

〈图像处理与仿真〉

一种基于多尺度的目标检测锚点构造方法

邵延华¹, 黄琦梦¹, 梅艳莹¹, 张晓强¹, 楚红雨¹, 吴亚东²

(1. 西南科技大学 信息工程学院, 四川 绵阳 621010;

2. 四川轻化工大学 计算机科学与工程学院, 四川 自贡 643000)

摘要: 目标检测是计算机视觉领域的研究热点和基础任务, 其中基于锚点 (Anchor) 的目标检测已在众多领域得到广泛应用。当前锚点选取方法主要面临两个问题: 基于特定数据集的先验取值尺寸固定、面对不同场景泛化能力弱。计算锚框的无监督 K-means 算法, 受初始值影响较大, 对目标尺寸较单一的数据集聚类产生的锚点差异较小, 无法充分体现网络多尺度输出的特点。针对上述问题, 本文提出一种基于多尺度的目标检测锚点构造方法 (multi-scale-anchor, MSA), 将聚类产生的锚点根据数据集本身的特性进行尺度的缩放和拉伸, 优化的锚点即保留原数据集的特点也体现了模型多尺度的优势。另外, 本方法应用在训练的预处理阶段, 不增加模型推理时间。最后, 选取单阶段主流算法 YOLO (You Only Look Once), 在多个不同场景的红外或工业场景数据集上进行丰富的实验。结果表明, 多尺度锚点优化方法 MSA 能显著提高小样本场景的检测精度。

关键词: 目标检测; 锚点; 红外; YOLO (You Only Look Once); 多尺度分析

中图分类号: TN911.73 **文献标识码:** A **文章编号:** 1001-8891(2024)02-162-06

Multi-scale Anchor Construction Method for Object Detection

SHAO Yanhua¹, HUANG Qimeng¹, MEI Yanying¹, ZHANG Xiaoqiang¹, CHU Hongyu¹, WU Yadong²

(1. School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China;

2. School of Computer Science and Engineering, Sichuan University of Science & Engineering, Zigong 643000, China)

Abstract: Object detection is a popular research topic and fundamental task in computer vision. Anchor-based object detection has been widely used in many fields. Current anchor selection methods face two main problems: a fixed size of a priori values based on a specific dataset and a weak generalization ability in different scenarios. The unsupervised K-means algorithm for calculating anchor frames, which is significantly influenced by initial values, generates less variation in anchor points for clustering datasets with a single object size and cannot reflect the multiscale output of the network. In this study, a multiscale anchor (MSA) method that introduces multiscale optimization was developed to address these issues. This method scales and stretches the anchor points generated by clustering according to the dataset characteristics. The optimized anchor points retain the characteristics of the original dataset and reflect the advantages of the multiple scales of the model. In addition, this method was applied to the preprocessing phase of training without increasing the model inference time. Finally, the single-stage mainstream algorithm, You Only Look Once (YOLO), was selected to perform extensive experiments on different scenes of the infrared and industrial scene datasets. The results show that the MSA method can significantly improve the detection accuracy of small-sample scenes.

Key words: object detection, anchor, YOLO (You Only Look Once), infrared, multi-scale analysis

收稿日期: 2022-04-07; 修订日期: 2022-05-11.

作者简介: 邵延华 (1982-), 男, 四川省绵阳市人, 博士, 讲师, 主要研究方向为计算机视觉与图像处理、深度学习。E-mail: syh@alu.cqu.edu.cn.

基金项目: 国家自然科学基金项目 (61601382); 国防科工项目; 四川省科技支撑计划项目 (2019YJ0325, 2020YFG0148)。

0 引言

目标检测是当前计算机视觉领域的研究热点,其中目标尺度多变是最具挑战性的问题之一^[1-2]。针对上述问题,国内外学者提出多尺度检测方法^[3],利用检测器构建特征金字塔和多尺度特征输出获取更优的预测结果。为实现多尺度输出,锚点(anchor)被著名的 Faster-RCNN^[3]首次引入目标检测:即对图像中的目标进行分类检测之前,预先放置一批已知长宽的候选框,以便网络进行分类和识别。从此,基于锚点的检测算法成为目标检测领域的一个重要分支^[1]。

