

基于改进 YOLOv5 的安全绳目标检测

王 猛, 高树静, 张俊虎, 李海涛

(青岛科技大学 信息科学与技术学院, 山东 青岛 266061)

摘要: 在工业施工过程中, 工人安全已成为一个日益重要的问题, 佩戴安全绳等安全装备是保护工人在高处工作时生命安全的重要措施; 在现代生产施工过程中, 通过使用监控摄像设备结合人工智能算法的方式来检测工人佩戴安全绳等设备越发普遍, 但安全绳由于细长、形状多变以及环境变化等因素较难以准确识别; 为解决以上问题, 并确保能够在不同环境下准确识别安全绳, 现提出一种使用 YOLOv5 的目标检测算法, 首先通过改进的 FasterNet 模块进行上下文信息提取, 在 Neck 网络中使用改进的多维动态卷积保留更多特征信息, 使用 WIoU_Loss 损失函数来提高定位精度, 在训练过程中使用动态调整学习率的策略; 实验结果表明, 改进后的算法在降低计算复杂度的情况下提高了 3.0% 的检测精度, $mAP@0.5$ 提高了 4.3%, 经过在实际场景应用, 满足项目对实时检测精度及速度的要求。

关键词: 安全绳目标检测; YOLOv5; FasterNet; 多维动态卷积; WIoU_Loss

Safe Rope Target Detection Based on Improved YOLOv5

WANG Meng, GAO Shujing, ZHANG Junhu, LI Haitao

(College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: In the process of industrial construction, the safety of workers has become an increasingly important issue, wearing safety equipment such as safety rope is an important guarantee for the safety of workers during high-altitude operations. In the process of modern production and construction, surveillance camera equipment are widely used to detect workers wearing safety ropes and other equipment combined with artificial intelligence algorithm, but it is difficult for the safety rope to accurately identify due to factors such as slender, changeable shape and environmental changes. In order to solve the above problems and ensure that the safety rope can be accurately identified in different environments, an object detection algorithm based on YOLOv5 is proposed. Firstly, the improved FasterNet module is used to extract the context information, and the improved multidimensional dynamic convolution is used to preserve the more feature information in the Neck network. The WIoU_Loss loss function is used to improve the positioning accuracy, and dynamically adjust the learning rate in the training process. Experimental results show that under reducing the computational complexity, the improved algorithm improves the detection accuracy by 3.0%, and $mAP@0.5$ by 4.3%. By the application in actual scenarios, the proposed algorithm can meet the requirements of real-time detection accuracy and speed in the project.

Keywords: safety rope target detection; YOLOv5; FasterNet; multidimensional dynamic convolution; WIoU_Loss

0 引言

2022 年上半年, 全国共发生各类生产安全事故 11 076 起, 死亡 8 870 人, 在众多工程安全事故中, 高处坠落为主要事故类型, 据高处坠落事故原因统计, 安全防护装备缺失(未戴安全帽、未系安全带)是事故发生的主要原因^[1]。在实际的装卸作业时, 工人需要在油罐车顶部进行货物的装卸工作, 安全绳可以为工人提供额外的保护, 防止意外坠落和受伤。如果没有安全绳, 工人可能会因为失去平衡、工具掉落或其他原因而坠落。因此, 通过摄像头结合深度学习目标检测算法对作业现场实时监测管理是更加高效而且必要的方式。目标检测是计算机视觉中的一项重要任务, 它的目的是识别图像中的物体并确定它们在图像中的位置。

早期的目标检测方法主要基于手工特征提取和分类器的组合, 例如 Haar Cascades 和 HOG 等方法。这些方法需要大量的人工设计和调整, 并且对复杂场景的适应性差。近年来, 随着深度学习的兴起, 越来越多的研究者开始使用神经网络进行目标检测, 基于深度学习的方法在准确率和速度方面都取得了显著的提高, 成为当前目标检测领域的主流方法。基于深度学习的目标检测方法可以分为单阶段和双阶段方法。

双阶段方法(如 R-CNN 系列)先使用区域提取算法(例如 Selective Search)生成候选框, 然后对每个候选框进行分类和回归。自 2014 年 Girshick 等人^[2]提出 R-CNN 算法后, 双阶段目标检测算法的研究更加迅速, 各种衍生算法不断产出。Fast R-CNN^[3]、Faster R-CNN^[4]、Mask R-

收稿日期: 2023-06-16; 修回日期: 2023-07-20。

基金项目: 山东省重点研发计划(2021SFRC0701); 青岛市海洋科技创新专项(22-3-3-hygg-3-hy)。

作者简介: 王 猛(1998-), 男, 硕士研究生。

通讯作者: 高树静(1976-), 女, 博士, 副教授, 硕士生导师。

引用格式: 王 猛, 高树静, 张俊虎, 等. 基于改进 YOLOv5 的安全绳目标检测[J]. 计算机测量与控制, 2024, 32(6): 42-50.

CNN^[5]等, Faster R-CNN 引入了 RPN 网络, 把候选框任务从 R-CNN 中分离出来, 使得网络变成两部分, RPN 网络用于候选框的生成, 分类网络用于候选框的分类和回归, Faster R-CNN 通过统一的损失函数同时训练 RPN 网络和分类网络, 相较于 R-CNN 有更快的速度, 更高的检测精度, 降低了候选框的错误率, 是目前目标检测领域的主流算法之一, 张彦凯等人^[6]在 Faster R-CNN 的基础上采用 K-Means++ 聚类优化算法调高对安全帽等小目标的检测能力。Mask R-CNN 在 Faster R-CNN 的基础上添加了掩码任务, 通过添加一个分支利用 ROI 池化层提取特征, 并输出一个二元掩码来确定物体是否在某个像素位置上, 使得模型能够对感兴趣的物体进行更精细的分割和定位。

