

## 基于 YOLO 网络的行人检测方法

高 宗<sup>a</sup>, 李少波<sup>a,b</sup>, 陈济楠<sup>a</sup>, 李政杰<sup>a</sup>

(贵州大学 a. 现代制造技术教育部重点实验室; b. 机械工程学院, 贵阳 550025)

**摘 要:** 针对基于人工提取特征的行人检测器鲁棒性差的问题, 借鉴目标检测的研究成果, 提出一种行人目标实时检测方法。以 YOLO 网络结构为基础, 结合行人在图像中呈现宽高比小的特点, 聚类选取合适的候选框数量和规格, 改进 YOLO 网络结构, 调整候选框在 X、Y 轴方向的分布密度, 形成适用于行人检测的网络结构。实验结果表明, 与 HOG、LatSVM-v2 等行人检测方法相比, 该方法降低了漏检率和误检率, 提高了定位准确性, 检测速度满足实时性要求。

**关键词:** YOLO 网络; 行人检测; 深度网络; 聚类; 特征重组

**中文引用格式:** 高 宗, 李少波, 陈济楠, 等. 基于 YOLO 网络的行人检测方法[J]. 计算机工程, 2018, 44(5): 215-219, 226.

**英文引用格式:** GAO Zong, LI Shaobo, CHEN Jinan, et al. Pedestrian Detection Method Based on YOLO Network[J]. Computer Engineering, 2018, 44(5): 215-219, 226.

## Pedestrian Detection Method Based on YOLO Network

GAO Zong<sup>a</sup>, LI Shaobo<sup>a,b</sup>, CHEN Jinan<sup>a</sup>, LI Zhengjie<sup>a</sup>

(a. Key Laboratory of the Ministry of Education of Modern Manufacturing Technology;

b. College of Mechanical Engineering, Guizhou University, Guiyang 550025, China)

**【Abstract】** For improving the poor robustness of pedestrian detectors based on artificial extraction characteristics, learns from the advanced research achievements in the object detection field, proposes a method of real-time pedestrian detection. Being based on the YOLO network structure, combines with characteristics that pedestrians' aspect ratio is small in images, selects appropriate numbers and scales of the initial candidate boxes by clustering, improves the YOLO network structure, adjusts density of candidate boxes distributed on the X and Y axis, forms a network structure which is suitable for pedestrian detection. Experiments in INRIA pedestrian datasets show that comparing with some pedestrian detection methods, such as HOG, LatSVM-v2 and so on, this method reduces the missed and false detection, improves the positioning accuracy and satisfies the real-time requirements.

**【Key words】** YOLO network; pedestrian detection; deep network; clustering; features reorganization

**DOI:** 10.19678/j.issn.1000-3428.0046885

### 0 概述

行人检测是目标检测领域的重要分支。近年来在视频监控、汽车辅助驾驶和智能机器人等方面得到了广泛关注。行人检测技术经过几十年的研究发展, 在检测精度和速度上都取得了很大的进展。主流的行人检测模型<sup>[1]</sup>主要有可变形部件模型(Deformable Part Models, DPM)、深度网络(Deep Network, DN)、决策树(Decision Forest, DF)。基于这 3 种结构的检测方法各有特点, 均可在行人检测上得到相似的最优结果。目前对于这 3 种模型孰优孰劣尚无定论。

传统的行人检测方法基于人工设计的特征提

取器, 通过提取 Haar 特征、方向梯度直方图(Histogram of Oriented Gradient, HOG)、局部二值模式(Local Binary Pattern, LBP)等训练分类器, 以达到行人检测的目的, 在行人检测任务中取得了令人瞩目成果。例如文献[2]利用 HOG + LBP 特征处理行人遮挡, 提高检测准确率。文献[3]和文献[4]分别提出积分通道特征(Integral Channel Features, ICF)和聚合通道特征(Aggregated Channel Features, ACF), 融合梯度直方图、LUV 色彩特征和梯度幅值特征, 获得了性能更好的行人特征表达。但人工设计的行人特征很难适应行人的大幅度变化。深度网络模型可以从图像像素中学习特征, 提高行人检测器性能。文献[5]提出的卷积神经网络