YOLO (you only look once) 除 YOLO v1 与 YOLOX 外,均为典型的单阶段基于锚点的检测模型^[4]。YOLO v1 使用全连接层直接预测边界框,由于空间信息丢失较多,导致定位不准。YOLO v2 受 Faster-RCNN 启发,通过引入大量锚点来代替 v1 的全连接层进行边界框预测,并使用交并比(intersection over Union, IoU)^[5]值最高的锚点结合预测偏移量得到预测框,提高了检测精度。由于原版 YOLO 模型的锚点是通过聚类 MS-COCO^[6]数据集得到的,对于特定场景数据集来说,部分锚点取值可能并不合理。因此,为避免上述问题,部分研究者提出了无锚点的检测模型,但该类方法由于图像金字塔或特征金字塔层数较多,导致检测速度有所降低^[7-8]。另有研究者在训练基于锚点的目标检测模型前使用基于欧式距离的 K-means 聚类方法自动找到较为合适的先验框。但出现第二个问题:K-means 的初始值对其结果影响较大,导致结果仅为局部最优^[8]。同时,通过聚类产生的锚点与数据集密切相关,当场景数据集类别单一或目标尺寸较集中时,聚类产生的锚点差异较小,无法体现 YOLO 多尺度输出的优势^[4]。

研究发现,不同场景的数据集具有不同的特点^[5],也影响锚点的取值。相较于可见光数据集(如 MS-COCO),红外数据集多具有小样本、目标少等特点,更容易导致聚类产生的锚点尺寸相似、缺乏多样性的问题。此外,热红外图像对比度低且纹理特征弱,会影响检测精度。故提升红外数据的检测精度是当前目标检测领域的热点之一。

针对人工设置锚点尺寸固定,泛化性差的问题,本文对锚点的优化进行深入讨论,提出一种多尺度的目标检测锚点构造方法(multi-scale-anchor, MSA)。该方法产生的锚点可根据数据集本身的特性进行线性尺度的缩放和拉伸,既保证模型多尺度的优势又体现不同数据集的特点。通过在 YOLO 系列的多个典型网络上进行测试,并分别使用中国空气动力研究与

发展中心的涡核(Vox)检测^[9]和南京航空航天大学 ComNet 红外人车航拍检测^[10]等 4 种不同场景的数据集验证了该方法的有效性。

1 锚点的研究进展

深度学习已成为目标检测任务的主流方法,基于锚点的目标检测器也广泛应用于各个领域^[1]。本章根据基于锚点模型的发展现状,进一步总结锚点机制的优缺点,并列举分析近几年改进的基于锚点检测器典型范式。

1.1 锚点机制的优势与局限

在目标检测中,可能出现极小、极大或者极端形状(如高窄型、宽扁型等)的物体,导致网络训练时对目标的识别和定位较困难^[11]。将多尺度的锚点应用于检测模型中,被证明是解决网络尺度问题的有效方法,但基于锚点系列模型存在以下 3 方面的问题^[1]:

①锚点策略存在尺寸固定、模型鲁棒性差等问题。锚点的相关超参数(尺寸大小、纵横比、IoU^[11]阈值)对预测结果的鲁棒性影响较为明显。由于超参数的设置与场景数据集密切相关,预置锚点的大小、比例在检测尺寸差异较大的物体时泛化能力弱。

②大量的锚点会导致网络参数增多,运算复杂度增大。由于基于锚点检测方法本质上是对图像进行密集采样,故导致产生冗余锚点。例如 DSSD (deconvolutional single shot detector)^[12]中锚点的设置超过 4 万个, RetinaNet^[13]超过 10 万。

③基于锚点的检测模型中只有少部分锚点会与真实目标重合,多数锚点为仅包含背景信息的负样本。这种情况会导致训练时正样本与负样本的比例失衡,易使模型的分类能力受负样本影响。

1.2 锚点的相关优化

针对当前锚点机制存在的问题,一些研究人员从尺度、数量等方面改进锚点,达到优化目的。在尺度方面,Cai 等^[14]在不同的特征层中设计不同尺度的检测器,低层的卷积网络用于检测包含小目标,高层的卷积网络用于检测大目标。Zhu 等^[15]提出一种基于步长缩减方法生成锚点的新策略,在检测包含小目标的高分辨特征图时防止漏检。在数量方面,Ke 等^[16]提出一种多锚点学习方法,基于交并比筛选部分优质的锚点,并构造属于固定目标的锚点袋,再结合分类评估锚点袋中正样本的训练损失。

