

文章编号: 2095-2163(2022)03-0123-05

中图分类号: TP391.4

文献标志码: A

基于 YOLOv4-Efficient 的目标检测和识别

史健婷¹, 李旭^{1,2}, 刘文斌¹, 安祥泽^{1,2}

(1 黑龙江科技大学 计算机与信息工程学院, 哈尔滨 150022; 2 黑龙江科技大学 研究生学院, 哈尔滨 150022)

摘要: 在自动驾驶场景中, 能够实现对前方车辆准确和及时的识别是至关重要的。本文对 YOLOv4 的主干网络进行修改, 形成了 YOLOv4-Efficient 目标检测算法, 在保证检测准确率的情况下, 提高了检测速度。通过使用一种轻量级算法 EfficientNet 代替 YOLOv4 的主干提取网络 CSPDarknet53。改进后的模型参数量少了近 4 倍。经实验对比, YOLOv4-Efficient 平均准确率达到了 96.27%, 参数量也仅仅有 60.52 M。结果表明, 改进后的算法有较好的识别能力和检测速度, 更适合移植到性能一般的设备上, 整体性能优于其他算法。

关键词: 自动驾驶; YOLOv4; 目标检测; 轻量级

Target detection and recognition based on YOLOv4-Efficient

SHI Jianting¹, LI Xu^{1,2}, LIU Wenbin¹, AN Xiangze^{1,2}

(1 Department of Computer and Information Engineering, Heilongjiang University of Science and Technology, Harbin 150022, China; 2 Graduate School, Heilongjiang University of Science and Technology, Harbin 150022, China)

[Abstract] In the automatic driving scene, it is very important to realize the accurate and timely identification of the vehicle in front. In this paper, the backbone network of YOLOv4 is modified to form a YOLOv4-Efficient target detection algorithm, which improves the detection speed under the condition of ensuring the detection accuracy. By using a lightweight algorithm EfficientNet to replace the backbone of YOLOv4, the network CSPDarknet53 is extracted. The parameters of the improved model are nearly four times less. Through experimental comparison, the average accuracy of YOLOv4-Efficient is 96.27%, and the parameter quantity is only 60.52 M. The results show that the improved algorithm has better recognition ability and detection speed, is more suitable for transplantation to devices with general performance, and the overall performance is better than other algorithms.

[Key words] automatic driving; YOLOv4; object detection; lightweight

0 引言

目标检测是计算机视觉和数字图像处理的一个热门方向, 广泛应用于视频监控、自动驾驶、航空航天等诸多领域。近年来目标检测成为研究和应用的主攻领域, 再加上深度学习的广泛运用, 目标检测得到了更加快速的发展。

目前主流的目标检测算法大概可以分为两大类:

(1) One-Stage 目标检测算法, 这类算法不需要候选区域阶段, 直接对物体进行检测和识别。YOLO 系列就是一阶段的代表性算法, 速度较快, 但精确度较差;

(2) Two-Stage 目标检测算法, 这类算法主要分为两个阶段, 第一阶段是产生候选区域, 也就是目标的大概位置; 第二阶段对所产生的候选区域进行类

别和位置信息的预测识别。R-CNN^[1] 是两阶段的代表性算法, 虽速度较慢, 但有较高的精度。两类算法各有优势, 随着研究的不断深入, 两类算法都在做着改进, 并能在速度和准确度上取得很好的表现。

随着网络的深度不断加深, 为了获得更多的目标信息, 势必要扩大网络的通道数, 这样就会带来大量的参数运算。本文将 EfficientNet 网络和 YOLOv4^[2-3] 中的检测网络相结合, 以达到参数量的大量减少, 检测速度尽可能提高的目标检测算法。

1 相关算法

1.1 YOLOv4 算法

YOLOv4 是 YOLOv3^[4] 的改进版, 在其基础上融合了许多创新性的小技巧。YOLOv4 可以分为 4 部分, 其中包括: 输入端、BackBone 主干网络、Neck、Prediction。

基金项目: 黑龙江省 2020 年大学生创新创业训练计划项目(202010219006); 2020 年度黑龙江省省属本科高校基本科研业务费项目(YJSCX2020-212HKD)的阶段成果。

