海洋目标识别

李扬1 江海龙1 杨睿昕1

**摘要** 海上周围信息的感知和获取是智能船舶自主安全航行的核心技术之一。对其他船舶的智能自主避让和动态路径规划实现的最重要前提是对本船舶可视范围内其他船舶的实时性检测和高精度识别，这对推动船舶的无人化和智能化具有重要的研究意义。我们采用改进的yolov5算法来进行海洋船舶目标识别。通过手动对数据集进行重新标注，优化初始anchor参数和损失函数，可以达到在较短的时间的完成较高精度的目标识别。

**关键词** 目标识别, 船舶检测, Yolov5

**Ocean target recognition**

LiYang1, JiangHailong1, YangRuixin1

**Abstract** The perception and acquisition of information around the sea is one of the core technologies for autonomous and safe navigation of smart ships. The most important prerequisite for the realization of intelligent autonomous avoidance and dynamic path planning for other ships is the real-time detection and high-precision recognition of other ships within the visible range of the ship, which has important research significance for promoting the unmanned and intelligent ships. We use an improved yolov5 algorithm to identify marine ship targets. By manually relabeling the data set and optimizing the initial anchor parameters and loss function, high-precision target recognition can be achieved in a shorter time.

**Key words** Target recognition, Ship detection, Yolov5

# 引言

随着世界经济的发展，各个国家之间的贸易量不断增加，而海洋运输业作为贸易的主要手段受到的关注度随之越来越高。船舶识别技术也得到了很大程度的发展，经历了由传统到现代、由人工识别信息到机器识别信息的转变[1]。

早期时候，雷达标绘仪(APPA)船舶定位识别技术是最有效的技术手段之一，该方法通过确定船舶所在位置以及不同船之间的间距评估发生碰撞的可能性。然而该系统存在无法识别船型、存在盲区、定位精确度较低等问题[2]。为了解决这些问题，目前，监控海上船舶主要是通过船舶自动识别系统(Automatic Identification System)。船舶自动识别技术是利用GPS导航定位技术，以自组织时分多址(SOTDMA)方式，在海上甚高频(VHF)频段(30-300MHz)传输静态数据(包括：船名、船籍、船舶长度、宽度、高度、类型等信息)、动态数据(包括：船舶航速、航向、航行状态等信息)、航行相关信息(包括：船舶吃水程度、目的地等信息)以及船舶安全信息，同时也可以自动的收到周围船舶发出的信息，并于海岸监控端进行信息交换。然而，该技术仍被指出存在以下缺陷：

1) 船舶自动识别技术设备传输的数据以及航行信息需要依靠船舶工作人员手动输入到设备中，这也降低了数据的可靠性。

2) 成本较高。尽管2011年中国海事局已正式要求200t以上船舶必须安装AIS设备，但AIS设备依然无法做到全部普及。

随着人工智能的快速发展，各类智能算法、设备在人们的日常生活中扮演着越来越重要的角色。在船舶检测研究中，国外方面，Salembier等人[3]提出了一种基于多特征和多尺度视觉显着性的舰船检测方法。该方法充分利用了可以同时在多个通道中同时操作的超复合图像的特征，节省了操作时间，并保证了不同尺度特征的特征；国内方面，王方超等[5]利用图像金字塔的方法提取相应的特征信息，然后对图片进行降维操作计算，结合 sobel 算法来实现船舶的检测，实验证明此方法具有鲁棒性，但检测速度上仍无法达到实时处理。在目标检测研究中，由 Redmon J[6]提出的一种端到端（End-to-End）的 Yolo 检测算法能够快速的从视频（图片流）中准确的标定出检测物体的类别、位置、颜色等信息。Yolo算法在目标检测领域表现出了极佳的效果。

