

基于改进卷积网络的航拍行人检测

贵向泉¹, 蒋家俊¹⁺, 李立¹, 高磊²

(1. 兰州理工大学 计算机与通信学院, 甘肃 兰州 730050;

2. 兰州三维大数据标准化研究院有限公司 大数据研发中心, 甘肃 兰州 730050)

摘要: 为对航拍图像下的行人进行检测, 提出一种改进的卷积神经网络模型。通过预处理航拍图像, 采用中值滤波和对比度增强方法提升检测有效性; 使用改进的卷积神经网络模型中的克罗内克积对航拍图像上采样, 将原图像在保证清晰度的情况下, 长和宽放大4倍; 提出分块区域提取算法, 对放大图像的每个图像块检测行人。实验结果表明, 该方法在航拍图像上的行人检测有效。

关键词: 卷积网络; 行人检测; 航拍; 分块; 克罗内克积

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1000-7024(2018)09-2947-05

doi: 10.16208/j.issn1000-7024.2018.09.043

Aerial pedestrian detection based on improved convolution network

GUI Xiang-quan¹, JIANG Jia-jun¹⁺, LI Li¹, GAO Lei²

(1. School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China;

2. Large Data Research and Development Center, Lanzhou Three Dimension Big Data Standardization Research Institute Limited Company, Lanzhou 730050, China)

Abstract: To detect the pedestrian in the aerial image, an improved convolution neural network model was presented. The aerial image was pre-processed, and the effectiveness of the detection was improved using the median filter and the contrast enhancement method. The Kronecker product of the improved convolution network model was used to up-sample the aerial image, and the original image was enlarged 4 times in length and width under the condition of ensuring clarity. A block region extraction algorithm was proposed to detect pedestrians for each image block of the enlarged image. Experimental results show that the proposed method is effective in pedestrian detection in aerial images.

Key words: convolution network; pedestrian detection; aerial images; block; Kronecker product

0 引言

近年来, 行人检测的技术不断趋于成熟。行人检测方法众多, 大致分为3类: 第一类采用原始特征如HOG、SIFT等算子结合SVM分类器检测行人^[1]; 第二类是基于Adboost级联分类器的行人检测^[2]; 第三类是用神经网络和深度学习技术对行人进行检测, 效果较人工提取特征的方式有很大改善^[3]。然而, 图像中的小目标行人检测问题依然未被很好解决^[4,5]。尤其在航拍图像中对行人的检测仍

是亟待解决的问题。

同时, 行人检测研究中视角的种类各异, 大致有3类: 一是平视下的行人检测, 二是监控式斜上角的行人检测^[6], 三是唐春辉教授所采用的俯视行人检测方法^[7-10]。这些行人检测研究多应用于目标分类、安防视频监控、车辆辅助驾驶或无人车驾驶等场合^[11,12]。在实际应用中, 航拍影像极大弥补了地面传统测量设备覆盖范围小、数据获取受限不足等情况^[13,14]。因此, 航拍图像下的行人检测应用前景广泛。

收稿日期: 2017-07-06; 修订日期: 2017-09-15

基金项目: 国家科技支撑计划课题基金项目(2012BAF12B19); 甘肃省计算中心爱客家开放基金项目(1505JTCA007); 兰州理工大学红柳青年教师培养计划基金项目(Q201312)

作者简介: 贵向泉(1981-), 男, 甘肃定西人, 博士, 副教授, CCF会员, 研究方向为复杂网络理论和应用、大数据处理分析; +通讯作者: 蒋家俊(1992-), 男, 江苏宜兴人, 硕士研究生, CCF学生会员, 研究方向为模式识别与图像处理、机器学习; 李立(1983-), 女, 甘肃天水人, 硕士, 讲师, 研究方向为模式识别与图像处理; 高磊(1983-), 男, 甘肃兰州人, 硕士研究生, 研究方向为大数据处理。
E-mail: 15961578209@139.com

由于航拍图像在空间上有更广阔的地理信息，因此检测范围更大，适合对大范围区域监测与管理。航拍图像中的检测目标较小，导致原有行人检测方法对行人的检测效果并不理想。通过采用改进卷积网络，解决了航拍行人的小目标检测问题。使用的航拍数据主要视角是距地面 50 m 到 70 m 范围的高空俯视地表环境，以检测航拍图像中的行人。

