

# 基于神经网络的工业机器人视觉抓取系统设计

燕 硕, 李建松, 唐昌松

(徐州工业职业技术学院 机电工程学院, 江苏 徐州 221140)

**摘要:** 针对机器人示教编程方法导致的工件位置固定、抓取效率低下的问题, 研究神经网络在机器人视觉识别与抓取规划中的应用, 建立了视觉引导方案, 通过 YOLOV5 神经网络模型开发视觉识别系统, 识别物体的种类, 同时获取待抓取物体定位点坐标; 提出了机器人六点手眼标定原理并进行标定实验, 提出了针对俯视图为圆形或长方形物体的定位方法; 最后针对 3 种物体进行了 180 次的抓取实验, 实验的综合平均抓取成功率约为 92.8%, 验证了视觉识别和抓取机器人系统具备实际应用的可能性, 有效提高了抓取效率。

**关键词:** 神经网络; 目标定位; 机器人抓取; 机器人标定; 视觉引导

## Design of Industrial Robot Visual Grasping System Based on Neural Network

YAN Shuo, LI Jiansong, TANG Changsong

(School of Mechatronic Engineering, Xuzhou College of Industrial Technology, Xuzhou 221140, China)

**Abstract:** Aiming at the problems of fixed workpiece position and low grasping efficiency caused by robot teaching programming method, this paper studies the application of neural network in robot vision recognition and grasping planning, establishes a vision guidance scheme, develops a vision recognition system through the YOLOV5 neural network model, identifies the types of objects, obtains the coordinates of the positioning points of the objects to be grasped, puts forward the robot six-point hand-eye calibration principle, carries out the calibration experiment, and proposes the positioning method for objects with circular or rectangular top view. Finally, the grabbing experiment is completed 180 times for three kinds of objects, and the overall average success rate of the experiments is about 92.8%, which verifies the possibility of practical application in the vision recognition and grabbing robot system and effectively improves the grabbing efficiency.

**Keywords:** neural network; target location; robot grasping; robot calibration; visual guidance

## 0 引言

在智能化要求越来越高的情况下, 传统的示教编程规划机器人路径的方法已经不能满足多样化的生产需求<sup>[1-2]</sup>。具备视觉检测功能的工业机器人替代传统的示教方法, 一定程度上发挥人眼检测的功能。基于神经网络的视觉识别系统调用摄像头后, 将图像信号传送给训练好的神经网络识别模型, 该模型将图像信息转变成数字信号<sup>[3]</sup>, 再利用特征提取方法提取目标物体的特征<sup>[4]</sup>, 从而实现目标检测和定位, 进而控制现场的机器人实施抓取动作。

## 1 视觉引导方案

引导系统如图 1 所示, 首先由摄像头采集物体的图像信息, 将此图像信息传送至 PC 中的神经网络模型, 神经网络模型开始识别图像信息, 包括种类、建议框位置等, 然后将该图像信息传输至机器人端的位置寄存器 PR [i], 机器人端读取位置寄存器中的位置信息并执行抓取程序, 实施抓取动作。在机器人视觉<sup>[5]</sup>系统引导方案中, 为了能够

实现准确而快速地执行完整流程, 需要对摄像头进行正确的选型并进行相机标定<sup>[6-7]</sup>, 同时为了统一相机的坐标系和机器人的坐标系还需对两者进行手眼标定<sup>[8-10]</sup>, 在下文中将介绍这些内容。

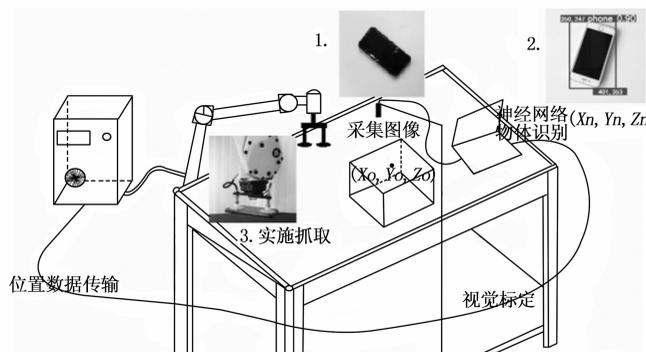


图 1 引导方案示意图

收稿日期: 2023-05-10; 修回日期: 2023-06-14。

基金项目: 江苏高校“青蓝工程”项目(苏教师函(2022)29号)。

作者简介: 燕 硕(1993-), 男, 硕士研究生。

引用格式: 燕 硕, 李建松, 唐昌松. 基于神经网络的工业机器人视觉抓取系统设计[J]. 计算机测量与控制, 2024, 32(4): 201-209.

## 2 特征提取方法与目标检测模型

### 2.1 特征提取方法

#### 2.1.1 卷积和池化

卷积 (Conv) 和池化 (Pool)<sup>[11-12]</sup>是卷积神经网络中特征提取的核心操作, 一个特征图包含若干个特征平面 (Feature Map), 如输入彩色

图像有  $R$ 、 $G$ 、 $B$  三个通道, 即 3 个特征平面, 特征平面的数量由卷积核卷积操作后得到, 每个特征单元呈平面的形状。在每个特征平面内, 卷积时的权值是共享的, 这种共享也称之为卷积核, 卷积核实际上为一个方形矩阵,

尺寸不固定, 如图 2 所示,  $3 \times 3$  的卷积核  $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$  在图

片的像素上以步长 ( $s$ ) 为 1 依次往右、往下滑动 (称为滑动窗口), 特征提取就是加权的過程, 在  $5 \times 5$  的像素上做卷积可得到  $3 \times 3$  的特征平面, 特征平面尺寸是按照式 (1) 计算得到 (假设输入  $n \times n$  图像):

$$\frac{n + 2p - f}{s} + 1 \quad (1)$$

其中:  $P$  为填充 (在特征平面边缘部分增加空白填充以适应卷积核的大小),  $s$  为步长,  $f$  为卷积核尺寸。神经网络学习时, 会更新卷积核内的参数。

