技术报告

蒋成鑫1 杨皓杰1 李志豪1

(1 武汉理工大学交通学院，湖北 武汉 430064)

**摘要** 本团队针对比赛方提供的数据集，设计了基于mmdetection深度学习目标检测平台的高精度海洋目标智能感知算法。该算法在对数据集进行统计和分析的基础上，从网络模型设计、数据增强、参数设置、模型训练和测试等方面提出一系列针对性的改进方法和策略，然后通过渐进迭代式的方式对提出的方法和策略进行验证，最终在A榜和B榜上分别取得了64.4和61.51的成绩。

**关键词** 海洋目标；数据增强；卷积神经网络；目标检测；深度学习；

**中图分类号**：TP319.56

**Technical Report**

Jiang Cheng-Xin1 Yang Hao-Jie1 Li Zhi-Hao1

(1 School of Transportation，Wuhan University of Technology，Hubei，Wuhan 430064，China)

**Abstract** The team designed a high-precision ocean target intelligence sensing algorithm based on the mmdetection deep learning target detection platform for the data set provided by the competition party. Based on the statistics and analysis of the data set, the algorithm puts forward a series of targeted improvement methods and strategies from the aspects of network model design, data enhancement, parameter setting, model training and testing, and then verifies the proposed methods and strategies by gradual iterative way, and finally achieves 64.4 and 61.51 results in the A and B lists, respectively.

**Key words** Ocean targets; Data augmentation; Convolutional Neural Networks; Object detection; Deep learning;

# 引言

工业和信息化部、交通运输部、国防科工局联合颁布《智能船舶发展行动计划（2019-2021年）》，为我国船舶工业提出了一个“三年目标”：形成智能船舶发展顶层规划，初步建立智能船舶规范标准体系，突破航行态势智能感知、自动靠离泊等核心技术，完成相关重点智能设备系统研制，实现远程遥控、自主航行等功能的典型场景试点示范，扩大典型智能船舶“一个平台+N个智能应用”的示范推广，初步形成智能船舶虚实结合、岸海一体的综合测试与验证能力，保持我国智能船舶发展与世界先进水平同步。

根据中国船级社（CCS）在2020年新颁布的《智能船舶规范》中的定义：“智能船舶系指利用传感器、通信、物联网、互联网等技术手段，自动感知和获得船舶自身、海洋环境、物流、港口等方面的信息和数据，并基于计算机技术、自动控制技术和大数据处理和分析技术，在船舶航行、管理、维护保养、货物运输等方面实现智能化运行的船舶，以使船舶更加安全、更加环保、更加经济和更加高效。

其中，摄像头作为应用最为广泛的传感器，已成为越来越多人的研究热点。此次海洋目标智能感知国际挑战赛以“逐梦海洋，感知智能，突破自我，创新未来”为主题，通过对海洋目标的智能感知测评，发现一批先进算法和技术，旨在推进国内外人工智能与智慧海洋技术领域研究热潮，搭建人工智能与智慧海洋的人才交流平台，加速智能技术在海洋领域的发展和应用创新，助力我国海洋强国建设。

本次比赛要从摄像头采集的可见光图像中，提取各类目标的各种特征，以达到目标识别和检测的效果。挑战赛的主办方提供了精心设计的数据样本，以此数据为训练集基础，每个团队按照竞赛规则，设计合适的算法，来实现不同类别船舶以及岛屿的检测。大赛对增强大学生海洋意识，创新精神，引导大学生树立努力奋斗、刻苦钻研，促进其科技创新及实践能力具有重要的意义。通过大赛，能够集众人智慧，实现海上目标感知的技术突破，从而为我国船舶航行态势智能感知领域的技术发展助力。

# 整体设计思路

**2.1赛题分析**

2.1.1 规则分析

该挑战赛采用平均精度均值mAP（mean Average Precision）作为评价标准，即交并比IoU(Intersection over Union)取值在0.50到0.95，间隔为0.05的10个阈值下计算得到的，在所有类别下的mAP的均值。

