

# Research on Target Tracking Algorithm Based on YOLO and KCF

Jianfang Liu, Chengjian Li

Computer Department, Pingdingshan University, Pingdingshan Henan  
Email: 61779746@qq.com

Received: May 15<sup>th</sup>, 2020; accepted: May 28<sup>th</sup>, 2020; published: Jun. 4<sup>th</sup>, 2020

## Abstract

In order to solve the problem of tracking offset and target loss caused by the offset of the recording device during target tracking, a target tracking algorithm based on YOLO and KCF is proposed. This paper uses the YOLO algorithm based on the end-to-end design idea of regression to achieve target detection. Before target detection, the video image frame is grayed and average filtered to achieve image enhancement, reduce redundant data and remove noise. Afterwards, the tracking frame of the KCF algorithm is initialized by the YOLO algorithm, and the offset error rate (OER) is set in the tracking process to adjust the tracking window position in real time to achieve real-time tracking of the target. Experimental results show that the method proposed in this paper has higher tracking accuracy and robustness than KCF algorithm and Camshift algorithm when facing the recording device.

## Keywords

YOLO Algorithm, KCF Algorithm, Image Enhancement Method, Target Tracking

# 基于YOLO和KCF的目标跟踪算法研究

刘建芳, 李成建

平顶山学院计算机学院, 河南 平顶山  
Email: 61779746@qq.com

收稿日期: 2020年5月15日; 录用日期: 2020年5月28日; 发布日期: 2020年6月4日

## 摘要

为了解决在目标跟踪过程中因录制设备发生偏移带来的跟踪偏移和目标丢失问题, 提出了基于YOLO和KCF的目标跟踪算法。本文使用基于回归的端到端设计思想的YOLO算法实现目标检测, 在目标检测前, 对视频图像帧进行灰度化和均值滤波法实现图像增强, 减少冗余数据, 去除噪点。随后通过YOLO算法进行KCF算法跟踪框的初始化, 在跟踪过程中设置偏移误差率(Offset error rate, OER)实时调整跟踪窗口位置, 实现目标实时准确跟踪。实验结果表明, 本文所提出的方法在面对录制设备发生偏移时相比于KCF

算法、Camshift算法有较高的跟踪准确率和鲁棒性。

## 关键词

YOLO算法, KCF算法, 图像增强法, 目标跟踪

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着 AI 技术的不断发展, 其子领域计算机视觉技术也获得了突飞猛进的进步, 计算机视觉即通过机器实现“人眼”对事物的测量和判别能力。目前, 计算机视觉技术主要应用于智能视频监控、智能交通、军事应用等领域, 其中智能视频监控是通过摄像机获取场景和目标的视频信息, 使用算法实现目标检测和跟踪, 对目标行为进行识别和场景理解。

本文主要对目标跟踪技术进行研究, 目标跟踪是计算机视觉领域的研究热点, 在军事制导、视觉导航等领域中有着广阔的应用前景, 跟踪算法的研究具有重要的实际意义和理论价值[1]。

目标跟踪技术就是在给定视频序列初始帧的目标大小与位置的情况下, 预测后续帧中该目标的大小与位置。为了提高目标跟踪技术的跟踪准确率和鲁棒性, 首先进行目标检测获取目标位置和大小, 目前常见的深度学习目标检测算法有两类, 一类是基于候选区域的目标检测算法, 典型算法如: R-CNN [2]、Fast R-CNN [3]、Faster R-CNN [4]; 另一类是基于回归的目标检测算法, 典型算法如: YOLO [5]、SSD [6]等。其中 R-CNN 算法是基于候选区域算法的基础, 后续此类算法都是对 R-CNN 算法进行改进以获得更高的识别速度和识别率的算法。R-CNN 的基本思想是利用选择搜索算法(Selective Search)在图像中提取可能包含目标的候选区域, 然后用 CNN 提取特征, 实现目标跟踪[7]。而基于回归的 YOLO 和 SSD 算法, 将分类和目标预测同时进行, 其中 YOLO 算法每秒可处理 45 帧图像能够满足目标跟踪的实时性。

视频图像帧目标检测后, 虽然获得了目标位置但缺乏对视频帧目标运动信息的连续性。传统的目标跟踪算法如 KCF (Kernel Correlation Filter)算法具有较强的实时性, 但当录制设备发生偏移时, 会出现目标跟踪偏移, 甚至丢失目标。