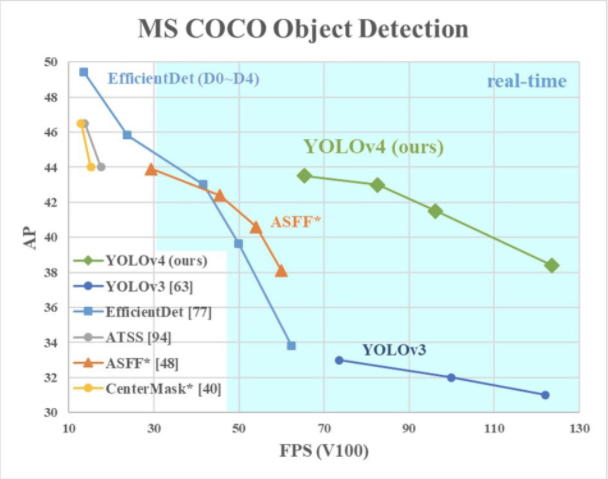


图1 Yolo性能评估

从图1中可以看出，YoloV4相较于YoloV3已取得了极大的效果上的提升[8]。而最近新公布的YoloV5算法检测速度可以达到250FPS，同时具有较高的mAP已经极大的超越了人眼的识别频率[7]。

本文将借鉴YoloV5的算法对船舶进行检测，摒弃传统的机器学习算法，证实了基于深度学习的方法能够使得检测结果拥有更高的准确度。

# 相关工作

数据集：对原有数据集进行清洗，删除模糊不清的分类元素，并对部分数据集进行重新标定和边框贴合。标注工具采用LabelImg进行标注。由于标注文件xml的框是,,,的形式，但是在yolo中是,,,需要将数据格式进行转换。对于小物体的边框进行了更加细粒度的重新标定；对于island reef这一分类进行统一划分，确保可以有更好的统一特征。

1. 网络模型：基于yolov5改进的网络模型，由于分类任务较轻，总共分为6分类。在本次比赛中不考虑实时性，只考虑精度问题。所以将网络模型加深，提高最后的精度；
2. 锚框大小的优化：由于实际的过程中，锚框的大小要随着数据集的变化而进行改变，这也是影响精度的主要问题，但本次比赛中，采用密度聚类KMeans来确定锚框的大小，并将其设置到配置文件中，但是在本次比赛中，密度聚类的正确率为64%，这也就限制了最后的最高分数，我们怀疑是数据集中同一分类的锚框并不具备统一性，尤其是island reef这一分类，导致了锚框的设置困难。

