

# 基于改进型 Faster R-CNN 的 仓储环境物体识别技术研究

周诗捷<sup>1</sup>, 王玉槐<sup>1†</sup>, 沈思橙<sup>1</sup>, 陈在娥<sup>1</sup>, 韩江涛<sup>1</sup>, 陈昱臻<sup>2</sup>

(1. 杭州师范大学 钱江学院, 浙江 杭州 310036; 2. 浙江省测绘科学技术研究院, 浙江 杭州 311100)

**摘要:**为解决传统目标检测精确度不高、有效性差、难以适应仓储环境下多目标识别应用场景的问题,提出了一种改进型 Faster R-CNN 目标检测算法。首先,采用 ResNet50 替换 VGG16 作为特征提取网络,以提高模型的检测精度;同时,为兼顾多尺度及小目标物体的检测,引入了特征金字塔网络,形成了残差金字塔特征提取网络 ResFPN;其次,引入了注意力机制,提高输入特征的空间和通道有效信息利用率;最后,使用 RoI Align 代替原有的 RoI Pooling,以消除因量化取整而产生的预测框回归误差。在经图像增广处理的自建数据集上进行实验测试,结果表明,提出的改进型 Faster R-CNN 算法在仓储环境下能满足对人员、叉车和托盘的目标检测需求,其平均检测精确度能达到 90.2%。

**关键词:** 仓储环境; 目标检测; 注意力机制; Faster R-CNN

**中图分类号:** TP391.41, TP183

**文献标识码:** A

## Research on Object Recognition in Warehouse Environment Based on Improved Faster R-CNN

ZHOU Shijie<sup>1</sup>, WANG Yuhuai<sup>1†</sup>, SHEN Sicheng<sup>1</sup>, CHEN Zaie<sup>1</sup>, HAN Jiangtao<sup>1</sup>, CHEN Yuzhen<sup>2</sup>

(1. Hangzhou Normal University Qianjiang College, Hangzhou, Zhejiang 310036, China;

2. Zhejiang Academy of Surveying and Mapping, Hangzhou, Zhejiang 311100, China)

**Abstract:** In order to solve the problem of low accuracy and poor effectiveness of traditional target detection, which is difficult to adapt to the application scenarios of multi-target recognition in warehouse environment, an improved Faster R-CNN target detection algorithm is proposed. Firstly, ResNet50 is used to replace VGG16 as the feature extraction network to improve the detection accuracy of the model. At the same time, in order to take into account the detection of multi-scale and small target objects, a feature pyramid network is introduced to form a residual pyramid feature extraction network called ResFPN. Secondly, attention mechanism is introduced to improve the effective information utilization rate of the input feature space and channels. Finally, ROI Align is used to replace the original ROI Pooling to eliminate the prediction box regression error caused by quantization rounding. The experimental tests were conducted on the self-built data set with data augmentation. The experimental results show that the improved Faster R-CNN algorithm proposed in this paper can meet the detection requirements of targets such as people, forklifts and pallets in the warehouse environment with an average detection accuracy of 90.2%.

**Key words:** warehouse environment; object detection; attention mechanism; Faster R-CNN

收稿日期: 2023-03-26

基金项目: 杭州师范大学钱江学院学生科研项目(2022QJXS16); 杭州师范大学钱江学院理工分院学生课外科技项目(2022LGXS13)

作者简介: 周诗捷(2001—), 女, 浙江温州人, 本科, 研究方向: 深度学习, 图像处理技术及应用。

† 通信联系人, E-mail: 5472728@qq.com

针对仓储环境的智能搬运系统是智慧物流和智能制造的重要组成部分,其中托盘和叉车等的智能识别是制造生产和仓储环境智能搬运的重要环节之一。其检测的准确率直接决定着整个仓储系统的容错率,因此对托盘、叉车等目标物体的高精度智能识别与检测研究具有重要意义。