作者简介: 史健婷(1981-), 女, 博士, 副教授, 主要研究方向: 图像处理与模式识别、多维数据分析与可视化。

收稿日期: 2021-10-29

输入端的创新,主要是训练时对输入端的改进,包括 Mosaic 数据增强等;BackBone 主干网络是将各种创新结合起来,其中包括: CSPDarknet53、Mish 激活函数、Dropblock 等;Neck 部分是在 Backbone 和最

后的输出层之间插入一些层,如 YOLOv4 中的 SPP 模块、FPN+PAN 结构;Prediction 部分主要改进训练时的损失函数 $CIOU_Loss$ 。YOLOv4 的网络结构如图 1 所示。

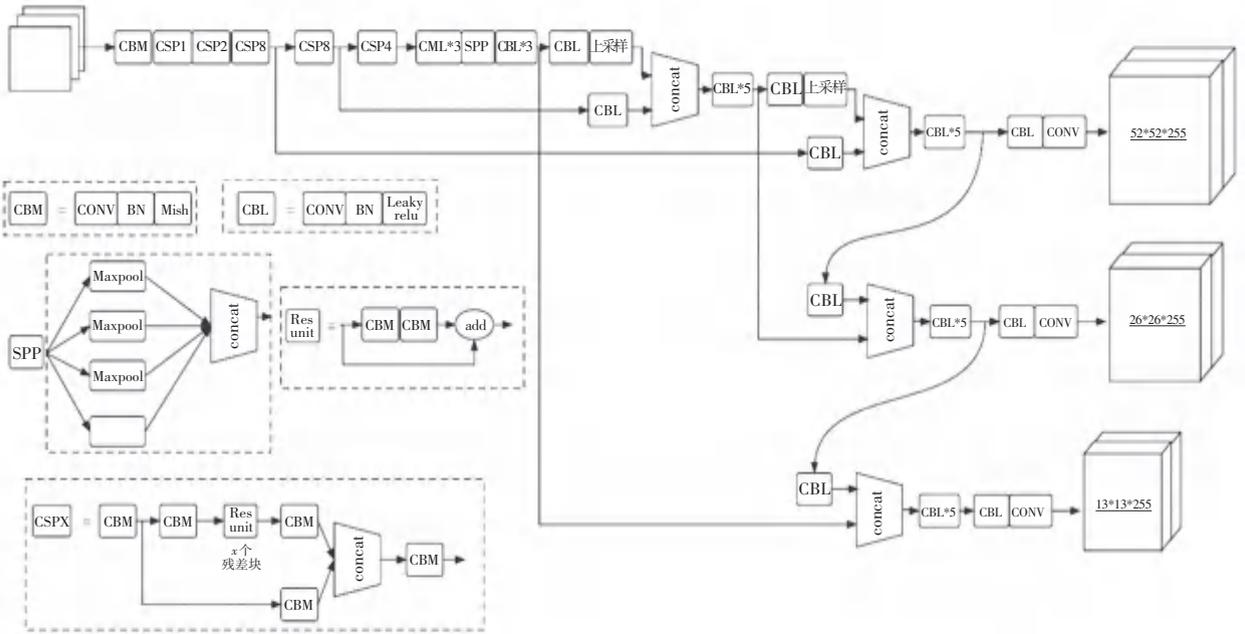


图 1 YOLOv4 网络结构图

Fig. 1 YOLOv4 network structure diagram

若以输入的图像尺寸为 416×416 为例,其处理过程如下:

图像进行 Mosaic 数据增强后,主干提取网络 CSPDarknet53^[5]进行特征提取,将提取到的特征进行 SPP 3 次不同尺度的最大池化,并通过 PANet 进行特征融合;将不同尺度的特征图进行融合,获得 3 种尺度的特征图 (52×52 、 26×26 、 13×13);最后分别对这 3 种不同尺度的特征图进行分类回归预测结果。

1.2 EfficientNet 算法

EfficientNet^[6-8]算法具有的特点:利用残差结构来提高网络的深度,通过更深层次的神经网络实现更多特征信息的提取;改变每一层提取的特征层数,实现更多层的特征提取,来提升宽度;通过增大输入图片的分辨率,使网络学习到更加丰富的内容,达到提高精确度的目的。

