

# 基于目标检测的施工钢材物料智能实时计数

陈 隽<sup>1,2</sup>, 陈文豪<sup>1,3</sup>, 李 洋<sup>1</sup>

(1. 同济大学 土木工程学院, 上海 200092; 2. 同济大学 土木工程防灾减灾全国重点实验室, 上海 200092;

3. 中南建筑设计院股份有限公司, 湖北 武汉 430071)

**摘要:** 研究了基于深度学习目标检测技术的钢材智能计数方法。通过拍摄并标注大量施工现场的钢筋、圆钢管、方钢管图片, 构建了包含近 40 万个计数点的数据集。基于 YOLOv4 目标检测算法, 建立了钢材智能计数模型, 并通过改进其网络结构、损失函数, 以及采用合适的训练策略, 提高了模型对于钢材的计数精度。检验表明, 模型的平均精度为 91.41%, 平均绝对误差为 4.07。利用上述成果开发的 APP 软件, 可通过手机拍照、上传, 完成实时计数。

**关键词:** 目标检测; 智能建造; 施工物料; 实时计数

**中图分类号:** TU712; TP391

**文献标志码:** A

## Intelligent Real-Time Counting of Construction Materials Based on Object Detection

CHEN Jun<sup>1,2</sup>, CHEN Wenhao<sup>1,3</sup>, LI Yang<sup>1</sup>

(1. College of Civil Engineering, Tongji University, Shanghai 200092, China; 2. State Key Laboratory of Disaster Reduction in Civil Engineering, Tongji University, Shanghai 200092, China; 3. Central South Architectural Design Institute, Wuhan 430071, China)

**Abstract:** The intelligent counting method of steel based on the deep learning target detection technology is studied in this paper. First, many pictures of steel bars, round steel pipe, and square steel pipe at the construction site are obtained and marked, and a data set containing nearly 400 000 counting points is constructed. Based on the YOLOv4 object detection algorithm, an intelligent steel counting model is established, and the counting accuracy of the steel model is improved by improving its network structure, loss function, and adopting appropriate training strategies. The maximum average precision of the model is 91.41%, with a mean absolute error of 4.07. Based on the above achievements, an APP software is developed, and the real-time counting is completed by

taking photos and uploading them.

**Key words:** object detection; intelligent construction; construction materials; real-time counting

建筑业是我国国民经济的支柱产业, 但效益低、欠环保、耗能大的现状, 距离高质量发展要求仍有巨大差距, 与世界上发达国家和地区相比, 亟需通过科技创新提升行业的智能化水平解决上述问题。智能建造是信息技术与工程建造深度融合的新模式, 从根本上改变了建筑工程生产与组织方式, 通过促进建造过程中信息的流动和融合来显著提高施工效率。“智能建造”作为一个新专业在 2018 年由同济大学提出, 由于与当今科技发展趋势以及行业的现实需求高度吻合, 目前已有 100 多所高校新增了此专业, 反映出很高的社会认可度, 也是建筑工业化协同发展未来 15 年的主要发展目标之一<sup>[1]</sup>。

施工过程智能化是智能建造全过程的重要环节之一, 而施工中钢筋(主要建材)、钢管(如用于脚手架的圆钢管、支撑体系的方钢管)等物料计数是施工过程管理、成本控制的关键环节。然而, 当前施工实践中对于上述钢材的管理仍主要依赖于人工计数方式, 效率低、成本高、自动化水平差。因此, 钢材物料的智能计数是实现施工智能建造需要解决的关键问题之一, 对照片进行图像处理实现钢材计数是目前最常见的思路。例如, Zhang 等<sup>[2]</sup>提出了一种基于模板匹配和可变阈值分割的在线计数与自动分选系统, 该方法拍摄图像时需要适当角度的辅助光源, 主要适用于工厂等光环境可控的场所。Ying 等<sup>[3]</sup>结合边缘检测器和图像处理算法将钢筋从背景中分离出来, 然后采用霍夫变换对圆形钢筋进行定位。Zhao 等<sup>[4]</sup>使用改进的边缘检测、图像处理算法和边缘聚

收稿日期: 2022-04-07

基金项目: 国家自然科学基金(52178151); 同济大学学科交叉攻关项目(2022-3-YB-06)

第一作者: 陈 隽(1972—), 男, 教授, 博士生导师, 工学博士, 主要研究方向为工程结构大数据防灾。

E-mail: cejchen@tongji.edu.cn



论文  
拓展  
介绍

类算法检测钢筋的数量,该方法需要稳定的检测环境。Su等<sup>[5]</sup>针对拍摄的钢筋图像,采用改进的梯度霍夫圆变换结合最大内切圆算法获取的半径进行钢筋定位。Wu等<sup>[6]</sup>提出一种利用凹点匹配进行分割,K级容错进行计数,最后利用视觉反馈进行多次分切的钢筋在线计数方法。Liu等<sup>[7]</sup>利用Canny算子结合形态学的边缘增强算法来提取感兴趣区域并去除噪声,以此实现圆钢管自动计数。上述工作的核心思想是分离出图像中每根钢材的边界,因此对图像的光照、指定钢筋截面形状、图像背景等条件都有严格的要求。然而,在施工现场所拍摄的钢材图像大多存在缩进不均匀、端部氧化修饰或遮挡、光照不均匀等干扰因素,使得上述方法很难实用。同时,目前尚未开展针对方钢管计数的研究工作,因为方钢管具有双轴对称的特点,其图像处理比轴对称的圆形钢筋和钢管要困难。

