

改进的 CNN 用于单帧红外图像行人检测的方法

崔少华, 李素文, 黄金乐, 单巍

(淮北师范大学 物理与电子信息学院, 安徽 淮北 235000)

摘要: 针对全卷积神经网络对单帧红外图像行人检测计算量大、检测率较低等问题, 提出了一种改进的 LeNet-7 系统对红外图像行人检测的方法。该系统包含 3 个卷积层、3 个池化层, 通过错误率最小的试选法确定每层参数, 以波士顿大学建立的 BU-TIV 数据库训练系统。首先, 以俄亥俄州立大学建立的 OTCBVS 和 Terravic Motion IR Database 红外数据库作为测试图像; 然后, 采用自适应阈值的垂直和水平投影法得到感兴趣区域 (regions of interest, ROI); 最后, 将得到的 ROI 输入训练好的系统进行测试。3 个测试集检测实验表明, 本文方法具有良好的识别能力, 与不同实验方法相比, 本文方法能有效提高检测率。

关键词: 图像处理; LeNet-7 系统; 单帧红外图像; 检测率

中图分类号: TP391 文献标志码: A 文章编号: 1001-8891(2020)05-0238-07

A Method of Pedestrian Detection Based on Improved CNN in Single-frame Infrared Images

CUI Shaohua, LI Suwen, HUANG Jinle, SHAN Wei

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

Abstract: We proposed an improved method of pedestrian detection in infrared images based on the LeNet-7 system, to address the problems of large computation and low detection rates in traditional methods based on a full convolution neural network. The system consists of three convolution layers and three pooling layers. The trail selection method with the smallest error rate is used to determine the parameters of each layer, while the BU-TIV database, established by Boston University, is used to train the system. Firstly, the Object Tracking and Classification in and Beyond the Visible Spectrum(OTCBVS) and Terravic Motion IR Database, established by Ohio State University, are used to test images. Then, the region of interest (ROI) is obtained by vertical and horizontal projection with adaptive thresholds. Finally, the ROI is input into the trained system for testing. Experiments on three test sets demonstrate that the proposed method has good recognition ability. Compared with different experimental methods, the proposed method can effectively improve the detection rate.

Key words: image processing, LeNet-7 system, single-frame infrared image, detection rate

0 引言

行人检测是机器视觉领域的重要分支, 目前已经得到广泛应用, 红外视频监控作为当今社会安防的重要手段, 采用行人检测技术对其进行分析与捕捉具有极高的应用价值。对红外视频的行人检测就是对组成视频的单帧红外图像行人检测。由于人体在图像中的大小和出现的位置都无法确定, 所以在行人检测研究的早期, 往往采取对图像进行多尺度遍历搜索的方式

检测是否存在人体目标。例如, Nanda 等^[1]通过人体亮度分布概率设计了一种亮度概率模板, 但由于人体姿态的多样性, 该方法仍需在不同尺度下使用多个模板进行匹配。Bertozzi 等^[2]构建了基于正面人体对称性的形态学人体模型, 并对人体在图像中的大小进行了预估, 然后在图像中进行多尺度的搜索以确定可能存在人体的候选区域, 最后将候选区域与构建的人体模型进行匹配完成红外图像中的行人检测。这两种方法虽然不易出现漏检, 具有较好的鲁棒性, 但在实时

收稿日期: 2019-06-25; 修订日期: 2019-12-18。

作者简介: 崔少华 (1983-), 女, 硕士, 讲师, 主要从事信号去噪、图像处理等方面的研究。E-mail: flower0804@126.com。

基金项目: 国家自然科学基金面上项目 (41875040); 安徽省教育厅项目 (2018jyxm0530, 2017kfk044, KJ2017B008)。

性上都大打折扣。因此,之后的红外行人检测方法中,都采用了感兴趣区域分割和目标识别的方法,以避免对图像进行多尺度遍历搜索,提高系统的实时性。例如, Gao 等^[3]采用基于立体视觉的方法获取 ROI,根据提取行的位置和姿态对其进行分类,用于检测路面、障碍物等,然而,该方法对获取的 ROI 进行分类仍然依赖人工,准确率较低。

