赵健,王奕,王海峰,等. 基于TDConv与统一注意力检测头的异物检测算法[J]. 矿业安全与环保,2024,51(4):26-34.
ZHAO Jian, WANG Yi, WANG Haifeng, et al. Foreign object detection algorithm based on TDConv and unified attention detection head[J]. Mining Safety & Environmental Protection, 2024, 51(4): 26-34.
DOI: 10.19835/j.issn.1008-4495.20240548

扫码阅读下载

基于TDConv与统一注意力检测头的异物检测算法

赵 健[1],王 奕[2],王海峰[1],程德强[2],李自豪[2]
(1. 内蒙古白音华蒙东露天煤业有限公司,内蒙古 二连浩特 012600; 2. 中国矿业大学,江苏 徐州 221116)

摘要:针对矿井输送带在输送煤流的过程中,大块煤矸石和锚杆存在不同尺寸和形状,图像特征信息难以提取,传统目标检测算法检测效果不理想的问题,提出一种基于TDConv与统一注意力检测头的异物检测算法。该算法通过设计并行卷积的组合方式形成TDConv卷积模块,能有效保持图像特征原有信息,帮助更深的卷积层提取有效细节信息;在检测头部分加入统一注意力模块,有效提取和识别不同尺寸物体、不同空间位置之间的特征信息;基于煤矿井下不同场景的输送带制作了10万张矿用异物数据集(MFID),为矿井煤流输送过程中异物检测的深入研究和实际应用提供资源支持。实验结果表明,该算法在矿用数据集MFID上的平均精度均值(mAP)与YOLOv5目标检测算法相比提升了2.1%;在具备高精度检测能力的同时,能有效减少异物检测网络模型参数量,使网络结构更加轻量化,适用于煤矿井下边缘计算设备。

关键词:物料输送;目标检测;异物识别;深度学习;注意力机制;YOLOv5
中图分类号:TP391.4;TD528.1 **文献标志码**:A **文章编号**:1008-4495(2024)04-0026-09

Foreign object detection algorithm based on TDConv and unified attention detection head

ZHAO Jian[1], *WANG Yi*[2], *WANG Haifeng*[1], *CHENG Deqiang*[2], *LI Zihao*[2]
(*1. Inner Mongolia Baiyinhua Mengdong Open-Pit Coal Industry Co., Ltd., Erenhot 012600, China*;
2. China University of Mining and Technology, Xuzhou 221116, China)

Abstract: In the process of coal folw transportation on mine conveyors, there are different sizes and shapes of rock bolt and large coal gangue, so it is difficult to extract image feature imformation, and the inspection effect of traditional objecet detection algorithm is not deal. To solve this problem, a foreign object detection algorithm based on TDConv and a unified attention detection head is proposed. This algorithm designs a TDConv convolution module by combining parallel convolution methods, effectively preserving the original information of image features and assisting deeper convolutional layers in extracting detailed information. An unified attention module is incorporated into the detection head to effectively extract and recognize feature information from objects of different sizes and spatial positions. A dataset of 100 000 images (MFID) of mining foreign objects was created based on various underground coal mine scenarios, providing resources for in-depth research and practical application of foreign object detection during coal flow transportation. Experimental results demonstrate that this algorithm improves the mean Average Precision (mAP) by 2.1% compared to the YOLOv5 object detection algorithm on the MFID mining dataset. Furthermore, it effectively reduces the parameter count of the foreign object detection network model while maintaining high detection accuracy, resulting in a more lightweight network structure suitable for magrinal computing devices in underground coal mines.

Keywords: material transport; object detection; foreign object recognition; deep learning; attention mechanism; YOLOv5

收稿日期:2024-06-14;2024-07-28修订
基金项目:国家自然科学基金面上项目(51774281);徐州市推动科技创新专项资金项目(KC23401)

作者简介:赵 健(1981—),男,内蒙古通辽人,高级工程师,主要从事矿山机电一体化、信息化与智能化等技术工作。E-mail:38847510@qq.com。
通信作者:程德强(1979—),男,河南洛阳人,教授,博士研究生导师,主要研究方向为机器视觉与模式识别、图像智能检测与信息处理,E-mail:chengdq@cumt.edu.cn。李自豪(2001—),男,湖南浏阳人,硕士研究生,主要研究方向为图像处理,E-mail:1355526134@qq.com。

目标检测中的异物检测,在矿井物料输送中起着至关重要的作用[1],其利用深度学习技术来识别和排除输送带上的异物,确保矿井的安全生产。

目标检测技术的发展经历了几个阶段。最初,

主要依赖人工巡查,这种方法不仅效率低,且容易受到人为因素的影响,导致检测准确性不高。随着计算机视觉和图像处理技术的发展,传统的图像处理算法开始被应用于异物检测,但这种方法对于复杂场景和多样化的异物识别能力有限[2]。

近年来,随着深度学习技术的兴起,基于深度学习的目标检测模型逐渐被广泛应用[3]。这些模型能够自动学习图像中的特征,提高了异物检测的准确性和鲁棒性。其中,卷积神经网络是应用广泛的深度学习模型之一,在图像分类、目标检测[4]、深度估计及语义分割等任务中取得了显著的成效。

在矿井物料输送场景下,基于深度学习的异物检测模型可以实时监控输送带上的物料,准确识别出异物,如大块煤矸石、锚杆等,并及时发出警报或触发相应的排除机制。这种智能检测方式[5]大大提高了矿井的安全性和运行效率。然而,矿井环境复杂多变,异物的种类和形态也千差万别,这给异物检测[6]带来了一定的挑战。目前,深度学习领域的研究主要集中在提升网络的性能方面,这通常需要构建更庞大和复杂的模型,随之而来的是模型参数和所需计算资源量显著增加。虽然在嵌入式设备和移动平台上运用这些高级深度模型的需求很强烈,但对内存、功耗和处理能力有着较高的要求,现有的硬件资源往往难以满足。因此,受硬件限制,许多复杂的算法模型并不能被有效地部署在井下等资源受限的移动环境中。现阶段关于井下输送带异物识别的不足之处在于:①矿井下环境恶劣,难以有效提取图像的更深层信息,检测精度不理想;②输送带异物检测过程中,当同一图像中存在不同尺度的物体时,检测精度会受到影响;③井下样本数据不足,且收集的数据集相对于真实矿井环境较为理想化,网络模型对井下复杂工况环境的鲁棒性差。

