

DOI:10.19651/j.cnki.emt.2208810

# 基于深度学习的钢筋绑扎机器人目标识别定位<sup>\*</sup>

董国梁 张雷 辛山

(北京建筑大学电气与信息工程学院 北京 100044)

**摘要:**为了解决钢筋绑扎机器人对绑扎点识别准确率低,定位精度差的问题,提出一种基于深度学习的钢筋绑扎机器人目标识别与定位方法。首先采用YOLOv4算法对绑扎点目标框识别和裁剪,完成绑扎点初始定位;其次设计轮廓角点选取方法,利用角点计算绑扎点的图像坐标;之后通过融入CBAM注意力机制改进Monodepth算法的特征提取部分,解码部分引入路径增强PAN结构,以提高模型的特征提取能力,进一步提高立体匹配精度;最后通过双目立体视觉定位技术获得绑扎点深度信息,并由坐标变换求解钢筋绑扎机器人手眼坐标系映射关系,从而实现绑扎点的精确识别和定位。实验结果表明:该方法针对绑扎点目标框的识别准确率达到99.75%,每秒传输帧数达到54.65;在空间中的定位精度最大误差为11.6mm。可较好地识别定位绑扎点位置,为自动绑扎工作提供有力支持。

**关键词:**钢筋绑扎机器人;深度学习;目标识别;特征提取;深度估计;双目视觉

**中图分类号:**TP391.4 **文献标识码:**A **国家标准学科分类代码:**520.6040

## Target recognition and location of steel bar binding robot based on deep learning

Dong Guoliang Zhang Lei Xin Shan

(School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China)

**Abstract:** In order to solve the problem of low recognition accuracy and poor positioning accuracy of steel bar binding robot, a target recognition and positioning method of steel bar binding robot based on deep learning is proposed. Firstly, YOLOv4 algorithm is used to identify and cut the target frame of binding point, and the initial positioning of binding point is completed. Secondly, the contour corner selection method is designed to calculate the image coordinates of the binding points by corner points. Then the feature extraction part of Monodepth algorithm is improved by integrating CBAM attention mechanism, and the decoding part introduces path aggregation network structure to improve the feature extraction ability of the model and further improve the stereo matching accuracy. Finally, the depth information of binding points is obtained by binocular stereo vision positioning technology, and the coordinate transformation is used to solve the mapping relationship between the hand-eye coordinate system of the steel bar binding robot, so as to realize the accurate identification and positioning of binding points. The experimental results show that the recognition accuracy of this method for binding point target frame reaches 99.75%, and the number of frames per second reaches 54.65. The Maximum error of positioning accuracy in space is 11.6 mm. It can better identify and locate the binding point position, and provide strong support for automatic binding work.

**Keywords:** steel bar binding robot; deep learning; object identification; feature extraction; depth estimation; binocular vision

## 0 引言

近年来建筑业存在建筑工人紧缺、老龄化等问题,目前绑扎钢筋仍需要大量的人力完成,且最常见的建筑工人伤害类型就是因为长时间绑扎钢筋造成的肌肉骨骼损伤<sup>[1]</sup>,

因此设计使用机器人代替人工变得尤为重要。

自动钢筋绑扎机器人系统通过视觉识别钢筋交叉中心点,即钢筋绑扎点,并定位绑扎点空间位置,由绑扎模块完成绑扎。其中,计算机视觉识别定位绑扎点是钢筋绑扎的关键问题之一。目前,在自动钢筋绑扎机器人工作方面,美

收稿日期:2022-01-09

<sup>\*</sup> 基金项目:智能机器人与系统高精尖创新中心建设项目(00921917001)资助

国 Advanced Construction Robotics 公司的 Cardno<sup>[2]</sup> 提出了 Tybot 钢筋绑扎机器人,通过二维码标签识别定位钢筋绑扎点;日本千叶工业大学的“T-iROBO Rebar”自动绑扎机器人通过激光检测器识别定位钢筋绑扎点;在我国, Jin 等<sup>[3]</sup> 提出了采用深度图像来识别定位绑扎位置。相比于二维码标签和激光识别方法因图像信息缺失易产生较大误差问题,使用计算机视觉可以获得更丰富的图像信息和更直观的可视化结果,并且双目相机比深度相机对室外施工的光照强度抗干扰效果更好。虽然目前使用计算机视觉识别定位钢筋绑扎点的应用研究还较少,但是针对于建筑机器人方面,国内外研究学者提出了诸多相关实现方法。其中, Wang 等<sup>[4]</sup> 设计了一种现场拆建垃圾分类回收机器人,利用 Mask R-CNN 语义分割算法完成建筑拆迁废物的识别和姿态的确定,以此完成拣选工作; Wang 等<sup>[6]</sup> 设计了一种建筑垃圾钉子螺丝回收机器人,使用 Faster R-CNN 方法实时识别发现散落的钉子和螺丝钉,假设摄像机位于机器人的位置与被测对象的位置深度固定,以此完成回收工作。

钢筋绑扎点精确定位需要对绑扎点进行深度估计。在传统方法中, Feng 等<sup>[6]</sup> 提出了通过提取立体图像的 ORB 特征点计算视差,依照检测结果选取特征点估计交通场景中目标深度; Hu 等<sup>[7]</sup> 提出利用反投影法和模板匹配法获得柑橘深度信息,实现柑橘定位。将深度学习应用于深度估计,速度和精度都有较大提升,从而受到广泛应用<sup>[8]</sup>。以 Godard 等<sup>[9]</sup> 提出的 Monodepth 算法为代表的无监督学习方式避免了使用真实背景深度标签训练,利用未做标记的图片序列作为数据集训练进行深度估计,数据集容易获取,适用场景和泛化能力更强。Wen 等<sup>[10]</sup> 提出将 Monodepth 算法应用于室内自主导航中,对障碍物深度进行预测,提高了预测精度和速度。

因此,为了解决识别准确率和定位精度不理想的问题,本文提出一种基于深度学习的双目视觉钢筋绑扎机器人识别定位方案来识别定位绑扎点空间位置。采用 YOLOv4 算法<sup>[11]</sup> 进行绑扎点目标框识别,在轮廓提取中设计了一种角点选取方法确定绑扎点图像坐标,利用在残差网络中融入卷积注意力模块(convolutional block attention module, CBAM)<sup>[12]</sup> 和解码过程加入 PAN 结构(path aggregation network)的改进 Monodepth 算法进行视差估计,最后通过双目立体视觉得到绑扎点的深度信息,并利用坐标转换,实现了绑扎机器人对绑扎点的识别与定位,通过实验验证了提出算法的有效性。

