

DOI:10.19651/j.cnki.emt.2210228

面向密集型钢筋计数的 GCA-MobilenetV2-YOLOv4 算法*

刘浩 辛山

(北京建筑大学电气与信息工程学院 北京 102627)

摘要: 为提高建筑工地的钢筋计数效率,围绕施工单位硬件设备算力不足,钢筋图像物体密集遮挡严重的情况,提出一种改进的轻量化 YOLOv4 算法。提出 GCA-MobilenetV2 轻量级网络替换 CSPDarknet53,作为 YOLOv4 算法的主干特征网络。针对钢筋图像密集,物体间遮挡严重的情况,提出融合通道注意力机制的 attention-CSP-PANet 结构。针对深层网络 SPP 结构参数量大,模型训练时梯度消失的问题,提出 DepthLite-SPP 结构,增强深层网络的感受野,提高算法的检测速度。针对一阶段回归的算法正负样本失衡问题,设计 CIOU-Focal 损失函数。实验证明,在自建钢筋数据集中检测精度为 98.78%,对比原算法精度提升了 3.36%,检测速度 FPS 提升了 7.6,参数量仅为原算法的 1/3。

关键词: 钢筋计数;YOLOv4 算法改进;GCA-MobilenetV2 网络;attention-CSP-PANet 结构;DepthLite-SPP 结构

中图分类号: TP399 **文献标识码:** A **国家标准学科分类代码:** 520.6040

GCA-MobilenetV2-YOLOv4 algorithm for intensive rebar counting

Liu Hao Xin Shan

(School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 102627, China)

Abstract: To improve the counting efficiency of steel bars in construction sites, an improved lightweight YOLOv4 algorithm is proposed based on the insufficient computing power of hardware equipment in construction units and the dense occlusion of steel bar image objects. GCA-MobilenetV2 lightweight network is proposed to replace CSPDarknet53 as the main feature network of YOLOv4 algorithm. Aiming at the situation of dense steel bar images and serious occlusion between objects, attention-CSP-PANet structure integrating channel attention mechanism is proposed. Aiming at the large number of SPP structure parameters in deep network, DepthLite-SPP structure is proposed to enhance the receptive field of deep network and improve the detection speed of the algorithm. In view of the imbalance between positive and negative samples of the one-stage regression algorithm, CIOU-Focal loss function is designed. The experimental results show that the detection accuracy of the steel bar data set is 98.78%, which is 3.36% higher than that of the original algorithm. The detection speed FPS is 7.6, and the number of parameters is only 1/3 of the original algorithm.

Keywords: steel counting; improved YOLOv4; GCA-MobilenetV2; attention-CSP-PANet structure; DepthLite-SPP structure

0 引言

随着新兴信息化技术的发展,智慧化的工地管理应用不断出现。目标检测(object detection)算法^[1]在智慧工地中已获得广泛的应用,如安全帽检测系统^[2]、零件器缺陷检测系统^[3]、车辆监控系统^[4]、人脸检测系统^[5]等。这些 AI 应用不仅可以提高工地的施工效率,保障施工人员的人身

安全,还能更好的发展绿色施工理念。

在传统的建筑施工现场,工人清点钢筋是一项繁琐而又非常重要的工作。由于钢筋成捆放置,密集程度高,数目多,清点时很容易出错,且非常耗时耗力。为解决这一问题,人们发明了各种钢筋自动计数技术,其中最有效的是利用传统的图像处理方法进行钢筋识别、统计工作,在一定程度上解决了计数问题。但当实际场景比较复杂时,传统的

收稿日期:2022-06-06

* 基金项目:智能机器人与系统高精尖创新中心建设项目(00921917001)、北京市重点实验室项目(BZ0337)资助

图像处理方法往往存在较大的误差。

近年来,大多数钢筋计数任务都是使用基于卷积神经网络(convolutional neural network,CNN)的目标检测算法实现的。在钢筋计数方面,王志丹^[6]利用Faster-RCNN^[7]目标检测算法,在小数据的钢筋数据集中进行实验,采用数据增强方法提升计数精度,但改进的Faster-RCNN^[7]算法,模型参数计算量大,检测速度慢,并不适用于实时检测任务。Zhu等^[8]在卷积神经网络中,结合Inception,增强感受野(receptive field block net,RFB),特征金字塔(feature Pyramid network,FPN)结构,解决了钢筋密集,目标小尺度,光照变化大,多噪声等问题。在钢筋检测中,与Cascade-RCNN-ResNet101算法相比精度提升明显,且具有更快的运行速度。明宏宇等^[9]针对密集型钢筋计数效果不理想的情况,提出了一种基于RetinaNet目标检测框架,结合高斯混合模型(Gaussian mixture model,GMM)和期望最大化(expectation maximization,EM)算法,以此提高计数精度。Yang等^[10]提出了一种基于Yolov3检测器的深度学习方法,用于自动钢筋检测和计数。采用额外的特征金字塔结构进行特征提取,利用IOU-Loss结合Focal Loss训练,以及Bag of Freebies优化方法,以改善钢筋检测和计数性能。Ana等^[11],通过神经网络和聚类两种方式相结合,进行钢筋计数。利用聚类优化钢筋数据集,解决由于目标多导致训练泛化性能弱的缺陷,设计出SA-CNN-DC网络,由三个神经网络和集群技术Nique组成,来提升对钢筋数据的预测和分类能力。