**基金项目:** 国家自然科学基金(51475097)。

**作者简介:** 高 宗(1992—), 男, 硕士研究生, 主研方向为图像处理、机器学习、机器视觉; 李少波, 教授; 陈济楠、李政杰, 硕士研究生。

**收稿日期:** 2017-04-19      **修回日期:** 2017-06-09      **E-mail:** zong209@163.com

络检测方法,运用基于卷积稀疏编码的无监督方法预训练卷积网络,用训练好的模型进行行人检测。文献[6]通过探究不同的网络深度、卷积核大小以及特征维度对行人检测结果的影响,构建了基于卷积神经网络的行人分类器,但该检测器不具备目标定位的功能。文献[7-8]提出运用深度模型学习不同身体部位特征来解决行人遮挡问题,并根据不同部位间的相互约束完成行人检测。深度网络模型也在行人检测领域得到了深入运用,随着大规模训练数据集的构建以及硬件计算能力的不断增强,深度网络结构在不同的视觉任务中取得了巨大的成功。在目标检测方面,从 RCNN<sup>[9]</sup>、SPP-Net<sup>[10]</sup>、Fast-RCNN<sup>[11]</sup>、Faster-RCNN<sup>[12]</sup> 到 YOLO<sup>[13]</sup>,目标检测的准确率和速度都达到了新的高度。其中 YOLO 网络是目前最优秀的目标检测架构之一,在检测实时性方面表现尤为突出。

本文借鉴目标检测领域先进的研究成果,提出将 YOLO 网络结构用于行人检测,聚类选取初始候选框,重组特征图,扩展横向候选框数量,构建基于 YOLO 网络的行人检测器 YOLO-P。

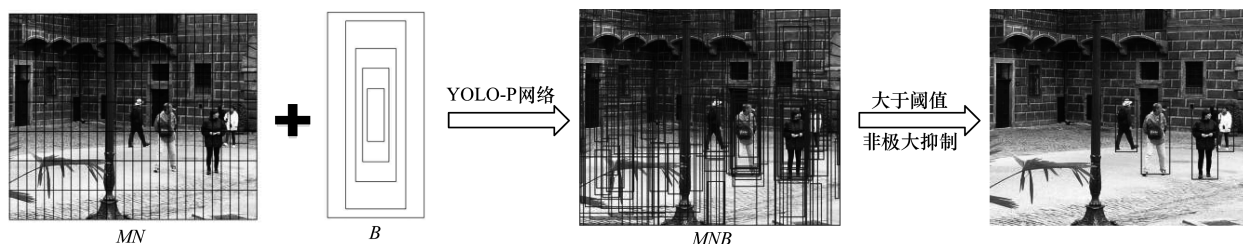


图1 YOLO-P 行人检测过程

## 1.2 待测目标检测

首先对候选框进行目标检测,预测每个候选框的中存在待判别目标的置信度  $Conf(Object)$ ,将不存在目标物的候选框置信度置为 0。

$$Conf(Object) = Pr(Object) \times IOU_{Pred}^{Truth}$$

$Pr(Object)$ :是否有目标物落入候选框对应的单元格中。若有,单元格对应的候选框的目标置信度为  $Conf(Object) = IOU_{Pred}^{Truth}$ ;否则,认定候选框中没有目标物,  $Conf(Object) = 0$ 。

$$Pr(Object) = \begin{cases} 0, & \text{单元格中无目标物} \\ 1, & \text{单元格中有目标物} \end{cases}$$

$IOU_{Pred}^{Truth}$ :预测框与实际框的交集面积与并集面积之比,即:

$$IOU_{Pred}^{Truth} = \frac{area(box(Truth) \cap box(Pred))}{area(box(Truth) \cup box(Pred))}$$

