

基于自编码神经网络与 AdaBoost 的快速行人检测算法

韩宪忠 李得锋 王克俭 周利亚

(河北农业大学 信息科学与技术学院 保定 071000)

摘要 针对传统基于 HOG 特征与 AdaBoost 算法分类器在目标检测中存在检测速度慢、误差率大的问题,提出了一种基于自编码神经网络与 AdaBoost 的快速行人检测算法。该算法首先利用基于 ACF 模型的目标检测算法对图像进行预处理,获得疑似目标区域;然后对获取的子区域进行尺度归一化,提取 HOG 特征,并输入到自编码神经网络中进行降维;最后利用 AdaBoost 分类器对分类检测,输出检测到的行人区域。实验结果表明:文中所提算法的行人检测性能超过现有的检测算法,其检测速度也超过大多数算法。

关键词 行人检测;HOG 特征;AdaBoost 算法;自编码网络;ACF 模型

中图分类号 TP391.9 **文献标识码** A **文章编号** 1672-4321(2018)01-0108-06

Fast Pedestrian Detection Algorithm Based on Auto-encoder Neural Network and AdaBoost

Han Xianzhong, Li Defeng, Wang Kejian, Zhou Liya

(College of Information Science and Technology, Hebei Agricultural University, Baoding, Hebei 071000, China)

Abstract Since the traditional algorithm has the shortages of slow detection rate and large error rate in pedestrian detection, a fast pedestrian detection algorithm based on auto-encoder neural network and AdaBoost is proposed. Firstly, the pedestrian detection algorithm based on ACF model is used to process the image to obtain the suspected object area. Then the acquired sub-region is normalized and the HOG feature is extracted and input into the auto-encoder neural network. Finally, the AdaBoost classifier is used to detect the classification and output the detected pedestrian area. The experimental results show that the proposed method has more performance than the existing detection algorithm for pedestrian detection, and its detection speed is also faster than most of the algorithms.

Keywords pedestrian detection; HOG feature; AdaBoost algorithm; auto-encoder network; ACF model

行人检测是利用计算机视觉技术判断图像或者视频序列中是否存在行人并给出精确定位,广泛应用于视频监控、无人驾驶、机器人导航等领域^[1]。尽管近年来对行人检测应用进行了大量的研究,取得了重大的进展。然而,由于行人外观易受穿着、尺度、遮挡、姿态和视角等影响,使得行人检测成为计算机视觉领域中一个既具有研究价值同时又极具挑战性的热门课题。

现有的行人检测算法大致上可分为两类:背景

建模和统计学习^[2]。前者旨在利用背景建模相关方法,提取出前景运动的目标,在目标区域内进行特征提取,然后利用分类器进行分类,判断是否包含行人。常见的模型有高斯混合 GMM 模型、Codebook 模型、低秩稀疏模型等。文献[3]提出了融合光流速度场的自适应背景建模的行人检测算法,该算法能够精准提取交通场景语义信息,但对光照影响的行人检测效果不太理想。文献[4]则提出将传统的行人分类器融入到 Codebook 背景建模算法中,通过背景建

收稿日期 2017-10-29

作者简介 韩宪忠(1964-)男,教授,研究方向:计算机网络与数据库、信息资源管理,Email:13832252366@163.com

基金项目 河北省科技项目基金资助项目(14227404D);河北农业大学理工基金项目(LG201407;ZD201407;LG20140703);河北省高等学校科学技术研究项目(ZD2015054)

模算法为行人检测提供候选区域,减少搜索范围,降低了行人误检率,解决了背景建模算法不能应对光照突变的问题。统计学习是目前行人检测最常用的方法,通过对图像中候选区域的特征进行提取,然后输入由大量样本训练的分类器中进行分类。常见的分类器主要包括神经网络、SVM、AdaBoost 以及深度学习。基于统计学习的行人检测中最关键部分是如何提取有效的行人特征,其检测的整体性能取决于其特征的表征能力。现有的优秀算法表明融合多种能够描述人不同特点的特征,能够获得比仅利用单一特征类型时更好的检测性能,如 HOG+LBP、Haar+Edgelet 等。Zhang 等人提出了聚合信道特征(ACF)与 HOG 特征来检测行人^[5]; Guo 等人^[6]引入 AdaBoost 与 SVM 提出的两阶段分类模型可以进一步降低检测的错误率,该系统比单级算法具有更好的性能,但这类算法的时间复杂度太大。