随着 2012 年神经网络在计算机科学技术方面的应用,其高效的网络结构和识别能力得到广泛公认。学者们尝试将神经网络应用于行人检测领域,其中, Girshick 等^[4]提出基于建议区域 ROI 获取的卷积神经网络 (Convolution Neural Network, CNN) 行人检测方法,利用 CNN 网络代替了传统的人工分类检测。许若等^[5]采用包含 12 个卷积层的 CNN 将感兴趣区域图像输入网络,由行人目标概率图生成红外图像中的行人目标标记框,用以训练和检测 CNN。谭康霞等^[6]提出基于 YOLO 模型的红外图像行人检测方法,利用 23 个卷积层和 5 个池化层构成卷积神经网络,对实际道路采集的红外数据进行训练与测试。陈恩加等^[7]将 CNN 网络和再识别模块联合,采用 64 个卷积层的全卷积网络实现红外图像的行人检测。上述方法虽然能有效完成行人检测,但是均采用多层卷积的网络对 ROI 特征进行分类,网络包含参数过多、容易过拟合,在环境较差的单帧红外图像行人检测中正确率相对较低。

基于前人的研究基础和上述文献的缺陷,本文提出一种基于 LeNet-7 的卷积神经网络对单帧红外图像行人检测的方法。引入自适应阈值的垂直和水平方向投影得到 ROI 的方法,将 ROI 图像输入本文构建的

LeNet-7 系统,该系统共包含 3 个卷积层、3 个池化层和 1 个输出层,每个卷积层所用卷积核大小和特征图个数并不单一固定,而是以错误率最小的实验选法,通过实验数据确定。采用波士顿大学建立的 BU-TIV (Thermal Infrared Video) Benchmark 热红外视频数据库训练本文系统,俄亥俄州立大学建立的 OTCBVS 和 Terravic Motion IR Database 红外数据库测试系统,与不同方法进行对比,本文方法更好地提高了检测率。

1 “方向投影”的 ROI 分割方法

一般而言,人体温度高于背景温度,人体在红外图像中体现为高灰度区域^[8]。然而,在城市环境中,车辆发动机、热水管和空调外机等非人体目标在红外图像中也体现为高灰度区域,因此仅靠目标灰度信息直接进行行人检测是不现实的。但是,与背景目标相比较,人体在红外图像中灰度仍然较高,通过搜索红外图像中的“热点区域”可以实现 ROI 的分割。因此本文采用一种“方向投影”的方法对人体可能存在的区域进行分割。

方向投影的基本方法是:首先,选择一定的阈值对红外图像进行分割,将分割后的图像向 X 轴作垂直投影,记录下像素点的数量,此时图像被分割为一系列垂直的亮度带,然后将亮度带向 Y 轴作水平投影,最后,得到高灰度区域的垂直位置与水平位置。具体的实现过程如图 1 所示。