1　基于 TDConv 与统一注意力检测头的异物检测算法

1.1　矿井异物数据集构建

目前,多数目标检测算法都是基于包括 COCO[7]、PASCAL VOC[8]和 KITTI[9]等公共数据集进行模型训练的,这些数据集包括飞机、自行车、鸟、船、瓶子、公交车、猫、椅子、牛、餐桌、狗、马、摩托车、人、盆栽、羊、沙发和火车等实物图像,然而当面对灰暗模糊、难以提取特征信息的矿井图片时,目标检测具有一定的局限性。因此,需要构建一种矿井异物数据集(MFID)用于网络训练模型,使模型在真实矿井异物检测中具有更好的泛化能力。

煤矿井下图像数据集是通过 KBA12B 矿用本安型摄像仪在多个煤矿井下拍摄获得的,该摄像仪最高分辨率为 2 560 像素×1 920 像素,补光距离为 30 m,具有结构紧凑、体积小、防爆和防潮等特点,非常适合在煤矿井下使用:①使用 KBA12B 摄像仪在多个煤矿井下的不同场景采集视频图像,包括井下巷道、井下工作车间和下沉井筒等场景;②对图像集进行筛选处理,剔除损坏的和极其模糊的图像,并将图像尺寸统一裁剪为 2 040 像素×1 368 像素;③从初选的 100 000 张数据集中筛选出 39 849 张矿井图像,包括 25 456 张大块煤矸石图像和 14 393 张锚杆图像,部分图像如图 1 所示。图像数据集基本反映了实际生产过程中各种常见场景下的矿井输送带异物,具有较强的适应性。

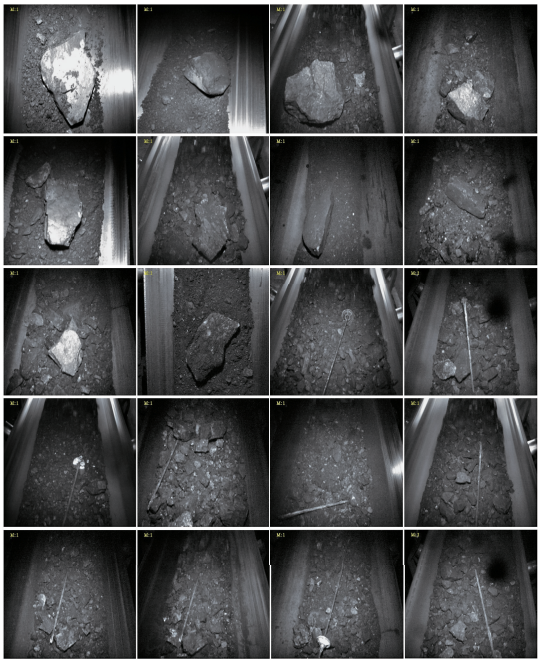


图 1　矿井异物数据集的部分图像展示
Fig. 1　Partial image presentation of MFID

1.2　算法框架

基于 TDConv 与统一注意力检测头的异物检测算法(以下简称“本文算法”),主要在 YOLOv5[10]主干网络的第 2、3、4、5 个 CBS 模块(由卷积、批归一化、SiLU 激活函数构成)中加入了并行双卷积核(TDConv),可更有效地保持图像特征原有信息,帮助更深的卷积层提取有效细节信息;其次在检测头中加入统一注意力模块,通过尺度感知注意力机制有效提取和识别不同大小物体的特征,通过空间感知注意力机制更好地提取空间位置之间的特征信息,以及加入任务感知注意力机制能够根据对象引起的各种卷积核响应,指导各个特征通道专注于各自的任务,从而进一步提高检测精度。本文算法网络结构如图 2 所示。

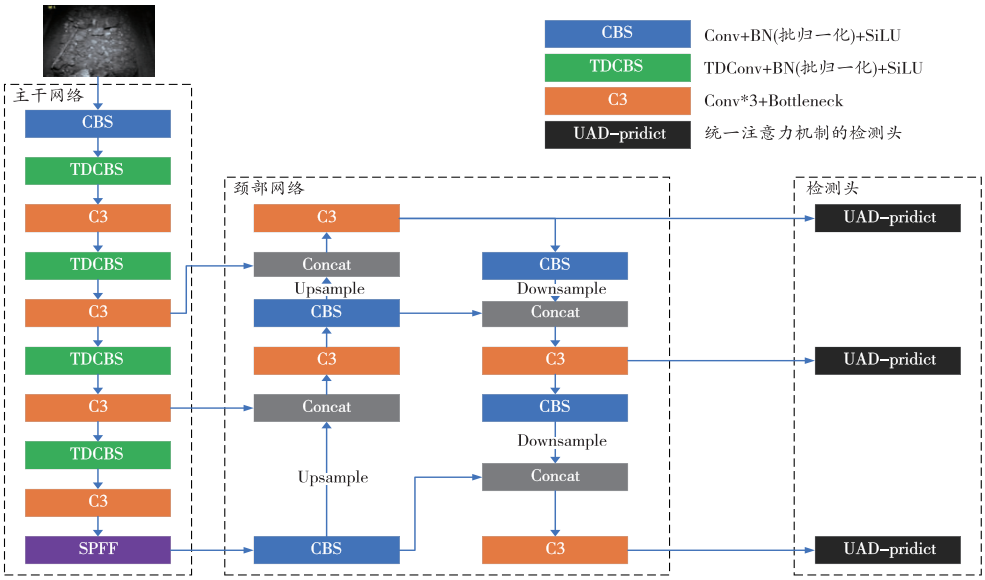


图 2 基于 TDConv 与统一注意力检测头的异物检测算法网络结构图

Fig. 2 Network structure diagram based on TDConv with unified attention detection head algorithm

