

# 基于深度学习的客流量统计方法<sup>①</sup>



韩晓微, 王雨薇, 谢英红, 齐晓轩

(沈阳大学 信息工程学院, 沈阳 110044)

通讯作者: 韩晓微, E-mail: hwx69@163.com

**摘要:** 针对目标快速移动和遮挡情况导致的客流统计存在误差的问题, 设计目标的检测、跟踪、进门行为判断等策略, 提出基于深度学习的餐饮业客流统计方法. 首先, 通过多数据集对 YOLOv3-tiny 模型进行训练, 实现对于小目标的准确检测; 进而设计多通道特征融合的目标跟踪算法, 完成目标快速移动情况下的稳定跟踪; 最后设计目标进门行为的判断方法, 通过重叠率对目标的进门行为进行判断, 实现对进门客流量的准确统计. 最终通过实验验证, 客流量统计的平均准确率达到 93.5%.

**关键词:** 深度学习; 目标检测; 目标跟踪; 客流统计

引用格式: 韩晓微, 王雨薇, 谢英红, 齐晓轩. 基于深度学习的客流量统计方法. 计算机系统应用, 2020, 29(4): 24-31. <http://www.c-s-a.org.cn/1003-3254/7311.html>

## Customer Flow Statistics Method Based on Deep Learning

HAN Xiao-Wei, WANG Yu-Wei, XIE Ying-Hong, QI Xiao-Xuan

(School of Information Engineering, Shenyang University, Shenyang 110044, China)

**Abstract:** Aiming at the problem of error in customer flow statistics caused by fast moving and shielding of the target when entering the door, this study designs the strategies of target detection, tracking, and behavior judgment of entering the door, and puts forward the method of customer flow statistics in catering industry based on deep learning. The YOLOv3-tiny model is trained by multi-data set, and the accurate detection of small target is realized. The target tracking algorithm of multi-channel feature fusion is designed to achieve the stable tracking in the case of fast target movement. In this study, we design a method to judge the entry behavior of the target through overlapping rate, and realize the accurate statistics of the entrance passenger flow. The experimental results show that the average accuracy rate of passenger flow statistics is 93.5%.

**Key words:** deep learning; target detection; target tracking; customer flow statistics

当前, 智能视频监控是计算机视觉的重要应用领域, 其不同于原有的视频监控方法, 智能视频监控通过计算机视觉方法对各种图像序列进行处理, 从而自动的实现对行人、交通等情况的有效监控<sup>[1,2]</sup>, 而人流统计则是近年来此领域的关键技术之一<sup>[3]</sup>. 人流统计主要通过目标检测、跟踪等多方法结合实现, 此技术在当

前已有了一定的研究成果, 但由于目标的快速移动与遮挡等情况的存在, 使得客流统计仍存在着严峻的挑战.

对于目标检测与跟踪方法, Felzenszwalb 等人<sup>[4]</sup>提出了多类目标检测方法, 其先通过 DPM 的人工特征提取, 再进行 Latent SVM 分类, 此方法成为在深度卷积神经网络出现之前最有效的识别算法. Girshick 等人<sup>[5]</sup>

① 基金项目: 辽宁省重点研发计划 (2018104012); 沈阳市科技计划 (18013015); 沈阳市双百工程计划 (Z18-5-013)

Foundation item: Key Research and Development Program of Liaoning Province (2018104012); Science and Technology Plan of Shenyang Municipality (18013015); Double Hundred Project of Shenyang Municipality (Z18-5-013)

收稿时间: 2019-08-14; 修改时间: 2019-09-06; 采用时间: 2019-09-10; csa 在线出版时间: 2020-04-05

首次首次在目标检测中使用深度学习理论,其通过卷积神经网络实现目标候选框特征的提取,之后利用SVM实现分类工作. Ren等人<sup>[6]</sup>提出了用于实时目标检测的候选框网络RPN,其可以从任意尺寸的图像中得到多个候选区,能够与卷积神经网络结合实现端到端的训练,此网络与Fast R-CNN<sup>[7]</sup>结合并共享卷积层特征之后的网络称为Faster R-CNN. Henriques等人<sup>[8]</sup>提出了一种核相关滤波器,其通过加入正则项建立模型,利用循环矩阵进行稠密采样,又通过多通道特征的提取,实现了较为鲁棒的跟踪. Mueller等人<sup>[9]</sup>提出了上下文感知滤波跟踪算法,实现了背景干扰下的有效跟踪.