对基于锚点机制的检测模型来说,无论是多尺度还是大数量,都会增加计算复杂度以及内存的消耗。故部分研究人员提出无锚点机制,其中典型的工作为 Law 等^[7]提出的基于关键点的无锚点检测模型

CornerNet, 使用特定点配对构造的预测框检测目标, 但使用特定点进行预测导致缺乏物体内部信息, 检测的准确性有待提升。

随着基于锚点检测和无锚点检测两种不同思路的发展, Zhang 等^[5]探究了二者之间产生差异的根本原因是标签分配策略, 该策略可进一步提升目标检测器的精度。故相较于无锚点检测模型, 针对基于锚点检测模型进行再优化的思想更为合理。

2 多尺度锚点 (MSA)

在基于锚点的检测模型中, 锚点的取值尤为重要, 检测之前要预设锚点的尺寸和比例, 检测时根据预设的锚点进行回归。最终得到预测框, 锚点的取值关乎检测模型的精度。

2.1 锚点的取值及作用

锚点为一批具有预定义位置、比例和长宽比的框。基于锚点的检测模型通常需要大量的锚点, 以确保与真值框 (ground truth) 有足够高的交并比。交并比是目标检测中的重要指标之一, 通过预测框和 Ground truth 间的交集与并集的比例进行计算, 常用于评价预测框的优劣, 其定义如下^[11]:

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (1)$$

式中: B^{gt} 为 Ground truth 的面积; B 为预测框的面积。在训练过程中, 当锚点与 Ground truth 有足够高的 IoU 时, 该锚点结合偏移后作为最终的预测框。该方法搜索范围小、易收敛且降低了训练难度。

一般来说, 当数据集类别单一且目标尺寸较集中时, 数据集样本量也相对较少^[17]。如红外数据集目标对象单一或采集不易, 导致数据集规模相对较小。此时, 采用 YOLO 系列的轻量级网络更为省时省力。典型的 YOLOv3-tiny 网络, 共有两个输出层, 分别为 13×13 和 26×26 , 每个输出层产生 3 个锚点, 共有 6 个锚点。这些锚点值通过聚类 MS-COCO^[6]数据集得到。由于 MS-COCO 数据集包含 80 类目标, 大到火车、轮船, 小到苹果、橘子, 目标尺寸差距较大, 不适用所有数据集, 因此在检测前需对锚点进行重新聚类。但 K-means 聚类算法采用迭代计算, 导致在数据集目标类型较单一或目标尺寸比较集中时, K-means 聚类产生的锚点大小差异不显著, 无法体现 YOLO 多尺度输出的优势。故本文将聚类产生的锚点再次根据数据集本身的特性进行线性的尺度缩放和拉伸, 进而获取更优化的多尺度锚点取值。

2.2 MSA 锚点优化

研究可知, 数据集目标尺寸过于单一时, 聚类方法产生新的锚点尺寸差异亦较小。故本文将聚类所产生的锚点再次根据数据集本身的特性进行线性尺度的缩放和拉伸来优化锚点的取值。涡核 (Vox) 数据集作为典型的工业现场数据集^[10], 来源于中国空气动力研究与发展中心进行的直升机旋翼流场测量试验。该数据集类别较为单一且目标尺寸较集中, Ground truth 纵横比约为 1, 优化实例如图 1 所示。

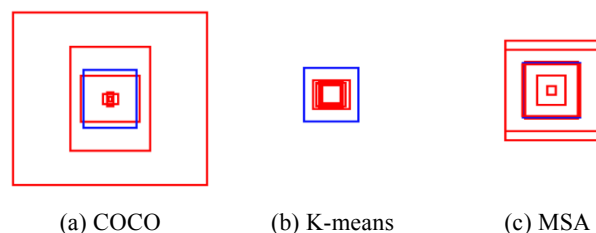


图 1 Vox 数据集上不同锚点与 Ground truth 对比

Fig.1 Comparison between Ground truth and different anchors on Vox data sets

注: 蓝色框为 Ground truth; 红色框为 6 个锚点

Note: The blue box is Ground truth; The red box shows six anchors

图 1(a)是由 MS-COCO 数据集聚类而得到的锚点, 其值与 Vox 数据集的目标相差较大, 导致面向特定目标时, 交并比几乎只由少数几个锚点来决定, 其余取值并不合理。图 1(b)是由 K-means 聚类所产生的锚点, 该方法产生的锚点受初始值影响较大且对噪音和异常点比较的敏感, 导致大多数 Ground truth 比得到的锚点尺寸略大, 在训练时对模型要求更高。图 1(c)为本文所提出的 MSA 方法产生的锚点, 既结合数据集的特性也保留了模型多尺度输出的优势, 提高了聚类对噪音和异常点的泛化能力。

