一种面向轨道车厢底 部的复杂构件检测方法

靳兴来1,2，张育龙2

（1.浙江大学机械工程学院，浙江 杭州 310000；2.杭州国辰机器人科技有限公司，浙江 杭州 310000）

摘 要：本文提出了一种添加了形状注意力机制的目标检测方法，通过注意力模块来预测目标物的尺度和纵横比，引导锚框的选择。与传统方法相比，本文所提出的方法在平均查全率和精度上有较好的提升，并且并没有过多的提高模型的复杂度，保证了实时性，为解决现有目标检测算法在复杂场景中精度低、计算量大、鲁棒性不足的问题提供可能的思路。实验结果表明，所提出的方法可以在不增加计算量的情况下快速精确地定位出目标构件，很好地应用于轨道车辆吹扫作业等复杂环境中。

关键词：xx

中图分类号：TP242

A detection method for complex components facing the bottom of rail cars

JIN Xinglai1,2, Zhang Yulong1

(1.Hangzhou Guochen Robot Technology Co., Ltd, Zhejiang Hangzhou 310000, China; 2.School of Mechanical Engineering Zhejiang University, Hangzhou 310000, China)

**Abstract**: xx

**Key words**: xx;

1. Introduction

轨道交通车辆在相对封闭的地下空间（空气中长期悬浮着大量粉尘）长期运行，由于摩擦、静电等原因，在车厢底部会积聚大量粉尘，而这些粉尘具有比电阻不高，吸水性强等特点。粉尘在列车底部的牵引电机、配电箱等设备表面吸附、凝结、沉降，甚至进入电器箱内[1]。聚集在电器表面的粉尘会导致电气元件的绝缘等级和灵敏度降低，同时对电气设备工作时的散热能力造成影响，若长期得不到有效清除将导致电气故障，具有较大的危害性，所以必须定期进行吹扫除尘[2]。

传统的轨道车辆吹扫作业多采用手动人工吹扫，然而其工作效率较低，劳动强度较大，吹扫扬尘多为无组织排放，并且当作业人员处于疲劳状态时，容易吹扫不干净、不彻底。现有的人工吹扫方式已不能满足高密度作业的需求[3]。因此，有必要实现吹扫作业的自动化与智能化，从而解决人工吹扫劳动强度和难度大的问题，提高车辆维护作业质量、作业效率、作业智能化水平。

5G网络和人工智能的发展为自动化检测创造了有利的条件，一种可行的做法是通过机辆车底智能巡检机器人自动化操作，通过多自由度机械臂、机器视觉技术、图像识别技术等替代传统人工检测作业，实现对关键物件的检测、定位，从而有助于精确吹扫，高效解决人工作业存在的问题。然而,由于轨道车厢底部环境的特殊性，现有的目标检测算法如[4,5,6]在应用于轨道车厢底部复杂构件检测时，会遇到以下问题：

1. 由于轨道拍摄时光线较弱，可见度低，导致所拍摄图像的清晰度不够或发生形变，进而影响后续算法的精度；
2. 在进行目标物定位时，由于目标物类别复杂，特征不明显，导致特征点提取不精准并且计算速度较慢；
3. 由于视场环境的复杂性，存在障碍物遮挡、高斯模糊等问题，在算法上还存在识别效率低，无法理解图片语义等问题。

轨道车辆自动化吹扫作业亟需一种精准高效的目标检测方法，来实现复杂轨道场景中的自动化检测。

因此，本文的研究针对上述提到的问题，提出了一种基于卷积神经网络两步检测算法的注意力机制候选区域网络（RPN），运用注意力引导模块对传统的目标识别网络进行改进，对候选区域网络的锚机制进行引导，使得对于轨道车厢底部复杂环境下的构件检测具有更高的精度和鲁棒性。实验结果表明，所提出的方法可以在不增加计算量的情况下快速精确地定位出目标构件，很好地应用于轨道车辆吹扫作业等复杂环境中。

本文的贡献主要为以下两个：

1. 提出了一种添加了形状注意力机制的目标检测方法，解决现有目标检测算法在复杂场景中精度低、计算量大、鲁棒性不足的问题。
2. 由于目前并没有轨道车厢底部的复杂构件训练数据集，因此首次创建了在轨道车厢底部可见度差，干扰物多的环境下的多目标物构件数据集，以获得更好的训练结果，并进行测试。
3. Related work

