

基于注意力与特征融合的工程机械目标检测方法

吴靖, 赵尔敦, 林卓成, 秦文清
(华中师范大学, 武汉 430079)

摘要: **目的** 针对施工环境中工程机械目标大小不一、相互遮挡、工作形态各异等问题, 提出一种基于注意力与特征融合的目标检测方法(AT-FFRCNN)。**方法** 在主干网络中采用 ResNet50 和特征路径聚合网络 PFPN, 融合不同尺度的特征信息, 在区域建议网络(RPN)和全连接层引入注意力机制, 提高目标识别的能力, 在损失函数中使用广义交并比(GIoU), 提高目标框的准确性。**结果** 实验表明, 文中提出方法检测准确率比其他方法有较大提高, 检测平均准确率(mAP)达到90%以上。**结论** 能够较好地完成工程机械目标的检测任务。

关键词: 目标检测; Faster R-CNN; 注意力机制; 特征融合

中图分类号: TP391 **文献标识码:** A **文章编号:** 1001-3563(2022)15-0061-07

DOI: 10.19554/j.cnki.1001-3563.2022.15.007

Object Detection Method of Construction Machinery Based on Attention and Feature Fusion

WU Jing, ZHAO Er-dun, LIN Zhuo-cheng, QIN Wen-qing

(Central China Normal University, Wuhan 430079, China)

ABSTRACT: The work aims to propose an object detection method based on attention and feature fusion (AT-FFRCNN) aiming at the problems of different size, mutual occlusion and different working forms of construction machinery objects in the construction environment. ResNet50 and feature path aggregation network PFPN were used in the backbone network to fuse feature information of different scales, and an attention mechanism was introduced into the region proposal network (RPN) and fully connected layer to improve the ability of target recognition, and generalized intersection over union (GIoU) was used in the loss function to improve the accuracy of the object box. Experiments indicated that the detection accuracy of the proposed method was greatly improved compared with other methods, and the average detection accuracy (mAP) reached more than 90%. The proposed method can complete the detection task of the construction machinery better.

KEY WORDS: object detection; Faster R-CNN; attention mechanism; feature fusion

在施工现场中, 工程机械目标常常是酿成安全事故的隐患, 存在倒塌、侧翻等不稳定的因素, 因此对工程机械进行目标检测是实现生产安全的重要一环。智能化的视频监控和影像分析可以实时了解施工现场的施工进度、秩序以及安全性。

近年来, 基于深度学习目标检测方法的研究发展十分迅速, 主要分为一阶段和两阶段的检测方法。一阶段的检测方法主要代表有 YOLO^[1]系列方法、CenterNet^[2]、RetinaNet^[3]、SSD^[4]等, 其特点是速度较快, 候选框的生成和分类同时进行, 将目标图片直

收稿日期: 2022-06-09

作者简介: 吴靖(1997—), 男, 华中师范大学硕士生, 主攻深度学习、机器视觉。

通信作者: 赵尔敦(1972—), 男, 博士, 华中师范大学副教授, 主要研究方向为机器视觉。

接送入网络即可生成预测框；两阶段的检测方法有 R-CNN^[5]、Fast R-CNN^[6]、Faster R-CNN^[7]等，不同于一阶段的检测方法，两阶段方法将候选框的生成和分类分开进行，整体网络由提取区域建议网络 RPN 和公共特征层二者相结合而构成，检测精度相对较高。

施工场景的目标检测有以下难点：塔吊、挖机、吊车等工程机械目标形态、颜色各异；真实施工场景监控摄像头的位置距离目标较远，拍摄到的目标尺寸相差较大；目标与建筑物、目标与目标之间会出现相互遮挡的情况；受天气的影响，会存在光照不足或者背光的情况，导致目标不明显。

为了解决上述检测难点，文中基于 Faster R-CNN 提出一种基于注意力与特征融合的工程机械目标检测方法 (AT-FFRCNN)，其特点如下：在主干特征提取网络采用 ResNet50，在特征提取网络之后加上特征路径聚合网络 (PFPN)，让网络能更好地学习不同尺度的特征；在区域建议网络 (RPN) 中引入注意力机制模块 CABM (AT-RPN)，实现对与目标有关的特征通道和空间位置赋予更多的关注，并在全连接层引入注意力机制进行增强 (AT-FC)，对目标建议框周围的建议框的特征向量赋予不同的权值并且进行融合，来更新目标建议框的特征向量；采用 K-means 聚类算法对样本集进行聚类，采用更适用于此问题的先验锚框 (Anchor Box) 宽高比 (该问题场景计算出宽高比为 1:1.4、1:1.7、1.5:1)；回归损失函数采用为广义交并比损失 (GIoU)，解决 SmoothL1 没有考虑预测框与真实框的重叠问题。