目前,国内外主流的基于深度学习的目标检测方法有单阶段检测算法和双阶段检测算法。前者无需单独寻找候选区域,典型算法有 SSD、YOLO<sup>[1]</sup>等。该类算法端到端检测速度快,但同时存在检测精度偏低的不足。相反,后者通过候选区域生成网络(Region Proposal Network, RPN)获取候选区域,再进行分类,提高检测精度的同时牺牲了一定的检测速度,如 R-CNN 系列等。为了进一步提高其检测速度,Ren 等<sup>[2]</sup>提出了 Faster R-CNN 网络,首次采用 RPN 网络代替分割算法生成候选框并与 Fast R-CNN 结合的方式,使其检测速度提升近 10 倍。

在仓储环境目标检测应用中,李天剑等<sup>[3]</sup>通过融合 SSD 和 DenseNet,实现了对托盘等目标物体的识别,但检测精度较有限仅有 69.5%;刘江玉等<sup>[4]</sup>使用以 VGG16 为特征提取网络的 Fast R-CNN 算法进行托盘检测,但在检测精度提升至 79.7%的同时检测速度变慢;张亚辉等<sup>[5]</sup>在 Faster R-CNN 网络中引入 K-means 算法和 Soft-NMS 算法,优化了预测框的选取。仓储物体识别极易受环境干扰导致模型特征提取不准确。近年来,注意力机制<sup>[6-7]</sup>受到研究者的大量关注。通过引入注意力机制,能有效抑制干扰,增强有效特征。应用注意力机制为环境干扰下仓储目标检测提供了新思路。

综上所述,本文提出一种针对仓储环境下多类物体高精度识别的改进型 Faster R-CNN 算法,采用自建数据集训练模型,并将其应用到仓储环境托盘、叉车及人员检测识别中。

## 1 Faster R-CNN 算法及其改进

### 1.1 Faster R-CNN 算法

Faster R-CNN 是一种两阶段目标检测算法,其主要由特征提取网络、候选区域生成网络、目标检测网络组成。

Faster R-CNN 算法流程可分为三个步骤:一是先将图像输入到特征提取网络中得到相应的特征图;二是将特征图共享至 RPN 结构生成候选框,再将该候选框投影到特征图上获得相应的特征矩阵;三是将每个特征矩阵通过感兴趣区域池化层(RoI Pooling)调整特征图大小,再将特征图展平并通过一系列全连接层得到预测结果。

### 1.2 Faster R-CNN 算法的改进与实现

为了提高仓储环境目标检测的精确度,结合目标易被遮挡等实际问题,本文对 Faster R-CNN 算法的改进主要包括以下四个部分:一是特征提取网络更换为 ResNet50 网络,以增强特征提取能力;二是引入特征金字塔网络(Feature Pyramid Network, FPN),以实现多尺度特征图融合;三是引入注意力机制(Convolutional Block Attention Module, CBAM),以使特征覆盖到待识别物体的更多部位<sup>[8]</sup>;四是 RoI Pooling 改用采用了双线性插值法的 RoI Align,以解决 RoI Pooling 操作中两次量化造成的区域不匹配的问题<sup>[9]</sup>。所提出的改进型 Faster R-CNN 网络结构如图 1 所示。