传统的目标检测技术一般是根据图像的特征点或基于滑动窗口的框架进行匹配。首先使用图像预处理方法对输入图像进行去噪、增强、裁剪等操作，然后使用滑动窗口方法筛选图像的候选区域，再使用经典的特征提取方法，如方向梯度直方图(histogram of oriented gradient, HOG) [7]、Sift [8]等，从中提取特征候选区域，最后利用AdaBoost [9]和支持向量机（support vector machine，SVM [10]等机器学习算法对得到的特征进行分类，然后通过目标类别对目标进行边界回归。然而，这些方法更多地依赖于视觉特征和图像特征的统计分析，需要针对不同的特征选择合适的分类器，难以处理在多种构件叠加的复杂轨道环境中获取的图像，通用性差，鲁棒性不足。

随着计算机视觉技术迅速发展，基于卷积神经网络（CNN）的目标检测方法被证明在大多数情况下具有更高的精度和更好的鲁棒性[11]。目前主流的基于卷积神经网络的目标检测方法主要可以分成两类：一类是单阶段(One-Stage)的目标检测方法，比如YOLO[12]、SSD(Singleshot multibox Detector) [13]、FCOS(Fully Convolutional One-Stage) [14]等，其典型特征是模型的训练是端到端的，训练和判别的速度更快。它们遵循滑动窗口的提取形式，使用手工标定的特征来表示图像块， 如果网格和支持图像在特征空间中具有高相似性，则网格可能包含对象实例，并且对于不同比例和纵横比的对象不灵活。 此外，手工标定的特征往往在视点和同类目标物变化方面表现不佳

与单阶段的目标检测方法相比，双阶段(Two-Stage)的目标检测方法，如Ｒ-CNN(Region-based Convolution Neural Network) [15]、Fast-ＲCNN[16]、Faster-ＲCNN[17]等模型，有更高的检测框准确率和灵活性。其通过一个阶段的图像特征图(Feature Map)提取，再在另一阶段的网络中利用区域提名(Region Proposal)的方式得到准确的目标在图上的位置。早期的双阶段检测器通常直接根据从骨干网络中提取的金字塔特征层次结构进行预测[18,19]。作为开创性工作之一，特征金字塔网络 (FPN) [20]提出了一种自上而下的途径来组合多尺度特征。随后，PANet  [21]在 FPN 之上添加了一个额外的自下而上的路径聚合网络； STDL[22]提出了一个尺度转移模块来利用跨尺度特征；M2det [23]提出了一个 U 形模块来融合多尺度特征，最近，NAS-FPN [24]利用神经架构搜索自动设计特征网络拓扑。虽然它取得了更好的性能，但 这些方法如NAS-FPN 在搜索过程中需要数千 GPU 小时，并且生成的特征网络是不规则的，因此难以解释。在本文中，我们旨在以更直观和更有原则的方式优化双阶段的目标检测方法。

1. Approach

在本节中，我们首先提出设想的场景，然后介绍算法的实际需求。 在本节末尾，我们介绍了我们提出的添加了形状注意力机制的目标检测方法。

**3.1 Scenarios**

让我们假设一种轨道交通车辆底部的粉尘吹扫作业。由于摩擦、静电等原因，在车厢底部会积聚大量粉尘，导致电气元件的绝缘等级和灵敏度降低，同时对电气设备工作时的散热能力造成影响，若长期得不到有效清除将导致电气故障，具有较大的危害性。场景中目标吹扫构件多为多为矩形机箱、电机、圆形仪表盘、三角形各类按钮等规则形状图形，如图1所示。在该系统中，机器人利用激光传感器，对场景进行离线建模，构建2D栅格地图，并通过视觉系统实现对特定目标物的检测，提取位姿，动态修改机械臂进行聚焦和下一步作业。本文重点描述视觉系统。

 Fig. 1. Proposed System Architecture

**3.2 Requirements**

基于上述场景，我们可以总结出实现面向轨道车厢底部的复杂构件检测方法的需求：

* 检测精度高：因为吹扫区域的容许误差较小，该算法需要准确判断目标物位置和大小，因此我们需要提高识别物体坐标点的精度，以帮助机械臂进一步地位姿规划。
* 低计算量：由于机器人设备的计算资源有限，算法需要在不影响图像处理的情况下使用尽可能低的计算资源来实现检测。
* 低处理时间：考虑到机器人作业的实时性需求，算法需要降低计算复杂性并提供低延迟的服务。
* 可扩展性：根据项目需求处理增加的新目标物也是需要考虑的问题之一，因此我们需要在提出的模型中增加算法的可扩展性。