近年来,计算硬件性能的飞速提升以及算法研究的突破,使得深度学习在许多领域得到了成功应用,极大地提高了图像中目标检测的准确性。深度学习通过组合简单非线性模块学习非常复杂的函数,与传统的机器学习方法相比具有更强的特征提取能力和泛化能力,从而更易发现海量数据中的复杂结构<sup>[8]</sup>。在众多深度学习网络中,卷积神经网络在处理图像方面有极大的优越性,基于深度学习的目标检测算法能够快速而准确地在图像中给定目标的位置和类别<sup>[9]</sup>。其中,YOLOv4<sup>[10]</sup>是目前广泛使用的单阶段目标检测算法,相比于双阶段目标检测器,在准确率不显著降低的前提下具有更快的响应速

度,适合于实时检测的需求。因此,本文采用目标检测的思路研究钢材智能计数,通过对YOLOv4的改进实现对方钢管、钢筋和圆钢管的实时计数,并开发面向工程使用的云端APP工具。

## 1 模型训练数据集和评价指标

### 1.1 钢材断面照片数据集

基于深度学习的目标检测是典型的数据驱动模型,需要大量的真实样本进行模型训练和性能检验。面向最终的实际应用场景(即实际的光照条件和背景噪声),本研究利用普通智能手机拍摄了大量的实际施工工地的钢筋、圆钢管和方钢管的横截面照片,并对其中的钢材进行标注(图1)。其中,采用水平矩形框来标定照片中每根钢筋、圆钢管的中心坐标及长宽,采用旋转矩形框标定每根方钢管横截面的中心坐标、长宽以及角度。本文最终使用的钢筋、圆钢管和方钢管的照片总数量分别是991、1 019和538张,总计数点分别是181 375、154 044和56 970个(表1)。

先验框是模型训练前预设的接近目标真实大小的框,算法是通过预测先验框的偏移量实现目标检测,合理的先验框设置能够加快模型的训练并提高检测的精度。因此,利用K-means聚类算法将数据集中每种钢材真实框的长宽分为最具代表性的9类,它们的聚类中心作为模型先验框,结果如表1和图2所示。

表1 各类钢材数量统计

Tab. 1 Statistics of all kinds of steel

图片类别	图片数量/ 张	钢材数量/ 根	聚类中心(长,宽)
钢筋	991	181 375	(32,32), (57,58), (89,90), (121,120), (153,152), (189,186), (232,229), (292,290), (395,396)
圆钢管	1 019	154 044	(25,24), (42,41), (60,59), (82,81), (110,108), (147,144), (187,182), (243,235), (316,312)
方钢管	538	56 970	(23,23), (45,40), (60,66), (70,122), (87,49), (102,87), (151,110), (198,201), (325,299)

### 1.2 评价指标选取

目标检测中常用平均精度AP(average precision)、平均最大绝对误差MAE(mean absolute error)或平均均方根误差RMSE(root mean squared error)等指标来评价模型的性能。本文实践表明,采用单一指标评价不适合于钢材计数模型。原因在于,AP值是包含定位和计数信息的一个综合指标,不能直接区分两个因素各自的影响,而MAE、RMSE只包含计数信息,存在虽然数量准确(如漏了

$n$ 个钢筋,而把 $n$ 个其他物品计为钢筋)但实际是误检的情况。因此,本研究综合采用AP、MAE和RMSE这3个指标,各自的定义如下:

$$AP = \sum_{r=0}^1 (r_{n+1} - r_n) \max_{r \geq r_{n+1}} p(\bar{r}) \quad (1)$$

式中: $n$ 为测试集中的图片数量; $p(\bar{r})$ 为在召回率为 $\bar{r}$ 时的测试精度,使用预测框与真实框之间的交并比衡量定位的准确性,在本研究中使用 $AP_{50}$ ,即IOU(intersection over union)阈值为0.5来测试模型的检测效果。

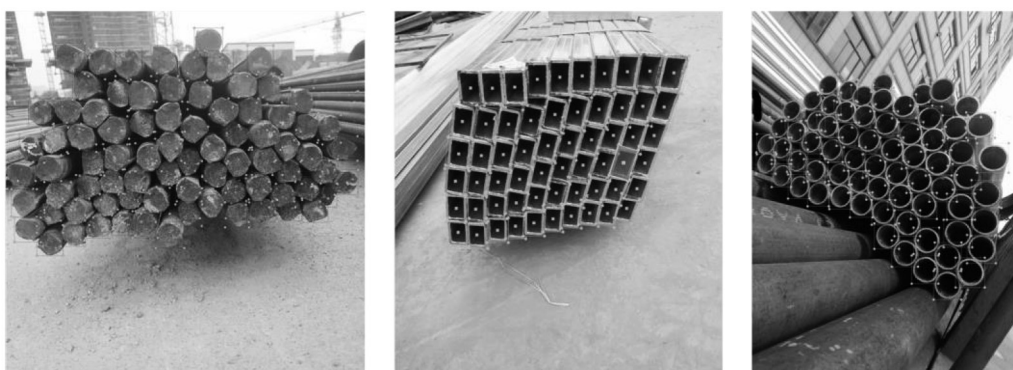


图 1 钢材标注示例

Fig. 1 Example of steel labeling

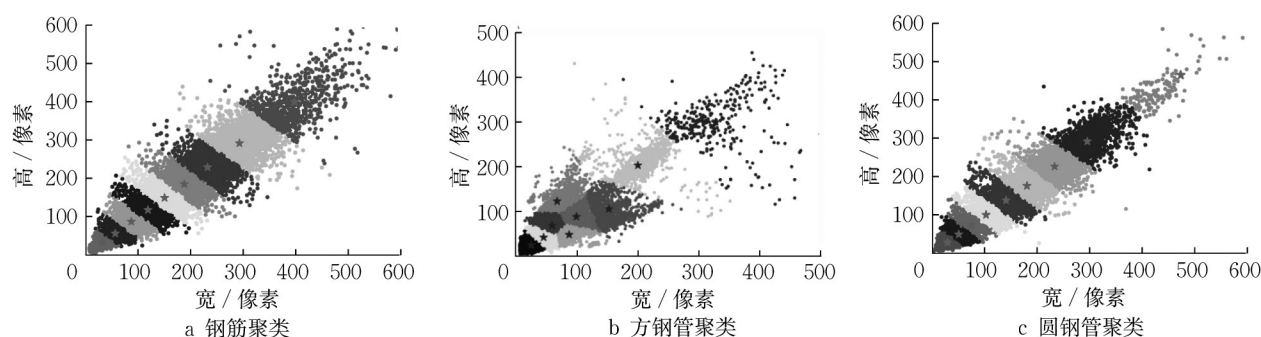


图 2 钢材聚类结果

Fig. 2 Result of steel clustering

MAE用于测试模型计数的准确性,而RMSE则用于测试模型计数的稳定性,定义如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