对于人流统计方法, Raheja等人<sup>[10]</sup>采用高斯混合模型对背景像素进行描述,提出了一种多模态背景的双向人员计数算法,实现了在人口密度较低情况下的准确计数. Wateosot等人<sup>[11]</sup>通过对对象分割对人的头部进行识别,然后使用粒子滤波方法跟踪头部的深度颜色特征,进而实现对于人头的统计工作. Songchenchen等人<sup>[12]</sup>提出了一种基于特征融合的人群计数方法,其通过图像特征提取和纹理特征分析,利用多源数据对人群数量进行估计,实现了高密度、静态图像中的人群统计. Chauhan等人<sup>[13]</sup>提出了一种基于深度学习框

架的数学计数回归方法,该方法首先估计出每个区域内的人数,之后通过各个区域的人群数量估计之和求得整个图像中总人数. 李笑月<sup>[14]</sup>通过SVM对样本进行训练得到行人分类器,利用Mean-Shift算法进行目标跟踪,判断目标是否离开场景来实现计数. 这些算法多是使用垂直于地面的摄像头进行监控,通过这种方式能够有效的应对目标遮挡情况,但当监控视频角度发生偏转时,算法的准确性会受到影响.

本文以餐饮业为背景,为实现对进店客流量的统计,设计基于深度学习的餐饮业客流量统计方法. 此方法首先通过多数据集对YOLOv3-tiny方法进行训练,实现对于目标的准确检测,然后设计多通道特征融合的核相关滤波目标跟踪算法,实现对于快速移动目标的稳定跟踪,最后设计基于双标准框的目标进门行为判断方法,实现了对于客流量的有效统计.

## 1 本文算法设计

### 1.1 YOLOv3 目标检测算法

YOLOv3<sup>[15]</sup>在原始两个版本<sup>[16,17]</sup>的基础上进行了多尺度预测与网络结构等方面的改进,使其检测的速度更快,准确率更高. YOLO各个版本的性能对比见表1所示.

表1 YOLO各版本性能对比

算法	机制	优点	缺点	适用范围
YOLOv1	设置图像输入尺寸,图像经过24层卷积与一层全连接后得到特征图,将图片划分为 $S \times S$ 个网格,每个网格预测 $B$ 个边界框的置信度和目标位置,每个网格根据训练数据种类的不同预测 $C$ 个条件类别的概率.	算法简洁速度快;能很好的避免背景错误;可以学到物体的泛化特征.	输入的尺寸是固定的;对小的目标检测效果不好;容易产生物体的定位错误.	快速移动的目标,抽象艺术品.
YOLOv2	在YOLOv1基础上对网络结构进行了修改,依赖于DarkNet-19的结构,取消原有的全连接层,改成了全卷积结构,同时去掉了YOLOv1的最后一个池化层,提增了特征的分辨率.将检测和分类问题做成了一个统一的框架.	比YOLOv1速度更快;适应不同输入尺.	未解决重叠的分类问题.	快速移动的目标.
YOLOv3	在YOLOv2基础上对网络结构进行了修改,在Darknet-19中添加残差网络的混合方式,使用连续的 $3 \times 3$ 和 $1 \times 1$ 卷积层,构建了Darknet-53结构.使用了标签预测和跨尺度预测,实现模型性能的进一步提升.	对于重叠的标签能够更好的检测;对小目标实现准确检测.	因网络加深导致速度降低.	小目标,快速移动的目标.
YOLOv3-tiny	YOLOv3-tiny版本对YOLOv3网络进行了压缩,减少了YOLOv3中的卷积层,没有使用残差层,只使用了两个不同尺度的输出层.	速度比YOLOv3快.	准确度低于YOLOv3.	小目标,快速移动的目标.

#### 1.1.1 YOLOv3 算法原理

YOLOv3的输入图片尺寸为 $416 \times 416$ 像素,通道数为3,其采用了Darknet-53的网络结构,主要实现YOLOv3的特征提取工作,其共包含53个卷积层,卷

积层中使用的卷积为 $3 \times 3$ 和 $6 \times 6$ ,具体结构<sup>[18]</sup>见图1所示.

Darknet-53之后通过类似于FPN网络<sup>[19]</sup>实现特征融合,网络总共划分3个尺度,分别为 $13 \times 13$ 、 $26 \times 26$ 、

52×52, 其中每个尺度先构成各自尺寸的特征图, 再使用卷积核实现各特征图间的特征融合. 此网络通过在多个尺度的融合特征图上进行独立的检测工作, 最终实现对于小目标检测性能的有效提升.