**3.3 Approach**

基于上述需求，我们提出了一种添加了形状注意力机制的目标检测方法，算法的执行过程如图2所示。首先输入的特定尺度的图片，利用所提出的形状特征完整度检测判断图片是否足够清晰；其次将清晰的图片输入构建的特征提取网络得到形状边缘清晰的特征图；随后将特征图输入形状注意力机制引导的RPN网络寻找锚点和锚框的候选集合，并结合选择性搜索学习对应规则形状的权重；接着将上述候选框集合映射到对应原图区域和特征图区域；最后通过精检测产生预测结果。在本文中，我们利用选择性搜索代替传统的滑窗搜索，产生自学习的区域形状特征融合权重，节省学习时间，并使模型更好地适应不同复杂情况下的目标检测任务；同时构建加入形状注意力机制的RPN网络，使用区域形状特征进行预测，强化模型对提取目标局部特征的学习能力，从而实现轨道车厢底部复杂构件的目标识别，解决传统目标识别算法在轨道车厢底部的复杂场景中，由于障碍物遮挡、模糊等问题导致的识别效率低，无法理解图片语义等问题，从而应用于实际工程中。



Fig. 2. Overview of the algorithm flow

1. Design and Implementation

**4.1 Algorithm Architecture**

在本节中，我们介绍了我们提出添加了形状注意力机制的目标检测模型架构设计。该框架由三个部分组成：特征模糊度判断，粗检测和精检测。特征模糊度判断利用形状特征完整度判断图片是否足够清晰，并进行预处理，便于下一步的粗检测；粗检测实现对候选区域进行前景与背景的分离，定位候选区域的位置；最后的精检测输出目标物的类别和坐标，模型具体结构如图3所示。具体来说，首先网络通过输入的特定尺度图片的特征图，提取待检测图片的形状特征，通过欧氏距离计算所述形状特征和规则形状的平均欧氏距离，根据平均欧氏距离的数量大小来确定形状特征完整度。随后将清晰的图片输入特征提取网络，包括特征编码层、稀疏卷积层和特征金字塔网络，其中，特征编码层将所述清晰的待检测图片按照R、G、B划分为等间距的三维通道，对三维通道中的每一层进行局部特征提取，将像素点转换为统一的特征表示；稀疏卷积层进一步抽象特征，增大感受野并学习特征的几何空间表示；特征金字塔网络从不同尺度的语义空间中聚合和拼接特征表示，得到形状边缘清晰的特征图。随后将形状边缘清晰的特征图输入形状注意力机制引导的RPN网络寻找锚点和锚框的候选集合，并结合选择性搜索学习对应规则形状的权重，生成锚点和锚框的候选集合。其中形状注意力机制包括两个注意力分支，分别用于赋予形状特征所包围的前景区域更高的权重，赋予背景区域更低的权重，以创建形状先验；RPN网络根据所述形状先验分割出各种形状特征补丁，并创建锚点和锚框的候选集合；选择性搜索根据锚点和锚框候选集合提供的信息，学习对应规则形状的权重，由于在训练过程中面临前景和背景区域之间极度不平衡的问题，引入了focal loss，最后获得锚点和锚框的候选集合。最后将所述锚点和锚框的候选集合映射到对应原图区域和特征图区域中，通过非极大值抑制算法NMS来搜索局部极大值，对冗余的锚框进行去冗余，获取感兴趣的特征图区域，并对所述感兴趣的特征图区域进行池化和全卷积操作，获得目标检测结果。



Fig. 3. Overview of the algorithm architecture

**4.2 Integrity Judgment of Shape Features**

在本节中，我们将介绍了我们提出的形状特征完整度判断方式。由于视场环境的复杂性，存在光线较弱，障碍物遮挡、模糊等问题，导致所拍摄图像的清晰度不够或发生形变，进而影响后续算法的精度，考虑到系统对于精度性的需求，因此有必要对于图片进行形状特征完整度判断，将足够清晰的图片传入神经网络的后续处理，从而得到更为精确的结果。在本文中，我们创新性地考虑到吹扫场景中目标物多为矩形机箱、电机、圆形仪表盘、三角形各类按钮等规则形状图形，因此预先定义规则形状为三角形、圆形和矩形，最后通过描述欧氏距离，找到输出形状特征和规则形状的平均欧氏距离，根据平均欧氏距离的数量大小来判断图像特征的模糊程度，具体如下：

