



# 煤炭科学技术

煤炭科学研究院 COAL SCIENCE AND TECHNOLOGY

## 融合金字塔结构与注意力机制的煤矿井下巡检机器人PT目标检测算法

王茂森 鲍久圣 鲍周洋 阴妍 王祥赛 葛世荣

引用本文:

王茂森, 鲍久圣, 鲍周洋, 等. 融合金字塔结构与注意力机制的煤矿井下巡检机器人PT目标检测算法[J]. 煤炭科学技术, 2024, 52(6): 206–215.

WANG Maosen, BAO Jiusheng, BAO Zhouyang. Research on mine underground inspection robot target detection algorithm based on pyramid structure and attention mechanism coupling[J]. Coal Science and Technology, 2024, 52(6): 206–215.

在线阅读 View online: <https://doi.org/10.12438/cst.2023-1071>

### 您可能感兴趣的其他文章

#### Articles you may be interested in

##### 煤矿水泵房巡检机器人路径规划与跟踪算法的研究

Research on path planning and tracking algorithm of inspection robot in coal mine water

煤炭科学技术. 2022, 50(5) <http://www.mtkxjs.com.cn/article/id/f7c33730-e477-4166-8fb1-c8f47410a519>

##### 煤矿综采工作面巡检机器人系统研究

Research on inspection robot system for fully-mechanized mining face in coal mine

煤炭科学技术. 2019(10) <http://www.mtkxjs.com.cn/article/id/0988492b-5856-4082-a405-ee00721d312b>

##### 煤矿带式输送机巡检机器人关键技术研究

Research on key technology of inspection robot for coal mine belt conveyor

煤炭科学技术. 2022, 50(5) <http://www.mtkxjs.com.cn/article/id/9cb74dfd-f1b2-4afc-89b2-2a92684bdaa2>

##### 改进YOLOv5s和DeepSORT的井下人员检测及跟踪算法

Underground personnel detection and tracking based on improved YOLOv5s and DeepSORT

煤炭科学技术. 2023, 51(10): 291–301 <https://doi.org/10.13199/j.cnki.cst.2022-1933>

##### 综采工作面自动巡检机器人系统设计

Design of automatic inspection robot system for fully-mechanized coal mining face

煤炭科学技术. 2020, 48(8): 145–149 <http://www.mtkxjs.com.cn/article/id/8b2026b0-3666-4b22-a290-41ab7b075fcb>

##### 综采工作面巡检机器人柔性轨道设计与运动学仿真

Design of flexible track and kinematics simulation of inspection robot in fully-mechanized mining face

煤炭科学技术. 2022, 50(12): 240–246 <https://doi.org/10.13199/j.cnki.cst.2020-0705>



关注微信公众号, 获得更多资讯信息



移动扫码阅读

王茂森, 鲍久圣, 鲍周洋, 等. 融合金字塔结构与注意力机制的煤矿井下巡检机器人PT目标检测算法[J]. 煤炭科学技术, 2024, 52(6): 206–215.

WANG Maosen, BAO Jiusheng, BAO Zhouyang, *et al.* Research on mine underground inspection robot target detection algorithm based on pyramid structure and attention mechanism coupling[J]. Coal Science and Technology, 2024, 52(6): 206–215.

## 融合金字塔结构与注意力机制的煤矿井下巡检机器人PT目标检测算法

王茂森<sup>1</sup>, 鲍久圣<sup>1</sup>, 鲍周洋<sup>1</sup>, 阴妍<sup>1</sup>, 王祥赛<sup>1</sup>, 葛世荣<sup>2</sup>

(1. 中国矿业大学机电工程学院, 江苏徐州 221116; 2. 中国矿业大学(北京)机械与电气工程学院, 北京 100083)

**摘要:**近年来, 煤矿机器人已成为现代煤机装备领域的研究热点, 多数煤矿的主煤流运输系统基本实现了连续化、机械化和自动化, 因此对主运输巷道内的安全监控与巡检效率提出了更高的要求, 而精准的目标检测是实现煤矿井下智能化安全监控的必要保障, 但现有的目标检测算法应用于复杂恶劣的煤矿井下巷道环境, 存在目标检测精度较低的问题。面向井下低照明、环境杂乱的特殊工况检测需求, 制作了井下巷道环境内目标物数据集, 完成数据集标注并展开多维度分析; 提出一种基于金字塔结构与注意力机制融合的PT目标检测算法, 利用注意力机制模块替换金字塔结构中的卷积模块, 在控制特征计算量的同时提高对全局特征的提取能力, 实现目标物局部特征与全局特征融合的提取效果, 提高了图像中目标感兴趣区域特征的表达能力。最后, 面向煤矿井下巡检机器人应用场景, 将提出的PT算法与传统经典的Faster R-CNN、YOLOv4算法进行对比分析。结果表明: 所提出的PT目标检测算法能够有效识别复杂环境下巷道内目标物, 相较于主流的Faster R-CNN、YOLOv4目标检测网络, PT算法有更好的综合识别能力, 识别煤矿人员的准确率分别提升了2.90%和4.30%, 识别井下障碍的准确率分别提升0.20%和4.80%, 识别矿井裂缝的准确率分别提升了4.40%和8.60%, 识别井下设备的准确率分别提升了3.00%和8.70%。因此, PT目标检测算法能够更好地适应井下环境, 目标检测算法较其他算法能够获得更高的准确率与检测速度, 可为井下巷道安控系统建设提供理论依据与技术支持。

**关键词:** 煤矿井下; 巡检机器人; 目标检测; 金字塔结构; 注意力机制

中图分类号: TD52

文献标志码: A

文章编号: 0253-2336(2024)06-0206-10

### Research on mine underground inspection robot target detection algorithm based on pyramid structure and attention mechanism coupling