# 模型和方法描述

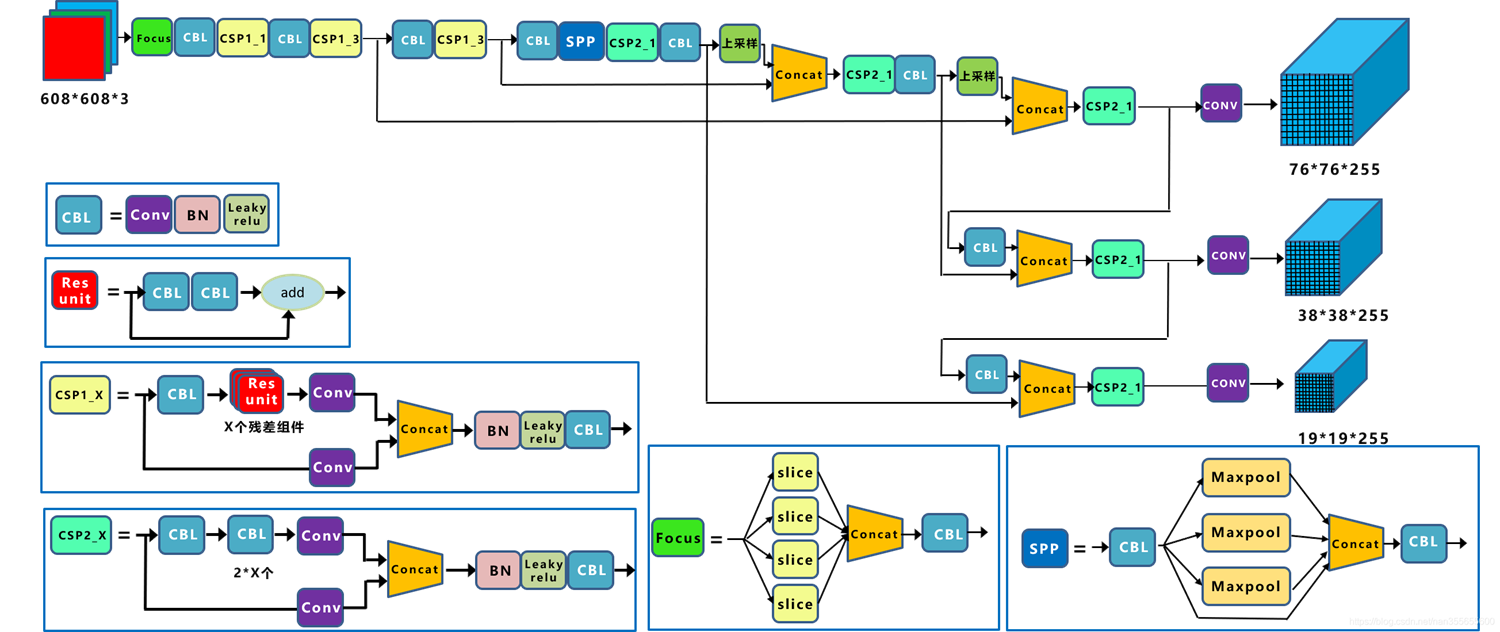


图2 YoloV5网络结构图

YoloV5网络的结构框架图如图2所示，该网络可以分为输入端、Backbone、Neck、输出端四个部分。下面将分别介绍四个组成部分。

## 输入端

1. Mosaic数据增强

Mosaic数据增强采用4张图片，进行随机缩放、随机裁剪、随机排布的方式进行拼接。该方式具有以下两个优点：

a.丰富数据集：随机使用4张图片，随机缩放，再随机分布进行拼接，大大丰富了检测数据集，特别是随机缩放增加了很多小目标，让网络的鲁棒性更好。

b.减少GPU：该方法进行增强训练时，可以直接计算4张图片的数据，使得Mini-batcch可以设置为较小的值，使用一块GPU就能达到很好的训练效果。

(2) 自适应锚框设计

在Yolo算法中，针对不同的数据集，都会有初始设定长宽的锚框。在网络训练中，网络在初始锚框的基础上输出预测框，进而和真实框groundtruth进行比对，计算两者差距，再反向更新，迭代网络参数。

(3) 自适应图片缩放

在传统的目标检测算法中，针对输入图片的大小不同的情况，常用方式是将原始图片统一缩放至同一个标准尺寸，然后将其输入至网络中进行处理。然而很多图片长宽比不同，完成缩放填充后，两边的黑边大小会不同，如果填充过多，会出现信息冗余的问题。模型中对此进行改进，大幅提升了推理速度。通过选择长宽中最小的缩放系数，再根据网络下采样次数算出图片高或宽需要填充的数值。这样使得图像填充的黑边减少，大幅缩短推理时间，进而显著提高目标检测、推理速度。

## Backbone

相对于yolov3&yolov4，在yolov5中新增了focus结构，它最主要的操作就是切片，即如下切片示意图所示，将4\*4\*3的图片切片后，可得到2\*2\*12的特征图。通过切片可以从高分辨率图像中，周期性的抽出像素点重构到低分辨率图像中，即将图像相邻的四个位置进行堆叠，聚焦宽高维度信息到c通道空间，提高每个点感受野，并减少原始信息的丢失，这样可以大幅度减少计算量并加快速度。

此外，相对于以前的yolo系列，yolov5设计了两种CSP结构，一种用于Backbone一种用于Neck中。CSP模块能将基础层的特征映射划分为两部分，然后通过跨阶段层次结构将它们合并，在减少了计算量的同时可以实现更丰富的梯度组合并保证了准确率。这样不仅能增强网络学习能力，使得在轻量化的同时保持准确性，而且还可以降低计算瓶颈与内存成本。