1. 设备采集到图片后，将待检测图片转换至多尺度空间，此处采取高斯模糊对图像多尺度的特征进行模拟。N维的高斯函数表达式为：

式中，表示的是尺度因子，表示的是模糊半径，也就是核元素与模板中心之间距离。

1. 构建高斯金字塔，以表征图像的多尺度空间，首先采取二维高斯函数（等比标准差）模板对图像开展卷积处理。在尺度下，代表的是图像表示，代表的是高斯核，代表的是输入图像函数，其卷积函数表达式是：

其中，表示尺度因子，经过大小的变化来对图像的尺度进行改变，代表的是像素点的坐标。

1. 随后下采样图像以获得一系列堆叠的塔状结构层，并进行空间中极值点的检测，其中金字塔层的数量是由原始图像的尺寸和顶部图像的尺寸确定的。函数表达式为：

式中，代表金字塔层的数量，和分别是原图像的大小和塔顶图像的大小。

1. 接着进行空间中极值点的检测，对高斯金字塔中相邻尺度图像做差，其构造公式如下：

式中，在尺度下，代表的是的差分图像，代表的是高斯核，代表的是输入图像的函数，代表的是尺度因子，代表的是缩放因子。

1. 最后对各像素同相邻区域中的26个点进行对比，观测上述点是否满足在同尺度域或者邻近尺度域中最值点的条件。 采取Hassion矩阵进一步去掉强边缘不稳定的相应点，并根据相邻像素梯度方向分布的特性来给特征点分配方向，其中任意点的梯度方向和大小的函数表达式为：

式中，表示的是梯度模值，代表的是相同的尺度图像内同关键点邻近的值，表示的是梯度方向。

通过描述欧氏距离，找到输出形状特征和规则形状的平均欧氏距离，根据平均欧氏距离的数量大小来判断图像特征的模糊程度。

**4.3 Feature Extraction Network Improvements**

在本节中，我们将介绍了我们构建的改进的特征提取网络。考虑到场景的复杂性以及目标物形状的特征，本文将传统的卷积神经网络结构优化为由特征编码层、稀疏卷积层和特征金字塔网络组成的联合特征提取网络。

特征编码层将输入图片按照R、G、B划分为等间距的三维通道，对每一层进行局部特征提取，将像素点转换为统一的特征表示，其主体是一个全连接网络，它由一个线性层、一个批量归一化层(BatchNorm)和一个ReLU层组成，用于提取逐点特征，然后利用逐点特征通过逐元素最大池化操作生成局部聚合特征，最后重复聚合特征并将像素点转换为统一的特征表示。

稀疏卷积层进一步抽象特征，增大感受野并学习特征的几何空间表示，本文所采用的稀疏卷积层的核大小为(3,3,3)。

特征金字塔网络(FPN)从不同尺度的语义空间中聚合和拼接特征表示，得到所需要的局部深层信息特征图，为探测器提供头部信息，允许获得每个对象的形状和颜色先验，其结构包括下采样卷积层、卷积层(Conv2D)、反卷积层和连接层，每个卷积层都包含一个 BatchNorm 层和一个 ReLU 层。 其中反卷积层的内核大小取决于扩展尺度，其余所有卷积层都使用3×3的内核。

**4.4 RPN Network Guided by Shape Attention Mechanism**

在本节中，我们将介绍了我们提出的形状注意力机制引导的RPN网络。本文提出的形状注意力机制引导的RPN网络目的是根据上述生成的形状边缘清晰的特征图，寻找锚点和锚框的候选集合，并结合选择性搜索快速精确学习对应规则形状的权重，准确地生产锚点和锚框的候选集合。

RPN网络最早是在Faster-RCNN中提出，传统的RPN算法的输入是深度卷积神经网络提取出的特征图，输出是候选建议区域，整个网络的作用是寻找目标可能存在的区域并进行粗检测。但是，传统RPN网络也存在这一些问题：1）候选框的分类和回归计算量很大，无法达到实时检测的目的；2）不同锚点所产生的锚框尺寸是相同的，缺乏多样性，进而导致小目标检测效果不好。