第一步: 阈值的选择

由于不同场景中的红外图像的亮度分布不同,因此在设置阈值时应采取自适应的方式。本文采用的分

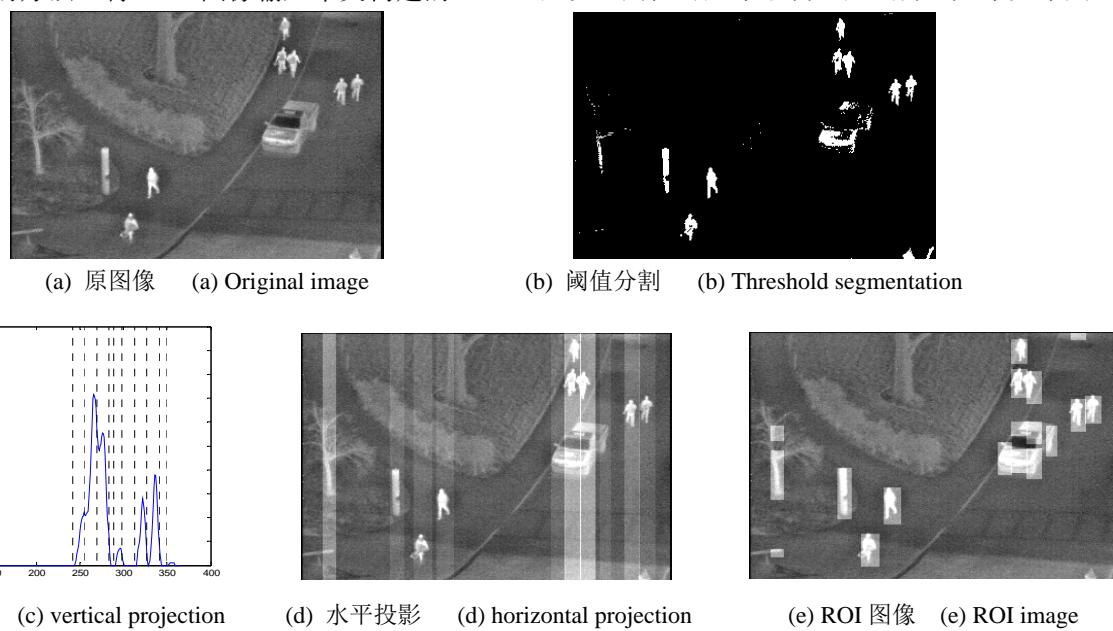


图 1 方向投影人体区域位置初定位 Fig.1 Preliminary location of human body region by directional projection

割阈值为：

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

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。将图像灰度最大值和图像灰度均值进行加权组合，可以增强分割阈值的自适应性。通过大量实验，本文最终将 ω 的取值定为 0.25。经过阈值分割以后，图像当中灰度较低的部分以及噪声基本被消除，如图 1(b) 所示。

第二步：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，曲线记录的是灰度值为 1 的像素点的数量，如图 1(c) 所示。由图 1(b) 可知，人体在图像中为高灰度区域，在图 1(c) 垂直投影曲线中表现为凸起的山峰，在投影曲线中凸起山峰的两侧分别寻找曲线的上升点与下降点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。

第三步：水平投影

将垂直投影得到的亮度带向 Y 轴做水平投影。与垂直投影相似，在水平投影曲线当中也能够得到一系列凸起的山峰，同样将每个山峰的上升点和下降点分别作为水平亮度带的起始点与结束点，即得到一系列平行于 X 轴的亮度带，结果如图 1(d) 所示。

第四步：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(e) 所示。这些矩形区域就是本文确定的 ROI，图 1(e) 中一些矩形区域包含人体目标，一些高灰度区域则包含非人体目标。

本文将确定的 ROI 作为卷积神经网络的输入，利用训练好的 CNN 网络进行二分类，从而检测出红外图像中的人体目标。因此接下来本文对 CNN 网络的建立进行探讨。

2 LeNet-7 网络的建立

2.1 CNN 网络的介绍

传统的人工神经网络结构中隐含层的神经元与前一层是全连通的，这意味着每一个神经元都与前一层的所有神经元存在参数关联，从而使得传统神经网络计算量大、速度慢，容易过拟合^[9]。而 CNN 是一个由单层卷积神经网络组成的多层次可训练监督学习网络，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）3 个阶段。若设每个神经元的输入为 x_i ，输出为 y_i ，每个神经元的输入和输出之间并非全连接，而是通过一定大小的区域相连接，同时 y_i 中的神经元通过一定大小的卷积核对这个区域进行特征提取，从而使得输入数据的权值数量大大降低（降维），提高了网络的训练速度，避免了过拟合^[10]。 y_i 和 x_i 之间的监督学习过程如式(2)所示：

$$y_j = \sum_i \omega_{ij} x_i + b_j \quad (2)$$

式中： ω_{ij} 是可训练的卷积核； b_j 是可训练的偏置参数。

2.2 LeNet-7 系统的介绍

CNN 最为经典的应用是 Y. Lecun 等提出的 LeNet-5 系统^[11]，该系统设计之初主要应用于手写数字识别，识别错误率仅达到 0.9%。由于红外图像采集环境复杂，单帧图像包含干扰目标（非行人目标）种类较多，本文将传统 LeNet-5 系统直接用于单帧红外图像行人检测，实验结果并不令人满意。因此本文对该系统进行了改进，提出 7 层网络：LeNet-7 系统，除去输入层，LeNet-7 系统仅仅包含 3 个卷积层、3 个池化层和 1 个输出层，具体结构如图 2 所示。图 2 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 分别为各个特征图的宽和高。如果是 LeNet-5 系统，则没有 C5 和 S6 层，S4 层直接全连接到输出层。由图 2 可知，影响 LeNet-7 系统的参数包含卷积核大小、特征图个数，目前对于二者的参数选择没有明确数学标准化，大多数依赖设计者的经验，本文采用错误率最小的试选法，通过实验数据分析，对系统的卷积核大小、特征图个数进行选取，使系统达到最佳检测效果。