1 主要方法

通过提出了一种改进的卷积网络模型对航拍图像下的行人进行检测。该模型先预处理图像，使用中值滤波去除噪声，使用直方图均衡化方法提升彩色图像的颜色对比度差异，便于检测。基于文献 [15] 中的 YOLO (you only look once) 模型，在其特征提取层前面添加克罗内克上采样层与分块提取层，并对模型后端参数改进，实现了在航拍图像中对行人提取特征再检测的效果。最终，实验验证了该方法在航拍图像中的行人检测效果。具体模型流程如图 1 所示。

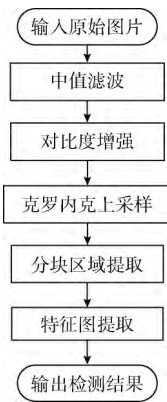


图 1 实验整体流程

1.1 图像预处理

1.1.1 中值滤波

由于高空下无人机航空拍摄得到的图像可能含有噪声，为了有利于检测行人，消除噪声，同时满足滤波处理后不能损坏图像轮廓及边缘等重要信息的情况下，实时采集的航拍图像需要采用中值滤波来处理，使得图像清晰，视觉效果良好。中值滤波将图像的每个像素用邻域像素的中值代替，公式如式 (1) 所示

$$y(i) = \text{med}[x(i-n), \dots, x(i), x(i+n)] \quad (1)$$

其中， $\text{med}[\cdot]$ 函数用于在所有值中选取中位数。当滤波核尺寸为 3×3 时，中值滤波即在图像 3×3 区域的像素中寻找中值。

1.1.2 对比度增强

直方图均衡化是采用拉伸像素强度分布范围的方法来增强图像对比度，它把一个分布映射到另一个更宽的分布

中，所以图像的对比度分布会在整个范围内展开。通过对滤波后的彩色图像在 R、G、B 这 3 个通道上采用直方图均衡化，得到每个通道输出再融合，实现了彩色图像对比度增强，使行人的轮廓清晰可见。

实现均衡化的映射函数是累积分布函数。对于直方图 $h(i)$ ，它的累计分布函数 $H(i)$ 为

$$H(i) = \sum_{0 \leq j < i} h(j) \quad (2)$$

当把 $H(i)$ 作为映射函数，需要对最大值的累积分布 $H(i)$ 归一化。使用一个简单的映射过程来获得均衡化后像素的强度值，具体公式如下所示

$$\text{equ}(x, y) = H(\text{src}(x, y)) \quad (3)$$

将彩色图像三通道灰度阈值都拉伸到 0-255 的范围中，再次合并三通道构成的彩色图像对比度有所提升，结果如图 2 所示。

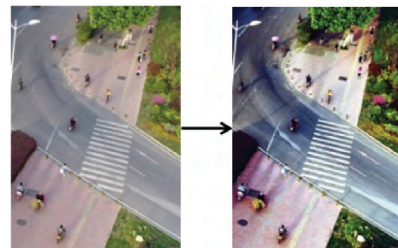


图 2 对比度增强

1.2 改进的卷积神经网络

在 YOLO 模型 [15] 的特征图卷积提取层前，先用克罗内克上采样层将原图像放大。因航拍高度为 50 m 至 70 m，放大倍数过小会难以检测出行人，放大倍数过大则导致轮廓信息不明显，行人块区域过大而难以检测。通过实验测试效果，选择对原图像放大 4 倍。

分块区域提取层是按照算法将原图像裁剪成大小相同的 16 份，每一张图像尺寸大小和原图像相等，这样方便了对每块图像的处理。循环对图像块特征提取以检测图像块内的行人。具体检测的流程如图 3 所示。

具体步骤如下所示：

(1) 克罗内克积上采样

如果 A 是原始图像矩阵， B 是自适应克罗内克积的核，则 $A * B$ 这个更高阶图像矩阵可以视为更高分辨率下的观察。因此，对给定航拍图像 A ，构造克罗内克核 B ，以生成高分辨率图像 $A * B$ ，实现对图像 A 的放大。

