

DOI: 10.13382/j.jemi.B2205954

# 注意力机制优化的全尺寸目标检测方法\*

刘鹏 毕誉轩 张天翼 史佳霖

(长春理工大学电子信息工程学院 长春 130022)

**摘要:**针对现有目标检测算法全尺寸目标检测精度低的问题,提出了一种改进的基于YOLOv3模型的全尺寸目标检测算法。该方法设计了一种全新的通道自适应递归FPN网络架构,提出了一种基于通道注意力的递归金字塔模型,提高了YOLOv3的特征提取能力和不同尺度目标的检测能力。同时在训练过程中引入损失函数转换,解决了训练过程中动态参数不优化的问题。与其他主流目标检测算法相比,本文提出的改进模型在小尺寸目标、大尺寸目标与复杂背景多尺寸目标的检测精度分别提高了5.6%、2.6%、1.6%。实验结果表明,本文提出的方法检测精度显著提升。

**关键词:**目标检测;YOLOv3;注意力机制;FPN

**中图分类号:**TP391.9 **文献标识码:**A **国家标准学科分类代码:**520.6040

## Full-size object detection method optimized by attention mechanism

Liu Peng Bi Yuxuan Zhang Tianyi Shi Jialin

(School of Electronic Information Engineering, Changchun University of Science and Technology, Changchun 130022, China)

**Abstract:** Aiming at the problem that existing object detection algorithms have low accuracy in full-size object detection, this paper proposes an improved full-size object detection algorithm based on the YOLOv3 model. In the method, a new adaptive recursive FPN network architecture is designed, and a recursive pyramid model based on channel attention is proposed to improve the feature extraction ability of YOLOv3 and the detection ability of objects at different scales. At the same time, loss function transformation is introduced in the training process to solve the problem of dynamic parameters that is not being optimized in the training process. Compared with other mainstream object detection algorithms, the accuracy of small-size objects, large-size objects and multi-size objects with complex backgrounds respectively improved by 5.6%, 2.6%, and 1.6%. Experimental results show that the detection accuracy of the proposed method is significantly improved.

**Keywords:** object detection; YOLOv3; attention mechanism; FPN

## 0 引言

随着计算机视觉的快速发展,目标检测技术被广泛应用于自动驾驶、视频监控、人脸识别等领域<sup>[1-3]</sup>。目标检测技术主要包括两个任务—目标识别和定位。首先,目标检测必须解决识别问题,将目标信息从背景信息中分离出来,确定准确的目标标签和预测值。其次,要解决定位问题,通过精确的锚框选择不同的目标。近年来,深度学习的快速发展为现有的目标检测技术提供了更显著

的性能提升<sup>[4]</sup>。

在过去的研究中,学者们对目标检测已经做了大量的研究。目前常见的目标检测方法有Fast RCNN<sup>[5]</sup>、SSD<sup>[6]</sup>和YOLO<sup>[7-11]</sup>。首先提出的Fast RCNN是一种基于VGG骨干网络的特征提取目标检测算法,该算法大大提高了目标检测的精度,但基于RPN的特征处理网络架构会产生大量的重复计算过程,导致检测速度缓慢。因此,后续提出的YOLOv1<sup>[7]</sup>和SSD算法放弃了RPN网络架构,采用端到端网络结构,同时实现了目标的识别和定位,满足了目标检测的实时性要求。在YOLOv1的基础

收稿日期:2022-11-02 Received Date: 2022-11-02

\* 基金项目:吉林省科技发展计划项目(20210201021GX)资助

上, YOLOv2<sup>[8]</sup> 提出了 Darknet-19 特征提取网络架构, 提高了检测精度, 实现了更快更准确的目标检测。但该算法对小目标和相邻多目标的检测精度不高。因此, Redmon J 在 YOLOv2 的基础上, 提出了目标检测算法 YOLOv3<sup>[9]</sup>。通过 Darknet-53 网络架构和 FPN 特征金字塔网络, 实现了目标特征提取和多尺度目标检测。2020 年, Bochkovskiy 等<sup>[10]</sup> 提出了 YOLOv4。YOLOv4 的骨干网为 CSPDarknet。YOLOv4 通过 CIOU\_LOSS 提高了检测精度。在 YOLOv4 的基础上, YOLOv5<sup>[11]</sup> 采用 PAN 结构提高检测精度。虽然提高了检测精度, 但降低了检测效率。由于现有检测算法模型针对不同尺寸大小目标有着不同的检测表现, 本文提出了一种全尺寸目标检测算法。相较于目前最先进的 YOLOv5 与 YOLOv7 检测算法模型, 成熟的 YOLOv3 模型检测性能良好, 卷积层数与网络规模更小, 更适用于硬件计算能力不高以及嵌入式设备计算能力有限的开发环境中。因此, 本文提出了一种改进的基于 YOLOv3 的全尺寸目标检测算法模型。本文主要做了两方面改进: 首先提出了一种全新的基于 FPN<sup>[12]</sup> 的递归特征金字塔网络架构, 该架构结合通道注意力机制, 提高了 YOLOv3 模型的检测精度。其次, 在训练过程中使用全新的损失函数转换, 解决了训练时预测框与真实框重合后动态参数无法继续优化的问题。通过实验数据证明改进的算法模型不仅提高了检测效率, 而且大大

提高了检测精度。

## 1 原始算法介绍

### 1.1 YOLOv3

在 YOLO 算法被提出之前, 最常用的目标检测算法是两级 R-CNN 序列算法, 但两级检测的特点使其无法满足检测速度的要求。为了满足目标的实时检测, YOLO 系列算法应运而生。YOLO 是一种单阶段目标检测算法。其核心思想是将原始的目标检测问题转化为回归问题。将整个图像作为输入, 通过网络得到边界框的位置和目标的类别。