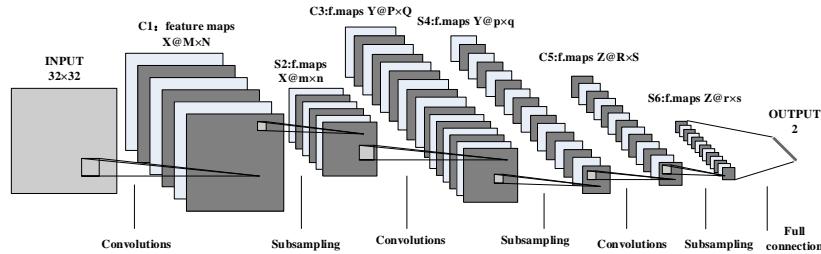


图 2 LeNet-7 系统结构

Fig.2 LeNet-7 System

2.3 LeNet-7 系统参数的确定

本文采用波士顿大学 2014 年提供的 BU-TIV (Thermal Infrared Video) Benchmark^[12]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。随机提取 3000 个行人样本, 2000 个用于训练, 1000 个用于测试; 随机提取 3000 个非行人样本中, 2000 个用于训练, 1000 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 8000 个 epoch (训练次数)。系统参数的选择, 采用错误率最小的试选法, 基本思想如下:

1) 在单层卷积网络中(图 2 中的 C1 层和 S2 层), 估计卷积核大小的范围, 固定某一个卷积核尺寸, 依次改变特征图个数, 通过实验数据分析, 选取该层最佳的卷积核尺寸和特征图个数。

2) 将上一层网络确定的参数作为 C3 层的输入, 联合 C1 层网络, 重复步骤 1), 选取 C3 层和 S4 层的最佳卷积核尺寸和特征图个数。

3) 将上一层网络确定的参数作为 C5 层的输入, 联合 C1、C3 层网络, 重复步骤 1), 选取 C5 层和 S6 层的最佳卷积核尺寸和特征图个数。

2.3.1 C1 层参数的确定

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, 因此本文选取特征图范围为 1~16。一般, 卷积核选取奇数, 由于行人在单帧红外图像中目标较小, 干扰目标较多, 为了保证检测的正确率, 卷积核不宜选取过大, 因此本文选取卷积核大小范围为 1×1 、 3×3 、 5×5 、 7×7 、 9×9 。采用检测错误率衡量网络提取特征的能力, 错误率的描述如公式(3)所示:

$$\text{错误率} = \frac{\text{ER}}{\text{TOTAL}} \times 100\% \quad (3)$$

式中: ER 为错误样本个数, TOTAL 为总样本个数。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。

本文首先构建单层卷积网络 (只保留图 2 中的 C1 层和 S2 层) 进行实验, 结果如表 1 所示。由于网络前向传播时特征图广度逐层减小, 后一卷积层中的卷积核大小不应大于上一卷积层中卷积核的大小, 因此, C1 层的卷积核选择 3×3 、 5×5 、 7×7 、 9×9 。由表 1 可知, 除个别情况外, 大部分不收敛 (错误率 0.5000) 情况出现在左下角呈阶梯状分布, 大部分收敛 (错误率较小) 情况出现在右上角。由此可知, 卷积核过小时, 增大特征图个数使得系统输出错误率过

大, 无法收敛; 卷积核过大时, 特征图个数过大, 也会使系统无法收敛。其中, 9×9 卷积核对应的 6 个特征图时, 单层网络的错误率最低, 因此, C1 层卷积层参数确定为 9×9 , 6 个特征图。一般, 该层卷积网络特征图尺寸为^[13]:

$$\text{map} = (n-k+1) \times (n-k+1) \quad (4)$$

式中: n 为该层网络的输入图像尺寸; k 为该层卷积核尺寸, 因此, C1 层 $M \times N = 24 \times 24$, 池化层 S2 中 $m \times n = 12 \times 12$ 。经过池化后系统的广度 (特征图的尺寸) 减小到原有的 $1/4$, 而数据的深度不变, 系统的参数个数减小到了 75%, 计算量大大降低, 池化层: S4 层、S6 层也是如此。

