

DOI: 10.13382/j.jemi.B1902762

高斯建模和卷积神经网络联合的 红外视频行人检测方法^{*}

崔少华 汪徐德 王江涛 单巍

(淮师范大学 物理与电子信息学院 淮北 235000)

摘要:针对传统红外视频中存在行人互相遮挡导致无法提取前景目标、检测率较低等问题,提出了高斯背景建模和卷积神经网络联合的红外视频行人检测方法。首先,对于连续序列红外图像,通过混合高斯模型提取前景目标;然后,对于行人互相遮挡的前景目标,以亮度曲线图中的谷底为分割点,通过方向投影分离出单行人目标区域;最后,将确定的感兴趣区域输入训练好的LeNet-7系统。3个不同测试集的检测实验表明,该方法具有良好的检测效果,与传统方法相比,该算法的检测率达到99%以上,虚警率低至0%。

关键词: 红外视频;图像处理;混合高斯模型;LeNet-7系统;检测率

中图分类号: TN219;TP391 **文献标识码:** A **国家标准学科分类代码:** 520.20

Infrared video pedestrian detection method via Gauss modeling and convolutional neural network

Cui Shaohua Wang Xude Wang Jiangtao Shan Wei

(College of Physics and Electronic Information, HuaiBei Normal University, HuaiBei 235000, China)

Abstract: A joint infrared video pedestrian method is proposed based on Gauss background modeling and convolution neural network, to address the problems of inability to extract foreground targets and low detection rate in traditional method. Firstly, for continuous sequence infrared images, the foreground targets are extracted by the mixture Gaussian model. Then, for the foreground targets which are occluded by pedestrians, the valley bottom of the luminance curve is used as the segmentation point. While, the single pedestrian target area is separated by directional projection. Finally, the determined region of interest is input into the trained LeNet-7 system. Experiments on three different test sets demonstrate that the proposed method has good detection effect. Compared with the traditional method, the detection rate of the proposed method is over 99%, and the false alarm rate is as low as 0%.

Keywords: infrared video; image processing; Gauss mixture model; LeNet-7 system; detection rate

0 引言

对红外视频进行行人检测,就是对序列红外图像进行行人检测,它是机器视觉的一个重要领域,应用极其广泛。早期,对红外图像进行目标检测,主要采用抑制背景、提高图像的信噪比,从而提取前景目标的方法,包含非自适应和自适应两种算法。Zeng等^[1]提出了顶帽形态学滤波器(top-hat)用于红外图像的小目标检测。李安冬等^[2]提出了局部滤波算法抑制背景检测前景目标的方法。此类方法自适应较差,红外目标提取效果欠佳。之

后,红外图像的目标检测大多采用自适应算法。Wang等^[3]提出了最小均方误差的自适应滤波法,但是对于背景复杂的、运动状态的红外图像,目标提取效果较差。

针对传统的自适应滤波提取候选目标的不足,学者们提出了基于候选目标提取和识别的方法。其中,候选目标常采用图像分割的方法提取感兴趣区域(regions of interest, ROI),将分割的ROI送入分类器进行判别。经典分类器主要有线性Fisher分类器^[4]、支持向量机(SVM)算法^[5]。然而传统分类器依赖人工提取特征进行分类,虽然具有较好的鲁棒性,但在复杂环境中检测率过低。随着2012年神经网络在计算机科学技术方面的

收稿日期: 2019-11-18 Received Date: 2019-11-18

^{*} 基金项目: 国家自然科学基金(11504121)、安徽省教育厅项目(2017kf044, 2018jyxm0530, 201910373104)资助

应用,其高效的识别能力得到广泛公认。学者们将神经网络应用于行人检测领域进行尝试, Ren 等^[6]采用基于区域的卷积神经网络(region convolution neural network, R-CNN)检测方法,但网络耗时过长,对于相互遮挡的行人无法检测。因此, Girshick 等^[7]提出了基于快速区域卷积神经网络(fast region convolution neural network, Fast R-CNN)模型,提高了训练的速度和检测率。然而,车凯等^[8]使用 Fast R-CNN 进行红外图像行人检测时发现, Fast R-CNN 采用搜索算法生成 ROI 数量多,计算量大,实时性差,检测率较低。为了改进 Fast R-CNN 的不足,陈恩加等^[9]采用卷积网络实现目标分类后,再识别模块分割 ROI 图像,降低了输入有效 ROI 的数量,但是网络包含 64 个卷积层,计算量大,容易过拟合。

因此,为了解决上述方法的缺陷,当行人互相遮挡时,本文在高斯背景建模基础上采用方向投影法分离 ROI。引入 LeNet-7 系统(卷积神经网络的一种)对 ROI 分类识别,测试结果表明,本文方法具有良好的检测效果,与传统方法进行对比,本文方法更好的提高了检测率。

1 高斯模型在线背景建模原理

Stauffer 等^[10]提出了多个高斯分布组合的混合高斯模型,以解决红外视频中行人背景亮度变化、重复运动等问题。对于红外图像,将同一个位置的像素点灰度值看成一个时间序列 $\{S_1, S_2, \dots, S_t\}$, 则 t 时刻该像素的概率观测值可用混合高斯模型表达为:

$$G(X_t) = \sum_{i=1}^k \omega_{i,t} \times p(S_t, \mu_{i,t}, \sigma_{i,t}^2) \quad (1)$$