YOLOv3 是 YOLO 系列的第 3 代改进算法。在特征提取部分采用 Darknet-53 网络结构替代原有的特征提取网络, 采用特征金字塔网络结构实现多尺度检测, 用 Logistic 回归代替 Softmax 分类方法。YOLOv3 通过 FPN 的多级检测结构, 解决了小物体难以检测、检测精度低等缺点。算法设计了 3 种不同尺度的网络输出, 每一层通过位置信息进行定位, 使用语义信息进行分类, 检测不同尺度的目标。特征提取网络架构与 Darknet-19 相比, Darknet-53 的网络更深, GPU 利用率更高。YOLOv3 的总体网络结构如图 1 所示。

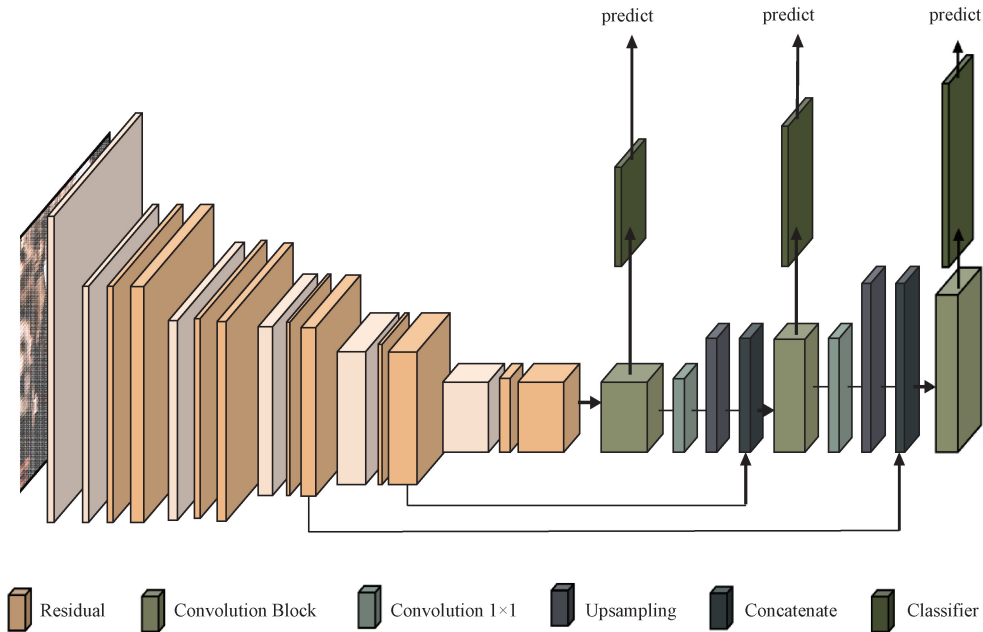


图 1 YOLOv3 网络结构  
Fig. 1 YOLOv3 architecture

### 1.2 FPN

FPN 将 3 个不同尺度卷积提取到的特征图构造特征金字塔结构, 通过金字塔进行简单的网络连接变换, 大

大大提高了小物体检测的精度。位于特征金字塔底部的特征图接近输入图像, 包含更多的位置信息, 但由于卷积层数较少, 包含的语义信息较少。高级特征映射图远离输

入图像,由于卷积层数较多,包含的语义信息较多,但位置信息较少。如图2所示,FPN由一条自下而上的连接(将语义信息分布到各层)和一条自上而下的连接(将位置信息分布到各层)之间的水平连接组成。通过对 $1 \times 1$ 卷积核进行水平连接补充,融合高层次的语义信息和底层的位置信息,得到本文最需要的目标特征,提高目标检测精度。

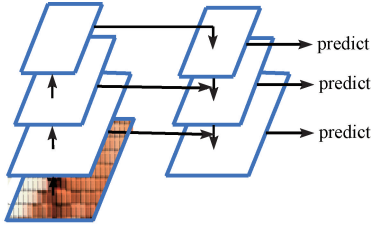


图2 FPN网络结构  
Fig.2 FPN architecture

### 1.3 注意力机制

近年来,注意力机制在图像处理领域取得了重要突破,并被证明有利于提高模型的性能<sup>[13]</sup>。在目标检测领域,注意力机制的应用有助于网络获取最重要的特征以及捕获特征之间的空间相关性。注意力机制专注于局部显著特征并会随着计算机任务的变化而改变关注的区域。目前主要有软注意力和强注意力。软注意力是通过学习通道之间的关系而产生的,通过神经网络前向传播和后向反馈计算梯度并学习计算通道权重<sup>[14]</sup>。强注意力是一个随机的预测过程,强调动态变化。在本文中,将软注意力添加到网络结构中,通过计算赋予特征最需要的通道最高的权值,以提高算法目标检测的精度。

## 2 算法改进

现有目标检测模型针对单一尺寸目标有着极高的检测精度,但对于包含大量复杂尺寸目标的待检测图像性能不佳。为了获得一种检测精度更高的全尺寸目标检测模型,本文提出了一种改进的基于YOLOv3的目标检测模型。该模型采用了一种基于注意机制的递归FPN网络结构,该结构由通道注意机制和特征金字塔递归层组成;同时提出了T-IOU损失函数,用该损失函数代替原模型中的GIOU损失函数。本节详细介绍了算法如何改进。

### 2.1 FPN

FPN提供了一种自顶向下的路径来融合多尺度特征图中的特征,利用高阶网络提取的语义信息和低阶网络提取的细节特征信息来预测多尺度对象。FPN的网络输出公式为:

$$f_i = F_i(f_{i+1}, X_i), X_i = B_i(X_{i-1}) \quad (1)$$

其中, $B_i$ 表示骨干网络从下至上路径的第*i*阶段, $X_i$ 表示第*i*阶段的输出, $F_i$ 表示对应的从上至下路径额外反馈的操作, $f_i$ 表示输出,其中*i*的取值范围为 $1 \sim S$ , $S$ 表示骨干网络中的阶段数,其中 $f_{(s+1)} = 0$ 。