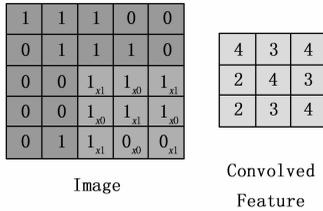


图 2 卷积过程

卷积提取特征时, 卷积窗口从左至右、从上至下滑动, 物体特征由具体到抽象、由浅层到深层依次提取, 如图 3 所示, 图像经过卷积后, 物体的特征逐渐抽象。卷积神经网络的另一个优点是可以降低参数数量<sup>[13]</sup>, 如图 4 所示, 假设图像有  $1\,000 \times 1\,000$  像素, 网络有 100 万个隐含层的神经元, 若将它们全连接会产生  $1\,000 \times 1\,000 \times 1\,000\,000$  共计  $10^{12}$  的权值参数。由于该卷积核进行卷积时的参数相同, 所以只提取了一种特征, 此时可以增加不同卷积核的数量来提取多种特征, 如颜色、边缘形状等, 若有 100 个不同的卷积核, 进行卷积操作后就会得到 100 个特征平面 (Feature Map), 此时的参数数量只有  $100 \times 100 = 10\,000$  个, 相比全连接方式有了较大程度的下降。

池化作为一种降维的方法, 一般跟在卷积层之后用来减小特征平面 (Feature Map) 的尺寸, 从而减少计算量, 和卷积一样通过感受野在特征平面上滑动逐步扫描全局。如图 5 所示, 最大池化是选取感受野中的最大值, 可获取局部信息。池化后的特征平面尺寸依据式 (2) 得到:

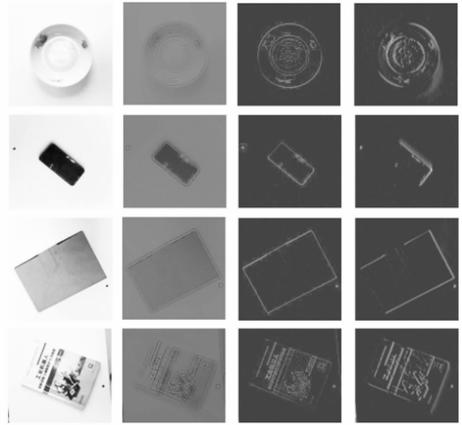


图 3 卷积特征提取过程

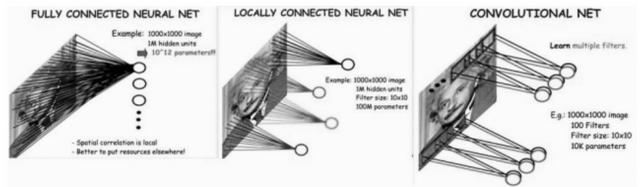


图 4 不同连接方式的参数数量

$$\frac{n - f}{s} + 1 \quad (2)$$

其中:  $n$  为池化前 (输入到池化层中的特征平面) 的特征平面的尺寸,  $f$  为池化窗口的尺寸,  $s$  为池化操作的步长 (每次池化窗口移动的距离)。经最大池化后图像尺寸减小, 但保留了主要的特征信息, 实际效果如图 6 所示。

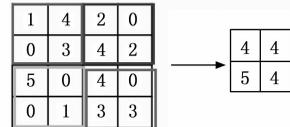


图 5 最大池化

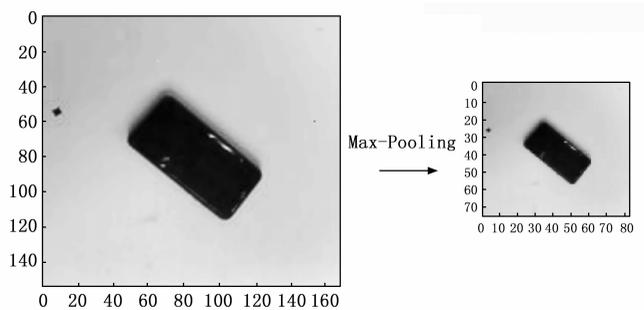


图 6 池化的实际效果

#### 2.1.2 特征融合方法

YOLOV5 中采用一种基于切片<sup>[14]</sup>操作 (Slice) 用来降低特征平面 (Feature Map) 尺寸的下采样方法, 切片可将特征平面切割成较小的尺寸, 把高分辨率的特征平面隔列切割成低分辨率的特征平面, 再进行拼接, 如图 7 所示,

不同于池化操作单纯缩小尺寸, 而是将原始图像按照纵横坐标方向各做二倍间隔采样得到的特征平面, 再经过路由 (Concat) 连接形成新的特征图。此操作能将二维平面特征向通道维度转换, 可减少下采样 (如池化操作) 带来的信息损失。

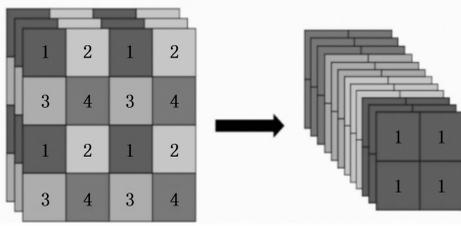


图 7 切片操作

### 2.1.3 激活函数

卷积的操作是通过参数的矩阵乘法来实现的, 所以纯卷积运算的本质是线性操作, 不具备非线性分类能力。激活函数的输入是神经网络的输入与权值的加权, 激活函数的作用是给神经网络带来非线性的特征, 特别是深层神经网络, 如图 8 所示, 激活函数可以实现复杂的非线性分类任务。

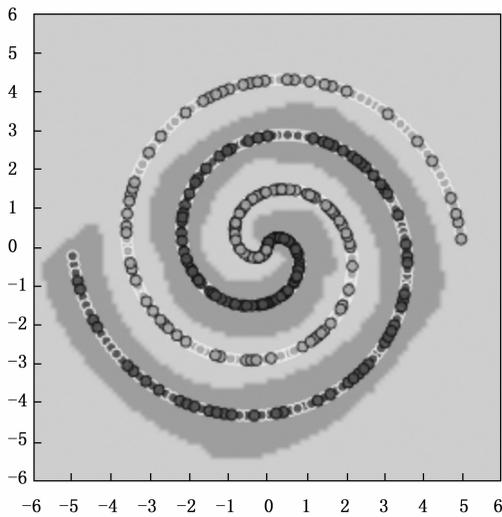


图 8 激活函数提供非线性分类能力

### 2.1.4 损失函数

损失函数用来描述真实值和预测值之间的差异, 香农 (Shannon) 认为信息量<sup>[15]</sup>与事件发生的概率成反比, 信息量越小则发生的概率越大, 反之亦然。假设某一事件发生的概率为  $p(x)$ , 则信息量表示为:

$$I(x) = -\log(P(x)) \quad (3)$$

信息熵又可称之为熵, 用所有信息量的期望来表示, 设随机变量  $X$ , 则信息熵可以表示为:

$$H(x) = -\sum_{i=1}^n P(x_i) \log(P(x_i)) \quad (4)$$

若  $X$  有两个独立的概率分布  $P(x)$ 、 $Q(x)$ , 那么可

使用相对熵 (KL 散度) 来描述  $P(x)$ 、 $Q(x)$  之间的差异, 相对熵 (KL 散度) 可表示为:

$$D_{KL}(p || q) = \sum_{i=1}^n p(x_i) \log \left[ \frac{p(x_i)}{q(x_i)} \right] \quad (5)$$

