

基于 Coordinate Attention 和空洞卷积的异物识别^①



王春霖, 吴春雷, 李灿伟, 朱明飞

(中国石油大学(华东) 计算机科学与技术学院, 青岛 266580)

通信作者: 吴春雷, E-mail: wuchunlei@upc.edu.cn

摘要: 在我国工厂的工业化生产中, 带式输送机占有重要的地位, 但是在其运输物料的过程中, 常有木板、金属管、大型金属片等混入物料中, 从而对带式输送机的传送带造成损毁, 引起巨大的经济损失. 为了检测出传送带上的不规则异物, 设计了一种新的异物检测方法. 针对传统异物检测方法中存在的对于图像特征提取能力不足以及网络感受野相对较小的问题, 我们提出了一种基于 coordinate attention 和空洞卷积的单阶段异物识别方法. 首先, 网络利用 coordinate attention 机制, 使网络更加关注图像的空间信息, 并对图像中的重要特征进行了增强, 增强了网络的性能; 其次, 在网络提取多尺度特征的部分, 将原网络的静态卷积变为空洞卷积, 有效减少了常规卷积造成的信息损失; 除此之外, 我们还使用了新的损失函数, 进一步提高了网络的性能. 实验结果证明, 我们提出的网络能有效识别出传送带上的异物, 较好地完成异物检测任务.

关键词: coordinate attention; 异物检测; 空洞卷积; 损失函数; 目标识别

引用格式: 王春霖, 吴春雷, 李灿伟, 朱明飞. 基于 Coordinate Attention 和空洞卷积的异物识别. 计算机系统应用, 2024, 33(3): 178-186. <http://www.c-s-a.org.cn/1003-3254/9416.html>

Foreign Object Recognition Based on Coordinated Attention and Atrous Convolution

WANG Chun-Lin, WU Chun-Lei, LI Can-Wei, ZHU Ming-Fei

(College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China)

Abstract: In the industrial production of factories in China, belt conveyors play an important role. However, in the process of transporting materials, wooden boards, metal pipes, large metal sheets, etc. are often mixed into the materials, causing damage to the conveyor belt of the belt conveyor and leading to huge economic losses. To detect irregular foreign objects on the conveyor belt, this study designs a new foreign object detection method. It proposes a single stage foreign object recognition method based on coordinated attention and atrous convolution to address the issues of insufficient image feature extraction ability and relatively small network receptive field in traditional foreign object detection methods. Firstly, the network utilizes the coordinated attention mechanism to make the network pay more attention to the spatial information of images and enhance important features in the images, improving the performance of the network. Secondly, while extracting multi-scale features from the network, the static convolution of the original network is transformed into an atrous convolution, effectively reducing the information loss caused by conventional convolution. In addition, the study also uses a new loss function, promoting the property of the network. The experimental results show that the proposed network can effectively identify foreign objects on the conveyor belt and effectively complete the foreign object detection task.

Key words: coordinate attention; foreign object detection; atrous convolution; loss function; target recognition

① 收稿时间: 2023-09-05; 修改时间: 2023-10-09; 采用时间: 2023-10-16; csa 在线出版时间: 2023-12-26

CNKI 网络首发时间: 2023-12-28

带式输送机是我国工厂物料运输的关键一环, 其具有载量大、经济成本低等优势, 现阶段还有没有其他物料输送方式的效益可以超过带式输送机运输, 因此, 未来一段时间内, 带式输送机将继续占据我国工厂工业化生产的重要地位。然而, 带式输送机在运送物料的过程中发生过各种各样的事故, 其中, 由于异物进入带式输送机的输送带而造成输送带损毁的事故占事故总数的 61%^[1]。异物的来源多种多样, 现阶段, 工厂中基本采用人工巡检的方式检查传送带上是否有异物, 但是这种方法无法对异物进行长期有效的检测, 且漏检的概率很大。因此, 利用深度学习构建异物检测网络, 对传送带进行全天候的检测, 是现阶段研究人员的主要研究目标。

目前, 基于深度学习的目标检测算法已经得到了较好的发展, 传统的单阶段检测网络, 例如 YOLOv5^[2]、single shot multibox detector^[3]、VarifocalNet^[4]等, 这些网络以牺牲精度为代价, 大幅度提高了网络的检测速度; 传统的双阶段检测网络, 以 Fast-RCNN^[5]、Mask-RCNN^[6]、R-FCN^[7]为代表, 虽然较好地提升了检测精度, 但随着带来的是模型复杂度的提升以及检测速度的降低。

