

# 改进的 YOLOv3 算法及其在军事目标检测中的应用

于博文, 吕明

(南京理工大学 自动化学院, 江苏 南京 210094)

**摘要:** 复杂环境下军事目标检测技术是提高战场态势生成、分析能力的基础和关键。针对军事目标检测任务在复杂环境下传统检测算法的检测性能较低问题, 提出一种基于改进 YOLOv3 的军事目标检测算法, 通过深度学习实现复杂环境下军事目标的自动检测。构建军事目标图像数据集, 为各类目标检测算法提供测试环境; 在网络结构上通过引入可形变卷积改进的 ResNet50-D 残差网络作为特征提取网络, 提高网络对形变目标的检测精度和速度; 在特征融合阶段引入双注意力机制和特征重构模块, 增强目标特征的表征能力, 抑制干扰, 提升检测精度; 利用 DIOU 损失函数和 Focal 损失函数重新设计目标检测器的损失函数, 进一步提高其对军事目标的检测精度; 在军事目标图像数据集中进行测试实验。实验结果表明, 改进的 YOLOv3 算法相比于原 YOLOv3 算法, 平均精度均值提高了 2.98%, 检测速度提高了 8.6 帧/s, 具有较好的检测性能, 可为战场态势生成、分析提供有效的辅助技术支持。

**关键词:** 目标检测; 可形变卷积; YOLOv3 算法; 特征融合; 注意力机制

**中图分类号:** TP391.413 **文献标志码:** A **文章编号:** 1000-1093(2022)02-0345-10

**DOI:** 10.3969/j.issn.1000-1093.2022.02.012

## Improved YOLOv3 Algorithm and Its Application in Military Target Detection

YU Bowen, LÜ Ming

(School of Automation, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China)

**Abstract:** Military target detection in a complex environment is the basis and key to improving battlefield situation generation and analysis capability. For the military target detection tasks, the detection performance of traditional detection algorithms in complex environment is low. A military target detection algorithm based on improved YOLOv3 algorithm is proposed to automatically detect the military targets in complex environment through deep learning. A military target image dataset is constructed to provide a testing environment for various target detection algorithms. The detection accuracy and speed of deformable target are improved by introducing the deformable convolutional improved ResNet50-D residual network as feature extraction network. In the stage of feature fusion, a dual-attention mechanism and feature reconstruction module are introduced to enhance the characterization ability of target features, suppress the interference, and improve the detection accuracy. The loss function of target detector is redesigned by using DIOU Loss functions and Focal Loss to further improve the detection accuracy of military targets. The experimental results show that the improved YOLOv3 algorithm improves the average detection accuracy by 2.98% and the detection speed by 8.6 frames/s compared with the original

收稿日期: 2021-02-20

基金项目: 江苏省自然科学基金项目(BK20180467)

作者简介: 于博文(1988—), 男, 博士研究生。E-mail: yubowen\_0403@njust.edu.cn

通信作者: 吕明(1980—), 女, 副研究员, 博士。E-mail: 18761654456@163.com

YOLOv3 algorithm. The improved YOLOv3 algorithm has better detection performance and can provide effective auxiliary technical support for battlefield situation generation and analysis.

**Keywords:** target detection; deformable convolution; YOLOv3 algorithm; feature fusion; attention mechanism

## 0 引言

战场态势感知是对作战部队、支援部队的部队部署、武器装备和战场环境(地形、气象、水文等)进行实时感知的过程,包括传统的侦察、监视、情报、航标指示、损伤评估、信息资源管理与控制等。在现阶段多兵种联合作战中通过战场态势进行作战形式解析和预判,进而控制整个作战过程,最终获得整场战争的作战优势。在未来的信息化战争中,提高战场态势感知能力,可以有效提高战争全局掌控能力,各军事大国都在加强相关技术的开发和研究<sup>[1]</sup>。目前,军事目标的识别与定位是影响战场态势感知的关键技术<sup>[2]</sup>。因此,开展针对复杂环境下战场军事目标自动检测技术的研究,对生成和分析战场态势具有重要意义。

近年来,学者们陆续将基于深度学习的方法应用于战场态势感知,能够较好地解决目前信息化战场中复杂战场态势生成和解析问题<sup>[3-4]</sup>。深度学习是一种具有多层感知的神经网络模型,具有非线性表达、多层学习和自主提取特征的特点。将基于深度学习的目标检测模型应用于战场军事目标识别和定位,可以为战场态势感知系统提供有效的技术支持<sup>[5]</sup>。包壮壮等<sup>[6]</sup>通过结合可形变卷积和反卷积的特性在单次多盒检测器(SSD)<sup>[7]</sup>的基础上设计了一种小目标检测模型,通过调整底层网络结构增强了网络对深层特征与浅层特征的融合能力,在进行小目标检测任务时具有较高的精度。戴文君等<sup>[8]</sup>提出的一种多尺度形变目标检测方法,通过基于可形变卷积改进的 ResNet101 残差网络<sup>[9]</sup>和感兴趣区域池化<sup>[10]</sup>,提高了对形变和多尺度目标的检测能力。杜泽星等<sup>[11]</sup>采用稠密连接网络结构,保证浅层网络的特征有效性,通过具有大尺寸感受野的扩张块和反卷积网络结构对浅层特征和深层次特征进行融合,大大减少网络推理时间和增强对多尺度目标的检测能力。杨传栋等<sup>[12]</sup>针对弹载图像目标检测精度不佳的问题,通过数据增强、尺度变换等处理方法,解决了在恶劣天气条件下目标形变、精度速度权衡等检测难题。王志等<sup>[13]</sup>针对复杂环境对静态目标检测有较大影响的问题,

提出了一种区域合并的方法进行端到端的交替训练,在对复杂环境背景下进行单、多目标检测具有较高的精度和鲁棒性。

目前,基于深度学习的目标检测主要有两个研究方向:多阶段目标检测算法(区域卷积神经网络系列<sup>[14-16]</sup>),该类方法通过提取候选区域的特征信息,来对预定义的候选区域进行分类和位置回归。该类检测模型精度较高,但是其检测速度较慢。单阶段目标检测算法(YOLO 系列算法<sup>[17-19]</sup>),该类模型利用初始锚点对目标区域进行定位,直接预测目标类别,从而在不使用区域建议网络的情况下完成整个检测过程,实现端到端的目标检测。该类检测模型速度较快,但精度相对较低。