实际上, 若相对熵 (KL 散度) 越小, 则  $P(x)$ 、 $Q(x)$  的分布便更接近, 因此可以通过训练  $Q(x)$  的分布使其逼近  $P(x)$  的分布。将相对熵的表达式可拆解为:

$$D_{KL}(p || q) = -H[p(x)] + \left\{ -\sum_{i=1}^n p(x_i) \log[q(x_i)] \right\} \quad (6)$$

其中:  $H[p(x)]$  表示信息熵, 后者为交叉熵, 交叉熵相比相对熵更加容易计算, 所以常使用交叉熵计算损失函数。

## 2.2 目标检测模型

### 2.2.1 YOLOV5 检测模型

YOLOV5<sup>[16]</sup>是集目标分类与目标追踪功能于一体的卷积神经网络模型<sup>[17]</sup>, 主要以卷积<sup>[18-19]</sup>和池化<sup>[20]</sup>提取物体的特征并进行特征融合。在目标追踪功能上, 以矩形框将目标框出, 如图 9 所示, 有两个矩形框  $B_1$ 、 $B_2$ , 根据两坐标点  $(x_1, y_1)$ 、 $(x_2, y_2)$ ,  $(x'_1, y'_1)$ 、 $(x'_2, y'_2)$  分别绘制得到。其中  $B_1$  为真实框 (Ground Truth Box),  $(x_1, y_1)$ 、 $(x_2, y_2)$  是人工标注的位置, 存放在标注文件中。 $B_2$  为预测框 (Prediction Box),  $(x'_1, y'_1)$ 、 $(x'_2, y'_2)$  是由目标检测模型计算输出的坐标, 为了拟合逼近两框, 定义  $IOU$ <sup>[21]</sup>表示为:

$$IOU = \frac{B_1 \cap B_2}{B_1 \cup B_2} \quad (7)$$

$IOU$  的值越大, 说明  $B_1$ 、 $B_2$  两框越接近, 所以定义损失函数  $IOU\_Loss$  为<sup>[22]</sup>:

$$IOU\_Loss = 1 - IOU = 1 - \frac{B_1 \cap B_2}{B_1 \cup B_2} \quad (8)$$

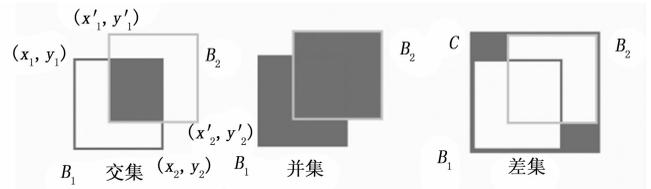


图 9  $B_1$  和  $B_2$  的交集、并集与差集

进一步使用  $CIOU\_Loss$ <sup>[23]</sup> 作为预测框的损失函数<sup>[24]</sup>,  $C$  代表矩形  $B_1$ 、 $B_2$  的最小外接矩形, 那么差集为并集减去交集, 则  $CIOU\_Loss$  可表示为:

$$CIOU\_Loss = 1 - IOU + \frac{\rho^2(b, b^w)}{c^2} + \alpha v \quad (9)$$

其中:

$$\alpha = \frac{v}{1 - IOU + v}, v = \frac{4}{\pi^2} \left( \arctan \frac{w^w}{h^w} - \arctan \frac{w}{h} \right)^2$$

$\rho^2(b, b^w)$  为预测框和真实框的中心点的欧式距离;  $c$  为外接矩形  $C$  的对角线距离;  $w$  为外接矩形框的宽度;  $h$  为外接矩形框的高度。

目标检测的时候，由于预测框采用滑动窗口的方式在原始图像上做检测，所以一个被检测物体上会出现多个预测框重复的情况。如图 10 所示，为了得到被检测目标上正确的预测框，使用非极大值抑制（NMS, non maximum suppression)<sup>[25]</sup>的方法抑制冗余的预测框，首先将所有的预测框根据得分（得分代表该框正确框出目标的概率）排序，将最高分对应的框保留，并与其他所有框依次计算 *IOU*，若 *IOU* 大于设定的阈值，则删除此框，将剩余的框（不含最高分对应的框）继续重复上述步骤即可得到被检测目标正确的预测框。

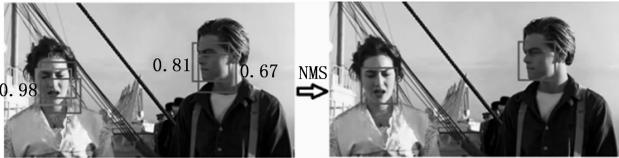


图 10 非极大值抑制

### 2.2.2 模型训练及评价指标

为实现对 4 种物体（书、盘子、手机、盒子）分类并检测，首先获取 4 种物体的相应图片共 3 200 张，将获取到的图片命名为：类别+编号的形式存放在数据集 images 中，采用精灵标注对图片标注，此步骤主要获取物体的目标框在图片中的位置坐标，精灵标注后将获得目标框的左上、右下顶点的像素坐标、类别、图像的尺寸、图片通道深度等，把所有的图像标注好后生成标签文件，与原图名称对应。将所有图像按照 8 : 1 : 1 的比例划分为训练集、验证集与测试集。本次模型训练在 Python3.7, Pytorch1.6.0, cuda10.0 环境中进行。设置训练的参数中，‘--weights’: 加载预训练的模型权重文件 yolov5m.pt; ‘--cfg’: 选择 yolov5m.yaml 模型配置; ‘--data’: 选择制作好的数据集; ‘--epochs’: 训练总轮次，本次训练 300 轮。‘--batch-size’: 一次训练选取的图片数，这里设置 16。