为了解决设备偏移带来的跟踪偏移和目标丢失问题, 本文采用基于 YOLO 和 KCF 的目标跟踪算法, 通过 YOLO 算法初始化 KCF 算法跟踪框, 设置偏移误差率判定何时需对 KCF 算法跟踪框进行更新。

## 2. 图像预处理和目标检测

本文主要通过灰度化和图像增强法中的均值滤波法实现图像预处理, 减少冗余信息, 去除噪点和 YOLO 算法实现目标检测。

### 2.1. 图像增强

图像增强的方法按照不同作用域可分为空域和频域增强, 空域有灰度变换、直方图均衡化和空域滤波等, 频域增强有高通、低通和同态滤波等[8]。本文使用灰度化和均值滤波法减少冗余信息, 去除噪点。

1) 灰度化, 在 RGB 模型中, 如果  $R = G = B$  时, 则彩色表示一种灰度颜色, 其中  $R = G = B$  的值叫灰度值, 因此, 灰度图像每个像素只需一个字节存放灰度值(又称强度值、亮度值), 灰度范围为 0~255。

灰度化如图 1 所示。

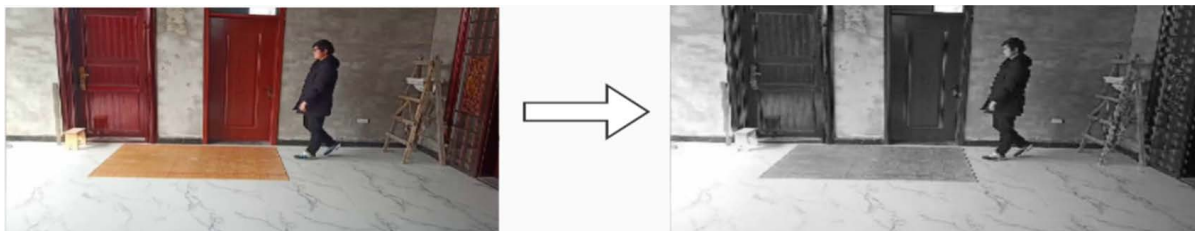


Figure 1. Grayscale image

图 1. 灰度化

2) 均值滤波是典型的线性滤波算法, 该算法在图像目标像素上给定一个模板, 该模板周围包括其周围邻近的像素, 再将模板中的全体像素进行求平均值, 用均值代替原像素值, 以实现像素的平滑, 突出细节信息, 忽略边缘, 达到去噪的目的。均值滤波法如图 2 所示。

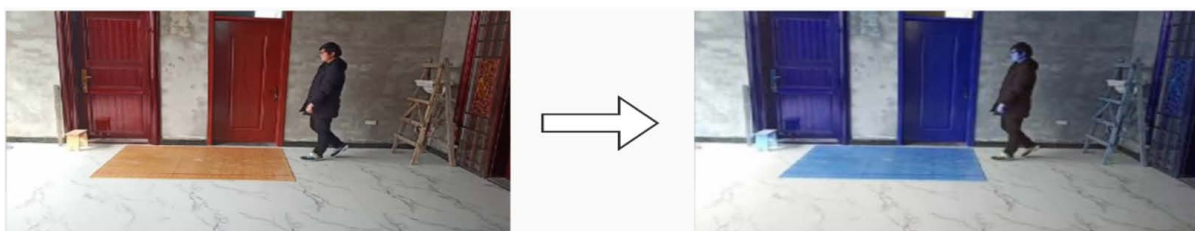


Figure 2. Mean filtering method

图 2. 均值滤波法

## 2.2. 目标检测

### 1) YOLO 算法原理

YOLO 算法在目标检测过程中采用了针对目标检测任务设计的 CNN 进行特征提取, 通过全连接层对识别出来的目标进行位置预测和分类, YOLO 算法的网络模型由输入层, 卷积层, 池化层, 全连接层构成, 如图 3 所示。