选手在公开测试集（A）和盲测试集（B）的竞赛成绩的加权平均作为总成绩。公开测试集（A）和训练集一同发放，盲测试集（B）在竞赛期间不公布。在公开测试集（A）排名前15%的选手将通过盲测试集（B）进一步区分，总加权成绩为0.4 x “A测试集mAP值”+ 0.6 x “B测试集mAP值”。

比赛的总成绩主要是以mAP值来计算，所以在模型设计的过程中，主要考虑精度的提升，期望能在竞赛中取得好成绩。

2.1.2 训练集分析

在拿到数据集后，我们对训练集进行了分析，首先进行了训练集中目标的统计分析，目标所占类别比例如图1所示：

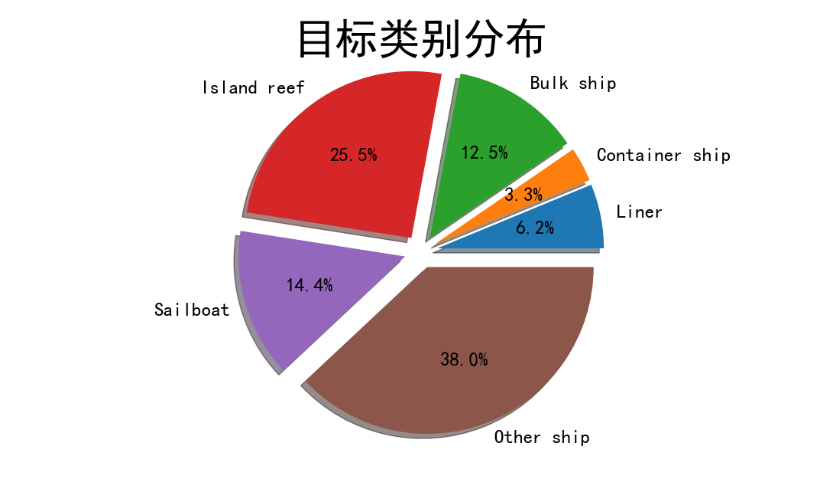


图1 训练集目标类别分布

从图1可以看出，训练集具有以下特点：1）在训练集的占比中，other ship 和 island reef所占的比例比较大，other ship占比为38%，island reef占比为25.5%；2）在分析训练集的标注特点时，发现island reef的标注存在打断现象，可能对模型训练产生干扰；3）other ship 所包含的类别太多，包括工程船，救援船，游艇等，可能使得机器学习的特征过于杂乱，使得模型学习效果较差。

然后，团队对训练集目标的尺度进行了分析，按照coco数据集尺度划分的方法：把物体分成small obj(面积小于32²像素)，medium obj(面积大于32²，小于96²像素)，和large obj(面积大于96²像素)。训练集的尺度分布如图2所示。