本次训练过程中，在训练至 0~50 轮之间准确率（Precision）快速上升，50~100 轮之间准确率（Precision）上升缓慢，100 轮后准确率达到 0.97 后开始小幅震荡。召回率在训练的 0~50 轮之间快速上升，达到 0.98 保持不变。目标框损失下降过程较缓慢，在训练 250 轮之后的损失为 0.015 左右，分类损失下降较快最终接近 0，可见分类比目标框损失更易收敛，验证集上也保持了相同的趋势。

### 2.2.3 目标检测实验

分别在亮光、暗光条件下拍摄手机、书、盘子、盒子 4 种物体的照片送入检测网络，网络调用训练好的权重文件进行检测，其检测结果如图 11 所示：对 4 种物体在亮光和暗光条件下各进行 50 次识别实验，结果如表 1 所示，在亮光条件下，YOLOV5 网络对 4 种物体的识别成功率高于暗光，对手机的识别成功率最高。（由于排版的原因图 11 的照片经过裁剪缩放等操作，故显示像素坐标与检测框大小不成比例）。

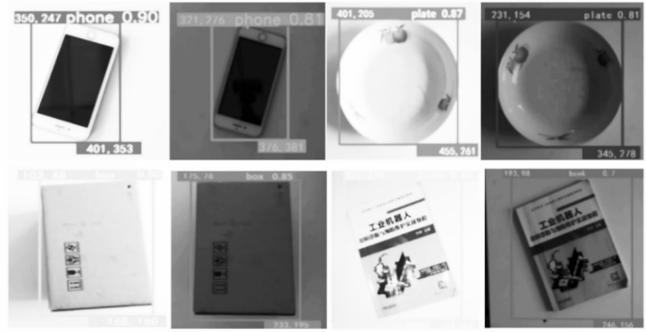


图 11 视觉检测结果

表 1 视觉识别成功率统计结果

光线	名称	次数	框中次数	成功率/%
亮光	盘子	50	49	94
	手机	50	50	98
	纸盒	50	47	92
	书	50	47	92
暗光	盘子	50	45	90
	手机	50	47	94
	纸盒	50	42	90
	书	50	43	90

## 3 机器人手眼标定

### 3.1 机器人手眼标定原理

为使机器人能顺利抓取物体，需将像素坐标转换成机器人坐标，这个过程为机器人手眼标定。工业机器人包括用户、工具、关节等在内的多个坐标系，如图 12 所示，不同类型的机器人坐标系有着不同的适用场景，在自动识别抓取机器人系统中选用用户坐标系建立实验平台，且摄像头的视野可覆盖整个坐标系。

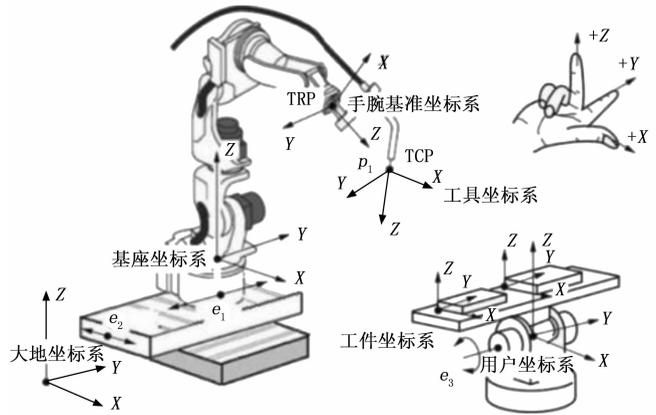


图 12 机器人坐标系

“手眼标定”中，“手”代表机械手，“眼”代表摄像头，根据相机的不同搭载方式可分为“eye in hand”、“eye to hand”两种，如图 13 所示，“eye in hand”方式中，摄像头跟随机器人的移动而移动，“eye to hand”方式中，摄像

头固定不动, 在本文自动识别抓取机器人的实验中, 将采用“eye to hand”方式。

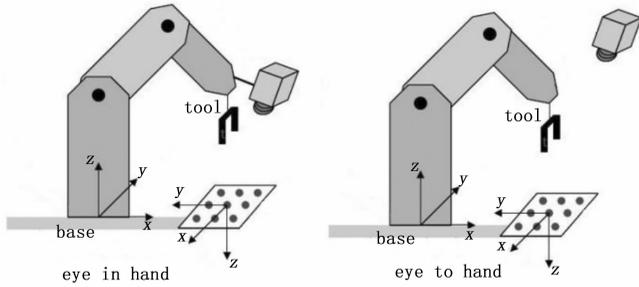


图 13 eye in hand 和 eye to hand

手眼标定有几种不同的方法, 本文抓取实验中由于被抓取物体的高度是一定的, 且采取同一种自上而下的抓取姿势, 这里提出简单的六点标定法。其思想为假设有一个变换矩阵  $T$ , 能将像素坐标系中的点  $X = [x, y, 1]^T$  转换为机器人用户坐标系中的点  $Y = [x', y', 1]^T$ , 即  $TX = Y$ , 可以表示为:

$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \Rightarrow \begin{cases} ax + by + c = x' \\ dx + ey + f = y' \end{cases} \quad (10)$$

### 3.2 机器人手眼标定实验

固定好相机确保视野覆盖工作台, 将标定针固定在机器人六轴法兰盘上, 并保持标定针的末端与被抓取物体的高度持平; 将标定板置于相机下方, 调整相机焦距确保拍摄清晰, 记录标定点在像素坐标系中的坐标; 将标定针移动到标定点上, 记录在机器人坐标系中的坐标。在本次实验中, 分别记录了相机像素坐标系中的 6 个随机点和其对应的机器人坐标系中的点, 本次实验记录的坐标如表 2 所示。得到转换矩阵后, 拍摄一组照片, 经转换矩阵转换后与机器人坐标系中的坐标对比, 结果如表 3 所示。

表 2 六点采集坐标

像素坐标/mm	机器人坐标/mm
(152.2, 131.4)	(312.5, 72.4)
(314.6, 131.2)	(382.4, 72.1)
(314.5, 276.8)	(382.1, 142.6)
(476.5, 276.4)	(452.8, 142.7)
(152.6, 421.5)	(312.3, 212.4)
(476.5, 421.4)	(452.2, 212.7)