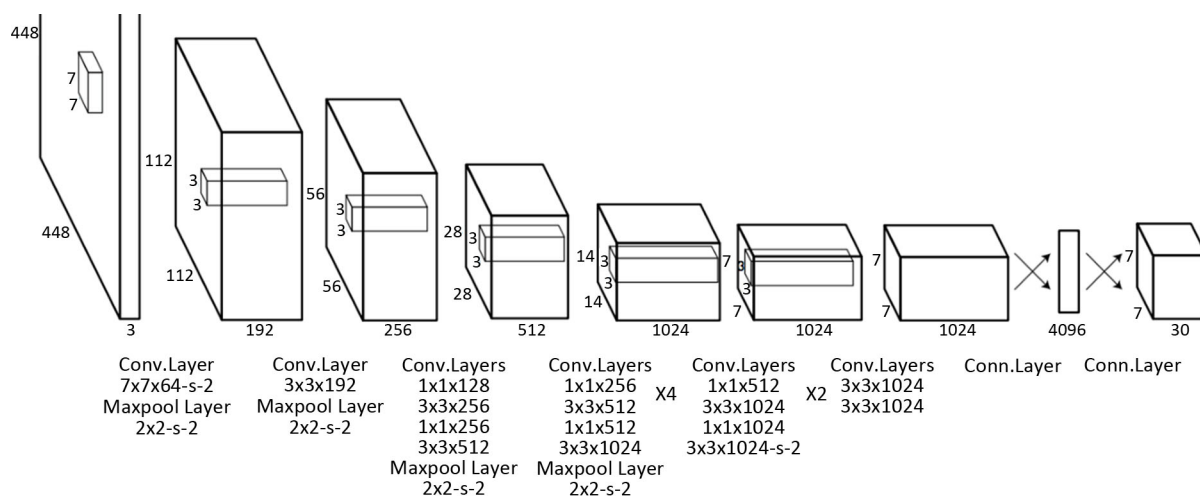


Figure 3. YOLO network [9]

图 3. YOLO 网络模型[9]

YOLO 算法的输入层, 将输入的图像剪切  $448*448*3$  的三通道图像, 三通道是指图像由红, 绿, 蓝三原色叠加组成。全连接层要求固定大小的向量作为输入, 在向前层网络反馈数据时要求原图像大小固定尺寸, 因此输入图像剪切为  $448*448$  固定大小。

输入层之后是 24 层卷积层, 通过对输入图像进行卷积运算获得特征图, 记录特征数据方便后续的分类和目标定位。YOLO 网络模型中采用  $3*3$  和  $1*1$  的卷积核, 其中  $1*1$  的卷积核主要用来降低卷积核通道数量, 减少后续运行参数。

YOLO 网络模型的池化层和卷积神经网络池化层功能相同, 用一个值代替对应区域像素点。YOLO 网络模型中使用最大池化, 用卷积后的图像区域最大值来代替原图像区域, 减少冗余数据, 防止过拟合。

YOLO 网络模型使用两个全连接层将卷积池化后多维数据“拉伸”成为一维向量, 为后面的分类和位置确定做铺垫。

YOLO 网络模型的最后一层输出层与 CNN 算法中的 SoftMax 分类器实现将全连接层数据分类输出相似, 输出的特征图个数就是目标的分类数, 但也有不同 YOLO 算法输出层输出一个  $7*7*30$  的张量,  $7*7$  与输入层的  $7*7$  网格相对应, 30 代表着图像中物体的分类结果以及位置信息编码。最后通过统一约定的方式对该向量进行解码在原图像中绘制出检测结果。

## 2) 检测流程

YOLO 算法将输入的图像划分为  $S*S$  的网格, 每个网格负责检测落在中心点的目标并获得目标边框。每个目标边框由五维预测参数组成即中心点偏移( $x,y$ ), 宽高( $w,h$ )和置信度(Confidence)。

置信度计算如公式(1-1)所示

$$C = \Pr(\text{Object}) * IoU \quad (1-1)$$

公式(1-1)中  $\Pr(\text{Object})$ 表示网格目标边框中目标存在的可能性, Object 代表目标对象, IoU (Intersection over Union, 交并比)用于展示当前模型预测到的目标边框位置的准确性, 表达式如公式(1-2)所示。

$$IoU_{pre}^{true} = \frac{\text{box}(pre) \cap \text{box}(true)}{\text{box}(pre) \cup \text{box}(true)} \quad (1-2)$$

公式(1-2)中  $\text{box}(pre)$ 表示预测的目标边框,  $\text{box}(true)$ 表示真实目标边框。

YOLO 算法检测流程如图 4 所示。

## 3. 联合跟踪算法

### 3.1. KCF 算法

KCF 算法的思想主要是鉴别式追踪方法, 该方法主要是在追踪过程中训练一个目标检测器, 使用所需的目标检测器去检测下一帧视频图像位置, 来预测该位置是否是所需要的目标, 然后再使用新的检测结果去更新训练集[10]。以此来实现更新目标检测器。通过在频域内的傅里叶变换和脊回归, 将目标区域作为正样本, 周围不包括目标的区域视为负样本, 选定的目标区域和周围区域的循环矩阵进行水平、垂直的循环移位来产生大量的样本, 实现目标检测器的训练。