与传统的目标检测任务既要求检测精度又要求预测框准确地将识别目标框住不同, 异物检测任务的主要目标是将图像中的非物料异物检测出来。而且, 异物的图像与传统图像的特点也不相同, 在传统图像中, 被识别目标一般相对面积较大, 且特征明显, 易于识别, 而异物图像中的异物面积通常相对较小, 且有些异

物与物料的特征相似, 所以异物图像中有关异物特征信息相对较少。在现有的异物检测方法中, 文献[8]通过计算原始图像与重构图像之间的重构误差来检测异物; 文献[9]提出了一种基于 single shot multibox detector 的网络, 通过对网络的损失函数进行改动来提高网络的性能; 文献[10]通过对图像之间的 NMS 差异特征的改动来提高网络的检测精度。然而, 上述网络基本都采用“单一 CNN”或“单一 CNN+new loss”的模式, 对网络的改动较小或者基本不改动, 导致模型的泛化能力不高, 无法针对多种多样的复杂场景, 而且并未考虑异物检测任务的核心问题。因为不规则异物相对较小, 特征信息不多, 因此, 网络必须拥有较强的特征提取能力并且减少信息损失。为了能够解决不规则异物识别任务的核心问题并且避免传统网络存在的问题, 本文提出了 CAFN (coordinate attention and atrous convolution foreign detection network) 模型, 通过在网络中加入 coordinate attention 机制, 提升网络的特征提取能力, 并且将网络 backbone 部分提取多尺度特征的常规卷积换为空洞卷积, 减少信息损失。除此之外, 我们使用了 CIoU 函数作为损失函数, 相较于原始网络的 GIoU 损失函数, CIoU 考虑得更全面, 可以使网络的检测精度进一步提升。

1 CAFN 模型

CAFN 模型的整体框架如图 1 所示。模型添加了 coordinate attention 机制, 来增强模型的特征提取能力, 该模型的具体流程如下。

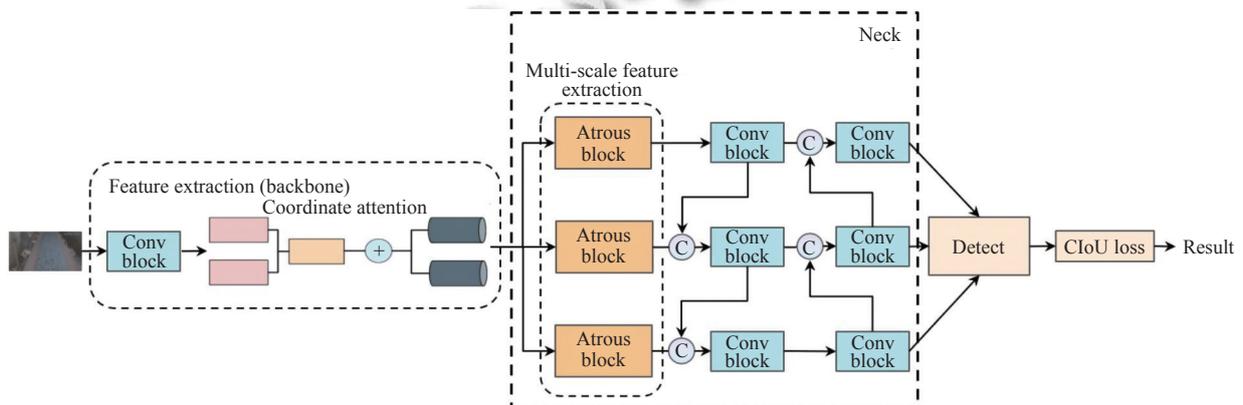


图 1 CAFN 网络结构

- 1) 输入待检测的图像。
- 2) 使用 conv block 提取图像特征, 利用 coordinate

attention 机制将图像空间特征嵌入到通道注意力中, 增强图像特征中有用的信息, 并形成主干特征。

3) 使用由空洞卷积组成的 atrous block 提取主干特征中的不同尺度的特征, 减少信息丢失。

4) 对来自不同分支的不同尺度的特征进行特征融合, 分为下融合和上融合, 然后将融合后的特征输入到检测模块中。

5) 使用新的 CIoU loss 函数计算边界框的损失, 并进行梯度更新。

6) 输出当前模型的预测结果。

网络中的 conv block 模块由多个 RepVGGBlock^[11] 构成, RepVGGBlock 的结构如图 2 所示。受 ResNet^[12] 启发, RepVGGBlock 是一种简单的残差结构, 它利用结构的重新参数化来将训练时的多分支拓扑结构与推理时的简单结构解耦, 与传统的多分支网络相比, 该结构显著降低了推理的时间成本。