计算可得转换矩阵:

$$T = \begin{bmatrix} -0.43 & 0 & 246.30 \\ 0 & 0.48 & 8.76 \end{bmatrix} \quad (11)$$

将转换后的横、纵坐标与机器人实际测量的横、纵坐标做均方误差, 可以得到横坐标的均方误差约为  $MSE(x) = 0.18 \text{ mm}$ , 纵坐标的均方误差约为  $MSE(y) = 0.60 \text{ mm}$ , 标定精度能够满足要求。

表 3 手眼标定结果

像素坐标/mm	理论坐标/mm	实际坐标/mm
(124.2, 89.4)	(299.1, 51.7)	(289.9, 51.6)
(158.4, 245.8)	(314.4, 126.7)	(314.5, 126.4)
(164.5, 179.2)	(317.0, 84.8)	(316.7, 84.6)
(315.1, 261.3)	(381.8, 134.2)	(382.0, 134.5)
(204.2, 368.7)	(334.1, 185.7)	(334.2, 185.3)
(289.4, 246.6)	(370.7, 127.1)	(370.6, 127.3)

## 4 目标定位方法研究

### 4.1 目标定位方法

对于工业机器人来说若想准确抓取物体, 则需要被测物体上的一个目标点与机器人的 TCP 点对应, 这里提出一种简便的目标定位方法, 可使神经网络识别坐标 (像素坐标) 转换为机器人坐标系中的定位坐标。3.2 节中将图像的像素坐标系和机器人的用户坐标系统一起来, 本节探索目标定位的方法。

在 2.2 节中, 物体经 YOLOV5 算法识别后会得到其在像素坐标系中的位置坐标信息, 即物体的最大外接矩形左上、右下顶点坐标, 图 14 为像素坐标系  $u-v$  中的两个像素坐标点 (401, 205)、(455, 261), 在一般情况下将其表示为  $(x_1, y_1)$ 、 $(x_2, y_2)$ , 将手眼标定得到的变换矩阵与该坐标相乘, 可把两坐标转换至机器人的用户坐标系中, 表示为:

$$\begin{bmatrix} x'_1 \\ y'_1 \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \quad \begin{bmatrix} x'_2 \\ y'_2 \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \quad (12)$$

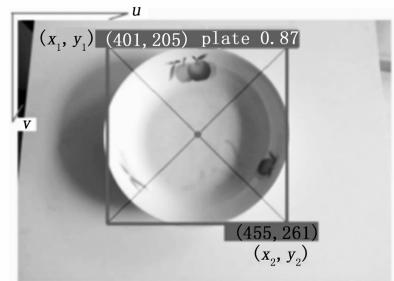


图 14 抓取点

在式 (12) 中,  $(x'_1, y'_1)$ 、 $(x'_2, y'_2)$  为转换后的机器人用户坐标系中的坐标, 机器人用户坐标系  $user\_xy$  中物体检测框如图 15 所示, 本文中工业机器人抓取实验对象为盘子, 由于盘子形状基本上为圆形, 所以抓取点可以定位为盘子中心点, 将该点表示为  $[(x'_1 + x'_2) / 2, (y'_1 + y'_2)]$ , 该点即为定位抓取点。FANUC 工业机器人的位置、位姿表示有 6 个参数, 分别为  $(x, y, z, \omega, p, r)$  其中  $x, y, z$  为位置信息, 用来描述物体在机器人坐标系中的位置;  $\omega, p, r$  为位姿信息, 用来描述机器人抓取时末端姿态, 包括旋转、倾斜角度等。前文说到被抓取物体的高度

是一定的,且采取同一种自上而下的抓取姿势,所以将  $z$ ,  $\omega$ ,  $p$ ,  $r$  设置为固定值。由于定位坐标由转换后的机器人坐标计算而来,因此定位点的实际偏差可参考表 3,其误差较低,能够获得较高的定位精度。

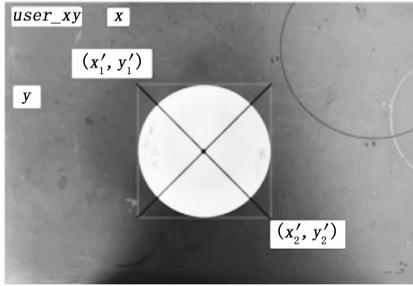


图 15 抓取点表示

### 4.2 物体坐标数据的传输研究

如图 16 所示,获取到定位点坐标后,该坐标经手眼标定矩阵计算后得到机器人坐标系坐标,并将该坐标实时保存至指定路径,上位机从该指定的路径中采集坐标,将其传输至机器人的 PR [i] 位置寄存器。为了能够顺利实现坐标的采集与发送,设计了数据传输模块上位机 FanucTransTool。

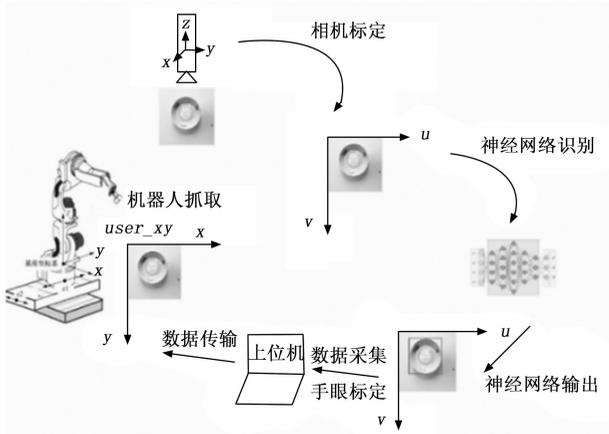


图 16 数据传输过程

#### 4.2.1 工作方案

数据传输模块是将 PC 端上位机视觉定位数据传输至机器人控制柜,并通过机器人编程执行相关动作。

基于 FANUC Robot Interface 接口通信可以直接调用 API 接口将数据写入机器人 DO 信号、PR 寄存器、R 寄存器等。根据实际系统设计的需要,数据传输模块分为几部分,分别是:机器人通信设置模块、实时数据显示模块、日志模块、读取文件数据模块以及实时传输模块。