### 1 钢筋绑扎点视觉识别定位模块

自动钢筋绑扎机器人系统的实物图如图 1 所示,包括视觉识别定位模块、机器人本体模块、控制模块以及绑扎模块。

对绑扎点进行识别定位的流程如图 2 所示,主要由以

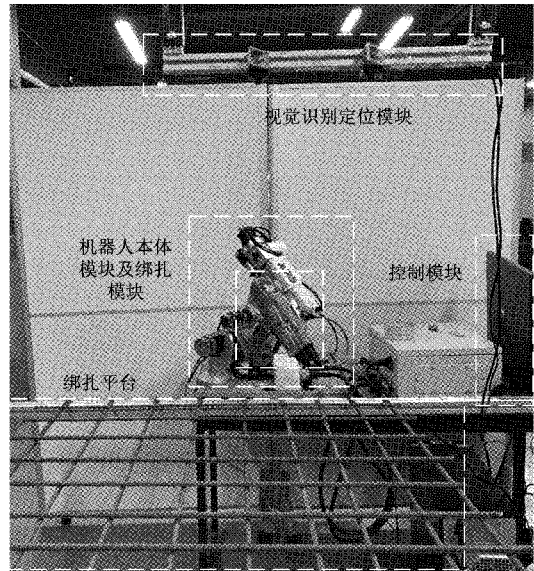


图 1 自动钢筋绑扎机器人系统实物图

下五部分组成:1)图像数据采集和处理,对采集的单双目图像制作数据集;2)检测识别目标,利用单目图像数据集训练网络,双目数据集的左图输入预测,获得绑扎点目标框;3)提取轮廓定位绑扎点在图像中位置,利用十字物体定位方法确定绑扎点目标框中的绑扎点坐标;4)估计视差,利用双目图像数据集训练网络,并根据左图绑扎点坐标估计绑扎点视差;5)空间信息计算,通过视差计算绑扎点深度,并利用坐标转换,以此获得绑扎点的空间信息。

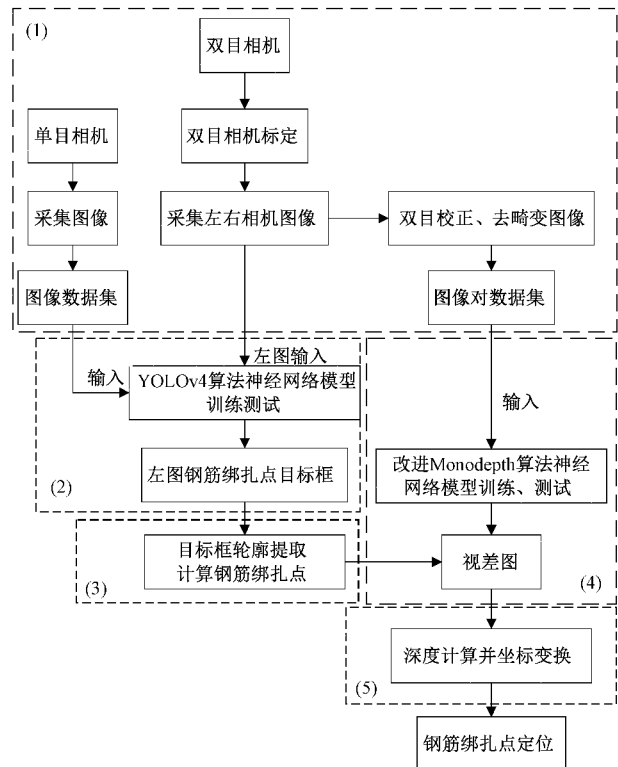


图 2 钢筋绑扎点识别与定位流程

## 2 钢筋绑扎点目标框检测与裁剪

对 YOLOv4 算法训练检测后的目标框进行提取裁剪,目的是为后续图像中钢筋交叉轮廓的提取减少噪声和不必要特征,提高轮廓提取和绑扎点定位精度。如图 3 所示,左相机采集到的区域是绑扎平台的虚线框部分,之后通过 YOLOv4 算法训练后预测绑扎点目标框,将目标框四条边相对于原始图像  $u-v$  坐标轴的数值(左边,上边,右边,下边)作为裁剪边,在原图中以目标框的左上角坐标为原点进行裁剪,最后得到只有目标框区域的图片。

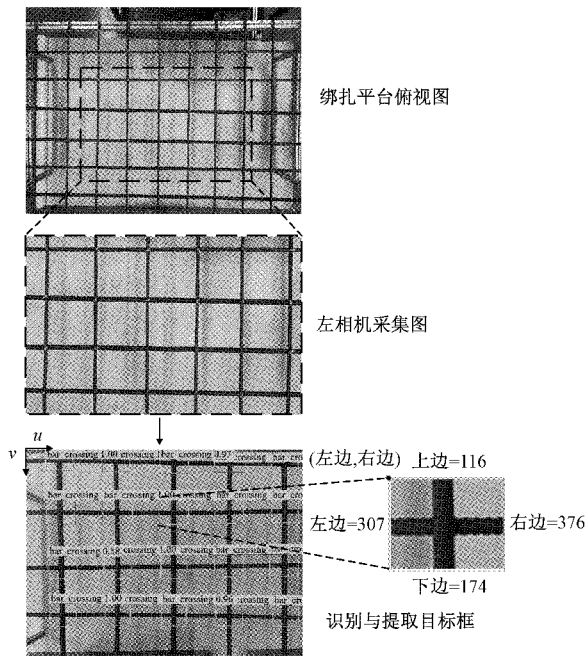


图 3 绑扎点目标框检测与裁剪

## 3 十字物体中心定位方法

本文提取钢筋交叉十字轮廓时,先对图像灰度化、中值滤波、二值化、开运算的预处理,以此来消除图像中的不必要特征和噪声干扰。

通常对于规则四边形物体,可以采用 FindContours 轮廓提取方法<sup>[13]</sup>,以矩形的四个角点坐标  $(x, y)$  来表示矩形轮廓。而钢筋是多边形,且有螺纹影响,针对这种边缘不规则十字物体提取轮廓,本文提出一种十字物体中心定位方法。首先采用 FindContours 方法提取钢筋十字交叉与图像四条边组成的四个边缘不规则的四边形轮廓,每个轮廓由四个角点周围的离散轮廓点组成。然后选取每个轮廓中最靠近十字物体中心位置的角周围的离散轮廓角点,通过式(1)求取各个角周围离散轮廓角点的均值点(如图中点 A、点 B、点 C、点 D),再由均值点代入式(1)求取十字物体中心点  $(x_{center}, y_{center})$ ,如图 4 所示。

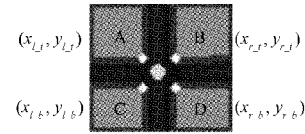


图 4 钢筋十字交叉轮廓、四角、绑扎点位置

均值公式如下:

$$\begin{cases} x = \frac{1}{N} \sum_{i=1}^N x_i \\ y = \frac{1}{N} \sum_{i=1}^N y_i \end{cases} \quad (1)$$