早期行人检测算法大都通过对行人全局外观进行建模,并采用 SVM、Boost 等方法进行检测分类。虽然此类行人检测算法能够取得较好效果,但应用于大规模的行人检测系统中时,通常受时间和空间复杂度约束,其本质原因是输入特征空间具有高维、非线性的特征,分类器不能对目标的特征进行完全拟合,同时也面临“维数灾害”的问题。近年来,由于深度模型包含了许多隐藏层,而隐藏层又具有许多隐藏结点,这样便使得模型的表达能力十分强大,因此国内外研究人员引入深度学习进行行人检测,以便提高检测性能。这类模型除了 Yann 等人^[7]提出的 ConvNet 之外,其他模型大都依赖人工特征。Hosang 等人^[8]利用基于 CNN 框架的多通道多模式特征对行人进行检测,取得了不错的效果; Girshick 等人^[9]使用基于卷积稀疏编码的无监督方法检测行人,该方法通过预先训练好的深层模型对疑似区域进行稀疏编码,克服区域特征的空间限制。虽然目前基于深度学习的检测方法取得了非常好的效果,但由于该模型训练耗时,同时对硬件要求较高,直接影响工业应用。

本文通过对传统模型与神经网络的分析,提出了一种基于自编码神经网络与 AdaBoost 的快速行人检测算法。为了降低特征训练分类的复杂度,该算法首先利用基于 ACF 模型的目标检测算法对图像进行预处理,获得疑似目标区域;然后对获取的子区域进行尺度归一化,提取 HOG 特征,并输入到自编码神经网络中进行降维;最后利用 AdaBoost 分类器对分类检测,输出检测到的行人区域。大量的定性定

量实验表明本文所提算法的行人检测性能超过现有的检测算法,其检测速度也超过大多数算法,适合工程应用。

1 相关技术

1.1 自编码网络

自编码网络是一种用于学习高效编码的人工神经网络,通过学习获得数据集的压缩编码,可以达到数据降维的目的^[10]。自编码网络能使具体的特征向量逐渐转化为抽象的特征向量。自编码网络能很好地满足高维数据空间和低维数据空间双向映射的非线性学习,它采用自适应、多层编码网络将高维原始数据转换成低维抽象数据,并利用类似的解码网络从低维抽象数据重构原始数据的高维特征。

自编码网络是一种非监督学习方法,图 1(a)展示了自编码网络的原理,其流程主要包括编码器和解码器两个过程。编码器用于降维,解码器用于重构,其视为编码器的逆过程。编码器和解码器之间还存在一个交叉部分,称为“隐藏层”。原始高维数据输入到编码器中,被压缩后表述为隐藏层,是输入数据的一个低维表示。图 1(b)最左侧节点是输入层,最右侧的一列神经元是输出层。输出层的神经元数量完全等于输入层神经元的数量,而隐藏层的神经元数量少于输出层。因此,自编码网络的作用是将输入样本压缩到隐藏层,再在输出端重建样本。也就是说,自编码网络输出层与输入层近似相等。

通过调整编码器和解码器的参数,使得重输出的重构数据和一开始的原始数据误差最小,意味着由抽象的特征向量构成的隐藏层数据是原输入数据的有效低维表示^[9]。因此对高维的 HOG 特征可以经过自编码网络进行降维,再输入到分类器中进行分类,提高后续检测的准确度与效率。

1.2 AdaBoost 分类器

AdaBoost 是一种迭代算法,其核心思想是针对同一个训练集训练不同的弱分类器,然后把这些弱分类器集合起来,构成一个更强的最终分类器(强分类器)。其算法本身是通过改变数据分布来实现的,它根据每次训练集之中每个样本的分类是否正确,以及上次的总体分类的准确率,来确定每个样本的权值。将修改过权值的新数据集送给下层分类器进行训练,最后将每次训练得到的分类器最后融合起来,作为最后的决策分类器。

