

# Video Pedestrian Detection Based on Deep Learning

Xiaomin Tong, Xiang Ji, Yin Tong

China Academy of Electronics and Information Technology, Beijing  
Email: 376095391@qq.com

Received: Oct. 6<sup>th</sup>, 2018; accepted: Oct. 17<sup>th</sup>, 2018; published: Oct. 24<sup>th</sup>, 2018

---

## Abstract

In recent years, with the extensive application of depth learning in the field of computer vision, video moving target detection based on deep learning has been favored by many scholars. The basic principle of this method is to train a classifier based on deep neural network using a large number of target sample data, and then detect the target online through the classifier. Since the depth neural network can describe the target features more deeply through multi-layer representation, the detection method based on deep learning has the advantage of being able to accurately detect targets with features from training data. For this particular application of moving target detection, the limitation of this method lies in that it does not use the target motion information and the detection results are prone to false alarm targets. In this paper, GMM method is combined with deep neural network to make full use of the target appearance and motion information in order to obtain more accurate detection results. The experimental verification was carried out on the monitoring data collected from the exhibition booth of the State-owned Enterprise Exhibition in 2017. The results show that compared with the detection method without fusion of motion information, the accuracy of pedestrian detection is improved by 3.8%.

## Keywords

Pedestrian Detection, Motion Detection, Deep Learning, GMM Modeling

---

# 基于深度学习的视频行人目标检测

仝小敏, 吉祥, 仝茵

中国电子科学研究院, 北京  
Email: 376095391@qq.com

收稿日期: 2018年10月6日; 录用日期: 2018年10月17日; 发布日期: 2018年10月24日

## 摘要

近年来,随着深度学习在计算机视觉领域的广泛应用,基于深度学习的视频运动目标检测受到广大学者的青睐。这种方法的基本原理是利用大量目标样本数据训练一个基于深度神经网络的分类器,然后通过分类器在线检测目标。由于神经网络能够通过多层表示的方式更加深刻的描述目标特征,基于深度学习的检测方法优点在于能够准确检测具有训练数据中目标特征的目标。针对视频运动目标检测这个特定的应用,这种方法的局限性在于没有利用目标运动信息,检测结果容易出现虚警目标。本文将GMM建模方法与神经网络相结合,充分利用目标外观特征和运动信息,以期获得更准确的检测结果。在2017年央企双创展实地采集的展台监控数据上进行了实验验证。结果表明,本文方法相比于不融合运动信息的检测方法,行人检测准确率提高3.8%。

## 关键词

行人检测, 运动检测, 深度学习, GMM建模

Copyright © 2018 by authors and Hans Publishers Inc.

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

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



Open Access

## 1. 引言

随着计算机计算技术和存储技术的快速发展,视频信息占人们接受信息的比重越来越大,对视频的智能分析也越来越重要。其中视频目标检测是对视频分析的重要切入点,因此不管是在学术界还是在商业界,目标检测都是研究和应用的一个热点。传统的视频运动目标检测方法包括背景差法、帧差法、光流法,这类方法[1] [2] [3]的主要原理是基于像素级分析来确定运动目标相对于背景图像的差异,检测运动目标所在的位置。尽管学者们在这类方法基础上进行了很多创新改进[4] [5] [6],但是这类自底向上的方法并未利用目标外观等宏观信息,检测结果容易受光照变化、目标交叉遮挡、目标与背景相似、阴影等因素影响。此外,针对视频行人检测这个特定的应用,仅仅利用目标运动信息并不能精确定位行人目标,尤其是目标之间有交叠、部分遮挡等情况下,无法区分不同的目标。

深度学习是目前机器学习在实际应用中最成功的一种方法,在自然图像分类、通用目标检测、语义分割等视觉领域取得了突破性的成绩。将深度学习用于视频运动目标检测的方法[7] [8] [9],能够有效描述目标外观、结构、色彩等视觉特征,从而检测定位目标。这类方法的局限性在于没有利用目标运动信息,导致与目标外观相似的虚警目标被误检。