其中,  $(x_i, y_i)$  为选取后的各轮廓中靠近十字物体中心位置的角周围的离散轮廓角点在图像中坐标,  $N$  为符合  $x$  和  $y$  阈值区域内的轮廓角点个数。

按照如下公式选取出每个四边形轮廓角点中靠近十字物体中心位置的角周围符合阈值范围内的轮廓角点。

1) 左上角轮廓:

$$\begin{cases} x_{\max} - \delta < x_i < x_{\max} \\ y_{\max} - \delta < y_i < y_{\max} \end{cases} \quad (2)$$

2) 右上角轮廓:

$$\begin{cases} x_{\min} < x_i < x_{\min} + \delta \\ y_{\max} - \delta < y_i < y_{\max} \end{cases} \quad (3)$$

3) 左下角轮廓:

$$\begin{cases} x_{\max} - \delta < x_i < x_{\max} \\ y_{\min} < y_i < y_{\min} + \delta \end{cases} \quad (4)$$

4) 右下角轮廓:

$$\begin{cases} x_{\min} < x_i < x_{\min} + \delta \\ y_{\min} < y_i < y_{\min} + \delta \end{cases} \quad (5)$$

其中,  $(x_i, y_i)$  为各个轮廓的轮廓角点,  $\delta$  为区域阈值,  $i, j$  的取值为  $0, 1, 2, \dots$ 。

## 4 改进 Monodepth 算法视差估计

### 4.1 Monodepth

Monodepth 算法是一种无监督的单目深度估计算法,通过在训练期间利用双目相机获得的右图来代替深度图作为网络的真实背景深度标签,解决了在实际训练中大量深度图样本标注困难的问题。整体实现过程如图 5 所示,将采集的左视图作为网络的输入,经过卷积神经网络(CNN)后输出分别对应双目左、右视图为基底的两张视差图  $(d^l, d^r)$ ,然后将右视图与预测得到的左视差图进行处理,得到重建后的左视图  $\tilde{I}^r$ ,右视图同理。将重建后的左右视图与真实的对比,利用损失函数得到损失值,从而通过反向传播训练网络。

### 4.2 改进 Monodepth 算法

在 Monodepth 算法中, ResNet50 网络引入残差结构,但随着网络层数增加,训练时间变长,误差变大,梯度消失现象明显,影响准确性。因此本文提出了一种改进方法,旨在增加较少网络层数的情况下,提高网络特征提取能力,弥补网络浅层信息表达能力不足的问题。首先在 Resnet50



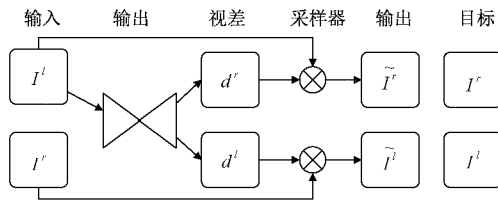


图 5 Monodepth 算法实现过程

网络模型的残差卷积层 resblock 中融入轻量级注意力模块 (CBAM), 解码过程中保留之前视差图上采样结构, 为视差图 1 提取更多特征, 之后加入 PAN 结构, 用 Concatenation 代替 Addition 作为特征融合方法, 将浅层信息和深层信息融合, 为后续视差图的输出提供更多特征, 增强网络对于特征的提取能力和对深浅层特征的融合能力。改进后的 Monodepth 算法的网络框架如图 6 所示。

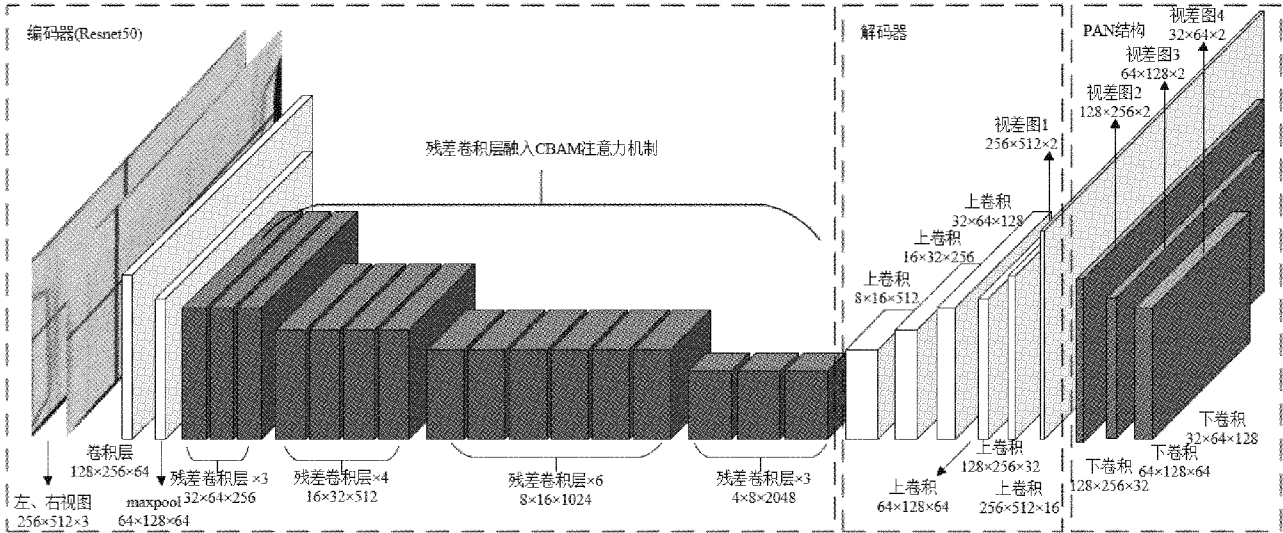


图 6 改进 Monodepth 算法网络框架结构

1) CBAM 注意力机制

CBAM 注意力机制由通道注意力和空间注意力组成, 通道注意力能为辨识信息更多的通道分配更多权重, 空间注意能够在通道注意力的基础上定位输入特征中的显著区域, 以此来提高在网络中的特征中提取关键信息的能力, 防止梯度爆炸或消失的问题。

Resnet50 网络模型的残差卷积层 resblock 由融入 CBAM 注意力机制的 resconv 残差结构组成, 如图 7 所示。残差结构 resconv 将输入特征图分别经过三个卷积层调整通道数和特征图尺寸, 再经过 CBAM 生成新的特征图, 并判断是否需要逆残差结构, 最后经过 BN 层 (Batch Normalization, 批量归一化) 和激活函数, 输出结果。