由于战场态势具有高动态性,对军事目标检测的实时性具有较高的要求。此外,在对感知系统(武器火控、无人机侦察等)获得的战场图像信息进行目标检测时,该过程中军事目标受到光照、成像视角、目标尺寸、伪装、部分目标存在遮挡等情况的影响,使得上述方法在复杂环境下军事目标检测结果不佳。基于此,本文将从战场态势感知的基本问题军事目标的识别和定位出发,通过深度学习方式对军事目标进行检测。YOLOv3 算法相对于其他检测算法具有较高的检测性能,故本文选取单阶段检测算法 YOLOv3 作为检测算法的基础结构,并对其进行改进,提出一种复杂环境下军事目标检测(YOLOv3-DAR)算法:针对复杂背景下的军事目标检测任务,构建了一个包含多种陆地作战中常见作战单元的军事目标数据集。使用性能更加优异的可形变卷积(DCN)<sup>[20]</sup>和 ResNet50-d 残差网络<sup>[21]</sup>相结合的网络作为检测算法的特征提取网络,在一定程度上提升检测精度和效率。在特征融合过程中引入双注意力机制模块和特征重构模块,引导相邻层次的特征进行自适应融合,使得各层级的特征信息可以相互增强、补充,增强各层级特征之间的信息交互。为了使检测算法的检测精度进一步提高,基于 DIOU 损失函数<sup>[22]</sup>和 Focal 损失函数<sup>[23]</sup>重新设计了目标检测器的损失函数。最后在本文所构建的数据集和通用数据集 PASCAL VOC<sup>[24]</sup>上进行对比实验,实验结果表明本文提出的目标检测算法 YOLOv3-

DAR 具有较好的检测性能,可为战场态势生成、分析提供有效的辅助技术支持。

### 1 构建军事目标数据集

根据陆地战场作战环境下常见的军事行动单元构建军事目标数据集,其中所包含的作战单元大体可以分为:地面机动单元,坦克、步兵战车、自行火炮等,其具有高机动性和高毁伤性,因此是地面战场中的主要检测目标;地面作战人员,地面作战部队是陆战场中不可忽略的一部分,其目标较小,发现概率较低;低空飞行单元,无人机、直升机等,其主要在战场中起到低空侦查、作战引导、直接攻击等作用,是陆战场中地面部队的直接威胁;导弹部队作战单元,具有较大的杀伤力,但是机动性较差,主要用来攻击重要军事目标。在真实作战环境中,军事行动单元通常采用军用迷彩或与复杂自然环境相近的伪装物进行伪装,从而大大提升军事行动单元的战场存活能力,有效降低了各类检测手段的发现几率。基于此,针对复杂环境下军事目标检测任务,从互联网上下载了大量复杂背景下的军事目标图像构建军事目标数据集。

数据集中总共包含了 4 种陆地作战场景(沙漠、丛林、城市、雪地)和陆地作战场景中常见的 7 类军事目标,同时考虑了影响目标检测结果的因素,例如前景遮挡、烟雾、目标尺寸、成像视角等,其中部分图像示例如图 1 所示。将数据集中各作战场景下的军事目标按照 7:2:1 的比例划分到训练集、测试集、验证集。然后,利用图像标记软件 LabelImg 对数据集中的样本数据进行标记,数据集标签格式与 PASCAL VOC 数据标签格式一致。数据集中总计有 17 700 幅图像,包含有 41 399 个目标标注框。各类军事目标的详细统计信息如表 1 所示。



图 1 数据集部分图像示例

Fig. 1 Examples of partial images of dataset

表 1 军事目标数据集信息明细

Tab. 1 Details of military target dataset information

军事目标	图像		目标标注框	
	数量	宽度与高度范围	数量	宽度与高度范围
坦克	3 962		8 556	
直升机	1 835		2 899	
作战士兵	1 735		9 628	
无人机	2 667	[600,1 280]	3 872	[14,1 280]
导弹	1 445		1 788	
步兵战车	3 381		6 824	
汽车	2 675		7 832	

在目标检测模型训练、应用的过程中会调整图像尺寸到某个固定的尺寸,通过计算目标标注框的宽度、高度与其所在图像的宽度、高度比值的最大值  $\gamma_{max}$  可以有效评价数据集尺度分布的情况,因此本文选取其作为衡量数据集尺度分布的标准,其计算公式为

$$\gamma_{max} = \max(a/A, b/B), \tag{1}$$

式中: $a$ 、 $b$  分别表示目标标注框的宽度、高度; $A$ 、 $B$  分别表示目标框所在图像的宽度、高度。

图 2 为本文构建数据集中各目标类别的尺度分布情况。图 3 为本文构建数据集中的汽车、步兵战车、坦克等主要目标与 PASCAL VOC 2007、2012 数据集中交通工具类目标的尺度分布对比图,可以看出本文构建的数据集相对于 PASCAL VOC 数据集具有更多小尺寸目标。

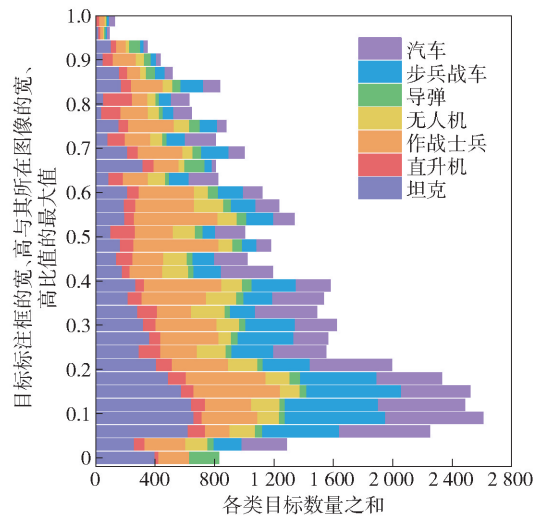


图 2 本文数据集中各类目标尺度分布图

Fig. 2 Distribution chart of various target scales in the dataset in the present paper