单阶段目标检测方法采用回归分析的思想, 将区域提取和分类回归合并在一起, 从而实现端到端的目标检测。它只需要一次向前计算就可以完成目标检测任务, 而且在处理小物体的时候表现更加出色。Liu 等人^[7]提出 SSD 算法, 采用多尺度特征图来检测不同大小的物体, 同时进行分类和边界框回归, 提高检测效率。Redmon 等人^[8]在 2016 年提出 YOLO 算法, YOLO 系列算法更加快速高效, 可以在一张图像中同时检测出多个目标, 并且速度非常快, 适用于实时检测的领域, YOLO 使用卷积神经网络, 很容易扩展到新的目标类别或者新的场景中, 目前该系列算法已经更新到 YOLOv8 版本, 很多特性得到了改进和增强, 比如: YOLOv3^[9]引入 FPN 和 PAN 结构对不同尺度物体具有更好的适应性, 采用更深的 DarkNet-53 网络结构; YOLOv4^[10]基于 CSPNet 改进 DarkNet, 大幅减少参数量和计算量, 使用 SPP 代替原来池化层, 提高感受野; YOLOv5^[11]采用 P5 作为下采样层, 将预处理图片分辨率提高至 640×640 , 通过自动混合精度训练技术, 将精度由 float32 提升到 float16, 使得模型在没有降低精度的情况下获得更快的训练速度等。目前 YOLO 算法已经成为实时检测领域中最常用的算法。在实际应用领域, 农元君等人^[12]以 Tiny-YOLOv3 检测网络为基础, 改进特征提取网络和多尺度预测优化网络结构, 引入空间金字塔池化模块改进对小目标的特征提取, 以提高安全帽检测精度; 郑楚伟等人^[13]融合 Swin Transformer 模块, 使用 K-means++ 聚类算法重新设计先验锚框尺寸, 提出新型跨尺度特征融合模块对安全帽佩戴进行检测, 更好地适应了不同尺度的目标检测任务; 顿伟超等人^[14]基于 YOLOv4, 采用 MobileNetv2^[15]作为主干网络, 改进空间金字塔池化结构, 提高对电力高空作业安全带的检测精度; 田坤等人^[16]基于 YOLO 和极限学习机相结合, 快速定位自驾区域并识别安全带, 检测速度大幅提升, 准确率更高。于硕等人^[17]使用改进后的损失函数并使用 DIoU-NMS 来替代原有 NMS 来检测口罩佩戴情况, 提高了检测的算法精度; 王彦生等人^[18]使用多头注意力机制替换 C3 模块提高算法学习能力, 关注图像全局, 添加 ECA 至于伊利机制提高特征提取能力, 有效地提高了对电工厂吸烟行为的检测能力。目标检

测技术已经广泛应用于各种领域, 包括自动驾驶、安防监控、工业检测等。未来随着技术不断发展, 目标检测将会在更多领域得到应用, 例如医学影像、智能家居等。

本论文是基于化工厂装卸作业进行研究, 实际作业过程中使用的是防爆摄像头, 防爆摄像头相较于普通摄像头, 清晰度会有一些的模糊, 而且背景比较复杂, 在夜间作业过程中, 安全绳由于细长等特点, 极易与背景物体重合难以区分, 因此, 本文通过使用 FasterNet 和多维动态卷积等方法来提升模型特征提取能力, 提高模型检测准确率, 降低模型计算量和参数量, 并保证算法实时检测效率。

1 算法原理介绍

YOLOv5 由 Ultralytics 公司于 2020 年推出, 有 5 种不同的网络深度和宽度的模型, 本文所使用的为 6.1 版本的 YOLOv5s 模型, 其模型网络结构如图 1 所示, 主要分为输入端、主干网络 backbone、Neck 网络、Head 共 4 部分。

在输入端 YOLOv5 采用 Mosaic 方式, 对输入图片进行随机裁剪、缩放、拼接进行数据增强, 这种方式对检测小目标物体效果显著, 采用自适应锚框, 对不同的数据集设定初始长宽锚框, 训练时计算相应的最佳锚框值。针对长宽不同的输入图片, 使用自适应缩放方法把输入图片缩放到标准尺寸, 再送入检测网络, 这种改进方法使得推理速度大幅提升。

在主干网络中采用 CSPDarkNet53 轻量级卷积神经网络结构, 引入 CSP 结构, 基于深度可分离卷积和通道分割思想, 把输入特征图分为两个部分, 将一部分进行卷积操作, 另一部分直接输出, 有效地减少了模型参数量和计算量, 实现了高效的特征提取和模型压缩。在 6.0 版本后用一个 6×6 大小的卷积层替换了原来的 Focus 模块, 具有比原来更高效的性能。使用 SPPF 替换原来的 SPP 结构, 在原有基础上加入 BN 层和 Relu 激活函数, 引入了跨尺度特征融合机制, 相较于之前有更出色的特征提取能力和计算效率。

YOLOv5 的 Neck 网络结构为 FPN+PAN, 实现了对不同尺度的特征图进行有效地融合和处理, 提高模型性能和泛化能力。FPN 使用 top-down 和 bottom-up 结构实现特征图的金字塔融合, 以适应不同大小的目标。PAN 则进一步提高了模型性能, 通过池化、卷积等操作将不同尺度的特征图融合成一个统一的特征图, 同时保留更多细节信息并提高计算效率。

YOLOv5 的 Head 网络由多层卷积和全连接层组成, 将 Neck 网络输出得到的特征图提取目标框的位置和分类信息, 使用预定义的 anchor box 生成一组预测框, 包括物体类别、bounding box 位置和大小。其 head 网络采用 yolact++ 设计, 使用自注意力机制和卷积操作进行实例分割。YOLOv5 使用 CIoU 损失函数, 能有效提高对遮挡目标的检测能力。