2) PAN 结构

由于解码过程中采用自顶向下结构虽然增强了语义特征, 但对定位信息传递较少。因此本文将 PAN 结构中的 Bottom-up Path Augmentation 策略与解码过程融合, 添加了自底向上的特征金字塔结构, 如图 8 所示。每次下采样过程中, 首先通过卷积核为  $3 \times 3$  的卷积层对特征图 N2 进行下采样, 之后与横向连接的下一层同级 P3 的上采样信息进行拼接, 再经过 5 次卷积层整合信息, 得到下层特征图 N3, 将浅层的强定位特征传递到深层, 增强多个尺度上的定位能力, 从而提高底层信息的利用率。

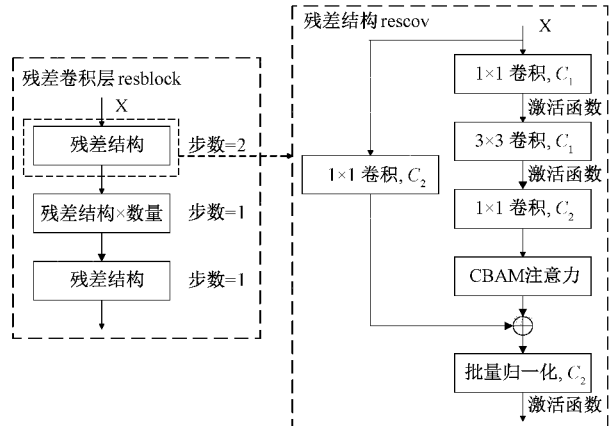


图 7 融入 CBAM 的残差结构 resconv 组成残差卷积层 resblock

4.3 钢筋绑扎点空间定位

1) 双目测距

本文采用双目视觉相机完成定位, 将两相机放置于同一水平面内, 校正后两相机光轴平行且基线平行于成像平面, 双目立体视觉系统中视差与深度的关系如图 9 所示。

$P(X_w, Y_w, Z_w)$  为空间某钢筋绑扎点,  $O_l, O_r$  分别为左右相机光心,  $Z_l, Z_r$  分别为左右相机的光轴, 相机焦距为  $f, f_x, f_y$  分别为在成像平面上  $x, y$  轴的焦距,  $O_l$  和  $O_r$  之间的距离  $B$  为基线距,  $x_l, x_r$  为  $P$  在左右成像平面的横

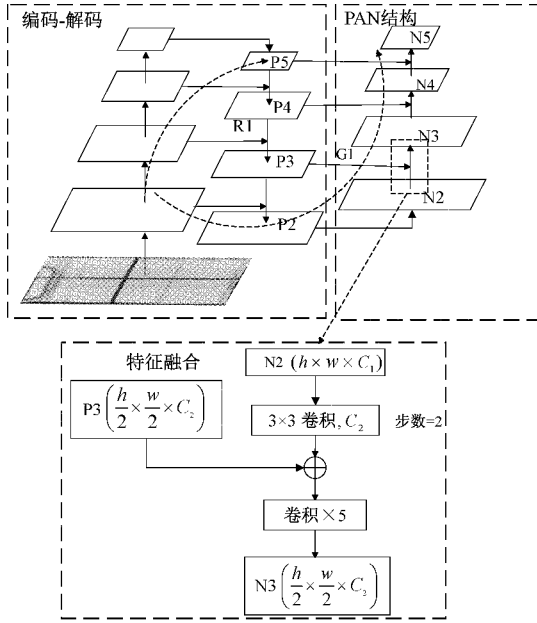


图8 PAN结构

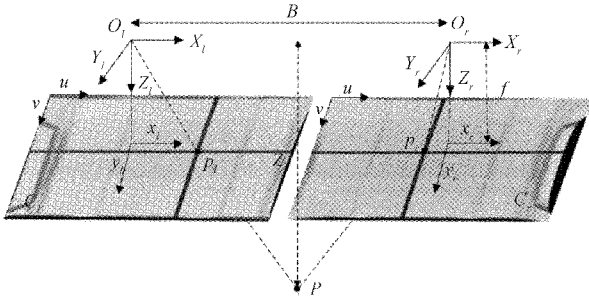


图9 双目立体视觉几何模型

坐标,  $p_l(u_l, v_l), p_r(u_r, v_r)$  为  $P$  点在左右像素平面上的投影点,  $(u_0, v_0)$  为成像平面原点在像素平面上的坐标, 两相机通过畸变矫正和立体校正后, 成为理想双目系统, 左右图像共面, 同一点的投影点在两个成像平面上  $y$  坐标相同。因此可以获得空间点  $P$  在两成像平面上的视差  $d$  为:

$$d = x_l - x_r \quad (6)$$

空间点的深度为:

$$Z = \frac{Bf}{d} \quad (7)$$

空间点的三维坐标信息中  $X, Y$  坐标为:

$$\begin{cases} X = \frac{Z}{f_x}(u_l - u_0) \\ Y = \frac{Z}{f_y}(v_l - v_0) \end{cases} \quad (8)$$

## 2) 坐标转换方法

由于通过双目立体视觉技术获得的是在左相机坐标系中的坐标, 而机械臂是处在三维世界坐标中。因此需要将相机三维坐标系中的绑扎点坐标统一变换为机械臂坐标系中的坐标, 即通过手眼标定实现机械臂坐标系与相机

坐标系之间的变换关系, 其中将相机坐标系中的  $x$  坐标取负, 方便计算。

将绑扎点在机械臂坐标系中的坐标设为  $P(X_w, Y_w, Z_w)$ , 对应相机坐标系中的坐标为  $P(X_l, Y_l, Z_l)$ , 如图10所示。

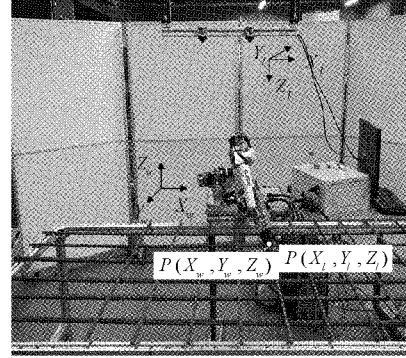


图10 机械臂坐标与相机坐标转换

其中机械臂坐标系原点为基座中心, 相机坐标系原点为左相机光心。

两坐标系转换可以表示为:

$$\begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = \mathbf{R} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \mathbf{T} \quad (9)$$

其中,  $\mathbf{R}(\alpha, \beta, \gamma)$  为  $3 \times 3$  的正交单位旋转矩阵,  $\mathbf{T}$  为  $3 \times 1$  的平移矩阵。