上述研究针对钢筋图像密集的问题提出了不同的研究方法,但是并未考虑面对钢筋密集造成物体间遮挡导致误检漏检的情况。此外,在施工现场实际应用时,需要额外的开销来配备强大的GPU算力设备,算法需要轻量化,因此上述研究还有很大的提升空间。对此,本文提出改进的YOLOv4^[12]目标检测算法,设计出针对密集型钢筋计数的轻量级网络,可应对硬件设备算力不足的情况,重点解决因钢筋密集造成的遮挡严重导致误检漏检的问题。本文研究如下:

1)采用MobilenetV2^[13]轻量级网络代替(cross stage partial Darknet53,CSPDarknet53)结构。在骨干网络为加强重要信息的特征提取,提出融合全局坐标注意力(global coordinate attention,GCA)的GCA-MobilenetV2结构,重点解决轻量级网络面对遮挡问题特征提取力不足导致误检漏检的情况。

2)为进一步解决遮挡问题,提出融合通道注意力的加强特征提取网络(attention-CSP-PANet),在特征融合中对网络进行注意力的加权,提高检测精度。

3)YOLOv4算法采用的深层感受野模型空间金字塔池化结构^[14](spatial Pyramid pooling,SPP)参数量大,面对算法轻量化的需求,本文提出融合深度可分离卷积的空间残差金字塔池化结构(DepthLite-SPP),融合深层感受野的

多尺度信息,扩大深层网络的感受野,降低模型参数量,并有效的缓解模型训练时梯度消失的问题。

4)YOLOv4是一阶段(one-stage)回归的算法,在模型训练时存在正负样本失衡问题,尤其是面对钢筋数据集物体种类单一数目多的情况,会存在大量的负样本。本文参考Focal Loss^[15],设计CIOU-Focal Loss损失函数,解决正负样本失衡问题。

1 改进的YOLOv4目标检测算法

1.1 YOLOv4算法概述

YOLOv4目标检测算法是Yolov3^[16]算法的改进,结构如图1所示。在图1的网络结构中分为主干特征网络(CSPDarknet53),加强特征提取网络(path aggregation network,PANet),分类器(head)3个部分。在加强特征提取网络中,PANet结构采用多尺度融合的方式,并在上采样与下采样过程中进行5次卷积,相比FPN结构,避免了特征信息丢失的缺陷。在Head结构中,将图像分别在 $13 \times 13, 26 \times 26, 52 \times 52$ 的3个特征图中预测锚框对应的物体的位置信息,置信度信息,分类信息。在训练过程中采用Mosaic数据增强的方法,并添加余弦退火衰减学习率。具有检测速度快,检测精度高的特点。

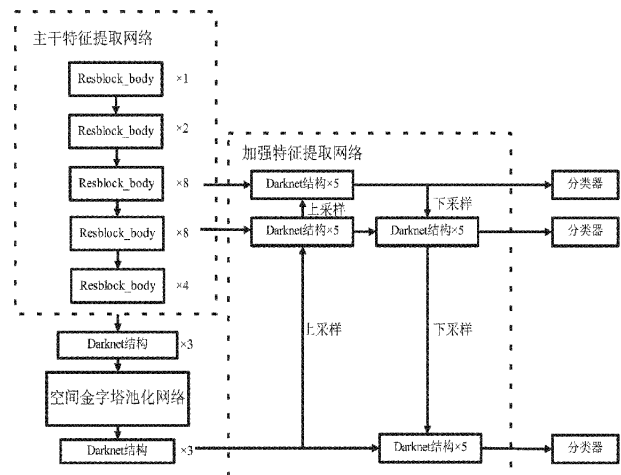


图1 YOLOv4算法结构

1.2 骨干网络改进

基于CNN方法的钢筋识别计数,丰富的特征信息有利于小尺度的锚框的回归预测。现有的主干特征提取网络,如稠密卷积网络(dense convolutional network,DenseNet),深层卷积网络(visual geometry group,VGG)等,利用大量的卷积计算获得的深层语义信息,提高识别的精度,但是施工现场很难配备强大的算力设备。对于施工单位,面对施工效率的需求,算法需要很快的检测速度。MobilenetV2结构采用了(3,3)大小的卷积核的深度可分离卷积,相比普通卷积减少了大量的参数计算,但是这种结构在下采样过程中容易造成特征信息的丢失,为保证钢筋

图像的计数精度,本文参考坐标注意力^[17](coordinate attention,CA),为解决遮挡问题,提出 GCA 注意力机制对存在物体的重要的通道信息重点关注,相比 CA 注意力,加强了全局信息的识别。通过在部分深度可分离卷积后添加 GCA 注意力,提出 Inverted GCA resblock 结构,如图 2 所示,弥补深度可分离卷积参数数量降低造成特征信息丢失的问题。