式(2)、(3)中: $y_i$ 为图片中目标的真实数量; $\hat{y}_i$ 为模型检测数量。

相比于相对误差,MAE能够直观地反映模型漏检的根数,同时不受单张图片中方钢管数量的影响,而且在实际使用场景中,统计人员也更关注漏检的根数。

## 2 方钢管计数模型

3类钢材计数中,由于方钢管种类繁多,拥有不同的壁厚、长宽比和排列方式,因此计数最为困难。考虑到篇幅,本节以方钢管为例说明计数模型的建立,前述方钢管数据集以4:1划分为训练集和测试集。

YOLOv4是目前使用很广泛、计算速度快的单

阶段目标检测算法,适合于钢材实时计数的需求。然而,虽然YOLOv4检测自然场景下水平物体的性能十分出色,但直接应用于具有较大纵横比且密集排列的方钢管却表现欠佳。这是因为,对施工现场非常常见的方钢管倾斜情况,照片采用水平标注框后产生的冗杂区域和各重叠区域将变得相对较大(相比之下,钢筋、圆钢管的重叠和冗杂区域小),如图3所示。冗杂区域会引入大量噪声,导致图像信息特征受到干扰甚至消失,较大的重叠区域则会导致在进行非极大值抑制后部分对象被抛弃,因此需要对标准YOLOv4(下称原模型)的目标检测算法进行改进,使其能够检测带有角度的方钢管并进行计数。

### 2.1 损失函数的改进

原模型将输出的特征图划分为不同数量的栅格单元,每个栅格单元有3个预测边界框,每个边界框包含坐标、置信度和类别信息,并采用CIOU(complete intersection over union)<sup>[11]</sup>计算位置损失,CIOU考虑了边界框回归过程中目标框的重叠区域、中心点的距离以及长宽比对训练的影响,如式(4)、(5)所示。

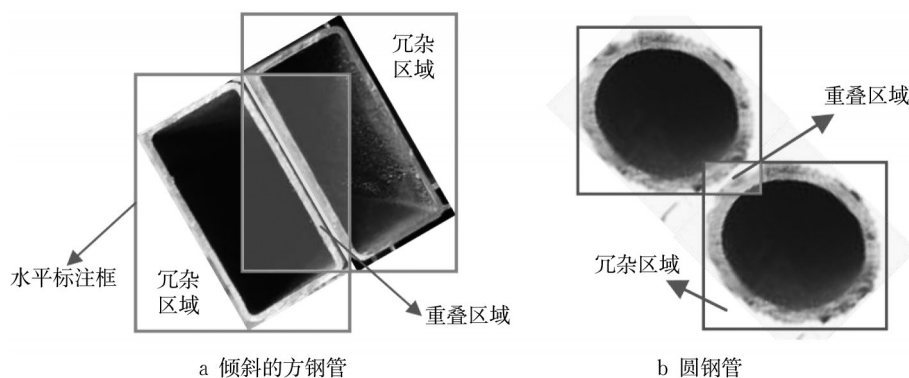


图3 水平框标注方钢管和圆钢管

Fig. 3 Detection of square steel tube and circle pipe using horizontal frame

$$L_{\text{ciou}} = 1 - \text{IOU} + \frac{d_c^2(b_p, b_s)}{d_e^2} + \frac{4\nu^2}{\pi^2((1 - \text{IOU}) + \nu)} \quad (4)$$

$$\nu = \arctan \frac{w_g}{h_g} - \arctan \frac{w_p}{h_p} \quad (5)$$

式(4)、(5)中:  $b_p$  和  $b_g$  分别为预测框和真实框的中心点;  $d_c$  为两个框中心点之间的距离;  $d_e$  为两个框相隔最远两点间的距离;  $w_g$  和  $h_g$  分别为预测框的高和宽;  $w_p$  和  $h_p$  分别为真实框的高和宽。各符号含义见图4。

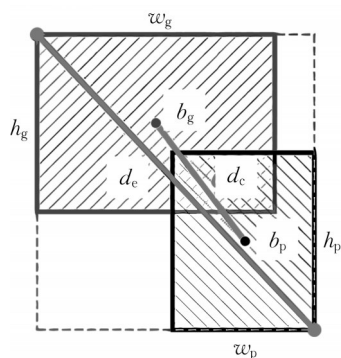


图4 CIOU 损失示意

Fig. 4 CIOU loss

式(4)表明CIOU损失中不包含角度信息,因此对旋转框进行边界框回归,必须对旋转框的角度进行定义并采用新的位置损失函数。在二维直角坐标系中旋转矩形框通常有两种定义方式,分别是OpenCV定义法和长边定义法,前者将与横轴成锐角的边定义为宽,角度范围为 $[0^\circ, 90^\circ]$ ,后者将较长的边定义为宽,角度范围为 $[-90^\circ, 90^\circ]$ 。经尝试,本文最终采用长边定义法。