WANG Maosen<sup>1</sup>, BAO Jiusheng<sup>1</sup>, BAO Zhouyang<sup>1</sup>, YIN Yan<sup>1</sup>, WANG Xiangsai<sup>1</sup>, GE Shirong<sup>2</sup>

(1. School of Mechanical and Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, China; 2. School of Mechanical and Electrical Engineering, China University of Mining and Technology-Beijing, Beijing 100083, China)

**Abstract:** In recent years, coal mine robots have become a research hotspot in the field of modern coal machine equipment, and the main coal flow transportation system of most coal mines has basically realized continuity, mechanization and automation, which also puts forward higher requirements for safety monitoring and inspection efficiency in the main transportation roadway, and accurate target detection is a necessary guarantee for intelligent safety monitoring in coal mines, but the existing object detection algorithm is applied to complex and harsh coal mine underground roadway environment, and there is a problem of low target detection accuracy. Aiming at the special working condition detection requirements of low lighting and chaotic environment in the downhole, the target data set in the underground

收稿日期: 2023-07-23

责任编辑: 周子博

DOI: 10.12438/cst.2023-1071

基金项目: 煤矿采掘机械装备国家工程实验室开放课题资助项目(GCZX-2023-01); 中央高校基本科研业务费专项资金资助项目(2022XSCX27); 江苏高校优势学科建设工程资助项目(PAPD)

作者简介: 王茂森(1994—), 男, 江苏宿迁人, 博士研究生。E-mail: tb20050018b0@cumt.edu.cn

通讯作者: 鲍久圣(1979—), 男, 安徽桐城人, 教授, 博士生导师, 博士。E-mail: cumtbjs@cumt.edu.cn

roadway environment was produced, and the dataset annotation was completed and multi-dimensional analysis was carried out. A PT target detection algorithm based on the fusion of pyramid structure and attention mechanism is proposed, and the attention mechanism module is used to replace the convolution module in the pyramid structure, which improves the extraction ability of global features while controlling the amount of feature calculation, realizes the extraction effect of the fusion of local features and global features of the target, and improves the expression ability of the features of the target area of interest in the image. Finally, for the application scenario of underground inspection robot in coal mine, the proposed PT algorithm is compared with the traditional Faster R-CNN and YOLOv4 algorithms. Compared with the mainstream Faster R-CNN and YOLOv4 target detection networks, the PT algorithm has better comprehensive recognition capabilities, and the accuracy of identifying coal mine personnel is increased by 2.90% and 4.30%, the accuracy of identifying underground obstacles is increased by 0.20% and 4.80%, and the accuracy of identifying mine cracks is increased by 4.40% and 8.60%, respectively. The accuracy rate of identifying downhole equipment was improved by 3.00% and 8.70%, respectively. Therefore, the PT target detection algorithm can better adapt to the underground environment, and the target detection algorithm can obtain higher accuracy and detection speed than other algorithms, which can provide theoretical basis and technical support for the construction of underground roadway security control system.

**Key words:** underground coal mine; inspection robot; target detection; pyramid structure; attention mechanism

## 0 引 言

我国是世界煤炭大国,在我国煤矿 250 万从业人员中,从事井下危险繁重岗位人员占比达 60% 以上,采用煤矿机器人下井代替矿工作业是我国煤矿智能化建设的必由之路<sup>[1-2]</sup>。其中,安控类煤矿机器人的应用不仅可以减轻巡检工人的劳动强度,而且对提高巡检质量,结合监控系统实现煤矿井下无死角、零盲区具有重要意义<sup>[3]</sup>。然而,目前国内安控类煤矿机器人的研究进展相对缓慢<sup>[4]</sup>,其中一个重要的限制因素是缺乏高效准确的井下巷道目标检测技术。

目标检测距今已有 50 多年的研究历史,国内外学者提出了各类理论和算法,目前基于深度学习的目标检测发展较为迅猛<sup>[5]</sup>。2014 年, Ross Girshick<sup>[6]</sup> 提出了区域卷积神经网络 (Regions with CNN features, RCNN), RCNN 首次使用深度学习进行目标检测,通过选框建议方法实现目标位置标定和分类,但因其候选框大量重叠,存在特征提取重复、计算成本高等问题。2015 年,何凯明<sup>[7]</sup> 基于 RCNN 提出了 SPP-Net,该网络对图片只进行一次特征提取,通过金字塔空间池化提取出各个候选框所对应的特征图,再利用 SVM 实现目标分类识别。同年, Ross Girshick<sup>[8]</sup> 在 RCNN 和 SPP-Net 的基础上提出了 Fast-RCNN,创新性的把目标识别与选框回归集成到卷积神经网络中。2016 年, Shaoqing Ren 等<sup>[9]</sup> 提出了 Fater-RCNN 目标检测框架, Faster-RCNN 首次引入了区域建议网络 (Region Proposal Network, RPN),把目标检测整个流程融入到一个网络中。目前, Faster-RCNN 目标检测框架在多个领域中有着广泛应用<sup>[10]</sup>。不同于卷积神经网络,2017 年 VASWANI 等<sup>[11]</sup> 首次提出以多头自注意力机制为核心的 Transformer 网络,为自然语言处理提供全新的网络架构。基于 Trans-

former 算法, DOSOVITSKIY 等<sup>[12]</sup> 于 2020 年提出了 ViT (Vision Transformer) 网络,创造性地将其应用至计算机视觉领域。ViT 在大数据集上的效果优于 CNN,但因全局自注意力运算导致其计算量十分庞大<sup>[13]</sup>。