### 2.2 基于通道自适应FPN递归层的YOLOv3

本文针对传统的YOLOv3算法,提出了一种全新的通道自适应FPN递归层。在网络结构中引入了ECA-Block通道注意力模块<sup>[15]</sup>,递归层中的通道自适应就是通过ECA-Block来实现的。ECA-Block模块融合了每层局部接受场的空间和通道信息,使网络能够构建信息特征,重点是自适应地改变通道权重关系。特征图通道对应的权值增大,对最终输出的影响也会增大。当权值较小时,通道的特征映射对应的值较小,对最终输出的影响较小,因此应用了注意力机制后比普通的卷积层更有效。

ECA-Block结构如图3所示。卷积块提取的图像特征输出为 $U \in R^{W \times H \times C}$ ,其中 $W, H, C$ 分别为宽度、高度和通道数。所有空间信息通过全局平均池化压缩到统计信息中,统计信息 $Z \in R^C$ 由空间维 $H \times W$ 收缩计算的卷积输出产生,所以 $Z$ 的第*c*元素(即 $Z^{1 \times 1 \times C}$ )如下公式平均池化生成:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (2)$$

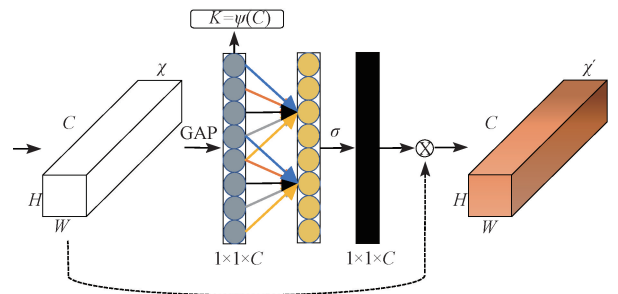


图3 ECA-Block网络结构  
Fig.3 ECA-Block architecture

得到聚合特征 $Z^{1 \times 1 \times C}$ 后,通过卷积核大小为*K*的一维卷积生成通道权值。一维卷积的初始特征矩阵如式(4)所示,其中一维卷积核*K*由信道维数*C*自适应确定,生成卷积核*K*的具体公式如下:

$$k = \psi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{odd} \quad (3)$$

其中, $K$ 为卷积核的大小, $\lfloor t \rfloor_{odd}$ 表示距离*t*最近的奇数值, $C$ 为通道的维数, $\gamma$ 与*b*分别为固定值2和1。

$$\begin{bmatrix} w^{1,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w^{C,C} \end{bmatrix} \quad (4)$$

ECA-Block通道的注意力模块计算权重后,每个通

道都有一个与之对应的数字,可以认定为通道的自适应权重(通道权重表示通道对全局特征提取的影响)。

自适应 FPN 递归层结构如图 4 所示。自适应 FPN 递归层以原有的 FPN 金字塔网络结构为基础,将 FPN 各层从上到下的附加反馈连接到自下而上的 Darknet-53 骨干特征提取层中,并通过自适应的权重模块计算各通道的权重来分配通道的权重,该网络架构将对目标图像进行两次的特征提取,递归提取增强了特征检测的强度和精度,输出特征依赖于前面步骤的输出。 $R_i$  表示自适应 FPN 递归层连接到骨干网之前的特征转换,输出特征

$f_i$  为:

$$f_i = F_i(f_{i+1}, x_i), x_i = u_i(x_{i-1}, R_i(f_i)) \quad (5)$$

式中:  $u_i$  表示的是骨干提取网络中的第  $i$  层卷积运算,  $x_i$  为卷积操作后的特征,  $F_i$  代表 FPN 对  $x_i$  的第  $i$  次运算。递归输出特征可以展开为一个输出顺序网络,即当  $\forall i = 1, \dots, S, t = 1, \dots, T$  时:

$$f_i^t = F_i^t(f_{i+1}^t, x_i^t), x_i^t = u_i^t(x_{i-1}^t, R_i^t(f_i^{t-1})) \quad (6)$$

其中,  $t$  表示递归特征提取的步数,本文  $t$  的取值为 2,  $f_i^0 = 0$ 。

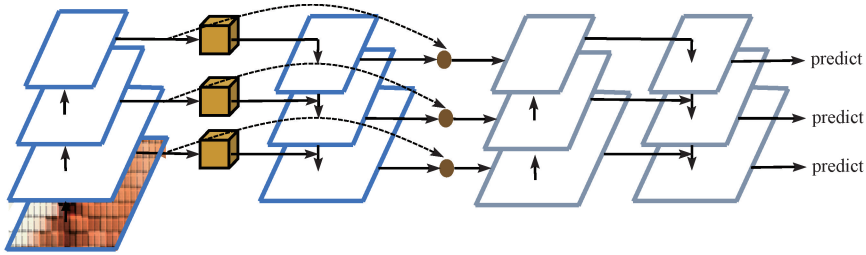


图 4 通道自适应 FPN 递归网络结构

Fig. 4 Adaptive FPN recursion layer architecture

完整的改进算法模型结构如图 5 所示。模型的骨干特征提取网络仍采用原有的 Darknet-53 结构,特征提取强化网络使用本文提出的通道自适应 FPN 递归层取代原有的 FPN。从骨干网中提取图像后,开始提取特征。首次特征提取过程的输出结果经过 4 次特征卷积,经过

通道注意模块自适应处理后与对应骨干网络的输出相结合作为附加反馈。经过组合特征层 5 次卷积处理后,通道权值再次进行自适应处理。经过处理后,通过 YOLOHead 得到预测结果和检测对象。