1.3 并行双卷积核(TDConv)模块

假设输入特征图的尺寸为 $D_0 \times D_1 \times M$(其中 D_0 和 D_1 分别为输入特征图的宽度和高度,M 为卷积输入层的通道数),N 为卷积输出层的通道数(也是卷积核的个数),标准卷积核大小为 $K \times K$。

受 HetConv[11] 与 GroupConv[12] 启发,在二者的基础上进行改进,提出一种并行双卷积核,TDConv 中一些卷积核同时执行 1×1 和 3×3 的卷积运算,其他卷积核只用来执行 1×1 的卷积运算。在输入特征图上进行连续的 1×1 的卷积,不仅降低了网络参数和计算复杂度,还提供了控制特征图深度的灵活度,实现跨通道融合。使用 1×1 的卷积对输入特征图进行滤波时,会将输入特征图的每个通道的原始信息融合到输出特征图中,可以帮助更深的卷积层更有效地提取信息。与 HetConv 相比,TDConv 不仅解决了 GroupConv 通信不良的问题,而且提高了深度神经网络的性能,基本消除了 HetConv 对输入特征图完整信息保存产生的负面影响。

并行双卷积核模块主要将 N 个卷积滤波器分成 T 组,每组处理完整的输入特征图,其中 M/T 个输入特征图通道由 3×3 和 1×1 卷积核同时处理,同时将 N/T 个卷积核平均放入 T 组中,每组中有 N/T^2 个 3×3 和 1×1 并行双卷积核,在每组中以图 3 的顺序进行排列,其余($M-M/T$)输入信道仅由 1×1 卷积核处理。将同时处理 3×3 和 1×1 卷积核的结果进行相加,如图 3 所示。

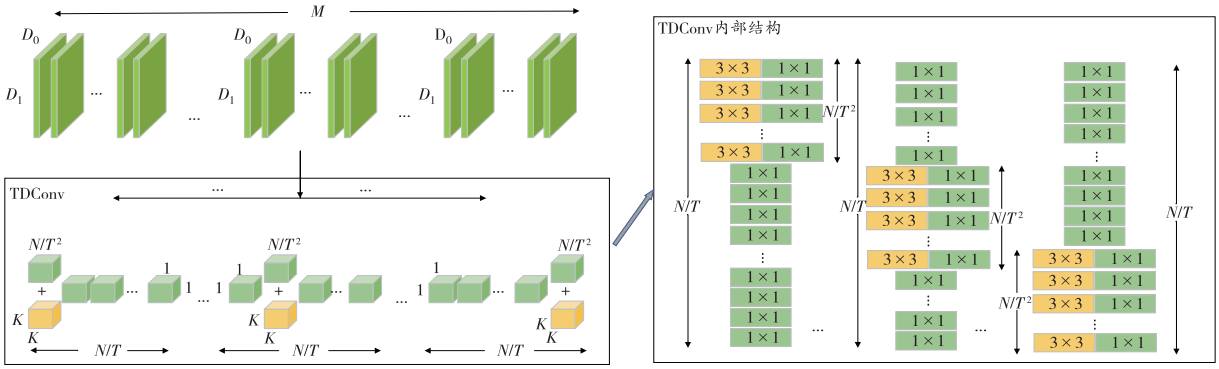


图 3 TDConv 模块结构图

Fig. 3 TDConv module structure diagram

TDConv 设计方案,通过分组卷积策略减少了原始骨干网络模型的参数,并通过保留输入特征图的原始信息和允许 M 个 1×1 卷积的最大交叉跨通道通信,促进了卷积层之间更好的信息共享。因此,可以在不需要通道随机混合的情况下构造 TDConv。

在 TDConv 中,卷积滤波器组的数量用于控制卷积滤波器中 $K \times K$ 卷积核的比例。对于给定的组数 T,大小为($K \times K+1 \times 1$)的组合卷积核的比例为所有通道的 $1/T$,而剩余的 1×1 卷积核比例为($1-1/T$)。因此,其中占 $1/T$ 比例的由并行双卷积核组成的卷

积滤波器的计算量 $F_{TD(K+1)}$ 如下：

$$F_{TD(K+1)}=(D_0\times D_1\times K^2\times M\times N+D_0\times D_1\times M\times N)/T \tag{1}$$

剩余 1×1 卷积核的 $F_{TD(1)}$ 为：

$$F_{TD(1)}=(D_0\times D_1\times M\times N)\times\left(1-\frac{1}{T}\right) \tag{2}$$

所以 TDConv 的计算量总数 F_{TDC} 为：

$$F_{TDC}=F_{TD(K+1)}+F_{TD(1)}=D_0\times D_1\times K^2\times M\times N/T+D_0\times D_1\times M\times N \tag{3}$$

标准卷积核计算量 F_K 为：

$$F_K=D_0\times D_1\times M\times N\times K\times K \tag{4}$$

将 TDConv 的计算量与标准卷积层进行对比，计算缩减比 $R_{TDC/K}$：

$$R_{TDC/K}=\frac{F_{TDC}}{F_K}=\frac{1}{T}+\frac{1}{K^2} \tag{5}$$

从式(5)中可以看出，当卷积核的尺寸 K 为 3、T 取较大值时，加速可以达到 8~9 倍。

1.4　融合统一注意力机制的预测头

在目标检测过程中，优秀的预测模块首先要能够识别图像中存在的多尺度物体，同时应具备对空间变化的认知，能够适应物体在不同视角下的形状、旋转和位置变化。此外，由于目标可能以不同形式出现，如边界框、中心点或角点，每种表现形式都有其独有的特征，因此预测模块还应具备任务适应性。

统一注意力检测头，融合了 3 种注意力机制[13]，包含尺度、空间和任务的注意力机制：通过尺度感知注意力机制有效提取和识别不同大小物体的特征；使用空间感知注意力机制可更好地提取空间位置之间的特征信息；加入任务感知注意力机制能够根据对象引起的各种卷积核响应，指导各个特征通道专注于各自的任务。

