叠结束的区间内, 以应用不同Mosaic方法的Yolov4-tiny算法将两艘船正确识别并分离的帧个数, 判断其对于重叠问题的识别能力, 具体实验结果如表7所示。

表 6 不同分辨率下的识别速度对比

Table 6 Comparison of recognition speed at different resolutions

| 网络类型 | 识别速度/fps | | | | 网络权重个数/个 |
|-------------|----------|------|-------|-------|--------------|
| | 1080p | 720p | 480p | 360p | |
| Yolov4-tiny | 6.34 | 8.68 | 10.48 | 11.38 | 5, 885, 666 |
| Yolov4 | 2.03 | 2.27 | 2.41 | 2.46 | 63, 964, 611 |

表 7 采用不同Mosaic方法在不同分辨率下的对于重叠问题的识别能力对比

Table 7 Comparison of the recognition ability of overlapping problems at different resolutions using different Mosaic methods

| 视频分辨率 | 采用的数据增强方法 | 无法分离开始的帧数 | 无法分离结束的帧数 | 无法分离的帧个数/个 | 无法分离的时间/s | 无法分离的时间相对于原始Mosaic方法(1:1:0)的比例 |
|-------|-------------|-----------|-----------|------------|-----------|--------------------------------|
| 1080p | 未采用Mosaic方法 | 654 | 804 | 151 | 6.29 | 196% |
| | 原始Mosaic方法 | 676 | 752 | 77 | 3.21 | 100% |
| | 改进Mosaic方法 | 692 | 755 | 64 | 2.67 | 83% |
| 720p | 未采用Mosaic方法 | 649 | 810 | 162 | 6.75 | 210% |
| | 原始Mosaic方法 | 673 | 752 | 80 | 3.33 | 104% |
| | 改进Mosaic方法 | 691 | 755 | 65 | 2.71 | 84% |
| 480p | 未采用Mosaic方法 | 650 | 912 | 263 | 10.96 | 342% |
| | 原始Mosaic方法 | 658 | 753 | 96 | 4.00 | 125% |
| | 改进Mosaic方法 | 692 | 755 | 64 | 2.67 | 83% |
| 360p | 未采用Mosaic方法 | 649 | 912 | 264 | 11.00 | 343% |
| | 原始Mosaic方法 | 656 | 754 | 99 | 4.13 | 129% |
| | 改进Mosaic方法 | 692 | 756 | 65 | 2.71 | 84% |

从表7中可以看出,随着分辨率的下降,各个算法对于重叠目标识别速度都有不同程度的减弱,其减弱的程度可以由无法分离的时间增长率来表示,如图11所示.由此可见,使用改进Mosaic方法的Yolov4-tiny算法在4种不同的分辨率下都取得较大优势,不仅在船舶目标产生重叠时,无法分离重叠目标的时间较短,而且在分辨率发生下降时也可以维持高识别率.

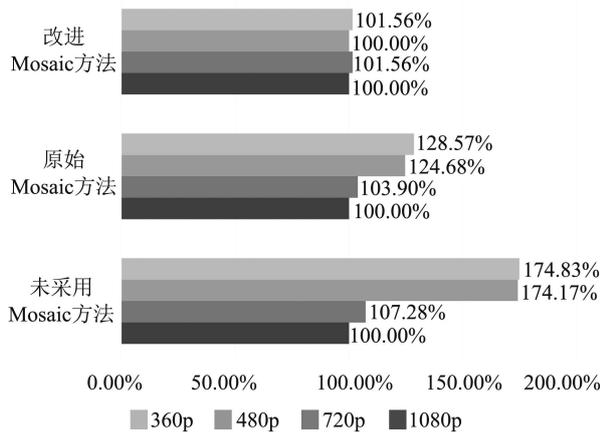


图 11 采用不同Mosaic方法在不同分辨率下无法分离的时间增长率

Fig. 11 Time growth rates that cannot be separated at different resolutions using different Mosaic methods

实验证明,改进Mosaic数据增强方法对船舶重叠目标的识别效果有一定的提升,应用于Yolov4-tiny算法可以部署于小型移动设备,可以灵活地装配在各类平台上,从而实现在离线状态下对海面船舶目标的实时监测.

4 结论

本文以Yolov4-tiny算法为实验基础,提出了一种改进的马赛克数据增强方法,并对马赛克填充的不同比例进行了对比试验,探究其对船舶重叠目标检测准确率的影响.在仿真实验中,对船舶数据集进行分析,以数据集中的船舶重叠目标作为检测对象,相对于原始算法,改进后的方法在测试数据集的识别准确率上提高了2.5%,达到了与原始Yolov4算法同样的检测效果,提升准确率的同时降低了算力的消耗.在实测实验中,将算法部署在小型移动测试平台上进行实测,相对于原始算法,改进后方法的目标丢失时间减少了17%,在不同视频分辨率下的识别稳定性上提高了27.01%.因此,目标识别算法经过改进的马赛克数据增强方法训练后,可以提升其对船舶重叠目标的识别能力.

参考文献:

[1] GUO H, XI Y, WANG N. A CenterNet++ model for ship detection in SAR images. *Pattern Recognition*, 2021, 223: 107787.

[2] LI L, ZHOU Z, WANG B. A novel CNN-based method for accurate ship detection in HR optical remote sensing images via rotated bounding box. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(1): 686 – 699.