式中: $\omega_{i,t}$ 是第 i 个高斯分布在时间 t 的权值, $\mu_{i,t}$ 和 $\sigma_{i,t}$ 为第 i 个高斯函数的期望值与标准差。 k 是混合高斯模型中包含的高斯函数的个数, p 是高斯概率密度函数:

$$p(S_t, \mu_{i,t}, \sigma_{i,t}^2) = \frac{1}{\sqrt{2\pi}\sigma_{i,t}} e^{-\frac{(X_t - \mu_{i,t})^2}{2\sigma_{i,t}^2}} \quad (2)$$

采用上述混合高斯模型,可判断一个像素是否为背景,其过程如下。

1) 初始化高斯函数。

2) 分析一个新的像素。若该像素观测值位于混合高斯模型中某一高斯函数的 2.5 个标准差范围内,则该像素与对应高斯函数匹配;若不匹配,则将概率最低的高斯函数替换为一个新的高斯函数^[10]。

3) 若匹配,则对所有高斯函数进行权值更新。

$$\omega_{i,t} = (1 - \alpha)\omega_{i,t-1} + \alpha(M_{i,t}) \quad i = 1, 2, \dots, k \quad (3)$$

式中: α 为学习参数,通常取 0.05; $M_{i,t}$ 是该像素的高斯函数匹配标识,若存在高斯函数与该像素匹配,则取值为

1, 否则为 0。

4) 归一化。将新生成的所有权值归一化处理,将其匹配的高斯函数参数更新为:

$$\mu_i = (1 - \eta)\mu_{i-1} + \eta(X_t) \quad (4)$$

$$\sigma_i^2 = (1 - \eta)\sigma_{i-1}^2 + \eta(x_t - \mu_i)^T(x_t - \mu_i) \quad (5)$$

式中: $\eta = \alpha p(X_t | \mu_i, \sigma_i^2)$, 而其他不匹配的高斯函数参数值保持不变。

5) 对下一帧重复步骤 1)~4)。

设置背景阈值 ω_0 , 将权值之和大于 ω_0 的前 n 个高斯函数作为背景模型:

$$BA = \text{argmin} \left(\sum_{k=1}^n \omega_k > \omega_0 \right) \quad (6)$$

确定背景模型后,即可将图像中的像素点进行分类,如果观测的像素与前 n 个高斯函数中的某个匹配,则认为该像素为背景的,否则为前景的。

2 高斯建模联合方向投影法

红外视频中往往出现行人之间存在相互遮挡的情况,按照上述高斯背景建模的方法,得到的 ROI 发生粘连,使得行人检测准确率大大降低。因此本文采用自适应阈值的“方向投影法”对行人互相遮挡情况下 ROI 的确定进行处理。

2.1 方向投影法

1) 阈值的选择

设置阈值的时应采取自适应的方式,相应的分割阈值为:

$$T = \omega \times \max(im) + (1 - \omega) \cdot \text{mean}(im) \quad (7)$$

式中: ω 为加权系数 ($0 \leq \omega \leq 1$); im 为原始图像; $\max$ 为图像灰度的最大值; mean 为图像灰度的均值。分割结果如图 1(a)、(d) 所示。通过大量实验,本文最终确定 ω 的值为 0.25。

2) 垂直投影

将经过阈值分割的图像向 X 轴作垂直投影,得到图像的灰度垂直投影曲线,如图 1(b)、(e) 所示。图 1(b) 人体灰度区域表现为凸起的山峰,且无粘连现象。为了定量计算,在投影曲线中凸起山峰的两侧分别确定上升和下降曲线的中点,将这两个中点分别作为一条亮度带的起始点与结束点,可得一系列垂直于 X 轴的亮度带,而人体可能存在的区域则被包含在亮度带中。图 1(e) 垂直投影曲线发生粘连,需要另做处理。