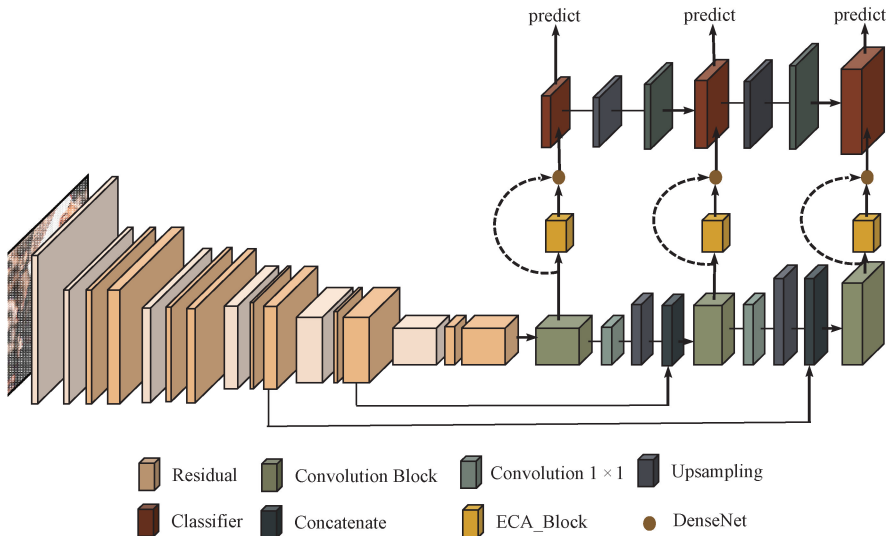


图 5 本文提出算法结构

Fig. 5 The proposed method architecture

### 2.3 损失函数

原 YOLOv3 模型中使用 GIOU 损失函数<sup>[9]</sup>作为边界

框回归损失函数来计算真实框与预测框之间的差异大小。由于 GIOU 损失函数在预测框包含于真实框的情况

时,参数优化就会暂停无法继续下去,使检测精度无法进一步提升,所以本文提出一种损失函数转换方案 T-IOU。

T-IOU 通过判断预测框与真实框之间的位置关系来自适应替换边界框回归损失函数。当两框不存在包含的情况时损失函数使用 GIOU,出现两框包含的情况就将损失函数替换为 CIOU<sup>[16]</sup>,通过两框中心点以及宽高纵横比之间的关系继续优化下去。如图 6(a)所示,预测框为  $B_p$ ,真实框为  $B_g$ ,两框相交区域为  $L$ ,两框组合区域为  $U$ ,可以包含两框的最小矩形区域为  $A$ ,则当前损失函数计算公式为:

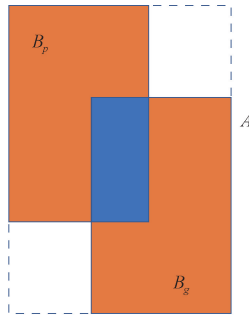
$$L_{GIOU} = 1 - GIOU \quad (7)$$

其中 GIOU 通过以下公式计算:

$$GIOU = IOU - \frac{A - U}{A} \quad (8)$$

式中:  $\frac{A - U}{A}$  为包含两框的最小矩形区域  $A$  减去两框组合区域  $U$  与最小矩形区域  $A$  的比值,这个比值在一定程度上可以区分预测框与真实框具有相同相交区域但不同相交模式的状态,即使训练过程中出现预测框与真实框并不存在相交区域的情况下,模型参数也可以继续优化下去。

虽然 GIOU 提升了损失优化性能,但是当预测框与真实框完全相交时,无法优化两框的位置关系。因此,当 GIOU 损失优化暂停更新时,转而使用 CIOU 损失函数。



(a) 两框相交互不包含  
(a) The intersection of two boxes does not contain

CIOU 损失函数通过预测框与真实框之间的中心点距离以及两框的纵横比来计算并优化位置关系。

CIOU 损失函数通过以下公式计算:

$$CIOU = IOU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v \quad (9)$$

式中:  $\rho^2$  表示两框中心点距离的平方,  $c^2$  表示包含预测框与真实框的最小矩形区域的对角线长度的平方,  $\alpha$  由式(10)计算。

$$\alpha = \frac{v}{(1 - IOU) + v'} \quad (10)$$

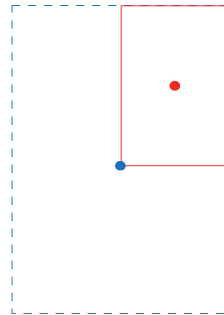
其中,参数  $v$  代表的是预测框与真实框之间纵横比的一致性,定义如式(11)所示。

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w^p}{h^p} \right)^2 \quad (11)$$

其中,  $w^{gt}$  与  $h^{gt}$  分别代表真实框的宽与高,  $w^p$  与  $h^p$  代表预测框的宽与高。所以 CIOU 的完整定义为:

$$LCIOU = 1 - CIOU = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (12)$$

如图 6(b)所示,实线预测框完全包含于虚线真实框,宽与高的纵横比均为 1 : 2,经过计算, GIOU 为 0.25, CIOU 为 0.187 5, LCIOU 为 0.812 5,由此可以证明,本文提出的 T-IOU 损失转换函数更准确的预估了预测框与真实框之间的匹配度,从而有效地提高了检测精度。



(b) 预测框包含于真实框  
(b) The prediction box is contained within the true box

图 6 预测框与真实框相交情况

Fig. 6 The intersection between the prediction box and the real box

### 3 实验和讨论

为了验证本文改进模型的性能,将小尺寸目标、大尺寸目标与复杂多尺寸目标的图像数据集进行训练。算法训练迭代次数为 20 000,初始学习率为 0.004,动量为 0.9,采用随机梯度下降(SGD)优化算法。最后,通过比较算法的 P-R 曲线、检测效果和 mAP 值,验证算法的检测性能。