目标检测的位置损失函数用于计算预测框与真实框之间差异。具体实现时,将旋转框的参数转化

为高斯分布的数字特征(即均值和方差),采用KL (Kullback-Leibler) 散度和高斯-瓦瑟斯坦距离 (Gaussian-Wasserstein distance, GWD)<sup>[12]</sup> 衡量两个二维高斯分布之间的差异作为YOLOv4的位置损失,并采用公式(6)~(7)将旋转框的定义参数转化为高斯分布的数字特征(图5),这种转换方式能够有效解决因角度周期性导致的损失不连续问题和长短边互换导致的边界不连续问题<sup>[13]</sup>。

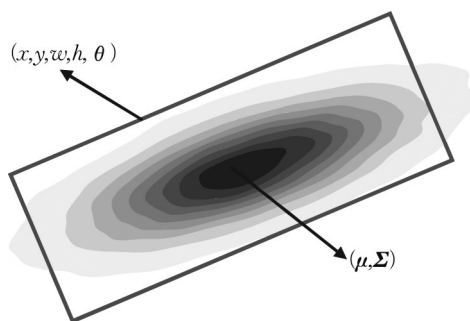


图5 二维高斯分布模拟旋转矩形框

Fig. 5 Modeling of a rotating bounding box by two-dimensional Gaussian distribution

$$\Sigma^{1/2} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \frac{w}{2} & 0 \\ 0 & \frac{h}{2} \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} = \begin{pmatrix} \frac{w}{2} \cos^2 \theta + \frac{h}{2} \sin^2 \theta & \frac{w-h}{2} \cos \theta \sin \theta \\ \frac{w-h}{2} \cos \theta \sin \theta & \frac{w}{2} \sin^2 \theta + \frac{h}{2} \cos^2 \theta \end{pmatrix} \quad (6)$$

$$\mu = (x, y) \quad (7)$$

式中:  $\mu$ 、 $\Sigma$  分别为二维高斯分布的均值向量、协方差矩阵;  $x$ 、 $y$ 、 $w$ 、 $h$ 、 $\theta$  分别为旋转矩形框的横坐标、纵坐标、宽、高、角度。二维高斯分布转化完成后,用公式

(8)~(11)分别计算KL散度 $D_{kl}$ 和GWD距离 $D_{gw}$ 对应的损失。

$$D_{kl}(N_p||N_t) = \frac{1}{2}(\mu_p - \mu_t)^T \Sigma_t^{-1}(\mu_p - \mu_t) + \frac{1}{2}\text{Tr}(\Sigma_t^{-1}\Sigma_p) + \frac{1}{2}\ln\frac{|\Sigma_t|}{|\Sigma_p|} - 1 \quad (8)$$

$$l_{kl} = 1 - \frac{1}{\tau + \sqrt{D_{kl}}} \quad (9)$$

$$D_{gw}(N_p, N_t)^2 = \|\mu_p - \mu_t\|_2^2 + \text{Tr}(\Sigma_p + \Sigma_t - 2(\Sigma_p^{1/2}\Sigma_t\Sigma_p^{1/2})^{1/2}) \quad (10)$$

$$l_{gw} = 1 - \frac{1}{\tau + \sqrt{D_{gw}}} \quad (11)$$

式中: $N$ 表示二维正态分布; $\mu$ 、 $\Sigma$ 分别为对应分布的均值向量和协方差矩阵;下标 $p$ 、 $t$ 分别表示预测框和真实框; $\text{Tr}$ 表示取矩阵的迹; $\|\cdot\|_2$ 表示向量的二范数; $l_{kl}$ 和 $l_{gw}$ 分别为基于KL散度、高斯-瓦瑟斯坦距离的损失函数; $\tau$ 为可调参数,在本文中 $\tau$ 取为2。

YOLOv4原模型的3个先验框均为水平,用来拟合方钢管的旋转真实框会加大模型的训练难度,导致检测精度降低。对此,本文将原来的每个先验框增加6种角度,分别为 $-60^\circ$ 、 $-30^\circ$ 、 $0^\circ$ 、 $30^\circ$ 、 $60^\circ$ 、 $90^\circ$ ,特征图上每个网格点共有18个先验框,虽然增加了网络检测头的厚度,但是能有效提高模型的检测精度。

此外,训练模型时需要从所有的先验框中区分正负样本,每个网格点上的正样本需要满足先验框与真实框的IOU大于某个阈值,并且其值也必须是该网格点上最大的。在水平目标检测中计算IOU是简单快速的,然而旋转框之间的IOU在训练阶段计算较为费时,故采用近似交并比ArIOU来计算训练阶段的IOU,其计算方法见式(12),同时对正样本 $S$ 的定义也有所改变,见式(13)。

$$\text{ArIOU}(T, A) = \text{IOU}(T, A^*)|\cos(\theta_T - \theta_A)| \quad (12)$$

式中: $T$ 表示真实框; $A$ 表示先验框; $\theta_T$ 和 $\theta_A$ 分别表示真实框和先验框的角度; $A^*$ 表示将角度调整为 $\theta_T$ 的先验框 $A$ ;IOU表示求交并比。

$$S = \begin{cases} 1 & \text{ArIOU} > \alpha \\ 1 & \beta < \text{ArIOU} < \alpha, |\theta_T - \theta_A| > \gamma \\ 0 & \text{其他} \end{cases} \quad (13)$$

式中:1表示正样本;0表示负样本; $\alpha$ 、 $\beta$ 、 $\gamma$ 为可调参数,本文分别取为0.6、0.4、 $15^\circ$ 。

未改进的计数模型在训练的过程中的置信度函数采用二分类交叉熵损失,为了减少正样本与负样本之间的不均衡性对模型训练的影响,并增强模型对困难样本的敏感度,将原置信度损失换为Focal损失<sup>[14]</sup>,其计算方式如下:

$$L_{\text{conf}} = \begin{cases} -(1-y')^\gamma \lg y' & y=1(\text{前景}) \\ -y'^\gamma \lg(1-y') & y=0(\text{背景}) \end{cases} \quad (14)$$