本文致力于复杂场景下的视频行人目标检测,将深度学习方法引入传统视频运动目标检测方法,利用深度学习对行人目标准确、全面的外观描述的同时,挖掘行人目标运动信息,克服传统运动检测出现空洞、易受阴影、光照变化影响等难题,提高行人检测准确率。运动检测采用背景建模法中比较成熟的混合高斯建模法 GMM [6],有效克服光照变化,深度学习采用 YOLOv3 神经网络模型[10],实现行人与背景的二分类,解决目标与背景相似的问题,二者充分融合,准确检测行人目标。

## 2. 本文算法

### 2.1. 算法流程

本文算法流程如图 1 所示。首先利用公开数据集离线训练 YOLOv3 行人检测模型参数,利用不同

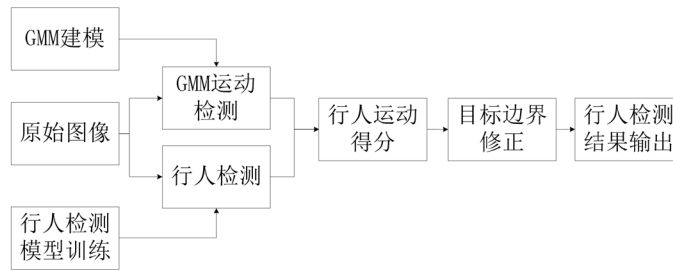


Figure 1. Algorithm flow chart of our detection  
图 1. 本文检测流程图

光照下采集的背景图像进行 GMM 背景建模。然后将原始图像分别输入 GMM 模型和行人检测模型，分别获得 GMM 运动检测结果和行人目标检测结果，并利用运动检测结果对 GMM 模型进行及时更新。根据运动检测结果计算每个检测到的行人的运动得分，最终获得行人检测结果。该方法的好处在于利用运动信息去除场景中检测到的虚警目标，提高检测精度。

## 2.2. GMM 运动检测

采用基于混合高斯模型的背景差算法对当前帧图像进行运动前景提取，得到的前景图记为  $I$ 。 $\langle w, Model \rangle$  表示一个权重为  $w$  的单高斯模型，假设对图像中坐标为  $(x, y)$  的点处混合高斯模型为  $Model = \{\langle w_i, Model_i \rangle, i = 1, 2, \dots, num\}$  ( $num$  表示混合高斯模型所包含的单高斯模型的数目)，那么前景提取公式如下：

$$I(x, y) = \begin{cases} 1 & \text{if } \exists i \in [1, num], \text{st } I_0(x, y) \in \langle w_i, Model_i \rangle \& w_i > \theta \\ 0 & \text{else} \end{cases} \quad (1)$$

式(1)中， $I_0$  为当前输入原始图像， $\theta$  为前景提取阈值，可以为固定阈值，也可取为自适应的(如取所有权重中的次小值)。

依据检测结果，按照如下公式对混合高斯模型进行及时更新。

$$w_i \leftarrow (1 - \alpha)w_i + \alpha M_i(x, y) \quad (2)$$

$$\mu_i \leftarrow (1 - \rho)\mu_i + \rho I(x, y) \quad (3)$$

$$\sigma_i^2 \leftarrow (1 - \rho)\sigma_i^2 + \rho(I(x, y) - \mu_i)^2 \quad (4)$$

其中  $\mu_i$  为第  $i$  个高斯模型  $\langle w_i, Model_i \rangle$  的均值， $\sigma_i^2$  为第  $i$  个高斯模型的方差。 $\alpha$  为混合高斯模型的学习率，且  $0 \leq \alpha \leq 1$ ，背景模型的更新速度取决于  $\alpha$  的取值。 $M_i(x, y)$  表示像素点与  $\langle w_i, Model_i \rangle$  匹配与否，匹配则为 1，否则为 0。当不匹配时，不对高斯模型的均值和方差进行更新。

$$M_i(x, y) = \begin{cases} 1 & I_0(x, y) \in \langle w_i, Model_i \rangle \\ 0 & \text{else} \end{cases} \quad (5)$$

$$\rho = \alpha \frac{1}{2\pi\sigma_i} \exp\left(-\frac{(I(x, y) - \mu_i)^2}{2\sigma_i^2}\right) \quad (6)$$