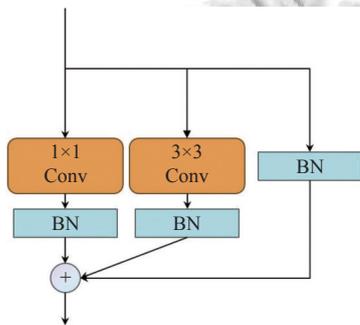


图 2 RepVGGBlock 结构

RepVGGBlock 的计算过程可被概括为式 (1):

$$Output = BN(1 \times 1 Conv(Input)) \oplus BN(3 \times 3 Conv(Input)) \oplus BN(Input) \quad (1)$$

其中, $Input$ 代表输入的特征向量, $1 \times 1 Conv$ 代表卷积核大小为 1×1 的卷积操作, $3 \times 3 Conv$ 代表卷积核大小为 3×3 的卷积操作, BN 代表 BN 函数。

1.1 Coordinate attention

近期的研究结果显示, 通道注意力, 例如 SENet^[13] 注意力机制对于目标检测模型的性能提升有着很大的作用, 但通道注意力通常忽略了图像的空间信息, 也就是位置信息, 而位置信息对于异物检测任务来说非常重要^[14], 因为该任务最关键的一点就是捕获图像中异物的位置。传统的注意力机制, 例如 BAM^[15] 和 CBAM^[16], 尝试通过减少特征向量的通道维度来利用空间位置信息, 然后使用常规卷积计算空间维度的注意力。然而, 常规卷积只能捕获局部特征的关系, 而无法对视觉任务所必需的远程依赖关系进行建模^[17,18]。

Coordinate attention 机制很好地避免了这个问题。Coordinate attention 机制利用两个 1D 的全局池化操作, 首先把沿水平和垂直方向输入的特征信息集成在 2 种单独的方向特征图中, 接着, 再把这 2 个具有嵌入的方向特征信号的特征图分别编码为两种注意力图, 每种注意力图捕获了输入特征图对沿同一个空间方向信息的长距离依赖性。这样, 位置信息就可以被存储到生成的注意力图中。这种方法在避免了二维全局池化引起的位置信息损失的同时, 也避免了产生显著的计算开销。

Coordinate attention 机制的流程图如图 3 所示 (X Avg pool 和 Y Avg pool 分别表示一维水平全局池化和一维垂直全局池化)。在 coordinate attention 机制中, 式 (2) 会被分解为一维的特征编码操作, 其目的是让注意力块利用特征图精确的位置信息捕获远程交互。

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (2)$$

其中, z_c 是与第 c 个通道相关的输出。

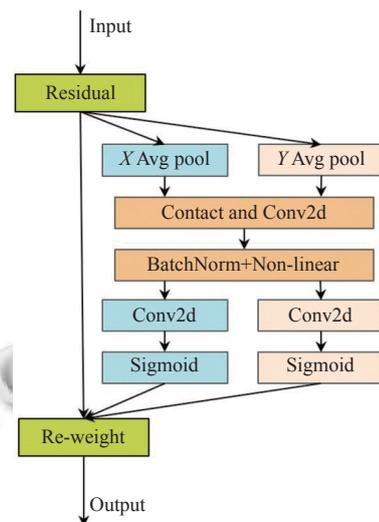


图 3 Coordinate attention 结构

因此, 对于给定输入 X , 通过 CA 机制可以在池化内核的 2 个空间范围 $(H, 1)$ 或 $(1, W)$ 内分别沿水平坐标系和垂直坐标方向对每个通道进行解码。所以, 第 c 个通道在高度 h 处的输出, 可表达为式 (3):

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (3)$$

类似的, 宽度为 w 的第 c 个通道的输出可表示为式 (4):

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq i < H} x_c(i, w) \quad (4)$$

与 SENet 不同,上述的两个沿不同方向聚合特征的池化操作,其会产生两个包含空间信息的特征图,它允许注意力模块捕获不同空间方向的远程依赖性,且另一个空间方向会保留精确的位置信息,这有助于网络定位感兴趣的对象。