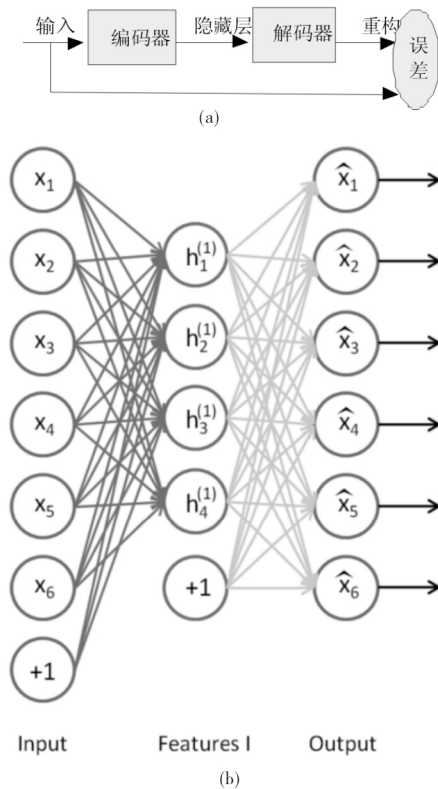


图 1 自编码网络结构及工作原理

Fig.1 Theory of auto-encoder network structure

2 本文提出的行人检测算法

2.1 问题描述

为了对模型进行训练,选取 N_p 个正样本 $I_p^i, i \in 1, \dots, N_p$ 和 N_q 个负样本 $I_q^i, i \in 1, \dots, N_q$, 其对应的 HOG 特征分别是 H_p^i 和 H_q^i . 同时, 经训练的正负样本的分类器向量分别表示为 $(w_p^i \in R^n, b_p^i \in R)$ 和 $(w_q^i \in R^n, b_q^i \in R)$. 为了便于描述, H_p^i, H_q^i 和 (w_p^i, b_p^i) 和 (w_q^i, b_q^i) 改写为 $x_p^i = [H_p^{iT}, 1]^T, x_q^i = [H_q^{iT}, 1]^T$ 和 $y_p^i = [H_p^{iT}, 1]^T, y_q^i = [H_q^{iT}, 1]^T$. 因此, 本文的目的就是希望能通过样本 x_E 学习出一个用于行人精确检测的回归函数 F 即分类向量 $y_E = F(x_E)$.

2.2 基于自编码神经网络的数据降维

由于使用传统的回归算法对高维向量进行分类预测是非常困难的, 本文采用自编码网络来减少回归目标函数的维度. 给定任意样本 $x_i, i \in 1, \dots, N$, 其编码与解码矩阵分别表示为 W_e 和 W_d . 在隐藏层与输出层的激活函数分别表示为 $f_h(\cdot)$ 和 $f_o(\cdot)$. $\{y_i\}$ 是所有样本的类标结果, 因此自编码网络应该确保其同步进行分类.

给定 $g_e(\theta, x)$ 和 $g_d(\theta, x)$ 分别是自编码网络对应

的编解码函数, θ 是模型的参数. 对于一个三层的网络, 我们可以获得 $g_e(\theta, x) = f_h(W_e^T x), g_d(\theta, x) = f_o(W_d^T x)$, 其中 $\theta = (W_e, W_d)$. 对于任意 $x^i, i = 1, \dots, N$, 其类标可以表示为 $l^i \in (-1, 1)$, 也就是说, 当 x^i 为正样本时, 其 $l^i = 1$; 当 x^i 为负样本时, 其 $l^i = -1$; 因此, 本文的自编码目标函数就可以写为如下形式:

$$L = \frac{1}{N} \sum_{i=1}^N \frac{1}{2} \|y^i - y_r^i\|_2^2 + h(1 - u^i, \rho) + \frac{\lambda}{2} \|y_r^i\|_2^2, \quad (1)$$

其中 $y_r^i = g_d \cdot g_e(\theta, y^i)$ 表示 y_i 的重构向量; $u^i = l^i (y_r^i)^T x^i; h(x) = \max(0, x)$; λ 是调节权重和偏置变化的幅度正则参数. 式(1)中, 第一项表示重建误差, 后两个项是由折叶损失函数表示的分类模型, 确保分类器 y_r^i 能够尽可能正确地对样本 $x^i, i \in 1, \dots, N$ 进行分类.