#### 1) 算法推导

设训练样本集为 $(x_i, y_i)$ , 则其线性回归函数为公式(2-1)所示。

$$f(x_i) = w^T x_i \quad (2-1)$$

其中,  $w$  代表列向量的权重系数, 为了降低损失函数, 优化算法, 于是采用最小二乘法来求解如公式(2-2)所示。

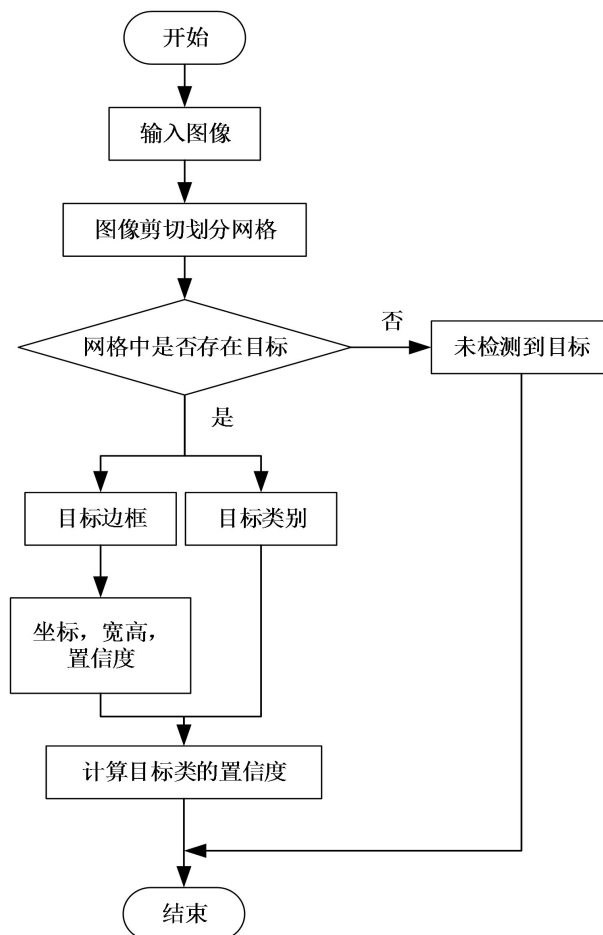


Figure 4. YOLO algorithm detection process  
图 4. YOLO 算法检测流程

$$\min_w \sum_i (f(x_i) - y_i)^2 + \lambda \|w\|^2 \quad (2-2)$$

其矩阵形式为

$$\min_w \|X_w - y\|^2 + \lambda \|w\|^2 \quad (2-3)$$

公式(2-2)又称脊回归, 其中  $\lambda \|w\|^2$  是最小化风险参数, 又称为惩罚项,  $\lambda$  是正则化参数, 可以防止过拟合, 计算权值  $w$ , 求解如公式(2-4)所示。

$$w = (X^T X + \lambda I)^{-1} X^T y \quad (2-4)$$

其中, 行向量矩阵表示为  $X = [x_1, x_2, x_3, x_4, \dots, x_n]^T$ ,  $y$  表示回归值,  $y_i$  是列向量, 因为考虑到傅里叶变化中有负数的存在, 所以要计算复数矩阵, 将式(2-4)转换为复数形式如公式(2-5)所示。

$$\bar{w} = (X^H X + \lambda I)^{-1} X^H y \quad (2-5)$$

其中,  $X$  的共轭复数转置矩阵为  $X^H$ 。

## 2) KCF 算法的目标检测

在当前帧选取与前一帧目标所在位置的目标区域坐标, 提取基本候选样本, 设样本大小为  $w * h$  的图像

块, 建立目标区域模型。通过循环移位获取一系列样本集合, 设为  $G_i (i \in \{0, 1, 2, \dots, w-1\} \times \{0, 1, 2, \dots, h-1\})$ 。对每个候选样本进行权值系数计算, 再根据  $S'_i = \sum_j a_i k(x_j, g_i)$  计算检测样本与目标相似度进行加权求和, 得到下一帧的响应值为了提高运行速度, 求解转换为频域如公式(2-6)所示。

$$S = F^{-1} [F(k) \cdot F(\vartheta)] \tag{2-6}$$

KCF 跟踪算法具有简洁、效果好、速度快的特点, 尤其是通过傅里叶变换的推导, 在频域范围内计算的速度非常快, 跟踪效果也非常显著[11]。但当因环境问题即本文讨论的录制设备发生偏移时导致无法准确跟踪或丢失目标。