模型更新后对权重进行归一化处理，并对多个高斯模型按照权重高低重新排序。

## 2.3. YOLO 行人检测

YOLO 是一个全新的方法，把一整张图片一下子应用到一个神经网络中去。网络把图片分成不同的



$$S_i = \frac{M_i}{A_i} \quad (7)$$

$$A_i = (x_{\max} - x_{\min} + 1) \cdot (y_{\max} - y_{\min} + 1) \quad (8)$$

$$M_i = \sum_{x=x_{\min}, y=y_{\min}}^{x_{\max}, y_{\max}} I(x, y) \quad (9)$$

若  $S_i$  大于给定阈值, 则认为  $T_i$  为正确检测到的行人目标, 反之, 则认为  $T_i$  为虚警目标, 从检测结果中去除。此处所用阈值的含义为前景面积占目标区域面积的比例, 本文实验中取 0.5。若为正确检测的行人目标, 则用当前前景区域边界作为目标检测结果输出, 实现对目标检测边界的修正, 并将修正后的检测结果进行输出。

### 3. 实验结果

本文方法在 2017 年央企双创展中实地应用, 主要通过统计各个展台前参观人数检测分析各个展台热度。数据源为 187 路海康网络摄像头, 通过对每一路摄像头轮流访问获取并处理当前捕获的图像帧, 统计当前展厅前参观人数。在 Intel(R) Core(TM) i5-3470 CPU@3.20GHz 的处理器上每帧图像处理耗时越 3 秒, 这样每个展台前摄像头的两次访问间隔约为 10 分钟, 能够满足展台热度分析需求。图 3 所示为行人检测结果, (a)~(f) 分别表示不同区域检测结果。橘色框为 YOLOv3 [10] 检测结果, 可以看出, 检测结果中有许多外观与行人特征相像的虚假目标, 并且正确检测的目标区域比真实目标区域偏大。红色框为本文方法检测结果, 利用运动检测信息, 去除了检测结果中的虚警目标, 同时, 使得检测到的行人区域更加精确。

对双创展举办期间的行人检测准确率进行统计, 并与 GMM 和 yolov3 方法从检测率、漏检率、虚警率等三方面进行对比分析, 如表 1 所示。可以看出, 由于受到阴影的影响, GMM 方法[6]检测结果中存在较多虚警目标, yolov3 方法[10]未利用目标运动信息, 容易检测到外观与行人特征相像的虚假目标。相比与 GMM 和 yolov3, 本文方法融合外观特征和运动特征, 能够有效降低虚警率, 从而提高检测准确率。本文方法的检测准确率相比于 YOLO, 提高了 3.8%。