P-R 曲线与 mAP 值均为目标检测模型性能的评价指标。P-R 曲线代表的是精确率与召回率的关系,其中  $P$  代表的是精确率,  $R$  为召回率。对于目标检测类的二元分类问题,可将样本分为正确的正例( $TP$ )、错误的正例( $FP$ )、正确的反例( $TN$ )以及错误的反例( $FN$ )。精确率  $P$  与召回率  $R$  定义如下:

$$P = \frac{TP}{TP + FP} \quad (13)$$

$$R = \frac{TP}{TP + FN} \tag{14}$$

以精确率  $P$  作为纵轴,召回率  $R$  为横轴,可以得到模型的性能 P-R 曲线。若一个模型的曲线被另一条曲线完全包含,可以认为后者性能优于前者;当两个模型的 P-R 曲线相交时,可以根据平衡点 (BEP) 来完成性能比较。平衡点是当  $P = R$  时的取值,平衡点 (BEP) 越大,可以认为该模型的性能越好。

mAP 是数据集中所有类的平均精度 (AP) 的均值, AP 的计算公式如下:

$$AP = \frac{1}{11} \sum_{r \in \{0,0.1,\dots,1\}} \rho_{interp}(r) \tag{15}$$

其中:

$$\rho_{interp} = \max_{r, r \geq r_n} \rho(\tilde{r}) \tag{16}$$

mAP 计算公式如下:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \tag{17}$$

mAP 值越高,可以认为模型的性能表现越好。

### 3.1 小目标检测效果

为了证明本文所提出的模型在小目标检测精度上的优越性,选取了 3 种主流算法 YOLOv3、CMTGAN<sup>[17]</sup> 和 Focud-Detctet<sup>[18]</sup> 进行检测结果比较。

根据小目标的定义,本文将相对面积小于 0.3% 的目标认定为小目标。传统的数据集以大目标图像为主,小目标图像占比很少。因此,本文选择了专注于小目标检测的 VisDrone-2019 数据集作为小目标训练数据集。小目标数据集训练后,将自采集小目标图像作为待测图片进行检测效果对比,对比结果如图 7 所示。

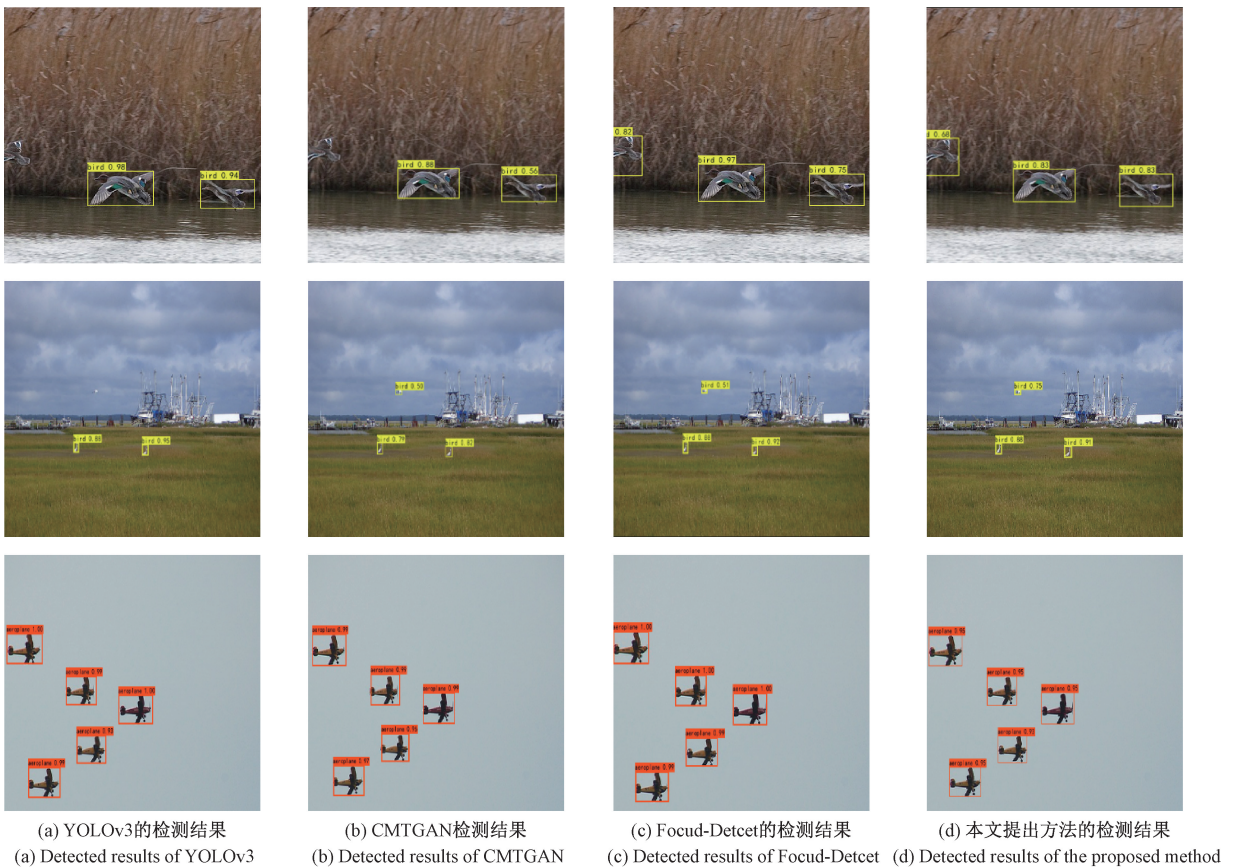


图 7 不同方法的小物体检测结果

Fig. 7 Small object detection results of different methods

图 7 (a)、(b)、(c)、(d) 分别显示了 YOLOv3、CMTGAN、Focud-Detctet 和本文方法的小目标检测结果。如图所示, YOLOv3 算法对小目标的检测精度并不高。虽然 CMTGAN 和 Focud-Detctet 提高了对小目标的检测精度,但对目标的瞄框标定并不精准。本文提出的方法具有较高的小目标检测精度和较高的瞄框标定精度,这表

明改进算法与其他算法相比具有较好的小目标检测性能。

在训练和测试过程中,算法模型的 P-R 曲线如图 8 所示。各模型的平衡点值分别为 0.678, 0.732, 0.765 与 0.823, 通过平衡点值可以看出本文所提改进算法相较于其他算法平衡点值最大,模型性能最好。