人的视觉系统可以对于人眼接收到的图像进行选择性处理，启发于此，基于注意力机制（Attention Mechanism）的深度神经网络被广泛应用到各个领域，模型可以利用注意力机制通过训练不断提取更为有效的注意区域，从而提升模型的性能。因此我们针对于以上问题，基于选择性搜索（Selective Search）算法的原理，引入形状注意力机制，采用边缘、形状等特征来寻找更为有效的候选网络区域，从而实现端到端的训练，在精度上和速度上满足应用实际需求。本文提出的基于形状注意力机制的候选区域网络，锚框大小和尺寸不再固定，并且能够更好地包围目标。

形状注意力机制需要训练大量的注意力权重，考虑到维度差异，本文将注意力机制分解为两个低维机制，并利用两个分支来实现它们，这两个注意力分支赋予形状特征所包围的区域更高的权重，背景区域更低的权重，以创建形状先验，为目标检测器提供先验信息。其中，每个注意力分支有三个模块组成：一个编码模块、一个类BEV注意力模块和一个垂直注意力模块，形状注意机制的输出特征图是这三个模块输出的串联。

编码模块用于保持原始检测器输入的特征，该特征对应于特征图的每个锚点。编码模块由一个1×1卷积层组成，输出特征图的大小为，其中是batch size，和分别是特征图的高和宽，代表通道。

类BEV注意力模块模仿两阶段检测器在3D目标检测中的BEV表示，为后续处理提供形状先验。根据规则形状对象在BEV中投影时占据的近似面积，基于当前锚点创建一个先验区域来包含对象，当该区域是对象的一部分时，该区域被赋予更大的权重。其由两个卷积层组成，即用于特征学习的类BEV编码层和用于提供权重的类BEV注意力层。若定义每个先验区域，其中,是的高度和宽度，类BEV编码特征图和类BEV注意力图的大小为，其中是batch size，，是特征图的高和宽，是通道数。定义是FPN的输出特征，表示类BEV编码层，表示类BEV注意力层，则每个anchor的类BEV注意力输出表示为：

式中，, 表示sigmoid层，操作是将特征图重塑为 ，操作指在特征通道中重复次,指逐元素乘法，的大小为。

垂直注意力模块用于关注对象区域，对检测范围内的物体由于视角问题产生的垂直误差进行修正。其包括两个卷积层，即垂直编码层和垂直注意力层，两个卷积层使用与类BEV注意力模块中相同的预定义区域，且具有相同的大小，定义是FPN的输出特征，和分别代表垂直编码层和垂直注意力层。对于每个anchor的垂直注意力模块可以表示为：

式中，,表示sigmoid层，操作与类BEV注意力模块相同,指逐元素乘法，的大小为。

RPN网络根据所得形状先验分割出各种特征补丁，并创建锚点和锚框的候选集合并结合选择性搜索学习对应形状特征的权重。由于在训练过程中面临前景和背景类之间极度不平衡的问题，对于分类任务应用了focal loss：

其中为focal loss，是标签的类的模型估计概率。和是焦点损失的参数。实验设置=0.25，=2。

对于回归任务，利用SmoothL1计算回归损失。假设一个回归框参数为，其中和是中心位置，和表示回归框的长度和宽度：

式中是锚点底部的对角线。

总损失函数为：

式中是正锚点（positive anchors）的数量，是负锚点（negative anchors）的数量， 是两个类的对数损失（对象与非对象），对应边界框回归器预测的对应类别的回归参数，是Anchor Location的数量，是小批量大小，对应真实目标的边界框回归参数。

**4.5 Fine Detection**

在本节中，我们将介绍了所提出的网络的精检测部分。精检测会将粗检测的锚点和锚框的候选集合映射到对应原图区域和特征图区域，获取感兴趣的特征图区域，并通过非极大值抑制算法NMS来搜索局部极大值，实现对冗余的锚框进行去冗余：

式中：为检测框的置信度因子；为检测框的置信度分数；为抑制框的置信度分数。

最后将映射后的感兴趣的特征图区域输入分类网络，获得目标检测结果分类网络。分类网络由softmax分类层和边框回归层组成，每个anchors回归框部分生成n个softmax输出值，softmax输出值根据构成的类别分布作为最后的预测结果，并定位候选区域的位置。

