

DOI: 10.13382/j.jemi.B2306668

面向多尺度多方向螺栓的检测算法*

唐心亮 刘子剑 于平平

(河北科技大学信息科学与工程学院 石家庄 050018)

摘要:在工业建设中,螺栓零件是关键的连接件之一,通常用于连接大型机器设备和构件,如钢结构、桥梁、高速公路、建筑、石油管道等。其安装状况的优劣将直接关系到整个设备或结构的稳定性和可靠性。然而,螺栓的安装位置常常处于狭窄、复杂的环境中,利用人工检测不仅难度大,效率低,而且容易出现误判和漏检的情况。为此本文以 Faster R-CNN 为基础开展螺栓零件的识别研究。针对螺栓零件检测的难点,提出了一种基于多尺度多方向螺栓的检测算法。首先对采集到的图像进行扩增,以提高数据集的多样性;其次,通过改变主干网络增强模型对特征信息的敏感程度,再利用多尺度融合模块加强模型对小目标的检测;在预测框生成阶段,提出自适应旋转区域建议网络,以获取最优预测框;最后,针对多方向检测中出现的边界不连续的问题,通过 Gaussian Wasserstein 距离和焦点损失作为损失函数来代替传统的 Smooth L1 损失函数。螺栓零件的识别实验结果表明,改进后的 Faster R-CNN 模型 mAP 值能达到 87.4%,相比于原始 Faster R-CNN 模型 mAP 值提升了 7.6%。通过消融实验可以得出,改进后的 ResNet50 网络相较于原始 ResNet50 网络的 AP 值提升了 0.2%。与其他旋转检测模型在相同数据集上进行比对得出,本文提出的模型 AP 值更高,鲁棒性更好。本文所提出的模型可以解决螺栓零件在识别任务中因拍摄角度和复杂环境出现的问题,缓解了因图像尺度和旋转边界不连续带来的问题。

关键词: 多方向检测;螺栓零件检测;卷积网络改进;模型改进

中图分类号: TP391.41 **文献标识码:** A **国家标准学科分类代码:** 520.2

Detection algorithm for multi-scale and multi-directional bolts

Tang Xinliang Liu Zijian Yu Pingping

(Hebei University of Science and Technology, Shijiazhuang 050018, China)

Abstract: In industrial construction, bolt components are key connectors commonly used to join large machinery and components such as steel structures, bridges, highways, buildings, and oil pipelines. The quality of their installation directly affects the stability and reliability of the entire equipment or structure. However, the installation of bolts often takes place in narrow and complex environments, making manual inspection difficult, inefficient, and prone to misjudgment and omissions. In this study, we conducted bolt component recognition research based on the Faster R-CNN framework, aiming to address the challenges in bolt detection, we propose a detection algorithm based on multi-scale and multi-directional bolts. Firstly, we augment the collected images to enhance the diversity of the dataset. Secondly, we enhance the sensitivity of the model to feature information by modifying the backbone network, and utilize a multi-scale fusion module to improve the detection of small targets. In the stage of generating bounding boxes, we introduce an adaptive rotation region proposal network to obtain optimal bounding boxes. Finally, we address the issue of discontinuous boundaries in multi-directional detection by employing the Gaussian Wasserstein distance and focal loss as the loss functions instead of the traditional Smooth L1 loss. Experimental results for bolt component recognition demonstrate that the improved Faster R-CNN model achieves a mAP (mean average precision) value of 87.4%, which is a 7.6% improvement over the original Faster R-CNN model. Through ablation experiments, it is observed that the improved ResNet50 network achieves a 0.2% increase in AP (average precision) compared to the original ResNet50 network. Comparisons with other rotation detection models on the same dataset reveal that the proposed model has

收稿日期: 2023-06-28 Received Date: 2023-06-28

* 基金项目: 河北省高等学校科学技术重点研究项目(ZD2020318)、河北省教育厅青年基金(QN2023185)项目资助

higher AP values and better robustness. The model presented in this study effectively addresses the challenges posed by the shooting angles and complex environments in bolt component recognition tasks, mitigating issues caused by image scale and discontinuous rotation boundaries.

Keywords: multi-directional detection; bolt component detection; convolutional network improvement; model improvement

0 引言

螺栓零件是工业建设过程中一些工程的重要组成部分,螺栓零件虽然制造工艺并不复杂,但是其安装环境多为隐蔽狭小的空间,针对螺栓位置的检测和识别较为困难。传统的检测方法主要有手工检测技术和图像检测技术,但是这些检测技术普遍存在几个问题:检测速度慢且检测周期较长,导致检测效率低下;检测难度高,对人力无法抵达的区域无法实施有效检测;检测过程繁琐,难以实现自动化识别。在铁路建设的接触网支柱基础螺栓质量检测中,手工检测速度慢,且检测精确度低。而实现接触网支柱基础螺栓的自动化质量检测首先就需要完成对螺栓的检测和识别。此外,螺栓检测还可作为施工现场重要连接件位置检测和形态检测智能化的基础。因此如何对螺栓零件进行自动化检测识别成为当前工程施工过程中值得深入研究的问题。

随着模式识别和计算机视觉技术的成熟应用,人工智能检测被广泛应用医疗、教育、金融、制造等行业^[1],为工业生产和人们的生活带来了快捷和便利^[2]。自 2012 年 AlexNet^[3] 在 ImageNet 图像分类竞赛中获得巨大成功以来,CNN 在计算机视觉领域得到了广泛的应用。近几年的目标检测算法层出不穷,基于 Faster R-CNN^[4] 的改进也多种多样,双阶段的目标检测算法基本都围绕 Faster R-CNN 展开。Wan 等^[5] 通过对特征提取网络的卷积层和池化层进行改进,构建了基于 Faster R-CNN 的两阶段多类水果目标检测算法,实现了多种类目标的检测,不过对于一些特定检测目标的检测精度提升十分有限。Cao 等^[6] 通过改进 Faster R-CNN 的交并比的损失函数进行包围盒回归,同时使用多尺度卷积特征融合模块,解决了传统损失函数不能准确反应预测区域建议网络与基准真相重叠的问题。Guo 等^[7] 利用注意力机制并设计中心点矩形损失(CR 损失)作为损失函数的方法对 Faster R-CNN 进行改进,完成了接触网悬挂系统滴管的故障检测,实现了对小目标的检测,但是对重叠率较高的小目标检测有一定的局限性。Liu^[8] 通过硬负样本挖掘和交替训练对 Faster R-CNN 进行改进,在检测精度方面有了很大的提升,在检测速度上仍有很大的提升空间。这些方法统称为水平区域检测方法。对于螺栓零件检测来说,水平检测方法并不能够反映各种方向的螺栓零件的真实形状,且容易受到旋转遮挡等因素的影响,导致检测结果与实