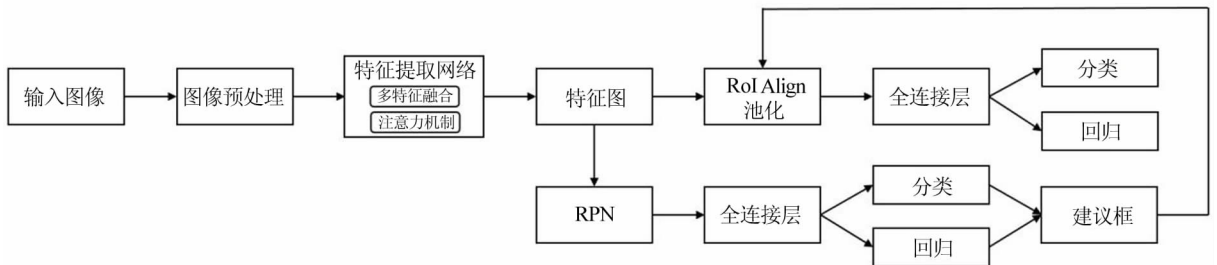


图 1 改进型 Faster R-CNN 网络结构图

#### 1.2.1 特征提取网络

Faster R-CNN 的特征提取网络 VGG16,它是通过小卷积核的堆叠来加深网络层数,从而提升网络性能,但随着网络层数的不断增加,梯度消失、梯

度爆炸和“退化”等问题也随之凸现,致使训练效果变差<sup>[10]</sup>。

为了缓解上述问题,训练更深的网络,得到更好的训练效果,选择具有三层残差模块的 Res-

Net50 作为特征提取网络进行特征提取。

### 1.2.2 残差多尺度特征提取网络

在特征提取网络中,图像经过深层网络的多层卷积后得到的特征图会丢失图片细节信息,导致小尺度目标不易被检测。针对这一问题,将 FPN 网络添加到残差网络中,使其在不显著增加原有模型计算量的情况下,能大幅度提升深度网络和提升对小目标检测的性能<sup>[11]</sup>。如图 2 所示,通过自下而上的路径、自上而下的路径和横向连接三个部分,将浅层细节特征和深层语义特征融合至各特征层,最终得到特征金字塔。

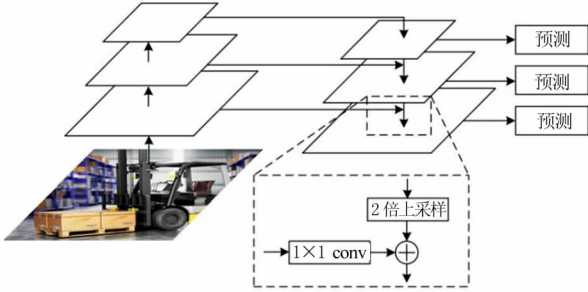


图 2 特征金字塔网络

在选取 ResNet50 网络进行特征提取的基础上,通过增加 FPN 构成残差多尺度特征提取网络 ResFPN,其结构如图 3 所示。{C2,C3,C4,C5} 分别表示为 ResNet50 第 2、3、4、5 层卷积输出的特征图,{P2,P3,P4,P5} 分别表示经 FPN 输出包含丰富细节和语义信息的多尺度特征图。

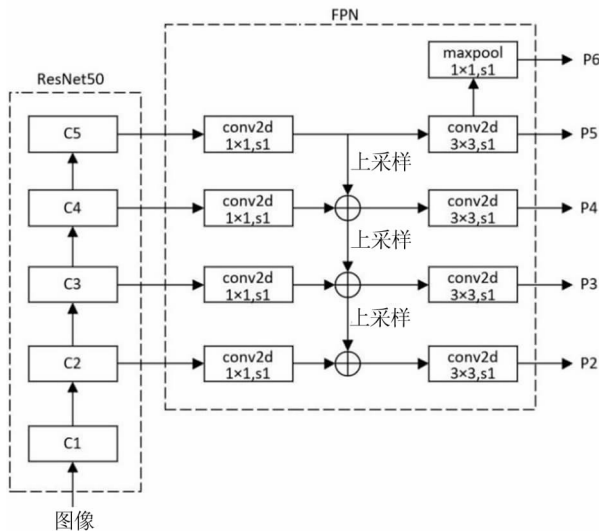


图 3 ResFPN 网络结构

### 1.2.3 注意力机制

为了提高网络模型对检测目标的注意力,抑制图像背景和其他噪声的特征干扰,引入了注意力机

制模块 CBAM 对特征提取网络进行优化。CBAM 由通道注意力(Channel Attention Module,CAM)和空间注意力(Spatial Attention Module,SAM)两个模块连接而成。其实现步骤包括:

首先,将输入特征图  $F \in R^{C \times H \times W}$  传入通道注意力模块,这里  $R$  表示实数空间, $C$ 、 $H$  和  $W$  分别表示输入特征图的通道数、高度和宽度,从而得到通道注意力图  $M_c \in R^{C \times 1 \times 1}$ ,如式(1)所示。