$\mathbf{R}$  为机械臂坐标系转换为相机坐标系的旋转矩阵, 由  $R_\alpha, R_\beta, R_\gamma$  相乘所得,  $\alpha, \beta, \gamma$  分别为机械臂坐标系  $X, Y, Z$  轴的旋转角度,  $R_\alpha, R_\beta, R_\gamma$  分别表示不同坐标轴的旋转分量, 表达式为:

$$R_\alpha = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} \quad (10)$$

$$R_\beta = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \quad (11)$$

$$R_\gamma = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (12)$$

$\mathbf{T}$  为机械臂坐标系转换为相机坐标系时的平移矩阵, 在  $X, Y, Z$  轴的平移的距离分别为  $u, v, p$ ,  $\mathbf{T}$  的表达式为:

$$\mathbf{T} = [u \quad v \quad p] \quad (13)$$

## 5 实验及结果分析

### 5.1 实验平台搭建

实验服务器软件环境: Windows10 1909 操作系统, 安装配置有 PyCharm2019 和 MATLAB2016 软件。卷积神经网络训练采用 PyTorch1.8.1 为深度学习开源框架, 通

用并行计算构架 CUDA10.2, 深度神经网络 GPU 加速库 CUDNN7.6.5, 选用 Python 编程语言; 硬件环境: 计算机内存 64 GB, 搭载 Inter (R) Core(TM) i7-9700K CPU@3.60 GHz x64 处理器, 显卡 NVIDIA GeForce RTX 2080Ti, 显卡内存 11 G, 硬盘 2 T。

实验硬件环境: 选用两个 CMOS 相机, 搭建了一个长 2 m, 宽 1 m, 高 0.8 m 的钢筋绑扎平台, 钢筋规格选用螺纹钢直径 8、10、12、14、16 mm 五种规格。

### 5.2 图像数据集制作

本文在实验室对 5 种不同规格的钢筋, 选用相同直径的钢筋排列组成 1/2/4/6/9 个绑扎点, 分两次采集图像。

利用单目相机, 以相同背景, 不同角度和距离进行拍摄, 共选取 780 张图片, 图像尺寸为 (800×800) 像素, 存储格式为 .jpg。为防止在目标框检测中, 因数据集样本容量过小, 在训练过程中出现过拟合问题, 采用了数据增强算法, 将原有图像数据集扩充至 4 680 张图片, 将其制作成目标检测所需要的图像数据集 Detection\_Datasets, 并用 Labelimg 标注工具对图像中钢筋交叉的位置进行标注(只适用于实验条件下), 如图 11 所示。

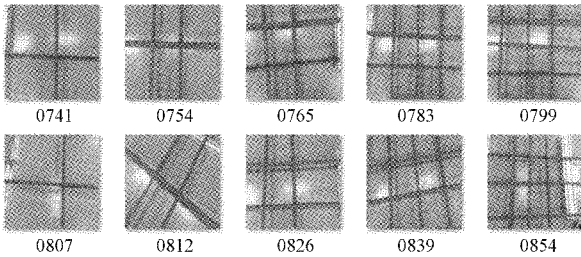
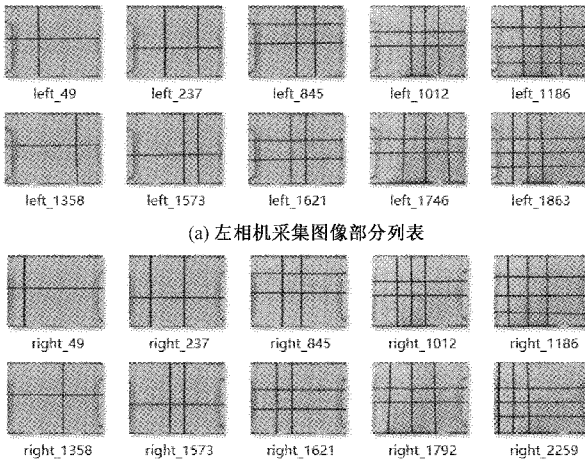


图 11 Detection\_Datasets 图像数据集部分列表

利用双目相机, 以相同背景, 同一平面位置、不同间距进行拍摄, 共选取 2 411 对图像。图像尺寸为 (640×480) 像素, 存储格式为 .jpg, 经双目视觉校正后, 制作成图像立体匹配训练的图像对数据集 Matching\_Datasets, 如图 12 所示。



(a) 左相机采集图像部分列表

(b) 右相机采集图像部分列表

图 12 Matching\_Datasets 图像数据集采集图像

### 5.3 标定实验

#### 1) 双目相机标定

本文的双目相机标定基于 OpenCV 与 MATLAB 结合, 选用张正友标定法对双目相机进行标定<sup>[14]</sup>, 如图 13 所示, 将标定误差控制在 0.17% 以内, 标定后获取双目相机的内外参数、畸变参数。

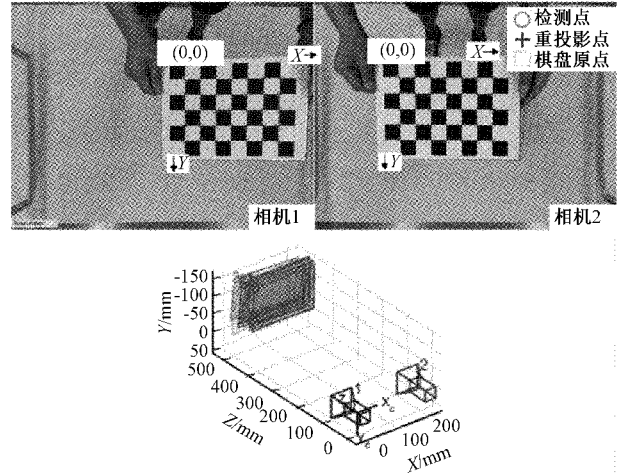


图 13 标定及相机位置

左相机内参矩阵为:

$$lcm = \begin{bmatrix} 486.8036 & 0.3308 & 332.8545 \\ 0.0 & 488.5720 & 263.4699 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$$

左相机畸变参数为:

$$ld = [0.0742 \quad 0.1261 \quad -0.0002 \quad -0.0063 \quad -0.5192]$$

右相机内参矩阵为:

$$rcm = \begin{bmatrix} 488.0967 & 0.3440 & 351.1817 \\ 0.0 & 493.7891 & 270.8836 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$$

右相机畸变参数为:

$$rd = [-0.0089 \quad 0.5542 \quad 0.0037 \quad -0.0071 \quad -1.1842]$$

旋转矩阵为:

$$R = \begin{bmatrix} 1.0000 & -0.0078 & 0.0059 \\ 0.0079 & 0.9999 & -0.0136 \\ -0.0058 & 0.0137 & 0.9999 \end{bmatrix}$$

平移矩阵为:

$$T = [-204.2231 \quad -2.5972 \quad 1.1684]$$

在实际操作中, 两个相机很难精确平行对准, 为消除左右平面畸变和保障其准确共面且行对准, 还需进行立体校正。通过 OpenCV 中的立体校正函数和计算更正 map, 完成本系统的双目校正和去畸变。