EfficientNet 使用一组固定的缩放系数统一缩放网络深度、宽度和分辨率,网络可以平衡缩放,进入网络中图像的分辨率、网络宽度和网络深度,减少了模型参数量,增强了特征提取能力,使网络的速度和精度达到最佳。EfficientNet 有多个 MBConv 模块。其总体的设计思路是 Inverted residuals 结构和残差结构,首先进行 1×1 卷积操作进行升维,在 3×3 或

5×5 的深度可分离卷积后,增加了一个通道的注意力机制,最后利用 1×1 卷积进行降维后,增加一个大的残差边,进行特征层的相加操作。MBConv 模块^[9]如图 2 所示。

1.3 CIOU 损失函数

YOLOv4 在损失函数方面,将 $CIOU$ ^[10-11]作为回归 $LOSS$ 。为了使目标框回归更加稳定,收敛的更快, $CIOU$ 考虑了多重因素。其中包括:目标框与锚框之间的距离,重叠率、尺度以及惩罚项,提高定位的精度,防止出现与 IOU (真实框和预测框的交集和并集之比)在训练过程中相同的发散问题。如式(1):

$$CIOU = IoU - (\rho^2(b, b^{gt}) / c^2 - \alpha v) \quad (1)$$

其中, α 和 v 的计算如式(2) ~ (3):

$$\alpha = v / ((1 - IoU) + v) \quad (2)$$

$$v = (4 / \pi^2) (\arctan(w^{gt} / h^{gt}) - \arctan(w / h))^2 \quad (3)$$

$CIOU$ 回归 $LOSS$ 的计算如公式(4):

$$LOSS_{CIOU} = 1 - IOU + (\rho^2(b, b^{gt}) / c^2 + \alpha v) \quad (4)$$

式中: b, b^{gt} 分别表示锚框和目标框的中心点; ρ 为计算两个中心点的欧式距离; α 为权重系数; v 为长宽比相似性; w, h 分别表示预测框的宽度和高度; w^{gt}, h^{gt} 分别表示真实框的宽度和高度。

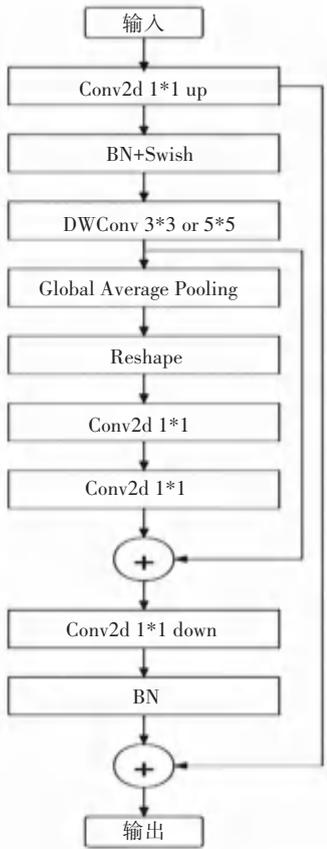


图 2 MBConv 模块结构

Fig. 2 MBConv module structure diagram

2 YOLOv4-Efficient 网络结构

2.1 主干提取网络

YOLOv4 中 CSPDarknet53 主干特征提取网络的提取能力和检测精度非常强, 在各个领域都有非常优秀的表现。但主干巨大的参数量对计算设备的性能有着较高的要求, 使其对车辆的实时监测和算法的移植有较高的难度。YOLOv4-tiny^[12] 虽然具有较低的参数量, 较快的检测速度, 但是特征提取能力和泛化能力较差, 在复杂的场景变化中检测能力较差, 对物体识别效果不太理想。

依据上述分析结果, 为了解决既提高速度又具备较强检测能力的问题, 将特征提取的主干换成 EfficientNet 轻量级网络, 因其具有较少的参数量和较高的特征提取能力, 可以大大减少运算时间提高实时监测能力, 并且该算法可以移植到较低算力的设备上。YOLOv4-Efficient 网络结构如图 3 所示。

EfficientNet 系列网络有 8 种不同类型的模型, 可以应对不同的场景。如: 应对自动驾驶中对车辆的检测时, 考虑到速度和精度, 可选择 EfficientNet-B2 模型, 将 EfficientNet-B2 最后的池化层和全连接层去掉之后, 代替 YOLOv4 原始的 CSPDarknet53 特征提取网络。EfficientNet-B2 网络结构见表 1。