为了求解式(1)的最小优化问题, 本文采用随机梯度下降法训练自动编码网络的参数^[10], 同时利用如下等式对参数 θ 更新直到收敛, 得到

$$\theta_{n+1} = \theta_n - \alpha \frac{\partial L}{\partial \theta}, \quad (2)$$

其中 n 与 α 分别表示迭代次数和学习因子. 目标函数的梯度可以表示为

$$\frac{\partial L}{\partial \theta} = \sum_{i=1}^N [- (y^i - y_r^i) - S(1 - u^i) l^i x^i + \lambda y_r^i] \frac{\partial y_r^i}{\partial \theta}, \quad (3)$$

其中 $S(x) = \frac{1}{2}(\text{sgn}(x) + 1)$, $\text{sgn}(x)$ 是一个符号函数. 显然, 计算 $\partial L / \partial \theta$ 的关键是获取 $\partial y_r^i / \partial \theta$ 的结果. 由于 y_r^i 是自动编码网络的输出, 其对 θ 的到导数可以使用误差反向传播(BP)神经网络进行计算^[11].

通过上面的分析可以看出, 一旦训练好自编码网络, 就可以将 y_i 转换成缩成低维向量 $z^i \in R^{n_c}, \forall i \in 1, \dots, N$ 即: $z^i = f_h(W_e^T y^i)$.

2.3 行人检测及其框架

基于滑动窗口的目标检测方法, 其框架主要包括两个步骤: (1) 生成候选窗口; (2) 对候选区域分类. 然而, 由于候选窗口的数量较多, 对每个候选窗口进行分类处理是非常低效的. 为了提高检测的效率, 本文采用文献[12]提到的目标检测算法(ACF)对图像中可能的区域进行预检测, 生成候选窗口. 实验仿真表明, 经过预处理后的候选区域数量大大降低. 该算法利用快速特征通道缩放来检测区域中的疑似目标, 对行人目标具有低漏检的检测效果.

对原始图像进行目标区域估计的目的是尽可能地获取包含行人的候选窗口. 为了便于分析, 假定获取的疑似区域个数为 M , 因此每一个候选块可以表示为 $\{B^i \in R^{m_i \times n_i} \mid i = 1, 2, \dots, M\}$. 由于分类器的输入参数大小必须相同, 所有的候选块经归一化处理到统一的大小, 即: $\{D^i \in R^{m \times n} \mid i = 1, 2, \dots, M\}$.

对 D^i 中的 HOG 特征进行计算, 并将其转换成特

征向量 d^i ; 然后利用提到的自编码神经网络对特征向量进行特征压缩, 得到向量 v^i . 最后利用训练好的自编码网络的解码函数就可以将压缩特征重构出分类向量 c^i , 其中 $c^i = g_d(\theta v^i)$.

本文提出的行人检测算法分为训练和测试两个阶段. 图 2 和图 3 直观地展示了模型两阶段的框架.