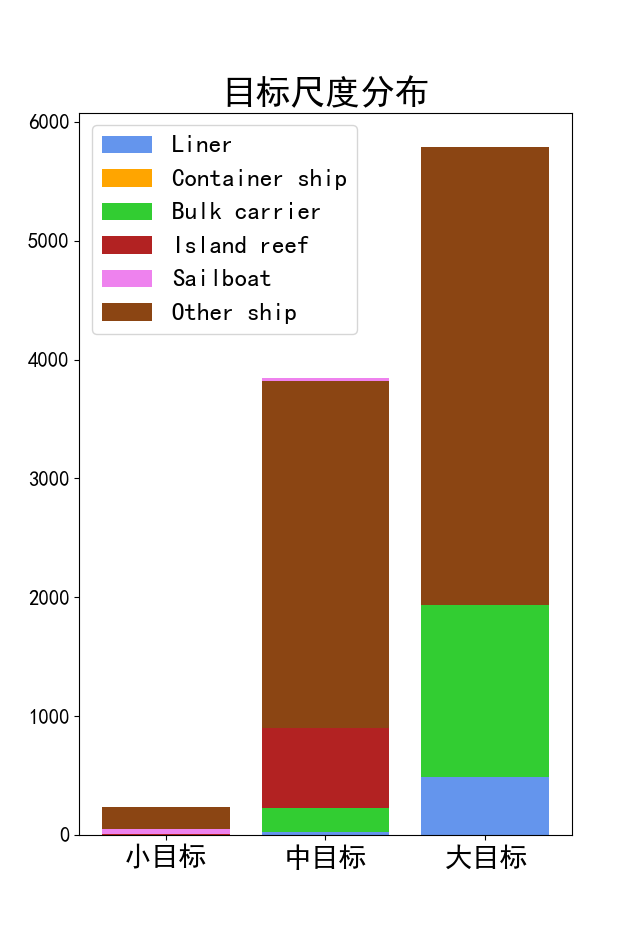


图2 训练集目标尺度分布

从图2可以分析得出，other ship尺度跨度比较大，在小、中和大尺度目标都有分布，识别可能比较困难；island reef主要分布在中尺度范围内；bulk carrier主要分布在大尺度范围内，在中尺度范围内也有少量占比；liner主要分布在大尺度范围内；sailboat主要分布在小尺度范围内；container ship主要分布在中尺度范围内。

**2.2设计思路**

根据对赛题的分析，本团队基于mmdetection[1]深度学习目标检测平台，设计了一套高精度海洋目标智能感知算法。

2.2.1 整体网络结构

随着RCNN进入目标检测领域，目标检测被构建为分类和回归的问题进行解决，在检测问题中是对图像中的所有候选框进行打分，通过IoU阈值来判定正负样本，因此IoU阈值的选取是一组需要精心挑选的超参数。一方面，IoU阈值选取的越高，则得到的正样本更接近目标，因此训练出的检测器定位更加精准，但是一味提高IoU阈值会引发两个问题：一是正样本过少导致训练的过拟合问题，二是训练和测试使用不一样的阈值导致评估性能的下降；另一方面，IoU阈值选取的越低，得到的正样本更为丰富，有利于检测器的训练，但势必会导致测试时出现大量的虚检。而Cascade RCNN目标检测算法便很好地解决了目标检测在训练过程中的IoU阈值的选取问题。因此本团队的主体使用Cascade RCNN[2] 目标检测算法。

对于骨干网络则是利用可变性卷积（DCN[3]）改进后的ResNet50[4]，该网络广泛用于目标分类等领域。Neck部分运用到了FPN [5]，提高了模型的多尺度检测能力。

2.2.2 数据处理技术

深度学习往往需要大量的样本数进行训练，而实际情况中数据并没有我们想象中的那么多，这时就要用到数据增强。

数据增强是用来增加训练的数据量，提高模型的泛化能力，增加噪声数据，提升模型的鲁棒性，避免过拟合。数据增强可以分为两类，一类是离线增强，一类是在线增强。在离线数据增强方面，我们对测试集采用了padding策略，通过补零，保证了图像的长宽比保持1：1，为后续的多尺度测试提供了便利。此外，本团队选择了在线的水平翻转策略（Flip），以0.5的概率随机水平翻转图像，以到达扩充数据集的效果，垂直翻转即上下翻转，不符合实际情况，可能对模型训练造成障碍，不予采用。

2.2.3 训练及后处理技术

训练及后处理技术对精度的提升也相当重要，本团队应用到的技巧如下：

多尺度训练[6]：通过分析训练集的图片，发现目标的面积跨度大，小目标也占了一部分比例，所以我们选用了多尺度训练，期望能提升对小目标的检测效果。

多尺度测试[6]：在测试阶段引入多尺度策略，也可获得大尺寸和多尺寸带来的增益。

Soft-NMS[7]：在传统的NMS中，跟得分最高预测框重合度超出一定阈值的预测框会被直接舍弃，不利于相邻物体的检测，所以采用了Soft-NMS策略。

混合精度训练[8]:因为训练集中有部分图片尺寸较大，所以我们采用了混合精度训练，用半精度浮点数fp16来代替单精度浮点数fp32,能够减小显存，增大训练批次。

# 具体模型和方法

3.1 数据处理

数据处理上，使用了离线数据增广、在线数据增广以及对先验框进行了统计分析，可以使先验框更好的匹配目标；

3.1.1 离线数据增广

离线数据增广即直接对数据集进行处理，本团队采用的策略为padding（补零填充）。在图像周围补零，使图片的长宽比为1:1。对于图像大小大于某一尺寸的的图片不做处理；对长宽比不为1:1的图片，固定他的长边，用长边减去短边，得到的差作为补零的数目。图3为没有经过padding操作的图片示例，图4为经过padding操作的图片示例。