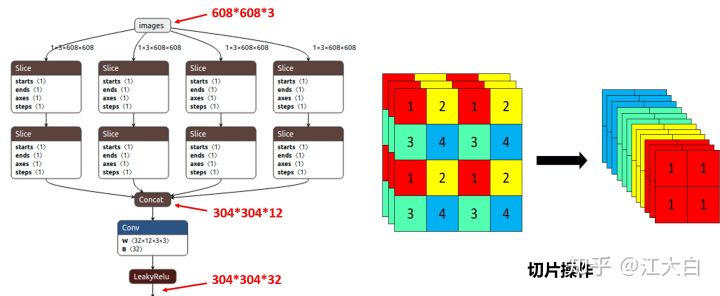


图3 切片操作

## Neck

Neck，相当于目标检测的颈部，在目标检测领域，为了更好的提取融合特征，通常在Backbone和输出层之间插入一些层，所插入的称为Neck。

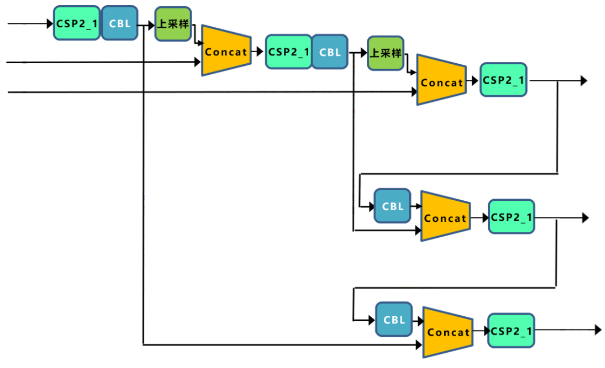


图4 YoloV5Neck层网络结构图

YoloV5的Neck采用FPN+PAN的结构。其中，PAN借鉴了刘等人提出的PANet[9]，针对底层特征流向高层特征的路线过长的问题，加入了Bottom-up的短线路；针对proposal只在一个特征层上进行预测的问题，提出了 adaptive feature pooling，进行特征融合。Yolov5的Neck结构中，采用借鉴CSPNet[10]设计的CSP2结构，降低了计算量，加强网络特征融合的能力。

## 输出端

回顾回归框损失函数的发展路线可以发现，早些年的IoU Loss不能反映两者的距离大小，IoU(A,B)=0时损失函数并不可导，而且此损失函数不能反映两个框是如何相交的；当目标框完全包裹预测框的时候，IoU=GIoU，此时的GIoU 退化为IoU，无法有效区分其相对位置关系；LDIoU虽然考虑了重叠面积和中心点距离，但没有考虑长宽比；CIoU则在DIoU的惩罚项基础上加了一个影响因子，把预测框长宽比拟合目标框的长宽比考虑进去。基于此分析，我们在模型中采用了ciou-Loss。

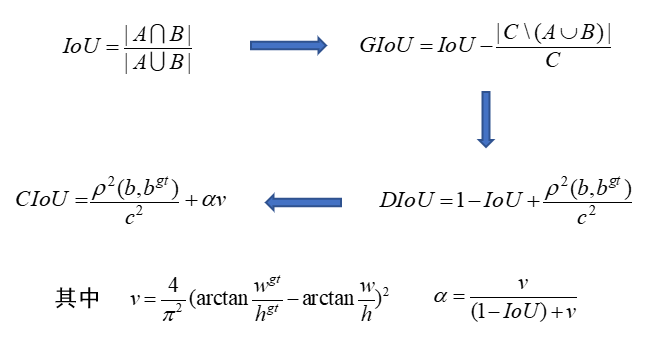


图5 回归框损失函数

# 实验结果

为了更快的进行收敛，我们将模型部署在linux系统上，使用两块2080Ti进行模型的训练。模型的box\_loss、cls\_loss和obj\_loss如图6所示。另外模型的mAP\_0.5:0.95、mAP\_0.5以及预测precision如图7所示。