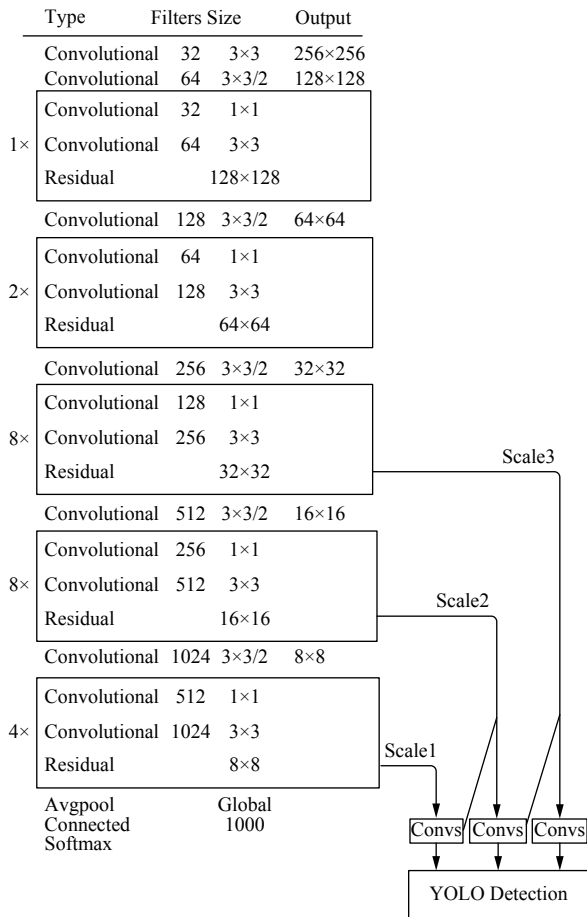


图1 YOLOv3 网络结构

YOLO 在训练时的损失函数计算公式为:

$$\lambda_{loss} = E_{coo} + E_{IOU} + E_{class} \quad (1)$$

其中,  $E_{coo}$  为检测结果与标定数据间的坐标误差,  $E_{IOU}$  表示两者间的 IOU 误差,  $E_{class}$  则表示分类误差.

$E_{coo}$  的计算公式为:

$$E_{coo} = \lambda_c \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \lambda_c \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} \left[ (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right] \quad (2)$$

其中,  $\lambda_c$  为权重误差,  $I_{ij}^{obj}$  表示在第  $i$  个网格的第  $j$  个预测框负责目标,  $S$  表示特征图的网格尺寸,  $B$  为每个网

格预测框的数量,  $w_i$  和  $h_i$  分别表示预测框的宽和高,  $\hat{w}_i$  和  $\hat{h}_i$  代表真实的宽高,  $(x_i, y_i)$  代表预测框的中心坐标,  $(\hat{x}_i, \hat{y}_i)$  则代表真实坐标.

$E_{IOU}$  的计算公式为:

$$E_{IOU} = \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (c_i - \hat{c}_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} (c_i - \hat{c}_i)^2 \quad (3)$$

其中,  $\lambda_{noobj}$  表示分类误差的权重,  $I_{ij}^{noobj}$  表示第  $i$  个网格的第  $j$  个预测框不负责目标,  $c_i$  表示第  $i$  个网格预测的置信度值,  $\hat{c}_i$  表示其真实置信度值.

$E_{class}$  的计算公式为:

$$E_{class} = \sum_{i=0}^{S^2} I_i^{obj} \sum_{k=0}^C (p_i(k) - \hat{p}_i(k))^2 \quad (4)$$

其中,  $I_i^{obj}$  表示在第  $i$  个单元格中存在目标,  $p_i(k)$  表示在第  $i$  个单元格中属于第  $k$  类别的预测概率,  $\hat{p}_i(k)$  则属于概率的真实值.

YOLOv3-tiny 的网络结构在 YOLOv3 的基础上进行了修改, 减少了卷积层、取消了残差层, 并且减少了输出尺度的数量. 此模型的整体性能虽不如 YOLOv3, 但由于 YOLOv3-tiny 的模型复杂度得到了减小, 使得其运算速度实现了进一步的提升, 而本文需要在 CPU 环境下实现, YOLOv3-tiny 能够保证整体算法的实时性, 因此本文选取了 YOLOv3 的轻量级版本 YOLOv3-tiny 作为最终的目标检测模型.

### 1.1.2 训练过程

#### (1) 样本准备

使用 LabelImg 对图片序列中的人群进行标注, 实验共涉及 45 家门店不同时间段的 5000 张图片, 具有普适性, 通过这些标注好数据集对模型进行训练. 本文算法针对目标为行人, 因此在模型训练时取消了对其它类别事物的识别, 只保留了 person 标签.

