

基于 VFNET 的船舶识别模型

廖章泽¹ 曹馨之¹ 朱世雄¹

摘要 海洋环境感知作为智慧海洋技术的一个重要组成部分，其技术的发展具有重要意义，本次海洋目标智能感知大赛旨在推动该技术的发展。面向本次大赛的海洋目标感知任务，我们有针对性的做了数据的分析与预处理的工作，并且针对训练数据量不足的问题进行了数据增广。基于数据的特征尝试了三种不同的目标识别的模型，最终确定使用 VFNet 作为基础模型。经过各种测试策略的实验结果得到基于 Soft-NMS 机制进行测试可得到最优精度，达到 64.17。

关键词 目标识别，数据增广，VFNet

中图分类号: TP391.41

Ship Detection Model Based On VFNet

Liao Zhangze¹ Cao Xinzhi¹ Zhu Shixiong¹

Abstract As an important part of smart ocean technology, the development of ocean environment perception technology is of great significance. This competition aims to promote the development of this technology. For the ocean target perception task of this competition, we have done the data analysis and preprocessing work, and have carried on the data expansion to solve the problem of insufficient training data. Based on the characteristics of the data, three different models of target recognition are tried, and finally vfnet is used as the basic model. According to the experimental results of various test strategies, the optimal accuracy of 64.17 can be obtained based on soft NMS mechanism.

Key words Object Detection, Data Expansion, VFNet

1 引言

海洋环境感知是当前海洋信息研究的前沿领域，是智慧海洋技术的一个重要组成部分。环境感知任务包含数据的获取与处理、模型的构建与评价。本文的工作基于本次海洋目标智能感知大赛，大赛数据集包含六类检测目标：liner（游轮），container ship（集装箱船），bulk carrier（散货船），island reef（岛礁），sailboat（帆船），other ship（其他船）。训练集中包含了9800张图片，公开测试集A中包含1000张图片。我们使用了一些预处理方法对训练集和测试集进行初步观察，这是我们在最终取得较好成绩的关键工作之一。我们的训练过程中共使用了三种基础目标检测模型，依次为：Faster R-CNN^[1]→YOLOv3^[2]→VFNet^[3]。在选定最终模型后，主要的工作为调整模型参数和训练数据增广，两者都在一定程度上提升了模型在测试集上的得分。除此之外，我们在生成测试集结果时采用Soft-NMS^[4]机制，将整体精度提高了0.5%。

1. 上海大学 计算机工程与科学学院 200444
1. School of Computer Science and Engineering, Shanghai University 200444

2 相关工作

2.1 数据增强技术

数据增强也叫数据扩增，意思是在不实质性的增加数据的情况下，让有限的的数据产生等价于更多数据的价值。传统的图像数据增强囊括了以下几种方法：水平翻转、竖直翻转、随机翻转角度、随机水平平移、随机竖直平移、随机错切变换、随机放大、颜色抖动和缩放等。使用这些有监督的数据增强方法的目的是为了得到多样性的船舶和岛屿数据，防止模型出现过拟合，提高模型泛化的能力。

2.2 目标检测模型

在我们的工作中共使用了三种目标检测模型，分别为 Faster R-CNN、YOLOv3 和 VFNet。

Faster R-CNN 是何凯明等在 2015 年提出目标检测算法，该算法在 2015 年的 ILSVRC^[5]和 COCO^[6]竞赛中获得多项第一。该算法在 Fast R-CNN^[7]基础上提出了 RPN 候选框生成算法，使得目标检测速度大大提高。RPN 是用来判断每个像素点对应的 k 个区域是不是包含目标，如果包含（那么先根据输出的坐标偏置修正 box 位置）则输给后面的 RCNN 做进一步判断，其中 k 为预定义常数。

YOLOv3 在保留划分单元格检测、采用 Leaky ReLU 作为激活函数、端到端进行训练、使用 BN 层（Batch Normalization）和多尺度训练的基础上，取消了 pooling 层并添加残差层加深了网络，使用多尺度输出以达到检测更多类别的目的。

VFNet 主要针对了两方面的问题做出了自己的优化。1) 分类分支或者联合分类分支和 IoU-based localization 分支作为检测候选框的排序依据不够可靠；2) Bounding Boxes 表示不够灵活，不够准确有效。对于问题 1) VFNet 提出了一个联合表示。

$$VFC(p, q) = \begin{cases} -q(q \log(p) + (1-q) \log(1-p)) & q > 0 \\ -ap^r \log(1-p) & q = 0 \end{cases} \quad (1)$$