## 1.3 行人检测与定位

对存在目标物的候选框进行行人判别,设预测目标物是行人的条件概率为  $Pr(Person|Object)$ ,则候选框中包含行人的置信度  $Conf$ :

$$Conf = Pr(Person|Object) \times Pr(Object) \times IOU_{Pred}^{Truth}$$

对每个候选框预测其中包含行人的概率以及边

## 1 检测方法

基于 YOLO 网络的检测方法将候选框提取、特征提取、目标分类、目标定位统一于一个神经网络中。神经网络直接从图像中提取候选区域,通过整幅图像特征来预测行人位置和概率。将行人检测问题转化为回归问题,真正实现端到端(end to end)的检测。

行人检测就是对输入的图像或视频,进行候选框提取,判断其中是否包含行人,若有给出其位置。而事实上,大部分的候选框中并不包含行人,如若对每个候选框都直接预测行人概率,无疑增加了网络学习的难度。在 YOLO-P 的检测方法中,将检测分为 3 个过程,即候选框的提取、待测目标检测、行人检测与定位。在待测目标检测的过程中,将部分预测框的置信度置为 0,以降低网络学习的难度。

### 1.1 候选框的提取

将输入的图像划分为  $M \times N$  个单元格,每个单元格给定  $B$  个不同规格的初始候选框,如图 1 所示,预测候选框经由卷积层网络提取出来,每幅图像候选框数量为  $M \times N \times B$ 。

界框的位置,则每个候选框输出的预测值为:

$$[X, Y, W, H, Conf(Object), Conf]$$

其中,  $X, Y$  为预测框中心相对于单元格边界的偏移,  $W, H$  为预测框宽高相对于整幅图像之比。对于输入的每幅图片,最终网络输出为向量:

$$M \times N \times B \times [X, Y, W, H, Conf(Object), Conf]$$

## 2 网络架构

本文网络以 YOLO 网络为原型,其采用区域建议网络(Regions Proposal Network, RPN)的思想,去掉全连接层,采用卷积层来预测目标框的偏移和置信度。对特征图中的每个位置预测这些偏移和置信度,以得到行人目标的概率和位置。

### 2.1 多级特征图融合

行人特征在深度学习网络中展现出结构性,高层的特征抽象程度高,表达出行人的整体特征,中层的特征相对具体,表达出行人的局部特征。结合不同细粒度特征,可以增加行人检测的鲁棒性。为提高对小目标的检测,融合多级特征图,让不同细粒度的特征参与行人检测,构成 YOLOv2<sup>[14]</sup> 网络,结构如图 2 所示。