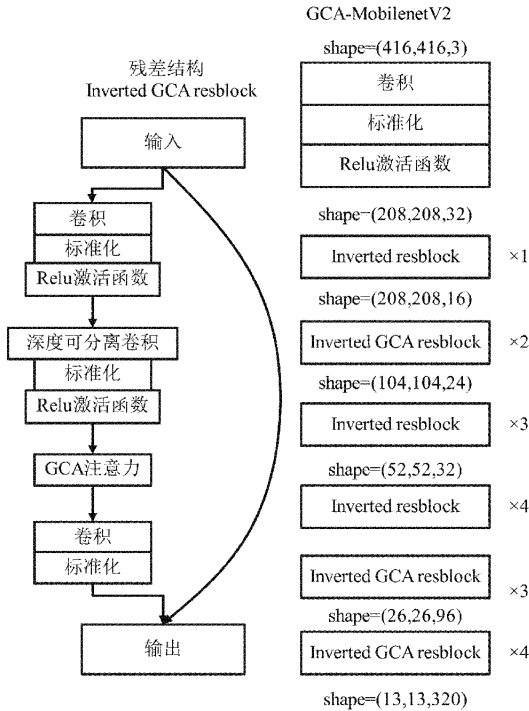


图 2 改进的 MobileNetV2

本文通过在 Inverted resblock 中添加 GCA 注意力,调整每个通道的权重,提高重要信息的表达,GCA 注意力如图 3 所示。图中 W, H 和 C 分别表示特征图的宽,高和通道数。

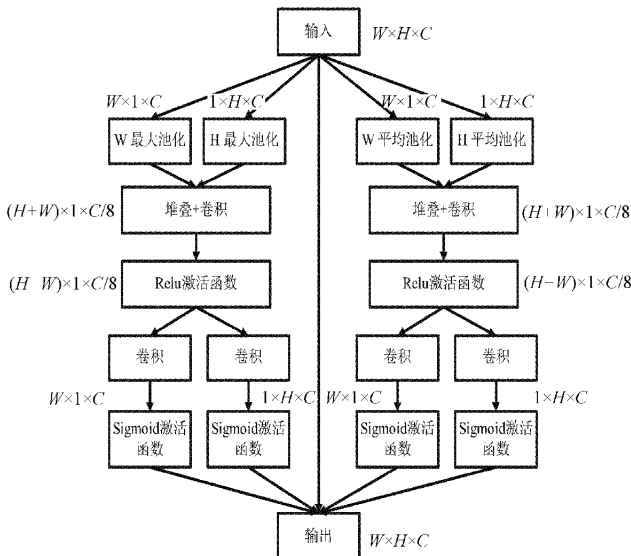


图 3 GCA 注意力机制

先将 Inverted resblock 中经深度可分离卷积提取的特征图在宽和高两个方向分别进行全局平均池化(global average pooling)和全局最大池化(global max pooling),公式如式(1),将特征图在宽和高两个方向进行压缩,分别得到特征图的平均值与最大值的感受野 $f(w)$ 与 $f(h)$ 。

$$f_{Avg}(w) = \frac{1}{w} \sum_{0 \leq i \leq w} f(h, i, c)$$

$$f_{Avg}(h) = \frac{1}{H} \sum_{0 \leq j \leq H} f(j, w, c)$$

$$f_{Max}(w) = \frac{1}{w} \sum_{0 \leq i \leq w} f(h, i, c)$$

$$f_{Max}(h) = \frac{1}{H} \sum_{0 \leq j \leq H} f(j, w, c)$$

将获得的平均值与最大值的感受野,沿道通在宽高两个方向堆叠,送入共享网络,公式如式(2)所示。

$$f_{Avg} = AvgPool [f_{Avg}(w); f_{Avg}(h)]$$

$$f_{Max} = MaxPool [f_{Max}(w); f_{Max}(h)]$$

将特征图 f 在通道方向先降维过 δ (Relu) 激活函数,再调整为原来的大小,并在宽高两个方向上过 σ (Sigmoid) 激活函数后得到特征图在宽度方向上的权重 f^w 和高度方向的权重 f^h 。公式如式(3)所示。

$$f_{Avg}^w = \sigma(\delta(f_{Avg}))$$

$$f_{Avg}^h = \sigma(\delta(f_{Avg}))$$

$$f_{Max}^w = \sigma(\delta(f_{Max}))$$

$$f_{Max}^h = \sigma(\delta(f_{Max}))$$

最后宽度注意力权重与高度注意力权重与输入特征图相乘,公式如式(4),分别获得了全局平均池化与全局最大池化后在宽和高两个方向的注意力特征图。

$$f_{Avg-out}(W, H, C) = f_{Avg}(w, h, c) \times f_{Avg}^w \times f_{Avg}^h$$

$$f_{Max-out}(W, H, C) = f_{Max}(w, h, c) \times f_{Max}^w \times f_{Max}^h$$

最后将注意力特征图的最大值与平均值相加,实现特征融合,获得图像的全局特征信息与空间信息公式如式(5)所示。

$$f_{out} = f_{Avg-out} + f_{Max-out}$$

本文将这一结构融合在 Inverted Resblock 中,面对钢筋密集造成的遮挡问题,通过对特征图在宽和高两个方向上的自适应学习,对物体的有用信息给予大的权重,增强了全局信息的识别与空间信息的捕捉能力,提高钢筋图像的检测精度。

1.3 融合通道注意力的 attention-CSP-PANet 结构

面对钢筋图像物体尺度小,密集程度高的特点。FPN 结构,利用自下而上的特征融合的方法,更关注浅层网络,获得更多的小物体目标的信息,提高小尺度物体的检测精度。但是 FPN 结构层与层之间跨度远,容易造成特征信息丢失,YOLOv4 算法的 PANet 结构利用三个特征层间的特征融合,进行 5 次卷积,更好的保存了特征信息。但是这种