2 算法改进

2.1 FasterNet 使网络更加轻量化

在深度学习领域, 一个拥有低延迟高吞吐量的快速化、

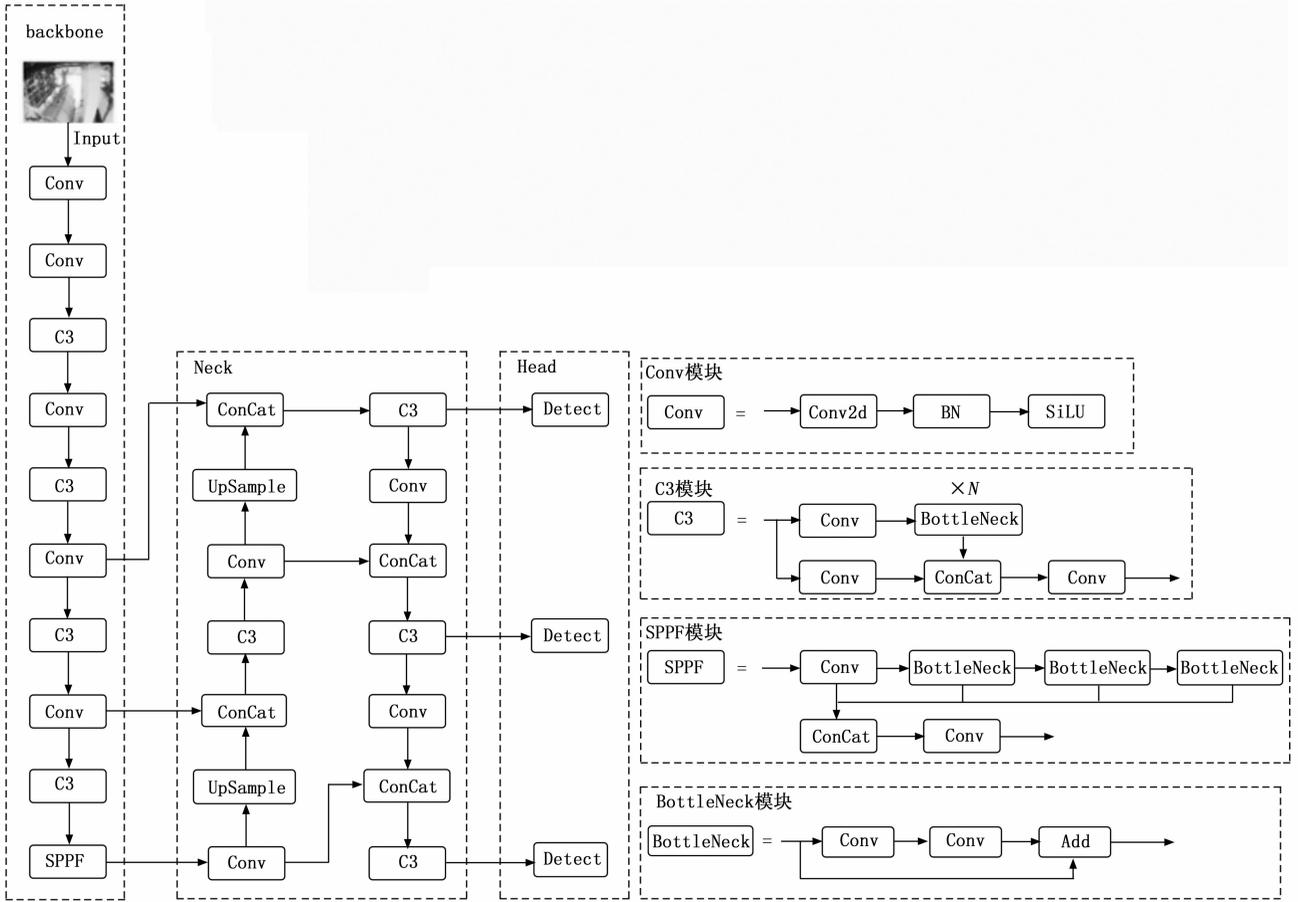


图 1 YOLOv5 网络结构

轻量化神经网络一直是研究者在寻求的方向，延迟跟计算量的关系可以表示为：

$$Latency = \frac{FLOPs}{FPS} \quad (1)$$

其中： $FLOPs$ 表示每秒浮点运算次数，用于衡量设备的计算速度和能力， $FLOPs$ 表示浮点运算次数，用于衡量模型的计算复杂度。为了减少网络的延迟，只追求减少 $FLOPs$ 不一定会使延迟降低，需要在降低 $FLOPs$ 的情况下保持较高的 $FLOPs$ ，在 YOLOv5 的主干网络的 C3 模块中使用了深度可分离卷积 DWConv 和 PWConv 结合的结构，这个结构可以减少计算量和参数数量，提高模型的准确性，但是 DWConv 不能简单代替标准卷积，否则会导致精度下降，为了补偿精度需要增加网络宽度，这就会产生更高的内存访问频率，导致较高的延迟并降低整体计算速度。

FasterNet 块的基本结构采用了部分卷积 PConv 加 PWConv 的结构，特征图之间可能包含非常相似的特征，或者存在高度重叠的感受野，利用特征映射中的冗余性，仅对部分输入通道使用常规卷积来进行特征提取，其余通道保持不变，部分卷积的结构如图 2 所示。

PConv 的算法流程可以简述为以下内容：假定输入特征图的通道数为 C ，我们设定一个参数 n 来表示将特征图分

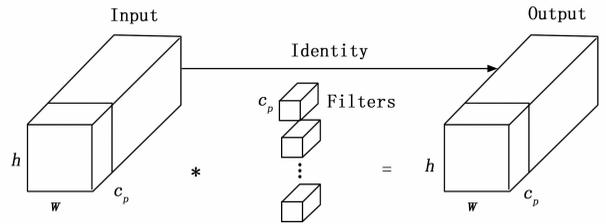


图 2 部分卷积示意图

割为部分卷积和未处理部分的比例，将输入特征图划分为两部分 X_1 和 X_2 ， X_1 的通道数为 C/n ， X_2 的通道数为 $C-C/n$ ，对 X_1 进行部分卷积，对 X_2 不做处理， X_1 得到的输出结果与 X_2 在通道维度上进行拼接得到输出特征图并返回进行之后的操作。