#### (2) 参数设置

本文对模型的训练阶段, 动量参数设置为 0.9, 其影响梯度下降到最优值时的速度, 训练使用小批量随机梯度下降进行优化. 初始学习率  $\eta = 0.001$ , 衰减系数设置为 0.005, 此系数的设置防止过拟合现象出现. 前 1000 次训练的学习率为  $\eta_{lr} = \eta \times (N_{batch}/1000)^2$ , 通过学



习率的设置保持网络的稳定,  $N_{\text{batch}}$  为模型运行的批次  
数, 之后的学习率为  $10^{-3}$ .

### 1.1.3 检测结果

本文将训练后的模型分别对 10 家不同门店的不

同时间进行测试, 图 2 为其中 3 个门店的识别效果, 其  
中门店 A 有强光影响, 门店 B 有较多汽车遮挡, 门店  
C 的视野不清晰. 通过图 2 可看出, 此算法对 3 个门店  
均能实现准确的检测, 有较强的鲁棒性.

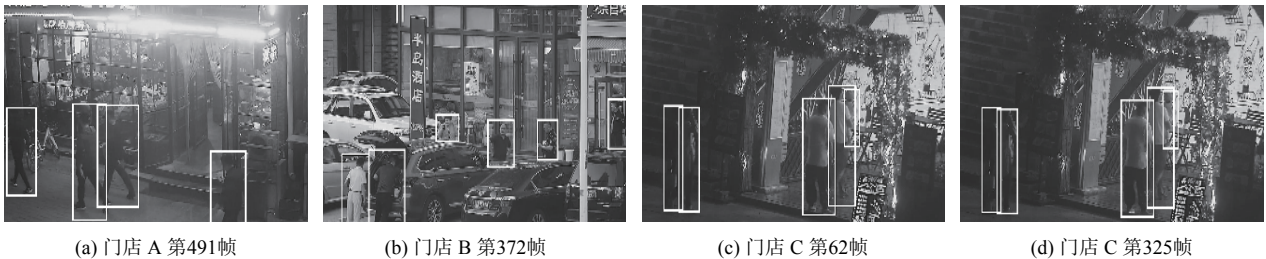


图 2 各门店识别结果

## 1.2 多通道特征融合目标跟踪算法设计

### 1.2.1 核相关滤波算法

核相关滤波器利用正则化的最小二乘法, 通过求  
极值的方法实现对滤波器的求解. 通过对每张图片中  
目标及其周围一定范围的图像块进行滤波器的训练,  
其中目标位置图像块  $x$  的尺寸为  $M \times N$ , 在滤波器训练  
过程中通过循环矩阵实现密集采样, 图像块  $x_i$  中  
 $i \in \{0, \dots, M-1\} \times \{0, \dots, N-1\}$ . 滤波器通过回归最小化  
进行求解, 具体表示为:

$$\min_w \sum_i \langle (w, x_i) - y_i \rangle^2 + \lambda \|w\|^2 \quad (5)$$

其中,  $\langle \cdot \rangle$  表示点积,  $w$  为滤波器的参数,  $y$  为期望的目标  
输出,  $\lambda$  为正则化参数, 能够防止模型过拟合现象, 使模  
型得到更好的泛化能力.

利用核函数  $\kappa(x, y) = \langle \varphi(x), \varphi(y) \rangle$ , 将高维非线性变  
换到低维函数实现加速计算, 其中  $\phi$  表示将特征向量  
映射到核空间的函数. 因此将式 (5) 的解  $w$  表示成输  
入的线性组合<sup>[20]</sup>, 即为:

$$w = \sum_i \alpha_i \phi(x_i) \quad (6)$$

其中,  $\alpha$  是  $w$  的对偶空间变量, 即为滤波器系数.

根据文献[21]可以得到  $\alpha$  的解, 具体求得  $\alpha$  为:

$$\alpha = (k + \lambda I)^{-1} y \quad (7)$$

通过循环矩阵性质可知  $k$  仍是循环矩阵, 结合循  
环矩阵的性质求解到傅里叶域中的  $\alpha$  为:

$$\hat{\alpha} = \frac{\hat{y}}{\hat{k}^{xx} + \lambda} \quad (8)$$

其中,  $\hat{\alpha}$  和  $\hat{y}$  表示他们各自的傅里叶变换.  $\hat{k}^{xx}$  为多通道  
的高斯核函数, 其计算公式为:

$$k^{xx'} = \exp\left(-\frac{1}{\sigma^2} \left(\|x\|^2 + \|x'\|^2 - 2F^{-1}(F(x) \odot F^*(x'))\right)\right) \quad (9)$$

其中,  $\odot$  表示各元素的对应相乘,  $F^{-1}$  表示傅里叶逆变  
换,  $*$  表示复共轭,  $\sigma$  为高斯核函数带宽参数.