图3 没有经过padding操作的图片示例



图4 经过padding操作的图片示例

3.1.1.2 在线数据增广

在线数据增广，即获得 batch 数据之后，对这个 batch 的数据进行增强，如旋转、平移、翻折等相应的变化。本团队采用随机水平翻转策略，以0.5的概率随机水平翻转图像，以到达扩充数据集的效果。

相较于垂直翻转，垂直翻转后会使船舶上下颠倒，水平翻转更符合工程实际。图5为未经水平翻转的图片示例，图6为经过水平翻转后的图片示例，图7为在线数据增广流程图。



图5 未经水平翻转的图片示例



图6 经过水平翻转的图片示例

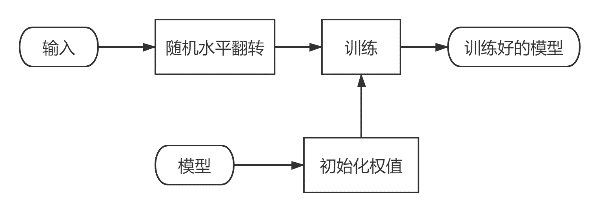


图7 在线数据增广流程图

模型中先验框的默认设置为ratios = [0.5，1.0，2.0] ，我们将检测框的参数从xml文件中提取出来，得到检测框的长宽比，对目标的长宽比进行统计分析，便于选择与ground truth有更大的IoU值的anchor boxes，可视化如图8所示：

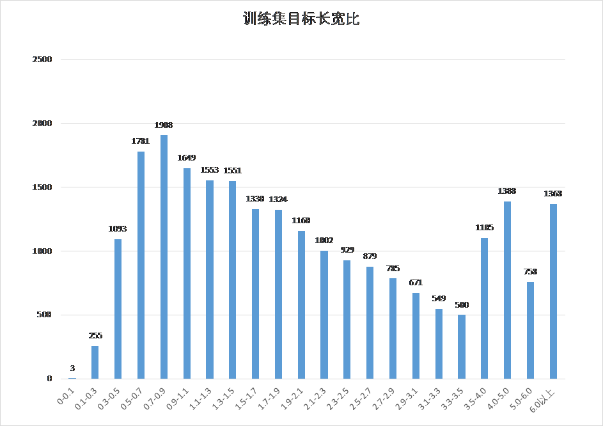


图8 训练集目标长宽比

分析目标特点后，本团队选择的先验框长宽比ratios = [0.4，0.7，1.0，2.0，5.0]，采用更合适的anchor设置，能够使网络更加容易去学习，能够得到更好的检测结果。

3.1.2 网络结构

以本团队最终在排行榜取的最终模型为例，本章将详细叙述该算法模型的设计细节。本团队采用的Cascade RCNN的网络结构图如图9所示：

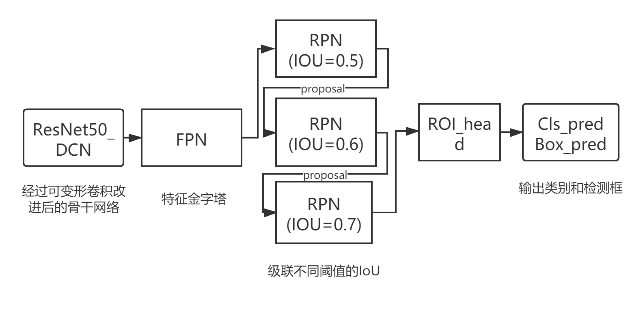


图9 网络结构图

3.1.2.1 骨干网络

本团队选择的骨干网络为ResNet50，ResNet系列在2015年由微软实验室提出，斩获了当年ImageNet竞赛中的分类任务第一名，目标检测第一名；获得COCO数据集中目标检测第一名，图像分割第一名。其模块性强，可移植性高，而且在计算量差不多的情况下，与VGG相比，精度提升了一个档次，被广泛运用于目标检测任务中。