### 2 基于改进的 YOLOv3 军事目标检测算法

图 4 为本文所提出的复杂环境下军事目标检测

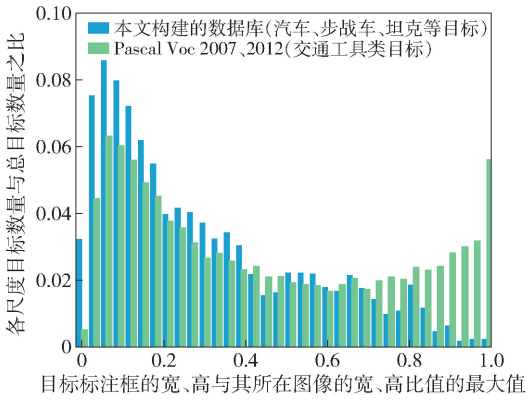


图 3 主要目标尺度分布对比图

Fig. 3 Comparison chart of main target scale distribution

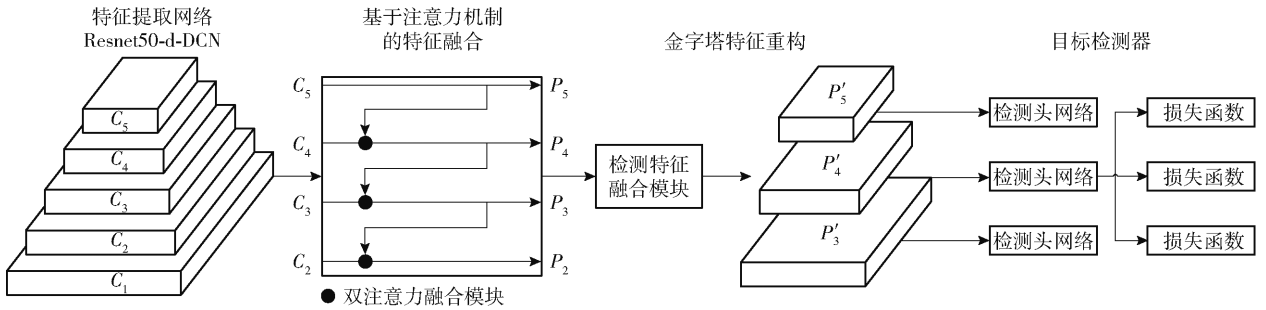


图 4 YOLOv3-DAR 检测算法的整体结构

Fig. 4 Framework of YOLOv3-DAR target detection algorithm

测特征。而传统的卷积网络由于卷积采样方式的限制,处理几何形变问题的能力相对有限,需要通过大量的图像训练来提高网络的泛化能力。在实际使用过程中,如果遇到数据集中没有的要素,极有可能出现漏检、误检的情况,从而影响目标检测的结果。

针对上述问题,本文引入可形变卷积网络和残差网络。可形变卷积的效果已在许多检测模型中得到了验证,如图 5 所示,相对于传统卷积,可形变卷积在采样过程中通过添加偏移量使其可以动态改变感受野,并根据目标的形变情况调整采样点的位置,有效提高网络的泛化能力,从而提高检测模型的检测精度。现阶段残差网络具有较为广泛的应用,其具有较多的变体可供选择,同时具有较强的可扩展性。通过综合考虑,本文应用可形变卷积对残差网络的变体 ResNet50-d<sup>[21]</sup>的部分卷积层进行改进得到 ResNet50-d-DCN 网络。可形变卷积本身不会显著的增加模型中参数的数量,但在实际应用中,过多的可形变卷积层会增加网络的推理时间。因此,为了平衡效率和有效性本文仅将 ResNet50-d 网络中的第 4 阶段、第 5 阶段中的 3×3 卷积替换为可形变卷积,其余层保持不变,选取第 2 阶段~第 5 阶段作为输出,记为 C<sub>2</sub>、C<sub>3</sub>、C<sub>4</sub>、C<sub>5</sub>。

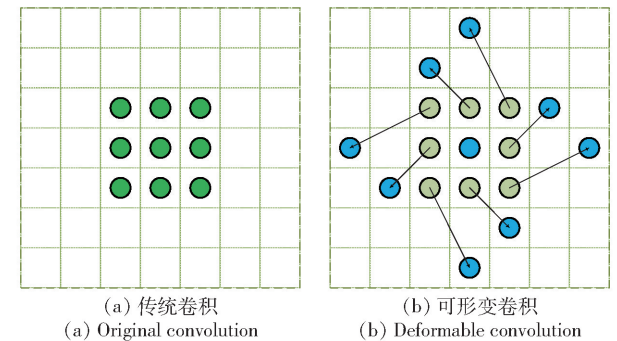


图 5 传统卷积与可形变卷积对比示意图

Fig. 5 Comparison diagram

传统的卷积过程可以表示为

$$R = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\}, \quad (2)$$

$$y(P_0) = \sum_{P_n \in R} (w(P_n) \cdot x(P_0 + P_n)), \quad (3)$$

式中:  $R$  为一个 3×3 卷积层网格;  $P_0$ 、 $P_n$  分别为特征图  $y$  和  $R$  中某个位置的坐标;  $w(P_n)$  为卷积核中的权值;  $P_0 + P_n$  为采样点的坐标,  $x(P_0 + P_n)$  为输入特征图  $x$  中对应的像素值。

可形变卷积层的输出可表示为

$$y(P_0) = \sum_{P_n \in R} (w(P_n) \cdot x(P_0 + P_n + \Delta P_n)), \quad (4)$$

$$\{\Delta P_n | n = 1, \dots, N\}, N = |R|. \quad (5)$$

偏移量添加到  $R$  中,则采样点的坐标变换为  $P_0 + P_n + \Delta P_n$ ,  $\Delta P_n$  一般为分数。输入特征图中的样本点会定位在一个分数的位置,通过双线性插值获取  $x(P_0 + P_n + \Delta P_n)$  的像素值,偏移量则通过额外的卷积层来学习。