1 基于注意力与特征融合的工程机械目标检测方法

文中提出的 AT-FFRCNN 算法框架见图 1，算法步骤如下。

1) 首先将图片送入特征提取模块，经过主干网络 ResNet-50 和特征路径聚合网络 PFPN 后得到特征图。

2) 接着将得到的特征图经过 3×3 的卷积之后得到特征图 F ，随后将特征图 F 送入融合注意力机制的区域建议网络 AT-RPN，得到区域建议框。

3) 将区域建议框映射到特征图 F 上，得到一系列的特征矩阵，再对特征矩阵进行感兴趣区域池化，固定其尺寸大小。后将池化后的特征矩阵与目标建议框送入基于关系增强的全连接层，融合周围目标建议框，获取位置信息更精准的目标建议框，最后再送入分类与回归模块进行目标检测。

1.1 特征路径聚合网络 PFPN

文中将 Faster R-CNN 的特征提取网络由 VGG16^[8]替换成 ResNet50^[9]，虽然能够提取到更为精细的特征，提高了检测大目标的检测效果。由于加深了网络的层数，也丢失了较多的浅层语义特征^[10]，最终导致无法精准地检测出小目标。为了解决这个问题，文中基于特征金字塔网络 FPN 和实例分割网络 (PANet) 的思想^[11]，构造一种特征路径聚合网络 (PFPN)，其结构见图 2。特征路径聚合网络 PFPN 主要由特征金字塔 (FPN) 和自下而上的路径聚合 (Path aggregation, PA) 组成。PFPN 的具体工作流程如下。