卷积神经网络在计算机视觉任务取得了显著的成果，但是经典卷积结构是固定的。由于图像不同位置的结构可能是不同的，因此，相同结构的卷积进行计算会使得对目标建模不充分。本团队采用通过可变形卷积改进后的ResNet50作为骨干网络，期望达到好的检测结果。

可变性卷积的原理如图10所示:

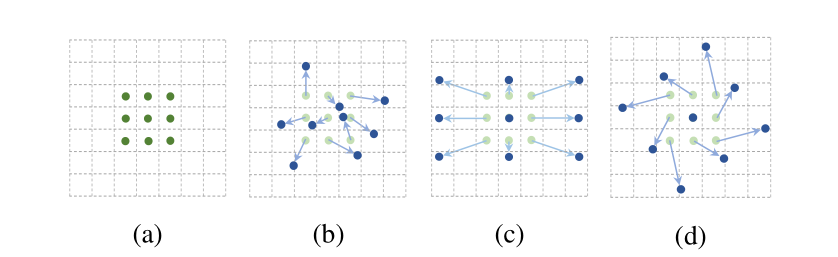
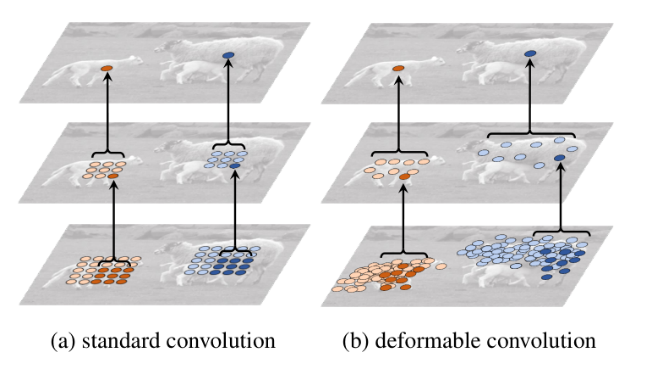


图10 卷积核的几种形式

图10(a)为3\*3经典卷积核，图10(c)是针对尺度和纵横比，改变之后的卷积核,图10(d)是针对图像旋转而改变的卷积核，图10(b)是原作者提出的，针对图像的任意变化，网络会学习到卷积核的偏移量，使网络可以适应目标不同的变化，在前向传播阶段，偏移量改变了之后的卷积核生成的特征图向下一层传播；在反向传播阶段中，要调整的参数除了原本的卷积核权值外，还有形变等等偏移量。

传统卷积和可变性卷积如图11所示，图11(a)为经典卷积，图11(b)为可变卷积。



(a)经典卷积操作 (b)可变形卷积操作

图11 两种卷积操作对比

3.1.2.2 特征金字塔

特征金字塔（FPN）主要是解决目标检测在处理多尺度变化问题时的不足，现在的很多网络只是利用了单个的高层特征，进行后续的物体的分类和bounding box的回归。这样的缺陷也很明显，在前面的目标尺度分析中可以看到，sailboat这一类别主要分布在小尺度范围内，本身具有的像素信息较少，在下采样过程中极易被丢失，所以引入特征金字塔，期望提高模型的精度。

在特征金字塔中，底层特征图的分辨率高，语义信息弱，顶层特征图的分辨率低，语义信息强。在通过一系列的卷积得到了不同层的feature map，然后我们通过上采样，又还原出尺寸大小不变的特征图，将此时的特征图与下采样过程中相应的feature map之间构件横向连接，以使得检测器可以更好地预测location。特征金字塔结构图如图12所示：