1. Experiment

在本节中，我们介绍我们的评估结果。

**5.1** **Loss Function**

对于所提出的框架，为了使其能够完成端到端方式的训练，本文以多任务损失函数完成对其的优化。在传统的分类损失和回归损失的基础上，引入了尺度损失和纵横比损失2种附加损失。它们共同优化模型的损失公式为：

式中：和为注意力引导模块的模型参数，和均采用smooth L1损失，分类损失和回归损失采用传统的损失函数。

**5.2** **Platform Environment**

处理器为Intel(R)Xeon(R) Silver4214R [CPU@2.40GHz](mailto:CPU@2.40GHz) 48cores，安装内存为64GB，操作系统为ubuntu20.04，显卡为NVIDIA GeForce RTX 3080，10GB，实验框架为Caffe。本文在经典的 MSCOCO2017数据集上进行实验

**5.3 Dataset**

本文在修改后的 MSCOCO2017数据集上进行实验。原MSCOCO总共包含91个类别，考虑到实际应用场景中目标吹扫构件多为矩形机箱、电机、圆形仪表盘、三角形各类按钮等规则形状图形，背景复杂，目标物较小，我们在原数据集的基础上，筛选出其中形状规则，背景较为复杂的20类，共40000张图片作为训练集，4000张图片作为测试集。

**5.4 Experiment setting**

本文选择广泛使用的ImageNet作为实验主干网络。对于尺度空间和纵横比空间分别设置为S=｛128，192，256，384，512｝和R=｛0.5，0.8，1，1.2，2｝。在NMS算法中，通过大量的试验性实验得到初始阈值选取0.6有着更好的效果。对于多任务损失函数设置注意力引导模型参数=0.5，=0.1，在训练中总共训练了200epoch，初始学习率设置为0.01。

**5.5 Evaluation metrics**

为获得更为全面有效的锚框选择准确率，本文以平均查全率（AR）作为评价候选区域的好坏。在不同的（0.5～0.9）进行平均查全率的比较。分别对应于每幅图像100、300和500个建议的平均查全率；分别对应于小型、中型以及大型目标选取100幅图像计算的平均查全率。在进行RPN网络和模型整体准确性性能的比较中，本文采用平均准确度（mAP）作为性能度量参数（下标的含义与AR相同）。

**5.6** **Experimental results and analysis**

首先对改进的RPN与原RPN以及RPN的一些变体进行AR的比较。“RPN-AT”为引入注意力引导模型的改进RPN网络，“RPN-9anchors”和“RPN-12anchors”分别对应于RPN固定的9种和固定的12种锚框，为了保持不变量，其中的NMS算法仍为传统贪心NMS算法，按照此方式对本文改进的RPN进行平均查全率的评估。最后，对提出的整体目标检测模型与现有一些先进方法进行比较，如CascadeR-CNN、MaskR-CNN和FPN。

如表1所示，本文方法对于候选区域的平均查全率性能明显要优于传统RPN和固定种类的RPN。从中也可以发现，本文方法对于小目标会有更好的效果，其原因是传统的方法由于固定的锚框尺度和纵横比的限制对于小目标不能做到很好的识别，而本文提出的方法对于不同目标，锚框的选择不再固定，根据定义的形状向量辅助对不同目标进行锚框选择，从而生成更符合目标的锚框种类。同时可以发现，本文方法虽然在运行速度上有所增加，但仍在可接受的范围内。

Table 1 Average Recall Rate of Candidate Regions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **方法** |  |  |  |  |  |  | **运行时间/s** |
| **RPN-9anchors** | 47.5 | 53.4 | 57.8 | 28.7 | 52.8 | 63.9 | 0.09 |
| **RPN-12anchors** | 50.2 | 56.6 | 58.3 | 33.9 | 58.2 | 67.5 | 0.09 |
| **RPN-AT** | **52.9** | **59.5** | **60.8** | **38.2** | **61.7** | **69.9** | **0.11** |

检测性能比较如图4和表2所示。在与上述3种先进算法的对比中可以发现，本文方法无论是在不同R\_IoU阈值下的准确度还是在平均准确度都有所提高。在小型、中型以及大型目标的测试中可以发现，本文方法对于小型和中型目标的检测查准率与先进算法相比有着很好的提升，对于大型目标也有着很高的准确率。