#### 2) 手眼标定

利用文献[15]的方法, 先通过在机械臂坐标系中设置 6 个已知绑扎点坐标, 并获得相机坐标系下绑扎点实际坐标, 将 6 组对应坐标代入式(9), 联立方程可求解出坐标转



换的旋转矩阵  $R$  和平移矩阵  $T$ 。

旋转矩阵为:

$$R = \begin{bmatrix} 0.9335 & -0.0358 & 0 \\ -0.0358 & -0.9336 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

平移矩阵为:

$$T = [-97.5559 \quad 577.2939 \quad 670.0491]$$

#### 5.4 识别钢筋绑扎点目标框结果

##### 1) 目标检测算法性能分析

本文选用 YOLOv4 作为目标检测算法,对绑扎点目标框进行识别,并与其他 4 种不同的目标检测算法对比,分别对数据增强后的图像数据集 Detection\_Datasets 采用迁移学习的方式进行训练,前 50 次迭代训练冻结部分预训练模型网络,后 50 次迭代训练解冻模型,共训练迭代 100 次,将数据集按 9:1 的比例划分为训练集和测试集,其中 4 212 张图片为训练集,468 张图片为测试集。各检测算法结果如表 1 所示。

表 1 钢筋绑扎点目标框检测算法实验结果

方法	准确率/%	每秒传输帧数
Faster R-CNN <sup>[16]</sup>	66.91	18.45
SSD <sup>[17]</sup>	99.15	67.79
Efficientdet <sup>[18]</sup>	99.35	14.35
YOLOv3 <sup>[19]</sup>	99.79	40.30
YOLOv4	99.75	54.65

针对于钢筋绑扎点数据集,YOLOv4 算法相比于其他 3 种算法准确率更高,为 99.75%,仅比 YOLOv3 算法的准确率低 0.04%。本实验数据集是基于绑扎点这一单分类物体的目标检测,并且不存在多尺度物体和遮挡问题,同图像中钢筋的尺寸统一,这也是各算法测试准确率较高的原因。从识别速度来看,针对于本文使用的实验设备,SSD 算法的识别速度最快,其次是 YOLOv4 算法,虽然 YOLOv3 算法的识别准确率高,但是在识别速度上比 YOLOv4 算法低 14.35 帧,YOLOv4 算法在识别准确率和运行速度更加平衡,且满足实时识别的要求。

##### 2) 识别结果分析

根据绑扎点在图像中的数量不同,随机选取部分图像进行 YOLOv4 算法识别绑扎点目标框的结果验证,如图 14 所示。YOLOv4 算法通过对数据增强数据集的训练,将其在双目数据集上进行测试,不论是单一还是多个绑扎点,均能以高识别率检测出来,并且置信度都在 0.94 以上,对于背景上钢筋的阴影部分也有很好的区别能力,表明 YOLOv4 具有良好的识别和抗干扰能力。

#### 5.5 钢筋绑扎点定位

本文首先对绑扎点目标框裁剪前后的图像进行轮廓提取的对比,如图 15 所示。对比可知,裁剪前对原图像

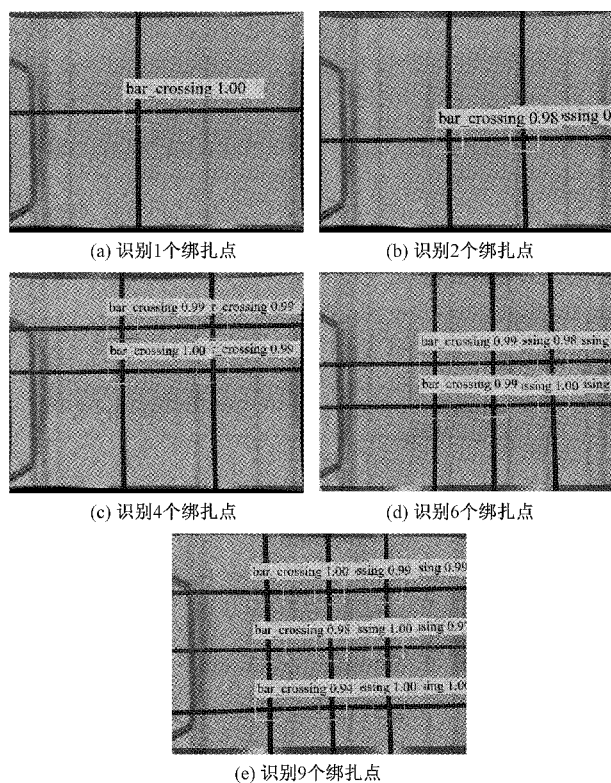


图 14 钢筋绑扎点目标框识别

轮廓提取时,易出现提取不完全或提取其他无用特征的问题,因此通过裁剪目标框,可有效避免其他特征干扰,只保留绑扎点周围的轮廓信息,有利于提高绑扎点定位精度。

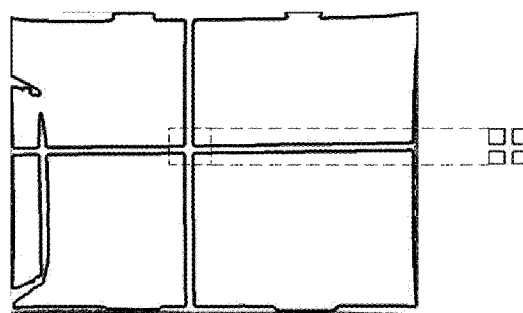


图 15 裁剪前轮廓(左)裁剪后轮廓(右)

分别采用十字物体中心定位方法和霍夫直线变换方法对绑扎点在图像上进行定位,共测试了 200 个绑扎点,如表 2 所示。其中十字物体中心定位方法的定位准确率为 96%,霍夫直线变换方法由于受钢筋边缘波浪状的影响,提取的直线与实际边缘误差较大,定位准确率仅为 63.5%。可以看出,十字物体中心定位方法对于绑扎点的定位精度更高。

本文采用十字物体中心定位方法,定位结果如图 16 所示。实验表明,本文采用的方法可以定位到每个绑扎点在图像中的位置,进一步验证了本方法的有效性。

表 2 钢筋绑扎点定位实验结果

方法	测试点 个数	正确定位 个数	漏检定位 个数	准确率/ %
十字物体中心 定位方法	200	192	8	96
霍夫直线 变换方法	200	127	73	63.5