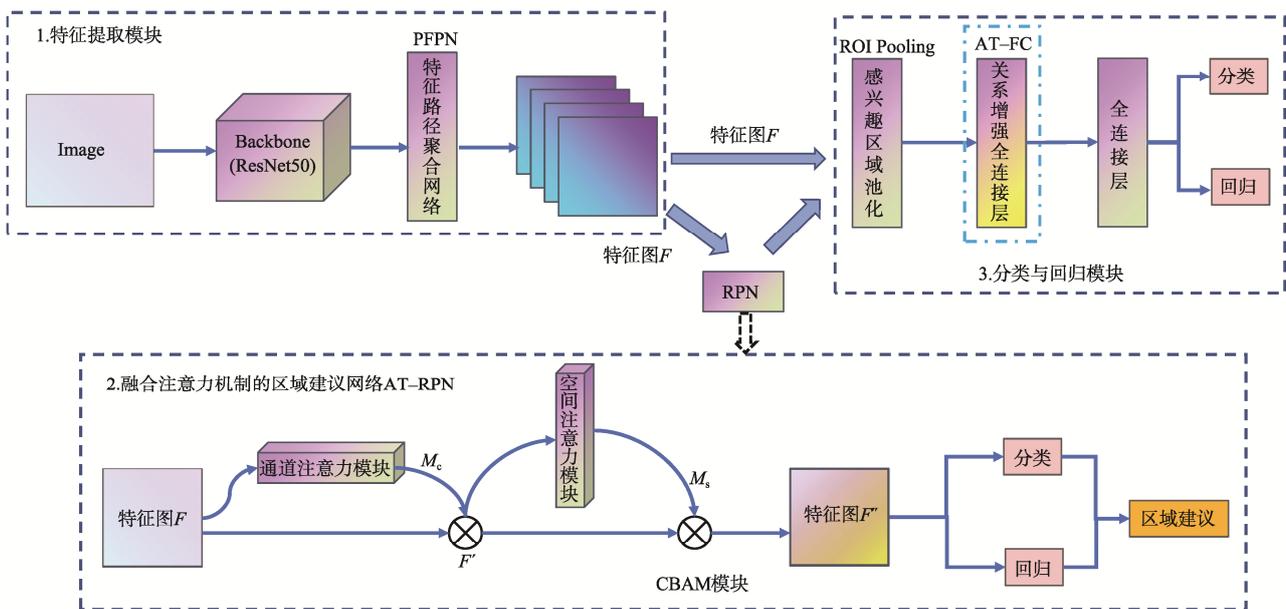


图 1 AT-FFRCNN 算法框架示意图

Fig.1 Schematic diagram of AT-FFRCNN algorithm framework

1) 构建特征金字塔。通过主干特征提取网络 ResNet50 提取图片特征, 得到不同层次、不同尺寸的特征图 $\{C_1, C_2, C_3, C_4, C_5\}$ 。

2) 自上而下特征融合。对特征金字塔 $\{C_1, C_2, C_3, C_4, C_5\}$ 进行特征融合操作得到 $\{P_2, P_3, P_4, P_5\}$, 其具体计算方法如图 2 的虚线框 A 所示。

3) 自下而上路径聚合。对 $\{P_2, P_3, P_4, P_5\}$ 进行路径聚合得 $\{N_2, N_3, N_4, N_5\}$, 其具体的计算方法如图 2 中虚线框 B 所示。

由于 PFPN 添加了自下而上的路径聚合过程, 相较于原始的 FPN, 能够更好地保留浅层特征信息, 增加特征的表达能力, 更有利于小目标的检测。这是因为 FPN 虽然增加了自上而下的特征融合操作, 能够保留一定的浅层特征信息, 但是由于主干网络较深, 浅层特征信息经过了太长的路径到达深层, 丢失的浅层信息仍然较多, 如图 2 中长点划线所示。在增加了自下而上的路径聚合后, 更好地保留了浅层特征信息, 如图 2 中圆点线所示。

1.2 K-means 锚框宽高比设置

为了生成更加适用于工程机械目标尺寸的锚框, 提高算法在目标检测中边界框回归的准确性, 文中算法不再使用 Faster R-CNN 的默认锚框宽高比 (1:1、1:2、12:1), 而是采用 K-means 聚类算法对锚框的宽高比进行设置指导。在文中设计的基于 K-means 的锚框宽高比聚类算法中, 设 m 为设定的锚框宽高比类别数, 簇

C_i 质心定义为 $(W_i, H_i), i \in [1, 2, 3, \dots, m]$, 样本真实目标框 (x_j, y_j, w_j, h_j) 与各簇 C_i 质心距离定义为 $d = 1 - F_{IoU}[(x_j, y_j, w_j, h_j), (x_j, y_j, W_i, H_i)], j \in [1, 2, \dots, N]$, 其中 N 为样本目标框数量, (x_j, y_j) 为目标框中心点, (w_j, h_j) 为目标框的宽和高, F_{IoU} 表示计算交并比的函数。文中取 $m=3$, 通过 K-means 聚类算法最终得到针对文中数据集的先验锚框 (anchor box) 合适的宽高比为 1:1.4、1:1.7、1.5:1。

1.3 融合注意力机制的 AT-RPN 网络

AT-RPN 网络在 RPN 网络的基础上引入了注意力模块 CBAM^[12], 提高 RPN 网络的目标分类与回归能力, 获取更精细的目标候选框。CBAM 包含通道注意力^[13-14]和空间注意力, 这两个部分相互连接、相互补充, 能够突出图像中的目标特征抑制非目标部分 (背景) 的特征, 而且还能突出目标的空间位置信息。

文中采用的 CBAM 模块见图 1, 特征图 $F \in R^{C \times H \times W}$ 通过通道注意力模块后得到通道注意力图 $M_c \in R^{C \times 1 \times 1}$, 将 M_c 与特征图 F 相对应的通道相乘进行加权, 得到通道信息精细的特征图 F' , 即 $F' = M_c(F) \odot F$; C, H, W 分别表示特征图的通道数、高度和宽度, \odot 表示特征图对应元素相乘; 然后再将 F' 送入空间注意力模块得到空间注意力图 $M_s \in R^{1 \times H \times W}$, 最后将 M_s 与特征图 F' 同一通道的不同像素点相乘进行位置加权, 得到位置信息精细的特征图 F'' , 即 $F'' = M_s(F') \odot F'$