#### 4.2.2 读写控制流程

基于 FANUC Robot Interface 接口读取 FANUC 机器人 IO 数据流程如图 17 所示,PC 端和 FANUC 机器人进行通信前,需创建 FRRJIF.Core object 对象,并设置读写 IO 的索引范围,连接通信后才可读写 IO 数据,程序执行结束后需断开连接并释放创建的对象。

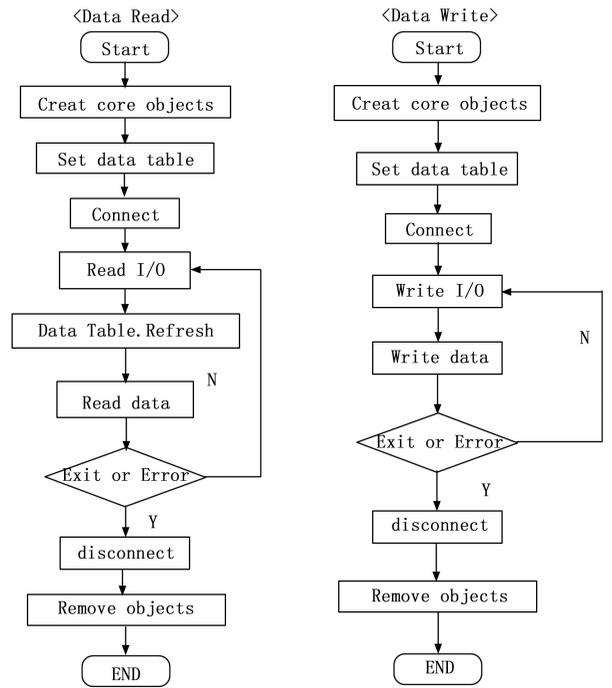


图 17 读写控制流程

#### 4.2.3 开发与运行环境

数据传输模块的开发与运行环境如表 4 所示, Visual Studio 2015 作为开发环境, C++ 编写程序代码。数据传输模块上位机 FanucTransTool 如图 18 所示。

表 4 开发与运行环境

操作系统	Windows7 及以上版本
硬件环境	CPU Intel Core i3 以上,内存 4 G 以上
开发环境	Visual Studio 2015
程序语言	C++
操作系统	Microsoft Windows98 以上版本



图 18 数据传输模块界面

## 5 机器人视觉抓取系统实验

### 5.1 机器人抓取实验平台搭建

本文的机器人视觉抓取系统如图 19 所示,选用深圳祺云公司的 FA0420C 型号 4 mm 定焦镜头,成像尺寸 7.2 mm

×5.3 mm, 靶面尺寸 1/1.8", 相机选择显微精工 XW500 型工业相机, 像元尺寸 2.2 mm×2.2 μm, 工业机器人选择 FANUC-M10IA-12 六轴工业机器人, 在机器人末端固定两个吸盘, 工业相机固定在工作台上, 视野可以覆盖整个工作台面, 相机通过 USB 接口和 PC 连接。机器人连同底座固定在地面上, 机器人的用户坐标系建立在工作台面上, 其  $x$  与相机像素坐标系的  $u$  方向相同,  $y$  与相机像素坐标系的  $v$  方向相同, 图中的工业机器人和工业相机构成了一个典型的 eye to hand 结构。除了上述设备外还有 PC 机、气泵、真空发生器、相机支架、工作台以及其他辅助设备。

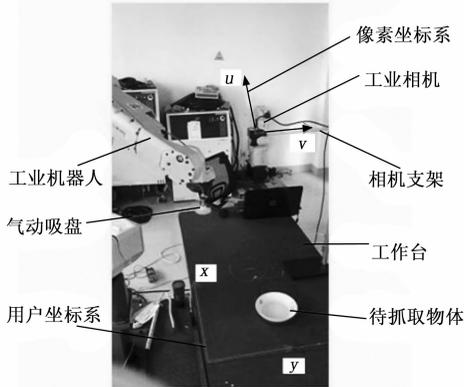


图 19 机器人视觉抓取系统

### 5.2 软件控制流程

软件逻辑流程如图 20 所示, 数据传输模块界面如图 21 所示, 通信建立之后, 数据传输模块会将 FANUC 机器人上线信号置 1, 会定时读取位姿数据, 若位姿数据变化, 则将数据发送至机器人, 完成后将机器人坐标完成信号置 1, 机器人执行后将坐标完成信号 DO 置 0, 完成整个数据传输流程。

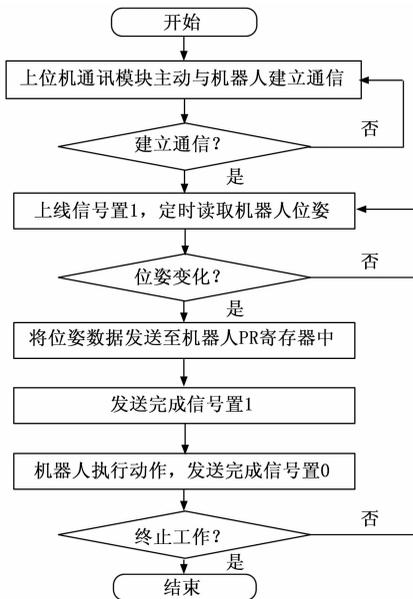


图 20 软件逻辑流程图

### 5.3 实验控制步骤

机器人抓取控制步骤如图 21 所示, 分为单目标抓取和多目标抓取。当工作台上只有一个物体时为单目标抓取, 此时工业相机实时获取图像信息, 将其送入识别算法, 视觉识别算法会实时刷新并输出定位坐标 (刷新周期可设置, 这里设置为 0.1 s), 输出坐标若与上一输出坐标相同, 或在给定的误差范围内 (误差范围可根据抓取精度要求而设定), 则判断为重复识别, 本次不抓取, 当输出坐标与上一输出坐标不同时则判定为新物体, 此时 FanucTransTool 会输出信号至工业机器人, 执行抓取指令, 当机器人执行抓取动作时, 视觉识别算法停止识别, 直至机器人完成当前物体的抓取并回归原点等待。当工作台上一次出现多个物体时, 为多目标抓取, 此时与获取单目标物体的定位坐标原理相同, 获取多个被抓取物体的定位坐标, 将这些坐标按照机器人用户坐标系中  $x$  的值从大到小将若干坐标点排序, 将最大坐标点传输至机器人端并抓取该物体, 抓取完成后重复上一步骤, 在抓取执行过程中, 视觉识别算法将不再重复识别, 直到当前抓取动作完成后, 再进行下一轮的识别。