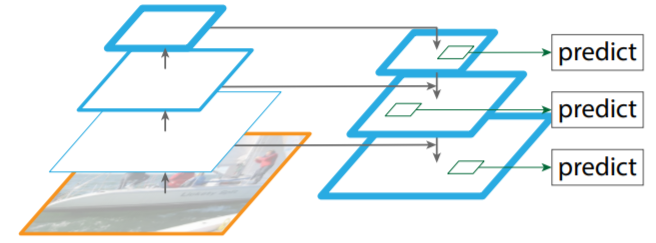


图 12 图像金字塔

3.1.2.3 Cascade RCNN

Cascade RCNN是在Faster RCNN上改进的，它能够级联不同IoU阈值（界定正负样本）的输出，使不同IoU值检测与其相对应的IoU值的目标。从图13中可以看出：

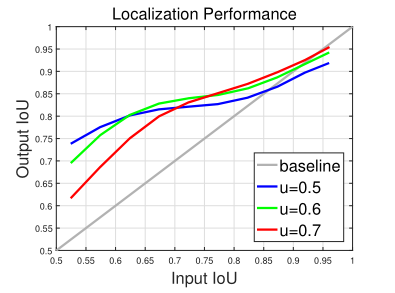


图13 不同阈值Input IoU与Output IoU曲线

图中的横轴Input Iou指的是RPN生成的proposal与ground truth的IoU值，纵轴指的是经过边界框回归之后的新的proposal与ground truth的IoU值。在图中横轴为0.5-0.6的范围内，可以看到设定的阈值为0.5的曲线，即为图中蓝色的线，输出的效果最好。如果输入IoU与设置的阈值接近，则这个模型的检测检测结果会比其他输入IoU的模型好。

因此，Cascade RCNN采用级联结构，级联各个检测模型，设置不同的IoU阈值界定样本训练模型，前一个检测模型的输出是后一个检测模型的输入，IoU值一直上升，能够有效提升模型的精度。下图为Cascade RCNN的网络结构图：

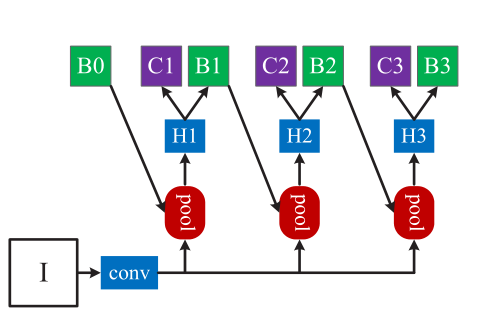


图14 Cascade RCNN的网络结构图

3.1.3 训练及后处理技术

模型的训练也是决定模型能够发挥其性能的重要因素，目前针对模型训练的技术研究也愈发火热。在训练及后处理阶段，本团队针对性地选择了五大技术，下面将详细介绍各技术的原理和参数设置。

3.1.3.1 多尺度训练

输入图片的尺寸对检测模型的性能影响相当明显，事实上，多尺度是提升精度最明显的技巧之一。在基础网络部分常常会生成比原图小数十倍的特征图，导致小物体的特征描述不容易被检测网络捕捉。通过输入更大、更多尺寸的图片进行训练，能够在一定程度上提高检测模型对物体大小的鲁棒性。

在训练前，我们通过padding处理后，保证了图片的长宽比为1。假设我们有一张图片的大小为（1500,1000），经过padding处理后，图片大小为（1500,1500），ratio为1；在训练过程中，我们设定了两个尺度，一个是800\*800，另一个是1200\*1200，首先我们将图像的短边固定到800到1200的范围内的某一个数值，假设为1100；保证长宽比不变，那么处理之后的图片长边为也1100，即图片尺寸大小为1100\*1100。

3.1.3.2 软非极大值抑制

NMS是一种获取局部最大值，抑制非极大值的算法，在计算机视觉中有着广泛的应用。在目标检测算法会输出多个检测边框，尤其是在真实目标周围会有很多置信度高的检测边框，NMS的作用就是为了去除重复的检测边框，达到每个物体有且只有一个检测结果。但是在我们分析训练集文件时，发现有很多船只叠加在一起，如果IoU阈值设得过高，那么可能造成重叠框过多；如果IoU设得过低，那么可能会造成漏检。而且传统的NMS方法是基于分类分数的，只有高分数的预测框能留下来，但是大多数的情况下IoU和分类分数不是强相关，很多分类标签置信度高的框位置并不是很准确。在这里本团队采用了Soft-NMS策略。