之后,CA 机制通过 Contact 操作将两个特征图连接起来,之后进行 1×1 卷积,并进行归一化操作和 BN 操作,过程如式 (5) 所示:

$$f = \delta(F_1([z^h, z^w])) \quad (5)$$

其中, $[\cdot, \cdot]$ 表示沿空间维度的串联操作, δ 是非线性激活函数, f 是对两个不同方向 (水平和垂直) 上的编码空间信息的特征图。然后沿着空间维度将 f 分解成两个独立的张量 f^w 和 f^h 。接下来的两个 1×1 的卷积变换 F_h 以及 F_w 用于分别将 f^w 和 f^h 变换为与输入 X 具有相同通道数的张量。过程可被分别概括为式 (6) 以及式 (7)。

$$g^h = \sigma(F_h(f^h)) \quad (6)$$

$$g^w = \sigma(F_w(f^w)) \quad (7)$$

其中, σ 是 Sigmoid 函数。

之后,输出 g^h 和 g^w 被扩展并分别用作注意力权重。最后,CA 机制的输出 Y 可被写为式 (8):

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (8)$$

Coordinate attention 机制的具体算法描述如算法 1 所示。

算法 1. Coordinate attention

输入: 图像浅层特征 *Input*。

输出: 增强后的图像特征 *Output*。

1. 将输入特征 *Input* 经过一维水平全局池化和一维垂直全局池化分别得到两个输出特征 z^w 和 z^h 。
2. 将 z^w 和 z^h 经过 Contact 以及卷积得到 f 。
3. 将 f 分别经过两个卷积得到 g^h 和 g^w 。
4. 将 g^h 和 g^w 分别经过 Sigmoid 函数以及残差连接得到最终输出 *Output*。

1.2 空洞卷积

基于空洞卷积的模型被积极的用于语义分割和目标检测。文献[19]实验了修改 dilation rate 以捕获提高网络捕获长距离信息的能力,文献[20]在 ResNet 的最后两个模块中采用了混合空洞率,而文献[21]进一步提出了可变形卷积,是空洞卷积概念的进一步延伸。文献[22]利用了图像标题和空洞卷积,进一步提高了分割

模型的准确性。

文献[23]认为,尽管目标检测近年来取得了很大的进展,但是针对小尺度物体的检测问题,现阶段仍然没有较好的解决办法,基于低级线索或 DCNN 的现有对象提议机制在小规模检测对象上的整体召回率较低,而且 DCNN 的重复下采样导致深度卷积丢失了特征图的部分信息,而这部分信息是高级特征信息,对于目标检测来说是非常重要的,为了解决这两个问题,文献[23]提出了一种基于 Fast-RCNN 的统一深度神经网络,这个网络是提出了一个空洞区域提议网络 ARPN,在最后一个卷积特征图上滑动一组空洞滤波器,并增加膨胀率,这个滤波器能够在不增加参数或操作数量的情况下扩大感受野,从而能使 ARPN 更高效地捕获小尺度对象的特征;文献[24]提出了一种基于空洞卷积的 A2SPP (atrous atrous spatial pyramid pooling) 模块。通过向原始 ASPP 的每个分支添加 CIEA 模块,生成的 A2SPP 可以自动增强每个 3D 位置的基本尺度并抑制噪声尺度。这种特征尺度聚合的方式提高了多尺度表示的能力。此外, A2SPP 添加了一个注意力残差分支来丰富多尺度特征。使用 A2SPP 和 CIEA 作为基本模块,构建了一个简单的编码器-解码器网络,即 A2SPPNet。除此之外,文献[25-27]将空洞卷积应用于目标检测。

图 4 分别展示了空洞率为 1 的空洞卷积 (图 4(a)) 和空洞率为 2 的空洞卷积 (图 4(b))。当空洞率为 1 时,空洞卷积其实就是普通卷积。本文将提取多尺度特征的卷积方式由常规卷积变为了空洞卷积。由于空洞卷积能扩大感受野,所以在提取多尺度特征时,空洞卷积可以有效地减少信息丢失。而且,空洞卷积不会增加参数量,因为卷积核大小和步长并未改变,也没有增加新的卷积操作。在我们的网络中,我们将空洞率设置为 2。

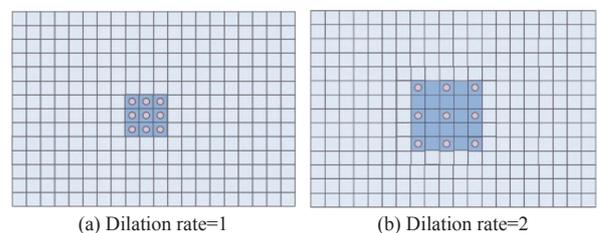


图 4 空洞卷积

1.3 损失函数

目标检测的损失函数一般由 3 部分组成: 分类损失、框回归损失以及对象损失。

框回归损失的主要目的是精确定位被检测对象的

边框. 其中, IoU 损失是现阶段常用且已经被公认有效的损失.