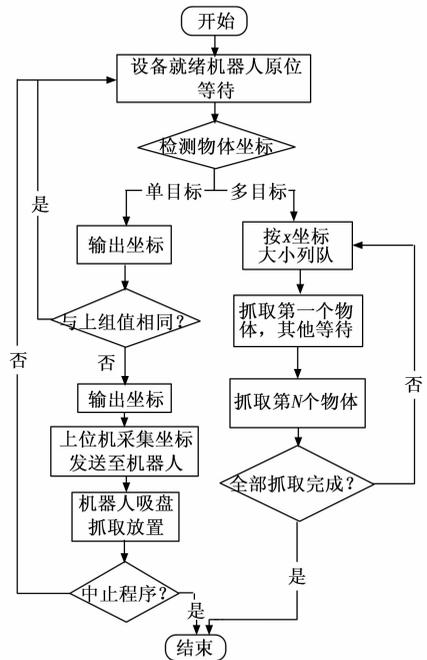


图 21 抓取控制流程

### 5.4 数据通讯实验

编写好数据传输界面后进行数据的收发实验, 在数据传输界面中输入机器人端的 IP 地址, 设置定时间隔 (数据传输间隔时间), 并指定机器人端的输出信号 DO [i] 为连接成功及数据传输完成的标志信号, 如图 22 所示, 指定一组坐标 [84, 85, 86, 87, 88, 89], 数据传输界面根据指定路径至该 TXT 文档中采集该坐标值, 并将其传输至工业机器人的 PR [1] 位置寄存器之中, 数据传输成功时机器人端的 DO [1] 会变成打开状态, 以便于调用机器人抓取

程序。

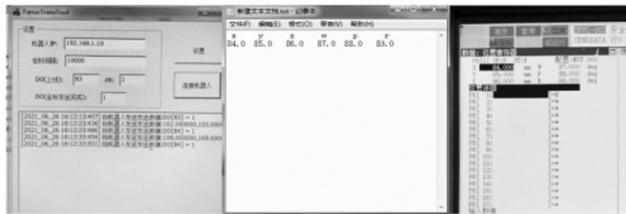


图 22 机器人端 PR [i] 接收数据

### 5.5 工业机器人识别与抓取实验验证

实验任务是实现对工作台上单一物体、多目标物体的识别抓取，并在指定位置码垛，搭建好所有的软硬件环境开始实验，启动识别算法、工业机器人原位等待，神经网络完成定位点坐标求解后，机器人计算各轴关节角进行轨迹规划，对物体实施抓取、搬运并码垛。当摄像机视野中出现多个物体时，会根据定位点横坐标大小排序，优先抓取横坐标小的目标，然后依次抓取。单一物体的识别与抓取效果如图 23 所示。

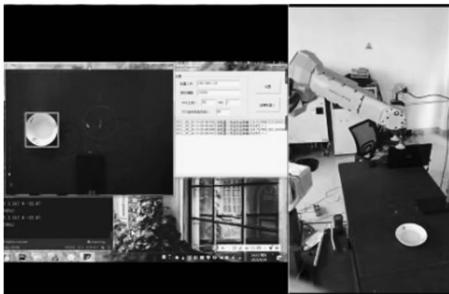


图 23 单一目标识别与抓取

多目标时物体的识别与抓取效果如图 24 所示，机器人首先移向第一个物体对其抓取并放置在指定位置，随后再依次抓取第二个、第三个、第 N 个物体，并在指定位置码垛。

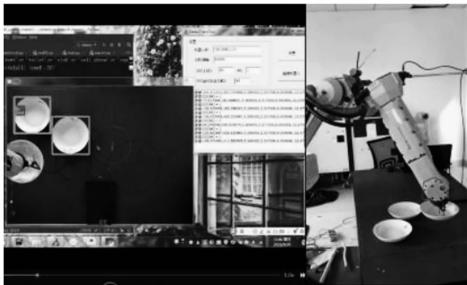


图 24 多目标识别与抓取

表 5 列出了本次实验的针对 3 种物体（盘子、手机、纸盒），在单一目标和多目标检测、定位、抓取中共 180 次的实验数据（正确识别物体的种类与定位坐标前提下），实验的综合平均抓取成功率约为 92.8%。

在三类物体的抓取实验中，盘子的成功率最高，原因是

表 5 抓取实验统计结果

目标名称		次数	成功率/%
单目标	盘子	30	100
	手机	30	90
	纸盒	30	93
多目标	盘子	30	97
	手机	30	90
	纸盒	30	87

盘子的特殊形状对机器人末端气动吸盘的位姿角度没有要求，但是由于手机和纸盒表面是矩形形状，对机器人末端吸盘的位姿有要求，因此导致实验的成功率略低于盘子的成功率。

### 6 结束语

在建立了机器人视觉引导方案的基础上，针对机器人抓取准确率低、反应速度慢的问题，提出了将 YOLOV5 为基础的神经网络模型应用于视觉识别。针对俯视圆形或长方形被抓取物体，提出了目标定位方法，可使神经网络识别坐标转换为机器人坐标系中的定位坐标。提出了六点手眼标定原理，通过六点标定法统一了机器人的用户坐标系与相机的像素坐标系，实验结果显示在 X, Y 方向上的抓取精度均满足工作要求，验证了神经网络工业机器人视觉抓取系统的准确性。

#### 参考文献:

- [1] 张 驰, 郭 媛, 黎 明. 神经网络模型发展及应用综述 [J]. 计算机工程与应用, 2021, 57 (11): 57-69.
- [2] 计时鸣, 黄希欢. 工业机器人技术的发展与应用综述 [J]. 机电工程, 2015, 32 (1): 1-13.
- [3] ORCHARD G, LAGORCE X, POSCH C, et al. Real-time event-driven spiking neural network object recognition on the SpiNNaker platform [C] //2015 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2015: 2413-2416.
- [4] 汤博宇, 焦良葆, 徐 逸, 等. 基于改进 ResNet-50 的图像特征提取网络 [J]. 计算机测量与控制, 2023, 31 (6): 162-167.
- [5] 高茂源, 王好臣, 赵锦泽, 等. 双目视觉引导机器人码垛定位技术的研究 [J]. 计算机测量与控制, 2020, 28 (1): 219-222.
- [6] ZHANG Z. Flexible camera calibration by viewing a plane from unknown orientations [C] //Proceedings of the Seventh IEEE International Conference on Computer Vision. IEEE, 1999: 666-673.
- [7] DATTA A, KIM J S, KANADE T. Accurate camera calibration using iterative refinement of control points [C] //2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops. IEEE, 2009: 1201-1208.
- [8] TSAI R Y, LENZ R K. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration [J]. IEEE