际情况偏差较大。其次对于长宽跨度比较大的螺栓来说,无法做到螺栓零件和背景的有效分离。因此,针对多方向螺栓零件的检测需要更加精细化的算法和方法。

目前大多数旋转检测算法主要应用于遥感图像,对于螺栓零件的多方向检测的研究并不是很多,但是随着人工智能技术的发展,旋转检测算法的精度和速度都得到了很大提升,从而使得螺栓零件的多方向检测成为可能。文献[9]在 Faster R-CNN 的 RPN 中引入旋转因子并改进了旋转感兴趣区域池化层,提出了用于多方向文本检测的旋转区域提取网络(region proposal network, RPN)。文献[10]通过将文本拆解再利用端对端训练的全卷积神经网络密集检测拆解的部分,实现了多方向长段非拉丁文本的检测。文献[11]对基线模型进行优化,完成了对航空影像的多方向检测。这些检测虽然都实现了目标图像的多方向检测,但是它们普遍存在损失不连续的问题,对于螺栓检测来说,还存在图像尺度不均衡的问题。而且目前的检测算法多数都是在常规数据集上进行实验的,由于螺栓零件图像具有一定的特殊性(尺度多样性,视角特殊性,位姿不定性等),基于常规数据集训练的模型在螺栓零件图像上的检测效果并不好。

本文结合螺栓零件自身的特点,提出了一种基于改进 Faster R-CNN 的螺栓零件的多方向检测方法。

1 图像采集与标注

1.1 螺栓零件图像采集和处理

螺栓零件图像的采集工作主要通过人工勘测采集来完成。为了能够更好的采集施工现场的螺栓图像,本文设计了接触网支柱基础螺栓图像采集设备,采集过程如图 1 所示。



图 1 接触网支柱基础螺栓图像采集设备

Fig. 1 Image acquisition equipment for contact wire support pillar foundation bolts

如图 2 所示,本文提出的多方向检测方法相较于传统的水平检测方法,能够最大程度减少背景信息对被检测目标图像的影响,这就要求数据集具有多样性。

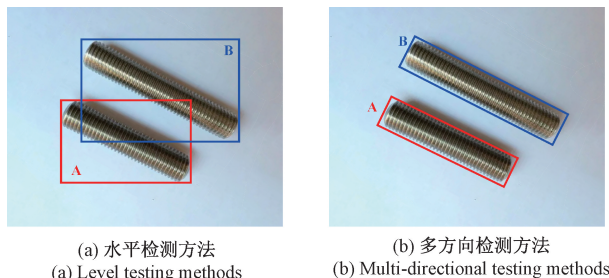


图 2 检测方法对比

Fig. 2 Comparison of detection methods

为了使检测模型更具实用性和普遍性,螺栓零件的采集地点主要集中于施工现场和实验室模拟环境,同时为了扩充螺栓零件数据集并兼顾多种环境下的检测,部分螺栓零件图片来源于网络。总共采集了 1 276 张不同环境下的螺栓图像,对数据集标注后按照进行随机划分,如图 3 所示。

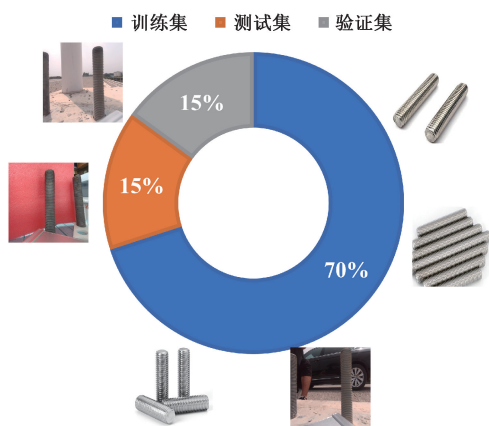


图 3 数据集划分

Fig. 3 Dataset partitioning

1.2 螺栓零件图像分析

除螺栓零件的长螺纹特征外,螺栓零件还会因为采集的区域地点和螺栓安装方式的不同造成采集的图像出现倾斜交叉重叠等情况,最终导致获得螺栓零件在整幅图像中的位置并不总是垂直的。因此,在进行图像标注时,如果用常规矩形框进行标注会造成特征信息混乱和特征信息丢失的问题。其标注文件内容如图 4(a) 所示。

本文在图像标注上采用了倾斜框进行标注,每张图像的标注信息都会储存在单独的文件中,相比于常规标注,使用本文的标注方法标注的位置信息更加贴合螺栓零件且标注文件内容更加丰富,能够在一定程度上减少特征信息的丢失的问题。螺栓零件标注情况如图 4(b) 所示。

```
<?xml version="1.0"?>
<annotation>
  <folder>Dateimg</folder>
  <filename>00 (1).jpg</filename>
  <path>E:\Dateimg\00 (1).jpg</path>
  <source>
    <database>Unknown</database>
  </source>
  <size>
    <width>1080</width>
    <height>1078</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
  <object>
    <name>bolt</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>332</xmin>
      <ymin>137</ymin>
      <xmax>824</xmax>
      <ymax>916</ymax>
    </bndbox>
  </object>
  <object>
    <name>bolt</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>3</xmin>
      <ymin>305</ymin>
      <xmax>1070</xmax>
      <ymax>698</ymax>
    </bndbox>
  </object>
</annotation>
```

(a) 常规标注文件内容
(a) Conventional annotation
file content

```
<?xml version="1.0"?>
<annotation verified="no">
  <folder>Dateimg</folder>
  <filename>00 (1)</filename>
  <path>E:\Dateimg\00 (1).jpg</path>
  <source>
    <database>Unknown</database>
  </source>
  <size>
    <width>1080</width>
    <height>1078</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
  <object>
    <type>robndbox</type>
    <name>bolt</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <robndbox>
      <cx>574.3887</cx>
      <cy>522.0037</cy>
      <w>129.135</w>
      <h>875.9841</h>
      <angle>2.641593</angle>
    </robndbox>
  </object>
  <object>
    <type>robndbox</type>
    <name>bolt</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <robndbox>
      <cx>540.538</cx>
      <cy>499.068</cy>
      <w>1085.1928</w>
      <h>116.9027</h>
      <angle>2.671593</angle>
    </robndbox>
  </object>
</annotation>
```