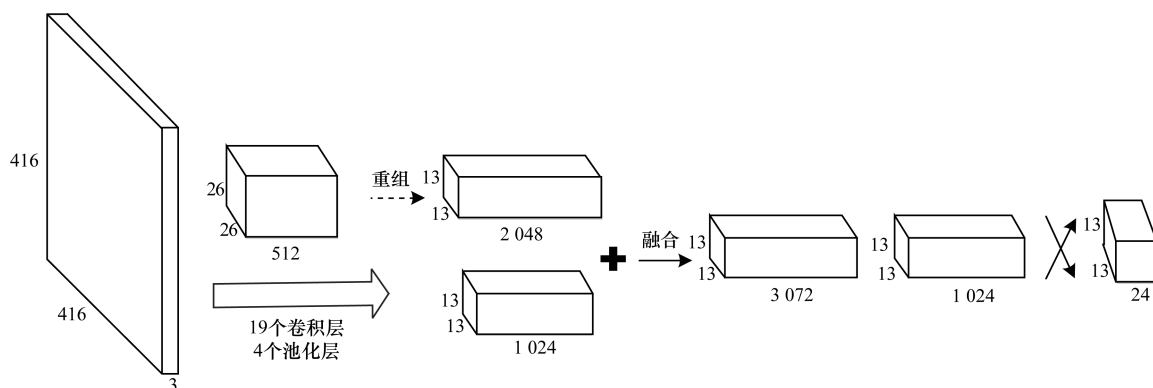


图2 YOLOv2 网络结构

## 2.2 横向候选框扩展

在YOLO检测方法中,图像被分成 $S \times S$ 的网格。候选框在 $X$ 和 $Y$ 轴上同等密度分布,对人群进行检测时,漏检率较高。实际上,行人在图像中呈现 $X$ 轴上分布更密集, $Y$ 轴上分布相对稀疏的特点,如图3所示。针对这一问题,本文提出在网络中增加一个reorg层,重组特征图,增加候选框在 $X$ 轴方向的密度,同时减少 $Y$ 轴方向候选框密度,构成YOLO-P网络(如图4所示)。



图3 行人在图像中分布情况

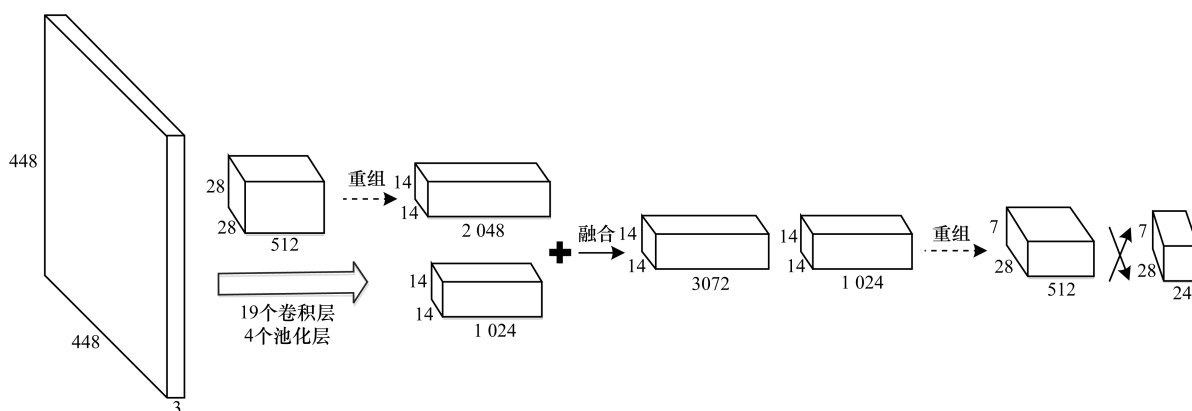


图4 YOLO-P 网络结构

用于行人检测的YOLO-P网络模型,输入图像分辨率为448像素 $\times$ 448像素,若为每个网格选定4个初始候选框,则YOLO-P网络输出是大小为 $28 \times 7 \times 24$ 的一组向量。

## 3 网络训练

以开源的神经网络框架Darknet为基础,以YOLO-P网络结构为模型,训练行人检测器。为加快训练速度和防止过拟合,选用的冲量常数0.9,权值衰减系数为0.0005。学习率采用多分步策略。

### 3.1 预训练

为减少训练时间,以Darknet19网络模型训练得到的网络参数初始化卷积层网络。该预训练模型由

图2中的前23层网络后接一个平均池化层和一个全连接层构成,在ImageNet1000数据集训练10个循环(epoch)后得到预训练参数。

### 3.2 训练集

基于卷积神经网络的行人检测方法需要从大量样本中学习行人特征,若样本集不具有代表性,很难选择出好的特征。本文以INRIA行人检测数据集作为训练和测试数据集。

INRIA行人检测数据集是目前使用最广泛的静态行人检测数据库之一,分为训练数据集和测试数据集。训练数据集包含正样本图像614张,行人数目为1237个;测试数据集包含正样本图像288张,行人数目为589个。图像光照变化明显,目标尺度