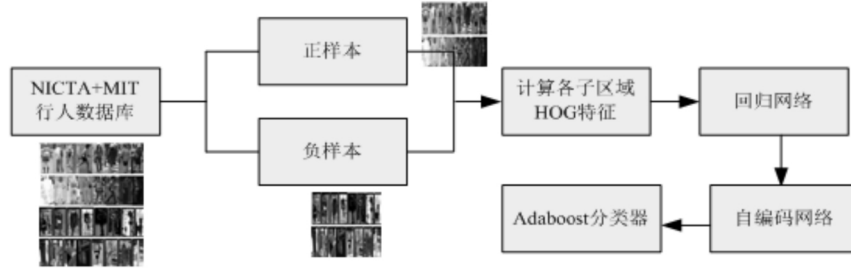


图 2 训练模型
Fig.2 Training Model

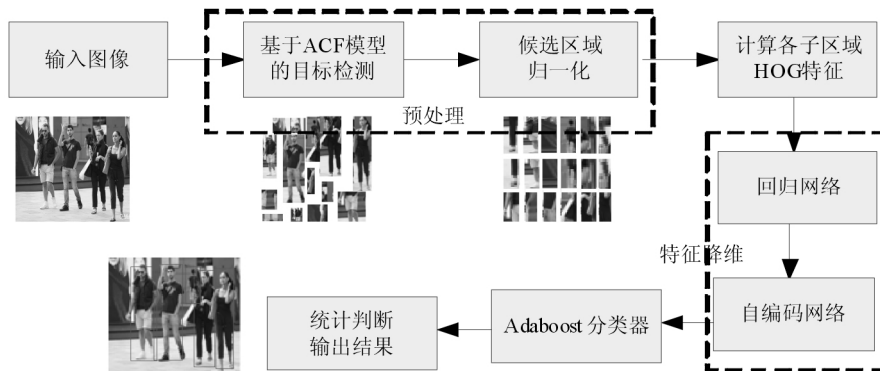


图 3 行人检测框架
Fig.3 Framework for Pedestrian detection

3 仿真实验

3.1 实验数据集

实验使用 NICAT 与 MIT 两个数据集图片来组成训练集. NICAT 数据库是目前规模较大的静态图像行人数据库, 25551 张含单人的图片, 5207 张高分辨率非行人图片. 数据库中已分好训练集和测试集, 方便不同分类器的比较; MIT 数据库共 924 张行人图片, 但只含正面和背面两个视角. 本文也采用了开源的一些手工标注了的互联网图片作为样本库. 由于包含人体的正样本数量远远小于对背景区域随机产生的负样本数量, 这种情况训练的模型效果并不好. 本文将采用文献 [13] 提出的方法选择具有代表性的负样本, 同时也去除仅有平滑区域的负样本. 训练集中正样本数量 12 679, 负样本数量 13 991, 其中所有样本都尺

度归一化为 64×128 .

为了评估本文提出算法的性能, 测试集选择国际上广泛使用且非常具有挑战的行人检测公共测试数据集 INRIA 数据库. 总共选择有 820 张图像, 包含不同光照、分辨率、姿态下的行人.

3.2 参数设置

为了验证本文提出的行人检测算法的有效性, 目前最优算法被采用作为对比算法 HOG^[14], ML^[15], CrossTalk^[16], Shapelet^[17], HOG-LPB^[18], RCNN^[19], 各算法的特点如表 2 所示. 其中 RCNN 则是一个基于深度学习的最优算法. 为了算法公平比较, 本文采用的对比算法都使用作者提供的源代码, 参数及其初始化值为默认值. 本文选择检测错误权衡图 (DET) 曲线和检测时间评价行人目标的检测效果, 其中 DET 表征检测率与每个图像误检率 (FPPI) 的关系. 实验硬件仿真平台采用 Visual Studio 2013, 并在 CPU 为 Core

i5 6600 处理器:主频 3.94GHZ,内存 16G,GPU GTX1050,操作系统:WINDOWS 10 上运行.HOG 特征采用积分图技术快速计算,其最优参数的设置如参考文献[20]所示. AdaBoost 级联分类器采用的开源模型,其参数设置保持一致.

3.3 定性定量分析

为了定性分析本文所提出的算法的行人检测性能,图 4 展示了检测率与 FPPI (False Positives per Image) 的关系曲线.显而易见,本文所提检测算法在 FPPI 一定的情况下,比传统 HOG 算法、VJ 算法、HOG-LPB 算法、Shapelet 算法、ML 算法和 CrossTalk 算法有更高的检测率,也比现有深度学习模型 RCNN 性能较好.为了便于分析,本文主要讨论当 FPPI 在一个固定值下的检测率.

表 3 中是 FPPI = 1 时各算法的检测结果.在 INRIA 数据库中,当 FPPI = 1 时,本文提出算法的检测率是 66.78%,而对比算法的最好结果是

RCNN 60.11%.总之,本文提出的算法在 INRIA 数据库获得了不错的检测效果,主要归功于本文采用的 CAF 预处理策略排除了非目标样本的干扰;同时,本文采用的自编码网络降维,增强了分类器的泛化能力.