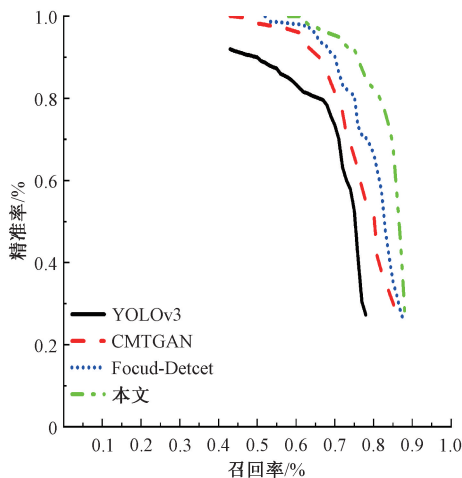


图8 几种小目标检测模型的 P-R 曲线  
Fig. 8 P-R curves of small object detection models in several categories

表1给出了计算后各算法的平均检测精度。从这些信息中,可以看出本文算法在小目标检测方面相对于其他算法的优越性,相较于小目标检测精度最高的检测模型,精度值提高了5.6%。

表1 各算法小目标检测 mAP 值比较  
Table 1 Comparison of mAP values of small object detection among algorithms

算法模型	AP@0.5	AP@[0.5,0.95]
YOLOv3	57.9%	33.0%
CMTGAN	60.2%	37.3%
Focud-Detctet	63.8%	39.7%
本文	67.2%	45.3%

### 3.2 大目标检测效果

本文选择了3种检测精度高的大目标检测算法与本文算法进行比较,分别是YOLOv5, SUCNN<sup>[19]</sup>和FMSSD<sup>[20]</sup>。

根据大目标定义,将相对面积大于3%的目标认定为大尺寸目标。为了验证模型检测性能,选取大尺寸目标图像较多的PASCAL VOC2007作为训练集,并在训练结束后使用不在训练集中的自采集待测图像进行检测效果对比,检测对比结果如图9所示。

图9(a)、(b)、(c)、(d)分别为YOLOv5、SUCNN、FMSSD和本文方法的大目标检测结果。从图中可以看出,虽然本文选择的3种算法在检测大目标时具有较高的准确率,但是在在大目标附近存在被遮挡了部分特征的小目标无法被检测到。同时,由于网络架构顶部缺乏位置信息,目标瞄框不准确。相比之下,本文算法既保证了大目标的检测精度,又兼顾了被大目标遮挡下的小目标

检测精度。

几种模型的P-R曲线如图10所示,4种模型的平衡点值分别为0.759,0.763,0.763与0.789,本文提出模型的大尺寸目标检测平衡点值最高,即性能表现最优。表2为计算后的各算法平均检测精度,从表中可以看出本文提出算法精度提高了2.6%。由此可见,自适应FPN递归网络架构可以有效解决长期以来难以解决的大尺寸目标检测问题。

表2 各算法大目标 mAP 值比较  
Table 2 Comparison of mAP values of large object detection among algorithms

算法模型	AP@0.5	AP@[0.5,0.95]
YOLOv5	64.9%	38.9%
SUCNN	65.2%	39.6%
FMSSD	65.8%	39.7%
本文	66.2%	42.3%

### 3.3 复杂多目标检测效果

为了验证模型复杂多目标检测效果,本文使用MS-COCO数据集作为训练集。MS-COCO数据集中包含了大量背景复杂、目标数量多的图像。本文选择YOLOv5-R<sup>[21]</sup>、DDH-YOLOv5<sup>[22]</sup>和YOLOv3-dense<sup>[23]</sup>3种检测精度高的算法与本文提出算法进行比较。训练结束后使用自采集待测图片进行效果对比,检测对比结果如图11所示。

图11(a)、(b)、(c)、(d)分别为YOLOv5-R、DDH-YOLOv5、YOLOv3-dense以及本文算法对复杂多目标的检测结果。从对比图中可以看出在复杂环境下的多目标检测中,本文算法比其他算法具有更高的检测精度。本文提出的自适应FPN递归网络结构更加关注被复杂环境遮挡部分特征的目标,从而对隐藏在后面被遮挡的目标进行检测,提高了算法复杂背景下多目标的检测精度。

几个模型的P-R曲线如图12所示,各模型平衡点值分别为0.763,0.752,0.757与0.778。本文提出改进模型的复杂背景多目标检测平衡点值最高,性能表现最好。

表3为各算法的mAP值,表内结果可以看出,本文提出的改进算法在复杂背景多目标检测方面的性能优于其他算法,检测精度相较于其他性能最优的模型提高了1.6%。

表3 各算法 mAP 值比较  
Table 3 Comparison of mAP values of each algorithms

算法模型	AP@0.5	AP@[0.5,0.95]
YOLOv3-dense	61.7%	38.6%
YOLOv5-R	61.9%	39.3%
DDH-YOLOv5	61.8%	39.7%
本文	63.2%	41.3%



(a) YOLOv5的检测结果 (b) SUCNN的检测结果 (c) FMSSD检测结果 (d) 建议方法的检测结果  
 (a) Detected results of YOLOv5 (b) Detected results of SUCNN (c) Detected results of FMSSD (d) Detected results of the proposed method

图 9 不同方法的大目标检测结果

Fig. 9 large object detection results of different methods

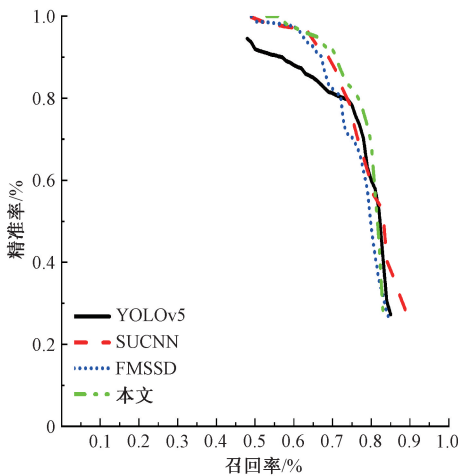


图 10 几种大目标检测模型的 P-R 曲线

Fig. 10 P-R curves of large object detection models in several categories

从以上 3 组实验结果可以看出,本文提出的改进算法在小目标检测、大目标检测和复杂多目标检测方面都优于目前主流的检测算法。因此,可以认为本文提出的

通道自适应递归 YOLOv3 算法模型显著提高了全尺度的目标检测精度。

### 4 结 论

为了提高全尺寸目标检测精度,本文设计了一种通道自适应 FPN 递归网络结构,该网络结构基于通道注意机制和递归特征提取,将该网络结构与 YOLOv3 相结合,提高了目标检测的精度。同时本文使用了一种全新的替换损失函数(T-IOU),解决了预测框包含于真实框时模型参数不优化的问题。相较于各尺寸目标中精度最高的检测算法模型,本文提出算法在小目标检测领域精度提高了 5.6%,大目标精度提高了 2.6%,复杂背景多尺寸目标精度提高了 1.6%。定性和定量实验的结果证明了该模型的优越性和全尺寸目标检测的潜力。本研究可以启发目标检测算法与注意机制融合的新研究,增强目标检测的准确性。