变化大、行人姿态丰富、外观服饰变化多、背景复杂且有不同程度的遮挡。

### 3.3 初始候选框

在训练网络时,需要预设候选框的初始规格及数量。随着迭代次数不断增加,网络学习到行人特征,预测框参数不断调整,最终接近真实框。为加快收敛速度,提高行人检测的位置精度,分析图像中行人的宽高特点,用 K-means 方法进行聚类,得到与图像中行人边界最相近的初始候选框参数。

一般 K-means 聚类采用欧式距离衡量两点之间的距离。本文对候选框宽高与单位网格长度之比进行聚类。预测框和真实框的交并比 (Intersection-Over-Union, IOU) 是反映预测框与真实框差异的重要指标,IOU 值越大,表明两者差异越小,“距离”越近。聚类的目标函数为:

$$\min \sum_i \sum_j (1 - I_{\text{IOU}}(\text{Box}[i], \text{Truth}[j]))$$

其中,  $i$  为聚类的类别数,  $j$  为样本集数量,  $\text{Box}[i]$

表示聚类得到的预测框  $i$  的规格,规格在数值上表示为一组数值:(预测框宽/图像宽,预测框高/图像框高),  $\text{Truth}[j]$  表示样本  $j$  中行人定位框规格。

## 4 实验结果及分析

本文实验环境为: Intel Xeon CPU E5-1620 V3 3.5 GHz, 16 GB 内存, Nvidia Geforce GTX1080, Ubuntu16.04, 64 位操作系统。在该实验环境下, YOLO-P 检测器的检测速度达到 25 frame/s, 满足实时性要求。

### 4.1 聚类选取初始候选框的有效性验证

为验证聚类选取初始候选框的规格和数量对行人检测器的影响,选择 INRIA 的训练和测试数据作为实验数据,以 YOLOv2 网络训练行人检测器。

1) 对 INRIA 的行人数据的真实框以 3.3 节中所述 K-means 方法聚类,聚类类别数量  $B$  分别设定为 3、4、5、6,聚类值为行人真实框宽高分别与图像宽高之比。将聚类的结果(如表 1 所示)作为网络训练候选框的初始规格。

表 1 INRIA 数据集上行人真实框聚类结果

聚类数量	规格 1	规格 2	规格 3	规格 4	规格 5	规格 6
$B = 3$	(0.069, 0.277)	(0.125, 0.479)	(0.236, 0.732)	—	—	—
$B = 4$	(0.058, 0.217)	(0.096, 0.389)	(0.160, 0.603)	(0.267, 0.756)	—	—
$B = 5$	(0.054, 0.195)	(0.084, 0.347)	(0.129, 0.493)	(0.191, 0.697)	(0.308, 0.762)	—
$B = 6$	(0.054, 0.195)	(0.084, 0.347)	(0.129, 0.493)	(0.183, 0.743)	(0.224, 0.566)	(0.312, 0.784)

2) 每组实验均选用 YOLOv2 网络作为训练行人检测器的网络,以排除其他因素的干扰。

3) 每组实验均选用 INRIA 数据集进行训练和检测。检测结果如图 5 所示,以平均每张图片误检数 (False Positive Per Image, FPPI) 作为横坐标,漏检率 (Miss-rate) 作为纵坐标。

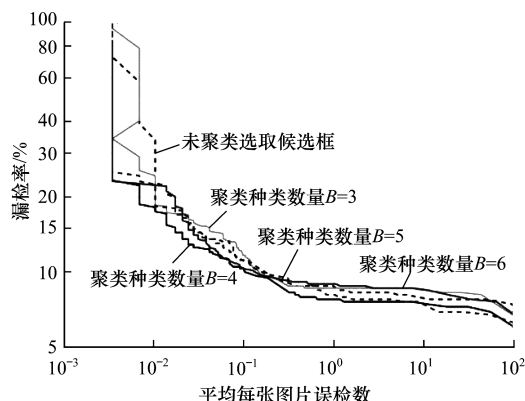


图 5 不同聚类数量对检测器的影响

从图 5 可以看出,通过聚类选取初始候选框后,检测器的漏检得到了明显改善。其中,将  $B = 4$  的聚类结果作为初始候选框,训练得到的检测器得到了最好的检测结果。但需要注意该聚类结果仅反映的是 INRIA 数据集上行人的宽高特点,对其他数据集可能并不适用。