## 2.2 基于注意力机制的特征融合算法

在YOLOv3算法中相邻的层级特征通过串联方式进行特征融合,而不同层级的特征分布差异悬殊,简单的特征融合方式会破坏原有的特征。低层次网络特征具有较为丰富的细粒度特征信息,而高层次网络特征具有较为丰富的语义信息,如何有效地将二者融合在一起是至关重要的。进行目标检测时不仅需要目标本身的尺度信息和细粒度信息,还需要依赖于周围的环境信息,基于此,本文基于压缩激励卷积网络<sup>[25]</sup>的思想,在特征融合阶段引入双注意力机制,使网络可以自适应过滤特征信息,强调对检测更为重要的特征信息,抑制干扰,加强检测特征的特征能力,特征融合过程如图6所示。图6中,  $F_s(\cdot)$ 、 $F_e(\cdot)$ 、 $F_r(\cdot)$  分别表示压缩操作、激励操作、权值分配操作;  $H$ 、 $W$ 、 $C$  分别表示卷积特征的高度、宽度、通道数。

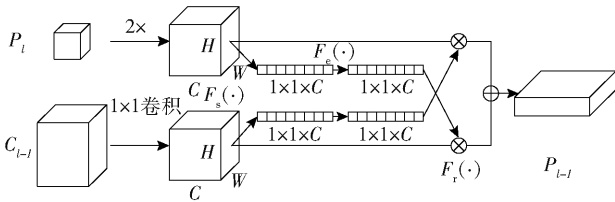


图6 特征自适应融合过程

Fig. 6 Process of adaptive feature fusion

如图6所示,在融合过程中引入双向选择机制,起到相互监督的作用。根据图4中的特征金字塔,通过  $P_l$  和  $C_{l-1}$  对特征融合过程进行说明。由于  $C_{l-1}$  分辨率是  $P_l$  的2倍,通过尺度为2的上采样层对  $P_l$  进行上采样,通过  $1 \times 1$  卷积对  $C_{l-1}$  进行降维操作,使得来自不同层级的特征图具有相同的尺寸。利用全局平均池化对特征图进行压缩操作,获得输出维度为  $1 \times 1 \times C$  的特征图  $z_l, z_{l-1}$ :

$$z_l = F_s(P_l) = \frac{1}{H_l \times W_l} \sum_{i=1}^{H_l} \sum_{j=1}^{W_l} P_l(i, j), \quad (6)$$

$$z_{l-1} = F_s(C_{l-1}) = \frac{1}{H_{l-1} \times W_{l-1}} \sum_{i=1}^{H_{l-1}} \sum_{j=1}^{W_{l-1}} C_{l-1}(i, j), \quad (7)$$

式中:  $H_l, W_l$  表示第  $l$  层级特征的高度、宽度。

由于全局平均池是基于通道的,  $z_l, z_{l-1}$  与  $P_l, C_{l-1}$  的通道数相同。然后利用 Sigmoid 函数、Relu 函数对压缩操作所获得的特征图  $z_l, z_{l-1}$  分别进行激励操作,获取特征通道间的非线性关系和特征通道的权值。权值  $s_l, s_{l-1}$  输出维度为  $1 \times 1 \times C$ , 如(8)式和(9)式所示:

$$s_l = F_e(z_l) = \sigma(\omega_l^2(\delta(\omega_l^1 z_l))), \quad (8)$$

$$s_{l-1} = F_e(z_{l-1}) = \sigma(\omega_{l-1}^2(\delta(\omega_{l-1}^1 z_{l-1}))), \quad (9)$$

式中:  $\delta(\cdot)$ 、 $\sigma(\cdot)$  分别为 Sigmoid 函数、Relu 函数;  $\omega^1, \omega^2$  分别为两个全连接层的权值矩阵,  $\omega^1 \in \mathbf{R}^{\frac{C}{16} \times C}$ ,  $\omega^2 \in \mathbf{R}^{C \times \frac{C}{16}}$ 。

最后是权值分配操作,将激励操作输出的权重通过乘法逐通道加权到先前的特征上,在通道维度上完成对原始特征的重新标定,让模型可以更加关注信息量最大的通道特征,抑制那些不重要的通道特征。本文在权值分配操作的基础上加入交叉监督机制,通过得到的特征通道权值  $s_l, s_{l-1}$ , 生成特征  $P_{l-1}$ :

$$P_{l-1} = F_r(s_l, C_{l-1}) + F_r(s_{l-1}, P_l) = s_l C_{l-1} + s_{l-1} P_l. \quad (10)$$

经过上述特征融合过程,可以有效地将高层次特征的语义信息和低层次特征的细粒度信息融合在一起,得到的融合特征具有较强的表征能力。YOLOv3 检测模型的训练过程中,高层次特征用于检测较大目标,低层次特征用于检测较小目标,这样就破坏了各个层级检测特征之间的依赖关系。根据文献[26]不同层级特征相互补充相互增强的思想,本文对检测特征进行重构,特征重构过程如图7所示。

首先将各层级检测特征  $P_2, P_3, P_4, P_5$  进行切片,每个层级特征分成4份。为了保证在接下来的重构过程中各层级特征维度是一致的,需要对检测特征中的各尺度的切片进行调整。如果切片与目标特征的尺寸一致,则保持原切片大小不变;如果切片较目标特征的尺寸小,则采用上采样层对切片特征进行上采样;如果切片较目标特征的尺寸大,则采用最大池化对切片特征进行下采样。然后将分辨率一致的切片通过串联的方式合并成新的特征。通过尺度缩放、重新配置,将各层级的语义信息和细粒度信息重新组合在一起。最后,将重构后的检测特征  $P'_3, P'_4, P'_5$  输入到目标检测器。