式中: $y$ 和 $y'$ 分别为真实置信度和预测置信度;可调参数 $\gamma$ 用于平衡难易样本的重要性,在训练过程中能够自动降低简单背景对训练权重的贡献度,使模型在训练时快速聚焦于学习困难的负样本,在训练过程中取为2。

## 2.2 网络结构的改进

在计算机视觉的目标检测任务中,一旦某个目标与某个特定图相关联,其他的特征图的相应位置将被视为背景,不同层次的特征就会产生冲突,干扰模型训练并降低特征提取的有效性。对此,采用注意力机制能使网络聚焦于重要的特征从而提高模型的精度。ASFF(adaptively spatial feature fusion)<sup>[15]</sup>能够将不同分辨率的特征图融合为一个固定分辨率的特征图,从而减少由于大物体与小分辨率的特征图相关而小物体又与大分辨率的特征图相关造成的不同尺度特征之间的不一致性。因此,本研究在模型中加入了ASFF使其直接在不同分辨率上对特征进行筛选,选择有效信息进行组合,从而提升模型的性能。通过实验对比了不同注意力机制下模型的性能,最终确定采用SE(squeeze and excitation)<sup>[16]</sup>注意力机制并加入ASFF的情况下性能最优,整个网络结构如图6所示,其中SE模型如图7所示,ASFF图形如图8所示。

## 3 模型训练策略选择及实现

不同的深度学习模型通常采用的训练策略各不相同,好的训练策略能够极大提高模型的检测效果。常用的训练策略有数据增强、学习率调整计划、迁移学习以及多尺度训练。

数据增强通过翻转、平移、旋转等对原始图片进行几何变换或通过颜色空间变换进行色彩增强,用于丰富训练集,使模型能够对不同的图片拥有更强的鲁棒性,如图9所示操作。由于本研究收集到的方钢管的图片有限,而实际场景中方钢管会有各种

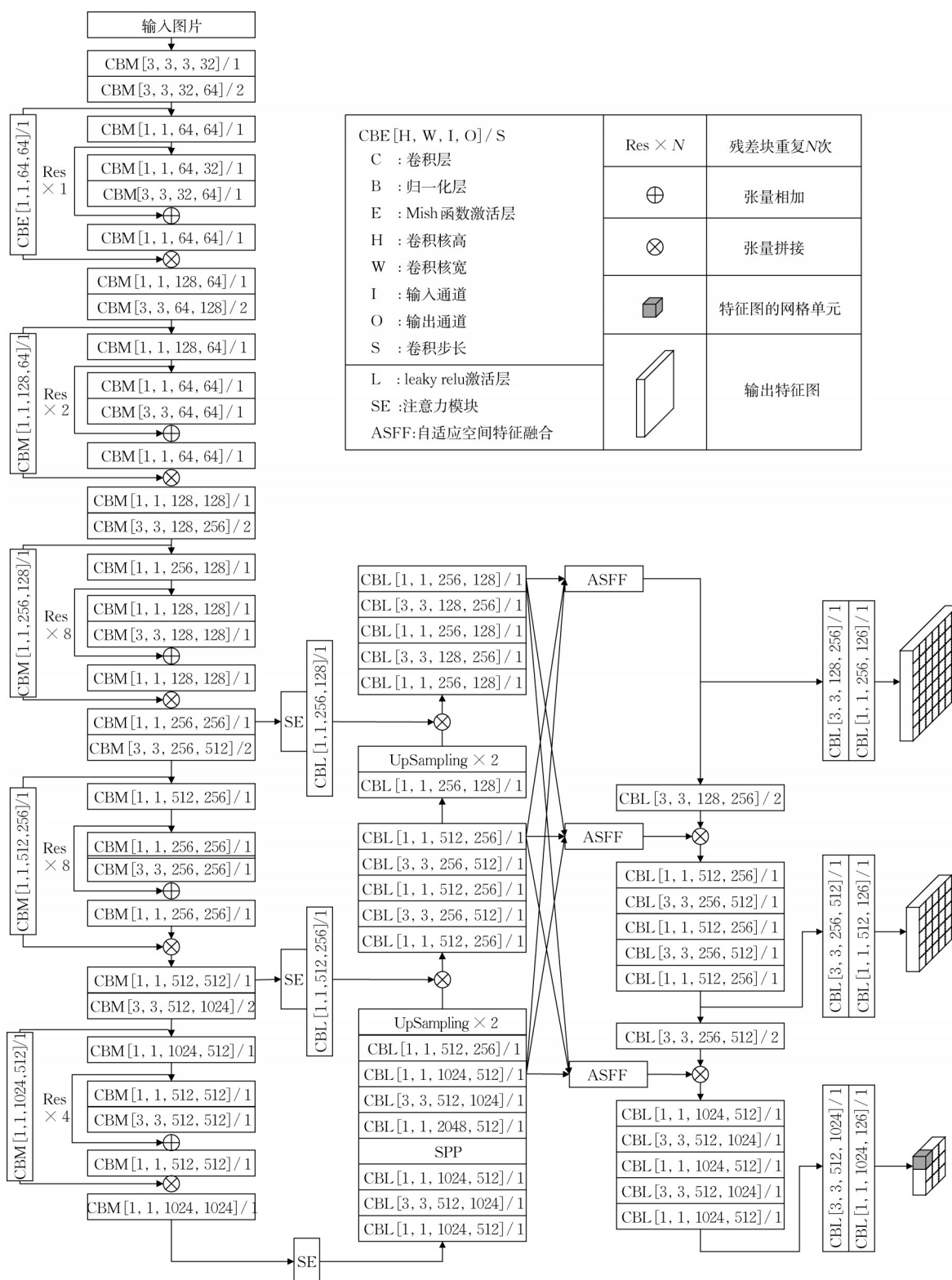


图 6 方钢管计数模型结构图

Fig. 6 Network of counting model of square steel pipe

角度的摆放,使用旋转进行增强能让模型在有限的数据下训练出检测各种角度方钢管的能力,提高模型的泛化能力。本研究在模型训练前先对数据集中一半的图片进行了随机的几何变换,然后再随机进