近几年,基于深度学习的目标检测算法逐步向煤矿井下发展,单阶段目标检测算法性能的大幅提升使得煤矿井下智能化巡检成为可能,其中最典型的是 YOLO 算法<sup>[14]</sup>。崔铁军等<sup>[15]</sup> 在 Keras 框架下将 YOLOv4 算法应用到煤矿井下人员佩戴防尘口罩情况的检测,获得了较好效果。同时众多学者尝试通过改进 YOLO 算法实现对井下人员的实时检测<sup>[16-19]</sup>。针对检测的尺寸存在不平衡现象 YOLOv4 的 3 个尺寸的预测头可以提高多尺寸目标的识别精度。YOLOv4 的边界框回归的损失函数采用 CIOU\_LOSS,综合考虑边界框宽高比的尺度信息,以及预测框是筛选的 NMS 为 DIOU\_NMS,考虑到边界框中心点的位置信息,使得预测框回归的精度更高一些。此外, YOLOv4 通过引入更多的数据增强技术和正则化方法,提高了模型的鲁棒性,能够更好的适应矿井下光亮不稳定、遮挡等复杂工况。由于井下环境复杂,存在光照低且分布不均、设备人员遮挡特征不完整等导致图像质量下降的问题,现有目标检测算法难以满足井下智能巡检的应用需求。

针对以上问题,基于金字塔结构与自注意力机制提出了 PT(Pyramid Transformer, PT) 目标检测算法,通过引入自注意力机制模块,增强对全局特征的提取能力,利用金字塔结构减少大型特征图的计算量,提高特征提取效率。

## 1 PT 目标检测算法原理

介绍基于金字塔结构的 CNN 算法框架和基于自注意力机制的 ViT 算法框架,针对煤矿井下检测

对象的特点,结合金字塔结构与自注意力机制,设计本文 PT 特征提取网络。

### 1.1 常见目标检测算法框架

CNN(Convolutional Neural Network)是深度神经网络在视觉识别领域中的主要代表,该模型包含具有特定接受域的卷积核,能够捕捉良好的视觉特征<sup>[20-21]</sup>。为提供平移等方差,卷积核的权值在整个图像空间中被共享,由图1可知,随着层数的不断提高,CNN在金字塔结构大框架下特征不断减小,目标特征尺寸不断缩减,尤其是小目标更容易丢失信息特征,导致检测性能下降,无法满足实际的应用需求。

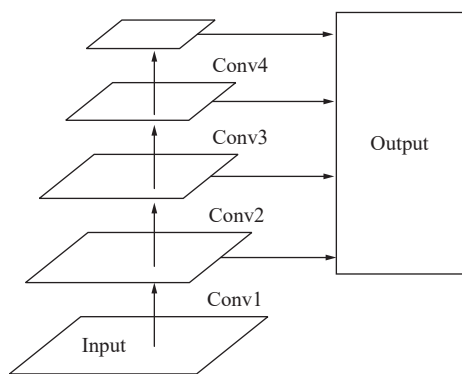


图1 CNN结构框架图

Fig.1 Structural framework of CNN

如图1、图2所示,Conv和TF-E分别代表Convolution和Transformer encoder。许多CNN的主干使用金字塔结构进行密集的预测任务,如对象检测、实例和语义分割。最近提出的ViT是一个专为图像分类设计的柱状结构<sup>[22]</sup>,Transformer编码器作为目标检测任务特定的主干部分,其中的核心部分为自注意力运算模块,虽然一些现有技术也将注意力模块纳入到CNN中,但目前探索一种干净和无卷积的Transformer的主干来解决计算机视觉中的预测分类任务的研究很少。

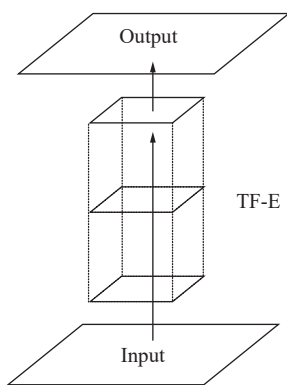


图2 Vision Transformer结构框架

Fig.2 Structural framework of vision transformer

### 1.2 PT特征提取网络设计

目前卷积神经网络在机器视觉领域仍然是研究的主流,但由于煤矿井下环境恶劣,普通神经网络很难直接应用于井下环境。为实现井下场景内多目标高精度巡检目标检测,整合了CNN的金字塔结构和ViT的自注意力机制,提出基于金字塔结构与自注意力机制结合的PT特征提取网络。

首先利用自注意力机制无需卷积的特性,以ViT中的Transformer编码器取代CNN中Backbone的多卷积层,缩减卷积过程,增强模型对图像特征的提取能力。虽然ViT适用于图像分类,但由于计算成本较高,很难将其直接应用于目标检测。为克服ViT高计算成本的限制,参考金字塔结构越深化卷积层越小的特点,将金字塔结构引入PT框架,使需要处理的序列长度变短,大幅减少目标检测过程中的计算成本。以上即为本文基于金字塔结构与注意力机制的PT特征提取网络的由来,其结构框架如图3所示。

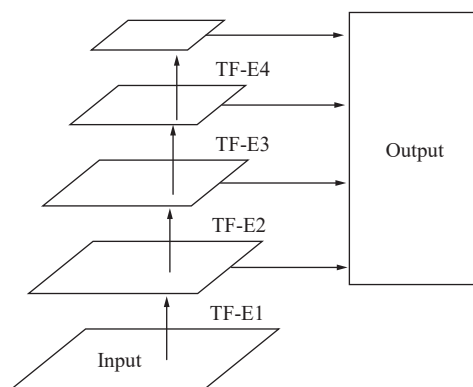


图3 PT特征提取网络结构框架

Fig.3 PT feature extraction network structure framework

不同于卷积神经网络的Backbone,使用不同的卷积获取特征图,使用一种金字塔缓慢收缩的策略,通过嵌入层来控制特征图的尺度。如图4所示,将嵌入层大小设置为 $P_i$ 。在第 $i$ 阶段,将输入的特征值 $Val_i$ 平均划分为 $\frac{H_i W_i}{P_i^2}$ 个值,然后将每个嵌入层进行扩张并投影到 $C_i$ 维嵌入中。经过线性投影后,嵌入的patch的形状为 $\frac{H_{i-1}}{P_i} \times \frac{W_{i-1}}{P_i} \times C_i$ ,其高度和宽度都是 $P_i$ 且比输入小1倍。这样,就构造了PT特征提取网络的金字塔结构。