Table 2 Detection Performance Comparison

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **算法** |  |  |  |  |  |  |
| **FPN** | 36.2 | 59.1 | 39.0 | 18.2 | 39.0 | 48.2 |
| **Mask R-CNN** | 39.8 | 62.3 | 43.4 | 22.1 | 43.2 | 51.2 |
| **Cascade R-CNN** | 42.8 | 62.1 | 46.3 | 23.7 | 45.7 | 55.2 |
| **AT＋ Faster R-CNN** | **43.3** | **63.0** | **47.8** | **24.3** | **46.4** | **55.3** |



Fig.4 Sample of test results

1. Conclusion

本文提出了一种添加了形状注意力机制的目标检测方法，通过注意力模块来预测目标物的尺度和纵横比，引导锚框的选择。与传统方法相比，本文所提出的方法在平均查全率和精度上有较好的提升，并且并没有过多的提高模型的复杂度，保证了实时性，为解决现有目标检测算法在复杂场景中精度低、计算量大、鲁棒性不足的问题提供可能的思路。在未来，我们希望改进方法使 RPN在寻找感兴趣区域时不过分依赖于参数设定，并创建更符合实际应用需求的轨道吹扫场景数据集来优化模型。

**Reference**：

[1] 刘海娇, 薛强, 卢晓东,等. 轨道车辆吹扫工艺现状分析[J]. 铁路节能环保与安全卫生, 2019, 9(6):4. (Liu Haijiao, Xue Qiang, Lu Xiaodong, et al. Analysis of the status quo of rail vehicle purging technology [J]. Railway Energy Saving, Environmental Protection, Safety and Sanitation, 2019, 9(6):4.)

[2] 陆志宏, 陆渭歧, 高绪昊,等. 车辆基地新型列车吹扫除尘技术研究[J]. 哈尔滨铁道科技, 2022(2):5(Lu Zhihong, Lu Weiqi, Gao Xuhao, et al. Research on New Type Train Purging and Dust Removal Technology in Vehicle Base[J]. Harbin Railway Science and Technology, 2022(2):5)

[3] Wang G, Yu Y, Liu H, et al. Progress on design and development of polymer electrolyte membrane fuel cell systems for vehicle applications: A review[J]. Fuel Processing Technology, 2018, 179: 203-228.

[4] Lee J, Hwang K. YOLO with adaptive frame control for real-time object detection applications[J]. Multimedia Tools and Applications, 2022, 81(25): 36375-36396.

[5] Galvez R L, Bandala A A, Dadios E P, et al. Object detection using convolutional neural networks[C]//TENCON 2018-2018 IEEE Region 10 Conference. IEEE, 2018: 2023-2027.

[6] Buckchash H, Raman B. A robust object detector: application to detection of visual knives[C]//2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). IEEE, 2017: 633-638.

[7] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Ieee, 2005, 1: 886-893.

[8] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91-110.

[9] Freund Y, Schapire R E. Experiments with a new boosting algorithm[C]//icml. 1996, 96: 148-156.

[10] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Ieee, 2005, 1: 886-893.

[11] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

[12] 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.

[13] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.

[14] Tian Z, Shen C, Chen H, et al. Fcos: Fully convolutional one-stage object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 9627-9636.

[15] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[16] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[17] 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.

[18] Cai Z, Fan Q, Feris R S, et al. A unified multi-scale deep convolutional neural network for fast object detection[C]//European conference on computer vision. Springer, Cham, 2016: 354-370.

[19] Sermanet P, Eigen D, Zhang X, et al. Overfeat: Integrated recognition, localization and detection using convolutional networks[J]. arXiv preprint arXiv:1312.6229, 2013.

[20] 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. 2017: 2117-2125.

[21] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.

[22] Zhou P, Ni B, Geng C, et al. Scale-transferrable object detection[C]//proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 528-537.

[23] Zhao Q, Sheng T, Wang Y, et al. M2det: A single-shot object detector based on multi-level feature pyramid network[C]//Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 9259-9266.

[24] Ghiasi G, Lin T Y, Le Q V. Nas-fpn: Learning scalable feature pyramid architecture for object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 7036-7045.