由图 1 可知, 对于 YOLOv3-tiny 等锚点较少的轻量级模型来说, 锚点个数较少, 聚类后的锚点更无法体现多尺度的优点。因此, 本文引入多尺度优化, 具体来讲将最大的锚点再扩大相应的倍数, 将最小的锚点再缩小相应的倍数。具体如式(2)所示:

$$\begin{cases} x_s' = \alpha x_s \\ y_s' = \alpha y_s \end{cases}; \begin{cases} x_m' = \beta x_m \\ y_m' = \beta y_m \end{cases} \quad (2)$$

式中: x_s 和 y_s 为最小锚点的长和宽; x_m 和 y_m 为最大锚点的长和宽; x_s' 、 y_s' 、 x_m' 、 y_m' 为缩放后的锚点长宽值; α 为缩小倍数, β 为放大倍数, 实验观察二者基本满足 $\alpha \approx 2 - \beta$, 其值对不同类型的数据集可进行微调。建议当数据集目标尺寸较单一时 α 和 β 取差异较大的值, 如 0.4 与 1.6; 若数据集目标尺寸的尺度仍有变化, 则 α 和 β 可取差异较大的数值, 如 0.9 与 1.1。

总之,依本文方法,可进一步结合传统的网格寻优技术自动获取更有效的锚点值。

当锚点与 Ground truth 有足够高的交并比时,该锚点结合偏移即可更有效地得到预测框。不同锚点中, IoU 值最大的锚点与数据集的 Ground truth 对比如图 2 所示。

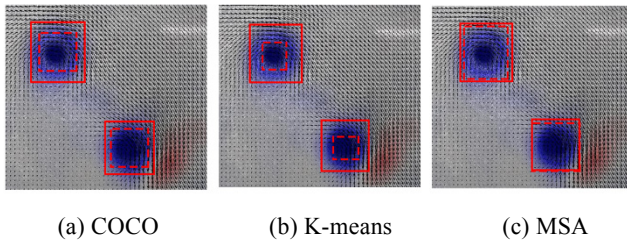


图 2 IoU 值最大的锚点与 Ground truth 的对比示例
Fig.2 Example of comparison between the anchor point with the maximum IoU value and ground truth

注: 实线为 Ground truth; 虚线为 IoU 值最大的锚点
Note: The solid line is Ground truth; Dotted line is the anchor point with the maximum IoU value

由图 1 和 2 可知,当数据集的目标尺寸较集中或异常点个数较多时,部分目标的 Ground truth 比聚类得到的锚点尺寸略大,导致训练时对模型要求较

高。因此将缩放后的值进行线性尺度的拉伸,可改善此问题,其计算公式如式(3)所示。

$$\begin{cases} x_i' = \frac{(x_i - x_s)}{(x_m - x_s)}(x_m' - x_s') + x_s' \\ y_i' = x_i' \cdot \frac{y_i}{x_i} \end{cases} \quad (3)$$

式中: x_i 与 y_i 为变化前的锚点值; x_i' 与 y_i' 为变化后的锚点。通过式(2)和式(3)可将聚类后尺度变化不显著的锚点进行优化,从而既体现 YOLO 算法多尺度输出的优势,又保留了对应场景数据集的特点。

3 数据集

数据集是衡量算法性能的基础,不同数据集会带来不同的挑战^[18]。红外图像对比度低、纹理特征不足,使检测面临更大的挑战。同时,红外现场实验不易,数据集难以获得,故大部分的红外数据集具有小样本、目标少等特点。为使训练得到的网络更加具有鲁棒性,且验证 MSA 优化方法对于不同特性的目标具有较强的泛化能力,选择 4 种不同场景的数据集进行多次测试。数据集的部分示例图像如图 3 所示。