PConv 部分卷积仅对部分输入通道应用常规 Conv 进行空间特征提取，而对其余通道不做任何处理。设定只对其中 $1/4$ 的通道进行部分卷积，即 $C_p/C=1/4$ 时，对于连续或常规内存访问，我们将第一个或最后一个连续的 C_p 通道视为整个特征图的代表进行计算。在不失一般性的前提下，我们认为输入和输出特征图的通道数相同，因此 PConv 的 $FLOPs$ 仅为：

$$FLOPs = h * w * k^2 * C_p^2 \quad (2)$$

其中: h 、 w 为特征的高宽, k 为卷积核的大小, 此时 PConv 的 FLOPs 仅为常规卷积的 1/16, 内存访问量为:

$$h \times w \times 2C_p + k^2 \times C_p^2 \approx h \times w \times 2C_p \quad (3)$$

内存访问量为常规卷积的 1/4, 这样我们就可以利用特征图的冗余性进一步优化成本, 减少计算冗余和内存访问。

部分卷积具有高度并行化, 使所有像素都可以同时处理, 在 GPU 上可以实现高效加速, 使用部分卷积可以使用更少的操作次数来达到相同的精度。通过减少冗余计算和内存访问, 能够更有效地提取特征。虽然部分卷积可以使用少量参数来建立部分连接的卷积, 减少了局部区域感受野过小的问题, 通过并行计算可以增加计算性能, 但是对于较深较复杂的神经网络仍然可能出现瓶颈, 所以在部分卷积后面加入 PWConv, PWConv 模块只包含 1×1 大小的卷积核, 可以在不改变特征图大小的情况下改变特征通道数量, 并有效减少计算复杂度。

FasterNet 的基本结构如图 3 所示, FasterNet 模块的实现主要有以下几步: 首先将输入特征图通过上述中的部分卷积操作, 在部分卷积层后面接着两个 PWConv, 第一个 PWConv 用于减少特征图的通道数, 从而降低计算成本和延迟, 这样 PConv 和第一个 PWConv 的 FLOPs 为:

$$h * w * (k^2 * C_p^2 + C_p^2) \quad (4)$$

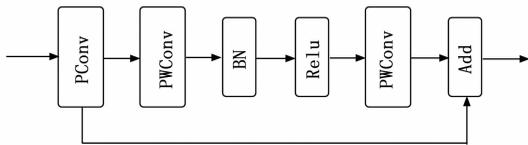


图 3 FasterNet 块结构

第二个 PWConv 层用于恢复特征图的通道数, 并且它可以对通道之间的关系进行编码, 将前面 PConv 中产生的空间信息和通道信息进行整合, 以提高网络的表达能力, 最后将输入特征图与输出特征图相加得到最终的输出结果。多个 FasterNet 模块组成子模块 m , 在前向传播过程中将输入张量卷积后的结果与输入张量拼接起来通过序列模块 m , 最后通过卷积得到最终输出。为了避免在神经网络中过度使用归一化层和激活层导致损害模型性能、限制特征多样性, 只把它们放在两个 PWConv 之间来保证特征多样性并实现更低的延迟, 在此处放置归一化层可以合并到相邻的卷积层中, 可以减少运算次数, 实现更快的推理速度。

改进后的 backbone 如图 4 所示。

2.2 多维动态卷积

传统的标准卷积是现代卷积神经网络的常见训练范式, 它的优点是简单、快速、易实现, 在标准卷积计算过程中, 静态卷积核会与输入的特征图的每个位置重合得到一个感受野区域, 并将卷积核的每个参数与对应位置的输入值相乘, 将所有结果相加得到输出值, 但是它的卷积核大小固定, 这些权重在训练过程中固定不变, 不会随着输入数据的变化而变化, 不同的数据需要使用不同的卷积核, 卷积

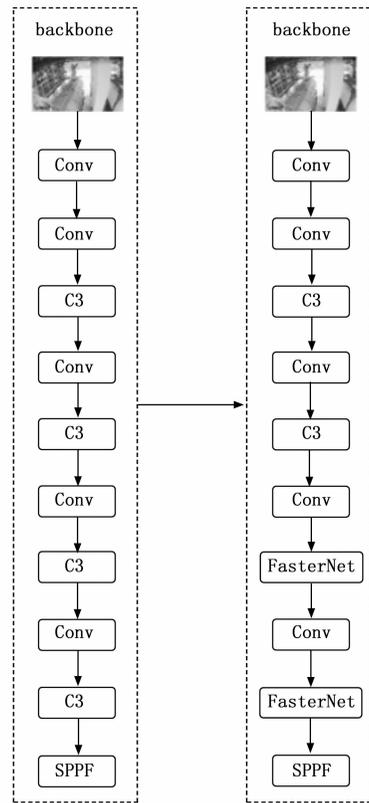


图 4 改进后的 backbone

核大小需要足够大才能捕捉图像的细节特征, 且对每个位置的输入特征采取相同的卷积操作, 这就导致一些局部特征细节会被忽略, 尤其是在边缘角落区域会丢失信息。

相比标准卷积, 动态卷积会根据输入特征图自适应调整卷积核的权重, 其形状大小可以根据输入数据的不同而变化, 在每个感受野内采用不同的卷积核进行卷积, 将输入数据划分为不同的空间分组和通道分组, 选择不同卷积核进行卷积运算, 生成输出值, 允许更多的计算共享, 使计算效率得到提高。在动态卷积中, 将注意力机制应用到卷积核上来动态地调整卷积核权重, 能够捕捉到更多的局部细节, 保持有效的推理效率, 应对更加复杂的任务和数据, 从而提高模型的性能。