3) 水平投影

将垂直投影得到的亮度带向 Y 轴做水平投影,亮度带起始和结束点选取方法与垂直投影相同。

4) ROI 的确定

将垂直投影和水平投影得到的亮度带同时放入原始

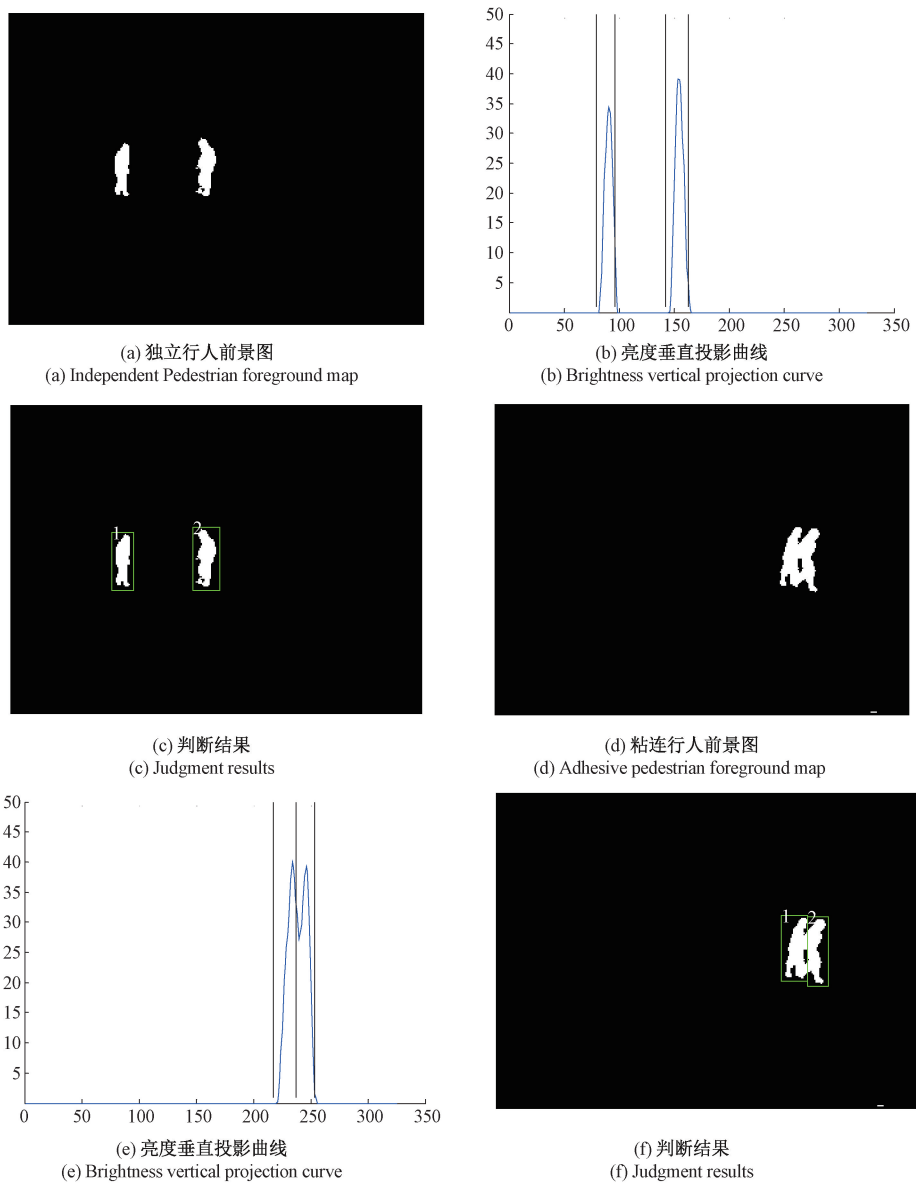


图 1 行人特征区别
Fig. 1 Pedestrian characteristics

图像中相应的位置,此时原始图像可以被分割为许多高亮度的矩形区域,如图 1(c) 所示。

上述方法适合独立且无遮挡物体 ROI 的确定,将所得 ROI 直接输入训练好的卷积神经网络即可对行人进行检测。然而该方法并不能准确处理行人互相遮挡的情况(图 1(d)),此时系统的检测率较低。因此,本文采用基于高斯背景建模的方向投影法确定遮挡行人的 ROI。

2.2 高斯背景建模的方向投影法

由图 1 可知,两个相互独立的行人,其亮度垂直投影曲线也是相互独立的;存在遮挡的行人,其亮度垂直投影曲线也发生了粘连。观察图 1(e) 可以发现,粘连行人的头部分别位于两条亮度投影曲线的峰顶,而粘连部位则