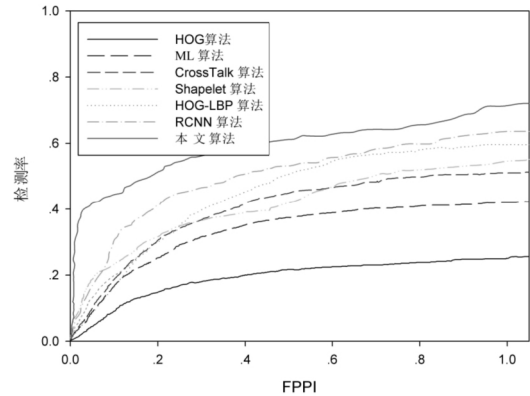


图 4 INRIA 数据库中检测性能对比

Fig.4 Performance comparison for different algorithms

表 2 对比算法特点

Tab.2 Comparison feature for different algorithms

| 算法 | HOG | ML | CrossTalk | Shapelet | HOG-LPB |
|-----|------------|------------|-----------|-----------|------------|
| 特征 | HOG | Multiple | Channels | Gradients | HOG-LPB |
| 分类器 | Linear SVM | Linear SVM | AdaBoost | AdaBoost | Linear SVM |

表 3 INRIA 库的检测率(FPPI=1)

Tab.3 Detection rate for INRIA (FPPI=1)

| 算法 | HOG | ML | CrossTalk | Shapelet | HOG-LPB | RCNN | 本文算法 |
|-------|-------|-------|-----------|----------|---------|-------|-------|
| 检测率/% | 27.91 | 41.90 | 49.75 | 53.28 | 60.01 | 60.11 | 66.78 |

表 4 单帧平均检测时间

Tab.4 Average detection time on single frame

| 算法 | HOG | ML | CrossTalk | Shapelet | HOG-LPB | 本文算法 |
|--------|-------|------|-----------|----------|---------|------|
| 检测时间/s | 13.77 | 8.91 | 9.27 | 3.11 | 10.23 | 1.85 |

由于各对比算法给出的源代码编程语言的不一,这对分析算法的计算时间带来了一定的困难.为了便于计算性能的分析,本文采用仅仅对 HOG、ML、CrossTalk、Shapelet 和 HOG-LPB 进行分析.各算法的当前帧平均检测时间如表 4 所示.虽然各算法单帧的计算时间与其他论文的差异较大,主要是因为本文采用的硬件环境的问题,但也能表明本文算法的有效性.虽然本文提出的算法增加了 CAF 预处理模块对图像中的疑似目标进行检测,但这反而降低了算法的时间复杂度.这是因为模型主要的计算复杂度在于 AdaBoost 与自编码网络的训练与分类计算,CAF 排除了大量非目标区域,大大减少了后续分类的时间.实验结果也表明,本文模型的单帧计算时间最低.在对视频中目标检测时,预处理模块、HOG 模块与分类模块可以并行计算,这进一步减少了处理时间.

本文所提算法的检测结果如图 5 所示,其中白色的矩形框是检测到的行人,黑色为未检测到的目标;灰色是检测到的错误目标.值得注意的是,本文提出的算法具有强垂直边缘(与人类肢体相似)的物体更有可能误认为人体,如图 5(a)路边竖立的架子;另一方面,本文提出的模型对携带行李、衣着宽松等情况的检测效果也相当理想,说明在压缩后的特征能保留最强特征,滤除干扰因素,提升检测精度.

3 结语

现有基于特征分类的行人检测中存在检测速度慢、误差率大的问题,本文提出了一种基于自编码神经网络与 AdaBoost 的快速行人检测算法.该算法利用自编码神经网络将高维原始特征转换成低维抽象特征,同时利用 ACF 模型的目标检测算法对图像进行

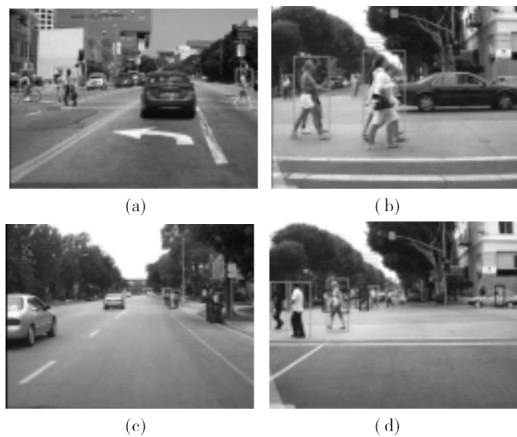


图 5 本文算法的检测结果

Fig.5 Detection results of our proposed algorithm

预处理 获得疑似目标区域 排除大量非目标区域 本文提出的算法可以大大降低单帧图像检测的时间.实验表明本文所提出的模型在具有挑战性的行人检测数据集上的行人检测性能超过现有的检测算法,其检测速度也超过大多数算法,适合工程应用.实验结果也表明本文提出的模型在遮挡情况下的效果还不太理想,下一步工作将重点研究行人的关联特征和上下文信息,提升模型的抗干扰能力.