常见的 DyConv 动态卷积只在一种维度上学习特征, 忽略了卷积核空间大小、输入通道数、输出通道数这 3 种维度, 每个卷积核的权重共享相同的注意力标量, 限制了捕获上下文信息的能力, 为了改进这一问题, 本文中采用动态卷积的方法, 在动态卷积模块中使用多种维度的注意力机制并行计算的策略, 分别沿输入通道维度、输出通道维度、空间维度、卷积核维度这 4 种类型的注意力进行计算。

多维动态卷积如图 5 所示, 其原理为输入特征图经过多维动态卷积生成多个不同维度的特征图, 对于每个特征图通过注意力机制计算每个位置上每个卷积核的权重, 并将它们组合成一个加权卷积核, 自适应地选择不同的卷积

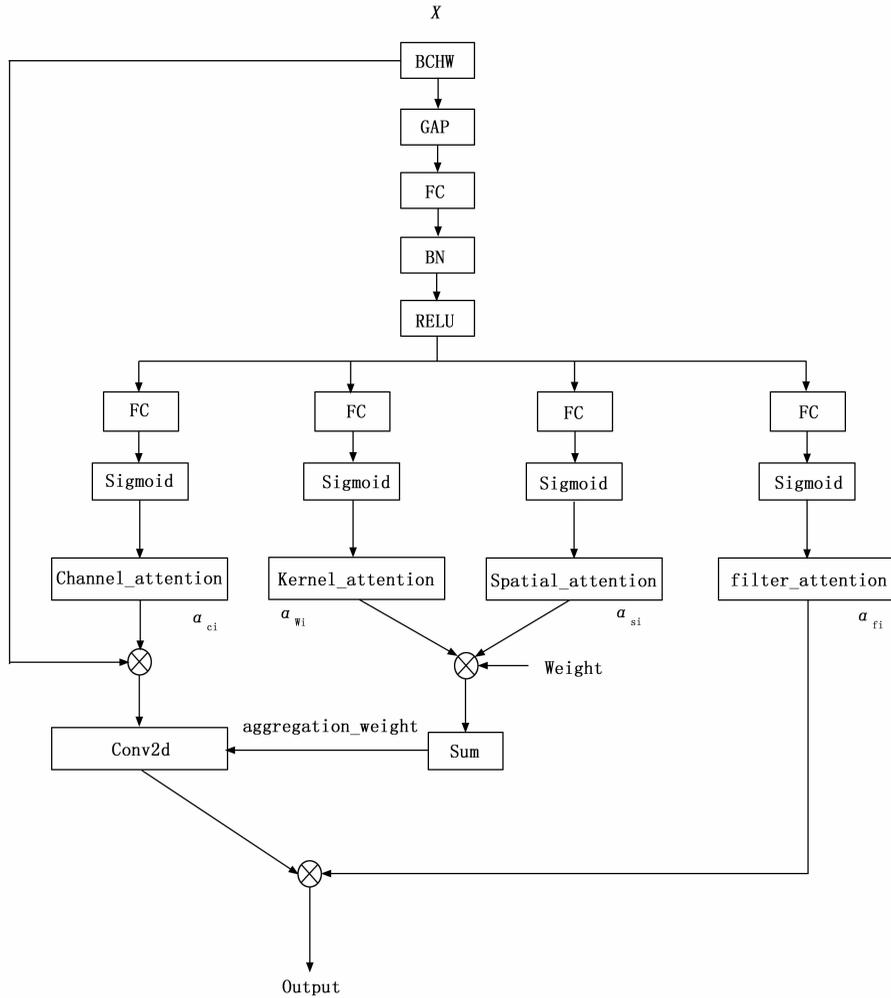


图 5 多维动态卷积示意图

核来处理不同位置上的特征信息，使用组合后的加权卷积核对每个位置上的特征进行卷积操作，得到最终的输出特征。

这 4 种注意力机制分别用于计算输入特征图、输出特征图和卷积核的动态权重，这些动态权重被用于计算卷积的输出，公式可以简化如下：

$$Output = (\alpha_{ci} \odot \alpha_{wi} \odot \alpha_{si} \odot \alpha_{fi} \odot W_1 + \dots + \alpha_{ci} \odot \alpha_{wi} \odot \alpha_{si} \odot \alpha_{fi} \odot W_n) * x \quad (5)$$

其中： W_i 表示第 i 个卷积核，可以表示为一个大小为 $n \times k \times k \times C_{in}$ 的四维数组，是由 n 个卷积滤波器组成的， W_{mi} 表示第 i 个卷积核的第 m 个滤波器，是 W_i 中的一个元素，大小为 $k * k * C_{in}$ ， C_{in} 为输入通道数， n 为卷积核数量， k 表示卷积核大小； α_{si} 表示卷积核 W_i 的 $k \times k$ 个空间位置上每个滤波器分配不同的注意力权重； α_{ci} 表示每个卷积滤波器 W_{mi} 的 C_{in} 个通道分配不同的注意力权重； α_{fi} 表示 C_{out} 个卷积滤波器分配不同的注意力权重， C_{out} 为输出通道数； α_{wi} 表示整个卷积核 W_i 分配一个注意力权重， \odot 表示沿内核空间的不同维度的乘法运算，Output 表示最终的输出特征图。这几

种注意力是互补的，将其与卷积核逐一相乘，使卷积存在各个维度的差异性，可以根据输入特征图的内容自适应地调整卷积核的形状和权重，更好地获取上下文信息，提高卷积神经网络的性能。