如图5所示, $Q$ 代表查询向量,后续会和每一个 $K$ 进行匹配; $K$ 代表键值,后续会被每个 $Q$ 匹配; $V$ 代表特征提取信息。与MHA(Multi-head attention)不同,采用的SRA(Spatial-reduction attention)在注意



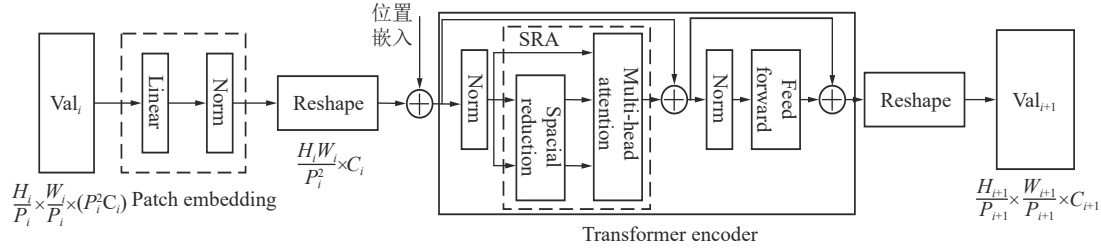


图 4 PT 特征提取模块(PTFEM)示意

Fig.4 Schematic of the PT feature extraction module

力操作前降低了  $K$  和  $V$  的空间尺度, 有利于减少计算量, 可以在有限的资源下处理更大的输入特征映射。

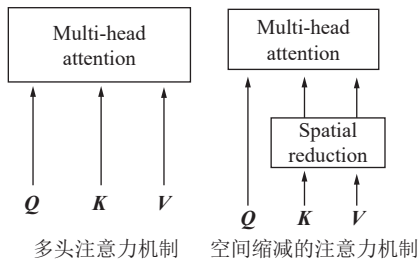


图 5 MHA 与 SRA 注意力机制对比

Fig.5 Comparison of attention mechanism between MHA and SRA

与目前现有的算法相比, PT 特征提取网络有以下优点:

1) 不同于卷积神经网络需要较大量的计算和内

存, PT 可以通过逐步缩小金字塔减少大型特征图的计算量, 提高特征提取的工作效率。

2) 为解决井下机器人检测效率的问题, 通过引入自注意力机制, 增强全局特征提取能力, 减少大量的卷积过程, 提高算法的检测速度。

## 2 PT 目标检测算法框架搭建

如图 6 所示, 优化的目标检测算法 PT 算法可以分为 3 个部分: 构建基于金字塔结构与注意力机制的 PT 特征提取网络, 作为本框架的 Backbone; 构建 FPN (Feature Pyramid Network) 模块进行加强特征提取, 作为本框架的 Neck; 利用 YOLO Head 对结果进行处理预测, 作为本框架的 Head。结构 Backbone 部分主要用于图像特征的提取, 通过结合注意力机制实现检测图像中感兴趣区域的重要特征提取, Neck

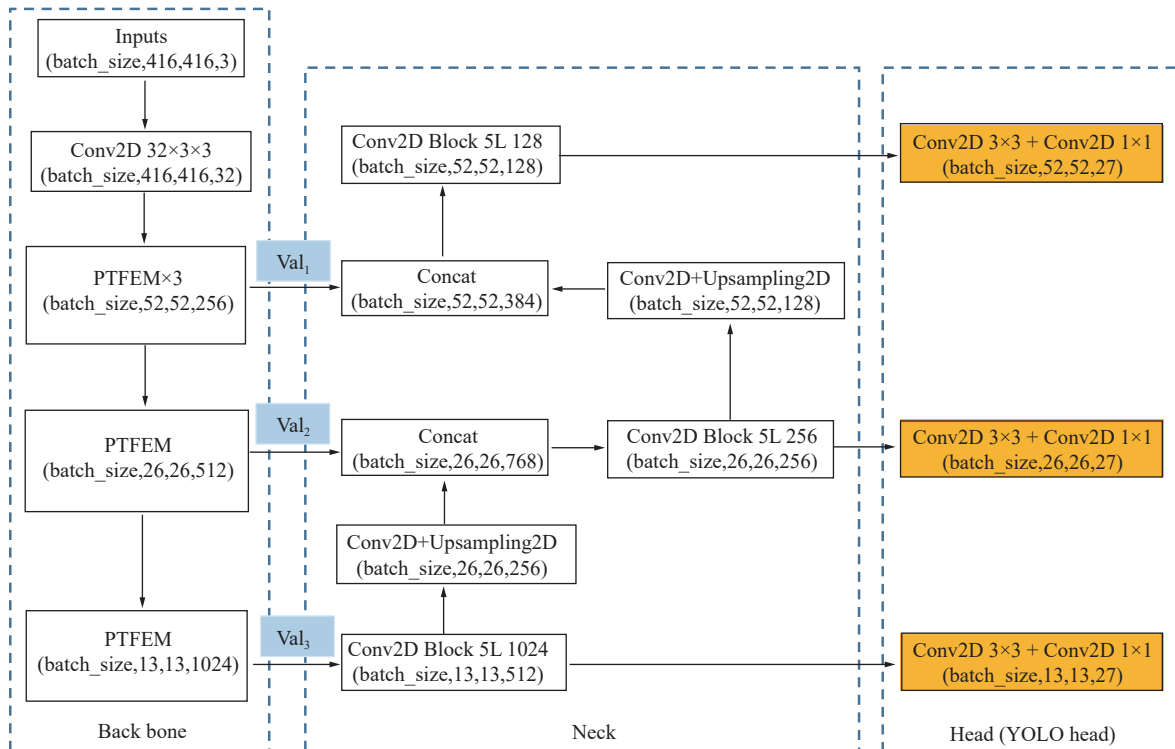


图 6 目标检测 PT 算法结构

Fig.6 Structure of target detection PT algorithm

部分主要用于对特征的处理,以实现更好的特征处理工作,最后 Head 部分用于对提取特征的预测,目标物体由于尺寸存在不平衡,所以采用大、中、小 3 个预测头以更好的实现目标检测。