结构庞大的参数量在训练过程容易导致梯度消失。对此,本文将PANet结构上采样与下采样过程中进行的五次卷积残差化,残差化的5次卷积结构如图4所示。在特征融合的上采样和下采样过程中,先将通道进行一个大的残差,在主干部分,进行5个小的残差结构,分别进行(1,1)和(3,3)的卷积,之后与小的残差边的特征信息相加,最后与大残差边的特征信息堆叠,这种结构不仅可以防止网络训练中的梯度消失,还减少了参数量,有效的提高训练和检测的速度。

在特征融合中,为进一步解决钢筋图像遮挡严重的问题,加强特征融合中重要的通道信息,考虑到输入的特征图已经过主干特征网络的空间信息捕捉,我们在特征金字塔的特征融合前后加入通道注意力机制。通道注意力结构如图5所示。图中W,H和C分别表示特征图的宽,高和通道数。

通过对特征图像中的通道分别采用全局平均池化和全局最大池化,在每个通道的特征图中计算出最大值和平均值。最大值与平均值公式如式(6)、(7)。式中 δ 是Relu激活函数, σ 是Sigmoid激活函数,U是特征图的通道信息, W_1 与 W_2 分别表示第1个全连接层和第2个全连接层的

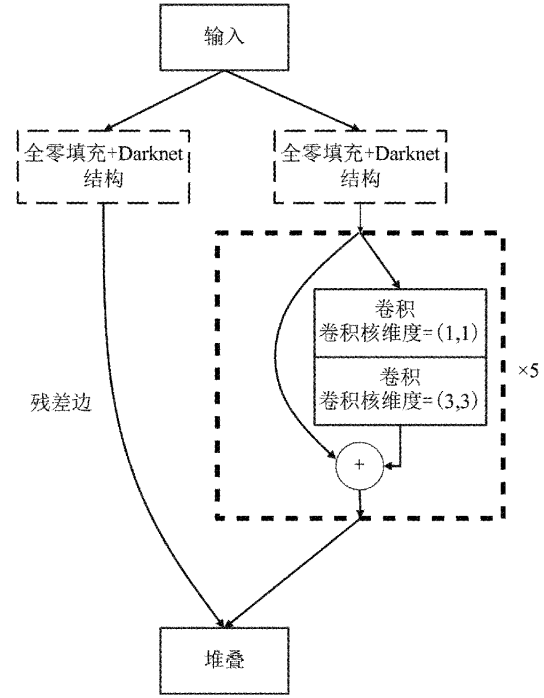


图4 残差化的5次卷积结构

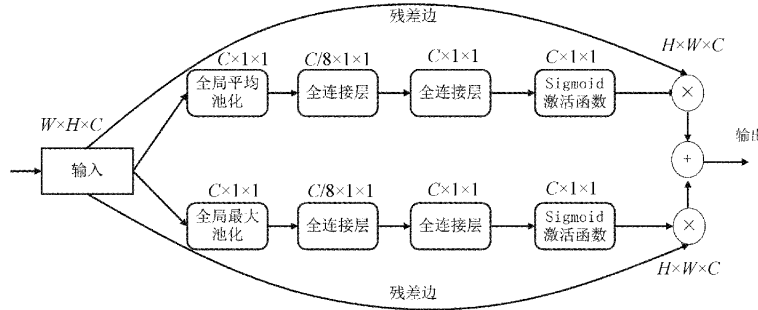


图5 通道注意力机制结构

通道特征信息的权值系数。将聚合特征映射的所有全局信息,通过压缩输入通道图的空间维数,逐元素求和,形成通道注意力图,公式如式(8)所示。

$$F_{Avg} = GlobalAvg(F)$$

$$F_{Avg}(U) = \frac{1}{H \times W} \left(\sum_{i=1}^W \sum_{j=1}^H U(i, j) \right)$$

$$F(F_{Avg}, W) = \sigma(W_2 \delta(W_1 F_{Avg})), W_1 \in R^{\frac{c}{r} \times c}, W_2 \in R^{\frac{c}{r} \times c}$$

$$F_{Avg} = F_{Avg} \times (H, W, 1)$$

$$F_{Max} = GlobalAvg(F)$$

$$F_{Max}(U) = \frac{1}{H \times W} \left(\sum_{i=1}^W \sum_{j=1}^H U(i, j) \right)$$

$$F(F_{Max}, W) = \sigma(W_2 \delta(W_1 F_{Max})), W_1 \in R^{\frac{c}{r} \times c}, W_2 \in R^{\frac{c}{r} \times c}$$

$$F_{Max} = F_{Max} \times (H, W, 1)$$

$$F_{out} = F_{Avg} + F_{Max}$$

本文将通道注意力机制嵌入到加强特征提取网络中,增强不同网络输出的有用信息的识别,更关注特征图中重要的通道信息,弱化了背景信息,进而解决了特征融合过程因钢筋密集造成的遮挡问题。

综上,本文基于YOLOv4的加强特征提取网络,提出如图6所示的attention-CSP-PANet网络结构。通过对GCA-MobileNetV2主干特征网络的(52, 52, 32), (26, 26, 96), (13, 13, 20)维度的特征图,先经过通道注意力的通道信息加权,再经过残差化的5次卷积的特征提取,完成特征融合,以此提高钢筋计数任务的检测精度。

1.4 DepthLite-SPP结构

改进的PANet结构在大小目标不同的特征层上存在正负样本冲突问题,参考当下加强深层感受野的网络结构,围绕视觉传感器实时拍摄钢筋端部,算法的参数量过大,会影响检测的实时性。为进一步降低模型的参数量同时扩大深层感受野,本文提出DepthLite-SPP结构。如图7所示。SPP结构采用 $5 \times 5, 9 \times 9, 13 \times 13$ 的卷积核进