表 1 C1 层不同卷积核和特征图对应错误率

Table 1 Corresponding error rates of different convolution kernels and characteristic map in layer C1

Feature map	Convolution kernel			
	3×3	5×5	7×7	9×9
1	0.0327	0.0330	0.0420	0.0618
2	0.0260	0.5000	0.0473	0.0523
3	0.0420	0.0202	0.0395	0.0310
4	0.0422	0.0268	0.0415	0.0299
5	0.0289	0.0305	0.0411	0.0310
6	0.0431	0.3500	0.0353	0.0175
7	0.5000	0.0360	0.5000	0.0405
8	0.5000	0.5000	0.0377	0.0225
9	0.5000	0.0310	0.5000	0.5000
10	0.5000	0.5000	0.5000	0.0274
11	0.5000	0.5000	0.5000	0.5000
12	0.5000	0.5000	0.5000	0.0331
13	0.5000	0.5000	0.5000	0.5000
14	0.5000	0.5000	0.5000	0.5000
15	0.5000	0.5000	0.5000	0.5000
16	0.5000	0.5000	0.5000	0.5000

2.3.2 C3 层参数的确定

联合 C1 层网络, 再次通过实验试选法进行 C3 层参数的确定。其中, C3 层卷积核大小选取 1×1 、 3×3 、 5×5 、 7×7 、 9×9 。由表 1 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, 因此本次实验选取特征图范围为 1~10。实验结果如表 2 所示, 由表 2 可知, 在非单层卷积网络中, 特征图个数和卷积核大小并无规律可寻, 5×5 卷积核对应 5 个特征图时, 系统的错误率最低, 因此, C3 层卷积层参数确定为 5×5 , 5 个特征图, C3 层 $P \times Q = 8 \times 8$, 池化层 S4 中 $p \times q = 4 \times 4$ 。

表2 C3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 C3

Feature map	Convolution kernel				
	1×1	3×3	5×5	7×7	9×9
1	0.0455	0.0370	0.0331	0.0510	0.0500
2	0.0490	0.0330	0.0481	0.0510	0.0532
3	0.0520	0.0480	0.0405	0.0546	0.0332
4	0.0365	0.0355	0.0493	0.0340	0.0352
5	0.0380	0.0432	0.0196	0.0392	0.0330
6	0.0416	0.0236	0.0387	0.4020	0.0449
7	0.0411	0.0375	0.0309	0.0366	0.0562
8	0.0510	0.0410	0.0415	0.0510	0.0429
9	0.0380	0.0419	0.0521	0.0507	0.0389
10	0.0331	0.0500	0.0483	0.0428	0.0505

2.3.3 C5层参数的确定

联合C1层、C3层网络,最后通过实验试选法确定C5层参数。一般,后一层的卷积核尺寸不大于上一层卷积核尺寸,因此本次实验选取卷积核大小为1×1、3×3、5×5,特征图个数选取1~10,实验结果如表3所示。由表3可知,卷积核3×3时7个特征图对应的错误率最小,此时C5层 $R \times S = 2 \times 2$, $r \times s = 2 \times 2$ 。因此,C5层卷积核大小确定为3×3,特征图个数为7。最终,LeNet-7系统参数确定为:9×9、6个特征图;5×5、5个特征图;3×3、7个特征图。

表3 C5层不同卷积核和特征图对应错误率

Table 3 Corresponding error rates of different convolution kernels and characteristic map in layer C5

Feature map	Convolution kernel		
	1×1	3×3	5×5
1	0.0276	0.0264	0.0452
2	0.0256	0.0229	0.0505
3	0.0420	0.0310	0.0412
4	0.0360	0.0304	0.0406
5	0.0268	0.0338	0.0307
6	0.0305	0.0301	0.0392
7	0.0340	0.0207	0.0403
8	0.0450	0.0307	0.0501
9	0.0442	0.0394	0.0501
10	0.0502	0.0406	0.0472