虽然本文大大提高了目标检测的精度,但是网络结构计算变得复杂。未来将探索更精确可靠的算法改进。未来的工作方向是将自适应递归 FPN 网络架构应用于

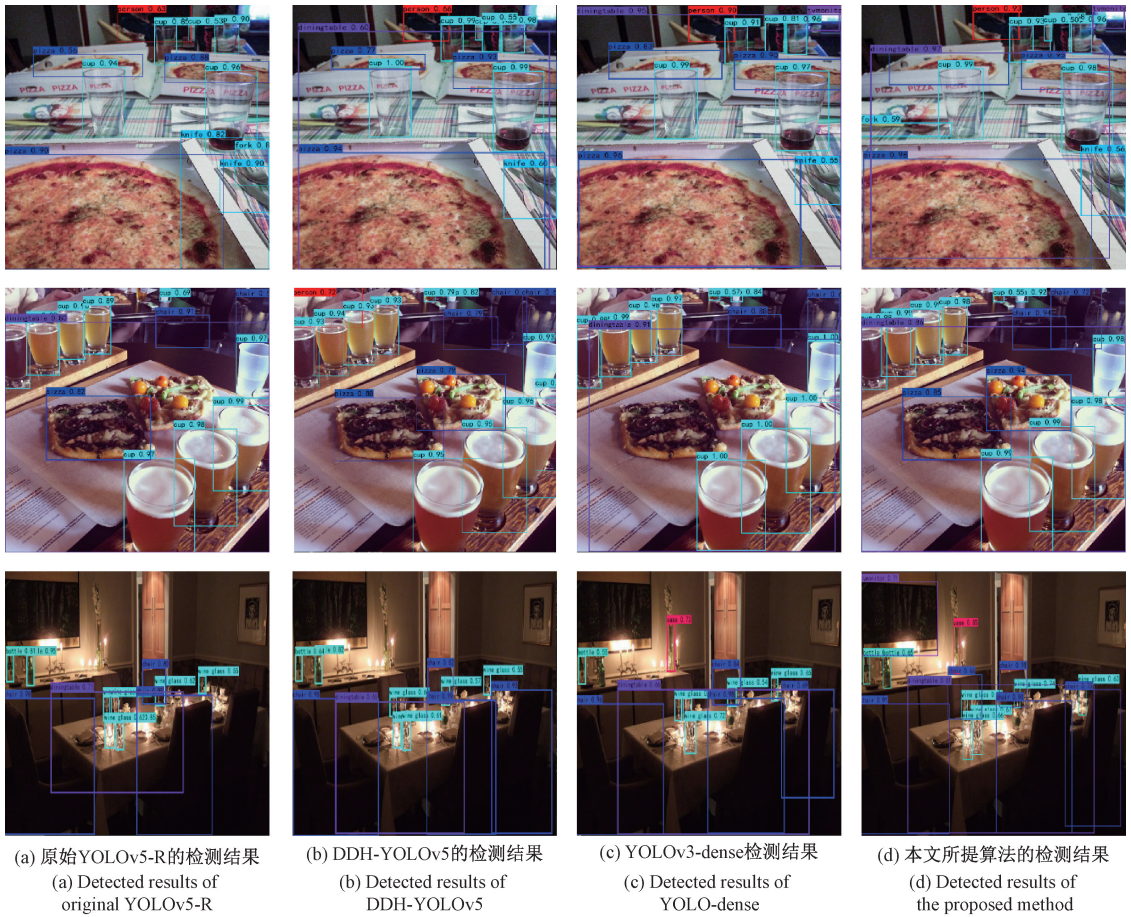


图 11 不同方法的检测结果

Fig. 11 Detected results of different methods

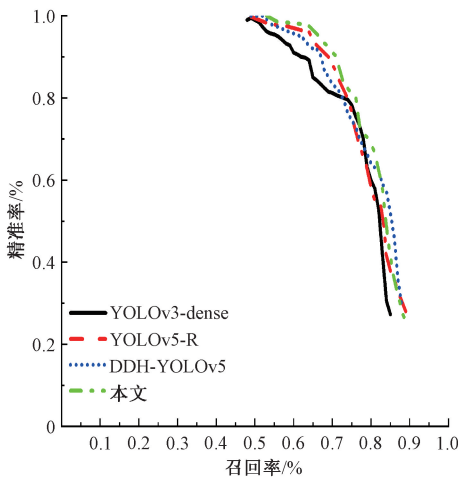


图 12 几种复杂多目标检测模型的 P-R 曲线  
Fig. 12 P-R curves of complex multiple object detection models in several categories

其他目标检测,消除检测计算过程复杂的缺点。

参考文献

[ 1 ] 石欣,卢灏,秦鹏杰,等. 一种远距离行人小目标检测方法[J]. 仪器仪表学报,2022,43(5):136-146.  
SHI X, LU H, QIN P J, et al. A long-distance pedestrian small target detection method [J]. Chinese Journal of Scientific Instrument, 2022, 43(5): 136-146.