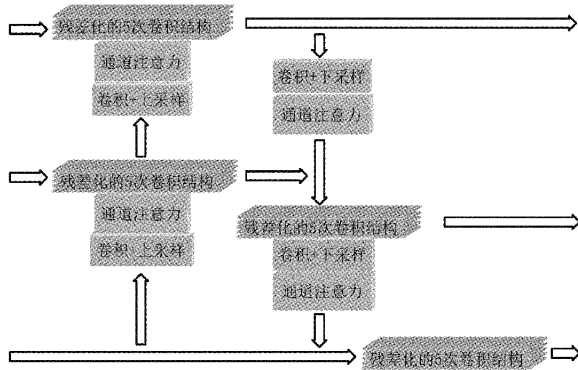


图 6 attention-CSP-PANet 结构

行最大池化,整合来自不同特征金字塔的特征,在一定程度上缓解大小目标在不同特征层上的正负样本冲突。但是面对物体种类单一,数目多的钢筋图像,SPP 结构在模型训练时易出现梯度消失问题。对此,我们根据 CSP 结构,将 SPP 结构先进行一个大的残差,将主干部分进行一个小的残差,将特征图 1/4 的通道进行 SPP 的最大池化,增强感受野,最后与残差边堆叠。同时本文将 3×3 大小的普通卷积替换为深度可分离卷积,与 GCA-MobilenetV2 网络的深度卷积保持一致,避免冗余信息的出现。在 CSP 化之后,新的结构减少了 60% 的参数量。SPP 模块最初被插入到 CSPDarknet53 颈部第一个计算列表组的中间位置。本文将 DepthLite-SPP 模块插入到 GCA-MobilenetV2 颈部的第一个计算列表组的中间位置。采用这种结构,可以有效的解决模型训练时梯度消失问题,达到轻量化的目的。

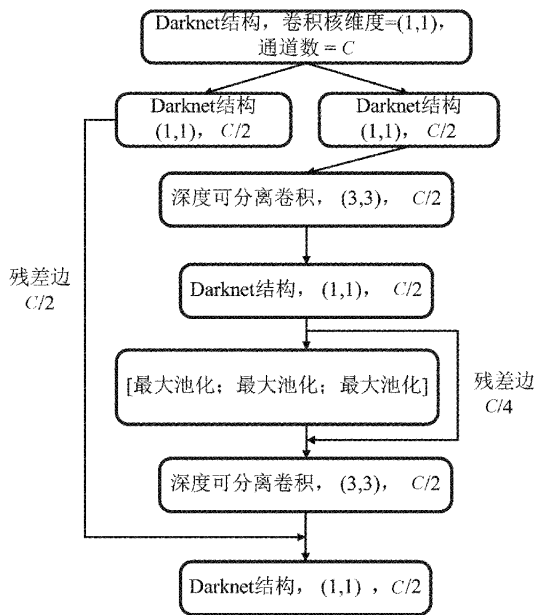


图 7 DepthLite-SPP 结构

1.5 模型训练优化

YOLOv4 算法在模型训练过程中,将梯度函数分为了

锚框损失,分类损失和置信度损失。

锚框损失函数通过计算预测框与真实框之间的差距程度,来进行梯度下降,更新预测值。锚框的真实值与预测值之间的差距越小,模型的拟合度越高。CIOU Loss 通过计算预测框与真实框的中心点距离损失,重叠面积损失和长宽比,使预测框的数学描述接近真实框。但是在分类损失中,一阶段回归的目标检测算法存在正负样本失衡问题,大量的负样本会使输出的损失增加。对此本文参考 Focal-Loss,公式如式(9),式中 y' 为激活函数的输出, f^w 为超参数。

$$l_{cls} = \begin{cases} -\alpha(1-y')^{\gamma} \log y', & y = 1 \\ -(1-\alpha)y'^{\gamma} \log(1-y'), & y = 0 \end{cases} \quad (9)$$

本文将 Focal Loss 作为分类损失函数,增大正样本比例,减小负样本的输出,使损失函数在大量简单样本的迭代过程中不断优化。式(9)中的函数可以通过减少易分类样本的权重,使得模型在训练时更专注于难分类的样本。 f^w 的取值可以调节简单样本权重降低的速率。

在置信度损失中,本文采用二元交叉熵损失,公式如式(10)。 y_i, \tilde{y}_i 分别为置信度的真实值与预测值。

$$l_{conf} = -\sum_{i=1}^n y_i \log \tilde{y}_i + (1-y_i) \log(1-\tilde{y}_i) \quad (10)$$

在锚框损失中,CIOU Loss 在 IOU Loss 的基础上增加了重叠面积损失、中心点距离损失、纵横比,但是未考虑难易样本的平衡问题,因此本文采用 CIOU Loss 与 Focal Loss 相结合的方式,更精确的得到真实框与预测框的距离,更有效的平衡正负样本。总损失如式(11)所示。锚框损失 f^w , 置信度损失 f^w , 分类损失 f^w 。

$$Loss = l_{box} + l_{cls} + l_{conf} \quad (11)$$

2 实验设计

2.1 实验设备及环境

本文采用笔记本电脑 y7000p, win10 操作系统,配置显卡 RTX2060。系统环境 CUDA10.1, GPU 加速 Cudnn7.6.1, python3.7, 深度学习框架 tensorflow2.3, keras2.1.5, 计算机视觉 Opencv-python4.1.1.26。利用 python 中的 socket 包,构建 udp 协议,实现图像视频传输。

