

Neck: YOLOv5 采用与 YOLOv4[1]相同的 FPN+PAN 的结构, 并采用了 CSP2_X 结构, 加强网络特征融合的能力。结构图如图所示:

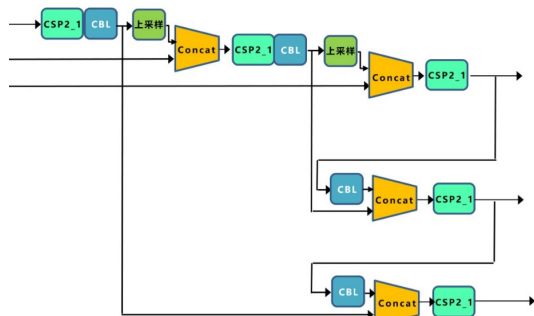
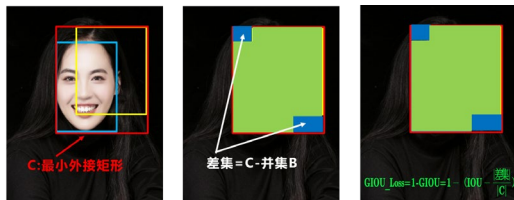


图 5 YOLOv5 Neck 结构图

Head: 在 YOLOv5 中采用了 GIoU_Loss[2]做为预测 Bounding box 的损失函数。由于传统的 IOU loss ($\text{IOU loss} = 1 - \text{IOU}$) 无法估计两个不相交框的距离，因此 2019 年提出的 GIoU_Loss 增加了相交尺度的衡量方式，其计算方法如图所示：



同时 YOLOv5 还采用了加权 nms 来筛选冗余预测框。

(2) 先验框配置

针对竞赛数据集，通过 K-means 聚类算法对训练集中已标注的 bounding box 进行聚类分析，综合考虑 6 种类别目标的大小和尺度变化范围，对三种尺度的特征图分别设计了 3 种尺度共 9 种尺度的先验框，如下表所示：

| Feature map | Anchor ratio |
|-------------|---------------------------------|
| P3/8 | (29,21), (71,33), (68,89) |
| P4/16 | (202,55), (181,133), (442,104) |
| P5/32 | (224,288), (549,199), (520,334) |

表 1 先验框配置表

2.2 在现有技术之上的创新之处

(1) 数据增广

经统计，比赛官方提供的训练集各类别数量为：

| 类别 | 数量 |
|-----------|------|
| Container | 871 |
| Liner | 1447 |
| Bulk | 3121 |
| Island | 5910 |
| Sailboat | 3350 |
| Other | 8843 |

表 2 官方训练集类别及数量统计表

由于存在较为明显的样本不均衡问题，因此团队额外采用了两批训练数据：

a. 额外人工标注的 3041 张船舶图片，Container、Liner、Bulk 各 1000 张左右，用于弥补这三类船舶训练样本较少的问题。

b. 自动生成 86400 张贴图数据，使用 Mask-RCNN 预训练模型对已有的训练集船舶图片进行目标分割提取，并将其贴在海洋背景图片上生成较为粗糙的贴图数据，实现流程如图所示：



图 6 贴图数据生成流程图

(2) 训练策略

为了更好地利用三部分训练数据（官方训练集、额外人工标注训练集、额外贴图训练集），团队有针对性地采取了特殊的训练策略。

a. 由于贴图训练集中的样本是“抠图+粘贴”的方式自动生成的，与真实图片差异很大，但可以快速生成大批量数据，因此选用了全部的额外的贴图数据（86400 个样本）被用于预训练（在 COCO 上预训练的 YOLOv5x 模型上进一步预训练）。

b. 官方训练集和部分贴图训练集（21000 个样本）合并后被用于对步骤 a 中所得的模型进行微调训练得到最终模型 1。

c. 官方训练集和额外人工标注训练集（3041 个样本）合并后同样被用于对步骤 a 中所得的模型

进行微调训练得到最终模型 2。

(3) 模型测试

a. 测试时增强 (Test Time Augmentation, TTA): 在测试阶段对每个样本进行了左右翻转、尺度缩放等图像变换, 得到多个增强样本, 输入模型并得到多个输出结果, 将这些预测结果合并后应用 NMS。TTA 技术的应用使模型能够更好适应不同尺度范围的目标以及提高模型的泛化能力。

b. 集成测试 (model ensembling): 对上述两个最终模型, 采用集成测试对两个模型最终的预测结果取平均, 获得最终的测试结果。

3 实验结果

首先使用 YOLOv5x 默认配置在官方提供训练集上进行训练并上传测试结果, 取得分数在 56.48, 并以此作为 baseline 进行改进。

从比赛训练集中挑选难度不一的 100 张样本, 作为验证集不参与训练, 用于辅助筛选模型。

通过进行 k-means 聚类算法针对官方训练集进行初始 anchor 配置, 并加入用于预训练的贴图训练集, 微调之后的模型测试结果为 63 分左右。

经过分析训练集的样本不均衡问题, 针对样本类别较少的三类船舶添加额外人工标注数据, 并在预训练模型上进行微调, 测试结果为 64 分左右。

最后根据公开测试集反馈, 多次筛选两个训练集上表现最优的训练结果模型, 并进行集成测试, 团队最终取得分数 64.95。

4 结论

本次竞赛中, 团队通过分析官方提供的数据集, 采用了创新性的数据增广方式, 特别是以低成本自动生成的 86400 张贴图数据大大提高了模型的泛化能力。同时, 团队采用有效的训练策略以及模型集成等测试阶段的技巧, 最终在竞赛中取得了较高的成绩。

参考文献

[1] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[2] 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. 2019: 658-666.

[3] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.

[4] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.