多维动态卷积模块的算法如下所示：

算法 1：多维动态卷积模块算法

- 1) 获取输入特征图 X 的批处理大小、输入通道数、图像尺寸；
- 2) 将输入特征图 X 经过通道式全局平均降维得到 X_1 ；
- 3) 通过 FC、BN、ReLU 层处理得到 X_2 ；
- 4) 分别通过 4 条分支计算得到 4 种注意力 $\alpha_{si}, \alpha_{ci}, \alpha_{fi}, \alpha_{wi}$ ；
- 5) α_{si}, α_{ci} 与卷积核权重相乘，通过 sum 函数按通道累加得到聚合权重 aggregation_weight；
- 6) 输入特征图 X 与 α_{ci} 相乘后得到的聚合权重进行卷积得到 Output1；
- 7) 将 Output1 与 α_{fi} 相乘并返回输出特征图 Output。

多维动态卷积的算法如上所示，首先经过 GAP 全局平

$$\beta = \frac{L_{IoU}^*}{L_{IoU}} \quad (10)$$

L_{IoU}^* 表示预测的边界框和真实边界框之间的 IoU 阈值。当 β 越小时, 锚框质量越高, 当 β 越大时, 锚框质量越差, 为了让边界框更加关注普通质量的锚框, 当 β 较大和较小时都分配一个较小的梯度增益, 构造一个非单调聚焦系数 r , 通过动态聚焦机制来实现, 用于控制不同位置之间的重要性权重:

$$r = \frac{\beta}{\delta \alpha^{\beta-\delta}} \quad (11)$$

δ 是指控制样本权重的缩放常数, α 是表示焦点系数的一个超参数, 用于控制损失函数中不同样本的权重, 当 β 等于一个常数时, 锚框具有最高梯度增益, 此时损失函数可定义为:

$$L_{wIoU} = rL_{wIoU_i} \quad (12)$$

L_{IoU} 是动态的, 锚框的质量划分标准也是动态的, 这让 $WIoU$ 损失函数能一直做出最适合当前状况的梯度增益分配策略。为了确保训练早期阶段模型能更快地学习到高质量的锚框, 需要设置一个动量 m 来延迟 L_{IoU} 值逼近真实的 $L_{IoU_{real}}$ 的时间, 使训练在初始阶段更多地关注到就有高质量的锚框。在训练的后面阶段, 为低质量的锚框分配较小的权重, 减少有害的梯度。

3 实验结果与分析

3.1 实验数据集

本文实验所用数据集为化工厂区装卸车作业区的防爆摄像头所拍摄的数据, 包含白天与夜间时间段, 晴天、阴天、雨天等天气状况下的实际作业场景, 将拍摄下的数据进行手动筛选后标注, 安全绳十分细长, 且背景较为复杂, 极易与背景中的物体重合, 相比安全帽等类物体的目标检测工作更加困难, 因此在标注时尽量贴合安全绳边界进行标注, 最大程度地降低背景对特征提取的影响, 提高对安全绳细小特征的提取, 最终得到的数据共计 5 000 张, 按照 8 : 1 : 1 的比例分成训练集、验证集、测试集三部分。

3.2 实验环境

本文实验是基于 PyTorch 和 CUDA 框架在 Linux 操作系统下进行的, 使用的预训练模型为 YOLOv5s, 具体的硬件配置和环境配置如表 1 所示。

表 1 实验平台参数

参数	版本
操作系统	Ubuntu 18.04
CPU	Intel(R) Core(TM) i7-12700KF CPU @ 3.60 GHz
GPU	NVIDIA GeForce RTX 4090
Python	3.7.1
PyTorch	PyTorch 1.10.1+cu113
CUDA	11.3
CuDNN	8.2

3.3 评估指标

本文实验采用准确率 (P, precision)、召回率 (R,

recall)、计算量 (GFLOPs)、平均准确度 (AP, average precision)、平均精度均值 (mAP, mean average precision)、参数量 (parameters)、平均检测时间作为评估指标。

其中准确率和召回率的公式分别为式 (13) (14), P 表示安全绳识别准确率, TP 表示将安全绳识别正确的样本数量, FP 表示将不是安全绳的目标识别成安全绳的数目, FN 表示将安全绳识别成不是安全绳的样本数目, R 表示安全绳目标检测的召回率。

$$P = \frac{TP}{TP + FP} \quad (13)$$

$$R = \frac{TP}{TP + FN} \quad (14)$$

平均准确度的公式为式 (15), AP 表示精度的平均值, m_{AP} 的公式为式 (16):

$$X_{AP} = \int_0^1 P(R) \quad (15)$$

$$m_{AP} = \frac{\sum_{i=1}^K X_{AP_i}}{K} \quad (16)$$

X_{AP_i} 表示对第 i 类物体的检测平均准确率, K 为类别。

在本文实验中以 IoU 阈值为 0.5 的情况作为评价指标, 表示为 $mAP@0.5$, 公式如式 (17), $X_{AP@0.5i}$ 代表第 i 类物体在 IoU 阈值为 0.5 时的平均准确度。

$$mAP@0.5 = \frac{\sum_{i=1}^K X_{AP@0.5i}}{K} \quad (17)$$

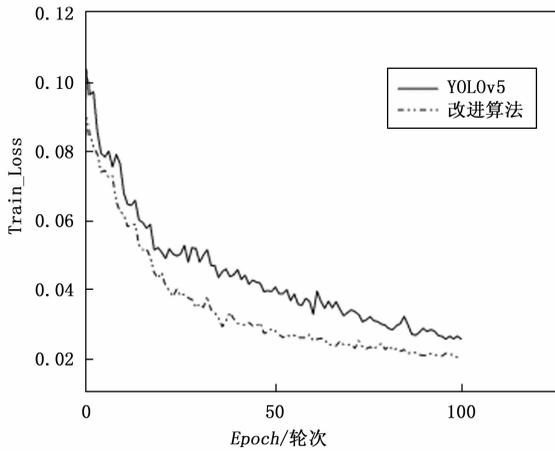
3.4 模型训练

在模型训练中, 设置的迭代次数为 100 轮, 通过实验看到在迭代 100 轮时模型已经达到收敛, 每次训练选取的样本数为 32, 权重衰减系数设置为 0.000 5, 动量因子为 0.8, 初始学习率为 0.01, 在夜间安全绳检测相对困难, 如果置信度阈值过高, 只有那些非常可信的目标候选框才会被认定为有效目标, 可能会出现漏检, 设置过低较低置信度得分的目标候选框也会被认定为有效目标。这可能导致模型出现大量误检, 我们将置信度阈值设置为常用数值 0.5, 在进行分类预测时, 只有当置信度高于或等于 0.5 时, 才会将该预测结果视为有效结果, 使用 SGD 优化器, SGD 优化器收敛速度快, 每次只更新一部分参数, 占用内存小, 泛化性好, 训练损失变化如图 7 所示。