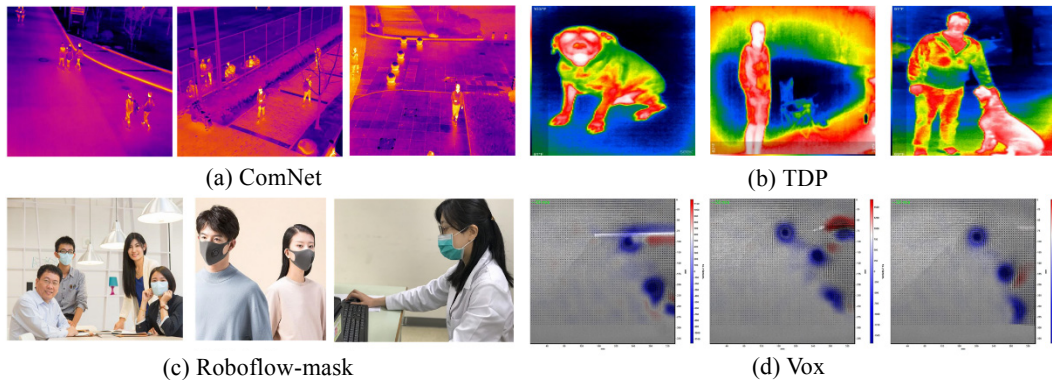


图 3 四类数据集的部分示例图像

Fig.3 Partial sample images of four types of datasets

3.1 南航 ComNet 红外人车检测数据集

ComNet 数据集来自南京航空航天大学 Minglei Li^[9]等所拍摄标注的红外人车检测图像数据集,由搭载在大疆 M600Pro 无人机上的红外 Vue Pro 热成像相机采集。数据集分别在白天和晚上捕获了包括校园和街道的各种常规交通场景,使用 Labelme 对这些图像进行手工标注,生成所需的训练数据和评估数据。ComNet 数据集包含 2975 张热红外图像,包括 4768 个行人实例和 3856 个车辆实例,即包括非刚体目标和刚体目标,图片大小为 640×512 。该数据集的部分示例图像如图 3(a)所示。

3.2 Roboflow TDP 数据集

Roboflow TDP (thermal dogs and people dataset) 数据集,包括人和狗在不同距离上的 203 张热红外

图像,由 Seek Compact XR Extra Range 红外相机拍摄。数据集中既有纵向又有横向目标,且保留一些不含目标的负样本图像。基于红外成像的人体和动物目标检测在安全、野生动物探测、狩猎和户外娱乐方面都具有广泛应用。其部分示例图像如图 3(b)所示。

3.3 Roboflow 口罩 (Roboflow-mask) 数据集

近期全球新型冠状病毒肆虐,在大型的公共场所对各种人员进行口罩检测能够减少人员之间交叉感染的风险。Roboflow-mask 口罩数据集包括 149 张戴口罩的人和未戴口罩的人。其部分示例图像如图 3(c)所示。

3.4 涡核 (Vox) 数据集

Vox 数据集^[10]是具有专业背景的涡核检测数据集,该数据集来源于中国空气动力研究与发展中心进

行的直升机旋翼流场测量试验。Vox 数据集的训练集包括 500 张涡核样本图像；测试集包括 106 张涡核样本图像。数据集类别较为单一且目标尺寸较集中，其 Ground truth 纵横比约为 1。其部分示例图像如图 3(d)所示。

4 实验与分析

本文实现和测试算法使用计算机硬件配置为 i7-9700K，GPU 版本为 NVIDIA GeForce RTX 2080Ti；计算机软件配置为 CUDA10.2，PyTorch1.10，Python3.8。

4.1 精度测试

表 1 不同 Anchor 在四类数据集上的对比结果

Table 1 Comparison results of different anchors on four data sets

Anchor	ComNet/(%)	TDP/(%)	Roboflow-mask(%)	Vox/(%)
COCO	96.86	90.03	66.20	89.36
K-means	97.12(+0.26)	90.90(+0.87)	66.60(+0.40)	91.03(+1.67)
MSA	97.24(+0.38)	91.74(+1.77)	67.00(+0.80)	91.44(+2.08)

由表 1 可知，改进后的锚点在 4 个数据集上的测试结果 mAP@0.5 指标均有所提升，针对不同特点的数据集缩放倍数只需微调即可有效提升目标检测的精度。其中，Roboflow-mask 数据集样本较少且两个类别的 Ground truth 数量差异较大，故 mAP@0.5 相对较低，但使用改进锚点后检测精度仍有提高。若数据集中未正确检测样本难例较多，例如模糊目标、较暗目标等难例，可通过对难例适当引入旋转、裁剪、缩放等数据增强^[20]方法进一步挖掘锚点的多尺度信息，再利用本方法进行优化。