3 LeNet-7系统应用于单帧红外图像

3.1 行人检测流程

通过第1、2章的探讨,本文建立了“方向投影”的ROI分割方法和7层的卷积神经网络,单帧红外图

像行人检测的具体方法为:将分割得到的ROI依次送入训练好的LeNet-7系统,经由系统二分类,最终输出检测率数值。具体流程如图3所示。

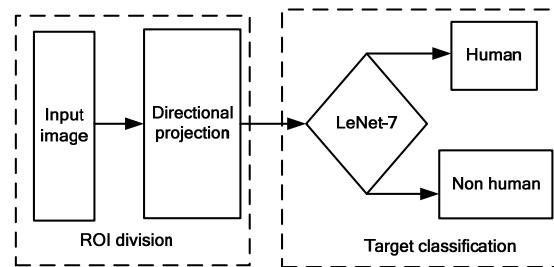


Fig.3 Pedestrian detection flow of single frame infrared image

需要指出的是:

1) 数据库的选择

为了避免测试时ROI中出现训练系统所用BU-TIV数据库中的人体目标,本文采用俄亥俄州立大学提供的OTCBVS Benchmark Dataset数据库^[14]和Terravic Motion IR Database数据库^[15]对LeNet-7网络进行测试。这样固然可以避免测试集和训练集相互重合,但是对深度卷积神经网络的场景迁移能力提出了挑战。

2) 人体目标的大小

本文训练样本来自BU-TIV(Thermal Infrared Video) Benchmark数据库,根据数据库提供的ground truth文件可以发现大部分人体目标大小为22×32,有一些特殊姿态的(如骑自行车或摩托车)的人体样本,给出的目标大小则为32×32。在制作训练样本集和测试样本集时,本文将所有的样本统一归一化为32×32。为了保证训练好的系统有效用于测试,本文将ROI全部归一化为32×32,以此保证测试集与训练集大小相同。

3.2 实验结果与分析

采用3个不同的红外图像测试集进行红外行人检测实验,测试集1来源于俄亥俄州立大学的OSU Thermal Pedestrian Database数据库,测试集2来源于OSU Color-Thermal Database数据库,测试集3来源于Terravic Motion IR Database数据库。其中,测试集1由23副图像组成,共含有101个人体目标,为多人体测试集;测试集2由54副图像组成,每幅图像含有一个目标,为单人体测试集;测试集3共有127幅携带武器的图像,每幅图像中含有2个人体目标,共含有254个人体目标。实验结果如图4所示。