对于问题 2)，VFNet 提出了 Bbox Refinement，星型 bbox 特征提取和 refine 操作。

3 模型和方法描述

3.1 数据分析和处理

我们在构建模型前先对数据集进行分析，首先是统计了每一类目标在数据集中的图像数量，（如果在一张图像中存在一个或多个该类目标物，则数量加一，遍历所有图像）统计结果如表 1 所示。然后我们对所有的标注文件进行了分析，主要从 Ground-truth 的类别的分布和图片的横纵比这两方面进行展开，统计结果如表 2、3、4 所示。通过对数据的预处理我们得到了很多信息，例如在标注数据中存在像素值面积为 1 的 Ground-truth（错误标注）；训练集中 other ship 的数量要远远高于其他船

类；island reef 标注的不仅仅是传统意义上的岛屿，岸边也被囊括在了其中（房子，码头等）；大部分图片的横纵比在 1~2 之间，相对常规。）

这些信息对我们接下来选择模型、训练模型和调整参数都起到了至关重要的影响。在观察各类目标的分布后可以发现分布是不均匀的，为了使得各类目标在特征空间中达到相对平衡，我们使用了数据增强的方法对增广了出现次数相对较少的目标类别，希望这样的方式平衡网络。除此之外我们还发现 Ground-truth 所包含的像素数量的取值范围很大，说明 Ground-truth 的尺度存在很大的区别。因此模型需要适应多种尺度的目标物，我们认为基于特征金字塔的目标检测模型在这个任务是合适的。

3.2 VFNet

在目标检测中 NMS 需要依据候选检测目标的排序来进行筛选框，因此对候选框进行准确的排序对于优异的目标检测器来说是至关重要的。但是仅基于 IoU 分支(IoU-Net)与 Centerness(FCOS)来作为大量候选检测的排序依据是不可靠的，会降低检测的准确度。于是 IoU 感知分类得分 (IACS) 的出现可以同时表示物体的存在置信度和定位精度，并且不需要采用预测一个额外的定位精确度得分，而是将其融入到分类得分中，即预测一个可以同时代表目标存在和定位进入的定位感知或 IoU 感知的得分，从而在训练密集的物体检测器中产生更准确的检测排序。

VarifocalNet(VFNet) 是一个基于 FCOS^[8]与 ATSS 架构的 IoU 感知分类的密集目标检测器，使得检测的结果更加准确，这也是我们本次比赛采用该模型的原因。与 FCOS 和 ATSS 相比，有三个创新点：Varifocal 损失，新的高效星形边界框特征表示以及优化边界框。其中 p 是预测的 IACS 得分， q 是目标 IoU 分数。对于训练中的正样本，将 q 设置为生成的 bbox 和 gtbox 之间的 IoU，而对于训练中的负样本，所有类别的训练目标 q 均为 0。高效星形边界框特征表示用于估算 IACS 得分并改进边界框。

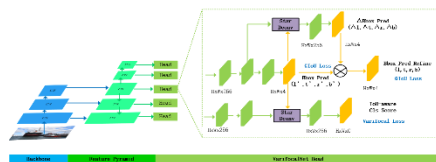


图 3.2.1 目标检测模型图

VFNet 是基于特征金字塔网络建立的，它的头网络是由两个子网构成的，其中一个子网是用来对初始边框进行回归并对其进行不断改进，另一个子网的作用是预测 IoU 感知分类得分。

3.3 Soft-NMS

NMS 是最常用的目标检测后处理的算法，但是 NMS 对于密集目标的检测就会出现一些问题，因为 NMS 是对其他框的 IoU 与当前最大置信度的框的 IoU 进行比较，如果大于一定的阈值，就将当前这个置信度最大的框周围满足条件的框去掉。通常情况普通的 NMS 是可行的，但是有些情况比如与相似目标的距离比较近时，可能会导致误检，那么这个阈值的设置永远就不可能完美，所以通过较低的 NMS 重叠阈值来移除所有相邻检测框并不是最优解，并且很容易漏掉被检测物体，尤其是在物体高度重叠的地方。所以由于在检测是会出现目标密集的情况，所以我们采用了 Soft-NMS 的方法。