(b) 本文标注文件内容
(b) Annotation file content for
this article

图 4 标注文件内容

Fig. 4 Annotation file content

2 面向多方向的自适应螺栓检测

2.1 原始模型及应用分析

Faster R-CNN 是在 Fast RCNN 模型基础上进行优化而来的模型。其主要特点是将 Fast RCNN 与 RPN 模型相结合,进一步提高模型的检测速度和准确度。本文主要针对 Faster R-CNN 中的主干网络、区域建议网络以及损失函数进行改进。

2.2 改进的特征提取网络

1) 原始主干网络

Faster R-CNN 模型是以 VGG16 网络作为主干特征提取网络的。但是在 VGG16 网络中存在网络的层数越深,越容易出现梯度弥散或者梯度爆炸的问题^[12]。当网络深度达到饱和后,模型会降低对螺栓零件图像的识别精准性。VGG16 算法模型如图 5 所示。

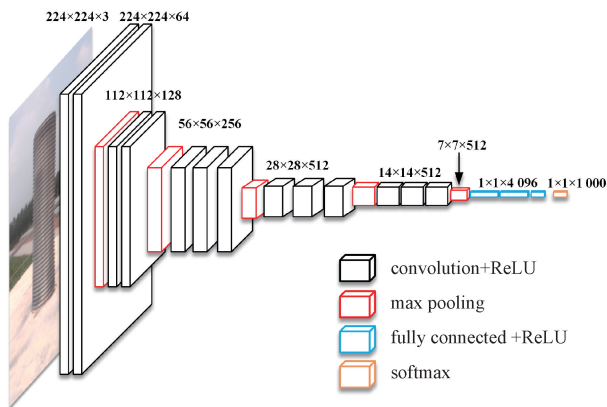


图 5 VGG16 结构

Fig. 5 Structure diagram of VGG16

2) 改进的网络

网络模型结构的调整一般不会增加模型计算的时间成本,但是可以大幅度提升模型的整体性能^[13],因此,本文采用改进的 ResNet50 网络作为训练模型的主干网络。ResNet50 网络中的深度残差网络^[14]可以有效解决 VGG16 网络中因网络层数加深而出现的梯度弥散现象。

ResNet50 网络如图 6 所示。

为了提高 ResNet50 网络在螺栓零件检测中的性能,本文依照卷积层中步长大小与获得输入信息多少的关系,针对网络各个阶段中卷积层的步长做了调整。首先,将下采样块第 1 个卷积层的步长与第 2 个卷积层的步长进行调换,以增强下采样块中第 1 输入信息的完整性。

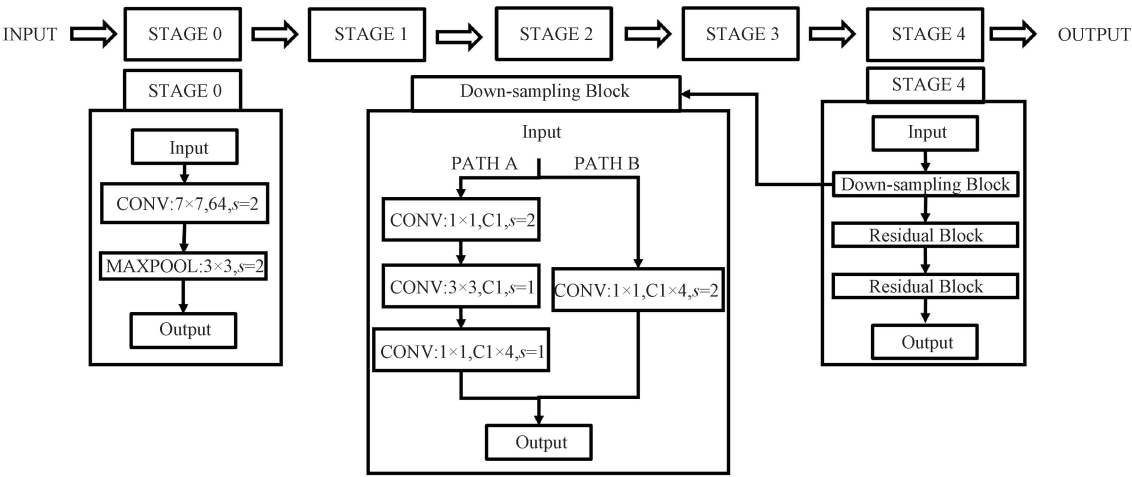


图 6 ResNet50 原始网络
Fig. 6 Network of ResNet50

其次,在路径 B 中增加了一个 2x2 的平均池化层,平均池化层不仅可以减少路径 B 中参数的维度,更有利于下一卷积层的特征信息提取,其结构如图 7 所示。

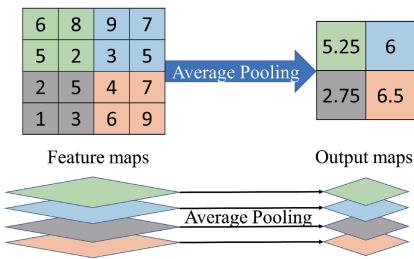


图 7 平均池化
Fig. 7 Average pooling

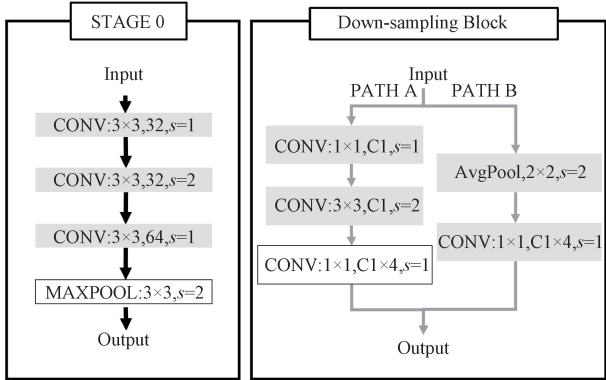


图 8 改进后的 ResNet50 网络
Fig. 8 The improved network of ResNet50

然后调整路径 A 和路径 B 的输出维度,将两个路径的输出相加作为下采样块的整体输出。最后为了简化卷积过程中计算的复杂度,将阶段 0 中卷积核为 7x7 的卷积层简化成 3 个卷积核为 3x3 的卷积层,卷积总步长不变。模型调整如图 8 所示。

2.3 多尺度融合分析

图像尺度的不平衡性如图 9 所示。在特征提取过程中,网络从不同尺度图像中提取出的细节信息非常有限,容易缺失特征信息,不利于螺栓零件图像的识别分类。

由于螺栓零件目标特征的尺度具有多样性,而原始模型在使用单尺度特征进行多尺度识别时,容易出现漏检和错检的情况^[25]。针对上述情况,本文采用特征金字塔 (feature pyramid networks, FPN)^[15]以增强尺度特征的鲁棒性,提升模型对多尺度目标的检测精度。