在构建特征金字塔时，定义一个由 L 层不同级别特征组合的特征序列 $\xi_{in}=\{F_i\}_{i=1}^{L}$。利用上采样或下采样技术，通过调整邻近层级上的特征尺寸以匹配中间层级特征的大小。经过这种调整，特征金字塔可以被表达为一个四维的张量 $\boldsymbol{\xi}\in R^{L\times H\times W\times C}$，这里的 L 代表金字塔的层数，而 H、W、C 分别指代中间层特征的高度、宽度和通道数量。为了简化这个四维张量，引入 $S=H\times W$ 的概念，将其转换为三维张量 $\boldsymbol{\xi}\in R^{L\times S\times C}$。

1.4.1　尺度感知注意力机制

在物体检测领域的研究进展中，尺度感知的重要性已经被广泛认同和证实。自然场景中的物体往往呈现出广泛的尺度多样性，给精确检测带来了挑战。

初步提出一种尺度感知注意力机制，旨在依据不同尺度特征的语义重要性实现其动态融合：

$$\pi_L(\boldsymbol{\xi})\cdot\boldsymbol{\xi}=\sigma\left(f\left(\frac{1}{SC}\sum_{S,C}\boldsymbol{\xi}\right)\right)\boldsymbol{\xi} \tag{6}$$

式中：$\pi_L(\cdot)$ 为应用于 L 维度的注意力函数；$\sigma(x)$ 为 Hard-Sigmoid 函数，$\sigma(x)=\max(0,\min(1,(x+1)/2))$；$f(\cdot)$ 为 1×1 卷积层近似的线性函数。

尺度感知注意力机制的结构如图 4 所示。首先采用平均池化降低特征维度，对特征信息进行合理整合与提炼；其次通过 1×1 卷积，实现不同输入通道信息的线性组合，从而完成通道间的信息交互；ReLU 和 Hard-Sigmoid 函数的主要作用是为神经网络引入非线性特性[14]，并解决梯度消失问题，同时提升计算效率和模型的稀疏性。

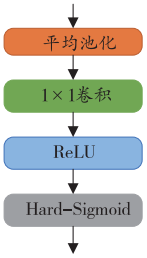


图 4　尺度感知注意力机制的结构

Fig. 4　Structure of scale aware attention mechanism

1.4.2　空间感知注意力机制

先前的研究致力于增强目标检测中的空间感知能力，以便更有效地执行语义识别。卷积神经网络在掌握图像中的空间变换方面存在限制。为克服这一挑战，引入一个基于融合特征的空间感知注意力模块，专注于识别在空间位置和特征层级间持续存在的判别性区域。考虑到特征空间 S 的高维特点，该模块被分为 2 个阶段实施：①通过可变形卷积实现对注意力的稀疏学习；②在相同的空间位置上进行跨层级特征的聚合。

$$\pi_S(\boldsymbol{\xi})\cdot\boldsymbol{\xi}=\frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}\omega_{l,k}\cdot\boldsymbol{\xi}(l;p_k+\Delta p_k;c)\cdot\Delta m_k \tag{7}$$

式中：$\pi_S(\cdot)$ 为应用于 S 维度的注意力函数；K 为稀疏采样位置的数量；$p_k+\Delta p_k$ 为通过自学得到的空间偏移量 Δp_k 进行偏移的位置，以便集中在一个具有区别性的区域上，并且是在位置 p_k 上通过自学习得到的重要性标量，两者都是从中值水平的输入特征中学习的。

空间注意力机制的结构如图 5 所示。首先通过 Index 函数帮助模型更有效地处理序列数据，特别是在处理长序列时，提高模型的性能和泛化能力；其次通过 3×3 的卷积实现对注意力的稀疏学习，Sigmoid 是一个非线性激活函数，用来对输出进行归一化处理，使得注意力权重能够适应不同的输入数据和任务需求；通过偏移量使位置集中在一个具有区别性

的区域上,并且是在位置上通过自学习得到的重要性标量;最终达到在相同的空间位置上进行跨层级特征的聚合。

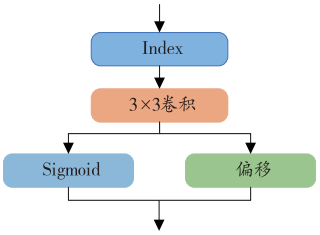


图 5 空间注意力机制的结构

Fig. 5 Structure of spatial attention mechanism

1.4.3 任务感知注意力机制

目标检测的发展历程始于两阶段检测方法,这一方法首先生成候选的目标区域建议,接着对这些建议进行分类,区分成不同的目标类别及背景。REN 等通过提出区域提议网络(RPN)的概念,实现了在一个卷积网络中同时进行目标区域的提议和分类,确立了当代两阶段目标检测框架的标准。随着时间的推移,由于其高效性,单阶段目标检测方法开始广泛采用。在此基础上,LIN 等[15]通过增加专门针对检测任务的分支结构,提升了模型的准确度,使其不仅达到了两阶段检测方法的水平,还保持了单阶段检测器的快速处理速度。

统一注意力模块在检测头中应用的任务感知注意力机制,其允许在通道上分配注意力,能够自适应地支持各种检测任务,无论是盒/中心检测器还是关键点检测器。

为了实现联合学习并推广对不同对象的表示,算法中部署了一种任务感知的注意力机制,可以动态地开启和关闭特征通道,以便适应不同的任务:

$$\pi_C(\boldsymbol{\xi})\cdot\boldsymbol{\xi}=\max(\alpha^1(\boldsymbol{\xi})\cdot\xi_C+\beta^1(\boldsymbol{\xi}),\alpha^2(\boldsymbol{\xi})\cdot\xi_C+\beta^2(\boldsymbol{\xi})) \tag{8}$$

式中:$\pi_C(\cdot)$为应用于 C 维度的注意力函数;ξ_C 代表第 C 个通道的特征切片;$[\alpha^1,\alpha^2,\beta^1,\beta^2]^{\mathrm{T}}=\theta(\cdot)$定义了一个超级函数,旨在学习激活阈值的调控,与动态 ReLU 相似。