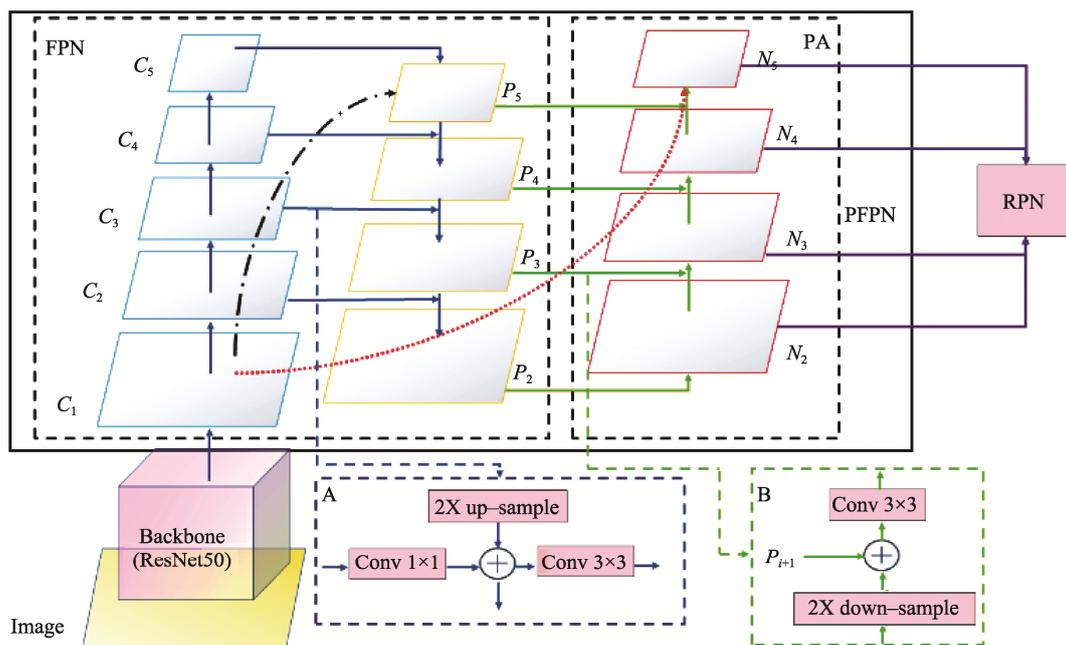


图 2 PFPN 网络结构
Fig.2 Structure of PFPN network

$M_c(F)$ 的计算过程如下, 首先将特征图 F 进行基于宽度 W 和高度 H 的最大池化和平均池化, 得到 $F_{max}^c \in R^{c \times 1 \times 1}$ 和 $F_{avg}^c \in R^{c \times 1 \times 1}$, 然后将这 2 个特征送入共享全连接层, 最后将输出进行对应元素相加, 经过 sigmoid 激活得到 M_c , 见式 (1)。

$$M_c(F) = \sigma(P_{MLP}(P_{avg}(F)) + P_{MLP}(P_{max}(F))) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \quad (1)$$

式中: P_{MLP} 为中间的共享网络层, 是一个多层感知机; W_0 、 W_1 为多层感知机的 2 层参数; $P_{avg}()$ 、 $P_{max}()$ 分别为平均池化操作和最大池化操作; $\sigma()$ 为激活函数。

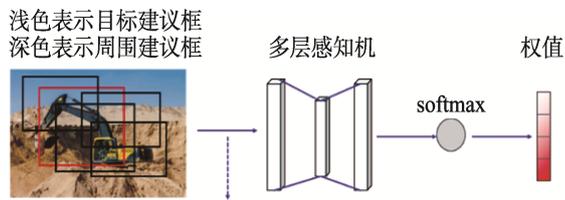
对于 $M_s(F)$ 的计算, 其输入是经过通道注意力机制之后的特征图 F' , 首先对 F' 进行基于通道的平均池化和最大池化得到特征图 $F_{avg}^s \in R^{1 \times H \times W}$ 和 $F_{max}^s \in R^{1 \times H \times W}$, 然后经过 7×7 的卷积操作与 sigmoid 激活操作得到空间注意力图 M_s , 见式 (2)。

$$M_s(F') = \sigma(f^{7 \times 7}([P_{avg}(F'); P_{max}(F')])) = \sigma(f^{7 \times 7}(F_{avg}^s; F_{max}^s)) \quad (2)$$