设克罗内克核 B 大小为 4×4 ，其可以表示为

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \end{bmatrix}$$

对分辨率为 $N * M$ 的图像 A 中每个像素 $a_{i,j} (i = 1, 2,$

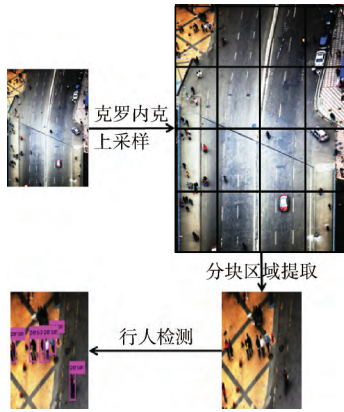


图 3 卷积网络检测流程

```
for(l=1;l<5;l++){
    for(i=1;i<height/4;i++){
        for(j=1;j<width/4;j++){
            B[i][j]=A[4*k+i][4*l+j]
        }
    }
}
```

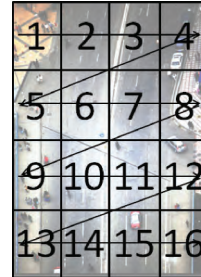


图 4 分块区域提取

..., N; j = 1, 2, ..., M), 构造克罗内克核 B 的方法为:

首先令 $b_{11} = 1$, 其次比较矩阵

$$a_{i,j} = \begin{bmatrix} a_{i,j} & a_{i,j+1} & a_{i,j+2} & a_{i,j+3} \\ a_{i+1,j} & a_{i+1,j+1} & a_{i+1,j+2} & a_{i+1,j+3} \\ a_{i+2,j} & a_{i+2,j+1} & a_{i+2,j+2} & a_{i+2,j+3} \\ a_{i+3,j} & a_{i+3,j+1} & a_{i+3,j+2} & a_{i+3,j+3} \end{bmatrix}$$

其中, 各元素 $a_{u,v}$ ($u \neq v$ 且 $v \neq j$) 与 $a_{i,j}$ 的关系: 如果 $a_{u,v} < a_{i,j}$, 取卷积核 B 中与 $a_{u,v}$ 位置对应元素 $b_{m,n} = 1 - \omega * \text{rand}(1)$; 如果 $a_{u,v} > a_{i,j}$, 取 $b_{m,n} = 1 + \omega * \text{rand}(1)$; 如果 $a_{u,v} = a_{i,j}$, 则 $b_{m,n} = 1$ 。这里 $\text{rand}(1)$ 表示在 $[0, 1]$ 上的高斯分布函数随机值, ω 是正常数, 取值为 0.05。使用该方法将原图像的长度和宽度都扩大了 4 倍, 但图像清晰度并没有太大的损失。

能量梯度函数可以对图像进行清晰度评价, 该函数定义如下

$$D(f) = \sum_y \sum_x (|f(x+1,y) - f(x,y)|^2 + |f(x,y+1) - f(x,y)|^2) \quad (4)$$

其中, $f(x, y)$ 表示图像 f 对应像素点 (x, y) 的灰度值, $D(f)$ 为图像清晰度计算结果。画面越模糊, 则 $D(f)$ 值越小。在原始数据集中采样计算, 得到 $D(f)$ 的均值为 935, 上采样图像的 $D(f)$ 均值为 783。将原始图像简单放大相同倍数的 $D(f)$ 均值为 437。说明了通过克罗内克积上采样既放大了图像以便于行人检测, 又对图像的清晰度未造成过多的损失。

(2) 分块区域提取

分块区域提取层对上采样的图像等分裁剪, 并以图像块为单位组成图片矩阵, 将每块图片提取到检测模型的输入层, 其中 k 和 l 表示图像块所在图像位置的行号和列号, i 和 j 代表每个图像块中的像素所在位置。算法将图像矩阵 $A * B$ 分成 16 份, 依次输入到检测层提取检测, 具体如图 4 所示。