- Transactions on Robotics & Automation, 1989, 5 (3): 345 - 358.
- [9] LIU Y, WANG Q L, LI Y. Calibration of a robot hand-eye system with a concentric circles target [C] //2015 12th International Bhurban Conference on Applied Sciences and Technology (IBCAST). IEEE, 2015: 204 - 209.
- [10] LI W L, XIE H, ZHANG G, et al. Hand-eye calibration in visually-guided robot grinding [J]. IEEE Transactions on Cybernetics, 2015, 46 (11): 2634 - 2642.
- [11] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39 (4): 640 - 651.
- [12] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37 (9): 1904 - 1916.
- [13] 李传浩. 基于卷积神经网络的机器人自动抓取规划研究 [D]. 哈尔滨: 哈尔滨工业大学, 2018.
- [14] SHROFF P, CHEN T, WEI Y, et al. Focus longer to see better: Recursively refined attention for fine-grained image classification [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 868 - 869.
- [15] SHANNON C E. Communication theory of secrecy systems [J]. The Bell System Technical Journal, 1949, 28 (4): 656 - 715.
- [16] 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: 3156 - 3164.
- [6] 卢思安, 侯国庆. 基于大数据分析技术的云计算资源预测研究 [J]. 计算机仿真, 2022, 39 (10): 502 - 505.
- [7] 姬晓鹏, 任鑫芳, 周震震, 等. 三维全景漫游中的动态物体消除与视图修补方法 [J]. 计算机辅助设计与图形学学报, 2022, 34 (8): 1147 - 1159.
- [8] 林晓玲, 王志强, 郭岩岩, 等. 基于多约束场景的BFO-ACO漫游路径规划 [J]. 深圳大学学报 (理工版), 2022, 39 (4): 463 - 471.
- [9] 方卫宁, 王健新, 陈悦源. 复杂系统人机交互中的多任务调度策略综述 [J]. 包装工程, 2021, 42 (18): 73 - 83.
- [10] 蒋成龙, 杨应迪, 黄建达. 基于OpenGL的矿井三维可视化通风系统构建 [J]. 煤矿安全, 2022, 53 (1): 128 - 133.
- [11] 刘峰, 居昊, 于宗良. 基于全卷积孪生神经网络的复杂监控场景下前景提取方法 [J]. 南京航空航天大学学报, 2021, 53 (5): 743 - 750.
- [12] 朱瑞金, 郝东光, 胡石峰. 小样本条件下基于卷积孪生网络的变压器故障诊断 [J]. 电力系统及其自动化学报, 2021, 33 (1): 64 - 69.
- [13] 官国宇, 王运豪, 别一鸣. 周期数据时段划分的双截断高斯混合模型及其EM算法 [J]. 数理统计与管理, 2022, 41 (1): 108 - 123.
- Recognition, 2016: 779 - 788.
- [17] OZA P, PATEL V M. One-class convolutional neural network [J]. IEEE Signal Processing Letters, 2018, 26 (2): 277 - 281.
- [18] CHEN D, JACOB L, MAIRAL J. Convolutional kernel networks for graph-structured data [C] //International Conference on Machine Learning. PMLR, 2020: 1576 - 1586.
- [19] CHEN M, WEI Z, HUANG Z, et al. Simple and deep graph convolutional networks [C] //International Conference on Machine Learning, PMLR, 2020: 1725 - 1735.
- [20] HYVÄRINEN A, KÖSTER U. Complex cell pooling and the statistics of natural images [J]. Network: Computation in Neural Systems, 2007, 18 (2): 81 - 100.
- [21] JIANG B, LUO R, MAO J, et al. Acquisition of localization confidence for accurate object detection [C] //Proceedings of the European Conference on Computer Vision (ECCV), 2018: 784 - 799.
- [22] YU J, JIANG Y, WANG Z, et al. Unitbox: an advanced object detection network [C] //Proceedings of the 24th ACM International Conference on Multimedia, 2016: 516 - 520.
- [23] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506: 146 - 157.
- [24] 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, 2020, 34 (7): 12993 - 13000.
- [25] NEUBECK A, VAN GOOL L. Efficient non-maximum suppression [C] //18th International Conference on Pattern Recognition (ICPR'06). IEEE, 2006, 3: 850 - 855.
- [14] 钟友坤, 莫海宁. 基于深度自编码-高斯混合模型的视频异常检测方法 [J]. 红外与激光工程, 2022, 51 (6): 375 - 381.
- [15] 朱庆, 张利国, 丁雨淋, 等. 从实景三维建模到数字孪生建模 [J]. 测绘学报, 2022, 51 (6): 1040 - 1049.
- [16] 张国义, 刘华, 高翔霄, 等. 基于"软硬结合"平台的导航战体系对抗与评估系统研究 [J]. 计算机测量与控制, 2022, 30 (1): 266 - 281.
- [17] 申小虎, 韩春润, 戚志如. 伪装对抗人脸识别综合实验的教学设计 [J]. 实验技术与管理, 2022, 39 (1): 167 - 172.
- [18] 袁得崙, 陈世聪, 高见, 等. 基于斯塔克尔伯格博弈的在线社交网络扭曲信息干预算法 [J]. 计算机科学, 2021, 48 (3): 313 - 319.
- [19] 张立红. 社会物理学视域下网络谣言传播模型及仿真 [J]. 自然辩证法通讯, 2021, 43 (11): 104 - 109.
- [20] 郑仟. 基于多旋翼无人机的架空输电线路状态检测系统设计 [J]. 计算机测量与控制, 2022, 30 (12): 332 - 338.
- [21] 郭芷含, 孔令谦, 李斌, 等. 基于单片机的灭火器数据采集实验装置设计 [J]. 实验技术与管理, 2022, 39 (6): 115 - 119.