行色彩增强。

迁移学习是从某个领域的任务中提取有用的特征并将其应用到新的任务中。直接训练深度学习模型而不使用迁移学习往往会训练出效果不理想的模

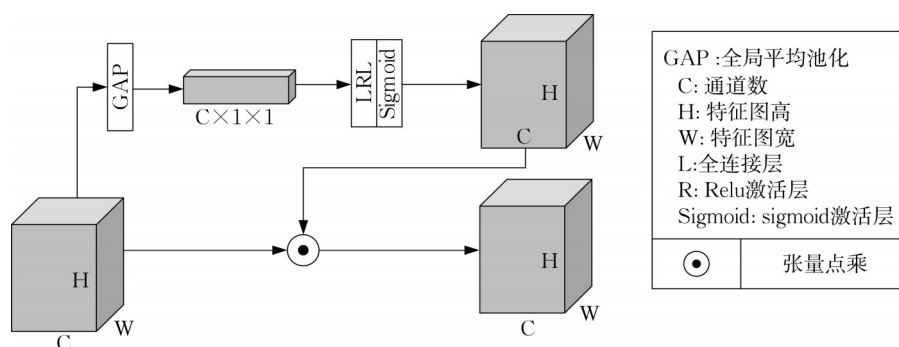


图 7 SE 模块的网络结构

Fig. 7 Network of SE module

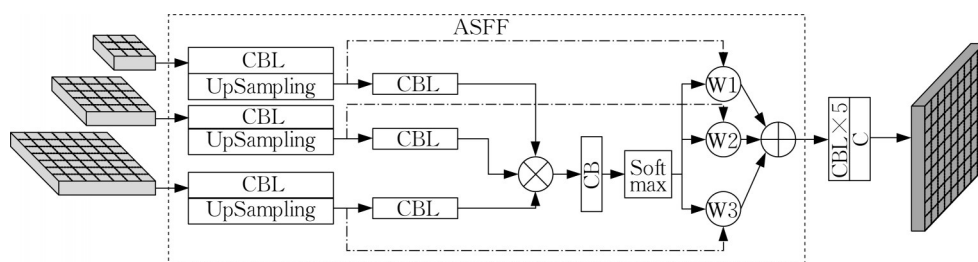


图 8 ASFF 的网络结构

Fig. 8 Network of ASFF module

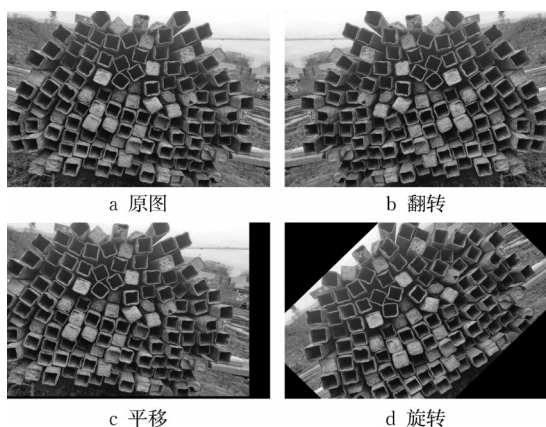


图 9 图片数据增强

Fig. 9 Data augmentation of photo

型,因此在训练方钢管计数模型过程中采用 COCO 数据集上的预训练权重来进行权重初始化。由于方钢管数据集的图片数量不够充足,以及方钢管种类和形态上的较大差异,同时考虑到圆钢管和方钢管在特征上具有一定的相似性,将模型在圆钢管数据集上训练的权重作为方钢管计数模型的初始权重。

学习率是深度学习模型训练时控制模型权重进行梯度更新的一个超参数,取值过高会使模型在训练过程中难以收敛,太低又会使模型收敛的速度过慢,制定合适的学习率调整策略对模型的训练十分关键。本研究采用余弦退火的方式进行学习率的调

整,如图 10 所示,余弦退火函数如下:

$$\eta_i = \eta_{\min}^i + \frac{1}{2}(\eta_{\max}^i - \eta_{\min}^i) \left( 1 + \cos \left( \frac{T_{\text{cur}}}{T_i} \pi \right) \right) \quad (15)$$

式中: $\eta_i$  为当前的学习率; $\eta_{\max}^i$  和  $\eta_{\min}^i$  分别为学习率的最大值和最小值; $i$  为循环序号; $T_{\text{cur}}$  为当前已经执行完的训练次数; $T_i$  为第  $i$  次循环时总的训练轮次。在本研究将最大、最小学习率分别设置为 0.000 2 和 0。

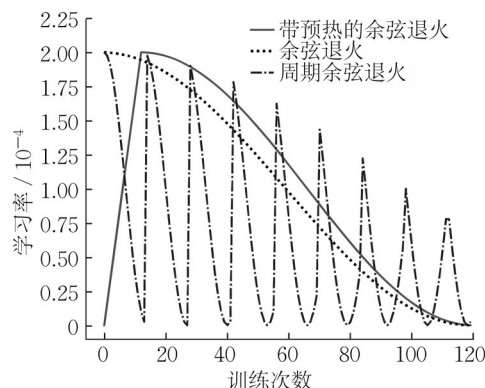


图 10 学习率曲线

Fig. 10 Learning rate

表 2 是如图 6 所示的模型在训练时是否采用数据增强以及学习率调整时的 AP 值,通过对比可以看出在模型训练时采用数据增强和合适学习率调整策

表2 不同训练策略下的模型AP值

Tab. 2 AP value of model in different training strategies

数据增强	学习率调整	AP
N	N	77.20
Y	N	86.80
N	Y	88.40
Y	Y	91.41

注:表中Y、N分别表示训练时是否加入相应的方式。

略能够明显提升模型的性能,AP值提升幅度达22%。图片被输入模型之前会被调整到某一固定尺度,尺度太大会使机器的显存溢出,尺度太小难以满足训练精度的要求,因此需要将图片调整到适当的尺度(即分辨率大小)。实践表明,采用固定尺度训练的模型泛化能力较差,因此本文使用了多尺度训练来提高模型的泛化能力,每次输入前从{416,448,