2.2 消融实验

本文收集,整理,并标注了 1 245 张 VOC 格式的钢筋数据集。我们采用数据增强的方法进行数据扩充至 6 225 张,以 8:2 的比例划分训练集和测试集。在训练过程中,采用 Adam 优化器,学习率 0.01,使用 Tensorflow 框架下的指数衰减学习率,在非极大值抑制(NMS)中,阈值统一采用 Score:0.01,IOU:0.5。Map 精度计算统一 AP0.5。

为证明本文提出的 GCA-MobilenetV2 结构, attention-CSP-PANet 结构, DepthLite-SPP 结构和 CIOU-Focal Loss 的有效性,在相同环境进行了表 1 所示的消融实验。实验以算法的参数量,平均准确率(Map),算法每秒

表1 本文提出改进方法的效果对比

算法	输入	α	Backbone	参数量	Map%	FPS	
1	YOLOv4	416	CSPDarknet53	64 430 501	95.42	21.44	
2	YOLOv4	416	1	MobilenetV2	38 699 126	94.93	28.00
3	YOLOv4	416	1	CA-MobilenetV2	38 882 562	95.74	27.55
4	YOLOv4	416	1	GCA-MobilenetV2	39 065 998	96.45	27.34
5	YOLOv4	416	1	GCA-MobilenetV2-(attention-CSP-PANet)	28 418 990	98.53	28.27
6	Ours	416	1	GCA-MobilenetV2-(attention-CSP-PANet)-(DepthLite-SPP)	18 133 742	98.66	29.04
7	Ours	416	1	GCA-MobilenetV2-(attention-CSP-PANet)-(DepthLite-SPP)	18 133 742	98.78	29.04

内可以处理的图片数量(FPS),作为评价指标。

表1中,实验1是YOLOv4算法,实验2是采用主干特征网络MobilenetV2的检测结果,MobilenetV2结构中的宽度因子 α 设置为1,相比CSPDarknet53结构检测精度下降了0.49%,检测速度FPS提高了6.56,有效的降低了模型的参数量。

为解决钢筋密集造成的遮挡问题,本文利用改进的CA注意力,加强全局信息的识别与通道信息的捕捉,为证明本文提出的GCA注意力比CA注意力在钢筋数据集中的检测效果更好。实验3、4通过在MobilenetV2网络的同一位置中分别添加CA注意力与GCA注意力。实验表明,在MobilenetV2结构中添加CA注意力,检测精度提高了0.81%。添加GCA注意力后因增加了全局信息的识别,重要性高的通道信息被给予更大的权重,精度提高了1.52%。我们的方法更好地提升了钢筋计数任务的检测精度。

在特征融合中,为进一步加强特征图中重要信息的识别,解决遮挡问题,并降低模型参数量,本文提出了attention-CSP-PANet结构。实验5是在实验4的基础上,将PANet结构替换为本文提出的融合通道注意力的

attention-CSP-PANet结构。本文的方法,使精度提高了2.08%,检测速度FPS提升了0.93。

针对YOLOv4的增强感受野SPP结构参数量大,面对物体数目多,种类单一的情况,在训练时存在梯度消失问题,实验6是本文采用提出的DepthLite-SPP结构增强深层感受野,与实验5的SPP结构相比,精度提升了0.13%,FPS提高了0.77,参数量减少了约11M。

实验7是本文提出的改进算法采用CIoU-Focal Loss的训练方法,与实验6的CIoU Loss相比解决了一阶段回归的目标检测算法正负样本失衡问题,检测精度提高了0.12%。

综上,本文提出的改进算法与YOLOv4算法相比,精度提高了3.36%,检测速度FPS提高了7.6,参数量仅为原算法的1/3。

为进一步证明本文算法在钢筋计数任务中的检测效果,我们在不同的算法中进行比较。如表2所示,表2中的算法为我们在Tensorflow2.3,Keras2.1.5框架下复现的算法。算法均采用Adam优化器,初始学习率0.01,利用指数衰减学习率,优化训练过程。在检测过程,NMS参数设置Score:0.01,IOU:0.5。

表2 当前主流算法的精度对比

算法	输入	网络结构	参数量	Map0.5%	FPS	
1	Faster-RCNN	600	ResNet50	13 331 373	69.73	7.53
2	Faster-RCNN	600	VGG	136 668 019	69.15	6.24
3	Faster-RCNN	600	Se-VGG	136 712 559	69.53	5.93
4	Faster-RCNN	600	ResNet101	38 589 357	70.22	7.08
5	Yolox	640	Darknet	99 096 338	76.27	28.78
6	Yolov3	416	Darknet	62 001 757	84.38	20.91
7	Retinanet	512	ResNet50-fpn	36 382 957	95.81	25.28
8	Multi-YOLOv4	416	Multi-CSP	113 017 037	90.64	12.54
9	YOLOv4	416	MobilenetV1	40 580 342	94.78	27.68
10	YOLOv4	416	MobilenetV3	39 617 914	94.56	27.91
11	Yolov5_x	416	CSP	87 244 374	98.81	20.94
12	Ours	416	GCA-MobilenetV2	18 133 742	98.78	29.04
13	Ours	800	GCA-MobilenetV2	18 133 742	99.34	23.02