参 考 文 献

- [1] 王 晓 陈 军. 基于显著区域的行人检测算法[J]. 计算机工程与设计, 2015, 36(2): 17-26.
- [2] 黄成都,黄文广,闫 斌. 基于 Codebook 背景建模的视频行人检测[J]. 传感器与微系统, 2017, 36(3): 144:147.
- [3] 王爱丽,董宝田,王泽胜 等. 融合光流速度场自适应背景建模的交通场景中运动行人检测算法[J], 长安大学学报(自然科学版), 2015, S1:120-134.
- [4] 张文强 路 红 陈义东. 动态场景中的自适应背景建模研究[J]. 中国图像图形学报, 2009, 14(12): 2627-2630.
- [5] Zhang S, Benenson R, Schiele B, et al. Filtered channel features for pedestrian detection [J]. computer vision and pattern recognition, 2015: 1751-1760.
- [6] Guo L, Ge P S, Zhang M H et al. Pedestrian detection for intelligent transportation systems combining AdaBoost algorithm and support vector machine [J]. Expert Syst. Appl. 2012, 39(4): 4274-4286.
- [7] Hosang J H, Benenson R, Dollar P, et al. What Makes for Effective Detection Proposals [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(4): 814-830.

- [8] Girshick R B, Donahue J, Darrell T, et al. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(1): 142-158.
- [9] Lecun Yann, Kavukcuoglu K, Farabet C, et al. Convolutional networks and applications in vision [C]// IEEE. International symposium on circuits and systems. Paris: IEEE, 2010: 253-256.
- [10] Zhang Jie, Shan Shiguang, Kan Meina, et al. Coarse-to-Fine Auto-Encoder Networks (CFAN) for Real-Time Face Alignment [C]//Springer. Proceedings of ECCV. Zurich, Switzerland: Springer, 2014: 1-16.
- [11] Mathur, Mridul Kumar. Image Compression using Robust Error Back Propagation Neural Network with GDAL [J]. International Journal of Research and Engineering, 2017, 14(4): 133-136.
- [12] Dollár P, Appel R, Belongie S, et al. Fast feature pyramids for object detection [J]. IEEE Tran Pattern Anal Mach Intell, 2014, 36(8): 1532-1545.
- [13] Felzenszwalb P, Girshick R, Ramanan D. Object detection with discriminatively trained part based models [J], IEEE Trans Pattern Anal Mach Intell, 2010, 32(9): 1627-1645.
- [14] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]//IEEE. in Proc. CVPR. San Diego: IEEE, 2005: 886-893.
- [15] Walk S, Majer N, Schindler K, et al. New features and insights for pedestrian detection [C]//IEEE, in Proc. CVPR. Shanghai: IEEE, 2010: 1030-1037.
- [16] Dollar P, Appel R, Kienzl W. Crosstalk cascades for frame-rate pedestrian detection [C]//IEEE. in Proc. ECCV. Florence: Spring 2012: 645-659.
- [17] P. Sabzmeydani and G. Mor. Detecting pedestrians by learning shapelet [C]//IEEE. in Proc. CVPR. Minneapolis: IEEE, 2007: 1-8.
- [18] X. Wang, T. X. Han, and S. Yan, An HOG-LBP human detector with partial occlusion handling [C]//IEEE. in Proc. ICCV. Kyoto: IEEE, 2009: 32-39.
- [19] Girshick. Fast R-CNN [C]//IEEE. International Conference on Computer Vision. Massachusetts: IEEE, 2015: 1440-1448.
- [20] Viola P, Jones M. Robust real-time face detection [J], Int. J. Comput. Vis., 2004, 57(2): 137-154.

(责任编辑 雷建云)