### 4.2 横轴候选框扩展的有效性验证

为验证本文提出的增加横轴方向候选框数量可以提高检测器性能的有效性。以 INRIA 数据集作为实验数据,比较 YOLOv2 网络和 YOLO-P 网络所训练的检测器的效果。

从图 6 可以看出,在 FPPI 相同时, YOLO-P 检测器的漏检率明显低于 YOLOv2 检测器。当 FPPI 为 0.1 时, YOLO-P 检测器的漏检率为 8.7%, 较 YOLOv2 检测器降低了 2.5%。

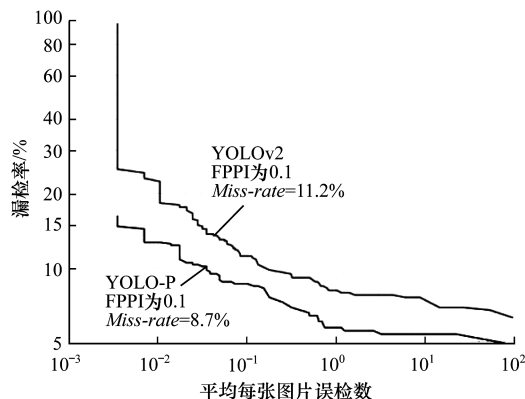


图 6 YOLOv2 与 YOLO-P 检测结果对比

### 4.3 结果对比

与当前具有代表性的行人检测方法<sup>[15-18]</sup>进行比较,验证 YOLO-P 检测方法的有效性。所有检测结

果均在 INRIA 数据集上训练测试得到。检测器的优劣可以通过 LAMR<sup>[19]</sup> (Log-Average Miss Rate) 指标评判。LAMR 反映的是 FPPI 在  $[10^{-2}, 10^2]$  上的总体漏检情况。

$$L_{\text{LAMR}} = e^{\frac{1}{9} \left( \sum_{i=1}^9 \lg(\text{Miss-Rate}(10^{-2.25+0.25i})) \right)}$$

从图7可以看出, YOLO-P 检测方法在 LAMR 指标评价中取得了最好的结果。在 FPPI 小于 0.1 时, 漏检率较其他检测方法显著降低。原因在于大多检测方法采用的是人工设计的特征提取器(如 HOG、LBP 特征提取), 行人特征表达不够充分, 在进行行人检测时, “虚警”更多; 而 YOLO-P 检测方法由卷积网络自主学习行人特征, 对行人特征表达更优秀。

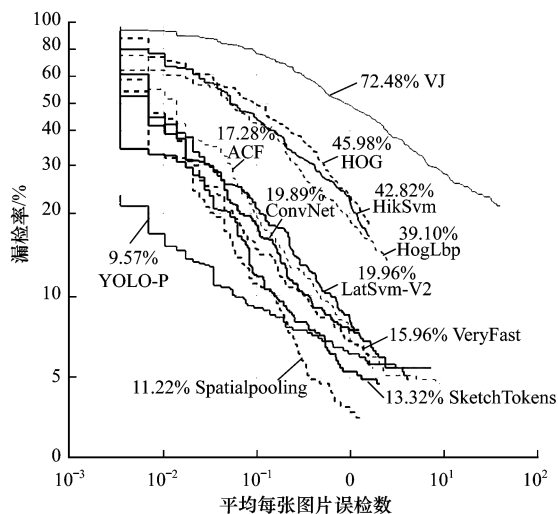


图7 各检测方法在 INRIA 数据集上测试结果对比

真实框与预测框的交并比 (IOU) 是反映行人检测定位准确性的重要指标, 值越大, 表明定位准确性越高。比较各检测方法预测结果的平均交并比 (Average IOU), 如图8所示, YOLO-P 检测器表现出了最好的定位准确性。