分块区域提取算法如下:

```
for(k=1;k<5;k++){
```

(3) 特征图提取与检测

卷积层作为深度学习提取特征的一种方法, 能够从数据中自动学习权重。将图像中某个区域与卷积核对应相乘得到卷积后的特征图, 该方法已成功用于物体分类、检测和识别等应用。卷积神经网络是一种多层有监督学习的复杂深度网络, 网络中需要由输入层输入原始图像, 经卷积层操作并下采样得到输出。学习算法采用误差反向传播优化网络结构。卷积层特征提取操作的公式定义为

$$\alpha'_j = f\left(\sum_k \omega'_{jk} \alpha_k^{l-1} + b'_j\right) \quad (5)$$

其中, $\sum_k \omega'_{jk} \alpha_k^{l-1} + b'_j$ 是矩阵操作, 对应相乘后相加, 并加上一个偏置值 b'_j , 通过激活函数 f 得到输出结果; ω'_{jk} 表示第 $l-1$ 层的第 k 个神经元连接到第 l 层的第 j 个神经元的权重; b'_j 表示第 l 层的第 j 个神经元的偏置; α'_j 表示第 l 层的第 j 个神经元的输出。

改进模型采用了 YOLO 模型作为模型后端, 对上采样后的分块图像卷积检测出其中的行人。特征图提取检测的模型段部分为了更好地适应对行人的检测, 修改了分类的参数, 将原有检测类别从 20 类变为 1 类以专门检测行人, 并将过滤层大小设为 30, 并将原模型的阈值设为 0.3, 优化了检测结果。

2 实验数据与结果分析

2.1 样本采集

实验样本采集于市内的不同路口, 使用无人机在高度为 50 m 至 70 m 的范围内拍摄, 示例如图 5 所示。这些图像普遍具有空间范围广阔, 行人尺寸小的特点。图像中的行人往往只有一个很小的轮廓, 且行人姿势不一, 具体有: 直立行人, 骑自行车行人, 穿雨具行人, 以及拎袋行人等。行人衣着颜色各不相同, 有些衣着颜色与背景相融合。在这些图像中, 颜色与背景相融合的行人会使检测效果的召回率降低。



图 5 原始图像

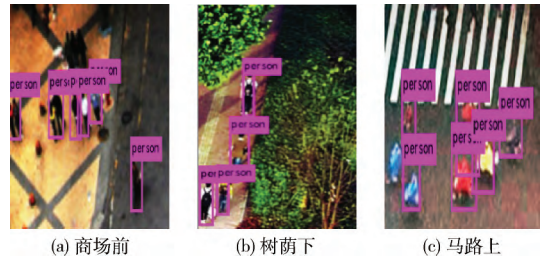


图 7 分块区域提取检测结果

2.2 图像预处理

2.2.1 中值滤波

滤波后，图像中的噪声都被消除，图像更清晰。对图像中需要检测的行人来说，噪声的消除使检测精度更高。

2.2.2 提升图像对比度

通过对彩色图像增强 R、G、B 三通道的对比度，使航拍图像中行人和背景的对对比度增大，轮廓信息更突出，图像色调更分明，更易于检测图像中的行人。

2.3 实验结果

2.3.1 克罗内克上采样图像

改进模型使用克罗内克积对原图上采样处理，得到长和宽都放大了 4 倍的图像。对比简单放大图像中某个区域的图像质量效果，该方法明显提高了清晰度。结果如图 6 (a) 与图 6 (b) 所示。

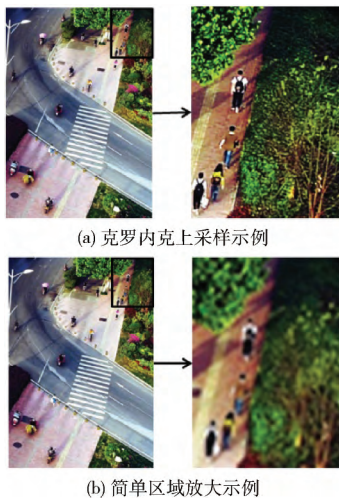


图 6 上采样与简单放大效果对比

2.3.2 分块检测结果

对航拍图像分块区域提取并对图像块逐个人检测。对比 YOLO 和 SSD(single shot multibox detector)^[16] 模型的检测结果，最终说明改进模型的精度和召回率都有所提升。对比效果如图 7 与图 8 所示。