任务注意力机制的结构如图 6 所示。首先通过在 $L\times S$ 维度上执行全局平均池化操作以实现降维;随后采用 2 层全连接层和 1 个规范化层进行处理;最终通过一个平移的 S 形函数把输出值规范至[-1,1]内。

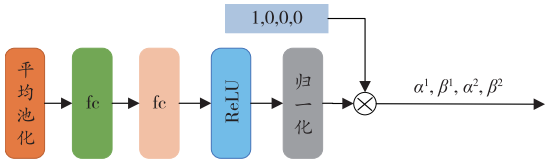


图 6 任务注意力机制的结构

Fig. 6 Structure of task attention mechanism

给定特征张量 $\boldsymbol{\xi}\in R^{L\times S\times C}$,应用自注意力的一般公式如下:

$$W(\boldsymbol{\xi})=\pi(\boldsymbol{\xi})\cdot\boldsymbol{\xi} \tag{9}$$

式中 $\pi(\cdot)$为一个注意力函数,实现其一种简单方法是通过全连接层。

然而,由于涉及的张量维数较高,在所有维度上直接计算注意力函数在计算资源消耗上是极其昂贵的,以至于在实践中几乎不可行。取而代之,统一注意力模块将注意力函数转换为 3 个顺序的注意力,每个注意力只关注 1 个视角,由于以上 3 种注意力机制是顺序应用的,统一注意力模块可以多次嵌套方程(9),以有效地将多个 π_L、π_S 和 π_C 块堆叠在一起:

$$W(\boldsymbol{\xi})=\pi_C(\pi_S(\pi_L(\boldsymbol{\xi})\cdot\boldsymbol{\xi})\cdot\boldsymbol{\xi})\cdot\boldsymbol{\xi} \tag{10}$$

式中:$W(\cdot)$为统一注意力函数;$\pi_C(\cdot)$、$\pi_S(\cdot)$、$\pi_L(\cdot)$分别为应用于维度 C、S、L 的 3 个不同的注意力函数。

统一注意力机制模块的结构如图 7 所示。初始阶段从主干网络得到的特征图含有较多噪声,这是因为其与在 ImageNet 上预训练的模型之间存在领域差异。经过尺度感知注意力模块处理后,特征图对于前景物体的尺度变化更加敏感。随后,空间感知注意力模块进一步提炼特征图,使其更加稀疏并集中于前景物体的关键空间位置。最终,任务感知注意力模块根据不同的下游任务要求调整特征图,形成特定的激活模式。这些视觉展示清晰地验证了各个注意力模块的作用和效果。

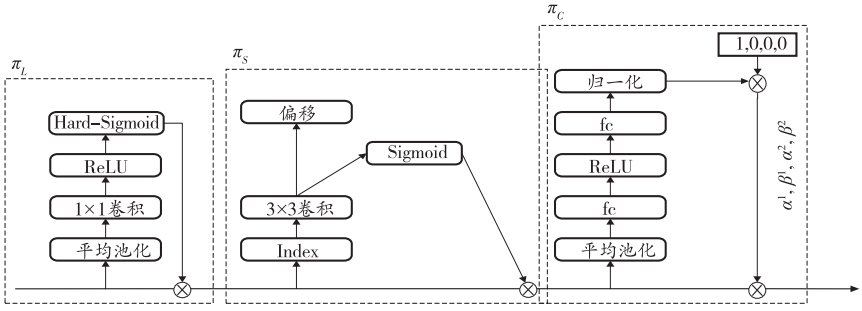


图 7 统一注意力机制模块的结构

Fig. 7 Structure of unified Attention Mechanism

单级检测器通过密集采样特征图来预测目标位置,这简化了检测器的设计。像 RetinaNet 这样的传统单级检测器包含 1 个提取密集特征的骨干网络和多个专为不同任务设计的子网络分支。以往的做法是让负责对象分类的子网络与负责边界框回归的子网络各自独立工作。与此不同的是,统一注意力模块仅将 1 个统一的分支连接到主干上,得益于多重注意力机制,这一分支能同时处理多种任务。这样不仅简化了结构,还提高了效率。近期,无锚点的单级检测器逐渐受到欢迎,例如 FCOS[16]、ATSS[17]和 RepPoint[18]等方法将对象表示为中心点或关键点以提升性能。这些方法需要在分类或回归分支中加入额外的中心点或关键点预测功能,这使构建针对特定任务的分支变得复杂。相较之下,采用统一注意力机制更为灵活,因为其可以在检测头的末端简单地附加不同类型的预测任务。检测头中统一注意力机制的结构如图 8 所示。

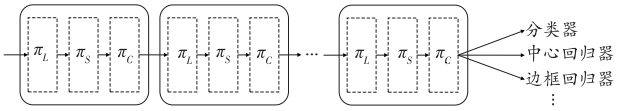


图 8　检测头中统一注意力机制的结构

Fig. 8　Structure of unified attention mechanism in detection head

2　实验结果与分析

在本次研究中,实验所用的计算硬件包括搭载 Intel(R) Core(TM) i9-10980XE 处理器的系统,该处理器频率 3.00 GHz,共有 18 个核心和 36 个线程;内存容量为 64 GB。图形处理单元采用的是 NVIDIA RTX 3090Ti 显卡,配备 24 GB 的视频内存。软件方面,操作系统为 Ubuntu 20.04 版;深度学习框架使用的是 Pytorch 的 1.10 版本;CUDA 11.4 用于加速学习过程;编程工作在 Python 3.9 环境下进行。所使用的模型的学习率初始设置为 0.000 125,并且模型训练迭代进行了 100 次。

2.1　TDConv 对检测效果的影响

为了尽可能减少模型参数量,提高检测精度,本文算法仅在主干网络的 ResNet50 部分将 3×3 的卷积模块换成了 TDConv 模块。为了探究 TDConv 分组的个数对模型检测效果的影响,对 TDConv 不同分组方案进行对比实验:

方案 A:将 TDConv 分为 2 组,并使用 TDConv 模块替换掉 ResNet50 中的部分标准 3×3 卷积;

方案 B:将 TDConv 分为 4 组,并使用 TDConv 模块替换掉 ResNet50 中的部分标准 3×3 卷积;