通过上述训练过程, 根据循环矩阵  $k$  构建训练样  
本和测试样本的核函数, 将分类器的求解转化到傅里  
叶域中, 对目标下一帧位置进行求解, 其目标响应  
 $f(z)$  为:

$$f(z) = F^{-1}(\hat{k}^{xz} \odot \hat{\alpha}) \quad (10)$$

### 1.2.2 多通道特征融合设计

颜色特征是一种全局特征, 其不受目标的旋转、  
平移等变化的影响. CN 特征<sup>[22]</sup>基于传统颜色特征的优  
势, 将原有的 RGB 的三通道特征进一步映射到 11 维,  
实现了从 3 种颜色到 11 种颜色的扩展. CN 颜色属性  
对于一张输入的目标图像  $I$ , 将  $a$  位置的像素值  $I(a)$  映  
射到 CN 空间, 进一步得到共 11 维的颜色概率特征图  
 $[a_1, a_2, \dots, a_{11}]$ . HOG 特征描述目标的方向梯度, 获取  
HOG 特征首先需进行灰度处理, 然后利用 Gamma 校  
正法对目标进行颜色空间归一化, 以此降低光照变化  
的影响, 从而实现降噪效果; 之后对每个像素点进行梯  
度方向和大小求解, 计算此点的梯度幅值和方向; 最  
后进行细胞单元 (cell)、块 (block) 直方图的构建, 通过  
对块的 HOG 特征进行整体串联, 实现目标框 HOG 特  
征的提取<sup>[23]</sup>.

本文所采用的 HOG 特征的每个 cell 为  $4 \times 4$  像素,

其获取到的特征为 31 维, CN 为 11 维特征, 由于单一的目标特征提取会导致跟踪不够准确的情况, 因此本文提出了一种可以将这两种特征进行融合的方法. 由于核相关函数只需进行向量间的点乘运算, 因此对于两种特征的融合工作就是将这些特征进行维度的串联, 所得的最终目标的特征维数为 42 维, 其融合方式具体表示为:

$$P = [P_C, P_H] \quad (11)$$

其中,  $P_C$ 、 $P_H$  分别表示 CN 和 HOG 特征.

### 1.3 进门行为判断方法设计

本文实现的是对进入门店的客流量统计, 现有的

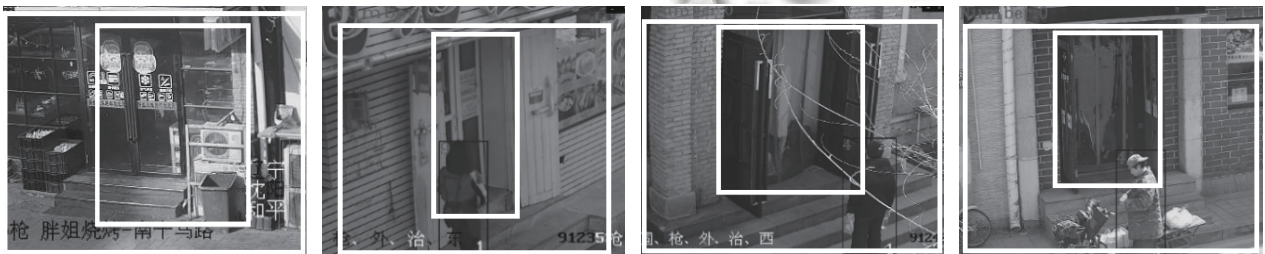


图3 双标准框标定示意图

在实际应用场景中, 不同摄像头距目标门店的距离各不相同, 因此需要对摄像头所拍摄的视频序列进行预处理, 确定目标感兴趣区域. 当目标进入感兴趣区域时开始对其进行整体的识别跟踪工作. 在双标准框确定之后, 本文将参数分别设为 1.2、1.5、2 作为双标准框外框尺寸同比例扩大的倍数, 并且在这些参数下分别对算法实现整体的测试. 经过实验得到当参数为 1.2 时, 目标从进入感兴趣区域到最终进入门店的过程过短, 算法在对目标进行跟踪的过程中会产生更多跟踪漂移的情况. 当参数为 2 时, 有更多的背景被圈定进去, 使得算法整体的运行效率降低. 因此通过实验, 本文将 1.5 作为最后感兴趣区域的参数, 在此参数下, 能够有效的减少其它背景区域的干扰, 提升算法运算效率. 感兴趣区域示意图见图 4 所示.