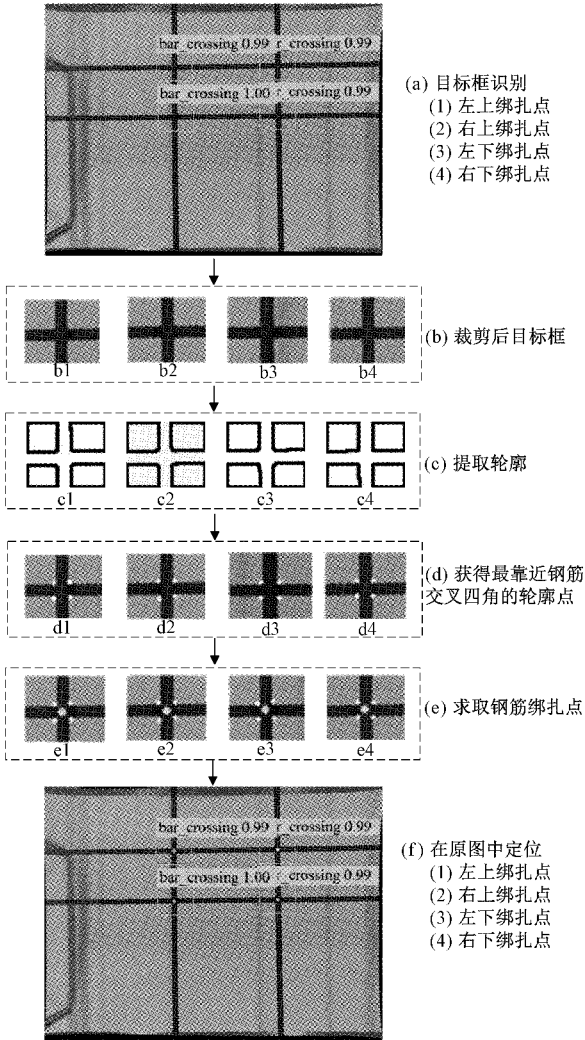


图 16 不同数量钢筋绑扎点的定位

5.6 钢筋绑扎点深度定位实验

分别对 Monodepth 算法采用 ResNet、ResNet + CBAM 和 ResNet + CBAM + PAN 模型,对自采集的 Matching\_Datasets 图像对数据集进行训练,对 3 种模型进行 500 次迭代训练,均已收敛。为了能更直观地分析改进 Monodepth 算法的模型性能,分别绘制了上述 3 种模型在训练过程的损失函数变化曲线,如图 17 所示。

3 种模型在前期迭代训练中损失值的波动均较大,但随着迭代次数的增加,损失值趋于平缓。可以看出,融入

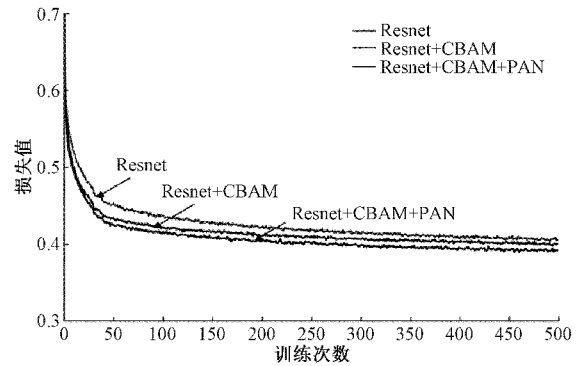


图 17 不同模型的性能曲线

CBAM 注意力机制和 PAN 结构的 ResNet 模型相对于原模型和仅融入 CBAM 注意力机制的 ResNet 模型,损失值有明显下降且损失值最小。

为了验证本文方法对绑扎点在相机坐标系下不同位置的定位效果,在钢筋绑扎台上调整绑扎点位置,以绑扎平台长为  $x$  轴,宽为  $y$  轴,固定  $y$  轴, $x$  轴上每隔 5 cm 做一次绑扎点测试,取 3 次实验结果为一组,共进行五组不同  $y$  坐标的实验,并进行误差估计。

利用 3 种模型分别对绑扎点进行视差估计,通过双目立体视觉方法完成在三维空间中深度的测定,实验结果如表 3 所示。采用十字物体中心定位方法,大部分左右视图定位绑扎点坐标相同,少数误差在 1 像素左右,主要是由于双目校正不能严格对准,存在误差。通过代入左视图定位的绑扎点坐标,采用 3 种模型获得不同视差,并通过双目立体视觉原理计算出深度,与实际距离相比较,分别获得平均误差为 5.2 mm、4.0 mm、3.2 mm,其中 ResNet + CBAM + PAN 模型相较于另外两种模型所估计的视差更小,比原算法平均提高了 2 mm 的精度。由此可知,验证了改进 Monodepth 算法的有效性,为后续钢筋绑扎点在空间中定位提供了有效支持。

5.7 钢筋绑扎点空间定位实验

为了验证本文所提方法的有效性,对已知机械臂坐标系下的钢筋绑扎点  $p(x_{wi}, y_{wi}, z_{wi})$ ,利用本文所提算法对已知绑扎点进行识别定位,求解在相机坐标系下的坐标,并利用坐标转换获得对应机械臂坐标系测量坐标  $p(x_{li}, y_{li}, z_{li})$ ,每个点测量 5 次取平均值,实验结果如表 4 所示。

其中将测量坐标与实际坐标的距离作为定位误差,如式表示为:

$$E_{xyz} = \sqrt{(x_{wi} - x_{li})^2 + (y_{wi} - y_{li})^2 + (z_{wi} - z_{li})^2} \quad (14)$$

利用本文所提算法所获得的空间坐标与实际坐标点的最大定位误差仅为 11.6 mm,具有较好的定位效果,主要是由于在立体匹配过程中通过对特征提取网络添加注意力机制以及在解码部分加入 PAN 结构,提高了定位特征利用率,从而实现了较高的定位精度。



表3 不同模型深度测试与实际距离比较

序号	左视图绑 扎点坐标	右视图绑 扎点坐标	ResNet 模型计算 视差/pixel	测试 距离/ mm	误差/ mm	ResNet+ CBAM 模型 计算视差/pixel	测试 距离/ mm	误差/ mm	ResNet+CBAM+ PAN 模型计算 视差/pixel	测试 距离/ mm	误差/ mm	实际 距离/ mm
1	[279,216]	[146,216]	133.34	751.3	5.7	133.05	753.0	4.0	132.73	754.2	2.8	757.0
2	[310,216]	[179,216]	133.16	752.3	4.7	132.57	755.7	1.3	132.43	755.9	1.1	757.0
3	[341,215]	[209,215]	132.97	753.4	3.6	132.92	753.7	3.3	132.68	754.5	2.5	757.0
4	[380,245]	[248,244]	133.00	753.2	3.8	132.83	754.2	2.8	132.61	754.9	2.1	757.0
5	[410,244]	[278,244]	132.89	753.9	3.1	132.73	754.8	2.2	132.48	755.6	1.4	757.0
6	[442,244]	[310,243]	132.66	755.1	1.9	132.12	758.3	-1.3	132.31	756.6	0.4	757.0
7	[219,310]	[86,309]	133.41	751.0	6.0	133.36	751.2	5.8	133.07	752.3	4.7	757.0
8	[250,309]	[117,309]	133.61	749.8	7.2	133.30	751.6	5.4	132.84	753.6	3.4	757.0
9	[284,309]	[151,308]	133.50	750.4	6.6	133.39	751.1	5.9	133.03	752.5	4.5	757.0
10	[281,338]	[148,338]	133.62	749.8	7.2	133.53	750.3	6.7	133.17	751.7	5.3	757.0
11	[314,337]	[181,337]	133.71	749.3	7.7	133.59	749.9	7.1	133.28	751.1	5.9	757.0
12	[344,337]	[211,337]	133.65	749.6	7.4	133.52	750.3	6.7	133.15	751.8	5.2	757.0
13	[375,372]	[243,372]	133.43	750.8	6.2	133.27	751.7	5.3	133.05	752.4	4.6	757.0
14	[409,371]	[276,371]	133.03	753.1	3.9	132.96	753.5	3.5	132.61	754.9	2.1	757.0
15	[445,371]	[312,372]	132.94	753.6	3.4	132.78	754.5	2.5	132.52	755.4	1.6	757.0