4.2 消融实验

为了验证不同取值的 α 和 β 对精度所产生的影响，本文在 TDP 数据集上针对 α 和 β 的取值进行了消融实验，结果如图 4 所示。

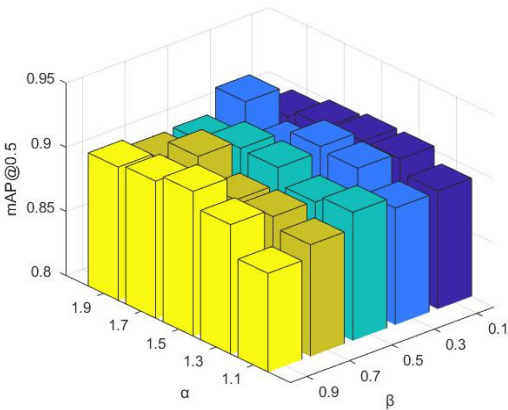


图 4 TDP 数据集中锚点缩放的消融测试

Fig.4 Ablative test of anchor point scaling on TDP dataset

基于 IoU 阈值为 0.5 的检测平均精度均值(mean average precision, mAP)已成为多年来目标检测问题最重要的实际度量标准之一^[19]。本文使用 YOLOv3-tiny 网络在 4 种不同的数据集上进行测试，mAP@0.5 结果如表 1 所示。其中 Vox 数据集和 ComNet 数据集由实验得出放大倍数 $\alpha=1.5$ ，缩小倍数 $\beta=0.5$ 时所得结果较为理想。TDP 数据集部分 Ground truth 的纵横比差距较大，本文针对缩放倍数进行多次实验寻优并微调，具体见图 4。Roboflow-mask 数据集图片较少且目标尺寸变化较为显著，故微调放大倍数 $\alpha=1.2$ ，缩小倍数 $\beta=0.5$ 时所得精度有所提高。

由图 4 可知，对于 TDP 数据集，放大倍数 $\alpha=1.9$ ，缩小倍数 $\beta=0.3$ 时 mAP@0.5 较高，可能由于该数据集的部分 Ground truth 的纵横比差距较大所造成。但对于大多数数据集来说放大倍数 $\alpha=1.5$ ，缩小倍数 $\beta=0.5$ 所得结果较为理想。此外，基于 3 种典型轻量级网络 YOLOv3-tiny, YOLOv4-tiny 和 YOLOv5s 以及大型网络 YOLOv3 在 Vox 数据集上测试验证了该方法的有效性。利用相同的锚点在同一网络上分别进行了 3 次测试并取 mAP@0.5 平均值，其结果如表 2 所示。

表 2 Vox 数据集在不同网络的表现

Table 2 Performance of different networks on Vox datasets

Network	Anchor	MAP@0.5(%)
YOLOv3-tiny	COCO	89.36
	K-means	91.03(+1.67)
	MSA	91.44(+2.08)
YOLOv4-tiny	COCO	90.16
	K-means	90.35(+0.19)
	MSA	91.56(+1.4)
YOLOv3	COCO	92.69
	K-means	93.32(+0.63)
	MSA	93.59(+0.90)
YOLOv5s	COCO	94.88
	K-means	95.02(+0.14)
	YOLOv5s-Aut	94.82(−0.06)
	MSA	95.63(+0.75)

由表2可知,所提出的多尺度优化锚值MSA方法适配多种网络结构, $mAP@0.5$ 指标都有提升,但随着网络结构变得复杂,精度的提升也有所减少,一个可能的原因是在网络变得复杂时,模型表达能力提升,此时锚点对网络的影响有所降低。其中, YOLO v3-tiny 和 YOLO v4-tiny 需6个锚点,而 YOLO v3 和 YOLO v5s 需9个锚点。同时, YOLO v5 作为 YOLO 家族的最新代表,其亮点之一是模型训练过程中可动态学习优化锚点值,可在一定程度上加速模型收敛。YOLO v5s-Aut 为 YOLO v5 模型在 Vox 数据集上开启自动学习优化锚点值的测试结果。相比而言,本文提出的尺度变化的锚值优化方法更适用于 YOLO v3-tiny 和 YOLO v4-tiny 等锚点较少的典型轻量级检测网络。