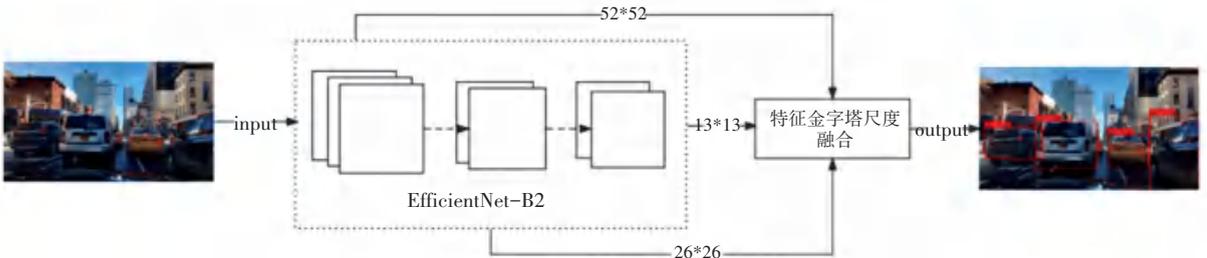


图 3 YOLOv4-Efficient 网络结构

Fig. 3 YOLOv4-Efficient network structure

表 1 EfficientNet-B2 网络结构

Tab. 1 EfficientNet-B2 network structure

阶段 (Stage)	卷积操作 (Operator)	通道数 (Channels)	输出 (Resolution)	网络层数 (Layers)
1	Conv3 * 3/2	32	208	1
2	MBConv1, k3 * 3	16	208	2
3	MBConv6, k3 * 3/2	24	104	3
4	MBConv6, k5 * 5/2	48	52	3
5	MBConv6, k3 * 3/2	88	26	4
6	MBConv6, k5 * 5	120	26	4
7	MBConv6, k5 * 5/2	208	13	5
8	MBConv6, k3 * 3	352	13	2
9	Conv1 * 1&Pooling&FC	1408	13	1

2.2 改进 PANet 特征融合

本文对 PANet 特征融合的改进主要涉及两个方面: 一是引入深度可分离卷积, 二是在进行特征融合上采样时, 加入了 SENet 注意力模块^[13]。

由于 PANet 中使用了大量的卷积, 会导致巨大的参数量, 所以在 PANet 中引入了深度可分离卷积, 使 PANet 特征金字塔部分的参数量大幅度减少, 提高整个模型的检测速度。

深度可分离卷积^[14]分为逐通道卷积^[15]和逐点卷积。在逐通道卷积中, 一个卷积核负责一个通道, 一个通道只被一个卷积核卷积。如图 4 所示。

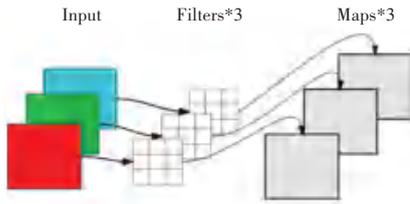


图4 逐通道卷积结构

Fig. 4 Channel by channel convolution structure

逐点卷积运算和常规卷积运算类似,其卷积核尺寸为 $1 \times 1 \times M$ (上一层的通道数)。这里的卷积运算会将逐通道卷积产生特征图,在深度方向上进行加权组合,生成新的特征图。有几个卷积核就输出几个特征图。

为了使不同的特征层进行融合时保证其关键信息得到加强,所以在 PANet 特征融合进行上采样时加入 SENet (Squeeze-and-Excitation Network, 即“压缩和激励”SE 块)。SE 块通过控制 scale 的大小,将一些重要特征进行增强,而对一些不重要的特征进行抑制,从而让提取到的特征有更强的指向性。SE 模块的结构如图 5 所示。

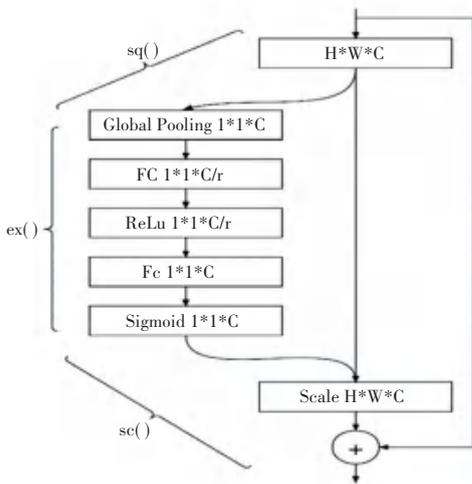


图5 SE 模块结构图

Fig. 5 SE module structure diagram

图中: $sq()$ 代表 Squeeze 过程; $ex()$ 代表 Excitation 过程; $sc()$ 是将 Excitation 得到的结果作为权重,乘到输入特征上。