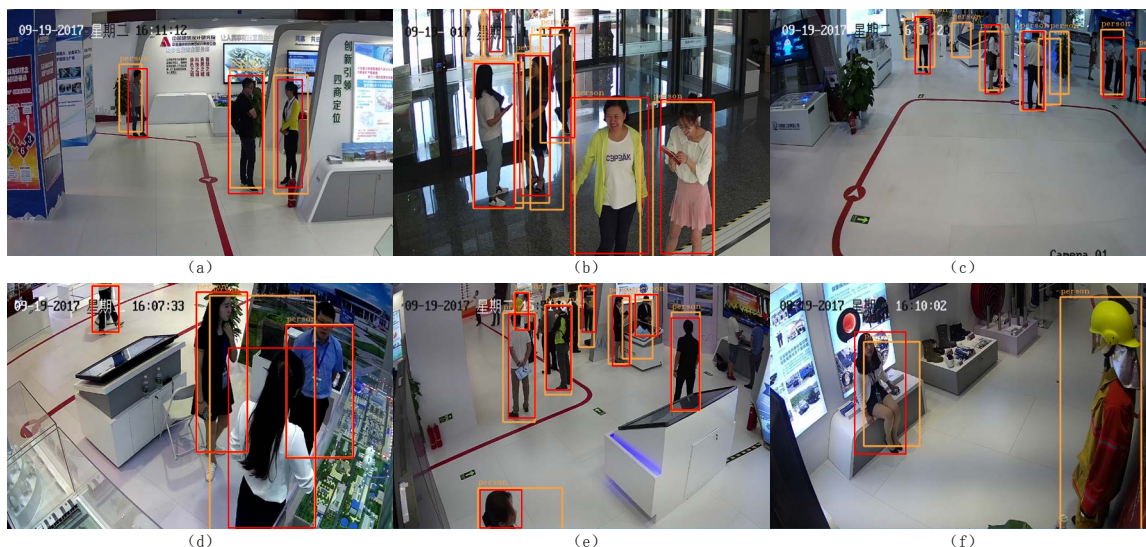


Figure 3. Pedestrian detection result  
图 3. 行人检测结果

**Table 1.** Detection result statistics and comparison by different algorithms**表 1.** 不同方法的检测结果统计对比

	GMM	YOLOv3	本文方法
检测率	91.4%	91.9%	95.7%
漏检率	8.6%	8.1%	4.3%
虚警率	5.1%	8.2%	2.5%

检测率  $P$ 、漏检率  $M$ 、虚警率  $W$  的计算公式如下。

$$P = \frac{\text{DetectionObjNum}}{\text{ObjNum}} \quad (10)$$

$$M = \frac{\text{MissObjNum}}{\text{ObjNum}} \quad (11)$$

$$W = \frac{\text{DetectionNum} - \text{DetectionObjNum}}{\text{DetectionNum}} \quad (12)$$

其中  $\text{DetectionNum}$  表示检测到目标的个数， $\text{DetectionObjNum}$  表示检测到的目标中真实目标的个数， $\text{ObjNum}$  表示场景中真实目标的个数， $\text{MissObjNum}$  表示未被检测到目标的个数。

#### 4. 结论

本文提出了一种融合运动检测信息和深度学习的视频行人检测方法，利用运动检测信息去除深度学习检测目标中的虚警目标。相比于只利用深度学习模型检测行人的方法，本文方法有效的去除了虚警目标，提高了行人目标的检测准确率。本文方法的局限性在于对于深度学习检测结果中的漏检没有做相应的处理，如何利用运动检测结果降低漏检率是下一步研究方向。

#### 致 谢

对本文的创作给予指导和帮助的同事朋友，在此一并感谢。

#### 基金项目

本文受到中电科集团创新项目“面向国家治理体系和治理能力的 XX 云平台”资助。

#### 参考文献

- [1] 陈海永, 鄯丽忠, 杨德东, 等. 基于超像素信息反馈的视觉背景提取算法[J]. 光学学报, 2017, 37(7): 0715001.
- [2] 宋涛, 李鸥, 崔弘亮. 基于场景感知的运动目标检测方法[J]. 电子学报, 2016, 44(11): 2625-2632.
- [3] 张应辉, 刘养硕. 基于帧差法和背景差法的运动目标检测[J]. 计算机技术与发展, 2017, 27(2): 25-28.
- [4] Prashant, D., Pratik, S. and Suman, K.M. (2015) Fast and Accurate Foreground Background Separation for Video Surveillance. *Lecture Notes in Computer Science*, **9124**, 96-103.
- [5] Zhu, W.J., Wang, G.L., Tian, J., Qiao, Z.T. and Gao, F.Q. (2018) Detection of Moving Objects in Complex Scenes Based on Multiple Features. *Acta Optica Sinica*, **38**, 0612004. <https://doi.org/10.3788/AOS201838.0612004>
- [6] 唐洪良, 黄颖, 黄淮, 杨成顺, 黄宵宁. 改进的自适应高斯混合模型运动目标检测算法[J]. 现代电子技术, 2017(11): 65-67.
- [7] Dai, J., Li, Y., He, K. and Sun, J. (2016) R-FCN: Object Detection via Region-Based Fully Convolutional Networks. CVPR.

- [8] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. CVPR.
- [9] Redmon, J. and Farhadi, A. (2016) YOLO9000: Better, Faster, Stronger. CoRR, abs-1612-08242.
- [10] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement. CoRR, abs-1804-02767.

**知网检索的两种方式:**

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>  
期刊邮箱: [csa@hanspub.org](mailto:csa@hanspub.org)