图4 感兴趣区域示意图

多数算法是通过垂直于地面的摄像头进行监控, 利用分割线对图像进行区域划分, 之后通过对目标经过各分割线的过程实现对垂直视野中的人数统计. 但本文使用的视频监控方式为水平方向或与地面夹角不超过 50 度, 现有的策略并不适用于本文所研究的对进入门店的客流统计背景. 因此本文对目标进门行为判断方法进行了设计, 从而实现对客户流量的统计.

首先初始化双标准框与目标感兴趣区域, 双标准框是根据不同餐饮门店的方位与角度进行手动标定, 其为之后判定目标是否进门提供依据, 标定示意图见图 3 所示.

本文对客户流量的统计是在双标准框基础上, 通过其与同一目标的初始位置与终止位置的相对关系进行确定. 整体判断流程为: 当目标初始位置在门外, 通过跟踪过程目标终止位置在门内, 则定义目标进入了门店, 实现人数累计, 若不满足此条件, 目标则未进门.

通过目标框与内标准框的重叠比例实现对目标与门店的位置关系判断. 标准框的设置能够获取其坐标信息, 目标识别与跟踪工作实现对于目标坐标信息的获取. 其中得到目标框面积为  $S_{obj}$ , 内标准框面积为  $S_{id}$ , 重叠比例即为:

$$R = \frac{S_{id} \cap S_{obj}}{\min(S_{obj}, S_{id})} \quad (12)$$

其中,  $\cap$  为取交集操作,  $\min()$  为求取两个目标框的较小值, 标准框与目标框的尺寸均存在不确定性, 因此保留较小值能有效防止比例漂移的情况.

当初始目标出现时判断目标与内标准框的重叠比例, 本文算法将初始重叠阈值设为 0.4,  $R > 0.4$  时表示目标在门内,  $R < 0.4$  时表示目标在门外. 当目标跟踪结束时, 终止重叠阈值设为 0.8, 当  $R > 0.8$  则表示目标进入门内, 反之目标则未在门内.



## 2 实验结果与分析

本文算法通过 Python 实现, 编辑平台为 Atom, 所使用的计算机为主频 3.40 GHz 的 CPU 以及 8 GB 的内存.

### 2.1 目标跟踪结果分析

为了验证算法的有效性, 本文选取了 10 个门店 30 天不同时间段的视频序列进行测试, 将本文算法与 KCF 算法的结果进行对比试验, 选取了其中两个门店的结果进行分析. 图 5 的门店 A 存在着光照强烈变化

的情况, 初始阶段两个算法均能实现稳定的跟踪, 而当目标弯腰时 KCF 算法产生了漂移现象, 而本文算法能够保持稳定的跟踪. 图 6 的门店 B 存在着车辆遮挡情况, 当目标侧身时 KCF 算法跟踪产生误差现象, 而本文算法能够一直准确的进行跟踪. 这是由于 CN 和 HOG 特征的融合丰富了目标表现特征描述, 能够应对目标表现变化情况, 并且相关滤波器算法通过循环矩阵与核化等方法的使用, 大大降低了算法运算的复杂度, 使得目标在快速移动时能保持稳定的跟踪效果.