$$\text{Average\_IOU} = \text{mean}(\text{IOU}_{\text{Truth}}^{\text{Tp}}(i))$$

$\text{IOU}_{\text{Truth}}^{\text{Tp}}(i)$  表示预测正确的预测框  $i$  与真实框  $i$  的交并比。

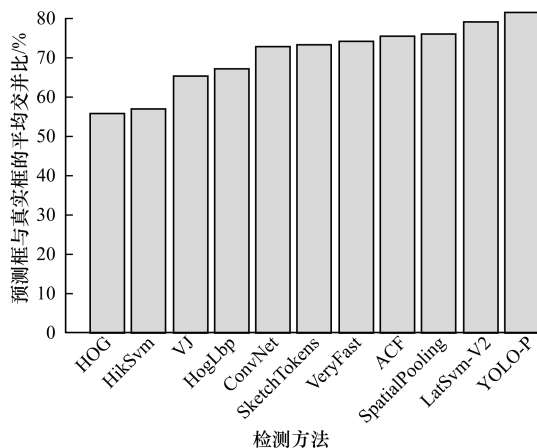


图8 各检测方法的平均交并比对比

## 5 结束语

本文提出将 YOLO 网络架构用于行人检测, 分析行人在图像中的分布特点, 改进网络后得到 YOLO-P 结构, 以达到行人目标实时检测的目的。实验结果表明, 该检测方法具有较好的检测准确率和定位准确性。但是, 该检测方法在 Caltech<sup>[20]</sup> 行人数据集上的检测效果并不理想。主要原因是: 检测方法仅将静态图片的信息作为检测依据, 信息来源较为单一; 将行人检测看作二值分类问题, 对行人动态变化的检测具有较大局限性。近年来, 很多学者努力提取更多信息辅助检测, 如光流信息、环境信息等, 提高特征表达能力, 这是行人检测发展的重要方向, 也是下一步的研究重点。

## 参考文献

- [1] BENENSON R, OMRAN M, HOSANG J, et al. Ten years of pedestrian detection, what have we learned? [J]. Computer Vision, 2014, 8926: 613-627.
- [2] WANG Xiaoyu, HAN T X, YAN Shuicheng. An HOG-LBP human detector with partial occlusion handling [C]// Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2009: 32-39.
- [3] DOLLÁR P, TU Z, PERONA P, et al. Integral channel features [C]// Proceedings of British Machine Vision Conference, BMVC '09. London, UK: [s. n.], 2009: 1-11.
- [4] DOLLAR P, APPEL R, BELONGIE S, et al. Fast feature pyramids for object detection [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 36 (8): 1532-1545.
- [5] SERMANET P, KAVUKCUOGLU K, CHINTALA S, et al. Pedestrian detection with unsupervised multi-stage feature learning [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Computer Society, 2013: 3626-3633.
- [6] 芮挺, 费建超, 周游, 等. 基于深度卷积神经网络的行人检测 [J]. 计算机工程与应用, 2016, 52 (13): 162-166.
- [7] WANG Xiaogang, OUYANG Wanli. A discriminative deep model for pedestrian detection with occlusion handling [C]// Proceedings of IEEE International Conference Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2012: 3258-3265.
- [8] WANG Xiaogang, OUYANG Wanli. Joint deep learning for pedestrian detection [C]// Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2013: 2056-2063.
- [9] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [EB/OL]. [2013-11-11]. <https://arxiv.org/abs/1311.2524>.