前 4 组实验是在两阶段回归的 Faster-RCNN 算法的基础上,通过替换不同的主干网络,验证 Faster-RCNN 算法在钢筋计数任务中的检测效果。实验 3 是我们参考文献^[18]改进的 VGG 算法结构,通过在 VGG 结构中添加通道注意力(squeeze-and-excitation, Se),增强通道信息的识别。通过实验对比,Faster-RCNN 这一两阶段回归的目标检测算法,在钢筋数据集中存在检测精度低,检测速度慢的缺点,不适合在施工现场完成钢筋计数任务。

实验 5 是在 COCO 数据集中效果显著的 YOLOx^[19] 算法,通过实验对比,在钢筋数据集中 YOLOx 算法无锚(anchor-free)的检测方法检测速度快,但检测精度低于实验 6 的 YOLOv3 算法。面对钢筋数据集物体数目多,种类单一的情况,无锚的检测方法增大了网络的召回率,导致检测精度低,因此锚框(anchor-based)检测的方法更适合钢筋计数任务。

实验 7 是参考文献^[9]采用的 Retinanet 算法,主干特征网络为 ResNet50-FPN 结构,预测过程采用柔性非极大值抑制(Soft-NMS^[20])算法评估质量分数。本文算法相比检测精度提高了 2.85%,检测速度 FPS 提高了 3.76。参数量降低了约 1.4 M,有更好的检测效果。

因钢筋规格多样,存在尺度变化,实验 8 是参考文献^[21]多尺度预测的 YOLOv4 算法,通过改进 PANet 的输出,由 3 个输出增加为了 5 个,增强多尺度物体的识别能力,但模型参数量大,检测速度慢,训练时存在梯度消失的问题。钢筋图像的尺度更符合 PANet 结构的 3 个特征图的输出,导致实验 8 的检测精度偏低,对于钢筋计数的效果并不理想。

实验 9、10 是基于 YOLOv4 算法采用的轻量级网络模型 MobilenetV1^[22] 和 MobilenetV3^[23],具有检测速度快, GPU 算力低的特点,与表 1 的实验 2 的检测效果相比, MobilenetV2 结构在钢筋数据集中检测效果更好。因此本文采用 MobilenetV2 结构作为算法的主干特征网络。

实验 11 是 YOLOv5_x^[24] 算法,通过对比本文提出的改进算法,算法检测精度高,检测速度快,参数量仅为 YOLOv5_x 算法的 1/4。当采用 800 的输入维度,精度高达 99.34%,检测速度依然高于 YOLOv5_x 算法。

综上,本文提出的 GCA-MobilenetV2-YOLOv4 算法相比当前的主流算法在钢筋数据集中有更好的检测效果,更适合应用于钢筋计数任务,作为视觉传感器的后端算法。

本文提出的改进方法检测效果提升明显,与原算法的实际效果对比如图 8 所示。我们的算法计数效果要优于

YOLOv4 算法。

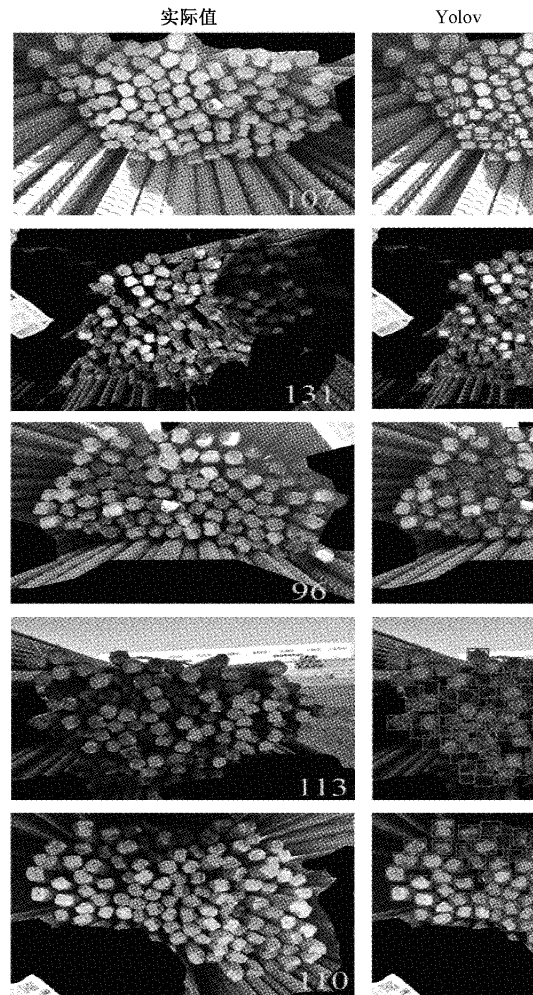


图 8 本文改进算法的实验效果

2.3 计数效果对比

本文采用式(12),平均误差的计算方式在测试集中比较不同算法钢筋计数的效果。 T_i 表示每张图片真实框的数目, P_i 表示每张图片预测框的数目。

$$MAE = \frac{1}{n} \sum_{i=1}^n (T_i - P_i) \tag{12}$$

在实际工作环境中,通过随机拍摄 230 张钢筋图片,共计 30 134 根钢筋,在表 3 不同的算法中进行计数效果对比。本文方法相比原算法 YOLOv4 在 MAE 计算上有 4.44 的提升,在当前的主流算法中有较好的检测效果。实验证明,我们的方法在实际应用时,计数效率高,准确率高,更适用于钢筋计数任务。