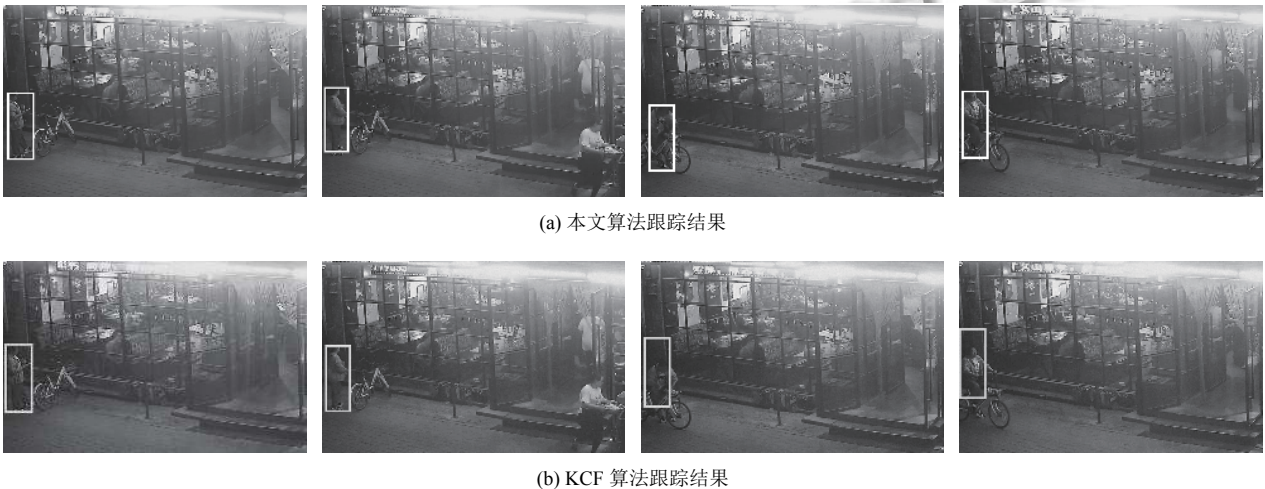


图 5 门店 A 跟踪结果



图 6 门店 B 跟踪结果

### 2.2 客流量统计结果分析

本文所使用的实验数据为多个餐饮业门店的实际监控视频录像, 视频格式为 AVI, 文本共选取了 10 家

门店 30 天的录像进行结果测试, 其中涉及白天与晚上. 图 7 所示为门店 15 一天的客流量统计情况, 其中将视频序列分成 8 段, 前 4 段为白天光照条件较好时的情况

况,后4段为晚上光线较差时的情况,通过该图可知,在多个视频序列下本文算法的统计结果与实际客流量基本一致,只有在晚上时有少量的出入,这是由于晚上的光线较差,目标背景间的差异较小导致的,但本算法在多目标进入门店发生遮挡等情况下有较好的表现,此门店整体统计的准确率为94.8%。图8为10家门店各自一天累计进店客流的统计结果,其中菱形标注本为算法的统计结果,正方形标注为实际人工客流的统计结果,本文方法整体客流统计平均准确率可达93.5%。

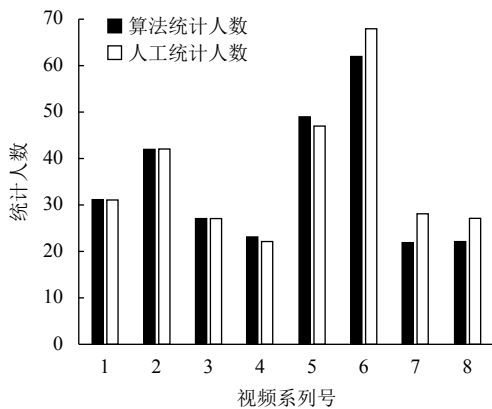


图7 门店15客流统计图

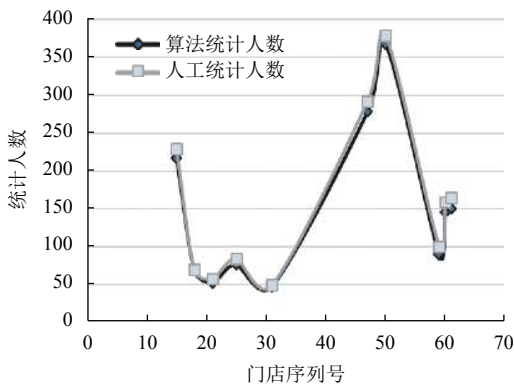


图8 各门店人数累计分布图

为了验证本为算法的准确性与鲁棒性,本文通过10个店铺各自的准确率及误算情况,进行对算法的评估。通过客流量检测值与真实值之间的误差情况进行准确率的计算,误算人数为算法整个过程中出现的错误情况,其中漏检人数包含在误算人数中。

从表2可看出本文算法的准确率几乎均在90%以上,能得到平均准确率为93.5%。表中出现漏检的情况为夜间光线差和特殊天气情况导致的背景与目标间

的差异较小,但在人群较多进入店铺时,本文算法能够得到准确的统计结果,这是由于本文目标进门行为判断算法的设计,使得目标在发生遮挡情况时能够进行稳定的统计。通过表3中本文算法与文献[14]算法的对比,可以得到本文算法有更高的准确率,这是由于本文使用了基于深度学习的目标检测算法,使得对目标的初始检测更加准确,设计了目标跟踪算法,使得当目标发生快速运动时跟踪器能有更好的稳定性。

表2 客流量统计结果准确率

门店序号	客流量检测值	客流量真实值	误算人数	漏检人数	准确率 (%)
15	217	229	20	8	94.8
18	69	70	2	1	98.6
21	53	58	7	2	91.4
25	77	85	13	5	90.6
31	48	50	3	1	96.0
47	278	292	20	6	95.2
50	367	378	25	14	97.1
59	90	101	17	6	89.1
60	146	160	20	6	91.3
61	150	165	18	3	90.9