式中: $\sigma()$ 为 sigmoid 激活操作; $f^{7 \times 7}()$ 为 7×7 的卷积操作。

1.4 基于关系增强的全连接层 AT-FC

在 AT-RPN 网络之后, 存在目标建议框与其交并比 IoU 较大的周围建议框, 这些周围建议框与目标建议框之间存在一定的关系, 其含有的位置信息能够帮助该目标建议框修正位置。为了使目标建议框的位置信息更精准, 文中提出在 RoI 池化之后的全连接层引入注意力机制^[15], 构造了一种基于关系增强的全连接层 AT-FC。AT-FC 的原理见图 3, 运用注意力机制计算得出周围建议框与目标建议框位置信息的关系大小 (权值), 然后将这些周围建议框与其对应的权值相乘, 以加权求和的方式融合并更新目标建议框。通过 AT-FC 更新之后, 目标建议框的位置信息将更加精确。



1. 根据设定的 IOU 阈值挑选周围建议框
2. 后将目标建议框与挑选出的周围建议框的特征向量进行展平连接

图 3 建议框融合算法
Fig.3 Suggestion box fusion algorithm

基于注意力机制的周围建议框融合算法步骤如下。

- 1) 选取经过 AT-RPN 网络得到的 K 个目标建议

框 $P = [p_1, p_2, \dots, p_K]$ 和其对应特征向量 $V = [v_1, v_2, \dots, v_K]$, 特征向量为 K 个目标建议框进行感兴趣区域池化和全连接所得到的。

2) 对步骤 1 中选取的每个目标建议框 p_i 计算其与周围建议框的 IoU 值, 选取 $F_{IoU} \geq T$ 的 J_i 个周围建议框, 用于对目标建议框 p_i 的更新。其中阈值 T 为超参数, 其值根据实验来选取;

3) 运用注意力机制计算每个目标建议框 p_i 与选取的 J_i 个周围建议框之间的权重 $\alpha_{ij}, j \in \{1, 2, 3, \dots, J_i\}$: 先将目标建议框与周围建议框的特征向量进行连接, 输入多层感知机, 并用 softmax 分类器得到权重, 见式 (3)。

$$a_{ij} = \text{softmax}(W_b \sigma(W_a [v_i \cdot v_j])) \quad (3)$$

式中: W_a 和 W_b 为多层感知机学习的权重参数; v_i 和 v_j 为与当前目标建议框的第 i 个和第 j 个周围建议框的对应全连接层的特征向量; $[\]$ 为连接特征向量。

4) 对周围建议框特征向量以加权求和的方式进行融合, 更新目标建议框 p_i 对应的特征向量 V_i , 计算见式 (4)。

$$V_i = \alpha_{i1} \cdot v_1 + \alpha_{i2} \cdot v_2 + \alpha_{i3} \cdot v_3 + \dots + \alpha_{iJ_i} \cdot v_{J_i} \quad (4)$$

5) 将更新后目标建议框特征向量送入后续全连接层, 进行目标分类和位置回归定位。

1.5 广义交并比 GIoU

在 Faster R-CNN 中, 采用 SmoothL1 损失作为目标框的回归损失, 该损失的计算是以预测边界框与真实目标边界框之间的距离度量的, 这种方法没有考虑到预测框与真实框之间的重叠部分, 不利于预测边界框的精确修正。为了解决 SmoothL1 损失的缺点, 回归损失可以使用预测框与真实框的 IoU 来计算, 但当预测边界框与真实边界框之间无重叠部分时 IoU 为 0, 此时无法进行优化, 另外当预测边界框与多个真实边界框 IoU 相同时, 会影响最终的预测结果。为了解决 SmoothL1 损失和 IoU 损失的缺点, 文中使用广义交并比 (GIoU)^[16] 作为回归损失, 其定义如下: 若矩形框 C 为包含 A 和 B 的最小矩形框, 则 A 和 B 的 GIoU 为 $GIoU(A, B) = IoU(A, B) - \frac{|C \setminus (A \cup B)|}{|C|}$ 。使用 GIoU 作为边界框计算回归损失, 其中 \setminus 表示两者之间的差集, 既能解决 SmoothL1 损失没有考虑预测框与真实边界框的重叠程度问题, 也能避免因 IoU 为 0 时无法优化的问题, 从而更有利于预测边界框的回归调整。