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (1)$$

其中,  $\sigma$  表示 Sigmoid 函数,  $MLP$  表示多层感知机,  $AvgPool$  表示平均池化,  $MaxPool$  表示最大池化。将  $M_c$  再与  $F$  进行矩阵相乘得到输出特征图  $F'$ ,如式(2)所示。

$$F' = M_c(F) \otimes F \quad (2)$$

然后,再将  $F'$  作为空间注意力模块的输入特征图,经过空间注意力模块,得到空间注意力图  $M_s \in R^{1 \times H \times W}$ ,如式(3)所示。

$$M_s(F') = \sigma(f^{7 \times 7}([AvgPool(F'); MaxPool(F')])) \quad (3)$$

其中,  $f^{7 \times 7}$  表示  $7 \times 7$  大小的卷积核。

最后,  $M_s$  再与  $F'$  进行相乘得到最终输出特征图  $F''$ ,如式(4)所示。

$$F'' = M_s(F') \otimes F' \quad (4)$$

### 1.2.4 RoI Align

RoI Pooling 是卷积神经网络在目标检测任务中被广泛使用的操作,是对不同尺寸的候选区域进行最大池化,从而得到固定尺寸的特征图,其间需经过两次量化操作。然而,量化会造成区域建议框原始回归位置与现在位置产生偏差,引起区域不匹配的问题,从而影响目标检测的精确性。

RoI Align 是一种区域特征聚集方式。相比于 RoI Pooling,其在映射和分割单元后均保留浮点数,再使用双线性插值法来计算各单元的四个采样点位置,最后经最大池化得到固定尺寸的特征图。RoI Align 的区域特征聚集特点使 RoI Pooling 引起的区域不匹配得以解决。因此,在本文检测任务中将 RoI Pooling 替换为 RoI Align,以提高模型检测的精确性。

## 2 数据集的准备与制作

### 2.1 数据集制作

在仓储环境中,所检测的目标主要包括人员、叉车和托盘三类。所有数据集图像来自网络公开

图像和实地现场拍摄,共有 1174 张,其中分别包含人员、叉车和托盘的图像 356 张、367 张和 802 张。数据集通过 LabelImg 软件进行标定标签,使每张图像生成对应的 xml 文件,其中包含图像名、图像宽高度、目标类别和目标框坐标等信息。

## 2.2 图像增广

网络模型训练依赖大量图像,然而自建数据集

总量过少且三类目标样本数不均衡。因而,需要进行图像增广,使模型具有更好的适应性和防止出现过拟合现象。利用 Python 语言和 OpenCV 库对图像进行像素增强和空间增强,具体包括图像亮度、饱和度、对比度、随机剪裁和随机翻转等处理,如图 4 所示。经图像增广后,数据集有 1574 张图像。



图 4 图像增广效果图

## 3 实验结果及分析

### 3.1 实验环境及参数设置

软件平台是 64 位 Windows 10 操作系统,硬件平台为 Intel Core i9-10900K 十核 3.70 GHz 的 CPU, NVIDIA GeForce RTX 3070 8 GB 的 GPU,内存 32 GB。深度学习框架为 Anaconda3+Py Torch,利用 Python 编程语言实现仓储环境下的目标检测模型的训练与测试。

本实验采用改进型 Faster R-CNN 作为目标检测模型,数据集按 7:3 比例随机划分成训练集和验证集,并以 PASCAL VOC2012 为样本数据集格式。改进型 Faster R-CNN 模型训练关键参数设置如表 1 所示。

表 1 模型训练所用关键参数

网络模型参数变量	参数描述	取值
classes num	目标类别(不含背景)	3
lr	初始学习率	0.01
batch size	批尺寸	8
epoch	训练次数	50
IOU	预测框与真实标签交并比阈值	0.5

### 3.2 实验分析

为了验证本文所提出的目标检测模型的有效

性,将包含人员、叉车和托盘三类的自建数据集在改进型 Faster R-CNN 模型上进行验证,并与传统 Faster R-CNN 模型实验结果进行对比。针对目标检测常用的 AP 和 mAP 两个指标,网络模型对比实验结果如表 2。

表 2 不同网络模型实验对比

改进方式	AP/%			mAP/%
	人员	叉车	托盘	
传统 Faster R-CNN	74.1	86.4	81.4	80.6
改进型 Faster R-CNN	87	96.8	86.8	90.2