## 2.3 损失函数

损失函数  $L_s$  由边界框损失函数  $L_l$ 、置信度损失

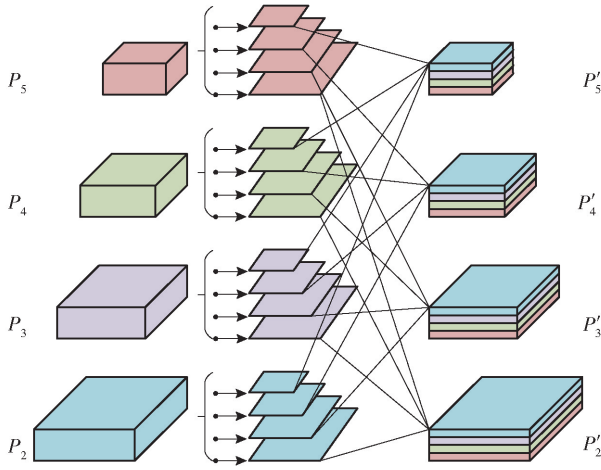


图 7 检测特征重构过程

Fig. 7 Process of detection feature reconstruction

函数  $L_d$  和类别损失函数  $L_c$  组成,如(11)式所示:

$$L_s = L_i + L_d + L_c. \quad (11)$$

交并比 (IOU) 在进行边界框的回归预测时,如果预测框与标注框之间没有重叠,会出现损失函数梯度为 0 的问题,为了解决这个问题,本文引入 DIOU 损失函数进行边界框的回归预测,其可以有效解决回归不准确和收敛慢的问题,其计算公式为

$$L_i = 1 - \frac{|B_p \cap B_g|}{|B_p \cup B_g|} + \frac{\rho^2(b_p, b_g)}{d^2}, \quad (12)$$

式中:  $b_p$  和  $b_g$  分别为预测框  $B_p = (x, y, w, h)$  和目标框  $B_g = (x_g, y_g, w_g, h_g)$  的中心点;  $\rho(\cdot)$  表示  $b_p$  与  $b_g$  之间欧几里得距离;  $d$  为覆盖目标框和预测框最小矩形的对角线长度。

在训练数据集中存在一部分较为容易分类的样本数据,因此会使得模型的优化方向与所预期的方向相反,为了降低负样本对模型优化的影响,本文引入 Focal 损失函数来降低负样本对模型优化影响的权重,使其更专注于更难更复杂的分类样本,其计算公式为

$$L_d = -(y_g - y_p)^2 \times y_g \ln y_p - (y_g - y_p)^2 (1 - y_g) \ln (1 - y_p), \quad (13)$$

式中:  $y_p$  和  $y_g$  分别为检测目标置信度的预测值和真实值。

分类损失通过交叉熵损失函数来计算,将输入图像划分成  $S \times S$  个网格,每个网格对应 3 个目标预测框,分类损失只对有识别对象的网格有效,其计算公式为

$$L_c = \sum_{i=0}^{S \times S} \sum_{j=0}^3 \sum_{r \in T} I_{ij} M(p_i(r), \hat{p}_i(r)), \quad (14)$$

$$M(a, \hat{a}) = -[a \ln \hat{a} + (1 - a) \ln (1 - \hat{a})], \quad (15)$$

式中:  $r, T$  分别表示检测目标的类别和数据集中所包含的目标类别;  $I_{ij}$  表示与第  $i$  个网格相关联的第  $j$  个预测框是否有检测到目标对象,如果有检测到目标  $I_{ij}$  的值为 1,如果没有检测到目标  $I_{ij}$  的值为 0;  $p_i(r)$ 、 $\hat{p}_i(r)$  分别为检测目标分类概率的真实值和预测值。为了保证复杂环境下军事目标的识别精度,在多个尺度上对军事目标进行检测,因此最终的损失函数  $L$  为

$$L = \sum_{i=1}^3 L_i + L_d + L_c, \quad (16)$$

式中:  $i$  表示检测特征的尺度。

### 3 实验结果分析

#### 3.1 实验参数配置与检测结果

##### 1) 实验条件

本文所有的实验均在操作系统为 Ubuntu18.04 的服务器上进行,中央处理器为至强银牌 4114 处理器 @ 10C \* 2.2 GHz, 图形处理器为 GeForce RTX 2080ti \* 4, 内存为 128 GB。采用 TensorFlow2.0.0 为基础框架搭建神经网络,使用 Python 语言进行编程。训练、测试数据集为本文构建的军事目标数据集。

##### 2) 模型训练细节

在训练过程中使用随机梯度下降 (SGD) 法进行检测模型训练,批处理大小设置为 64, 最大迭代次数设置为 100 k, 初始学习率、权重衰减系数、动量因子分别设置为 0.001、0.000 5、0.9。当迭代次数到达 60 000 和 85 000 时,学习率分别除以 10。同时结合多尺度训练策略<sup>[18]</sup>, 每隔 10 次迭代随机从 {320, 352, 384, 416, 448, 480, 512, 544, 576, 608} 中选取图片的输入尺寸。通过图像旋转图像、缩放、调整对比度、调整饱和度等方法生成更多可供网络训练的军事目标样本,增强网络的泛化性,提高对军事目标检测的精准度及稳定性。根据训练数据集中的图像样本通过 K-means 算法对目标标注框进行聚类,得到 9 个模板尺寸 (模型输入分辨率  $544 \times 544$ ), 分别为 (24, 29)、(41, 69)、(66, 153)、(79, 89)、(127, 283)、(158, 152)、(243, 398)、(351, 262)、(455, 468), 将运用这些先验框来进行目标检测实验。此外,在预测框评估阶段引入 DIOU 非极大值抑制 (DIOU-NMS)<sup>[22]</sup> 以提高模型的检测性能。

##### 3) 评价指标

本文选取平均查准率 (AP) 用来评价检测模型对单个目标类别的检测准确性。AP 从准确率、召回

率这两个方面来评价检测模型的精准性,可有效地评价模型对单个目标类别的检测性能,AP采用文献[24]的计算方式,其中IOU阈值、置信度阈值分别设置为0.5、0.001。AP均值(mAP)用来评价检测模型对多个目标类别的综合检测性能。通常AP和mAP数值越高,模型的检测性能越好,反之亦然。通过每秒可处理的图像数量(FPS)来测量检测模型的检测速度。