FPN 可以根据特征层和低特征层的特点对特征层信息进行调整和融合,从而提高中小型物体的检测精度。FPN 结构如图 10 所示。

FPN 将自下而上路径中最上层的螺栓零件特征图进行卷积后,利用卷积后得到的特征图作为自上而下路径

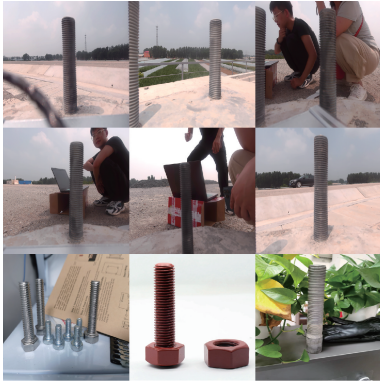


图 9 图像尺度的不均衡性

Fig. 9 The unevenness of image scale

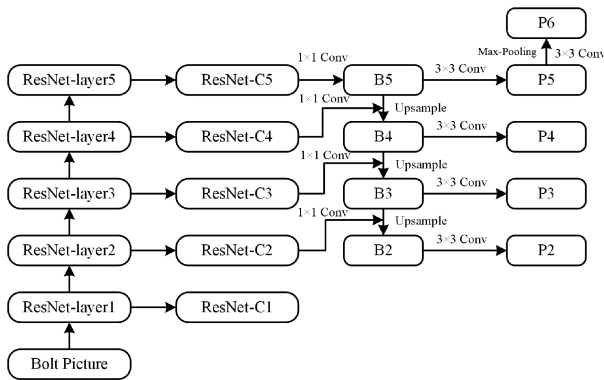


图 10 特征金字塔网络结构

Fig. 10 Structure diagram of feature pyramid network

的起点,自下而上的结构能够与改进后的特征提取网络 ResNest50 的前向传播过程相结合,将改进后的 ResNest50 在每个阶段最后一层的特征作为基层特征图 $\{C2, C3, C4, C5\}$ 。在自顶向下的过程中,网络对高层特征进行上采样来得到与低层特征相同空间分辨率的特征,然后使用横向连接将上采样得到的特征与基层特征 $\{C2, C3, C4, C5\}$ 相融合,最终生成 FPN 特征图 $\{P2, P3, P4, P5\}$, $\{P2, P3, P4, P5\}$ 与 $\{C2, C3, C4, C5\}$ 关系如式(1)所示。

$$P_i = \text{Conv}_{3 \times 3} \left[\sum_{j=i+1}^5 \text{upsample}(P_j) \oplus \text{Conv}_{1 \times 1}(C_i) \right] \quad (1)$$

同时对 $P5$ 进行最大池化得到特征 $P6$ 。 $P5$ 和 $P6$ 具体关系如式(2)所示。

$$P_6 = \text{maxpooling}(P_5) \quad (2)$$

由于多层特征图像分辨率不同,在执行多尺度特征映射时,针对不同尺度螺栓图像的区域预测框,选择最合适尺寸的特征图层来提取特征块。特征金字塔的级别 P_k 通过计算来确定,其计算公式如式(3)所示。

$$k = k_0 + \log_2 \left(\frac{\sqrt{wh}}{640} \right) \quad (3)$$

其中, k_0 是基准值, w 是生成预测框的宽, h 是生成预测框的长, 640 是输入图像的大小。

利用特征金字塔进行对比分析每层的特征信息然后综合得出包含丰富特征信息的特征图,进而逐步提高网络对不同尺度螺栓零件检测的能力。

2.4 自适应旋转区域建议网络

在采集到的螺栓零件图像中,螺栓零件的分布并不总是垂直均匀的,并且图像中有许多与检测目标无关的区域,例如树木、道路等。在非目标区域设置锚框不仅会增加计算量^[27],同时会造成目标信息混乱,从而干扰对目标的检测。为了在精准生成锚框的同时实现对螺栓的多方向检测,本文提出了自适应旋转区域建议网络(adaptive-oriented rotated region proposal network, AORRPN)。AORRPN 引入了角度感知机制,使得该算法能够检测具有不同方向和角度的目标。此外, AORRPN 还可以通过放缩卷积(scale-invariant convolution)对螺栓零件的检测框进行适当调整。

针对多方向的边界框,本文使用 5 个参数对边界框进行重新定义,即 (c_x, c_y, w, h, θ) , $\theta \in [-\pi/2, 0]$ 。其中, (c_x, c_y) 表示边界框的中心位置的坐标, (w, h) 为边界框的宽和高, θ 为 x 轴的正方向与其最近边界框的边之间的角度值^[9]。为了模型能够更好的契合螺栓图像,本文根据螺栓零件在图像中的特点设置了特定的宽高比 $[1:3, 1:5, 1:7, 1:9]$, 特定的尺寸值 $[4, 8, 16]$ 和特定的角度值 $[-10^\circ, -25^\circ, -40^\circ, -55^\circ, -70^\circ, -90^\circ]$, 如图 11 所示。特征图会根据设定的值在每个点生成 72 个旋转锚框(R-anchor)。对应的回归层则会生成 360 个输出值,分类层生成 144 个输出值,对于大小为 $H \times W$ 的特征图,可以产生 $H \times W \times 72$ 个旋转锚框^[28]。

为了防止旋转后的锚框继续沿用旋转前锚框的形状,本文在 AORRPN 中,提出了特征自适应模块。在 Faster R-CNN 的 RPN 中,由于最终形成的检测框具有角度属性,如果将其按照传统的方式直接在特征图上进行全卷积操作,可能会由于锚框大小与实际目标不匹配导致检测效果不佳。理想情况下的锚框大小应根据目标尺寸的大小调整到最佳位置,特征自适应模块参考了可变卷积^[16]的思想,提出了一种新的适合于多方向检测的放缩卷积。

放缩卷积可以将生成的特征图先通过一个 3×3 的卷积层,然后进行权重补偿和角度赋予,使最终获得的检测框更加贴合旋转目标。根据锚框特征生成新的特征图,可由式(4)表示。

$$y(F_0) = \text{Conv}_{3 \times 3} [x(F_n)] \cdot x(F_0 + F_q + F_\theta) \quad (4)$$

其中, $y(F_0)$ 为放缩卷积后的特征, $\text{Conv}_{3 \times 3}$ 为 3×3 的卷积层, $x(F_n)$ 为特征图中第 n 个位置的特征, F_0 为初始位置坐标, F_q 权重补偿值, F_θ 为赋予该位置对应锚框