因此本文采用 YOLO 和 KCF 算法联合的目标跟踪算法, 用于解决设备偏移追踪不准确或丢失目标的问题。

### 3.2. 联合跟踪算法设计

#### 1) 偏移误差率(Offset error rate, OER)

为了解决 KCF 算法因设备偏移带来的跟踪偏移和目标丢失问题, 使用 YOLO 算法进行 KCF 算法的初始化, 在跟踪过程中, 将 KCF 算法预测的目标位置  $(kx, ky)$  中的  $kx$  与 YOLO 算法目标检测位置  $(x, y)$  中的  $x$  进行比较, 两种算法在  $x$  轴方向上的比值定义为偏移误差率(Offset error rate, OER)如公式(2-7)所示。通过 OER 来判断是否使用 YOLO 算法对 KCF 算法跟踪框进行更新。

$$OER = x/kx \tag{2-7}$$

#### 2) 算法设计

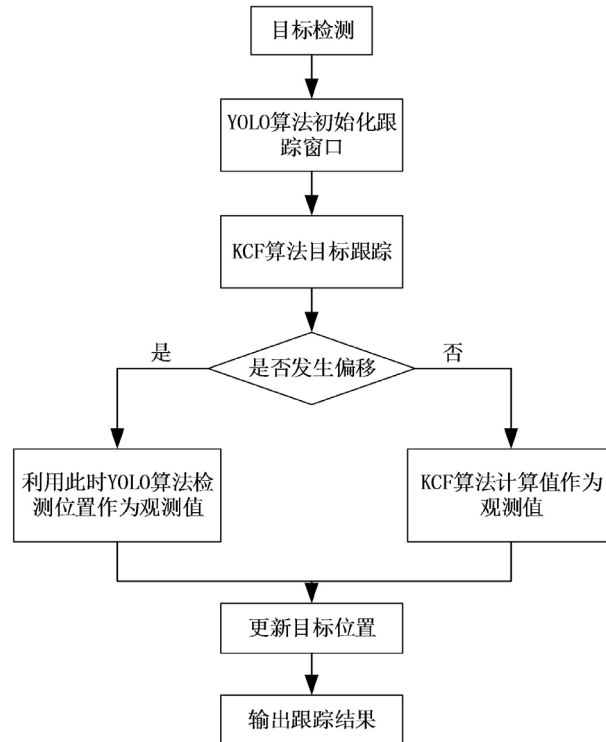


Figure 5. Flow chart of target tracking algorithm combined with YOLO and KCF

图 5. 基于 YOLO 和 KCF 的目标跟踪算法流程图

本文提出的基于 YOLO 和 KCF 的目标跟踪算法, 相较于传统跟踪算法提高了其跟踪准确率和鲁棒性。算法步骤如下:

Step1 初始化: YOLO 算法初始化 KCF 算法跟踪框;

Step2 KCF 算法跟踪: 利用 KCF 算法跟踪目标, 通过 OER 判断是否发生偏移或丢失目标现象;

1) 有偏移: 若  $OER \geq \beta$ , 其中  $\beta$  为偏移阈值, 则认为目标跟踪发生偏移。在判断出目标发生偏移后, 将此时的 YOLO 算法目标边框赋值 KCF 算法重新跟踪。

2) 无偏移: 若  $OER < \beta$ , 则认为此时目标被跟踪, 用 KCF 算法继续跟踪。

3) 丢失目标: 若  $OER = 0$ , 此时 KCF 算法丢失目标。通过 YOLO 算法对 KCF 算法再次进行初始化实现目标跟踪。

基于 YOLO 和 KCF 的目标跟踪算法流程图如图 5 所示。

#### 4. 实验结果及分析

为了更加直观的展示本文提出的基于 YOLO 和 KCF 的目标跟踪算法的跟踪效果, 将 KCF 算法、Camshift 算法和基于 YOLO 和 KCF 的目标跟踪算法进行实验对比分析。

实验 1: 选取录制设备未发生偏移的视频, 视频中背景环境对跟踪目标的干扰较小。图 6、图 7 和图 8 分别是视频序列的第 6、84、187 帧, 用矩形框表示被跟踪目标的位置和区域。