480,512,544,576,608}中随机选择一种尺度,然后将图片调整到该尺度下进行训练,实践表明,多尺度训练策略有效。

## 4 结果与应用扩展

### 4.1 结果分析

表3比较了在网络结构中是否采用自适应特征融合、不同注意力机制、不同损失函数下的方钢管计数模型的测试效果。在实验过程中,将538张方钢管图片按4:1划分为训练集和测试集,训练轮次设置为120,初始学习率设置为0.0002,实验过程中采用相同的数据增强和学习率调整方式,在测试阶段图片的输入尺度为608,置信度损失采用Focal损失。

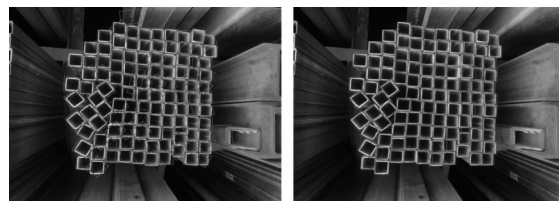
表3 各种网络结构的测试结果对比

Tab. 3 Comparison of test results of various neural network structures

序号	注意力模块	ASFF	GWD			KL		
			AP	MAE	RMSE	AP	MAE	RMSE
1	PSA <sup>[17]</sup>	N	85.67	4.72	7.36	85.01	3.87	6.79
2	CBAM <sup>[18]</sup>	N	88.42	3.73	5.62	86.60	4.22	6.18
3	SE <sup>[16]</sup>	N	86.84	4.37	6.62	85.48	4.25	6.00
4	Triple attention <sup>[19]</sup>	N	87.17	3.65	6.17	86.67	4.00	7.00
5	PSA	Y	84.45	3.92	5.79	84.78	4.50	7.26
6	CBAM	Y	88.50	3.65	5.88	85.76	3.98	6.63
7	SE	Y	91.41	4.07	5.85	83.80	4.45	6.74
8	triple attention	Y	84.99	4.70	7.40	83.20	4.21	6.80

注:表中Y、N分别表示在网络中是否加入ASFF机制;GWD和KL分别表示使用高斯距和KL散度作为位置损失函数。

表3结果表明,模型在采用不同位置损失函数和网络模块时的精度差异较大,AP差值最高达到8.21,MAE的差值则为1.05。其中,采用GWD作为位置损失函数时的AP值普遍高于采用KL散度作为位置损失时的精度。对比2.1节中两种损失函数的公式可知,KL散度AP值较低的主要原因是其不对称性,因此在预测框与真实框不变时对调其位置后损失值会发生改变。在模型采用SE注意力机制(图7)和ASFF(图8)并在位置损失函数采用GWD的情况下模型的最高平均精度达到了91.41%,平均绝对误差为4.07。上述结果表明本文针对YOLOv4原模型的改进措施非常有效,模型损失函数改进前后的检测效果如图11所示,可见更改损失函数前存在错检,并且检测框的大小和角度存在较大误差,而更改后模型的检测准确率有明显提升,错检漏检的情况得到极大改善。最终模型在两个较难场景(散乱和密集)的检测效果示例见图12。



a 损失函数修改前 b 损失函数修改后

图11 损失函数修改前后效果对比

Fig. 11 Comparison of result before and after modification of loss function

### 4.2 钢筋及圆钢管拓展

上述方钢管计数模型的建立思路可以直接拓展至钢筋和圆钢管,并分别开发了各自的实时计数模型<sup>[20-21]</sup>,建模过程不再赘述,计数结果示例见图13。

### 4.3 实用性检验

为检验上述理论研究成果的实用性,真正解决现实工程问题,研究开发了基于手机的“钢材智能计数”APP并已上线,用户只需从应用商店下载注册即

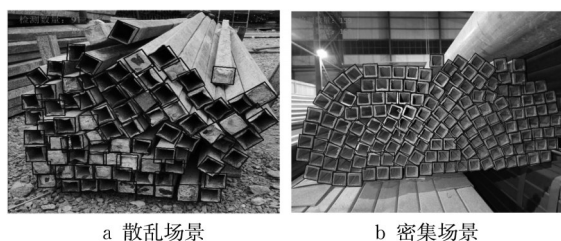


图 12 方钢管计数模型检测结果

Fig. 12 Detection results of square steel pipe counting model

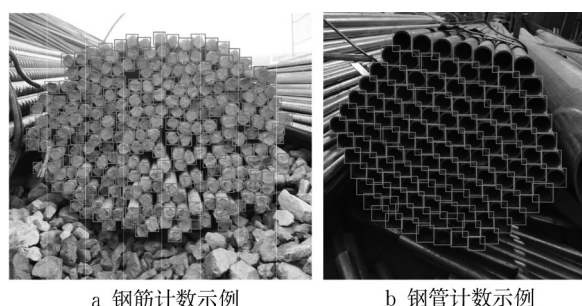


图 13 钢筋和钢管计数结果

Fig. 13 Counting results of rebar and steel pipe

可使用,APP 首页如图 14 所示。使用时,用户将手机拍摄相应钢材端面图像上传云端服务器,即可完成数量计算,整个计算和结果反馈过程一般在 1~2 s 内完成,满足实时性要求。APP 自上线以来,已吸引到 26 000 多注册用户,已完成 90 000 多照片的计数任务,实际使用效果良好。