应用YOLOv3-DAR目标检测模型对本文数据集中复杂环境下部分军事目标进行目标检测,其检测效果如图8所示。

### 3.2 军事目标检测结果分析

#### 3.2.1 目标检测性能分析

通过YOLOv3、YOLOv3-DAR、R-FCN<sup>[10]</sup>、Faster RCNN<sup>[27]</sup>算法对本文构建的数据集进行目标检测,其检测性能比对结果如表2所示。从表2中可以看出,本文提出的YOLOv3-DAR算法比YOLOv3算法



图8 YOLOv3-DAR算法的检测结果示例

Fig.8 Test results of YOLOv3-DAR algorithm

的平均精度均值高出了2.98%,FPS提高了8.6帧/s。YOLOv3-DAR算法比Faster RCNN、R-FCN算法的平均精度均值分别高出了1.70%、1.44%,同时YOLOv3-DAR算法的检测速度分别是Faster RCNN、R-FCN算法的13.88倍和5.51倍。YOLOv3-DAR算法相对其他3种检测算法具有较高的检测精度和处理速度。

表2 YOLOv3-DAR算法与其他3种检测算法的对比实验结果

Tab.2 Experimental results of YOLOv3-DAR and other three detection algorithms

检测算法	特征提取网络	各类目标的检测精准度 AP/%							mAP/ %	FPS/ (帧·s <sup>-1</sup> )
		坦克	直升飞机	作战士兵	无人机	导弹	步兵战车	汽车		
Faster RCNN	ResNet101 + FPN	78.85	77.09	79.09	72.06	77.36	67.21	84.77	76.63	2.5
R-FCN	ResNet101 + DCN	79.12	77.35	79.35	72.3	77.63	67.45	85.06	76.89	6.3
YOLOv3	Darknet53	77.53	75.81	77.75	70.86	76.07	66.09	83.34	75.35	26.1
YOLOv3-DAR	ResNet50-d + DCN	81.22	78.41	80.5	74.31	78.66	69.8	85.34	78.33	34.7

从数据集中选取若干幅图片进行YOLOv3算法与YOLOv3-DAR算法检测结果对比实验,二者的IOU阈值、置信度阈值分别设置为0.5、0.25,从中随机选取5幅图片进行说明,误检、漏检通过人工标注的方式对目标进行标注。YOLOv3-DAR算法与YOLOv3算法检测效果对比如表3所示。

测试图片1中一共有5个目标,其中2个目标在图像中所占像素较少,YOLOv3算法检测出其中3个目标,YOLOv3-DAR算法检测出全部5个目标。测试图片2~图片4为属于遮挡、形变较为严重的情况,YOLOv3算法存在漏检、误检的情况,YOLOv3-DAR算法可以检测出全部目标,可在置信度较低的情况下有效标注出待检测目标。测试图片5属于目标过于密集、遮挡较为严重的情况,YOLOv3-DAR算法出现一个漏检目标,YOLOv3算法出现两个漏检目标。综上所述,可以看出YOLOv3-DAR网络整体性能优于YOLOv3网络,可以有效降低目标受到光

照、成像视角、目标尺寸、目标之间遮挡等对军事目标检测的影响。通过对特征提取网络的改进、引入注意力机制和特征重构策略,增强了网络的表示能力,使网络具有自适应过滤干扰特征信息的能力,具有较强的鲁棒性,同时减少了网络参数,提高了复杂环境下军事目标的检测效率和精准度。

#### 3.2.2 模型分解实验

本文所提出的YOLOv3-DAR网络中应用通过可形变卷积改进的ResNet50-d作为特征提取网络,使用基于注意力机制的融合算法对各个尺度的检测特征进行融合,对融合后的检测特征进行重构,重新设计了多尺度目标检测器的损失函数,在训练过程中采用多尺度训练,为了验证各改进策略对检测性能的贡献度,在本文所构建的训练数据集中进行分解实验。当输入分辨率为544×544时,其检测结果如表4所示,可以看出本文对YOLOv3中各模块的改进策略对检测性能的提升均有一定的帮助。

表 3 YOLOv3 算法与 YOLOv3-DAR 算法识别效果对比

Tab. 3 Recognition effects of YOLOv3 and YOLOv3-DAR algorithms

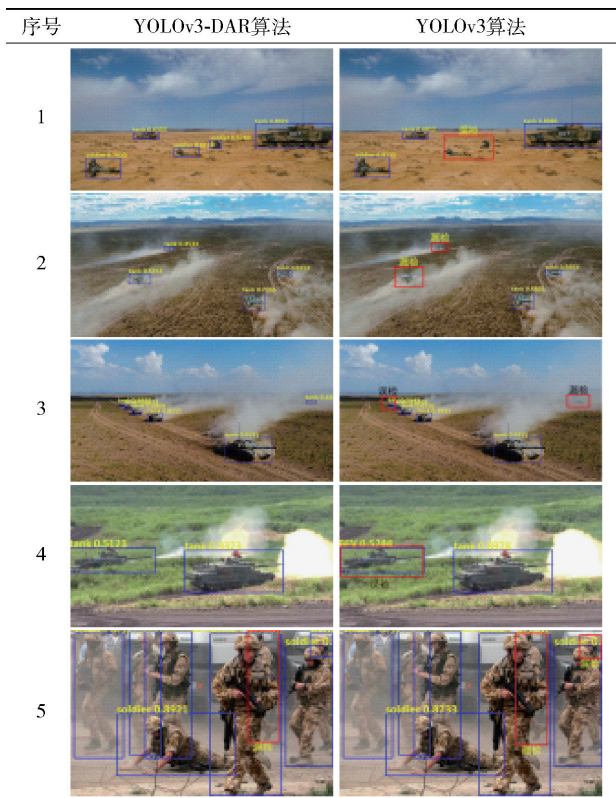


表 4 YOLOv3-DAR 模型分解实验结果

Tab. 4 Experimental results of YOLOv3-DAR model decomposition

模型	改进策略	mAP/%	FPS/ (帧·s <sup>-1</sup> )
A	DarkNet53 YOLOv3	75.35	26.1
B	ResNet50-d + DCN YOLOv3	75.57	37.8
C	模型 B + 双注意力机制	77.10	35.5
D	模型 C + 特征重构	77.56	34.7
E	模型 D + DIU 损失函数 + Focal 损失函数	77.85	34.7
F	模型 E + 多尺度训练	78.33	34.7