原始网络使用的是常规的 $GIoU$ 损失函数, 他的损失函数表达式如式 (9) 所示:

$$GIoU = IoU - \frac{C - (A \cup B)}{C} \quad (9)$$

其中, IoU 的表达式如式 (10) 所示, IoU 简称交并比. 图形化表达如图 5 所示.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (10)$$

其中, A 是网络预测框 (predict box), 也就是网络预测的物体位置; B 是真实框 (ground truth box), 也就是物体的正确位置.

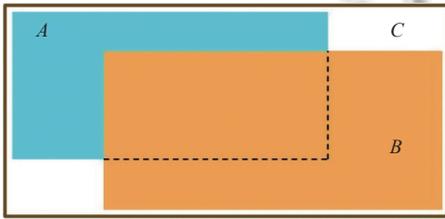


图 5 IoU 图形化表达

传统的 $GIoU$ 损失函数, 仅考虑重叠框的面积损失, 考虑元素较为单一, 严重依赖于 IoU 损失, 无法满足复杂的应用场景. 而 $CIoU$ ^[28] 损失函数, 不仅考虑了框的重叠面积损失, 还考虑了 predict 框与 ground truth 框中心点距离的归一化损失以及纵横比损失. $CIoU$ 损失函数的形式如式 (11) 所示:

$$L_{CIoU} = 1 - IoU + \frac{\rho(\mathbf{b}, \mathbf{b}^{gt})}{c^2} + \alpha v \quad (11)$$

其中, v 和 α 的公式如式 (12) 以及式 (13) 所示:

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (12)$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (13)$$

2 实验分析

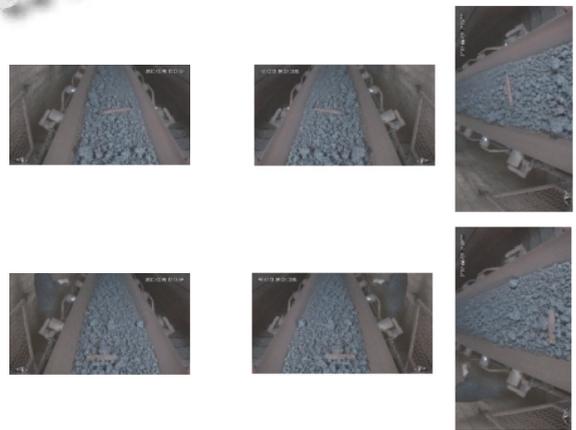
2.1 数据集

由于不规则异物检测任务比较特殊, 其数据通常都来自于工厂实地, 涉及企业机密, 因此现阶段并没有异物的公开数据集. 我们使用的异物数据集均来自于内蒙古某工厂, 我们通过现场采集和人工标注, 并对数

据进行了数据增强, 构成了最终的不规则异物检测数据集, 数据集的数量及划分如表 1 所示. 我们进行的数据增强包括左右变换、旋转操作, 如图 6 所示. 除此之外, 我们还是用了公共数据集 COCO 对网络性能进行了评估, 以证明 CAFN 模型在传统目标检测任务上的有效性. COCO 数据集是一个可以用于大规模目标检测任务的数据集, 它的 trainset 有 118000 张图像, valset 有 5000 张图像.

表 1 不规则异物数据集 (IFO) 划分

数据集	训练集	验证集	测试集	总数
原始数据集	275	79	39	393
增强后的数据集	825	236	118	1179



(a) 原图 (b) 增强后的图像

图 6 不规则异物数据集的数据增强

2.2 实验环境

本文实验环境配置如表 2 所示.

表 2 实验环境配置

环境	配置
系统	Ubuntu 20.04
显卡	NVIDIA Tesla P100 GPU
内存	128 GB DDR4
编辑环境	PyTorch 1.9.1
编程语言	Python 3.7
计算机架构	CUDA 11.7

2.3 评价指标

不规则异物识别任务的核心目标就是将异物的位置标出, 所以预测框的位置是否精确直接决定了任务的成功与失败. 因此, 我们选用了 AP 、 AP_{50} 、 AP_{75} 这 3 个指标来进行网络性能的评价. 其中, 3 个指标中 AP 的含义是平均精度 (average precision, AP), AP 指标的含义从 $IoU=0.5$ 到 $IoU=0.95$ 之间, 以 0.05 为步长, 所

有区间的 AP 值加起来然后取平均值; AP_{50} 是指当 IoU 大于 0.5 时, 网络的 AP 值; AP_{75} 的含义是 IoU 大于 0.75 时, 网络的 AP 值. AP (平均精度) 由召回率 (recall, R) 和准确率 (precision, P) 计算而来, AP 的值越大, 说明网络的检测精度越高. 除此之外, 我们还使用了 FPS 指标来衡量网络的检测速度, FPS 的含义是 1 s 的时间间隔内处理的图片数量.