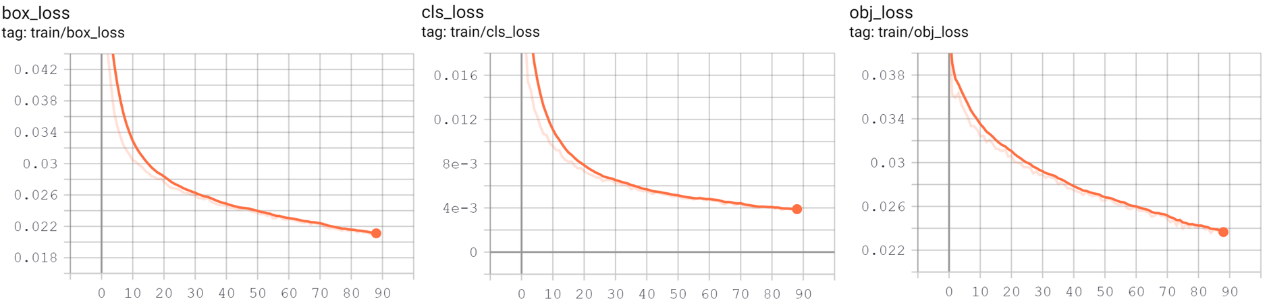


图6 相关损失

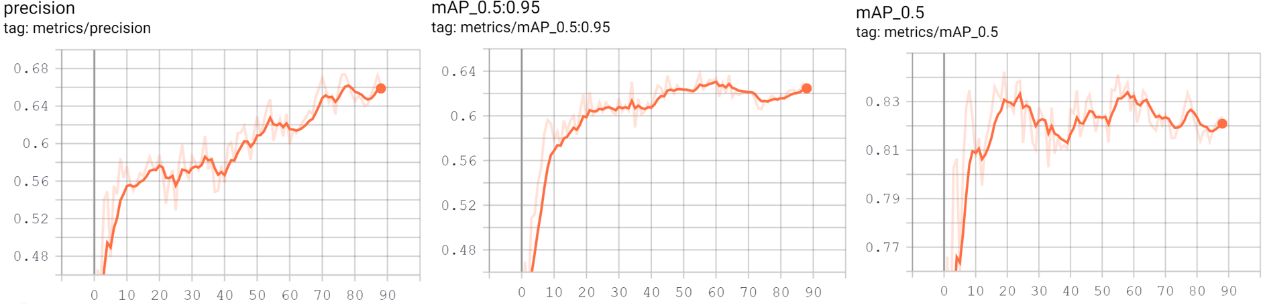


图7 置信度

将训练后模型进行测试，测试的结果如图8所示。

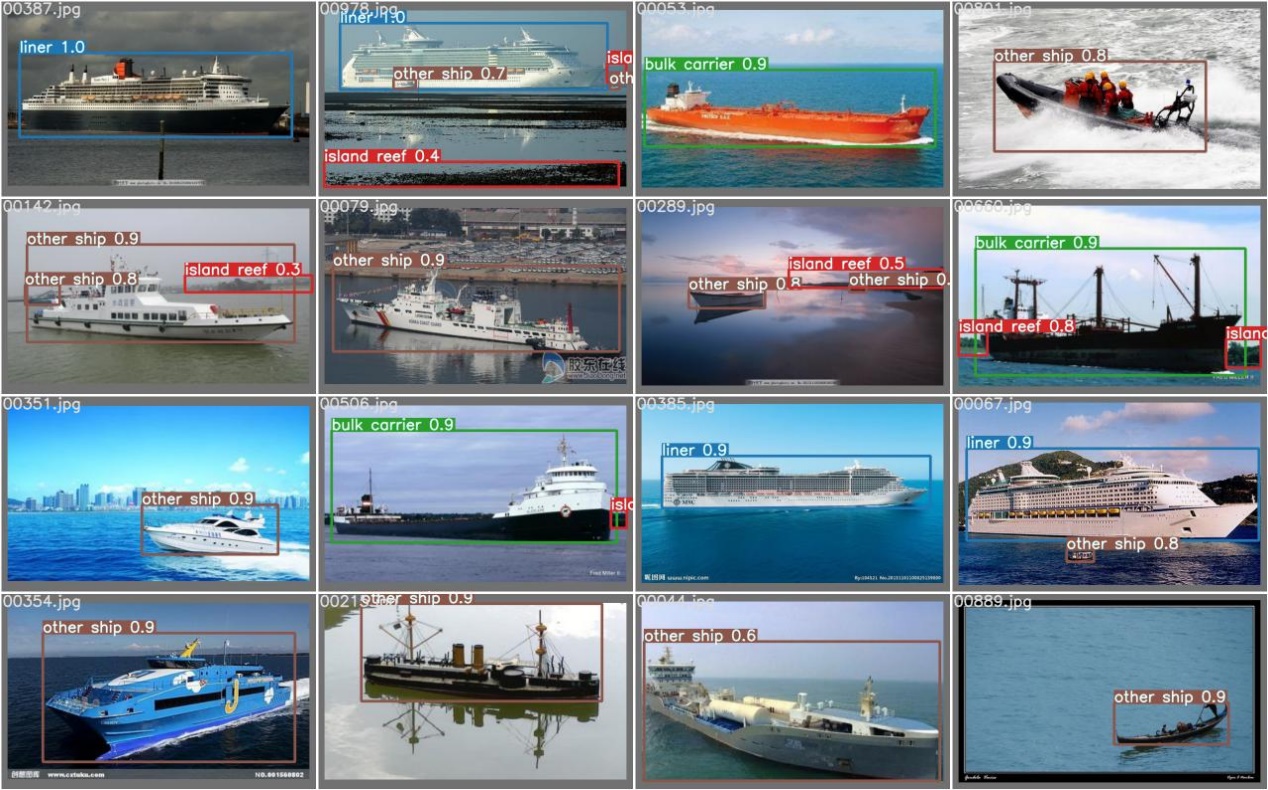


图8 测试集结果

可以看出在测试集上可以有效的结果进行识别。由于本次比赛只考虑了最终的精度，所以经过公开测试集的反馈，加深网络模型进行训练，从原来的60层加深到131层，提高最后的精度，尤其是在特征提取层面。在公开测试集上成绩由原来的60.4分提高到了63.97分。

不仅如此，根据在实际训练的过程中，迭代次数在50以后就已经接近于收敛，为了能搞更好的优化模型参数，我们采用动态学习率调整学习率，能够在迭代的后期可以更好的学习。在这里我们采用LambdaLR作为我们的动态调整学习率的方法。其中lambda表达式为式2。

最后由于不同的数据集对应着不同的锚框，如果采用官方自带的锚框参数训练出来的效果并不好，所以后期我们使用K-Means密度聚类来生成个性化的锚框，用以来适应本次船舶数据集。

# 结论

综上所述，本文在yolov5的基础上，提出了一种改进的yolov5算法实现了对海洋上船舶识别检测与自动分类。实验结果表明，该方法可以较为准确的识别出海面上船只及其的种类，并能识别出不易察觉到的暗礁，具有较高的准确率与较快的识别速度，能有效为相关者提供辅助决策信息。但由于受到训练时的数据库中样本数量与标注等问题影响，模型仍存在遗漏、错分类等问题。未来可基于此进一步改进与完善模型与算法。

# References

1. 张全发，蒲宝明，李天然等。基于HOG特征和机器学习的工程车辆检测[J]。计算机系统应用，2013，22(7)：104-107.
2. 赵磊，王斌，张立明。基于分割窗半监督支持向量机的遥感图像变化检测[J]。复旦学报（自然科学版），2010，49(2)：190-196.
3. Salembier P, Liesegang S, Lopez-Martinez C. Ship Detection in SAR Images Based on Maxtree Representation and Graph Signal Processing[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018:1-16.
4. Park S, Cho C J, Ku B, et al. Compact HF Surface Wave Radar Data Generating Simulator for Ship Detection and Tracking[J]. IEEE Geoscience & Remote Sensing Letters, 2017, PP(99):1-5..
5. 王方超，张旻，宫丽美等。海洋背景下的海面船舶快速检测算法[J]。激光与红外，2016,46(5)：602-606.
6. 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). IEEE, 2016.
7. CAI Ying Feng, WANG Hai. Using visual significance and depth of the convolution of the neural network robust visual vehicle identification algorithm[J]. Journal of Jiangsu university: Natural science edition, 2015, 36(3): 331-336.
8. REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]//IEEE Conference on Computer Vision & Pattern Recognition, 2017.
9. Liu S，Qi L，Qin H F，Shi J P and Jia J Y． 2018． Path aggregation network for instance segmentation/Proceedings of 2018 IEEE / CVF Conference on Computer Vision and Pattern Recognition． Salt Lake City: IEEE: 8759-8768.
10. Wang C Y, Mark Liao H Y, Wu Y H, et al. CSPNet: A new backbone that can enhance learning capability of cnn[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 390-391.