位于投影曲线中两个峰之间的谷底处。因此,基于高斯背景建模的方向投影法分割 ROI 具体步骤如下,流程如图 2 所示。

1) 由式(1)~(6)将红外视频所得序列图像进行前景划分。

2) 对前景目标进行阈值分割,此时前景中包含多个互相遮挡的行人,如图 1(d) 所示。

3) 垂直投影。如图 1(e) 所示,以粘连行人的波谷为分割点,分别对存在双峰特征的窗口进行 2.1 节所述垂直投影。

4) 水平投影。以 2.1 节所述水平投影法,得到多人的 Y 轴投影。

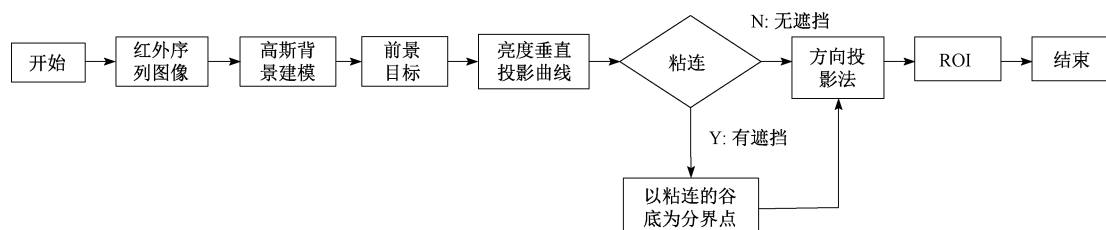


图2 本文分割 ROI 的流程

Fig. 2 Flow chart of segmentation ROI in this paper

5) ROI 确定。此时,相互遮挡、粘连的两人分别包含在两个不同的矩形亮度框中,即被划分为两个 ROI,如图 1(f) 所示。

将采用本文方法分割的遮挡和无遮挡 ROI,输入训练好的卷积神经网络进行分类识别,可实现红外视频的行人检测。因此,接下来将对卷积神经网络的构建和训练进行探讨。

3 LeNet-7 系统

2006 年以来,深度学习卷积神经网络在机器视觉方面得到广泛应用,它是由多个单层卷积神经网络组成的可训练的监督学习网络,每个单层卷积神经网络包含卷积、非线性变换和下采样(池化),处理能力都远超以往的浅层学习网络^[11]。卷积阶段的输入特征图(feature maps)记作 x_i ,将输出特征图记作 y_j 。输出特征图中的每个神经元与输入特征图中一定大小的区域相连接,这个区域被称为局部感受野或滑动窗口,神经元通过与感

受野相同大小的卷积核对感受野进行特征提取。连接 x_i 和 y_j 的权值记作 ω_{ij} ,该权值 ω_{ij} 就是可训练的卷积核。其关系表达式为:

$$y_j = \sum_i \omega_{ij} \cdot x_i + b_j \quad (8)$$

早在 1989 年,Lecun 等^[12]提出的 LeNet-5 系统被认为是深度学习最经典的应用。笔者将传统 LeNet-5 系统直接用于红外图像行人检测,实验结果并不令人满意,原因是红外图像采集环境复杂,包含干扰目标(非行人目标)种类较多。扩展后的 LeNet-7 系统结构如图 3 所示,除输入层外,该网络共包含 3 个卷积层、3 个池化层和 1 个输出层,其中 X, Y, Z 分别是各卷积层与池化层中特征图的个数, M, N, P, Q, R, S 和 m, n, p, q, r, s 分别为各特征图的宽和高。若是 LeNet-5 系统,则没有 C5 层和 S6 层, S4 层直接全连接到输出层。由图 3 可知,影响系统的参数包含网络层数、卷积核大小和特征图个数,目前对于这 3 个参数的选择没有明确数学标准化,大多数依赖设计者的经验,本文采用检测错误率最小的试选法,通过实验数据分析确定上述参数,使系统达到最佳检测效果。

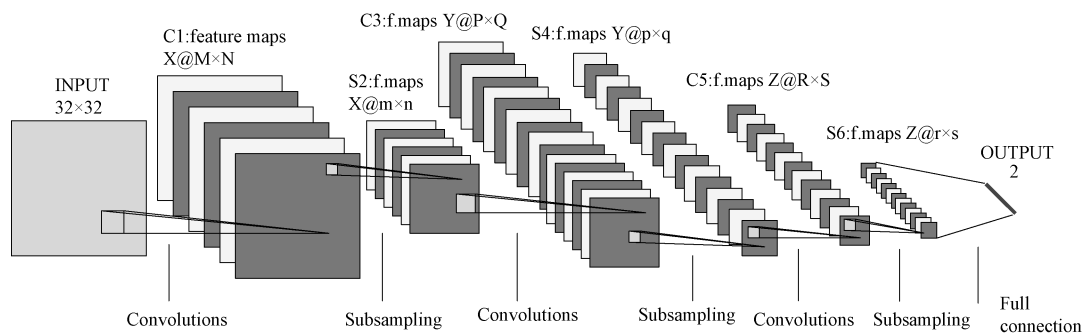


图3 LeNet-7 系统结构

Fig. 3 LeNet-7 System

3.1 网络层数的确定

本文采用波士顿大学 2014 年提供的 BU-TIV (thermal infrared video) Benchmark 热红外视频数据库训练系统,该数据库包含的图像环境多样,相对复杂,具有很高的实用性^[13]。随机提取 3 000 个行人样本,2 000 个用于训练,1 000 个用于测试;随机提取 3 000 个非行人

样本中,2 000 个用于训练,1 000 个用于测试。使用时所有样本归一化为 32×32 ,步长为 1,激活函数为 ReLU,历经 8 000 个训练次数。

卷积核必须保证输出特征图尺寸为整数,且后一层的卷积核尺寸不应大于上一层尺寸。最终列出图 3 的所有可能参数,如表 1 所示,可知组合 2、5、9、10、11、18 输

出特征图尺寸不是整数,组合不成立。组合 1、3、8、12 是单层卷积结构,不符合本文网络设计思想。组合 4、6、13、14、15、17 可以构成 5 层深度卷积网络,而组合 7、16 最多可以构成 7 层深度卷积网络。考虑网络层数越深,表达能力相应越强,因此,本文最终确定网络层数为 7 层。表 1 的组合 7、16 不仅确定了网络层数,还确定了各层的卷积核大小,目前只有各层的特征图个数未确定,接下来将对此参数进行选取。