如表 4 所示,模型 A 为原始的 YOLOv3 模型,模型 B 为引入 ResNet50-d + DCN 特征提取网络的 YOLOv3 模型,模型 C、D、E、F 分别是在模型 B 的基础上递进增加双注意力机制、特征重构模块、改进后的损失函数、多尺度训练。下面将对各改进策略逐一说明。

模型 A→模型 B:使用可形变卷积改进的 ResNet50-d 残差网络替换 YOLOv3 模型中的 DarkNet53

网络作为特征提取网络。ResNet50-d 的网络参数个数相对于 DarkNet53 小的很多,在 ResNet50-d 的基础上引入可形变卷积,增强了检测模型对检测目标不同姿态、目标遮挡、部分形变及尺度变化的处理能力,使得检测模型具有更快的检测速度和更高的检测精度。实验结果表明,相对于原始 YOLOv3 模型的检测结果 mAP 提高了 0.22%,FPS 提高了 11.7 帧/s,有效地提高了检测性能。

模型 B→模型 C:在特征融合阶段引入双注意力选择机制,通过相邻层级特征之间的相互监督,使得模型可以更加关注信息量最大的通道特征,自适应地抑制干扰因素对检测结果的影响,加强了目标检测特征的表征能力,使得 mAP 从 75.57% 提高到 77.10%,结果表明引入注意力机制可以有效提高检测精度。

模型 C→模型 D:特征重构将多个层级的检测特征融合在一起,促进各层级特征之间的信息传递,使不同层级的检测特征之间可以相互补充、相互增强,从而使得到的检测特征具有更强的表达能力,使得 mAP 从 77.10% 提高到 77.56%。

模型 D→模型 E:考虑修改 YOLOv3 模型的损失函数来提高模型的精准度,修改损失函数通常只对训练过程产生影响,而不会或很少影响网络的推理时间。本文通过引入 DIU 损失函数、Focal 损失函数重新设计损失函数,进一步提高了检测模型的定位和分类能力,在检测框筛选的过程中引入 DIU-NMS 使得到的结果更加合理、有效,检测结果 mAP 从 77.56% 提高到 77.85%。

模型 E→模型 F:每隔 10 次迭代随机选取一种尺度进行网络训练,结合多尺度训练将 YOLOv3-DAR 检测网络的 mAP 提高了 0.48%,这样可以使检测网络的鲁棒性更强,使得检测网络可以处理不同分辨率的目标图像。

同时,为了进一步验证可形变卷积、基于双向选择注意力机制的特征融合算法对本文数据集检测精度的影响,选取 Faster RCNN 算法、R-FCN 算法和本文算法进行对比实验,其中各算法均使用相同的特征提取网络,检测结果如表 5 所示。

从表 5 中的检测结果可以看出,单独使用可形变卷积可以有效提高模型的检测精度,而同时使用可形变卷积和基于双向选择注意力机制的特征融合则可以进一步提高检测模型的检测精度。实验证明通过上述改进策略的结合,可以有效提高检测模型的检测能力。

表 5 改进策略对 mAP 的影响

Tab.5 Impact of improvement strategies on mAP

检测算法	特征提取网络	特征融合	mAP/ %	FPS/ (帧·s <sup>-1</sup> )
Faster RCNN	ResNet50-d	×	74.32	5.1
	ResNet50-d + DCN	×	75.41	4.4
	ResNet50-d + DCN	√	77.86	3.6
R-FCN	ResNet50-d	×	74.87	12.6
	ResNet50-d + DCN	×	76.21	11.2
	ResNet50-d + DCN	√	78.11	9.3
本文算法 (分辨率 544 × 544)	ResNet50-d	×	74.18	38.6
	ResNet50-d + DCN	×	75.57	37.8
	ResNet50-d + DCN	√	77.85	34.7

注: ×表示未使用特征融合网络,√表示使用特征融合网络。

### 3.3 PASCAL VOC 数据集上的实验

除了在本文收集的数据集上进行实验外,还在通用数据集 PASCAL VOC 2007、2012 上进行了算法对比实验。为了公平起见,所有算法均使用 Tensorflow2.0 为基础构建模型,每个算法模型训练 100 个 Epochs。在训练过程中采用文献[21]的训练策略。各算法在 PASCAL VOC2007 测试集上的检测结果如表 6 所示。

表 6 在 PASCAL VOC 2007 测试集上的结果

Tab.6 Experimental results of PASCAL VOC 2007 dataset

检测算法	输入分辨率	特征提取网络	mAP/ %	FPS/ (帧·s <sup>-1</sup> )
YOLO	448 × 448		66.40	20.0
YOLOv2	544 × 544	DarkNet19	78.62	40.0
Faster RCNN		ResNet101 + FPN	81.37	4.3
R-FCN		ResNet101 + DCN	82.54	8.0
YOLOv3	544 × 544	Darknet53	79.56	26.7
YOLOv3-DAR	416 × 416	ResNet50-d + DCN	82.58	43.5
YOLOv3-DAR	544 × 544	ResNet50-d + DCN	83.32	35.6

从表 6 的实验数据可以看出,本文所提出的算法检测性能整体上优于其他检测模型。当输入分辨率为 544 × 544 时,YOLOv3-DAR 算法相对于原 YOLOv3 算法,mAP 提高了 3.76%,FPS 提高了 8.9,检测精度和速度均有一定的优势。当 YOLOv3-DAR、YOLOv3 算法的输入分辨率分别为 416 × 416、544 × 544 时,YOLOv3-DAR 算法相对于原 YOLOv3 算法,mAP 提高了 3.02%,FPS 达到了 43.5,可以看出 YOLOv3-DAR 算法相对于 YOLOv3 算法具有更为优秀的检测性能。综上所述可以证明本文所提出

的 YOLOv3-DAR 检测模型具有较好的检测性能。

## 4 结论

本文从影响战场态势生成和分析的关键问题军事目标的识别与定位出发,针对复杂战场环境下军事目标检测任务,构建了一个复杂场景下军事目标检测训练数据集,提出一种基于卷积神经网络 YOLOv3-DAR 的军事目标检测方法。得到以下主要结论:

1) ResNet50-d-DCN 特征提取网络相对于 DarkNet53 在检测形变目标时具有更好的检测性能。

2) 在特征融合阶段引入双注意力机制和特征融合模块,使网络具有自适应过滤功能,各层级特征可以有效地融合在一起,增强了特征的表达能力,使得检测模型的检测性能得到提升。

3) 基于 DIUO 损失函数和 Focal 损失函数对 YOLOv3 网络的损失函数进行改进,结合 DIUO-NMS,进一步提高了复杂环境下军事目标的检测精度。

4) YOLOv3-DAR 检测模型具有较高的检测精度、速度,可以满足实际作战需求,可以为战场态势生成、分析提供有效的辅助技术支持。

本文提出的检测方法的基础是需要大量的实际样本数据,按照标准对目标进行人工标注,供模型训练使用。目前,本文针对复杂环境下军事目标检测任务构建的数据集样本数据还不够完善,将对其进行补充扩展,丰富样本数据。此外,由于实际战场应用中有可能出现未知类别目标,针对该问题,后期将对数据集添加语义标签,通过修改模型结合零样本分类器进行未知类别目标检测,这将是本文进一步研究的重点。

## 参考文献 (References)

- [1] MILLER A. Situational awareness-from the battlefield to the corporation[J]. Computer Fraud & Security, 2006, 2006(9): 13-16.
- [2] PENG H, ZHANG Y F, YANG S, et al. Battlefield image situational awareness application based on deep learning[J]. IEEE Intelligent Systems, 2020, 35(1): 36-43.
- [3] HANRATTY T P, HAMMELL I R J, BODT B A, et al. Enhancing battlefield situational awareness through fuzzy based value of information[C] // Proceedings of the 46th Hawaii International Conference on System Sciences. Wailea, HI, US: IEEE, 2013: 1402-1411.
- [4] LIU S, LIU Z. Multi-channel CNN-based object detection for enhanced situation awareness [C] // Proceedings of the 9th NATO Military Sensing Symposium (2017-11-30) [2021-02-18]. https:

//arxiv.org/abs/1712.00075.

- [5] 戴文君, 常天庆, 褚凯轩, 等. 基于时空卷积特征记忆模型的坦克火控系统视频目标检测方法[J]. 兵工学报, 2020, 41(9):1708-1718.
- DAI W J, CHANG T Q, CHU K X, et al. Video object detection method for tank fire control system based on spatial-temporal convolution feature memory model [J]. Acta Armamentarii, 2020, 41(9):1708-1718. (in Chinese)
- [6] 包壮壮, 赵学军, 王明芳, 等. 脱离预训练的多尺度目标检测网络模型[J]. 计算机工程, 2020, 46(6):248-255.
- BAO Z Z, ZHAO X J, WANG M F, et al. Multi-Scale target detection network model trained from scratch [J]. Computer Engineering, 2020, 46(6):248-255. (in Chinese)
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] // Proceedings of the European Conference on Computer Vision. Amsterdam, the Netherlands: Springer, 2016: 21-37.
- [8] 戴文君, 常天庆, 张雷, 等. 面向坦克火控系统的多尺度形变目标检测方法[J]. 计算机辅助设计与图形学学报, 2019, 31(12):2082-2090.
- DAI W J, CHANG T Q, ZHANG L, et al. Multi-scale deformable target detection method for tank fire control system [J]. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(12):2082-2090. (in Chinese)
- [9] HE K M, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, US: IEEE, 2016: 770-778.
- [10] DAI J F, LI Y, HE K M, et al. R-FCN: object detection via region-based fully convolutional networks [C] // Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates, 2016: 379-387.
- [11] 杜泽星, 殷进勇, 杨建. 基于密集连接网络的遥感图像检测方法[J]. 激光与光电子学进展, 2019, 56(22):210-216.
- DU Z X, YIN J Y, YANG J. Remote sensing image detection based on dense connected networks [J]. Laser & Optoelectronics Progress, 2019, 56(22):210-216. (in Chinese)
- [12] 杨传栋, 刘桢, 石胜斌. 基于 CNN 的弹载图像目标检测方法研究[J]. 战术导弹技术, 2019(4):85-92.
- YANG C D, LIU Z, SHI S B. Research on object detection for missile-borne image based on CNN [J]. Tactical Missile Technology, 2019(4):85-92. (in Chinese)
- [13] 王志, 陈平, 潘晋孝. 基于深度学习的复杂背景下目标检测[J]. 重庆理工大学学报(自然科学), 2018, 32(4):171-176.
- WANG Z, CHEN P, PAN J X. Object detection based on deep learning in complex background [J]. Journal of Chongqing University of Technology (Natural Science), 2018, 32(4):171-176. (in Chinese)
- [14] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmenta-
- tion [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, US: IEEE, 2014: 580-587.
- [15] GIRSHICK R. Fast R-CNN [C] // Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 1440-1448.
- [16] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6):1137-1149.
- [17] 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. Las Vegas, NV, US: IEEE, 2016: 779-788.
- [18] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C] // Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, US: IEEE, 2017.
- [19] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. (2018-04-08) [2021-02-18]. https://arxiv.org/abs/1804.02767.
- [20] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks [C] // Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 764-773.
- [21] HE T, ZHANG Z, ZHANG H, et al. Bag of tricks for image classification with convolutional neural networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, US: IEEE, 2019: 558-567.
- [22] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression [C] // Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, NY, US: AAAI, 2020: 12993-13000.
- [23] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2):318-327.
- [24] EVERINGHAM M, GOOL L V, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge [J]. International Journal of Computer Vision, 2010, 88:303-338.
- [25] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8):2011-2023.
- [26] 鞠默然, 罗江宁, 王仲博, 等. 融合注意力机制的多尺度目标检测算法[J]. 光学学报, 2020, 40(13):126-134.
- JU M R, LUO J N, WANG Z B, et al. Multi-scale target detection algorithm based on attention mechanism [J]. Acta Optica Sinica, 2020, 40(13):126-134. (in Chinese)
- [27] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, US: IEEE, 2017: 2117-2125.