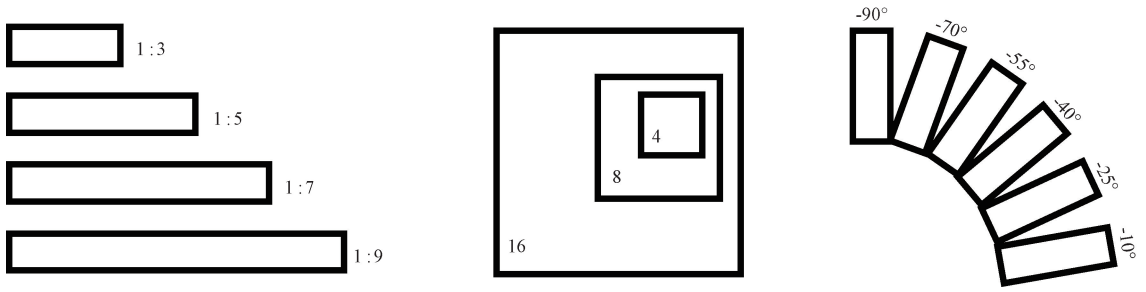


图 11 锚框的设置

Fig. 11 Angle value of RRPN

形状的角度值。

AORRPN 利用上式对原始特征图进行放缩卷积,以获取输出特征图,然后对其进行分类和边界框回归。放缩卷积的示意图如图 12 所示。

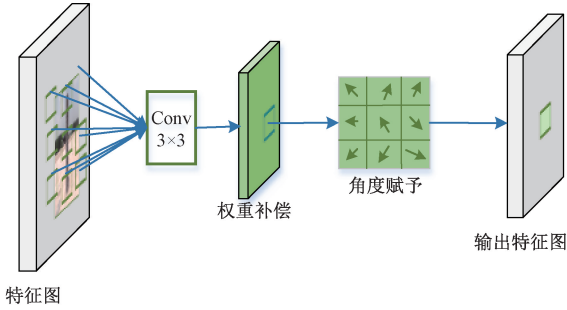


图 12 放缩卷积

Fig. 12 Scale-invariant convolution

在进行目标检测时通常要计算检测框的交并比(intersection-over-union, IoU),其主要作用是将正负样本进行分类以及后续非极大抑制过滤,传统的 IoU 是基于矩形框进行计算的。IoU 的计算公式如式(5)所示。

$$IoU = \frac{\text{area}(A \cap B)}{\text{area}(A \cup B)} \quad (5)$$

但是在计算多方向目标检测 IoU 时,由于旋转检测框具有角度信息,而传统的 IoU 只考虑了矩形框之间在水平和垂直方向上的重叠情况,并没有考虑旋转角度对结果产生的影响。在存在较大旋转角度时,使用传统 IoU 可能会导致误差增加。因此,本文提出了一种基于多方向目标检测的倾斜 IoU 计算方法。

倾斜 IoU 计算方法如图 13 所示,给定两个部分重叠的旋转检测框 S_{ABCD} 和 S_{EFGH} ,定义重叠部分的多边形为 φ ,以为图 13(b)为例,多边形区域 φ 的顶点为 $ILKJ$,以 I 为计算顶点连接其他各个顶点将多边形区域剖分成多个三角形,即 $\triangle_{ILK}$ 和 $\triangle_{IKJ}$,然后计算出多边形 φ 的面积,计算公式如式(6)所示。

$$\text{Area}(\varphi) = \text{Area}(\triangle_{ILK} + \triangle_{IKJ}) \quad (6)$$

转换为通用公式为:

$$\text{Area}(\varphi) = \sum_{i=1}^n \text{Area}(\triangle_i) \quad (7)$$

其中, $\text{Area}(\varphi)$ 为多边形的面积, n 为三角形个数, i 为三角形的索引, $\text{Area}(\triangle_i)$ 为第 i 个三角形的面积。联立式(6)、(7)可得倾斜 IoU 的计算公式如式(8)所示。

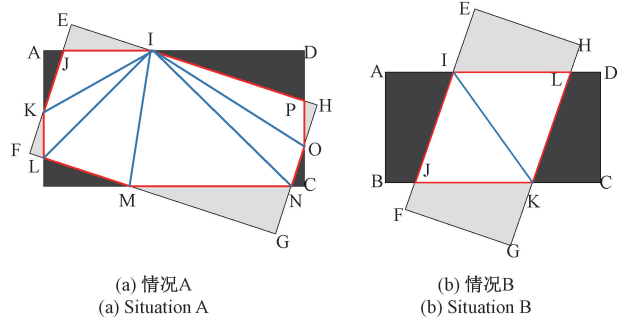


图 13 三角剖分法

Fig. 13 Triangulation

$$IoU(S_{ABCD} + S_{EFGH}) = \frac{\text{Area}(\varphi)}{\text{Area}(S_{ABCD}) + \text{Area}(S_{EFGH}) - \text{Area}(\varphi)} \quad (8)$$

2.5 损失函数设计

本文中的旋转矩形由 5 个参数 (c_x, c_y, w, h, θ) 表示,回归方程如下:

$$\begin{aligned} p_x &= \frac{(c_x - c_{x_a})}{w_a}, p_y = \frac{(c_y - c_{y_a})}{h_a} \\ p_w &= \log\left(\frac{w}{w_a}\right), p_h = \log\left(\frac{h}{h_a}\right), p_\theta = \theta - \theta_a \\ p_x^* &= \frac{(c_x^* - c_{x_a})}{w_a}, p_y^* = \frac{(c_y^* - c_{y_a})}{w_a} \\ p_w^* &= \log\left(\frac{w^*}{w_a}\right), p_h^* = \log\left(\frac{h^*}{h_a}\right), p_\theta^* = \theta^* - \theta_a \end{aligned} \quad (9)$$

其中, c_x, c_y, w, h, θ 分别表示框的中心坐标、宽度、高度和角度。变量 $\theta, \theta_a, \theta^*$ 分别用于真实框、锚框和预测框(同样用于 c_x, c_y, w, h)。传统的 Smooth L1 损失函数仅考虑到了水平检测的因素,用于多方向检测还存在一

定的问题。例如在进行多方向目标检测时会出现由边界框回归引起的损失不连续的情况,这种情况通常是由于在边界框进行回归时参数变化引起边界处损失急剧增加导致的。在进行边界回归时,需要将锚框进行旋转以获得与真实框重合的预测框,如图 14 所示。