表 1 可能的网络结构参数

Table 1 Possible network structure parameters																	
组合	宽度	高度	C1	M	N	m	n	C3	P	Q	p	q	C5	R	S	r	s
1	32	32	3×3	30	30	15	15										
2	32	32	3×3	30	30	15	15	3×3	13	13	6.5	6.5					
3	32	32	5×5	28	28	14	14										
4	32	32	5×5	28	28	14	14	5×5	10	10	5	5					
5	32	32	5×5	28	28	14	14	5×5	10	10	5	5	3×3	3	3	1.5	1.5
6	32	32	5×5	28	28	14	14	3×3	12	12	6	6					
7	32	32	5×5	28	28	14	14	3×3	12	12	6	6	3×3	4	4	2	2
8	32	32	7×7	26	26	13	13										
9	32	32	7×7	26	26	13	13	7×7	7	7	3.5	3.5					
10	32	32	7×7	26	26	13	13	5×5	9	9	4.5	4.5					
11	32	32	7×7	26	26	13	13	3×3	11	11	5.5	5.5					
12	32	32	9×9	24	24	12	12										
13	32	32	9×9	24	24	12	12	9×9	4	4	2	2					
14	32	32	9×9	24	24	12	12	7×7	6	6	3	3					
15	32	32	9×9	24	24	12	12	5×5	8	8	4	4					
16	32	32	9×9	24	24	12	12	5×5	8	8	4	4	3×3	2	2	1	1
17	32	32	9×9	24	24	12	12	3×3	10	10	5	5					
18	32	32	9×9	24	24	12	12	3×3	10	10	5	5	3×3	3	3	1.5	1.5

3.2 各层特征图的确定

1) C1 层特征图的确定

在单层卷积网络中(图 3 中的 C1 和 S2),固定 9×9 (对应组合 16)和 5×5 (对应组合 7)的卷积核尺寸,依次改变特征图个数进行实验。传统 LeNet-5 系统用于数字识别时,在最后一层使用 16 张特征图实现十分类(数字 0~9),行人检测只需实现二分类,因此实验选取特征图个数为 1~16。采用检测错误率衡量网络提取特征的能力,错误率的描述如式(9)所示,错误率越低,

网络提取特征的能力越强,错误率为 0.5 表示系统无法收敛。

$$\text{错误率} = \frac{ER}{TOTAL} \times 100\%$$

(9)

式中:ER 为错误样本个数,TOTAL 为总样本个数^[14]。

单层卷积网络实验结果如表 2 所示,可知 9×9 卷积核对应的 3 个特征图时,或 5×5 卷积核对应 2 个特征图时,系统错误率最低。因此确定组合 16 中 C1 层特征图个数为 3,组合 7 特征图个数为 2。

表 2 C1 层不同卷积核和特征图对应错误率

Table 2 Corresponding error rates of different convolution kernels and characteristic map in layer C1																
卷积核	特征图															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
9×9	0.061 8	0.052 3	0.019 7	0.029 9	0.031 0	0.031 0	0.040 5	0.022 5	0.500 0	0.027 4	0.500 0	0.500 0	0.500 0	0.500 0	0.500 0	0.500 0
5×5	0.033 0	0.020 2	0.030 4	0.026 8	0.030 5	0.350 0	0.036 0	0.500 0	0.031 0	0.500 0	0.500 0	0.500 0	0.500 0	0.500 0	0.500 0	0.500 0

2) C3 层特征图的确定

将上一层网络确定的参数作为 C3 层的输入,联合 C1 层构成两层网络,固定 C3 层 5×5 (组合 16)和 3×3 (组合 7)的卷积核尺寸,依次改变特征图个数 1~10 进行实验(由表 2 可知大部分不收敛的情况出现在特征图个数大于 10 的时候)。实验结果表明 5×5 卷积核对应 3 个特征图时,或 3×3 对应 8 个特征图时,系统错误率最低,因

此,确定组合 16 中 C3 层特征图个数为 3,组合 7 特征图个数为 8。

3) C5 层参数的确定

将上一层网络确定的参数作为 C5 层的输入,联合 C1、C3 层构成 3 层网络,固定 C5 层 3×3 (组合 16 和 7)卷积核尺寸,依次改变特征图个数为 1~10 进行实验。实验结果表明卷积核 3×3 时 7 个特征图对应的错误率

最小,因此确定组合 16 和组合 7 中 C5 层特征图个数为 7。

3.3 最终参数确定

针对上节探讨的组合 7 和组合 16,本文采用检测错误率衡量不同组合提取特征的能力,对二者进行最终筛选,实验结果如表 3 所示。由表 3 可知,组合 16 对应网络的错误率最小,组合 7 错误率最高。此时组合 16 的特征表达维数为 $1 \times 1 \times 7 \times 3 \times 3 = 63$,组合 7 的特征表达维数为 $2 \times 2 \times 7 \times 8 \times 2 = 448$ 。组合 7 由于特征向量维数过大,产生了过拟合使得错误率增大,而组合 16 只有 63 维特征向量,却具有相对强的表达能力。因此,本文最终确定 7 层网络的参数为表 1、3 中组合 16 对应参数。

表 3 7 层网络各组合最优参数与结果

Table 3 Optimal parameter and experimental results of all combination in seven layers.