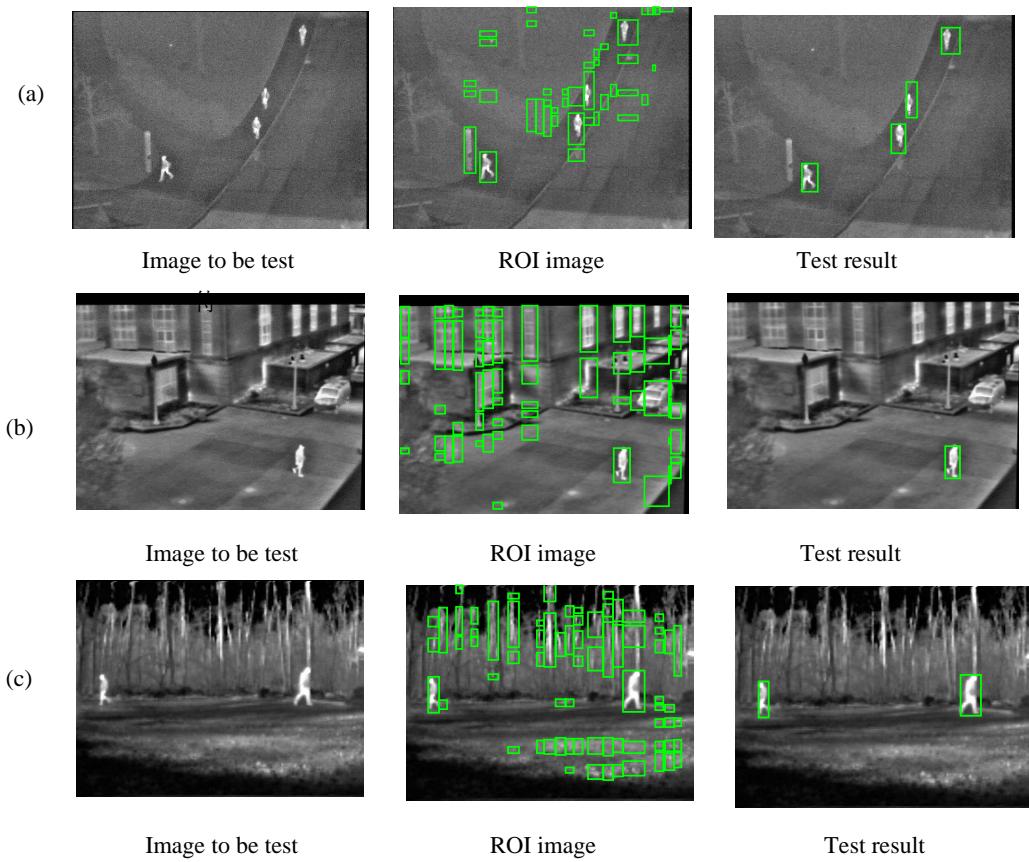


图4 不同测试集行人检测结果 (a)测试集1; (b)测试集2; (c)测试集3
Fig.4 Pedestrian test results of different test sets (a) test set 1; (b) test set 2; (c) test set 3

由图4可知,3个测试集的图像经过自适应阈值方向投影后得到的ROI中存在大量的非人体目标,经过本文构建的LeNet-7系统进行分类后,人体目标能够很好地被分离,这表明本文提出的LeNet-7系统对单帧红外图像的人体检测效果良好。为直观反映检测效果,采用检测率(accuracy rate, AR)和虚警率(false alarm rate, FAR)作为衡量指标,具体描述如下^[16]:

$$AR = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

$$FAR = \frac{FN}{TP + FN} \times 100\% \quad (6)$$

式(5)和式(6)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。将本文方法所得检测结果与传统“HOG+Fisher”(文献[17])方法、文献[6]方法对比,

结果如表4所示。其中,文献[17]对ROI的检索性能受限于手动设计的行人特征,而且滑动窗并未实现对于不同数据集的可伸缩性。文献[6]构建的卷积神经网络包含23个卷积层和5个池化层,参数多,计算复杂。

由表4可知,本文构建的LeNet-7系统在3个测试集中的检测率均高于其他方法,这表明基于LeNet-7系统的卷积神经网络针对行人检测系统具有很好的正确率、迁移性。在单人体测试集2中,本文方法检测率达到100%,同时没有产生虚警;测试集1和3中本文方法存在部分人体目标未被正确检测的问题,但虚警率为0%。经过分析,原因是没有正确检测的ROI中行人之间存在较为严重的相互遮挡,导致系统无法正确判断人体目标的个数。在以后的工作中,本文将进一步研究。

表4 不同方法的实验结果

Table 4 Experimental results of different methods

Test set	Methods	Pedestrian numbers	TP	FP	FN	AR	FAR
1	HOG+Fisher	101	96	5	1	95%	1%
	Paper[6]		99	2	1	98%	1%
	Proposed method		100	1	0	99%	0%
2	HOG+Fisher	54	53	1	0	98%	0%
	Paper[6]		54	0	0	100%	0%
	Proposed method		54	0	0	100%	0%
3	HOG+Fisher	254	249	5	0	98.0%	0%
	Paper[6]		251	3	0	98.8%	0%
	Proposed method		252	2	0	99.2%	0%

4 结论

本文提出一种基于 LeNet-7 的卷积神经网络对单帧红外图像行人检测的方法。首先，采用自适应阈值分割图像，然后将分割后的图像向 X 轴方向投影，再向 Y 轴投影，最后将得到的 ROI 图像输入训练好的 LeNet-7 系统进行测试。该系统仅包含 3 个卷积层，避免了全卷积神经网络参数多、计算量大等问题，每层的参数以错误率最小的试选法进行确定，避免了依赖经验选择参数的盲目性。其中，训练系统的 6000 个随机样本来自于 BU-TIV 数据库，测试样本来自于 OTCBVS 和 Terravic Motion IR Database 数据库。3 个不同测试集的实验结果表明，本文方法具有很高的红外图像行人检测率，与传统“HOG+Fisher”方法、采用多个卷积层的神经网络对比，本文算法的检测率更高，虚警率更低，具有较高的实用性。然而，本文方法还有进一步改进的空间，当 ROI 图像中行人之间存在较为严重的相互遮挡时，系统无法正确判断人体目标，在接下来的工作中，将进一步提高行人遮挡图像 ROI 的分割精度，提升系统的识别能力，为更深层、更复杂的序列红外图像行人检测作铺垫。