3 实验结果与分析

本次实验环境配置中,软件环境为:python3.6 编程语言、CUDA 版本为 11.2、深度学习框架 pytorch1.2;硬件环境配置:GPU 为 RTX1050,CPU 为 i7-7700HQ。通过标注和处理 1 521 张来自

BDD100K 自动驾驶的图片,构成本次实验的数据集。

实验中,将改进的 YOLOv4-Efficient 模型与 YOLOv4 模型、YOLOv4-tiny 进行对比。在输入图片分辨率为 416×416 的情况下,通过精度 (*precision*)、参数量 (*M*)、召回率 (*recall*) 和平均准确率 (*mAP*) 等性能指标进行评价。计算公式如式 (5) ~ (6):

$$recall = TP / (TP + FN) \quad (5)$$

$$precision = TP / (TP + FP) \quad (6)$$

式中: *TP* 为图像中车的区域,预测为车的正确情况; *FN* 为图像中为车的区域,预测为不是车错误情况; *FP* 为实际不是车的区域,但是预测此区域是车的情况。

将 YOLOv4-Efficient 与 YOLOv4, YOLOv4-tiny 模型进行对比测试,测试结果见表 2。

表2 模型检测性能对比

Tab. 2 Performance comparison of models

指标	平均准确率 (<i>mAP</i> = 0.5)	召回率 (<i>recall</i>)	参数量 <i>M</i>	每秒帧率 (FPS)	精度 (<i>precision</i>)
YOLOv4	97.59%	95.65%	224.29	10.70	91.19%
YOLOv4-tiny	78.26%	73.09%	23.10	67.03	98.75%
YOLOv4-Efficient (本文模型)	96.27%	93.67%	60.52	17.34	94.29%

通过表中数据可以看出,本文算法整体性要优于 YOLOv4 和 YOLOv4-tiny。在平均准确率 (*mAP*) 和召回率 (*recall*) 对比中,明显高于 YOLOv4-tiny; 在参数量和精度方面要优于 YOLOv4; 在平均准确率和召回率方面略低于 YOLOv4。进行综合对比后,在减少近 4 倍的参数量后,仍然有较好的检测性能。pr 曲线和精度曲线如图 6~7 所示。