对于NMS而言，其直接将与得分最大的框重合程度较高的其他预测框剔除。而在Soft-NMS算法中，对于与最高分框重叠部分大于阈值的框，我们不把他直接去掉，而是将他的置信度降低，这样的方法可以使多一些框被保留下来，从而一定程度上避免过度抑制的情况出现。

蓝色的部分表示原始NMS算法，绿色部分表示Soft-NMS算法，可以看出Soft-NMS只是把Si降低了，而不是把bi直接去掉。其中的f函数是为了降低目标框的置信度，满足条件，如果bi和M的IoU越大，f(iou(M, bi))就应该越小。图15为NMS、Soft—NMS的完整算法；且本团队应用了高斯加权函数（如下公式）：

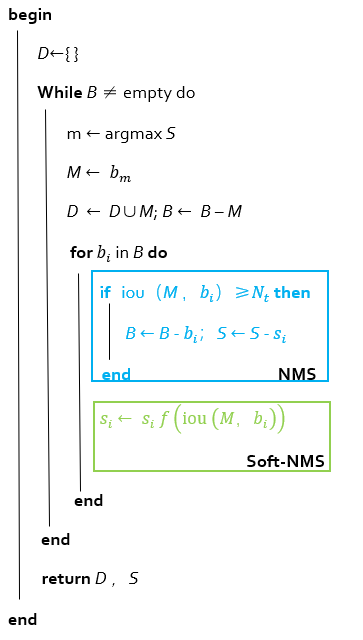


图15 NMS、Soft\_NMS算法流程

即以一个权重的形式，将获得的IoU值取高斯指数后乘上原得分，之后重新排序，继续循环。越接近高斯分布中心，惩罚力度越大。

最后Soft-NMS还是需要调整阈值的，由于没有“去掉”某些框的操作，因此，最后所有框都会被加入D集合中，只是那些被去掉的框的置信度明显降低，所以需要一个阈值（通常很小）来过滤D中的输出框，从而输出最后的预测框。

3.1.3.3 学习率调整

学习率调整包括预热策略和学习率阶梯式下降。预热策略即在训练开始时选择一个较小的学习率，然后经过一定的迭代次数，达到预先设定的学习率。在训练初期，模型的权重是随机初始化的，此时若选择一个较大的学习率,可能带来模型的不稳定。本团队设定的初始的学习率为0.001，预热周期为1000，预设的学习率为0.0015，在此周期内，学习率进行线性增长。在达到预先设定的学习率后，在epoch为[9，11]的范围，进行阶梯式衰减，这样有助于模型收敛速度变快，效果更佳。

3.1.3.4 多尺度测试

多尺度测试是在所有的尺度下逐一测试，然后融合所有尺度的测试结果，得到平均输出，作为图像的最后输出。与多尺度训练不同，在保证长宽比不变的参数设定下，测试集的图片输出尺寸可以为任意多个，本团队设计的尺寸为600\*600， 1000\*1000，1400\*1400；对不同尺寸的图像进行处理，得到不同尺寸的feature map，

对每个Region Proposal，在不同的feature map上也有不同的尺度。我们选择最接近某一固定尺寸（即检测头部的输入尺寸）的Region Proposal作为后续的输入。

3.1.3.5 混合精度

在日常中深度学习的系统，一般使用的是单精度浮点表示，即FP32，那么半精度浮点数就是FP16。混合精度训练，指的是单精度FP32和半精度FP16混合，这样可以使内存占用更少，训练的时候可以用更大的训练批次，计算更快。但是在训练期间，一些权重梯度具有非常小的指数，其FP16格式可能会变为零。因此，本团队将所有中间损失乘以512，以缓解对部分极小中间损失的忽略，保证模型训练过程中的精确度。