表4 定位测量结果

序号	实际坐标点/mm	ResNet+CBAM+PAN 模型测量坐标/mm	定位误差/mm
1	(275.0,654.0,-87.0)	(271.4,649.1,-83.8)	6.9
2	(125.0,653.0,-87.0)	(119.3,646.5,-85.1)	8.9
3	(-25.0,651.0,-87.0)	(-20.9,657.7,-82.5)	9.1
4	(277.0,500.0,-87.0)	(272.8,493.1,-83.7)	8.7
5	(124.0,502.0,-87.0)	(118.7,493.9,-81.9)	10.9
6	(-26.0,503.0,-87.0)	(-20.2,509.4,-81.1)	10.5
7	(276.0,357.0,-87.0)	(269.6,351.6,-82.4)	9.6
8	(122.0,359.0,-87.0)	(114.3,351.9,-84.5)	10.8
9	(-28.0,362.0,-87.0)	(-21.6,368.7,-80.0)	11.6

## 6 结 论

本文采用双目视觉增强了图像信息,基于深度学习提出了将YOLOv4目标检测算法和改进Monodepth深度估计算法结合的钢筋绑扎机器人识别定位方法。通过YOLOv4算法对绑扎点目标框进行识别和裁剪,减少无用特征干扰,方便轮廓提取;利用十字物体中心定位方法,对提取的多轮廓进行轮廓角点选取,通过角点定位绑扎点图像坐标;通过对Monodepth算法的残差网络融入CBAM,在解码过程中加入PAN结构,提升了网络的特征提取能力,相比于原算法损失值更小,视差更小,深度定位精度更高,利用坐标间转换关系,实现了钢筋绑扎机器人对钢筋绑扎点的精准识别和定位。实验结果表明,对于绑扎点目标框识别准确率为99.75%,绑扎点的图像定位准确率为96%,在空间中最大定位精度误差为11.6mm,为自动钢

筋绑扎机器人目标识别定位提供了新的识别定位方案,并取得较好定位精度效果。

## 参考文献

- [1] SAFA M, CARDENAS J G, LEBLANC D G, et al. Improving construction labor productivity using automatic rebar tying gun[C]. ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, IAARC Publications, 2016, 33:150-155.
- [2] CARDNO C A. Robotic rebar-tying system uses artificial intelligence[J]. Civil Engineering Magazine Archive, 2018, 88(1): 38-39.
- [3] JIN J, ZHANG W, LI F, et al. Robotic binding of rebar based on active perception and planning[J]. Automation in Construction, 2021, 132: 103939.

- [4] WANG Z, LI H, YANG X. Vision-based robotic system for on-site construction and demolition waste sorting and recycling [J]. *Journal of Building Engineering*, 2020, 32, DOI: 10. 1016/j. jobe. 2020. 101769.
- [5] WANG Z, LI H, ZHANG X. Construction waste recycling robot for nails and screws: Computer vision technology and neural network approach [J]. *Automation in Construction*, 2019, 97: 220-228.
- [6] FENG M, LIU Y, JIANG P, et al. Object detection and localization based on binocular vision for autonomous vehicles[C]. *Journal of Physics: Conference Series*. IOP Publishing, 2020, 1544(1): 012134.
- [7] HU X, WU J, YU H, et al. Research on the location of citrus based on RGB-D binocular camera[C]. 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), IEEE, 2021, 5: 2280-2284.
- [8] LAGA H, JOSPIN L V, BOUSSAID F, et al. A survey on deep learning techniques for stereo-based depth estimation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 44 (4): 1738-1764.
- [9] GODARD C, MAC AODHA O, BROSTOW G J. Unsupervised monocular depth estimation with left-right consistency [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 270-279.
- [10] WEN S, LV X, LAM H K, et al. Probability Dueling DQN active visual SLAM for autonomous navigation in indoor environment [J]. *Industrial Robot: the international journal of robotics research and application*, 2021, 48(3): 359-365.
- [11] 解尧婷, 张丕状. 基于改进的 YOLOv4 输电线路小目标检测[J]. *国外电子测量技术*, 2021, 40(2): 47-51.
- [12] 张宇, 张雷. 融入注意力机制的深度学习动作识别[J]. *电讯技术*, 2021, 61(10): 1205-1212.
- [13] 胡佳成, 陆枫, 陈爱军, 等. ABS 齿圈环形表面缺陷检测方法[J]. *电子测量与仪器学报*, 2017, 31(3): 408-414.
- [14] 韦溟, 张丽艳. 机器人自动钻铆系统的现场快速自定位方法[J]. *仪器仪表学报*, 2020, 41(6): 66-75.
- [15] 王芳, 崔丹丹, 李林. 基于深度学习的采摘机器人目标识别定位算法[J]. *电子测量技术*, 2021, 44 (20): 162-167.
- [16] 汤踊, 韩军, 魏文力, 等. 深度学习在输电线路中部件识别与缺陷检测的研究[J]. *电子测量技术*, 2018, 41(6): 60-65.
- [17] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]. *European Conference on Computer Vision Springer, Cham*, 2016: 21-37.
- [18] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 10781-10790.
- [19] 曾维, 尹生阳, 张凤. 基于计算机视觉的垃圾塑料瓶识别与定位算法研究[J]. *电子测量技术*, 2021, 44(23): 12-17.

### 作者简介

董国梁, 硕士研究生, 主要研究方向为工业机器人、计算机视觉等。

E-mail: 17340797057@163.com

张雷, 工学博士, 教授、硕士生导师, 主要研究方向为智能机器人与系统、智慧城市等。

辛山, 工学博士, 讲师, 主要研究方向为智能设备。