## 2.1 PT 特征提取网络搭建

第一阶段如图 7 所示,设置大小为  $416 \times 416 \times 3$  的输入图像,首先将其先卷积为一个  $416 \times 416 \times 32$  的特征矩阵,之后将其划分为  $(416 \times 416 / 16)$  个 patch,每个 patch 大小为  $4 \times 4 \times 32$ 。再将扁平的 patch 进行线性投影 Linear,得到大小为  $416 \times 416 \times C1$  的嵌入式 patch。然后,将嵌入的 patch 和位置嵌入通过一层的 Transformer 编码器,将输出重构为大小  $416 \times 416 / 16 \times C1$  的特征映射  $Val_1$ 。同样,使用前一阶段的特征映射作为输入,得到如下特征映射:  $Val_2$ 、 $Val_3$ , 它们相对于输入图像的步长分别为 8、16、32 像素。

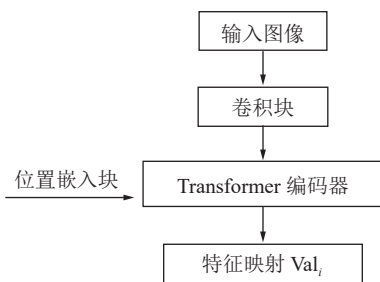


图 7 PT 特征提取网络原理示意

Fig.7 Schematic of PT feature extraction network

## 2.2 FPN 模块构建并预测

在特征提取过程中, FPN 模块围绕金字塔结构提取 3 个特征层进行目标检测。3 个特征金字塔层  $\{Val_1, Val_2, Val_3\}$  的大小分别为  $(52, 52, 256)$ 、 $(26, 26, 512)$ 、 $(13, 13, 1024)$ 。如图 8 所示, 利用这 3 个特征层对 FPN 模块进行构建:

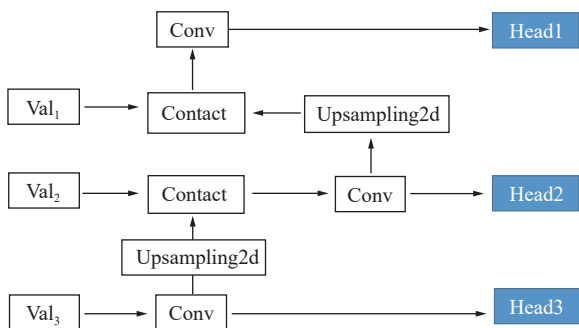


图 8 FPN 原理示意

Fig.8 Schematic of FPN principle

1) 对输出  $Val_3$  特征映射值进行卷积处理, 处理后直接送入 Head 获得预测结果。

2) 结合输出的  $Val_3$  特征映射值进行卷积处理,

一部分用于进行上采样  $UpSampling2d$  后, 与 FPN 特征层进行结合, 处理完后送入 Head 获得预测结果。

3) 结合输出的  $Val_2$  特征映射值再次进行卷积处理, 一部分用于进行上采样  $UpSampling2d$  后, 与 FPN 特征层进行结合, 处理完后送入 Head 获得预测结果。

文中 PT 算法的 FPN 模块对 3 个特征层进行处理后, 输出特征层大小分别为  $(13, 13, 27)$ 、 $(26, 26, 27)$ 、 $(52, 52, 27)$ , 对每一个特征点存在 3 个预选框, 利用 YOLO Head 进行处理预测。

## 3 煤矿井下巷道场景数据集制作

目前鲜有被公开使用的煤矿井下数据集, 而且受限于井下安全章程, 煤矿井下场景不易拍摄。在神东某矿的视频中截取视频帧, 以实现针对井下巡检机器人的图像数据集的制作与训练。

### 3.1 煤矿井下巡检内容分析

主要针对煤矿井下工作人员, 煤矿巷道壁裂缝, 煤矿设备, 废弃障碍物进行巡检检测, 通过图像处理技术对以上对象进行分析判定。此次采集工作共获得 3 000 张图片帧。根据图像内容将巡检机器人在巷道内的行驶状况主要分为以下 3 类:

1) 巷道内的工作人员。在巡检过程中, 需要遵循“机器人让人”的原则, 同时精确识别工作人员与远程定位系统形成定位融合冗余。

2) 巷道内的裂缝。矿井下支护的可靠性关乎矿井工作的安全运行, 在巡检时, 可根据巷道壁的表征对巷道安全性进行判定, 检测识别超出安全范围时, 应立即预警。

3) 巷道内的设备与障碍。对于机器人而言, 当前方出现设备与障碍, 需要规划路径避开行驶。对于巷道中的障碍物, 识别过后上传定位信息提醒工作人员进行清障作业。

### 3.2 数据标注与处理

针对确定的图像进行分类, 分别为 MKper(煤矿人员)、MKobs(井下障碍)、MKcre(煤矿裂缝)、MKdev(井下设备)。数据集采用 Labelimg 进行人工标注, 示例图更为清楚地对数据进行表达, 对目标物使用 Labelimg 进行了人工标注, 部分数据集如图 9 所示。

1) 打开煤矿井下巷道图片;

2) 利用 Labelimg 对场景内的目标物体进行标注;

3) 对一份图片标注完, 再将 .xml 文件打包至统一文件。



图 9 数据集图片例图  
Fig.9 Screenshot of data set

其中 3 000 张图片包含要素见表 1, 为统一标准数据集, 所有图像帧大小均调整为 416×416, 最后将标签文件中数据划分的 train percent 和 test percent 分别设置为 0.8 和 0.2, 然后进行随机划分, 从数据里获得 80% 样本的训练集和 20% 样本的测试集, 以及对应的图片也按照标签文件进行相应的划分, 最后在本地图生文件夹将划分好的数据图片和 txt 格式标签分别进行保存。