2.4 实验参数

我们实验的参数如表 3 所示, 除此之外, 我们使用随机梯度下降 (SGD) 优化器以及余弦衰减的学习率. 此外, 还利用了 Warm-up、分组权重衰减策略和平均指数移动 (EMA).

表 3 实验参数

实验参数	数值
batch size	128
epochs	600
backbone out channels	[64, 128, 256, 512, 1024]
neck out channels	[256, 128, 128, 256, 256, 512]
learning rate	0.0032

2.5 收敛曲线分析

使用 CAFN 模型在不规则异物数据集上经过 600 个轮次的训练后, loss 的变化趋势如图 7 所示. 从图中可以看出, 损失曲线在开始训练阶段, 损失从 1.18 左右急速下降到 0.7 左右, 以及 580 个 epochs 模型损失从 0.35 左右急速下降到 0.2 左右, 这两个节点表明模型在急速收敛. 除了这两个特殊节点外, 其余曲线均缓慢波动下降, 整体损失是下降的, 证明我们的模型是有效的.

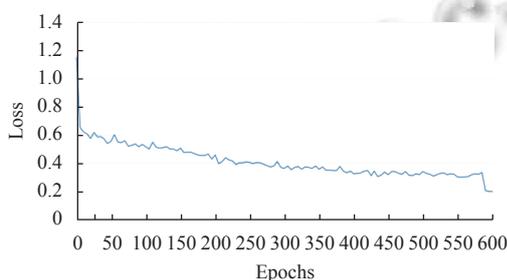


图 7 CAFN 模型损失变化曲线

2.6 数据增强实验及分析

为了证明数据增强的有效性, 我们使用原始网络和 CAFN 网络分别对 3 种数据增强情况下的数据集进行了实验, 3 种数据增强情况为: 1) 没有数据增强, 即使用原始数据集; 2) 仅左右变换, 即原始数据集+数据

增强 (左右变换); 3) 完整增强 (原始数据集+左右变换+旋转). 实验结果如表 4 所示, 其中, OM 代表原始模型, CAFN 代表本文提出的 CAFN 模型. Y 代表原始数据集, LR 代表数据增强 (左右变换), RO 代表数据增强 (旋转).

表 4 数据增强实验 (%)

方法	OM	CAFN	AP	AP_{50}	AP_{75}
Y	√	—	41.4	67.0	50.5
	—	√	41.8	67.4	51.3
Y+LR	√	—	68.7	83.4	75.1
	—	√	69.2	84.2	78.0
Y+LR+RO	√	—	64.3	87.4	74.5
	—	√	69.7	91.0	76.7

从表 4 中可以看出, 无论是原始模型还是 CAFN 模型, 经过数据增强后, 其模型的检测精度均有提升. 针对原始模型, 当仅使用原始数据集和左右旋转增强的数据集时, 相较于使用原始数据集, 原始模型的 AP 增加了 27.3%, AP_{50} 增加了 16.4%, AP_{75} 增加了 24.6%; 当使用完整增强过的数据集后, 相较于使用原始数据集, 原始模型的 AP 增加了 22.9%, AP_{50} 增加了 20.0%, AP_{75} 增加了 24.0%; 针对本文提出的 CAFN 模型, 当仅使用原始数据集和左右旋转增强的数据集时, 相较于使用原始数据集, CAFN 网络的 AP 增加了 27.4%, AP_{50} 增加了 16.8%, AP_{75} 增加了 26.7%, 使用完整增强过的数据集后, 相较于使用原始数据集, CAFN 网络的 AP 增加了 27.9%, AP_{50} 增加了 23.6%, AP_{75} 增加了 25.4%. 从结果可以看出, 本文进行的数据增强工作是有效的.

2.7 不同网络对比分析

如表 5 和表 6 所示, 我们选用 Faster-RCNN、Mask-RCNN、YOLOv6-S 以及 DETR 这 4 种网络与我们提出的 CAFN 模型在异物数据集以及公共数据集 COCO 上进行了对比实验. Faster-RCNN 是一种基于快速区域的卷积网络方法, Mask-RCNN 为每个实例生成高质量的分割掩码, 通过对特征图进行随机遮挡来提高网络的鲁棒性, YOLOv6-S 提出了一个基于高效可重新化参数的主干以及 PAN 拓扑的目标检测网络, 该方法较好地平衡了网络的速度与精度, DETR 提出了一种基于 Transformer 的端到端的检测网络. 通过与这些网络进行对比实验, 证明 CAFN 模型在不规则异物检测任务以及传统目标检测任务上的有效性.