组合	C1	X	C3	Y	C5	Z	错误率	维数
7	5×5	2	3×3	8	3×3	7	0.033 5	448
16	9×9	3	5×5	3	3×3	7	0.017 5	63

4 实验结果

本文设计之初,期望建立 CPU 可快速运行的卷积神经网络,在 Intel i7 9750H,4.5 GHz,中央处理器 32 GB 内存的环境下,将分割的序列红外图像的 ROI 输入构建的 LeNet-7 系统,用于检测系统的识别能力。本文采用俄亥俄立大学提供的 OTCBVS Benchmark Dataset 数据库^[15]和 Terravic Motion IR Database 数据库^[16]进行测试,并将 ROI 全部归一化为 32×32 ,使得测试集与训练集大小相同。

使用 3 个不同的红外图像测试集进行红外人体检测实验,测试集 1 来源于 OSU Thermal Pedestrian 数据库,共有 23 幅图像,共含有 101 个人体目标,为多人人体测试集,其中,包含行人互相粘连(遮挡)的情况。测试集 2 来源于 Terravic Motion IR Database Database 数据库,由 230 幅图像组成,每幅图像中含有 2 个携带武器人体目标,共含有 460 个人体目标,其中,包括行人互相遮挡的情况。测试集 3

来源于 OSU Color-Thermal Database 数据库,由 200 幅图像组成,每幅图像含有一个人体目标,为单人人体测试集。图 4 所示为 3 个测试集中的一部分检测结果,其中的绿色矩形框表示检测到的行人的位置区域。由图 4 可知,在人体之间发生轻度遮挡时,本文方法仍然能够正确检测出序列红外图像当中的人体目标,表明本文方法有效可行。

为直观反映检测效果,选取如下性能指标衡量系统能力^[17]:

$$\text{检测率} = \frac{TP}{TP + FP} \times 100\%$$

(10)

$$\text{虚警率} = \frac{FN}{TP + FN} \times 100\%$$

(11)

式中:TP 为正确检测的人体目标数量;FP 为未被正确检测的人体目标数量;FN 为被误判为行人的非人体目标数量。并将本文方法与传统的方向梯度直方图(HOG)结合 SVM 分类、Fast R-CNN 模型进行对比。其中 Fast R-CNN 的开源结构一般有 VGG-CNN-M1024、ZF、VGG16 三种^[18-20],其网络层次依次加深,选取最深层的 VGG16 网络与建立的 LeNet-7 网络进行对比。结果如表 4 所示。

由表 4 可知,相比两种传统方法,本文方法的检测率更高,虚警率更低。其中,在较简单的测试集 3(单行人目标)中,HOG+SVM 和 Fast R-CNN 与本文方法性能一致。在行人之间出现粘连、遮挡情况较多的测试集 2 中,本文方法的检测率更高,没有误判。在复杂环境的测试集 1(多行人目标)中,本文方法检测率最高,无误判行人,Fast R-CNN 模型受到环境等的干扰,检测率较低,且出现 1 个行人目标误判。

从检测时间来看,传统 HOG+SVM 由于采用固定尺寸的、单一形状的特征向量作为线性分类的输入,当图像处于复杂环境时遍历窗口数目过多,检测耗时最长。采用卷积神经网络替代传统特征向量的匹配后,检测时间大大减少,相比 Fast R-CNN,本文方法的检测时间更低,可将 Fast R-CNN 的检测缩短约一半时间。在单行人测试集 3 中,本文方法的检测时间约为 Fast R-CNN 的 1/3,说明本文方法对检测时间的提升效果明显。