4 实验结果

4.1 YOLOv3

我们首先以 YOLOv3 为初始模型，在训练集上训练 250 轮左右，并在训练集上查看一下训练的效果，这里的比较方式是以 IOU=0.5 的方式做的，如下图所示：

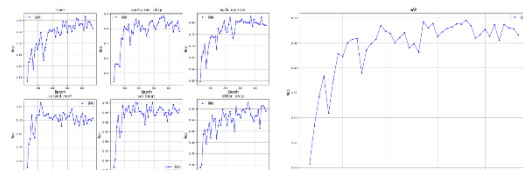


图 4.1.1 yolov3 训练模型在训练集上的效果

观察上述图像我们可以得到以下结论

- ✓ Liner、container ship、bulk carrier 上整体能获得较好的目标识别效果，且在 200 轮基本上已经达到最优值。
- ✓ 可以发现，拉低整个模型下限的两个类别分别是：island，other ship。

4.2 VFNet

有了上述经验之后，我们选择了最近的 VFNet 模型进行训练，并得到了 63.8 的结果。我们在训练集上可视化了部分船舶图像，发现部分船舶会发生被二次检测的问题，这是和非极大值抑制挂钩的问题。为此，我们将原始的 NMS 调整为 Soft-NMS，并调整阈值在 {0.6, 0.7, 0.8} 之间，最后上传结果显示在 0.7 时，整体的模型精度能够达到 64.0。此外，我们还设定了多尺度测试的方式，最后上传的结果显示，多尺度测试使得模型精度提升到 64.17。

5 结论

本次竞赛为 6 分类船舶识别，通过整体流程我们发现难点集中在：岛屿，其他船舶的识别上。在

模型选择方面，由于数据尺度多样性，必须选择基于特征金字塔的模型，且数据增强策略、Soft-NMS、多尺度测试能够大幅度的提升模型最后的性能。

表 1 基于图为单位的单目标统计表 （单位：张）

| | 游轮 | 集装箱船 | 散货船 | 岛礁 | 帆船 | 其他船 |
|-----|-------------|-----------|-------------|-------------|-------------|-------------|
| 训练集 | 1312(13.4%) | 842(8.6%) | 2864(29.2%) | 3353(34.2%) | 1679(17.1%) | 4514(46.1%) |

表 2 单目标标注数据统计表

| | 游轮 | 集装箱船 | 散货船 | 岛礁 | 帆船 | 其他船 |
|-----|------------|-----------|-------------|-------------|-------------|-------------|
| 训练集 | 1447(6.1%) | 871(3.7%) | 3121(13.3%) | 5910(25.1%) | 3350(14.2%) | 8843(37.6%) |

表 3 训练集图片横纵比统计表

| | 0~1 | 1~2 | 2~3 | 3~4 | 4~5 |
|-----|------------|--------------|------------|-----------|----------|
| 训练集 | 961(9.81%) | 8690(88.67%) | 133(1.40%) | 14(0.10%) | 2(0.02%) |

表 4 标注数据框内像素数量分级统计表

| | 0~100 | ~1000 | ~10000 | ~100000 | ~1000000 | ~100000000 |
|-----|------------|--------------|--------------|--------------|--------------|------------|
| 训练集 | 155(0.66%) | 2745(11.66%) | 7251(30.80%) | 9854(41.86%) | 2838(12.05%) | 699(2.97%) |

References:

1.

Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems; 2015; 2015. 91-99.

2.

Redmon J, Farhadi A. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 2018.

3.

Zhang H, Wang Y, Dayoub F, Sünderhauf N. VarifocalNet: An IoU-aware Dense Object Detector. arXiv preprint arXiv:2008.13367 2020.

4.

Bodla N, Singh B, Chellappa R, Davis LS. Soft-NMS--improving object detection with one line of code. Proceedings of the IEEE international conference on computer vision; 2017; 2017. 5561-5569.

5.

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M. Imagenet large scale visual recognition challenge. INT J COMPUT VISION 2015; 115(3): 211-252.

6.

Lin T, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft coco: Common objects in context. European conference on computer vision; 2014: Springer; 2014. 740-755.

7.

Girshick R. Fast r-cnn. Proceedings of the IEEE international conference on computer vision; 2015; 2015. 1440-1448.

8.

Tian Z, Shen C, Chen H, He T. Fcos: Fully convolutional one-stage object detection. Proceedings of the IEEE international conference on computer vision; 2019; 2019. 9627-9636.