由表可见,本文提出的改进型 Faster R-CNN 算法较于传统 Faster R-CNN 算法,在人员、叉车和托盘的识别精确度分别提升了 12.9%、10.4%、5.4%。三类物体识别精确度的最小提升发生在托盘检测中,分析其原因,在于托盘高度较低且经常存在被货物等遮挡情况,致使识别难度加大,精确度提升幅度受限。同时,改进型 Faster R-CNN 算法的 mAP 也提高了 9.6%,达到了 90.2%。

图 5 为本文改进型模型训练损失值和学习率曲线图,其横轴为训练的迭代次数,左纵轴为损失值,右纵轴为学习率。学习率采用等间隔下降方法进行更新,每间隔 3 个 epoch 使其学习率缩小为原来的 0.33。由图可见,损失值在前 5 个 epoch 快速下降,之后下降速度逐渐放缓,并于第 15 个 epoch 后趋于稳定,表明本文所提模型的有效性。

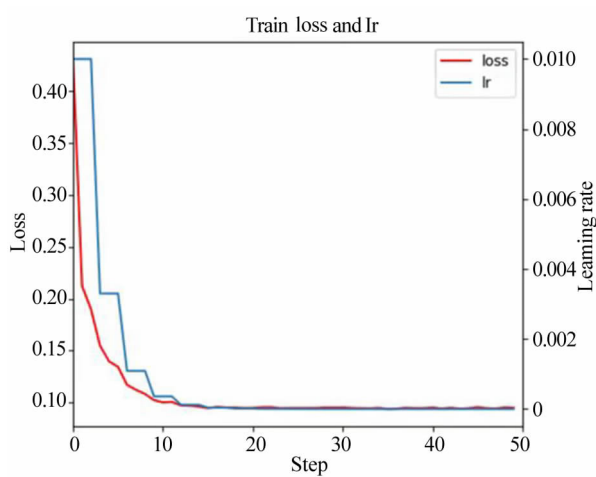


图5 损失值和学习率曲线图

## 4 结论

针对仓储环境下目标物体高精度检测的要求,提出了一种改进型 Faster R-CNN 算法。该算法通过 ResFPN 和注意力机制,实现了特征融合,增强了模型特征提取能力;加之使用 RoI Align,提高了预测框的精确度。经过对比实验发现,所提出改进型 Faster R-CNN 算法比传统 Faster R-CNN 算法在对仓储环境人员检测提升最大为 12.9%,对托盘检测提升最小为 5.4%,验证了本文算法在仓储环境物体检测识别的可行性和有效性。

## 参考文献

- [1] REDMON J, FARHADI A. YOLOv3: an incremental improvement [R]. arXiv e-prints: arXiv, 2018: 1804.02767.
- [2] 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.
- [3] 李天剑,黄斌,刘江玉,等.卷积神经网络物体检测算法在物流仓库中的应用[J].计算机工程,2018,44(6):176-181.
- [4] 刘江玉,李天剑.基于深度学习的仓储托盘检测算法研究[J].北京信息科技大学学报(自然科学版),2017,32(2):78-84+94.
- [5] 张亚辉,杨林,白雪.基于 Faster RCNN 网络的仓储托盘检测方法[J].现代信息科技,2021,5(2):57-62.
- [6] 陶磊,李天剑,胡欢.基于改进 Mask R-CNN 的纸箱堆放分割与定位方法[J].北京信息科技大学学报(自然科学版),2020,35(3):85-88.
- [7] WOO S, PARK J, LEE J Y, et al. Cham: convolutional block attention module [C]. Proceedings of the European Conference on Computer Vision (ECCV), Cham: Springer, 2018: 3-19.
- [8] 杨永波,李栋.改进 YOLOv5 的轻量级安全帽佩戴检测算法[J].计算机工程与应用,2022,58(9):201-207.
- [9] 李恒凯,肖松松,王秀丽,等.基于 Mask R-CNN 的高分遥感影像的稀土开采识别方法[J].中国矿业大学学报,2020,49(6):1215-1222.
- [10] 谭鹤群,李玉祥,朱明,等.通过图像增强与改进 Faster-RCNN 网络的重叠鱼群尾数检测[J].农业工程学报,2022,38(13):167-176.
- [11] 贺艺斌,田圣哲,兰贵龙.基于改进 Faster-RCNN 算法的行人检测[J].汽车实用技术,2022,47(5):34-37.