Figure 6. Tracking effect of KCF algorithm

图 6. KCF 算法跟踪效果



Figure 7. Target tracking effect of Camshift algorithm

图 7. Camshift 算法跟踪效果



Figure 8. Target tracking effect of YOLO and KCF algorithm

图 8. 基于 YOLO 和 KCF 算法的目标跟踪效果

实验 2: 选取录制设备发生偏移的视频, 视频中背景环境对跟踪目标的干扰较小。图 9、图 10 和图 11 分别是视频序列的第 6、84、187 帧, 用矩形框表示被跟踪目标的位置和区域。



Figure 9. Tracking effect of KCF algorithm  
图 9. KCF 算法跟踪效果



Figure 10. Target tracking effect of Camshift algorithm  
图 10. Camshift 算法跟踪效果



Figure 11. Target tracking effect of YOLO and KCF algorithm  
图 11. 基于 YOLO 和 KCF 算法的目标跟踪效果

KCF 算法在录制设备有无偏移的两种情况下目标跟踪效果如图 6 和图 9 所示。通过对图 9(b)和图 9(c)的观察, 当人体目标处于跟踪状态出现录制设备偏移情况时, 算法的跟踪结果不准确, 跟踪框发生了较大的偏移, 彻底地丢失人体目标。

Camshift 算法在实验 1 和实验 2 的效果分别如图 7 和图 10 所示。在未发生录制设备偏移时, 如图 7(b)所示, 跟踪效果没有影响。当录制设备发生偏移时, 如图 10 所示 Camshift 算法跟踪目标丢失。

基于 YOLO 和 KCF 算法在实验 1 和实验 2 的效果分别如图 8 和图 11 所示。在实验 2 的跟踪过程中, 当出现录像位置偏移时, 如图 11(b)和图 11(c)所示并未发生目标丢失或跟踪偏移现象。通过上述算法的比较证明了基于 YOLO 和 KCF 的目标跟踪算法具有较高的准确率和鲁棒性。

## 5. 结论

本文针对在目标跟踪过程中因录制设备发生偏移导致跟踪偏移或目标丢失的问题, 通过 YOLO 算法对 KCF 算法进行跟踪框初始化, 设置 OER 作为偏移误差率, 判定是否更新 KCF 算法跟踪框。基于 YOLO



和 KCF 的目标跟踪算法相较于 KCF 算法、Camshift 算法具有较高的跟踪准确率和鲁棒性。但本算法时间复杂度较高, 在未来会将研究重心放到设计更加合理的 YOLO 算法结构模型和 KCF 算法优化, 以提高算法运行效率。

## 基金项目

河南省科技厅科技发展计划科技攻关项目(No.182102310040); 平顶山学院青年科研基金项目(No. PXY-QNJJ-2018005)。

## 参考文献

- [1] 李均利, 尹宽, 储诚曦, 等. 视频目标跟踪技术综述研究[J]. 燕山大学学报, 2019, 43(3): 251-262.
- [2] Girshick, R, Donahue, J., Darrell, T., *et al.* (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/cvpr.2014.81>
- [3] Girshick, R. (2015) Fast R-CNN. 2015 *IEEE International Conference on Computer Vision (ICCV)*, Santiago, 7-13 December 2015, 1440-1448.
- [4] Ren, S., He, K., Girshick, R., *et al.* (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. <https://doi.org/10.1109/tpami.2016.2577031>
- [5] Redmon, J., Divvala, S., Girshick, R., *et al.* (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/cvpr.2016.91>
- [6] Liu, W., Anguelov, D., Erhan, D., *et al.* (2015) SSD: Single Shot Multi Box Detector. arXiv:1512.02325 [cs.CV]
- [7] 韩鹏, 沈建新, 江俊佳, 等. 联合 YOLO 和 Camshift 的目标跟踪算法研究[J]. 计算机系统应用, 2019, 28(9): 271-277.
- [8] 郝志成, 吴川, 杨航, 朱明. 基于双边纹理滤波的图像细节增强方法[J]. 中国光学, 2016, 9(4): 423-431.
- [9] 阮激扬. 基于 YOLO 的目标检测算法设计与实现[D]: [硕士学位论文]. 北京: 北京邮电大学, 2019.
- [10] Comaniciu, D., Ramesh, V. and Meer, P. (2000) Real-Time Tracking of Non-Rigid Objects Using Mean Shift. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, Hilton Head Island, 15 June 2000, 142-149. <https://doi.org/10.1109/cvpr.2000.854761>
- [11] 张岩. 室内场景下行人检测与跟踪技术的研究[D]: [硕士学位论文]. 北京: 北京工业大学, 2017.