表 3 不同算法的钢筋计数结果

算法	Faster-RCNN	Retinanet	Yolov3	YOLOv4	Yolov5_x	Yolox	Ours
MAE	37.66	7.50	13.28	6.72	3.37	34.69	2.28

3 结 论

为提高智慧工地的钢筋计数效率,进一步实践智慧工地理念。本文提出改进的 YOLOv4 目标检测算法进行钢筋计数。为解决钢筋图像多尺度,钢筋密集造成的遮挡问题,提出融合 GCA 注意力的 GCA-MobileNetV2 轻量化结构,与融合通道注意力的 attention-CSP-PANet 特征金字塔结构。在模型训练时,本文提出了 DepthLite-SPP 轻量化结构,改进 YOLOv4 算法的增强感受野 SPP 结构存在梯度消失的问题。在通过对比当前的主流算法,本文提出的改进方法检测精度高达 98.78%,精度提升明显。对比 YOLOv4 提高了 3.36%,当采用(800,800)的输入维度,检测精度高达 99.34%。YOLOv4 在本文的实验环境中 FPS 为 21.44,本文提出的改进算法检测速度为 29.04,同时与当前的研究相比精度和速度提升明显。将改进算法应用于建筑工地作为视觉传感器的后端算法,具有检测精度高,检测速度快,不需要很高的计算机硬件配置的特点,可以提高工地的生产效率。在实际应用中,视频拍摄须正对钢筋表面,如果拍摄角度倾斜严重容易出现漏检的情况,因此未来在视频序列的读取^[25]和倾斜化拍摄处理上有待提高。

参考文献

- [1] 周晓彦,王珂,李凌燕. 基于深度学习的目标检测算法综述[J]. 电子测量技术,2017,40(11):89-93.
- [2] YAN G B, SUN Q, HUANG J Y, et al. Helmet detection based on deep learning and random forest on UAV for power construction safety: Regular papers[J]. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2021,25(1):40-49.
- [3] 于重重,萨良兵,马先钦,等. 基于深度学习的小样本零件表面缺陷检测[J]. 仪器仪表学报,2020,41(7):214-223.
- [4] SANJANA S, SHRIVA V R, VAISHNAVIG, et al. A review on various methodologies used for vehicle classification, helmet detection and number plate recognition [J]. Evolutionary Intelligence, 2021, 14(6): 979-987.
- [5] LIU Y F, LIU H B, PENG J X, et al. Research on the use of YOLOv5 object detection algorithm in mask wearing recognition [J]. World Scientific Research Journal, 2020,6(11):230-238.
- [6] 王志丹. 基于小数据的钢筋检测方法的研究[J]. 邮电设计技术,2020,(2):39-44.
- [7] 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 & Machine Intelligence, 2017, 39 (6): 1137-1149.
- [8] ZHU Y, TANG C, LIU H, et al. End-face localization and segmentation of steel bar based on convolution neural network[J]. IEEE Access, 2020,(8): 74679-74690.
- [9] 明洪宇,陈春梅,刘桂华,等. 基于的密集型钢筋计数改进算法[J]. 传感器与微系统,2020,39(12):115-118.
- [10] YANG L A, YLA C, JCA B. A deep learning approach for real-time rebar counting on the construction site based on YOLOv3 detector [J]. Automation in Construction, 2021, 124 (4): 103602.1-103602.14.
- [11] ANA C H R, JA M N, JD B P. Steel bar counting from images with machine learning[J]. Electronics, 2021,10(4):402.
- [12] WANG C Y, LIAO H. YOLOv4: Optimal speed and accuracy of object detection[J]. Arxiv Preprint, 2020, ArXiv:2004.10934.
- [13] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted residuals and linear bottlenecks [C]. Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [14] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37 (9): 1904-16.
- [15] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for dense object detection[J]. Transactions on Pattern Analysis & Machine Intelligence, 2017(99): 2999-3007.
- [16] TIAN Y, YANG G, WANG Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model [J]. Computers and Electronics in Agriculture, 2019,(157):417-426.
- [17] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [C]. Computer Vision and Pattern Recognition, 2021: 1371-1372.
- [18] 李克文,李新宇. 基于 SENet 改进的 Faster R-CNN 行人检测模型[J]. 计算机系统与应用,2020,29(4): 266-271.
- [19] GE Z, LIU S, WANG F, et al. YOLOX: exceeding YOLO series in 2021 [J]. ArXiv Preprint, 2021, ArXiv:2107.08430.
- [20] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS—Improving object detection with one line of code [C]. 2017 IEEE International Conference on Computer Vision (ICCV), 2017:5561-5569.
- [21] WANG C Y, BOCHKOVSKIV A, LIAO H. Scaled-

- YOLOv4: Scaling cross stage partial network [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021:13029-13038.
- [22] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications [J]. Arxiv Preprint, 2017, ArXiv:1704.04861.
- [23] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3 [C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 2020:1314-1324.
- [24] 王超, 张运楚, 孙绍涵, 等. 改进 YOLOv5 算法的钢筋端面检测 [J]. 计算机系统与应用, 2022, 31(4): 68-80.
- [25] 任克强, 高小琳. 基于五帧差和二维 Renyi 熵的运动目标检测 [J]. 电子测量与仪器学报, 2015, 29(8): 1179-1186.

作者简介

刘浩, 硕士研究生, 主要研究方向为计算机视觉, 机械。

E-mail: 907191698@qq.com

辛山, 博士, 讲师, 主要研究方向为智能建造。