方案 C:将 TDConv 分为 8 组,并使用 TDConv 模块替换掉 ResNet50 中的部分标准 3×3 卷积;

方案 D:将 TDConv 分为 16 组,并使用 TDConv 模块替换掉 ResNet50 中的部分标准 3×3 卷积。

为了验证 TDConv 不同分组对检测性能和检测速度的影响,对上述配置 TDConv 的 4 种方案在 PASCAL VOC 数据集上进行实验,结果见表 1(表中的计算量是针对 YOLOv5 网络计算的)。

表 1　TDConv 配置实验

Table 1　TDConv configuration experiment

方案类别	平均精度均值 mAP/%	计算量/GFLOPs
Base	90.56	272.45
A	90.89	-20.45
B	91.07	-35.67
C	90.74	-47.24
D	89.98	-54.27

由表 1 可以看出,在同一结构下,方案 A 的计算量相对较大,方案 B 的 mAP 值最高,即其检测效果最好,检测速度也相对理想;方案 C 和 D 虽然在计算量上明显减少,但平均精度和检测速度并不理想。这是由于连续的 1×1 卷积,促进了卷积层之间的信息共享,适当数量的 3×3 卷积可以增强特征提取能力,因此使用方案 B 对 YOLOv5 进行改进。

2.2　统一注意力机制模块实验

为了验证不同模块的添加对统一注意力机制模块检测性能的影响,首先通过将不同注意力模块逐步添加到 YOLOv5 的原始模型中,以对照研究统一注意力机制模块中不同注意力机制的有效性,这里逐一加入的注意力模块选取的是最优效果层数下的模块。消融实验在 PASCAL VOC 数据集上进行,实验结果见表 2。

表 2　注意力机制综合实验结果

Table 2　Attention mechanisms comprehensive experiment

L	S	C	mAP/%
×	×	×	90.56
√	×	×	90.95
×	√	×	91.33
×	×	√	91.02
√	√	×	91.54
√	×	√	91.16
×	√	√	91.58
√	√	√	91.71

注:L、S、C 分别代表尺度感知注意力模块、空间感知注意力模块和任务感知模块;√表示使用该模块;×表示未使用该模块,下同。

由表 2 可以看出,将任一注意力机制模块独立地集成进基础网络架构中,均能实现性能的显著提升,mAP 分别提升了 0.43%、0.85%、0.51%。考虑到空间

感知注意力模块在 3 个增强模块中对模型主导性能维度的贡献最为突出,因此其实现最大幅度的性能增益是在预期之内的结果。同时将 L 和 S、L 和 C、S 和 C 添加到原始模型中时,mAP 性能分别提高了 1.08%、0.66%、1.12%。最后,将统一注意力机制放入检测头中时,相对于原始模型 mAP 提升了 1.27%。上述实验验证了本文算法的注意力机制的有效性。

除此之外,为了探究统一注意力模块数量对检测头的影响,通过控制深度(块的数量)来评估检测头的性能。改变使用的统一注意力模块的数量,分别对 1、2、4、6、8、10 个模块进行对比实验,并将其性能和计算量与原始模型进行比较,结果见表 3。

表 3 统一注意力机制配置实验
Table 3 Unified attention mechanism configuration experiment

模块数量	计算量/GFLOPs	mAP/%
Base	272.45	90.56
1	−84.69	89.24
2	−63.45	90.62
4	−20.97	91.43
6	+21.50	91.71
8	+63.98	91.58
10	+106.46	90.89

由表 3 可见,统一注意力模块可以通过堆叠模块来提高模型性能,在数量达到 6 之前其性能增强都受益于深度的增加。值得注意的是,统一注意力模块为 2 块时,在精度略微提高的基础上,已经以更低的计算量超过了原始模型。同时,即使有 6 个模块,与主干的计算量相比,增加的计算量也可以忽略不计,同时大大提高了精度。进一步证明了统一注意力模块的有效性。

2.3 在矿用数据集上的实验结果

为了验证本文算法在实际场景下的有效性,在自制矿用数据集中进行实验。矿用数据集采集自淮南煤矿选煤厂的带式输送系统环境,共采集 39 849 张图片,分为大块煤矸石图片和锚杆图片 2 类,其中训练验证集图片共 35 863 张,测试集图片为 3 986 张。为了更加直观地评价检测精度,采用平均精度均值和准确率对矿用数据集进行评估。

为了验证统一注意力机制和 TDConv 对检测性能的影响,在矿用数据集上进行了消融实验,结果见表 4。

表 4 YOLO-UAD 自制矿用数据集上的消融实验结果
Table 4 The ablation experiment of YOLO-UAD on a self-made mining dataset

Baseline	统一注意力机制	TDConv	mAP/%
√			81.67
√	√		83.04
√		√	82.14
√	√	√	83.35

由表 4 可以看出,与原始的 YOLOv5 算法相比,加入统一注意力机制可提高模型的检测能力,算法的 mAP 值由 81.67% 提升至 83.04%,提升了 1.68%;加入 TDConv 模块后,算法的 mAP 值进一步提升到 83.35%,提升了 2.06%。自制矿用数据集上的消融实验进一步验证了统一注意力机制和 TDConv 均可提升算法的精度。

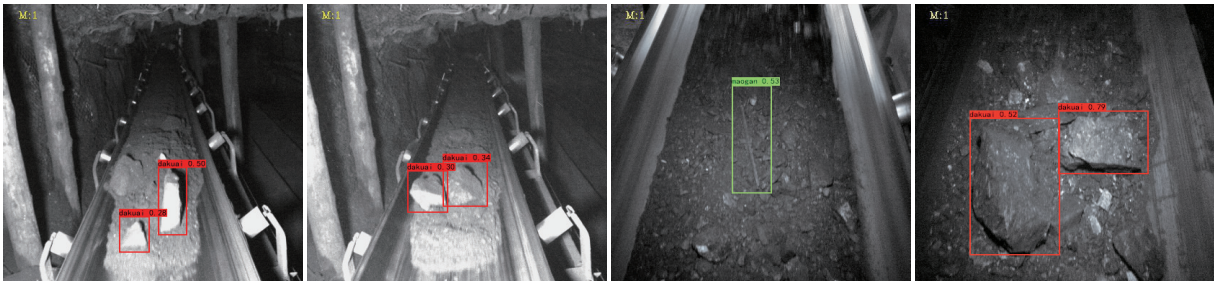
最终为了验证本文算法 YOLO-UAD 的性能,在矿用数据集上与其他主流算法 SSD、Faster R-CNN、RetinaNet、CenterNet、YOLOv3、YOLOv4、YOLOv5x、YOLOx-m,以及 Poly-YOLO 进行对比实验,结果见表 5。