表3 两种算法客流量统计结果准确率对比

算法	客流量检测值	客流量真实值	误算人数	漏检人数	准确率 (%)
本文算法	1495	1588	145	52	94.1
文献[14]算法	1453	1588	192	57	91.5

### 3 结论与展望

本文算法实现的是对于进门客流量的统计,算法整体流程首先是对获取的监控视频,利用卷积神经网络进行特征提取,训练YOLOv3-tiny模型实现目标的识别,之后通过多特征融合的核相关滤波器设计,实现对于快速移动目标的稳定跟踪,最后通过进门行为判断方法的设计实现对于客流量的准确统计。通过实验验证,本文算法的平均准确率能达到93.5%。在目标快速移动和遮挡情况下能够实现准确的统计,但在光线昏暗的情况下算法的准确率需要进一步的提高,这是由于在光线较差的情况下目标与背景间的差异较小,使得算法对目标特征的获取产生了误差,在之后的研究将进一步改进目标跟踪方法,改变获取特征的方式,将多特征融合进行特征提取的方式转换成深度卷积特征,实现对于目标特征更准确的描述,从而进一步提升算法的稳定性。

## 参考文献

- 1 陈旭, 孟朝晖. 基于深度学习的目标视频跟踪算法综述. 计算机系统应用, 2019, 28(1): 1–9. [doi: 10.15888/j.cnki.csa.006720]
- 2 王晓宁, 宫法明, 时念云, 等. 基于卷积神经网络的复杂场景目标检测算法. 计算机系统应用, 2019, 28(6): 153–158. [doi: 10.15888/j.cnki.csa.006941]
- 3 徐超, 高梦珠, 查宇锋, 等. 基于 HOG 和 SVM 的公交乘客人流量统计算法. 仪器仪表学报, 2015, 36(2): 446–452.
- 4 Felzenszwalb PF, Girshick RB, Mcallester D, *et al.* Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 32(9): 1627–1645.
- 5 Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA. 2014. 580–587.
- 6 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems*. Washington, DC, USA. 2015. 91–99.
- 7 Girshick R. Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*. Santiago, Chile. 2015. 1440–1448.
- 8 Henriques JF, Caseiro R, Martins P, *et al.* High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 37(3): 583–596.
- 9 Mueller M, Smith N, Ghanem B. Context-aware correlation filter tracking. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA. 2017. 1396–1404.
- 10 Raheja JL, Kalita S, Dutta PJ, *et al.* A robust real time people tracking and counting incorporating shadow detection and removal. *International Journal of Computer Applications*, 2012, 46(4): 51–58.
- 11 Wateosot C, Suvonvorn N. Top-view based people counting using mixture of depth and color information. *The Second Asian Conference on Information Systems*. ACIS. 2013.
- 12 Songchenchen G, Bourennane EB. Implementation of real time reconfigurable embedded architecture for people counting in a crowd area. *International Symposium on Modelling and Implementation of Complex Systems*. Cham. 2018. 219–229.
- 13 Rohit, Chauhan V, Kumar S, *et al.* Human count estimation in high density crowd images and videos. 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC). Wagnaghat, India. 2016. 343–347.
- 14 李笑月. 基于多视觉传感器的公交车乘客人数统计关键技术的研究[硕士学位论文]. 长春: 吉林大学, 2013. 26–53.
- 15 Redmon J, Farhadi A. YOLOv3: An incremental improvement. *arXiv preprint arXiv: 1804.02767*, 2018.
- 16 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Vegas, NV, USA. 2016. 779–788.
- 17 Redmon J, Farhadi A. YOLO9000: better, faster, stronger. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA. 2017. 7263–7271.
- 18 鞠默然, 罗海波, 王仲博, 等. 改进的 YOLO V3 及其在小目标检测中的应用. *光学学报*, 2019, 39(7): 0715004–1.
- 19 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA. 2017. 2117–2125.
- 20 Schölkopf B, Smola AJ. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge: MIT Press, 2001.
- 21 Rifkin R, Yeo G, Poggio T. Regularized least-squares classification. *Nato Science Series Sub Series III Computer and Systems Sciences*, 2003, 190: 131–154.
- 22 Danelljan M, Khan FS, Felsberg M, *et al.* Adaptive color attributes for real-time visual tracking. *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA. 2014. 1090–1097.
- 23 白中浩, 王鹏辉, 李智强. 基于 Stixel-world 及特征融合的双目立体视觉行人检测. *仪器仪表学报*, 2017, 38(11): 2822–2829. [doi: 10.3969/j.issn.0254-3087.2017.11.025]