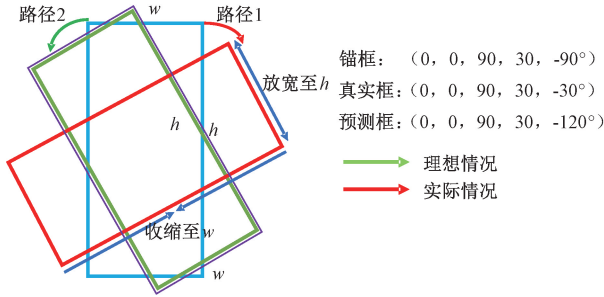


图 14 损失不连续问题

Fig. 14 Loss discontinuity problem

理想情况下,预测框会由锚框的位置转化到真实框的位置,路径 2 通过将预测框逆时针旋转 30° 来达到转换的目的,但是由于角度的周期性的^[13]问题,路径 2 中转换后的真实框角度超出了定义的范围,此时会出现相当大的 Smooth L1 损失,为了避免出现损失波动,模型必须以前述方式进行回归预测,如路径 1 所示,模型通过顺时针旋转并放缩长度和宽度来实现将预测框旋转到真实框的位置。此时的损耗值虽然接近于 0,但是模型会经历复杂的回归。这就要求较高的鲁棒性,增加了训练难度^[26]。

相较于 Smooth L1 损失函数而言, Gaussian Wasserstein 距离^[29]能够兼顾模型的预测分布和目标分布,通过计算这两个分布之间的 Gaussian Wasserstein 距离,可以衡量他们的差异程度。Gaussian Wasserstein 距离将这两个概率分布映射到高维空间,并计算它们之间的最小传输成本,从而考虑了整体结构和形状的差异。通过最小化 Gaussian Wasserstein 距离,我们可以鼓励模型生成更平滑的预测分布,特别是在边界处,从而缓解边界不连续问题。

本文在使用 Gaussian Wasserstein 距离的基础上加入了焦点损失函数^[24],进一步加强了模型的学习能力和对关键样本的关注程度。改进后的模型在处理更多样化的回归形式时能够更加准确地进行学习和预测。具体如式 (10) 所示

$$L = \frac{\lambda_1}{N} \sum_{n=1}^N obj_n \cdot L_{gwd}(b_n, gt_n) + \frac{\lambda_2}{N} \sum_{n=1}^N L_{cls}(p_n, t_n) \quad (10)$$

式中: λ_1, λ_2 控制权衡参数,默认设置为 $[2, 1]$ 。 N 表示锚点的数量, obj_n 表示前景和背景。 b_n 表示第 n 个预测框, gt_n 是第 n 个目标的真实框。 p_n 是通过 sigmoid 函数计算得到的各

类的第 n 个概率分布。 t_n 表示第 n 个对象的标签。

3 实验结果及分析

3.1 模型训练环境和参数设置

本文实验基于操作北京超算云计算中心平台, GPU 型号为 Tesla V100-SXM2, 显存大小为 32 GB。该模型是基于 Pytorch1.7.0 深度学习框架, 在 CUDA 10.1 和 CUDNN 7.6.3 下选择 GPU 模式完成模型的搭建以及网络训练, 训练参数如表 1 所示。

表 1 实验参数设置

Table 1 Experimental parameter setting

参数	数值
epochs	300
batch-size	16
learning rate	0.000 5
Mosaic	0.9
Weight-decay	0.000 1
Img size	640×640

3.2 评价指标

目标检测算法一般使用平均准确率 (mean average precision, mAP) 作为评价标准, 该指标综合考虑了准确率和召回率。准确率的计算如式 (11) 所示。

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

其中, TP 表示 IoU 大于某一阈值的预测框数量, FP 表示 IoU 小于某一阈值的预测框数量或冗余预测框数量。召回率的计算如式 (12) 所示。

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (12)$$

式中: FN 表示未被检测到的真实框的数量。

以准确率作为纵坐标, 召回率作为横坐标, 绘制准确率-召回率 (precision-recall, P-R) 曲线。曲线下的面积是某一分类的平均准确度 (average precision, AP), AP 计算方式如式 (13) 所示。

$$AP = \int_0^1 P(R) dR \quad (13)$$

式中: R 是召回率, P 是准确率。

mAP 是所有分类的平均 AP 值, 计算如式 (14) 所示。

$$mAP = \frac{1}{n} \sum_{k=1}^n AP(k) \quad (14)$$

3.3 消融实验分析

为了探究本文对 Faster R-CNN 网络所进行的各项改进的有效性, 本文使用消融实验对部分改进模块进行消融对比, 使用 AP 值作为目标检测模型检测精度的衡量指标。

如表 2 所示,改进后的 ResNet50 网络的 AP 值达到了 87.4,相比于原始 ResNet50 网络来说 AP 值提升了 0.2%。图 15 则展示了改进前后的 ResNet50 网络在训练中 PR 曲线的变化情况。这证实了改进后的 ResNet50 网络能够从特征图中获得更多的特征信息,3 个 3×3 的卷积层对特征信息的敏感程度要强于单个 7×7 的卷积层。说明改进后的 ResNet50 网络相对于原始 ResNet50 网络来说,更适合于多方向螺栓的检测。

表 2 改进前后 ResNet50 网络检测结果

Table 2 Detection results of different ResNet50 networks	
主干网络	AP/%
ResNet50	87.2
Improved ResNet50	87.4

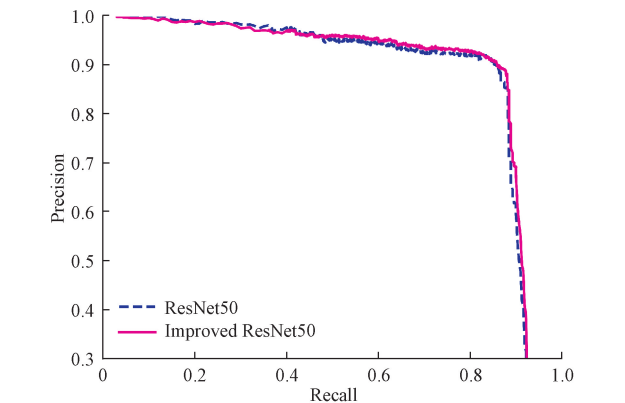


图 15 不同主干网络的准确召回曲线

Fig. 15 Precision-Recall curve of different backbone networks

如表 3 所示,本文所使用的损失函数相较于 Smooth L1 损失函数来说 AP 值提升了 3.2%。图 16 展示了两种损失函数在训练中损失值的变化情况。证实了本文所使用的损失函数可以显著提高边界框回归的准确率,具有更好的效果。这是因为本文所使用的损失函数可以通过将旋转边界框转换成二维高斯分布的方式来解决多方向目标检测中由边界框回归引起的损失不连续的问题。

表 3 不同损失函数检测结果

Table 3 Detection results of different loss functions	
损失函数	AP/%
Smooth L1 Loss	84.2
Ours Loss	87.4