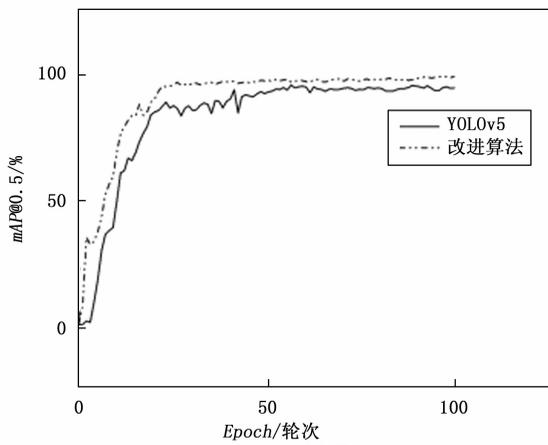
3.5 对比实验分析

为了验证本文改进算法的有效性, 在同等实验条件下, 与当前较为常用的目标检测算法如 Faster R-CNN、Fcos^[20]、YOLOX 等模型进行对比实验, 选取未参与训练过程的数据作为测试集。在对比实验中, 选择 $mAP@0.5$ 、计算量、参数量、推理速度等作为评价指标。

由表 2 可知, 改进后的算法与目前常见的目标检测算法相比, 改进算法有更强的识别性能, $mAP@0.5$ 比 yolox 高 5.2%; 在计算量 $FLOPs$ 和参数量指标比较中, 本文算法虽然分别略高于 yolox 和 Faster R-CNN, 但差距不大,



(a) 边界框回归损失对比图



(b) mAP@0.5对比图

图 7 改进后算法和 YOLOv5 训练过程曲线变化

比其他模型的计算量和参数量减少明显, 模型更加轻量化; 在推理速度指标比较中, 本文算法的推理速度更快, 比推理速度较快的 yolof 还要快 1.3 ms, FPS 也高出近 63。综合来看, 本文改进模型无论在识别性能还是在推理速度上都优于其他模型, 在安全绳的识别任务中有更好的识别效果。与 YOLOv3、YOLOv4 相比提升效果明显。

表 2 对比实验

模型	mAP@0.5	FLOPs	参数量/M	单张照片推理速度/ms	FPS
atss	89.6	110.96	50.88	11.7	85.5
Faster R-CNN	83.8	91.0	6.1	8.9	109.3
Fcos	87.1	78.63	31.84	9.1	110.1
gfl	92.7	81.77	32.03	9.5	105.8
yolox	92.6	13.4	8.97	6.1	165.6
yolof	84.4	39.27	39.27	5.2	193.8
YOLOv3	82.5	156.6	61.94	6.5	153.8
YOLOv3-SPP	83.5	157.5	62.9	6.7	149.2
YOLOv3-tiny	80.1	13.3	8.66	4.7	212.7
YOLOv4	87.6	110.5	40.2	7.1	140.8
改进算法	97.8	13.8	6.12	3.9	256.4

3.6 消融实验分析

为进一步验证对 YOLOv5 模型的改进方法的有效性, 进行了消融实验, 我们对每个实验进行 4 次训练, 选取结果中较为稳定的结果数值取平均数, 改进后的模型对安全绳的检测准确率为 94.6%, 召回率为 96.7%, FLOPs 为 12.8, mAP@0.5 达到 97.8%, 实验结果对比如表 3 所示。

表 3 消融实验

模型	准确率/%	召回率/%	mAP@0.5/%	FLOPs	单张照片推理速度/ms
YOLOv5	91.6	96.2	93.5	15.8	3.6
YOLOv5+FasterNet	96.7	97.2	97.2	14.8	3.6
YOLOv5+多维动态卷积	94.3	96.8	97.1	15.3	4.0
YOLOv5+WIOU	94.7	96.7	97.5	15.8	3.4
YOLOv5+多维动态卷积+WIOU	96.2	96.7	97.5	15.3	3.7
YOLOv5+FasterNet+WIOU	96.4	97.0	97.6	14.1	3.5
YOLOv5+FasterNet+多维动态卷积	95.8	96.9	97.2	13.6	3.8
改进算法	96.6	96.7	97.8	13.8	3.9
改进后与 YOLOv5 性能差异	+5.0	+0.5	+4.3	-2.0	+0.3

通过表中数据得到以下结论: 在单个模块的消融实验中, 通过改进 YOLOv5s 的主干网络, 使用 PConv 和 PW-Conv 结合的 FasterNet 模块后, 模型计算量减少较多, 同时准确率和 mAP 值都有一定提升, 用多维动态卷积后, 计算量有一定减少, 而且识别率提高了 2.8%, mAP@0.5 提升了 3.6%, 单张图片推理速度增加了 0.4 ms; 将损失函数由原来的 CIOU_Loss 替换为 WIOU_Loss 后, 准确率提高了 3.1%, 单张图片推理速度减少了 0.2 ms。随后我们又进行了任意两个模块组合的方式进行实验, 添加两个模块相较于单个模块在准确率和 mAP 值上也有少量的提升, 由于多个模块的加入, 在照片推理速度方面会有一定的增加。改进模块两两组合时 mAP 值总体会比逐一模块的 mAP 略有提高, 虽然 mAP 值与总体改进算法相差不大, 但在准确率、FLOPs 以及推理速度上有相对较大的变化。