1.6 损失函数

损失函数由 2 部分组成, 分别是区域建议网络 RPN 的损失和分类与回归模块的损失。

1.6.1 区域建议网络 RPN 损失

RPN 网络包含锚框的分类与回归, 其损失函数为:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} p_i^* L_{reg}(t_i, t_i^*) \quad (5)$$

$$L_{cls}(p_i, p_i^*) = -\lg[p_i p_i^* + (1 - p_i)(1 - p_i^*)] \quad (6)$$

$$L_{reg}(t_i, t_i^*) = 1 - GIoU(t_i, t_i^*) \quad (7)$$

式中: 分类损失 L_{cls} 为交叉熵损失; 回归损失 L_{reg} 为 GIoU 损失; p_i 为第 i 个锚框预测为真实标签的概率, 当为正样本时, p_i^* 的取值为 1, 否则为 0; t_i 为预测的第 i 个锚框; t_i^* 为第 i 个锚框对应的真实的边界框 (GTBox); N_{cls} 为一个小批量的样本数量; N_{reg} 为锚框位置的个数; λ 为加权平衡参数, 一般设置为 10。

1.6.2 分类与回归模块损失

文中使用的损失函数为:

$$L(p, u, t^u, v) = L_{cls}(p, u) + L_{loc}(t^u, v) \quad (8)$$

$$L_{cls}(p, u) = -\lg(p_u) \quad (9)$$

$$L_{loc}(t^u, v) = 1 - GIoU(t^u, v) \quad (10)$$

式中: L_{cls} 为目标分类损失; L_{loc} 为边界框回归损失; p 为预测目标类别的概率; u 为真实的类别; t^u 为预测对应类别为 u 的目标框; v 为真实的边界框。

2 实验结果与分析

2.1 数据集与参数配置

文中收集了施工现场中 3 类机械目标的样本数据集, 本数据集共包含有塔吊 (crane)、挖机 (excavator) 和吊车 (hoist) 3 类机械设备的图片共 1 415 张, 其中 991 张为训练集, 283 张为验证集, 141 张为测试集。

文中实验采用带动量的随机梯度下降法 (Stochastic Gradient Descent, SGD) 对网络进行训练, 初始学习率为 0.005, 学习率更新策略采用的是 Step,

动量系数为 0.9, 学习率调整倍数为 0.33, 建议框融合算法阈值为 0.7。

2.2 实验结果分析

2.2.1 结果对比

在目标检测任务中, 一般使用精确度 (Precision) 和召回率 (Recall) 来检测是否准确以及是否有漏检。

精确度的计算公式为 $P = \frac{T_P}{T_P + F_P}$, 召回率的计算公式

为 $R = \frac{T_P}{T_P + F_N}$, 其中, T_P 真正例, 在目标检测中也表示与真实框 $F_{IoU} > 0.5$ 的预测框数量, F_P 假正例, 表示与真实框 $IoU \leq 0.5$ 的预测框数量, T_N 为真反例, F_N 为假反例, 表示没有检测出的真实框数量。根据精确度和召回率绘制出 $P-R$ (Precision-Recall) 曲线, 就可以计算网络模型对于某一检测类别的平均精度 (Average Precision, AP), AP 的计算公式为 $P_{AP} = \sum_{r \in (0,1)} P(r) / Q$, Q 为将横坐标 r 划分成的区域数, 其中 r 为召回率, $P(r)$ 为 $P-R$ 曲线中当前区域召回率右侧的最大精确度。

为了评估文中实验的有效性, 文中使用平均精度均值 (Mean Average Precision, mAP) 来衡量模型检测的结果, 平均精度均值定义为 $P_{mAP} = \sum_i^N P_{AP_i} / N$ 其中 N 表示待检测的类别的个数。 P_{mAP} 的取值范围为 0 到 1, 其值越高代表网络模型检测性能越好。