表 5 本文算法与其他算法在矿用数据集上的对比结果
Table 5 Comparison results of YOLO-UAD algorithms on mining datasets

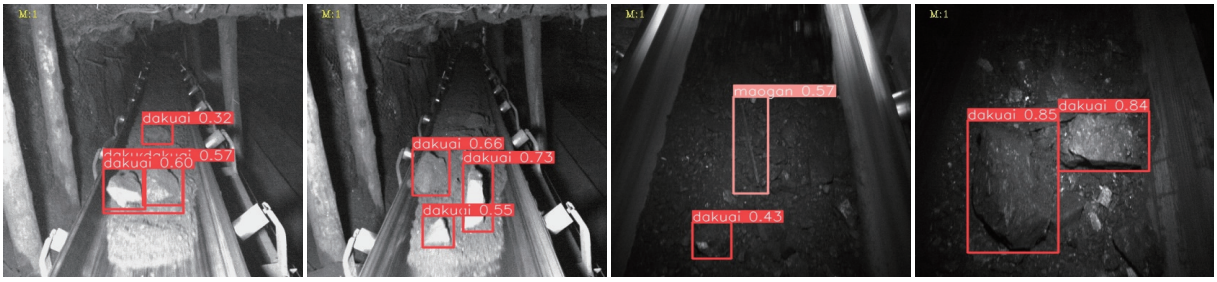
算法类别	图片尺寸/(像素×像素)	大块煤矸石 AP/%	锚杆 AP/%	mAP/%	帧率/(帧·s^{-1})	准确率/%
SSD[19]	416×416	73.56	79.92	76.74	59.70	82.79
Faster R-CNN[20]	416×416	68.15	72.47	70.31	19.40	74.98
RetinaNet[21]	416×416	73.98	78.69	76.33	29.40	86.21
CenterNet[22]	416×416	70.89	74.78	72.84	69.70	78.94
YOLOv3[23]	416×416	75.52	83.45	79.48	31.60	87.07
YOLOv4[24]	416×416	76.26	84.40	80.33	36.40	88.05
YOLOx-m[25]	640×640	73.26	78.37	75.82	43.20	85.36
YOLOv5x	416×416	77.83	85.51	81.67	53.26	89.88
Poly-YOLO[26]	352×608	77.56	85.37	81.47	39.80	89.35
YOLO-UAD	416×416	79.32	87.40	83.35	54.76	91.56

由表 5 可见,在输入图片尺寸为 416 像素×416 像素的情况下,本文算法 YOLO-UAD 大块煤矸石和锚杆的平均精度 AP 分别为 79.32%、87.40%。与相同输入的 YOLOv3、YOLOv4 及 YOLOv5x 算法相比,大块煤矸石和锚杆的精度分别上升了 5.0%、4.0%、1.9%, 4.7%、3.6%、2.2%。同时,本文算法的 mAP 较原始 YOLOv5x 算法提高了 2.1%。矿用数据集上,对比 SSD 和 CenterNet 算法,本文算法在检测速度方面略低于这 2 种算法的情况下,mAP 分别提高了 8.6%、18.5%;在对比针对嵌入式设备而优化的 Poly-YOLO 算法时,本文算法在准确性和检测速度方面都有了明显提升。与其他参照算法相比,本文算法在精度上达到了领先水平,进一步说明了其在井下等特定场景中可被高效应用,且适应性较强。

为了更加直观感受本文算法的有效性,对原始的 YOLOv5 与本文算法的检测结果进行可视化展示和热力图展示,如图 9 所示。



(a) YOLOv5 算法



(b) 本文算法

图 9　YOLOv5 和本文算法检测结果对比

Fig. 9　Comparison of detection results between YOLOv5 and ours algorithms

由图 9 可以看出,由于矿井下环境比较恶劣,当部分目标特征不明显或体型过小时,原始的 YOLOv5 算法会出现检测精度较低和漏检的情况。由图 9 中 a1、a2、a3 和 b1、b2、b3 可知,原始 YOLOv5 算法检测时漏检了图像中的小型煤矸石,而本文算法 YOLO-UAD 采用统一注意力机制解决目标尺度和多样性的问题,提高了检测精度并降低了漏检率。对比 a3、a4 和 b3、b4,在检测过程中,无论是对大块煤矸石还是锚杆的检测精度,本文算法都明显优于原始 YOLOv5 算法。

通过实验对比与效果展示,证明了本文算法可以有效减少模型参数量,提升检测精度,能够有效地对运煤输送带上的异物进行实时目标检测,从而提升了输送带的输送效率,消除了煤矿安全生产隐患。

3　结论

1)基于 TDConv 与统一注意力检测头的异物检测算法,首先以 YOLOv5 网络结构为基础模型框架,设计了 TDConv 卷积模块,利用 1×1 和 3×3 的并行卷积并以特殊的排列方式组合,更有效地保持了图像特征原有信息,帮助更深的卷积层提取有效细节信息;其次在检测头部分加入了统一注意力模块,通过尺度感知注意力机制有效提取和识别不同大小物体的特征,通过空间感知注意力机制更好地提取空间位置之间的特征信息,以及添加任务感知注意力机制能够根据对象引起的各种卷积核响应,指导各个特征通道专注于各自的任务。