(a) 测试集1
(a) Test set1

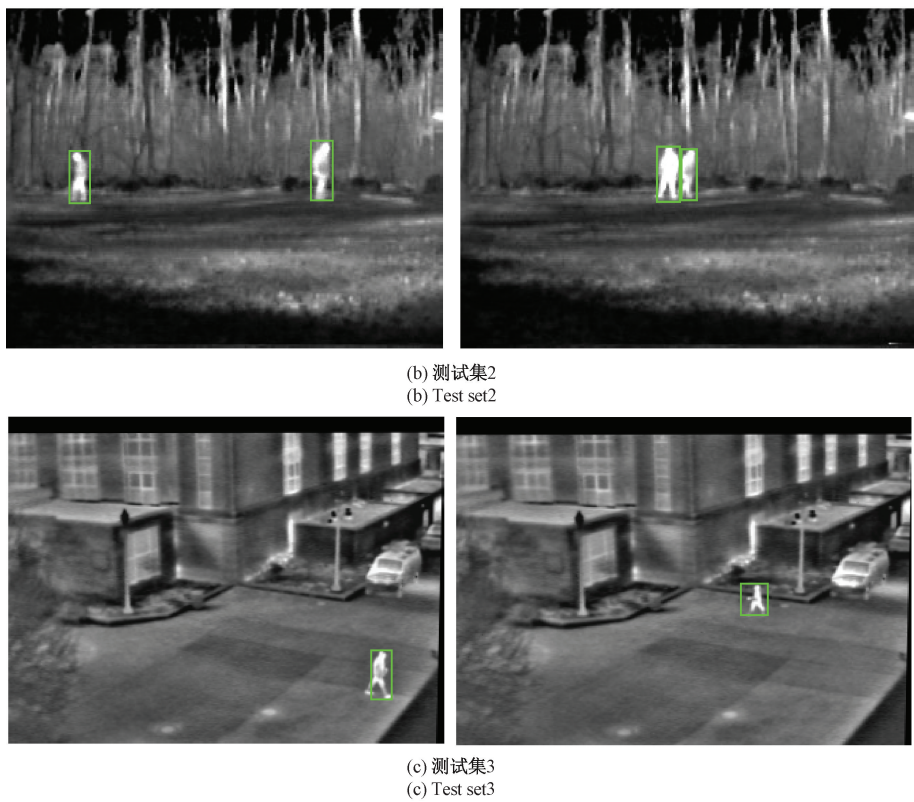


图 4 部分实验结果

Fig. 4 Part of the experimental results

表 4 实验结果

Table 4 Experimental results

测试集	不同方法	行人数量	<i>TP</i>	<i>FP</i>	<i>FN</i>	检测率/%	虚警率	检测时间/ms
1	HOG+SVM	101	98	3	2	97	1.9%	402.27
	Fast R-CNN		99	2	1	98	1.0%	31.05
	本文方法		100	1	0	99	0%	14.88
2	HOG+SVM	460	451	9	1	98	0.2%	318.03
	Fast R-CNN		453	7	0	98.4	0%	28.66
	本文方法		456	4	0	99.1	0%	14.30
3	HOG+SVM	200	200	0	0	100	0%	204.72
	Fast R-CNN		200	0	0	100	0%	27.39
	本文方法		200	0	0	100	0%	10.24

5 结 论

本文采用高斯背景建模的方法解决红外视频中快速变化、无法进行背景抽取的问题,针对行人互相遮挡的情况,在高斯背景建模的方法上结合方向投影法对粘连行人进行 ROI 分离,最后采用训练好的 LeNet-7 系统对 ROI 进行分类识别。对于 LeNet-7 系统,由于参数无明确的数学标准化,在最小检测错误率的实验数据中依次确定网络层数、卷积核大小、特征图个数,该参数不依赖设计者

经验,避免了传统卷积神经网络单一的特征提取模式。测试结果表明,与传统检测算法对比,本文构造的 LeNet-7 系统具有较高的检测率和很好的迁移性。然而,本文方法还有进一步改进的空间,从测试集 2 的实验结果来看,当行人之间存在重度遮挡时,本文方法仍然无法将前景目标中的 ROI 分离,导致系统无法正确检测,因此,在接下来的工作中,将进一步研究针对重度遮挡的行人 ROI 进行分离,并进一步改进深度卷积神经网络,提高红外视频的行人检测正确率。

参考文献

- [1] ZENG M, LI J X, PENG Z. The design of top-hat morphological filter and application to infrared target detection [J]. *Infrared Physics & Technology*, 2006, 48(1): 67-76.
- [2] 李安冬. 基于压缩感知的红外弱小目标检测技术研究[D]. 长沙: 国防科技大学, 2015: 20-30.
LI A D. The research on technology of infrared small target detection based on compressed sensing [D]. Changsha: National University of Defense Technology, 2015: 20-30.
- [3] WANG P, TIAN J W, GAO C Q. Infrared small target detection using directional high-pass filters based on LS-SVM [J]. *Electronics Letters*, 2009, 45(3): 156-158.
- [4] XU Y L, MA B P, HUANG R, et al. Person search in a scene by jointly modeling people commonness and person uniqueness [C]. *Proceedings of the 22nd ACM International Conference on Multimedia*. New York: ACM Press, 2014: 937-940.
- [5] 席林, 孙韶媛, 李琳娜, 等. 基于 SVM 模型的单目红外图像深度估计 [J]. *激光与红外*, 2012, 42(11): 1311-1315.
XI L, SUN S Y, LI L N, et al. Depth estimation of monocular infrared image based on SVM model [J]. *Infrared Physics & Technology*, 2012, 42(11): 1311-1315.
- [6] REN S, HE K, 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, 39(6): 1137-1149.
- [7] GIRSHICK R. Fast R-CNN [C]. *Proceedings of The IEEE International Conference on Computer Vision*, 2015: 1440-1448.
- [8] 车凯, 向郑涛, 陈宇峰, 等. 基于改进 Fast R-CNN 的红外图像行人检测研究 [J]. *红外技术*, 2018, 40(6): 578-584.
CHE K, XIANG Z T, CHEN Y F, et al. Research on infrared image pedestrian detection based on improved fast R-CNN [J]. *Infrared Technology*, 2018, 40(6): 578-584.
- [9] 陈恩加, 唐向宏, 傅博文. Faster R-CNN 行人检测与再识别为一体的行人检索算法 [J]. *计算机辅助设计与图形学学报*, 2019, 31(2): 332-339.
CHEN E G, TANG X H, FU B W. Pedestrian search method based on faster R-CNN with the integration of pedestrian detection and re-identification [J]. *Journal of Computer Aided Design & Graphics*, 2019, 31(2): 332-339.
- [10] STAUFFER C, GRIMSON W E L. Adaptive background mixture models for real-time tracking [C]. *Computer Vision and Pattern Recognition, IEEE*, 1999: 6346911.
- [11] 刘玉杰, 庞芸萍, 李宗民, 等. 融合抽象层级变换和卷积神经网络的手绘图像检索方法 [J]. *浙江大学学报 (理学版)*, 2016, 43(6): 657-663, 675.
LIU Y J, PANG Y P, LI Z M, et al. Sketch based image retrieval based on abstract-level transformation and convolution neural network [J]. *Journal of Zhejiang University (Science Edition)*, 2016, 43(6): 657-663, 675.
- [12] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [13] ZHENG W, FULLER N, THERIAULT D, et al. *IEEE Conference on Computer Vision and Pattern Recognition [DB/OL]*. (2014-6-24) [2019-12-18]. <http://csr.bu.edu/BU-TIV/BUTIV.html>.
- [14] 孙锐, 王慧慧, 叶子豪. 融合深度感知特征与核极限学习机的行人检测 [J]. *电子测量与仪器学报*, 2019, 33(2): 39-47.
SUN R, WANG H H, YE Z H. Pedestrian detection based on combining depth perception features with kernel extreme learning machine [J]. *Journal of Electronic Measurement and Instrumentation*, 2019, 33(2): 39-47.
- [15] HAMMOUD R I. *OTCBVS Benchmark Dataset Collection [DB/OL]*. (2014-6-22) [2019-12-18]. <http://vcip-lkstate.org/pbvs/bench/>
- [16] HAMMOUD R I. *Terravic Motion IR Database [DB/OL]*. (2014-6-22) [2019-12-18]. <http://vcip-lkstate.org/pbvs/bench/Data/05/download.html>.
- [17] 王文秀, 傅雨田, 董峰, 等. 基于深度卷积神经网络的红外船只目标检测方法 [J]. *光学学报*, 2018, 38(7): 160-166.
WANG W X, FU Y T, DONG F, et al. Infrared ship target detection method based on deep convolution neural network [J]. *Acta Optica Sinica*, 2018, 38(7): 160-166.
- [18] 叶国林, 孙韶媛, 高凯珺, 等. 基于加速区域卷积神经网络的夜间行人检测研究 [J]. *激光与光电子学进展*, 2017, 54(8): 081003.
- [19] YE G L, SUN S H, GAO K J, et al. Research on night pedestrian detection based on accelerated area convolution neural network [J]. *Laser & Optoelectronics Progress*, 2017, 54(8): 081003.
- [19] 谢晓蔚, 史健芳. 弱监督卷积神经网络的多目标图像检测研究 [J]. *电子测量与仪器学报*, 2019, 33(6): 31-37.
XIE X W, SHI J F. Research of convolutional neural