与 Faster R-CNN 不同, AT-FFRCNN 算法提出了多种模块, 在文中的数据集下进行消融实验, 探求各模块对实验效果的影响, 实验对比结果见表 1。表 1 中, mAP 表示目标检测的评价精确率。由表 1 可知, Faster R-CNN 网络的 mAP 值为 77.8%, 而将主干网络调整为 ResNet50 之后, mAP 值提升了 3.9%; 在此基础之上加入特征金字塔网络 FPN 后, mAP 值又提升了 1.8%; 引入文中构造的特征路径聚合网络 FPFN 后, mAP 值则提升了 1.9%。后继续通过 K-means 算法对样本聚类, 修改 RPN 网络的锚框宽高比后, mAP 值又提升了 2.8%, 最后在引入 GIoU 回归损失后, mAP 值达到了 90.2%。

表 1 不同改进下的 mAP 对比

Tab.1 mAP comparison under different improvements

主干网络	FPN	特征路径聚合网络 FPFN	K-means 锚框调整	GIoU	mAP 值/%
VGG16	—	—	—	—	77.8
ResNet50	—	—	—	—	81.7
	√	—	—	—	83.5
	√	√	—	—	85.4
	√	√	√	—	88.2
	√	√	√	√	90.2

2.2.2 不同锚框宽高比的实验结果对比

利用文中K-mean聚类算法对锚框宽高比进行调整的多组实验结果见表2。从表2中可以得出,锚框的宽高比例并不是越多越好,也即K-means对样本聚的簇不是越多越好,最适合文中问题场景的锚框宽高比为1:1.4、1:1.7、1.5:1,随着宽高比例数量的增加,模型的检测精度mAP值会变低。对比表格的第1行和第2行可知,相较于Faster R-CNN的默认锚框宽高比,使用K-means聚类算法对锚框宽高比进行调整后,网络模型的检测精度mAP值比初始Faster R-CNN提高了3.5%。

表2 不同锚框宽高比下的检测mAP
Tab.2 Detection mAP with different anchor box aspect ratios

网络主体	锚框的宽高比	mAP 值/ %
ResNet50+ PPFN+ GloU	1:1、1:2、2:1	86.7
	1:1.4、1:1.7、1.5:1	90.2
	1:1.4、1:1.3、1:1.2、1.6:1	86.4
	1:1.7、1:1.2、1:1.3、1.7:1、1.2:1	82.5

2.2.3 不同算法模型实验结果对比

为了验证文中提出AT-FFRCNN算法的效果,文中也与当前其他目标检测算法进行了对比,不同网络模型实验结果评价指标见表3。

从表3中可以看出,Faster R-CNN在使用文中数据集进行训练并在验证集上进行验证得到的AP50和AP75值分别为86.7%和68.9%,mAP值为77.8%。文中AT-FFRCNN算法的AP50和AP75值相较于

Faster R-CNN分别提高了10.5%、20.4%,并且mAP值达到了95.4%,在原来的基础上提升了17.8%。同其他目标检测算法比较,文中AT-FFRCNN算法的各项指标均提高了不少,相较于SSD,AP50提高了13%,AP75提高了22.5%,mAP提高了20.1%,相较于YOLOV3,AP50、AP75、mAP则分别提高了9.4%、19.8%、17.3%,这也充分证明该算法在Faster R-CNN基础上的改进是有效的,能够解决文中引言中提出的问题。

表3 不同网络模型实验结果评价指标对比
Tab.3 Comparison of evaluation indicators of experimental results of different network models

主干网络	AP50 值/%	AP75 值/%	mAP 值/%
Faster R-CNN	86.7	68.9	77.8
SSD	84.2	66.8	75.5
YOLOv3	87.8	69.5	78.3
AT-FFRCNN	97.2	89.3	95.6

注:评价指标AP50和AP75分别表示预测框与真实框交并比IoU值大于0.5和大于0.75的检测精度。

2.2.4 检测效果示例

图4是文中算法与Faster R-CNN检测效果图,其中图4a为Faster R-CNN算法的检测效果,图4b为AT-FFRCNN算法对应的检测效果。可以看出,AT-FFRCNN检测出的工程机械目标准确率更高,位置信息也更准确,并且能够识别出尺寸较小和不明显的目标,针对Faster R-CNN在复杂场景目标遮挡、交叉的特殊情况下的漏检、错检问题有显著改善。