3.4 提出的模型与其他模型对比

为了更好的验证改进的 Faster R-CNN 算法的检测性能,本文将改进的 Faster R-CNN 螺栓检测模型与现有的主流多方向目标检测模型 R²CNN^[17]、RRPN^[9]、R²PN^[18]、ROI-trans^[19]进行了比较,此外,为了更直观体现多方向检测与水平检测的区别,在对比的模型中加入

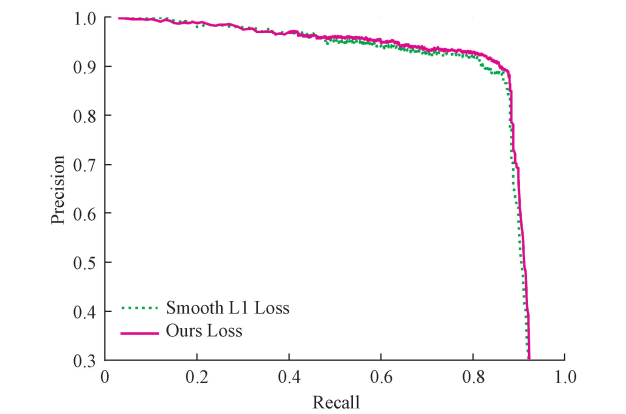


图 16 不同损失函数的准确召回曲线

Fig. 16 Precision-Recall curve of different loss functions

了水平检测模型 Faster R-CNN、SSD^[20]和 R-FCN^[21]。与水平检测方法比较,多方向检测算法的可以更好的处理螺栓零件目标检测中的方向多样化,尺度不均衡和密集分布的问题,从而提高检测效率。

从表 4 中可以看出本文方法训练的网络模型得到的 AP 值达到了 87.4%。本文提出的模型与水平检测的 Faster R-CNN 相比,AP 值提高了 7.6%。与其他多方向检测模型相比较(包括 R²CNN,RRPN,R²PN,ROI-trans,RSDE^[22],RTMDet^[23]等),本文提出的模型准确度更高,鲁棒性更强,具体表现在本文提出模型 AP 比 R²CNN 高 13.2%,比 RRPN 高 8.7%,比 R²PN 高 7.9%,比 RTMDet 高 2.9%,比 ROI-trans 高 0.8%,比 RSDE^[22]高 0.2%。本文提出的模型相对于其他多方向模型检测效果优异的原因是在加入自适应模块的同时引入了基于 Gaussian Wasserstein 距离的损失函数,解决了多方向检测普遍存在的损失不连续的问题。综上所述,本文基于改进的 Faster R-CNN 的算法在在多方向目标检测方面表现出卓越的效果。不仅提升了检测精度,同时也表现出了良好的准确性,这一结果充分证明了本文算法的优越性与可行性。

表 4 与其他模型的对比试验

Table 4 Comparison test with others model			
序号	模型	主干网络	AP/%
1	Faster R-CNN	VGG16	79.8
2	SSD	VGG16	82.3
3	R-FCN	ResNet50	86.3
4	R ² CNN	ResNet101	74.2
5	RRPN	ResNet101	78.7
6	R ² PN	VGG16	79.5
7	RTMDet	Improved CSPDarkNet	84.5
8	ROI-trans	ResNet101	86.3
9	RSDE ^[22]	ResNet50	87.2
10	Ours	Improved ResNet50	87.4

3.5 结果展示

为了更加直观的体现本文提出模型的检测效果,图 17 将部分有代表性的检测结果进行展示。

图 17(a)中,可以看到模型在处理多尺度图像和遮挡图像时表现出了卓越的检测能力。不论是多尺度的目标图像检测还是在螺栓零件遮挡的情况下,模型都能够准确地检测到目标并提供可靠的结果。

图 17(b)展示了模型在接触网支柱基础法兰识别应用中的检测结果。这个应用场景中存在着复杂的目标形态和背景干扰,然而模型仍然能够准确地识别出目标对象,显示出了出色的鲁棒性和适应能力。

图 17(c)呈现了当目标图像角度处于未定义角度区间时,模型面临损失不连续问题的检测结果。通过引入基于 Gaussian Wasserstein 距离的损失函数,本文的模型能够有效地解决这一问题,并展示出在这种情况下出色性能。这表明本文的模型具备处理复杂姿态和角度变化的能力。

图 17(d)显示了模型在处理大角度区间的目标图像时的检测结果。即使在目标图像存在较大旋转角度的情况下,模型仍然能够准确地定位和识别目标,显示出了对姿态变化的良好适应性。

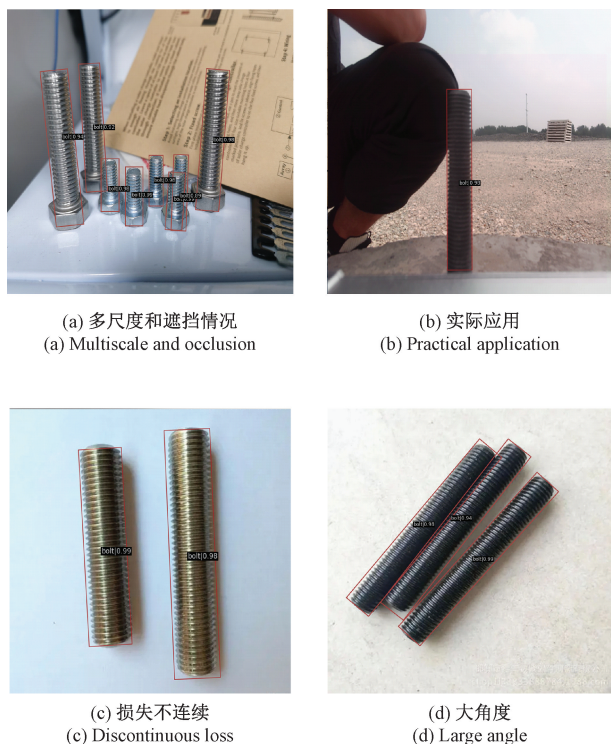


图 17 检测结果展示

Fig. 17 Result display

综上所述,通过图 17 中多个具体实例的展示,本文提出的模型在多个方面,包括多方向、多尺度、大角度和

密集分布的检测中都展现出了卓越的性能。

4 结 论

为了能够自动处理图像并检测图像中螺栓,减轻人工的压力,本文基于螺栓零件自身的特点和当前检测识别技术的发展提出了一种基于改进 Faster R-CNN 的多方向螺栓检测模型。首先利用改进后的 ResNet50 作为特征提取网络,同时引入特征金字塔,然后对区域建议网络进行改进,最后利用 Gaussian Wasserstein 距离和焦点损失的损失函数解决多方向检测中损失不连续的问题,最终得到本文改进后的 Faster R-CNN 模型。通过实验测试说明,本文提出的模型有效缓解了传统水平检测方法因安装位置偏僻,小目标,零件之间相互遮挡和不同位姿等干扰而检测识别鲁棒性差的问题,同时,对于其他多方向检测模型来说,本文提出的模型有效解决了因边界框回归导致的损失不连续的问题。综上所述,本文提出的模型在检测方面能达到更好的效果,检测精度更高。可以对工业建设中的螺栓零件进行智能检测识别,能够大大减轻人工检测的压力。