3.2 在现有技术之上的创新之处

第一，在数据处理方面，团队采用了padding策略。在多尺度训练中，需要从中选取一个尺度，然后重新定义图片的尺寸，先对其进行padding策略，能够保证在保持长宽比不变的配置下，尽可能多的保留图片的原始信息。

第二，在模型设计方面，在骨干网络ResNet50中将普通卷积改为可变形卷积。其卷积核在每一个元素上额外增加了一个可学习的偏移参数。这样的卷积核在学习过程中可以调整卷积的感受野，从而能够更好的提取图像特征，以达到提升目标检测精度的目的。

第三，采用Soft-NMS改进Cascade中的NMS算法。在进行训练集分析时，发现部分图像存在真值框重叠，如果采用NMS算法，则会过度抑制检测框的保存，团队采用Soft-NMS算法，对于与最高分框重叠部分大于阈值的框，不把他直接去掉，而是将他的置信度降低，这样的方法可以使多一些框被保留下来，从而一定程度上避免过度抑制的情况出现。

# 实验结果

表1为本团队消融实验结果表，我们确定基准模型为Cascade RCNN+ResNet50+FPN后，在数据处理、网络结构、训练及后处理技巧上，不断进行改进，最终在A榜和B榜上分别取得了64.4和61.51的成绩。



表1 消融实验结果表

# 结论

针对比赛方提供的数据集，本团队基于mmdetection深度学习目标检测平台，设计了一套高精度海洋目标智能感知算法，经过一系列的消融实验后，得到了最终的模型。模型在给定船型如bulk carrier等表现较好，在island reef即岛礁的表现欠佳。此问题将在后续的学习过程中，对其进行进一步研究。

# References

1. Kai Chen，Jiaqi Wang，Jiangmiao Pang，et al．MM-

Detection: Open MMLab Detection Toolbox and Bench-

mark[J/OL]．[2019-6-17]．http：arxiv.org/abs/1906.071

15．

1. Zhaowei Cai，Nuno Vasconcelos．Cascade R-CNN: Delving into High Quality Object Detection[J/OL]．[2017-12-3]．http：arxiv.org/abs/1712.00726．
2. Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, Yichen Wei. Deformable Convolutional Networks[J/OL].[2017-3-17]. https://arxiv.org/abs/1703.0

6211.

1. Kaiming He，Xiangyu Zhang，Shaoqing Ren，et al．Deep Residual Learning for Image Recognition[J/OL]．[2015-12-10]．http：arxiv.org/abs/1512.03385．
2. Tsung-Yi Lin，Piotr Dollár，Ross Girshick，et al．Feature Pyramid Networks for Object Detection[J/OL]．[2017-4-19]．http：arxiv.org/abs/1612.03144．
3. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J/OL].[2014-6-18]. https://arxiv.

org/abs/1406.4729.

1. Navaneeth Bodla, Bharat Singh, Rama Chellappa, Larry S. Davis. Soft-nms–improving object detection with one line

of code[J/OL]. [2017-4-14]. https://arxiv.org/abs/1704.04

503.

1. Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Gins-

burg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, Hao Wu. Mixed Precision Training[J/OL].

[2017-10-10]. https://arxiv.org/abs/1710.03740.

**蒋成鑫** 武汉理工大学交通学院本科生.未来研究方向为目标检测技术.本文通信作者.

E-mail: 488365347@qq.com

(Jiang Cheng-Xin Undergraduate student at School of

Tran-sportation, Wuhan University of Technology. His future research interest is the object detection technology. Corresponding author of the paper.)

**杨皓杰** 武汉理工大学交通学院本科生.未来研究方向为图像增强技术.

E-mail: 69475741@qq.com

(Yang Hao-Jie Undergraduate student at School of Transportation, Wuhan University of Technology. His future research interest is image enhancement technique.)

**李志豪** 武汉理工大学交通学院本科生.未来研究方向为目标检测技术.

E-mail: 1322525676@qq.com

(Li Zhi-Hao Undergraduate student at School of Transportation, Wuhan University of Technology. His future research interest is the object detection technology.)