(下转第226页)

## 参考文献

- [1] PANG S, SUN M, HU X, et al. SGM-based seamline determination for urban orthophoto mosaicking [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 112: 1-12.
- [2] 江 铁, 朱桂斌, 孙 奥. 全景图像拼接技术研究现状综述[J]. *重庆工商大学学报(自然科学版)*, 2012, 29(12): 60-71.
- [3] 盖进伟. 一种基于特征点的图像拼接技术研究[D]. 太原: 中北大学, 2012.
- [4] LI X, HUI N, SHEN H, et al. A robust mosaicking procedure for high spatial resolution remote sensing images [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, 109: 108-125.
- [5] 刘志文, 刘定生, 刘 鹏. 应用尺度不变特征变换的多源遥感图像特征点匹配[J]. *光学精密工程*, 2013, 21(8): 2146-2153.
- [6] LOWE D G. Distinctive image features from scale-invariant key points [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [7] GAO J, KIM S J, BROWN M S. Constructing image panoramas using dual-homography warping [C]// *Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2011: 49-56.
- [8] LIU F, GLEICHER M, JIN H, et al. Content-preserving warps for 3D video stabilization[J]. *ACM Transactions on Graphics*, 2009, 28(3): 44.
- [9] 张永梅, 张晨希, 郭 莎. 基于 SIFT 特征的彩色图像拼接方法研究[J]. *计算机测量与控制*, 2016, 24(8): 236-239.
- [10] 刘文进, 张 蕾, 孙劲光. 基于 AP 聚类图像分块的角点检测改进方法[J]. *计算机工程*, 2015, 41(4): 217-221.
- [11] ZARAGOZA J, CHIN T J, BROWN M S, et al. As-projective-as-possible image stitching with moving DLT[C]// *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2013: 2339-2346.
- [12] 刘德连, 张建奇. 基于 3 点匹配的图像拼接算法[J]. *计算机工程*, 2006, 32(13): 203-205.
- [13] ZHANG Z. A flexible new technique for camera calibration[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(11): 1330-1334.
- [14] van de SSNDE K E A, UIJLINGS J R R, GEVERS T, et al. Segmentation as selective search for object recognition [C]// *Proceedings of 2011 IEEE International Conference on Computer Vision*. Washington D. C., USA: IEEE Press, 2011: 1879-1886.
- [15] 李玉峰, 李广泽, 谷绍湖, 等. 基于区域分块与尺度不变特征变换的图像拼接算法[J]. *光学精密工程*, 2016, 24(5): 1197-1205.
- [16] 杜京义, 胡益民, 刘宇程. 基于区域分块的 SIFT 图像匹配技术研究及实现[J]. *光电工程*, 2013, 40(8): 52-58.
- [17] 韩 冰, 王永明. 基于一种快速归一化积相关算法的图像匹配研究[J]. *兵工学报*, 2010, 31(2): 160-165.
- [10] HE Kaiming, ZHANG Xianyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2014, 37(9): 1904-1916.
- [11] GIRSHICK R. Fast R-CNN[C]// *Proceedings of IEEE International Conference on Computer Vision*. Washington D. C., USA: IEEE Press, 2015: 1440-1448.
- [12] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2015: 1-1.
- [13] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]// *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Computer Society, 2016: 779-788.
- [14] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]// *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Computer Society, 2017: 6517-6525.
- [15] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// *Proceedings of Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2005: 886-893.
- [16] FELZENSZWALB P, MCALLESTER D, RAMANAN D. A discriminatively trained, multiscale, deformable part model [C]// *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2008: 1-8.
- [17] CHEN Guang, DING Yuanyuan, XIAO Jing, et al. Detection evolution with multi-order contextual co-occurrence [C]// *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Computer Society, 2013: 1798-1805.
- [18] ZHANG Shanshan, BENENSON R, OMRAN M, et al. How far are we from solving pedestrian detection [C]// *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Computer Society, 2016: 1259-1267.
- [19] WOJEK C, DOLLAR P, SCHIELE B, et al. Pedestrian detection; an evaluation of the state of the art [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2012, 34(4): 743-761.
- [20] DOLLAR P, WOJEK C, SCHIELE B, et al. Pedestrian detection; a benchmark [C]// *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2009: 304-311.

编辑 索书志

编辑 刘 冰

(上接第 219 页)