networks with weakly-supervised learning on multi-object image detection[J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(6): 31-37.

- [20] 陈丹, 林清泉. 基于级联式 Faster RCNN 的三维目标最优抓取方法研究[J]. 仪器仪表学报, 2019, 40(4): 229-237.

CHEN D, LIN Q Q. Research on 3D object optimal grasping method based on cascaded Faster RCNN [J]. Chinese Journal of Scientific Instrument, 2019, 40(4): 229-237.

作者简介



崔少华, 2007 年于安徽理工大学获得学士学位, 2012 年于西安交通大学获得硕士学位, 现为淮北师范大学讲师, 主要研究方向为图像处理、地震信号处理。

E-mail: flower0804@126.com

Cui Shaohua received B. Sc. degree from Anhui University of Technology in 2007, and M. Sc. degree from Xi'an Jiaotong University in 2012, respectively. Now she is a lecturer at Huaibei Normal University. Her main research interests include image processing and seismic signal processing.



汪徐德, 2004 年于安徽师范大学获得学士学位, 2007 年于华南师范大学获得硕士学位, 2016 年于华南师范大学获得博士学位, 现为淮北师范大学副教授, 主要研究方向为非线性超快脉冲技术以及被动锁模光纤激光器的研究。

E-mail: 99483950@qq.com

Wang Xude received B. Sc. degree from Anhui Normal University in 2004, M. Sc. degree from South China Normal University in 2007, and Ph. D. degree from South China Normal University in 2016, respectively. Now he is an associate professor at Huaibei Normal University. His main research interests include nonlinear ultrafast pulse technology and passive mode-locked fiber laser.



王江涛, 2002 年于青岛大学获得学士学位, 2005 年于青岛大学获硕士学位, 2008 年于南京理工大学获得博士学位, 现为淮北师范大学教授, 主要研究方向为图像处理、模式识别、计算机视觉。

E-mail: 18613540@qq.com

Wang Jiangtao received B. Sc. degree from Qingdao University in 2002, M. Sc. degree from Qingdao University in 2005, and Ph. D. degree from Nanjing University of Technology in 2008, respectively. Now he is a professor at Huaibei Normal University. His main research interests include image processing, pattern recognition, computer vision.



单巍, 2005 年于合肥工业大学获得学士学位, 2009 年于合肥工业大学获硕士学位, 现为淮北师范大学讲师, 主要研究方向为图像处理。

E-mail: 79733238@qq.com

Shan Wei received B. Sc. degree from Hefei Polytechnic University in 2005, M. Sc. degree from Hefei Polytechnic University in 2009. Now he is a lecturer at Huaibei Normal University. His main research interest is image processing.