表 1 数据集要素统计  
Table 1 Statistics of data set elements

要素	数量/个
MKper(煤矿人员)	1 407
MKobs(井下障碍)	945
MKcre(煤矿裂缝)	1 071
MKdev(井下设备)	1 239

4 目标检测试验与结果分析

在制作煤矿井下巷道场景数据集的基础上, 通过修改巡检机器人的目标检测算法尺度、瞄框数量与置信度对该模型进行试验分析。

4.1 井下巷道目标检测试验

模型训练开始前, 设置初始化动态因子为 0.8, batch 为 4, 以避免陷入最优解。将网络训练的 30 次

迭代的学习率设置为 0.01, 后续学习率设置为  $1 \times 10^{-3}$ 。将权重衰减正则项设置为 0.005, 以防止出现过拟合的现象。激活函数为 ReLU 函数, 最终确定迭代次数为 100 次, 并利用测试集进行测试。为验证本文所提出的 PT 算法在井下环境中的目标检测效果, 将其与 Faster R-CNN, YOLOv4 算法以及其他典型算法进行对比试验分析。目标检测训练的试验配置见表 2。试验采用的 RTX 3090 显卡, NVIDIA CUDA Core 核心数为 10 496, 加速频率为 1.7 GHz, 每秒浮点运算次数(FLOPS)为 35 686.4 GFLOPS。

表 2 目标检测试验基本配置  
Table 2 Basic configuration of target detection test

区域	项目	内容
软件	图像处理库	OpenCV3.4
	开发语言	Python3.7
	编程工具	Pycharm2020
	深度学习框架	Pytorch1.8
硬件	系统	Ubuntu18.04
	处理器	GHz Intel Core i7
	内存	32G
	GPU	NVIDIA GeForce 3090

在目标检测算法评价指标中, 常用  $P$ (准确率)、 $R$ (召回率)和  $mAP$ (均值平均精度)、FPS(每秒帧数)、

准确率 (Accuracy) 来对算法性能进行评价。准确率表示是在所有结果中, 预测结果与真实情况一致所占的比例, 表达式为:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

式中: TP: 预测为正, 实际为正; TN: 预测为负, 实际为负; FP: 预测为正, 实际为负; FN: 预测为负, 实际为正。

4.2 试验结果分析

对比 Faster R-CNN、YOLOv4 以及文中 PT 三种算法经 100 次迭代训练进行目标检测预测, 训练过程中, 记录算法迭代时的损失值, 如图 10 所示。文中 PT 算法结合金字塔结构与注意力机制, 减少了大量卷积, 初始的损失值介于 Faster-R-CNN 与 YOLO 之间, 在经过 20 次迭代后, 损失值低于二者, 最终稳定在 0.01 左右, 达到了较佳的 loss 效果。

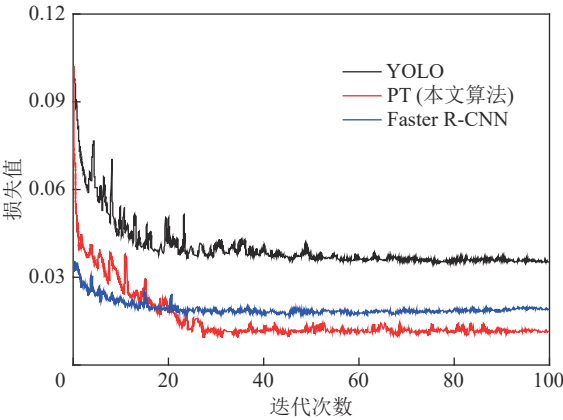


图 10 模型损失变化对比  
Fig.10 Comparison of model loss changes

由图 11 可以看出, YOLOv4 在迭代到 25 次左右准确率上升到 0.6, 准确率最终稳定在 0.81 左右, Faster-R-CNN 与 PT 算法在迭代 15 次左右达到 0.6 的准确率, Faster-R-CNN 准确率在迭代 20 次之后逐渐达到 0.85 左右。其中, PT 算法的准确率高出其他两种算法, 最终达到 0.88 左右。

对比 Faster R-CNN、YOLOv4 以及本文 PT3 种算法部分检测效果如图 12 所示, 3 种算法在本次目标检测中的平均准确率结果见表 3。

对比 3 种算法的检测数据: 煤矿人员准确率, PT 较 Faster R-CNN 提升 2.90%, 较 YOLOv4 提升了 4.30%; 井下障碍准确率, PT 较 Faster R-CNN 提升 0.20%, 较 YOLOv4 提升了 4.80%; 煤矿裂缝准确率, PT 较 Faster R-CNN 提升 4.40%, 较 YOLOv4 提升了 8.60%; 井下设备准确率, PT 较 Faster R-CNN 提升 3.00%, 较 YOLOv4 提升了 8.70%。所以,

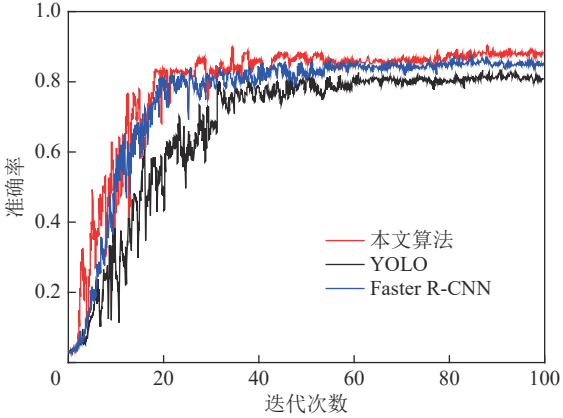


图 11 模型准确率变化对比  
Fig.11 Comparison of model accuracy change

表 3 模型平均准确率对比  
Table 3 Comparison of average accuracy of models

训练网络	模型准确率/%		
	Faster R-CNN	YOLOv4	本文PT算法
MKperson (煤矿人员)	88.20	86.80	91.10
MKobstacle (井下障碍)	85.70	81.10	85.90
MKcrevice (煤矿裂缝)	84.40	80.20	88.80
MKdevice (井下设备)	82.60	76.90	85.60