表5 不同网络在不规则异物数据集上的对比实验

Model	AP (%)	AP ₅₀ (%)	AP ₇₅ (%)	FPS
Faster-RCNN	52.4	74.1	60.8	—
Mask-RCNN	63.5	86.3	70.8	—
YOLOv6-S	64.3	87.4	74.5	260
DETR	68.9	90.0	77.5	22
ACFN	69.7	91.0	76.7	235

表6 不同网络在公共数据集 COCO 上的对比实验

Model	COCO val 2017			
	AP (%)	AP ₅₀ (%)	AP ₇₅ (%)	FPS
Faster-RCNN	21.2	41.5	—	—
Mask-RCNN	37.1	60.0	39.4	—
DETR	42.0	62.4	44.2	28
YOLOv6-S	43.5	60.4	47.5	358
ACFN	43.8	61.5	47.6	304

从表5中可以看出,在不规则异物数据集上,DETR 因为是以 Transformer 为基础结构的网络,所以检测速度很慢;我们提出的 ACFN 模型虽然识别速度不是最快的,但是识别效果是最好的.相较于每个主流网络,ACFN 模型在 AP、AP₅₀、AP₇₅ 这3个指标上均是最高,从而证明了 ACFN 网络在不规则异物检测任务上的有效性.

从表6可以看出,在公共数据集 COCO 上,DETR 因为是以 Transformer 为基础结构的网络,所以虽然检测结果高但是检测速度很慢,而且 DETR 网络只有 AP₅₀ 指标高于 ACFN 网络,其他所有网络的所有指标均低于 ACFN 网络,证明我们提出的网络在目标检测任务上尤其是高精度目标检测上优于传统网络.表6的实验结果证明我们提出的 ACFN 网络在传统目标检测任务上也是十分有效的.



图8 检测结果可视化实验

从结果中可以看出,当检测目标处于图像边缘时,将不同的模块加入到网络中,检测效果均有提升,当把所有的模块都加入到网络中,组成本文提出的

2.8 消融实验及分析

本节中,我们对网络进行了消融实验,因为本文一共提出了3个创新点,所以我们一共进行了7种不同组合的消融实验,通过把某个模块或某两个模块增量地添加到原始网络当中,来证明我们提出的模块是有效的.实验结果见表7,其中,CA 代表 coordinate attention 机制, Atrous 代表空洞卷积, CIOU 代表新的损失函数.

从表7中可以看出,当把 coordinate attention 模块、 atrous block 模块以及 CIOU 分别加入到网络中, AP 值分别提高了 2.2%、1.4%、1.3%, AP₅₀ 分别提高了 1.0%、0.2%、0.6%; AP₇₅ 分别提高了 0.1%、0.3%; 各个指标均有不同程度的提升,证明我们提出的各个模块是有效的.

表7 消融实验 (%)

CA	Atrous	CIOU	AP	AP ₅₀	AP ₇₅
×	×	×	64.30	87.40	74.50
√	×	×	67.50	88.40	74.60
×	√	×	66.70	87.60	74.30
×	×	√	67.30	88.90	74.80
√	√	×	68.20	89.40	74.50
√	×	√	68.60	90.10	75.80
×	√	√	68.90	89.50	76.10
√	√	√	69.70	91.00	76.70

2.9 检测结果可视化实验

可视化实验的结果如图8所示,其中,OM 代表原始模型; CA 代表 coordinate attention 机制; Atrous 代表空洞卷积; ACFN 代表本文提出的 ACFN 模型.最左边的上下两张图片是两个不同的原始图像,右边依次是添加不同模块后的可视化结果对比.

ACFN 模型,检测效果是最好的;当检测目标与物料的特征差异较小时,我们提出的模型检测效果也相对较好.

3 结论

针对传统异物识别网络对于异物图像特征提取能力不足的问题、信息丢失的问题以及损失函数考虑不全面的问题,我们提出了ACFN网络.其中,coordinate attention机制通过将空间信息嵌入到通道注意力中,使模型可以更好地获得图像的空间位置信息,这对于异物识别任务来说有重要的意义.空洞卷积由于其可以扩大感受野,所以可以很好地防止异物特征的丢失,CIoU函数因为其全面的考虑,所以拥有了比传统IoU损失函数更好的性能.

实验结果表明,增强网络的特征提取能力以及减少特征信息的丢失,不仅可以很好地解决异物识别的问题,同时在传统目标检测任务上也有较好的性能.