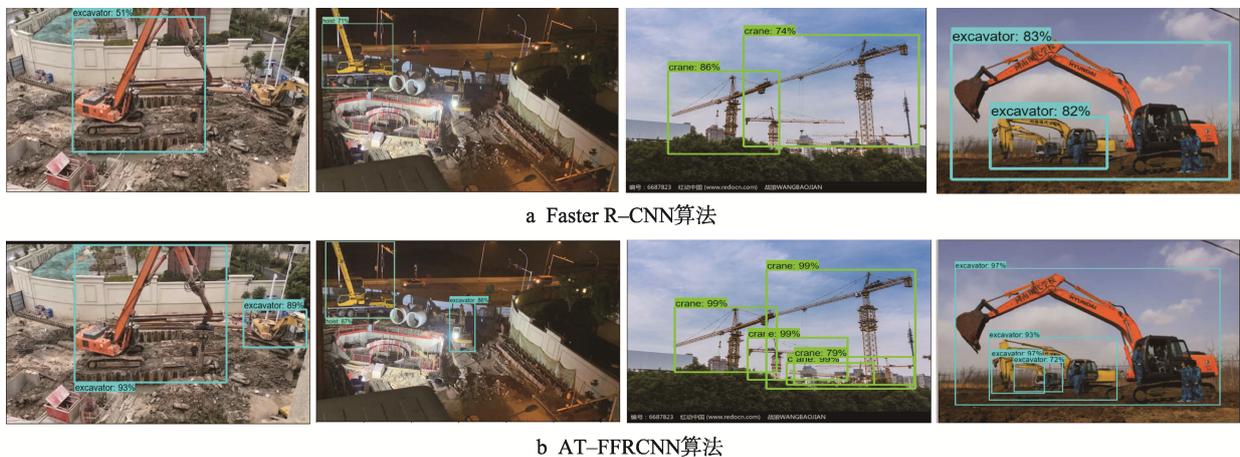


图4 2种算法的检测效果
Fig.4 Detection effect of two algorithms

3 结语

为实现对施工现场工程机械目标的准确检测,文中提出了一种基于多尺度特征融合的施工目标检测方法 AT-FFRCNN。该方法采用了 ResNet50 作为主干特征提取结构,结合构造的特征路径聚合网络 PFPN,提高了对不同尺寸目标的检测精度;使用 GIoU 和 K-means 锚框宽高比聚类算法,弥补了目标定位不准确的问题;在区域建议网络和全连接层引入注意力机制进行增强(AT-FC),提高了复杂环境中目标检测的能力。文中对设计的算法进行了相关对比实验,实验结果表明,文中提出的算法能实现对施工隐患目标更精准的检测,满足了实际应用的需求。

参考文献:

- [1] REDMON J, DIVVALA S, GIRSHICK R, et al. You only Look Once: Unified, Real-Time Object Detection[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, 10: 779-788.
- [2] DUAN KAIWEN, BAI SONG, XIE LINGXI, et al. CenterNet: Keypoint Triplets for Object Detection[C]// Proceedings of 2019 IEEE/CVF International Conference on Computer Vision, 2019: 6568-6577.
- [3] 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.
- [4] LIU WEI, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]// European conference on computer vision, Springer, Cham, 2016: 21-37.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [6] GIRSHICK R. Fast R-Cnn[C]// Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
- [7] REN S, HE K, 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.
- [8] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C]// International Conference on Learning Representations, 2015: 4011-4019.
- [9] HE KAIMING, ZHANG XIANGYU, REN SHAOQING, et al. Deep Residual Learning for Image Recognition[C]// Conference on Computer Vision and Pattern Recognition. IEEE, 2015: 1212-1231.
- [10] ZAGORUYKO S, LERER A, LIN T Y, et al. A Multi-Path Network for Object Detection[J]. Computer Vision and Pattern Recognition, 2016,23(4):1604.
- [11] LIU SHU, QI LU, QIN HAIFANG, et al. Path Aggregation Network for Instance Segmentation[C]// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [12] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block Attention Module[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [13] HU Jie, SHEN Li, SUN Gang, et al. Squeeze-and-Excitation Networks[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [14] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020, Issue: 11531-11539.
- [15] YANG JIANWEI, LU JIASEN, LEE S, et al. Graph R-CNN for Scene Graph Generation[C]// Proceedings of 15th European Conference on Computer Vision. Munich, Germany, 2018:670-685.
- [16] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized Intersection over Union: A Metric and a Loss for Bounding Box Regression[C]// Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, 2019: 658-666.