提出的改进算法 PT 针对煤矿井下环境目标检测具备较强的适应性。

4.3 结果讨论

论文对 Faster R-CNN、YOLOv4、VFNet、HVR-Net、Trans VOD 模型的  $P$ 、 $R$  和  $mAP$  数据展开了对比分析, 并增加对推理速度每秒帧数(FPS)的性能测试。检测结果见表 4, 结果表明: 提出的 PT 检测算法, 在检测精度和实时性两方面兼备较好的检测性能, 能够满足井下监测的实际应用需求, 并极大程度降低时间成本。

对于网络结构设计计算量控制部分, 为克服传统算法高计算成本的限制, 将金字塔结构引入 PT 框架, 使需要处理的序列长度变短, 大幅减少目标检测过程中的计算成本。此外, 如图 5 所示, 采用的 SRA 在注意力操作前降低了  $K$  和  $V$  的空间尺度, 有利于减少计算量。 $Q$  代表 query, 后续会去和每一个  $K$  进行匹配;  $K$  代表 key, 后续会被每个  $Q$  匹配;  $V$  代表特征提取信息。

与 MHA(Multi-head attention) 相似, 采用一款 SRA(Spatial-reduction attention) 接收查询  $Q$ , 键  $K$  和值  $V$  作为输入, 并且输出一个重构特征。与 MHA 不同的是, 在注意力操作之前, SRA 降低了  $K$  和  $V$





图 12 目标检测算法井下情况效果

Fig.12 Effect of target detection algorithm underground

表 4 不同网络模型检测的试验结果

Table 4 Test results of different network model detection

网络结构	Mkper			Mkobs			Mkdev			Mkere			mAP/%	FPS
	P	R	mAP	P	R	mAP	P	R	mAP	P	R	mAP		
Faster R-CNN	100	33.33	77.61	84.21	50	74.5	91.89	45.95	63.99	100	40.74	56.43	68.6	15.3
YOLOv4	93.33	70	84.55	88.89	76.19	82.33	84.71	75	80.12	83.87	70.27	74.96	80.3	18.0
VFNet	86.67	78	85.85	82.14	79.31	81.59	90	72.97	73.17	87.88	65.91	70.62	63.4	24.3
HVRNet	92	85.19	84.72	91.3	77.78	83.75	87.5	54.9	72.01	86.27	59.46	68.59	74.8	18.4
TransVOD	83.33	75	88.03	95.24	74.07	84.55	94.59	68.63	76.03	86.67	61.9	71.61	79.5	22.3
PT(本文)	85	89.47	95.49	91.3	84	90.46	95.19	85.19	88.25	91.67	81.48	86.15	89.85	21.0

的空间尺度,这大幅减少了计算。SRA 的具体情况如下:

$$\text{SRA}(Q, K, V) = \text{Concat}(\text{head}_0, \dots, \text{head}_i)W^o \quad (2)$$

$$\text{head}_j = \text{Attention}(QW_j^Q, SR(K)W_j^K, SR(V)W_j^V) \quad (3)$$

其中:  $W_j^Q$ ,  $W_j^K$ ,  $W_j^V$ , 以及  $W^o$  为线性投影参数;  $N_i$  为  $\text{Transformer}_i$  注意力层的 head 的数量。因此, 每个 head 的尺寸等于  $C_i/N_i$ 。SR() 为降低输入序列空间维数的操作。

$$\text{SR}(x) = \text{Norm}(\text{Reshape}(x, R_i)W^s) \quad (4)$$

式中:  $x$  为输入序列;  $R_i$  在 Transformer 中减少了注意

力层效率;  $\text{Reshape}()$  为变形公式, 从序列  $x$  到  $W^s$  的序列;  $W^s$  为一个线性投影, 它将输入序列的维数降至  $C_i$ 。Norm() 是将一个序列规范化的过程; 与最初的 Transformer 一样, Attention() 可以计算为:

$$\text{Attention}(q, k, v) = \text{Softmax}\left(\frac{qk^T}{\sqrt{d_{\text{head}}}}\right)v \quad (5)$$

通过这些公式可以发现, 注意力所操作的计算过程  $R_i^2$  时间比 MHA 低, 因此 SRA 可以降低参数数量和显存消耗。对模型的参数量进行了计算, 在使用 SRA 模块后, PT 算法没有 SRA 和 PT 算法的参数量分别为 62.7 M 和 53.4 M, PT 模型算法参数明显下降。

综上,结合金字塔结构与自注意力机制的主干网络的PT算法较其他典型算法在性能上获得较大提升,可以满足煤矿井下环境目标检测的需求。

## 5 结 论

1)构建井下巷道环境内目标物数据集,完成数据集标注并展开多维度分析,并实现了数据集的训练集与测试集合理划分构建。

2)提出一种基于金字塔结构与注意力机制的PT目标检测算法,将输入特征嵌入层进行扩展并逐层降维映射,改善对全局特征的提取能力,实现目标物局部特征与全局特征融合的提取效果,提高了目标物特征的表达能力及检测精度。

3)展开PT算法与Faster-R-CNN、VFNet和YOLOv4等算法对比试验,结果表明,所提出的PT目标检测算法在目标检测性能方面具备较大优势,更加适用于煤矿井下检测。

4)在接下来的研究过程中,为更好地提高检测效率,在检测的基础上进一步轻量化网络模型以提高检测速度,并且融合其他传感器提高机器人巡检的可靠性,进一步提高井下巡检机器人感知信息准确率。