参考文献:

- [1] Nanda H , Davis L. Probabilistic template based pedestrian detection in infrared videos[C]//Intelligent Vehicle Symposium, IEEE, 2002: 7712599.
- [2] Bertozzi M, Broggi A, Grisleri P, et al. Pedestrian detection in infrared images[C]//IEEE Intelligent Vehicles Symposium, IEEE, 2003: 7883392.
- [3] GAO Y , AI X , WANG Y , et al. U-V-Disparity based Obstacle Detection with 3D Camera and steerable filter[C]//Proc Intl Intelligent Vehicle Symposium, IEEE, 2011: 12095161.
- [4] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [5] 许茗, 于晓升, 陈东岳, 等. 复杂热红外监控场景下行人检测[J]. 中国图象图形学报, 2018, 23(12): 1829-1837.
XU M, YU X S, CHEN D Y, et al. Man detection in complex thermal infrared monitoring scenes[J]. *Chinese Journal of Image and Graphics*, 2018, 23(12): 1829-1837.
- [6] 谭康霞, 平鹏, 秦文虎. 基于 YOLO 模型的红外图像行人检测方法[J]. 激光与红外, 2018, 48(11): 1436-1442.
TAN K X, PING P, QIN W H. Infrared image pedestrian detection method based on YOLO model[J]. *Laser & Infrared*, 2018, 48(11): 1436-1442.
- [7] 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.
- [8] 刘智嘉, 贾鹏, 夏寅辉, 等. 基于红外与可见光图像融合技术发展与性能评价[J]. 激光与红外, 2019, 49(5): 633-640.
LIU Z J, JIA P, XIA Y H, et al. Development and performance evaluation of infrared and visible image fusion technology[J]. *Laser & Infrared*, 2019, 49(5): 633-640.
- [9] 吴志洋, 卓勇, 李军, 等. 基于卷积神经网络的单色布匹瑕疵快速检测算法[J]. 计算机辅助设计与图形学学报, 2018, 30(12): 2262-2270.
WU Z Y, ZHUO Y, LI J, et al. Fast detection algorithm of monochrome fabric defects based on convolution neural network[J]. *Journal of Computer Aided Design & Graphics*, 2018, 30(12): 2262-2270.
- [10] 欧攀, 张正, 路奎, 等. 基于卷积神经网络的遥感图像目标检测[J]. 激光与光电子学进展, 2019, 56(5): 74-80.
OU P, ZHANG Z, LU K, et al. Remote sensing image target detection based on convolution neural network[J]. *Laser & Optoelectronics Progress*, 2019, 56(5): 74-80.
- [11] Y. Lecun, L. Bottou, Y. Bengi, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [12] ZHENG Wu, Nathan Fuller, Diane Theriault, 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>.
- [13] 吕永标, 赵建伟, 曹飞龙. 基于复合卷积神经网络的图像去噪算法[J]. 模式识别与人工智能, 2017, 30(2): 97-105.
LU Y B, ZHAO J W, CAO F L. Image denoising algorithm based on compound convolution neural network[J]. *Pattern Recognition & Artificial Intelligence*, 2017, 30(2): 97-105.
- [14] Riad I. Hammoud. OTCBVS Benchmark Dataset Collection[DB/OL]. (2014-6-22)[2019-12-18]. <http://vcipl-okstate.org/pbvs/bench/>.
- [15] Riad I. Hammoud. Terravic Motion IR Database[DB/OL]. (2014-6-22) [2019-12-18]. <http://vcipl-okstate.org/pbvs/bench/Data/05/download.html>.
- [16] 苏育挺, 陈耀, 吕卫. 基于近红外图像的嵌入式人员在岗检测系统[J]. 红外技术, 2019, 41(4): 377-382.
SU Y T, CHEN Y, LU W. Embedded on-the-job detection system based on near infrared image[J]. *Infrared Technology*, 2019, 41(4): 377-382.
- [17] 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, 2014: 937-940.