目前本文所提出的模型和算法已经在接触网支柱基础法兰识别得到了初步的应用。但是模型仍存在一些不足之处。模型在处理大规模数据时的计算效率有待提高。此外,模型的泛化能力还需要进一步验证和加强。后续将增加更多的样本和场景多样性,进一步验证模型的泛化能力,以满足实际生产环境中的需求。

参考文献

- [1] LIU P, LU L, ZHANG J, et al. Application of artificial intelligence in medicine: An overview [J]. Current Medical Science, 2021; 1105-1115.
- [2] CHEN L, CHEN P, LIN Z. Artificial intelligence in education: A review [J]. IEEE Access, 2020; 75264-75278.
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017; 84-90.
- [4] 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.
- [5] WAN S, GOUDOS S. Faster R-CNN for multi-class fruit detection using a robotic vision system [J]. Computer Networks, 2020; 107036.
- [6] CAO C, WANG B, ZHANG W, et al. An improved faster R-CNN for small object detection [J]. IEEE Access, 2019; 106838-106846.
- [7] GUO Q, LIU L, XU W, et al. An improved faster R-

- CNN for high-speed railway dropper detection[J]. IEEE Access, 2020; 105622-105633.
- [8] LIU Y. An improved Faster R-CNN for object detection[C]. 2018 11th International Symposium on Computational Intelligence and Design (ISCID), 2018; 119-123.
- [9] MA J, SHAO W, YE H, et al. Arbitrary-oriented scene text detection via rotation proposals [J]. IEEE Transactions on Multimedia, 2018; 3111-3122.
- [10] SHI B, BAI X, BELONGIE S. Detecting oriented text in natural images by linking segments[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017; 2550-2558.
- [11] WOLF S, SOMMER L, SCHUMANN A. Fastaer det: Fast aerial embedded real-time detection [J]. Remote Sensing, 2021; 3088.
- [12] XIANG W, SONG Z, ZHANG G, et al. Birds detection in natural scenes based on improved Faster R-CNN[J]. Applied Sciences, 2022; 6094.
- [13] 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, 2019; 558-567.
- [14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016; 770-778.
- [15] 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.
- [16] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017; 764-773.
- [17] JIANG Y, ZHU X, WANG X, et al. R²CNN: Rotational region CNN for arbitrarily-oriented scene text detection[C]. 2018 24th International Conference on Pattern Recognition (ICPR), 2018; 3610-3615.
- [18] ZHANG Z, GUO W, ZHU S, et al. Toward arbitrary-oriented ship detection with rotated region proposal and discrimination networks [J]. IEEE Geoscience and Remote Sensing Letters, 2018; 1745-1749.
- [19] DING J, XUE N, LONG Y, et al. Learning roi transformer for oriented object detection in aerial images [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019; 2849-2858.
- [20] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C]. Computer Vision-ECCV 2016; 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14, 2016; 21-37.
- [21] DAI J, LI Y, HE K, et al. R-fcn: Object detection via region-based fully convolutional networks[J]. Advances in Neural Information Processing Systems, 2016, 29.
- [22] QIAN W, YANG X, PENG S, et al. Learning modulated loss for rotated object detection[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021; 2458-2466.
- [23] LYU C, ZHANG W, HUANG H, et al. RTMDet: An empirical study of designing real-time object detectors[J]. arXiv 2022, arXiv preprint arXiv:2212.07784.
- [24] REN J. Research on financial investment decision based on artificial intelligence algorithm [J]. IEEE Sensors Journal, 2021; 25190-25197.
- [25] BAI T, YANG J, XU G, et al. An optimized railway fastener detection method based on modified Faster R-CNN[J]. Measurement, 2021; 109742.
- [26] ZHAO L, LI C J, WU X D, et al. Improved damage characteristics identification method of concrete CT images based on region convolutional neural network[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2020; 2054018.
- [27] ZHOU Y, WEN S, WANG D, et al. Object detection in autonomous driving scenarios based on an improved faster-RCNN[J]. Applied Sciences, 2021; 11630.
- [28] YANG J, JI L, GENG X, et al. Building detection in high spatial resolution remote sensing imagery with the U-rotation detection network [J]. International Journal of Remote Sensing, 2019; 6036-6058.
- [29] YANG X, YAN J, MING Q, et al. Rethinking rotated object detection with Gaussian wasserstein distance loss [C]. International Conference on Machine Learning, 2021; 11830-11841.

作者简介



唐心亮, 1999 年于河北工业大学获得学士学位, 2005 年于燕山大学获得硕士学位, 2014 年于河北工业大学获得博士学位, 现为河北科技大学研究员, 主要研究方向为计算机控制、智能硬件与移动互联网应用。

E-mail: tangxinliang@hebust.edu.cn

Tang Xinliang received his B. Sc. degree from Hebei University of Technology in 1999, M. Sc. degree from Yanshan University in 2005 and Ph. D. degree from Hebei University of Technology in 2014. Now he is a researcher at Hebei University of Science and Technology. His main research interests include computer control, intelligent hardware and mobile internet applications.



刘子剑, 在 2021 年于河北科技大学理工学院获得学士学位, 现为河北科技大学硕士研究生, 主要研究方向为深度学习、计算机视觉。

E-mail: 2336617735@qq.com

Liu Zijian received his B. Sc. degree from Hebei University of Science and Technology College of Technology in 2021. Now he is a M. Sc. candidate at Hebei University of Science & Technology. His main research interests include deep learning, computer vision, etc.



于平平 (通信作者), 分别在 2006 年、2008 年和 2010 年于天津大学获得学士学位、硕士学位和博士学位, 现为河北科技大学副教授, 主要研究方向图像处理、计算机视觉和 VLSI 设计等。

E-mail: yppflx@aliyun.com

Yu Pingping (Corresponding author) received B. Sc., M. Sc., and Ph. D. degree all from Tianjin University in 2006, 2008 and 2010, respectively. Now she is a lecturer at Hebei University of Science & Technology. Her main research interests include image processing, computer vision and VLSI design, etc.