图 7 中 3 幅图都是经克罗内克上采样后分块提取的部分图像块，尺寸与原图相同。通过对区域提取图像块中的

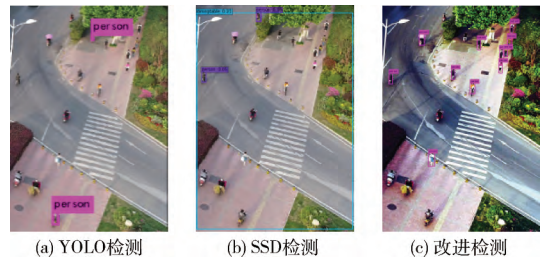


图 8 不同模型检测结果对比

行人进行检测，如图 7 (a) 所示，验证了该方法对衣着和背景有明显差异的行人检测有效性。图 7 (b) 中，树荫下的行人都能被检测。图 7 (c) 中，大部分穿雨具骑行的行人都能被检测，说明改进模型对多姿态行人检测的鲁棒性效果较好。

图 8 针对某个路口分别采用不同的模型检测，如图 8 (a) 所示，因行人尺度小，采用 YOLO 模型检测效果不理想；图 8 (b) 中采用 SSD 也很难检测到行人；图 8 (c) 中，将图像块的检测结果重新拼合成一张该路口的完整图像，采用该改进模型已能检测到图像中大多数行人。

2.4 精度和召回率分析

精度 P 反映了航拍图像中检测目标是行人的比重。具体如式 (6) 所示

$$P = \frac{TP}{TP + FP} \tag{6}$$

召回率 R 反映了航拍图像中被检测到的行人占行人总数的比重，具体如式 (7) 所示

$$R = \frac{TP}{TP + FN} = 1 - \frac{FN}{T} \tag{7}$$

其中， TP 表示检测出的目标是行人的数量， FP 表示检测出的目标不是行人的数量， FN 表示未被检测出的行人数量， T 表示真实行人的总数。

目前没有标准的航拍行人数据库，通过在市内不同路口，使用无人机拍摄了 5000 张图像实验，并与 YOLO 和 SSD 方法的实验效果进行了对比，验证了该方法的有效性，结果见表 1。

由表 1 可知，因 SSD 能在多尺度上检测行人，效果优于 YOLO 模型。而基于 YOLO 的改进模型使检测到的小目

表 1 实验结果

方法	总人数	检测值	漏检值	误检值	精度/%	召回率/%
YOLO ^[15]	27 263	4253	23 872	862	79.73	12.44
SSD ^[16]	27 263	11 462	12 160	1549	86.49	36.36
改进模型	27 263	22 478	7386	2541	88.70	73.13

标行人数量显著增多。

3 结束语

首先介绍了航拍图像中行人检测的应用场景, 给出了利用改进的卷积网络模型对航拍图像下行人检测的流程。采用滤波与彩图对比度增强的预处理操作使行人易于检测。使用改进的卷积网络对图像先克罗内克上采样, 并提出新的分块区域提取算法, 对上采样后的每个图像块使用改进参数的 YOLO 模型进行区域特征提取, 以提取出有效的行人特征。实验结果表明, 改进模型使航拍图像下行人检测的精度有所提高, 召回率有明显提升。但该模型针对接近于垂直角度拍摄的行人图像, 检测效果还有待提升, 将在后续研究中进一步完善。

参考文献:

- [1] YUE Xin. Pedestrian detection based on HOG feature and SVM classifier [J]. Innovation and Application of Science and Technology, 2016 (5): 62-63 (in Chinese). [岳鑫. 基于 HOG 特征和 SVM 分类器的行人检测研究 [J]. 科技创新与应用, 2016 (5): 62-63.]
- [2] JIANG Ailian, YANG Xingtong. Pedestrian detection based on AdaBoost-SVM cascade classifier [J]. Computer Engineering and Design, 2013, 34 (7): 2547-2550 (in Chinese). [蒋爱莲, 杨兴彤. 基于 AdaBoost-SVM 级联分类器的行人检测 [J]. 计算机工程与设计, 2013, 34 (7): 2547-2550.]
- [3] WANG Bin. Pedestrian detection based on deep learning [D]. Beijing: Beijing Jiaotong University, 2015 (in Chinese). [王斌. 基于深度学习的行人检测 [D]. 北京: 北京交通大学, 2015.]
- [4] ZHANG Chunfeng, SONG Jiatao, WANG Wanliang. Survey of pedestrian detection technology [J]. Television Technology, 2014, 38 (3): 157-162 (in Chinese). [张春风, 宋加涛, 王万良. 行人检测技术研究综述 [J]. 电视技术, 2014, 38 (3): 157-162.]
- [5] WANG Genling. Research on some key technologies of pedestrian detection and tracking [D]. Zhejiang: Zhejiang University of Technology, 2014 (in Chinese). [王根岭. 行人检测与跟踪中若干关键技术的研究 [D]. 浙江: 浙江理工大学, 2014.]
- [6] CHEN Huijie. Research on pedestrian detection and tracking in video surveillance [D]. Xiamen: Xiamen University, 2014 (in Chinese). [陈慧杰. 视频监控中的行人检测与跟踪方法研究 [D]. 厦门: 厦门大学, 2014.]
- [7] TANG Chunhui. Monocular blind pedestrian detection based on multi feature fusion [J]. Journal of systems simulation, 2016, 28 (9): 2146-2153 (in Chinese). [唐春晖. 一种基于多特征融合的单目俯视行人检测 [J]. 系统仿真学报, 2016, 28 (9): 2146-2153.]
- [8] TANG Chunhui. A method of pedestrian detection based on gradient orientation histogram [J]. Pattern Recognition and Artificial Intelligence, 2015, 28 (1): 19-26 (in Chinese). [唐春晖. 一种基于梯度方向直方图的俯视行人的检测方法 [J]. 模式识别与人工智能, 2015, 28 (1): 19-26.]
- [9] TANG Chunhui, CHEN Qijun. A multi-target detection and tracking method for pedestrians [J]. Journal of systems simulation, 2013, 25 (10): 2464-2467 (in Chinese). [唐春晖, 陈启军. 一种俯视行人的多目标检测和跟踪方法 [J]. 系统仿真学报, 2013, 25 (10): 2464-2467.]
- [10] TANG Chunhui, CHEN Qijun. A rapid method for pedestrian detection [J]. Journal of Systems Simulation, 2012, 24 (9): 253-256 (in Chinese). [唐春晖, 陈启军. 一种快速的俯视行人检测方法 [J]. 系统仿真学报, 2012, 24 (9): 253-256.]
- [11] GAO Chenxu. Pedestrian detection technology in monitoring video and its application [D]. Changsha: National University of Defense Technology, 2014 (in Chinese). [高晨旭. 监控视频中的行人检测技术及其应用 [D]. 长沙: 国防科学技术大学, 2014.]
- [12] CHE Zhifu, MIAO Zhenjiang, WANG Mengsi. Research and application of pedestrian detection in subway video surveillance system [J]. Modern Urban Rail Traffic, 2010 (2): 31-33 (in Chinese). [车志富, 苗振江, 王梦思. 地铁视频监控系统中的行人检测研究与应用 [J]. 现代城市轨道交通, 2010 (2): 31-33.]
- [13] TONG Guofeng, LI Yong, DING Weili, et al. Survey of remote sensing image change detection algorithms [J]. Chinese Journal of Image and Graphics, 2015, 20 (12): 1561-1571 (in Chinese). [佟国峰, 李勇, 丁伟利, 等. 遥感影像变化检测算法综述 [J]. 中国图象图形学报, 2015, 20 (12): 1561-1571.]
- [14] FENG Peng, GAO Feng. Survey of road information extraction methods in high resolution remote sensing images [J]. Modern Electronic Technology, 2015, 38 (17): 53-57 (in Chinese). [冯鹏, 高峰. 高分遥感影像中道路信息提取方法综述 [J]. 现代电子技术, 2015, 38 (17): 53-57.]
- [15] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection [C] //Computer Vision and Pattern Recognition. IEEE, 2016: 779-788.
- [16] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot MultiBox detector [C] //European Conference on Computer Vision. Springer, 2016: 21-37.