[ 2 ] 冯明驰,高小倩,汪静姝,等. 基于立体视觉与激光雷达的车辆目标外形位置融合算法研究[J]. 仪器仪表学报,2021,42(10):210-220.  
FENG M CH, GAO X Q, WANG J SH, et al. Research on the fusion algorithm of vehicle object shape-position based on stereo vision and lidar [J]. Chinese Journal of Scientific Instrument, 2021, 42(10): 210-220.

[ 3 ] WIJETHUNGA P, CHANDRAWANSA I A, RATHNAYAKE B, et al. Control of a 6DOF mobile manipulator with object detection and tracking using stereo vision[J]. Instrumentation, 2021.

[ 4 ] KONG T, SUN F, LIU H, et al. Foveabox: Beyond anchor-based object detection[J]. IEEE Transactions on

- Image Processing, 2020, 29: 7389-7398.
- [ 5 ] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2015, 28.
- [ 6 ] FUENTES A, YOON S, KIM S C, et al. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition [ J ]. Sensors, 2017, 17(9): 2022.
- [ 7 ] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [ 8 ] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [ 9 ] REDMON J, FARHADI A. YOLOv3: An incremental improvement [ J ]. arXiv preprint arXiv: 1804.02767, 2018.
- [ 10 ] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
- [ 11 ] 吕禾丰, 陆华才. 基于 YOLOv5 算法的交通标志识别技术研究[J]. 电子测量与仪器学报, 2021, 35(10): 137-144.
- LYU H F, LU H C. Research on traffic sign recognition technology based on YOLOv5 algorithm [ J ]. Journal of Electronic Measurement and Instrumentation, 2021, 35(10): 137-144.
- [ 12 ] 曹红燕, 沈小林, 刘长明, 等. 改进的 YOLOv3 的红外目标检测算法[J]. 电子测量与仪器学报, 2020, 34(8): 188-194.
- CAO H Y, SHEN X L, LIU CH M, et al. Improved infrared target detection algorithm of YOLOv3 [ J ]. Journal of Electronic Measurement and Instrumentation, 2020, 34(8): 188-194.
- [ 13 ] 梁继然, 陈壮, 董国军, 等. 结合注意力机制和密集连接网络的车辆检测方法[J]. 电子测量与仪器学报, 2022, 36(3): 210-216.
- LIANG J R, CHEN ZH, DONG G J, et al. Vehicle detection method combining attention mechanism and dense connection network [ J ]. Journal of Electronic Measurement and Instrumentation, 2022, 36 ( 3 ) : 210-216.
- [ 14 ] 徐健, 陆珍, 刘秀平, 等. 注意力机制优化 RetinaNet 的密集工件检测方法研究[J]. 电子测量与仪器学报, 2022, 36(1): 227-235.
- XU J, LU ZH, LIU X P, et al. Research on dense workpiece detection method based on attentional mechanism optimization RetinaNet [ J ]. Journal of Electronic Measurement and Instrumentation, 2022, 36(1): 227-235.
- [ 15 ] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [ C ]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.
- [ 16 ] 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/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658-666.
- [ 17 ] WANG H, WANG J, BAI K, et al. Centered multi-task generative adversarial network for small object detection[J]. Sensors, 2021, 21(15): 5194.
- [ 18 ] KOYUN O C, KESER R K, AKKAYA I B, et al. Focus-and-Detect: A small object detection framework for aerial images [ J ]. Signal Processing: Image Communication, 2022, 104: 116675.
- [ 19 ] HU Y, LI X, ZHOU N, et al. A sample update-based convolutional neural network framework for object detection in large-area remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(6): 947-951.
- [ 20 ] WANG P, SUN X, DIAO W, et al. FMSSD: Feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery [ J ]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 58(5): 3377-3390.
- [ 21 ] REN J, WANG Z, ZHANG Y, et al. YOLOv5-R: Lightweight real-time detection based on improved YOLOv5 [ J ]. Journal of Electronic Imaging, 2022, 31(3): 033033.
- [ 22 ] WANG H, JIN Y, KE H, et al. DDH-YOLOv5: Improved YOLOv5 based on double IoU-aware decoupled head for object detection[J]. Journal of Real-Time Image Processing, 2022, 19: 1023-1033.
- [ 23 ] TIAN Y, YANG G, WANG Z, et al. Apple detection during different growth stages in orchards using the improved YOLOv3 model[J]. Computers and Electronics in Agriculture, 2019, 157: 417-426.

## 作者简介



刘鹏(通信作者),2015年于长春理工大学获得博士学位,现为长春理工大学电子信息工程学院副研究员,主要研究方向为光电检测与智能信息处理与视频图像处理。

E-mail: liup@cust.edu.cn

**Liu Peng** (Corresponding author)

received his Ph. D. degree from Changchun University of Science and Technology in 2015. Now he is an associate researcher at the School of Electronic and Information Engineering, Changchun University of Science and Technology. His main research interests include photoelectric detection and intelligent information processing and video image processing.



毕誉轩,2020年于长春理工大学获得学士学位,现为长春理工大学电子信息工程学院博士研究生,主要研究方向为图像处理。

E-mail: b971224@126.com

**Bi Yuxuan** received his B. Sc. degree

from Changchun University of Science and Technology in 2020. Now he is a doctoral candidate at the School of Electronic Information Engineering, Changchun University of Science and

Technology. His main research interest includes image processing.



张天翼,2022年于长春理工大学获得学士学位,现为长春理工大学电子信息工程学院硕士研究生,主要研究方向为图像处理。

E-mail: 1131709484@qq.com

**Zhang Tianyi** received his B. Sc. degree

from Changchun University of Science and Technology in 2022. Now he is a M. Sc. candidate at the School of Electronic Information Engineering, Changchun University of Science and Technology. His main research interest includes image processing.



史佳霖,2022年于西南民族大学获得学士学位,现为长春理工大学电子信息工程学院硕士研究生,主要研究方向为图像处理。

E-mail: 1456325470@qq.com

**Shi Jialin** received his B. Sc. degree from

Southwest Minzu University in 2022. Now he is a M. Sc. candidate at the School of Electronic Information Engineering, Changchun University of Science and Technology. His main research interest includes image processing.