2)针对矿业领域数据资源不足的问题,创建了一个自制的矿井异物数据集(MFID),并从初选的 100 000 张数据集中筛选出 39 849 张矿井图像,包括 25 456 大块煤矸石图像和 14 393 张锚杆图像。实验结果表明,所提出的目标检测算法在矿用自制数据集上取得了较好的表现,能够有效地对输送带输送煤炭中的杂质异物进行识别,提高了检测精度和检测速度。同时本文算法减少了原始网络模型的

参数量,在具备高精度的检测能力的同时,使异物检测网络结构更加轻量化,适用于煤矿井下边缘计算设备。

3)本文算法能够增强对矿井异物识别的检测效果,提高对井下输送带的运输效率,保障煤矿生产安全;促进智能视频分析在矿井运输方面的应用,有助于智慧化矿山的建设。

参考文献(References):

[1] 程德强,寇旗旗,江鹤,等.全矿井智能视频分析关键技术综述[J].工矿自动化,2023,49(11):1-21.
CHENG Deqiang, KOU Qiqi, JIANG He, et al. Overview of key technologies for mine-wide intelligent video analysis[J]. Industry and Mine Automation,2023,49(11):1-21.

[2] 程德强,钱建生,郭星歌,等.煤矿安全生产视频AI识别关键技术研究综述[J].煤炭科学技术,2023,51(2):349-365.
CHENG Deqiang, QIAN Jiansheng, GUO Xingge, et al. Review on key technologies of AI recognition for videos in coal mine[J]. Coal Science and Technology,2023,51(2):349-365.

[3] 雷世威,肖兴美,张明.基于改进YOLOv3的煤矸识别方法研究[J].矿业安全与环保,2021,48(3):50-55.
LEI Shiwei, XIAO Xingmei, ZHANG Ming. Research on coal and gangue identification method based on improved YOLOv3[J]. Mining Safety & Environmental Protection,2021,48(3):50-55.

[4] BORNGRABER F. FOREIGN OBJECT DETECTION:14/436 999. US20150288214[P]. 2024-06-09.

[5] 王国法,庞义辉,任怀伟,等.矿山智能化建设的挑战与思考[J].智能矿山,2022,3(10):2-15.

[6] 程德强,徐进洋,寇旗旗,等.融合残差信息轻量级网络的运煤皮带异物分类[J].煤炭学报,2022,47(3):1361-1369.
CHENG Deqiang,XU Jinyang,KOU Qiqi,et al. Lightweight network based on residual information for foreign body classification on coal conveyor belt[J]. Journal of China Coal Society,2022,47(3):1361-1369.

[7] LIN T Y,MAIRE M,BELONGIE S,et al. Microsoft COCO:Common objects in context[M]//Computer Vision - ECCV 2014. Cham:Springer International Publishing,2014:740-755.

[8] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision,2010,88(2):303-338.

[9] GEIGER A,LENZ P,STILLER C,et al. Vision meets robotics:The KITTI dataset[J]. The International Journal of Robotics Research,2013,32(11):1231-1237.

[10] YAN Q, TONG Z, ZHIHUI L I. Improved helmet wear detection algorithm for YOLOv5[J]. Computer Engineering and Applications,2023,59(11):203-211.

[11] SINGH P, VERMA V K, RAI P, et al. HetConv: Beyond homogeneous convolution kernels for deep CNNs[J]. International Journal of Computer Vision,2020,128(8):2068-2088.

[12] LI B,HE Z,YE X,et al. Flow group convolution:Group convolution with channels information interaction[C]//Twelfth International Conference on Graphics and Image Processing (ICGIP 2020). SPIE,2021,11720:721-727.

[13] NIU Z Y,ZHONG G Q,YU H. A review on the attention mechanism of deep learning[J]. Neurocomputing,2021,452:48-62.

[14] 郭永存,张勇,李飞,等.嵌入空洞卷积和批归一化模块的智能煤矸识别算法[J].矿业安全与环保,2022,49(3):45-50.
GUO Yongcun, ZHANG Yong, LI Fei, et al. Intelligent coal and gangue identification algorithm embedded in dilated convolution and batch normalization module[J]. Mining Safety & Environmental Protection,2022,49(3):45-50.

[15] LIN T Y,GOYAL P,GIRSHICK R,et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision,2017:2980-2988.

[16] TIAN Z,SHEN C H,CHEN H,et al. FCOS:A simple and strong anchor-free object detector[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2022,44(4):1922-1933.

[17] BIFFI L J, MITISHITA E, LIESENBERG V, et al. ATSS deep learning-based approach to detect apple fruits[J]. Remote Sensing,2020,13(1):54.

[18] YANG Z,LIU S H,HU H,et al.. Reppoints:Point set representation for object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision,2019:9657-9666.

[19] LIU W,ANGUELOV D,ERHAN D,et al. SSD:Single shot MultiBox detector[M]//LEIBE B, MATAS J, SEBE N, et al. Computer Vision - ECCV 2016. Cham: Springer International Publishing,2016:21-37.

[20] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision,2015:1440-1448.

[21] WANG Y Y,WANG C,ZHANG H,et al. Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery[J]. Remote Sensing,2019,11(5):531.

[22] DUAN K W,BAI S,XIE L X,et al. CenterNet:Keypoint triplets for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul,Korea (South). IEEE,2019:6568-6577.

[23] REDMON J,FARHADI A. YOLOv3:An incremental improvement[EB/OL]. 2018:arXiv:1804.02767. http://arxiv.org/abs/1804.02767.

[24] BOCHKOVSKIY A,WANG C Y,LIAO H Y M. YOLOv4:Optimal speed and accuracy of object detection[EB/OL]. 2020: arXiv: 2004.10934. http://arxiv.org/abs/2004.10934.

[25] GE Z,LIU S T,WANG F,et al. YOLOX:Exceeding YOLO series in 2021[EB/OL]. 2021: arXiv: 2107.08430. http://arxiv.org/abs/2107.08430.

[26] HURTIK P,MOLEK V,HULA J,et al. Poly-YOLO:Higher speed, more precise detection and instance segmentation for YOLOv3[J]. Neural Computing and Applications,2022,34(10):8275-8290.

(责任编辑:熊云威)