5 结语

本文提出一种基于多尺度优化的目标检测锚点构造方法,针对场景数据集的特性将生成的锚点进行线性缩放和拉伸,解决了聚类产生的锚点大小差异不显著且泛化能力差与无法体现 YOLO 多尺度输出优势的问题。利用该方法在多个典型 YOLO 模型和多个场景数据集上进行了测试。数据难以获得、数据获取成本高、红外数据集样本小、目标少,该方法针对此类数据的优化效果明显,可显著提高检测精度。另外,本方法应用在训练的预处理阶段,不增加模型推理时间。

参考文献:

- [1] 伏轩仪, 张奎景, 梁文科, 等. 锚点机制在目标检测领域的发展综述[J]. 计算机科学与探索, 2022, 16(4): 791-805.
FU Xuanyi, ZHANG Luanjing, LIANG Wenke, et al. Review on the development of anchor mechanism in object detection[J]. *Journal of Frontiers of Computer Science and Technology*, 2022, 16(4): 791-805.
- [2] 易诗, 周思尧, 沈炼, 等. 基于增强型轻量级网络的车载热成像目标检测方法[J]. 红外技术, 2021, 43(3): 237-245.
YI Shi, ZHOU Siyao, SHEN Lian, et al. Vehicle-based thermal imaging object detection method based on enhanced lightweight network[J]. *Infrared Technology*, 2021, 43(3): 237-245.
- [3] 顾佼佼, 李炳臻, 刘克, 等. 基于改进 Faster R-CNN 的红外舰船目标检测算法[J]. 红外技术, 2021, 43(2): 170-178.
GU Jiaojiao, LI Bingzhen, LIU Ke, et al. Infrared ship object detection algorithm based on improved faster R-CNN[J]. *Infrared Technology*, 2021, 43(2): 170-178.
- [4] 邵延华, 张铎, 楚红雨, 等. 基于深度学习的 YOLO 目标检测综述[J]. 电子与信息学报, 2022, 44(10): 3697-3708.
SHAO Yanhua, ZHANG Duo, CHU Hongyu, et al. A review of YOLO object detection based on deep learning[J]. *Journal of Electronics & Information Technology*, 2022, 44(10): 3697-3708.
- [5] ZHANG S, CHI C, YAO Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 9759-9768.
- [6] LIN T Y, Maire M, Belongie S, et al. Microsoft coco: common objects in context[C]//*European Conference on Computer Vision*, 2014: 740-755.
- [7] LAW H, DENG J. Cornernet: detecting objects as paired keypoints[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 765-781.
- [8] YUAN C, YANG H. Research on K-value selection method of K-means clustering algorithm[J]. *Multidisciplinary Scientific Journal*, 2019, 2(2): 226-235.
- [9] LI M, ZHAO X, LI J, et al. ComNet: combinational neural network for object detection in UAV-Borne thermal images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(8): 6662-6673.
- [10] LUO Y, SHAO Y, CHU H, et al. CNN-based blade tip vortex region detection in flow field[C]//*Eleventh International Conference on Graphics and Image Processing*, 2020, 11373: 113730P.
- [11] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//*Association for the Advance of Artificial Intelligence(AAAI 2020)*, 2020: 12993-13000.
- [12] FU C Y, LIU W, Ranga A, et al. Dssd: Deconvolutional single shot detector[J/OL]. arXiv preprint arXiv:1701.06659, 2017.
- [13] LIN T, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 42(2): 318-327.
- [14] CAI Z, FAN Q, FERIS R S, et al. A unified multi-scale deep convolutional neural network for fast object detection[C]//*Proceedings of the 14th European Conference on Computer Vision*, 2016: 354-370.
- [15] ZHU C, TAO R, LU K, et al. Seeing small faces from robust anchor's perspective[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 5127-5136.
- [16] KE W, ZHANG T, HUANG Z, et al. Multiple anchor learning for visual object detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 10206-10215.
- [17] Ramachandran P, Zoph B, Le Q V. Searching for activation functions[J/OL]. arXiv preprint arXiv:1710.05941, 2017.
- [18] KONG T, SUN F, LIU H, et al. Foveabox: beyond anchor-based object detection[J]. *IEEE Transactions on Image Processing*, 2020, 29: 7389-7398.
- [19] ZOU Zhengxia, SHI Zhenwei, GUO Yuhong, et al. Object detection in 20 years: a survey[J/OL]. arXiv preprint arXiv: 1905.05055, 2019.
- [20] Zoph B, Cubuk E D, Ghiasi G, et al. Learning data augmentation strategies for object detection[C]//*European Conference on Computer Vision*, 2020: 566-583.