通过消融实验可知, 3 种改进方法都是有效的, 相较于 YOLOv5 原模型都有一定程度的提高, 最终的改进算法比 YOLOv5 在识别率上提高了 3.0%, 召回率提高了 0.5%, mAP@0.5 提高了 4.3%, 单张图片推理速度增加了 0.3 ms, 虽然推理速度增加了 0.3 ms, 仍旧满足实际作业过程中的实时检测的任务需求, 且识别准确率获得较大提高。

为了更加直观地看到模型改进前后的性能差异, 观察改进前后对不同时间、环境下和不同状态下的安全绳, 现选取未参与训练过程的数据进行测试, 分别验证原生与改进后的模型, 验证结果如图 8 所示, 根据图 8 (a) 可以看到, 在白天光线较暗的情况下, 改进后的算法明显检测性能更强, 当细长的安全绳与人物背景重合时, 改进算法能够更好地检测出与复杂背景重叠的目标; 从图 8 (b) 可以

看出,在夜间情况下,改进后的算法比原来的算法检测精度也有了不小的提升,在模型改进后对模糊目标的特征提取能力有明显提高,能够更加准确地识别夜间情况下的安全绳。由此可以得出结论,改进后的网络模型能够在不同场景下都有更好的识别性能,改进算法能够更好地适应不同场景下的检测任务。

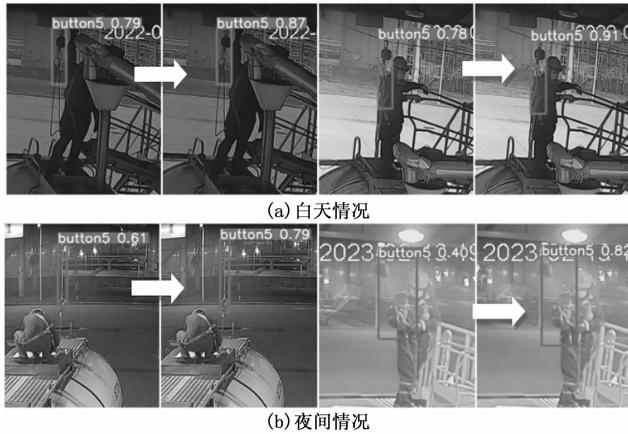


图 8 验证结果

4 结束语

本文对比原本的 YOLOv5 算法,本文所提出的在主干网络中使用 FasterNet 模块,能有效地减少计算量,降低延迟,加快推理速度,使网络更轻量化;然后使用与注意力结合的动态卷积,能更好地提高特征提取能力,最后把损失函数替换为 $WIoU_Loss$,加快边界框的回归,使模型更快收敛。实验结果表明,经过改进后的算法的准确率提高了,改进后的算法可以满足对工人安全绳的识别精度和实时检测的要求,提高了对于夜间和光照不充足的环境下的安全绳的识别效果。未来将继续研究如何提高安全绳的识别准确率,进一步提高检测性能并在实际业务流程中发挥作用。

参考文献:

- [1] 孙世梅, 赵金坤, 傅 贵. 基于“2~4”模型的高处坠落事故行为原因研究 [J]. 中国安全科学学报, 2019, 29 (8): 23 - 28.
- [2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580 - 587.
- [3] GIRSHICK R. Fast R-CNN [C] // Proceedings of IEEE International Conference on Computer Vision, 2015: 1440 - 1448.
- [4] REN S, HE K, 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: 1137 - 1149.
- [5] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020: 386 - 397.
- [6] 张彦凯, 余飞鸿, 夏 岩, 等. 一种高精度优化 Faster-RCNN 变电站安全帽检测方法 [J]. 四川电力技术, 2023, 46 (1): 62 - 68.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] //European Conference on Computer Vision. Springer, Cham, 2016: 21 - 37.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779 - 788.
- [9] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement [C] //IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018, 87 (8): 101 - 104.
- [10] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. (2024-03-01) [2020-06-05]. <https://arxiv.org/abs/2004.10934>.
- [11] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C] //2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2021.
- [12] 农元君, 王俊杰, 徐晓东, 等. 面向嵌入式平台的安全帽实时检测方法 [J]. 计算机工程与应用, 2022, 58 (9): 161 - 167.
- [13] 顿伟超, 王淑青, 张鹏飞, 等. 基于改进 YOLOv4 的电力高空作业安全带检测 [J]. 湖北工业大学学报, 2022, 37 (5): 6 - 11.
- [14] 郑楚伟, 林 辉. 基于 Swin Transformer 的 YOLOv5 安全帽佩戴检测方法 [J]. 计算机测量与控制, 2023, 31 (3): 15 - 21
- [15] HOWARD A G, SANDLER M B, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks [C] //Computer Vision and Pattern Recognition, 2018: 4510 - 4520.
- [16] 田 坤, 李 冠, 赵卫东. 基于 YOLO 和极限学习机的驾驶员安全带检测模型研究 [J]. 计算机应用与软件, 2019, 36 (11): 196 - 201.
- [17] 于 硕, 李 慧, 桂方俊, 等. 复杂场景下基于 YOLOv5 的口罩佩戴实时检测算法研究 [J]. 计算机测量与控制, 2021, 29 (12): 188 - 194.
- [18] 王彦生, 曹雪虹, 焦良葆, 等. 基于改进 YOLOv5 的电厂人员吸烟检测 [J]. 计算机测量与控制, 2023, 31 (5): 48 - 55.
- [19] 徐长凯. 基于优化边界框回归的目标检测研究 [D]. 青岛: 青岛科技大学, 2021.
- [20] KASHIKAWA N, AOKI K, ASAI R, et al. FOCAS: the faint object camera and spectrograph for the subaru telescope [J]. Publications of the Astronomical Society of Japan, 2002: 819 - 832.