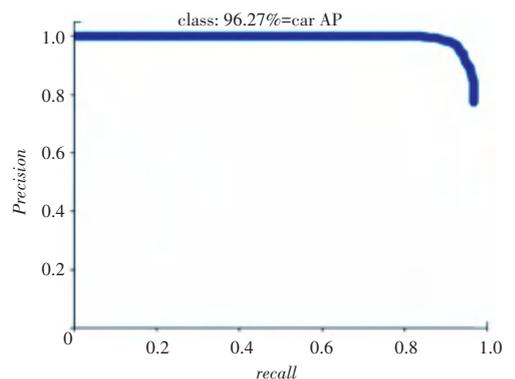


图6 pr 曲线

Fig. 6 Pr curve

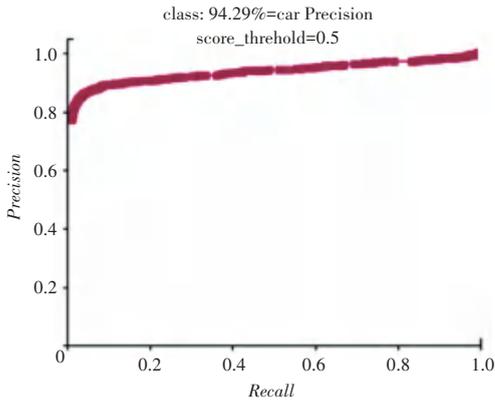


图7 精度曲线

Fig. 7 Precision curve

利用YOLOv4-Efficient测试结果如图8所示。

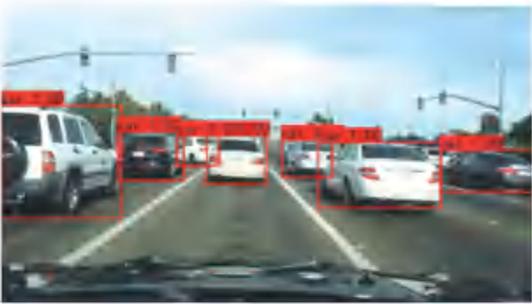


图8 测试结果

Fig. 8 Test result

4 结束语

本文在YOLOv4的基础上提出了一种轻量级的车辆检测算法YOLOv4-Efficient,将YOLOv4的主干提取网络改为EfficientNet-B2后,在保证检测精度的情况下,减少了模型的参数量,提高了检测速度。相比YOLOv4的参数量,改进后的模型参数量少了近4倍。经过实验对比,在YOLOv4-Efficient对车辆的检测中,参数量和精度方面要优于YOLOv4,其精度达到了94.29%,参数量也仅仅有60.52 M,改进后的模型更适用于移动端的设备。后期工作中,在

提高模型数据量来增强鲁棒性的同时,还要进一步研究在不同复杂环境和天气下对目标的检测。

参考文献

- [1] GIRSHICK R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
- [2] 钟志峰,夏一帆,周冬平,等. 基于改进YOLOv4的轻量化目标检测算法[J/OL]. 计算机应用;1-8[2022-02-08]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20210929.1334.012.html>.
- [3] 杨英彬. 基于改进YOLOv4的车辆检测及统计研究[D]. 武汉: 江汉大学, 2021.
- [4] 韩进,周昊. 树莓派的YOLOv3轻量化算法研究及应用[J]. 单片机与嵌入式系统应用, 2021, 21(5): 43-47.
- [5] YU Hailan, CHEN Weili. Motion Target Detection and Recognition Based on YOLOv4 Algorithm[J]. Journal of Physics: Conference Series, 2021, 2025(1): 012053.
- [6] 刘军,张艳迪. 基于Efficient高效网络的目标检测和识别[J]. 沈阳理工大学学报, 2020, 39(6): 1-5.
- [7] WU Tao, ZHU Hongjin, FAN Honghui, et al. An improved target detection algorithm based on EfficientNet[J]. Journal of Physics: Conference Series, 2021, 1983(1): 012017.
- [8] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//International Conference on Machine Learning. PMLR, 2019: 6105-6114.
- [9] 毛昭勇,王亦晨,王鑫,等. 面向高速公路的车辆视频监控分析系统[J]. 西安电子科技大学学报, 2021, 48(5): 178-189.
- [10] 刘昌华. 复杂交通场景下自动驾驶道路目标检测[D]. 大连: 大连理工大学, 2021.
- [11] ZHENG Z, WANG P, REN D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Transactions on Cybernetics, 2021: 1-13. DOI:10.1109/TCYB.2021.3095305.
- [12] 陈梦涛,余粟. 基于改进YOLOV4模型的交通标志识别研究[J/OL]. 微电子学与计算机;1-10[2022-02-08]. <http://kns.cnki.net/kcms/detail/61.1123.TN.20210923.0115.001.html>.
- [13] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
- [14] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.
- [15] 张宝朋,康谦泽,李佳萌,等. 轻量化改进的YOLOv4目标检测算法[J/OL]. 计算机工程;1-13[2022-02-08]. DOI: 10.19678/j.issn.1000-3428.0062216.

(上接第122页)

参考文献

- [1] 蔡炎斌. 对公安基层派出所职能定位的思考[J]. 湖南警察学院学报, 2018, 30(5): 122-128.
- [2] 王发曾. 城市公安机构布局调整的必要性与方法[J]. 城市问题, 1995(1): 11-14.
- [3] 王春梅,慕三英. 我国西部农村公安派出所规划布局问题研究[J]. 江西公安专科学校学报, 2006(2): 31-33.

- [4] 孙立,王中辉,孙立坚,等. 城市派出所空间位置优化及警力分配[J]. 地球信息科学学报, 2019, 21(3): 346-358.
- [5] 翁里,潘建珍. 城市规划与防控犯罪之理论初探[J]. 中国人民公安大学学报, 2003(2): 135-140.
- [6] 刘一恒,李文武. 快速出警的派出所最优选址模型研究[J]. 成都工业学院学报, 2020, 23(4): 58-62.
- [7] COOPER L. Heuristic methods for location-allocation problems[J]. SIAM Review, 1964, 6(1): 37-53.