参考文献

- 曹虎奇. 煤矿带式输送机撕带断带研究分析. 煤炭科学技术, 2015, 43(S2): 130-134.
- Wu WT, Liu H, Li LL, *et al.* Application of local fully convolutional neural network combined with YOLOv5 algorithm in small target detection of remote sensing image. PLoS One, 2021, 16(10): e0259283. [doi: [10.1371/journal.pone.0259283](https://doi.org/10.1371/journal.pone.0259283)]
- Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21-37.
- Zhang HY, Wang Y, Dayoub F, *et al.* VarifocalNet: An IoU-aware dense object detector. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 8510-8519.
- Girshick R. Fast R-CNN. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1440-1448.
- He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2980-2988.
- Dai JF, Li Y, He KM, *et al.* R-FCN: Object detection via region-based fully convolutional networks. Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016. 379-387.
- 卢学明, 于在川, 许升起. 基于深度生成模型的煤矿运输皮带异物检测. 计算机系统应用, 2022, 31(5): 358-363. [doi: [10.15888/j.cnki.csa.008468](https://doi.org/10.15888/j.cnki.csa.008468)]
- Wang YB, Wang YJ, Dang LF. Video detection of foreign objects on the surface of belt conveyor underground coal mine based on improved SSD. Journal of Ambient Intelligence and Humanized Computing, 2023, 14(5): 5507-5516. [doi: [10.1007/s12652-020-02495-w](https://doi.org/10.1007/s12652-020-02495-w)]
- Xiao CK, Sun B, Wang YL, *et al.* Foreign object detection of sintering transport belt based on CNN. IFAC-PapersOnLine, 2021, 54(21): 25-30. [doi: [10.1016/j.ifacol.2021.12.005](https://doi.org/10.1016/j.ifacol.2021.12.005)]
- Ding XH, Zhang XY, Ma NN, *et al.* RepVGG: Making VGG-style ConvNets great again. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13728-13737.
- Targ S, Almeida D, Lyman K. ResNet in ResNet: Generalizing residual architectures. arXiv:1603.08029, 2016.
- Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132-7141.
- Wang HY, Zhu YK, Green B, *et al.* Axial-DeepLab: Stand-alone axial-attention for panoptic segmentation. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 108-126.
- Park J, Woo S, Lee JY, *et al.* BAM: Bottleneck attention module. Proceedings of the 2018 British Machine Vision Conference. Newcastle: BMVA Press, 2018. 147.
- Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3-19.
- Zhao HS, Shi JP, Qi XJ, *et al.* Pyramid scene parsing network. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6230-6239.
- Hou QB, Zhang L, Cheng MM, *et al.* Strip pooling: Rethinking spatial pooling for scene parsing. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 4002-4011.
- Wu ZF, Shen CH, van den Hengel A. Bridging category-level and instance-level semantic image segmentation. arXiv: 1605.06885, 2016.
- Wang PQ, Chen PF, Yuan Y, *et al.* Understanding convolution for semantic segmentation. Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision. Lake Tahoe: IEEE, 2018. 1451-1460.
- Dai JF, Qi HZ, Xiong YW, *et al.* Deformable convolutional networks. Proceedings of the 2017 IEEE International

- Conference on Computer Vision. Venice: IEEE, 2017. 764–773.
- 22 Wang GR, Luo P, Lin L, *et al.* Learning object interactions and descriptions for semantic image segmentation. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 5235–5243.
- 23 Guan TF, Zhu H. Atrous Faster R-CNN for small scale object detection. Proceedings of the 2nd International Conference on Multimedia and Image Processing. Wuhan: IEEE, 2017. 16–21.
- 24 Qiu Y, Liu Y, Chen YN, *et al.* A2SPPNet: Attentive atrous spatial pyramid pooling network for salient object detection. IEEE Transactions on Multimedia, 2022, 25: 1991–2006.
- 25 Papandreou G, Kokkinos I, Savalle PA. Modeling local and global deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 390–399.
- 26 Wang B, Ding H, Chen C. AC-YOLOv4: An object detection model incorporating attention mechanism and atrous convolution for contraband detection in X-ray images. Multimedia Tools and Applications, 2023: 1–20.
- 27 Huang J, Rathod V, Sun C, *et al.* Speed/accuracy trade-offs for modern convolutional object detectors. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 3296–3297.
- 28 Zheng ZH, Wang P, Liu W, *et al.* Distance-IoU loss: Faster and better learning for bounding box regression. Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI, 2020. 12993–13000.

(校对责编: 孙君艳)