图 14 钢材智能计数 APP 首页

Fig. 14 Frontpage of intelligent steel counting APP

## 5 结论

本文利用深度学习中的目标检测技术研究钢材的智能计数,通过改进 YOLOv4 模型来提升精度,在此基础上开发了实用 APP,以解决建筑工地钢材人工统计费时费力的问题。研究得出以下结论:

(1)不同的钢材计数任务需要不同的检测模型,本研究针对钢筋、圆钢管以及方钢管开发了满足计数精度的不同模型。

(2)对于含有角度影响的方钢管计数,本研究在网络中加入了 SE 注意力机制、ASFF 模块以及针对有角度物体的损失函数,并结合数据增强和学习率调整等策略显著提升了模型精度,在普通场景下方钢管计数模型能够达到 90% 以上的平均精度和 4.07 的平均绝对误差。

(3)研究成果形成了实用的 APP,并已拥有大量的用户,可为实际的建设工程节省人力物力。

本文的研究仍有一些不足:因训练数据有限,方钢管计数模型在复杂场景下的计数能力仍有待提升;虽然实现了对建筑施工主要钢材的智能实时计数,但对其他材料(如工字钢、模板、砌块)的计数仍需新的模型,用同一个模型实现不同材料的计数是未来值得研究的问题。

### 作者贡献声明:

陈 隽:制定论文总体思路,论文文稿修改。

陈文豪:模型构建,论文撰写和修改。

李 洋:协助实验,提供撰写建议。

### 参考文献:

- [1] 廖玉平. 加快建筑业转型 推动高质量发展——解读《关于推动智能建造与建筑工业化协同发展的指导意见》[J]. 中国勘察设计, 2020(9): 20.  
LIAO Yupin. Speed up the transformation of construction industry and promote high-quality development-- interpretation of 《The Guidance on Promoting the Coordinated Development of Intelligent Construction and Construction Industrialization》 [J]. Survey and Design in China, 2020(9): 20.
- [2] ZHANG D, XIE Z, WANG C. Bar section image enhancement and positioning method in on-line steel bar counting and automatic separating system[C]//2008 Congress on Image and Signal Processing. Piscataway: IEEE, 2008, 2: 319-323.

- [3] YING X, WEI X, PEI X Y, *et al.* Research on an automatic counting method for steel bars' image [C]//2010 International Conference on Electrical and Control Engineering. Piscataway: IEEE, 2010: 1644-1647.
- [4] ZHAO J, XIA X, WANG H, *et al.* Design of real-time steel bars recognition system based on machine vision [C]//2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC). Piscataway: IEEE, 2016, 1: 505-509.
- [5] SU Z, FANG K, PENG Z, *et al.* Rebar automatically counting on the product line [C]//2010 IEEE International Conference on Progress in Informatics and Computing. Piscataway: IEEE, 2010, 2: 756-760.
- [6] WU Y, ZHOU X, ZHANG Y. Steel bars counting and splitting method based on machine vision [C]//2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER). Piscataway: IEEE, 2015: 420-425.
- [7] LIU Y, LIU Y, SUN Z. Research on stainless steel pipes auto-count algorithm based on image processing [C]//2012 Spring Congress on Engineering and Technology. Piscataway: IEEE, 2012: 1-3.
- [8] LECUN Y, BENGIO Y, HINTON G, *et al.* Deep learning [J]. *Nature*, 2015, 521(7553): 436.
- [9] ZHAO Z Q, ZHENG P, XU S T, *et al.* Object detection with deep learning: a review [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 30(11): 3212.
- [10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [J]. *arXiv*: 2004.10934, 2020.
- [11] ZHENG Z, WANG P, LIU W, *et al.* Distance-IoU loss: faster and better learning for bounding box regression [C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. Menlo Park: AAAI, 2020, 34(7): 12993-13000.
- [12] YANG X, YANG X, YANG J, *et al.* Learning high-precision bounding box for rotated object detection via kullback-leibler divergence [J]. *Advances in Neural Information Processing Systems*, 2021, 22: 18381.
- [13] YANG X, YAN J, MING Q, *et al.* Rethinking rotated object detection with gaussian wasserstein distance loss [C]// *International Conference on Machine Learning*. [S.l.]: PMLR, 2021: 11830-11841.
- [14] 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*. Piscataway: IEEE, 2017: 2980-2988.
- [15] LIU S, HUANG D, WANG Y. Learning spatial fusion for single-shot object detection [J]. *arXiv*: 1911.09516, 2019.
- [16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2018: 7132-7141.
- [17] ZHANG H, ZU K, LU J, *et al.* Epsanet: an efficient pyramid split attention block on convolutional neural network [J]. *arXiv*: 2105.14447, 2021.
- [18] WOO S, PARK J, LEE J Y, *et al.* Cbam: convolutional block attention module [C]// *Proceedings of the European Conference on Computer Vision*. Munich: ECCV, 2018: 3-19.
- [19] MISRA D, NALAMADA T, ARASANIPALAI A U, *et al.* Rotate to attend: convolutional triplet attention module [C]// *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. Piscataway: IEEE, 2021: 3139-3148.
- [20] LI Y, LU Y, CHEN J. A deep learning approach for real-time rebar counting on the construction site based on YOLOv3 detector [J]. *Automation in Construction*, 2021, 124: 103602.
- [21] LI Y, CHEN J. Computer vision - based counting model for dense steel pipe on construction sites [J]. *Journal of Construction Engineering and Management*, 2022, 148 (1) : 04021178.