## 参考文献(References):

- [1] 葛世荣,胡而已,裴文良.煤矿机器人体系及关键技术[J].煤炭学报,2020,45(1):455-463.  
GE Shirong, HU Eryi, PEI Wenliang. Classification system and key technology of coal mine robot[J]. Journal of China Coal Society, 2020, 45(1): 455-463.
- [2] 鲍久圣,章全利,葛世荣,等.煤矿井下无人化辅助运输系统关键基础研究及应用实践[J].煤炭学报,2023,48(2):1085-1098.  
BAO Jiusheng, ZHANG Quanli, GE Shirong, et al. Basic research and application practice of unmanned auxiliary transportation system in coal mine[J]. Journal of China Coal Society, 2023, 48(2): 1085-1098.
- [3] 王峰.8部委:推动智能化技术与煤炭产业融合发展[J].中国设备工程,2020(7):1.  
WANG Feng. 8 ministries and commissions: promote the integration and development of intelligent technology and coal industry[J]. China Plant Engineering, 2020(7): 1.
- [4] 鲍久圣,张牧野,葛世荣,等.基于改进A\*和人工势场算法的无轨胶轮车井下无人驾驶路径规划[J].煤炭学报,2022,47(3):1347-1360.  
BAO Jiusheng, ZHANG Muye, GE Shirong, et al. Underground driverless path planning of trackless rubber tyred vehicle based on improved A\* and artificial potential field algorithm[J]. Journal of China Coal Society, 2022, 47(3): 1347-1360.
- [5] 鲍久圣,刘琴,葛世荣,等.矿山运输装备智能化技术研究现状及发展趋势[J].智能矿山,2020,1(1):78-88.  
BAO Jiusheng, LIU Qin, GE Shirong, et al. Research status and development trend of intelligent technologies for mine transportation equipment[J]. Journal of Intelligent Mine, 2020, 1(1): 78-88.
- [6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. IEEE, 2014: 580-587.
- [7] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE, 2017: 2980-2988.
- [8] GIRSHICK, R. Fast R-CNN[J]. Computer Science, 2015, 14(1): 580-587.
- [9] 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.
- [10] WANG X L, SHRIVASTAVA A, GUPTA A. A-fast-RCNN: hard positive generation via adversary for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017: 3039-3048.
- [11] VASWANI A, SHAZEER N M, PARMAR N, et al. Attention is all you need[C]//NIP'17: Proceeding of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, 2017:6000-6010.
- [12] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[J]. arXiv, 2010.11929
- [13] 刘学思,聂瑞,张和华,等.基于Swin Transformer网络的肺结核影像自动分类效果评价[J].中国医疗设备,2022,37(8):25-31,65.  
LIU Xuesi, NIE Rui, ZHANG Hehua, et al. Effect evaluation of automatic classification of pulmonary tuberculosis images based on swin transformer network[J]. China Medical Devices, 2022, 37(8): 25-31, 65.
- [14] 李正明,章金龙.基于深度学习的抓取目标姿态检测与定位[J].信息与控制,2020,49(2):147-153.  
LI Zhengming, ZHANG Jinlong. Detection and positioning of grab target based on deep learning[J]. Information and Control, 2020, 49(2): 147-153.
- [15] 崔铁军,王凌霄.YOLOv4目标检测算法在煤矿工人口罩佩戴监测工作中的应用研究[J].中国安全生产科学技术,2021,17(10):66-71.  
CUI Tiejun, WANG Lingxiao. Research on application of YOLOv4 object detection algorithm in monitoring on masks wearing of coal miners[J]. Journal of Safety Science and Technology, 2021, 17(10): 66-71.
- [16] 鲍周洋.煤矿井下巷道全方位巡检轮式机器人设计与研究[D].徐州:中国矿业大学,2023.  
BAO Zhouyang. Design and research of wheeled robot for omnidirectional inspection of underground roadway in coal mine[D]. Xuzhou: China University of Mining and Technology, 2023.

- [ 17 ] 汝洪芳,王珂硕,王国新.改进 YOLOv4 网络的煤矿井下行人检测算法[J].[黑龙江科技大学学报](#),2022,32(4):557-562.  
RU Hongfang, WANG Keshuo, WANG Guoxin. Coal mine pedestrian detection algorithm based on improved YOLOv4 network[J]. [Journal of Heilongjiang University of Science and Technology](#), 2022, 32(4): 557-562.
- [ 18 ] 张明臻.基于 Dense-YOLO 网络的井下行人检测模型[J].[工矿自动化](#),2022,48(3):86-90.  
ZHANG Mingzhen. Underground pedestrian detection model based on Dense-YOLO network[J]. [Industry and Mine Automation](#), 2022, 48(3): 86-90.
- [ 19 ] 付燕,窦晓熠,叶鸥.基于 YOLO 的井下人员速度测量方法研究[J].[煤炭工程](#),2022,54(2):160-165.  
FU Yan, DOU Xiaoyi, YE Ou. A measurement method of underground workers speed based on YOLO[J]. [Coal Engineering](#), 2022, 54(2): 160-165.
- [ 20 ] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. [Nature](#), 2015, 521(7553): 436-444.
- [ 21 ] 叶鸥,窦晓熠,付燕,等.融合轻量级网络和双重注意力机制的煤块检测方法[J].[工矿自动化](#),2021,47(12):75-80.  
YE Ou, DOU Xiaoyi, FU Yan, *et al.* Coal block detection method integrating lightweight network and dual attention mechanism[J]. [Industry and Mine Automation](#), 2021, 47(12): 75-80.
- [ 22 ] ALEXEY Dosovitskiy, LUCAS Beyer, ALEXANDER Kolesnikov, *et al.* An image is worth 16 x 16 words: transformers for image recognition at scale[J]. [Computer Vision and Pattern Recognition](#), 2020, 2010.11929.