[3] FU J, SUN X, WANG Z. An anchor-free method based on feature balancing and refinement network for multi-scale ship detection in SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(2): 1331 – 1344.

[4] WAN J, LI J, LAI Z. Robust face alignment by cascaded regression and de-occlusion. *Neural Networks*, 2020, 123: 261 – 272.

[5] CHOWDHURY P, SHIVAKUMARA P, KANCHAN S. Graph attention network for detecting license plates in crowded street scenes. *Pattern Recognition Letters*, 2020, 140: 18 – 25.

[6] LIU T, LUO W, MA L. Coupled network for robust pedestrian detection with gated multi-layer feature extraction and deformable occlusion handling. *IEEE Transactions on Image Processing*, 2021, 30: 754 – 766.

[7] GIRSHICK R, DONAHUE J, DARRELL T. Rich feature hierarchies for accurate object detection and semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014: 580 – 587.

[8] GIRSHICK R. Fast R-CNN. *IEEE International Conference on Computer Vision (ICCV)*. Santiago: IEEE, 2015: 1440 – 1448.

[9] REN S, HE K, GIRSHICK R. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137 – 1149.

[10] HE K, GKIOXARI G, DOLLÁR P. Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 386 – 397.

[11] LIU W, ANGUELOV D, ERHAN D. SSD: Single shot multi-box detector. *European Conference on Computer Vision*. Cham: Springer, 2016: 21 – 37.

[12] REDMON J, DIVVALA S, GIRSHICK R. You only look once: Unified, real-time object detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas: IEEE, 2016: 779 – 788.

[13] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu: IEEE, 2017: 6517 – 6525.

[14] REDMON J, FARHADI A. Yolov3: An incremental improvement. *arXiv Preprint*. ArXiv: 1804.02767, 2018.

[15] BOCHKOVSKIY A, WANG C, LIAO H. Yolov4: optimal speed and accuracy of object detection. *arXiv Preprint*. ArXiv:2004.10934, 2020.

[16] JIANG M, RAO Y, ZHANG J. Automatic behavior recognition of group-housed goats using deep learning. *Computers and Electronics in Agriculture*, 2020, 177: 105706.

[17] YU A, SHEN Y, SHEN C. A real-time detection approach for bridge cracks based on Yolov4-FPM. *Automation in Construction*, 2021, 122: 103514.

[18] ALBAHLI S, NIDA N, IRTAZA A. Melanoma lesion detection and segmentation using Yolov4-DarkNet and active contour. *IEEE Access*, 2020, 8: 198403 – 198414.

[19] LOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv Preprint*. ArXiv:1502.03167, 2015.

[20] XU B, WANG N, CHEN T. Empirical evaluation of rectified activations in convolutional network. *arXiv Preprint*. ArXiv:1505.00853, 2015.

[21] WANG C, LIAO M, WU Y. CSPNet: A new backbone that can enhance learning capability of CNN. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Seattle: IEEE, 2020: 1571 – 1580.

- [22] LIN T, DOLLÁR P, GIRSHICK R. Feature pyramid networks for object detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu: IEEE, 2017: 936 – 941.
- [23] YUN S, HAN D, CHUN S. CutMix: Regularization strategy to train strong classifiers with localizable features. *IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul: IEEE, 2019: 6022 – 6031.
- [24] ZHENG Z, WANG P, LIU W. Distance-IoU loss: Faster and better learning for bounding box regression. *Proceedings of the AAAI Conference on Artificial Intelligence*. New York: AAAI, 2020, 34(7): 12993 – 13000.
- [25] EVERINGHAM M, VAN G, WILLIAMS C. The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 2010, 88(2): 303 – 338.
- [26] LIN T, MAIRE M, BELONGIE S. Microsoft coco: Common objects in context. *European Conference on Computer Vision*. Cham: Springer, 2014: 740 – 755.
- [27] SHAO Z, WU W, WANG Z. Seaships: A large-scale precisely annotated dataset for ship detection. *IEEE Transactions on Multimedia*, 2018, 20(10): 2593 – 2604.
- [28] ZHANG H, CISSE M, DAUPHIN Y N. Mixup: Beyond empirical risk minimization. *arXiv Preprint*. ArXiv: 1710.09412, 2017.
- [29] DEVRIES T, TAYLO G W. Improved regularization of convolutional neural networks with cutout. *arXiv Preprint*. ArXiv:1708.04552, 2017.
- [30] YUN S, HAN D, OH S J. Cutmix: Regularization strategy to train strong classifiers with localizable features. *IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul: IEEE, 2019: 6023 – 6032.
- [31] SELVARAJU R, COGSWELL M, DAS A. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *IEEE/CVF International Conference on Computer Vision (ICCV)*. Venice: IEEE, 2017: 618 – 626.

作者简介:

曾广森 博士, 目前研究方向为智能信息处理与计算机视觉,

E-mail: gm.zeng@foxmail.com;

俞万能 教授, 博士生导师, 目前研究方向为船舶智能控制及电力

推进, E-mail: wnyu2007@jmu.edu.cn;

王荣杰 教授, 博士生导师, 目前研究方向为智能信息处理、盲源分离和电力系统故障诊断, E-mail: roger811207@163.com;

林安辉 讲师, 博士, 目前研究方向为智能信息处理及控制技术, E-mail: linanhui@jmu.edu.cn.