DOI: 10.14135/j.cnki.1006-3080.20201110002

一种改进的适用于监控视频的轻量级入侵检测算法及其应用

陈 涛[1], 陈天宇[1], 万永菁[1], 王 嵘[1], 孙 静[2]
(1. 华东理工大学信息科学与工程学院, 上海 200237;
2. 上海卓希智能科技有限公司研发部, 上海 201611)

摘要: 由于传统的目标检测算法较为复杂, 在算力、存储空间有限的场景下无法实时检测, 因此本文提出了一种轻量级入侵检测算法。首先采用自适应更新率的混合高斯前景提取算法提取初筛目标, 然后基于改进的残差压缩网络(R-SqueezeNet)对初筛目标进行识别分类。实验结果表明, 该算法在不降低检测精度的前提下, 比传统算法的检测速度平均提升了30倍, 模型体积缩减至YOLOv3-tiny算法的1/40。

关键词: 监控视频; 入侵检测; 轻量级; 自适应更新率; R-SqueezeNet
中图分类号: TP391.4 **文献标志码:** A

视频监控系统一直以来都是公共安防的重要组成部分, 随着计算机视觉目标检测算法的发展, 自动化、智能化的视频监控系统应运而生[1]。

对于视频的入侵检测, 通常做法是将视频看成连续的帧图像, 通过算法对每一帧进行目标检测。回顾目标检测算法的发展历程, 最初主要为人工特征+分类器的基于机器学习的方法, 如梯度方向直方图(HOG)+支持向量机(SVM)、哈尔特征(Haar)+自适应增强学习算法(AdaBoost)等[2-3]。这类目标检测算法虽然应用广泛, 但存在着环境适应性低、计算量大、受训练样本影响大、精度相对较低的检测瓶颈。

2013年, 区域卷积神经网络特征算法(RCNN)[4]首次将深度学习卷积神经网络的概念引入目标检测领域, 检测性能比基于机器学习的算法大幅提高, 奠定了目标检测算法发展的基础。2016年, YOLOv1算法[5]通过单阶段检测的方法, 提升了检测速度。随后相继提出的单次多框检测算法(SSD)、视网膜网络算法(RetinaNet)、YOLOv2、YOLOv3等[6-9]主要在精度上逐步进行了改进。目前, 这些深度学习算法使用GPU能实现比较好的检测效果, 但因为算法过于复杂, 在算力、存储空间有限的场景下并不适用。

缩减主干分类网络可以提升算法检测速度, 但同时也意味着精度的降低, 如YOLOv3-tiny[9]。这也导致了类似于压缩网络(SqueezeNet)[10]的移动端轻量级分类网络无法直接应用于传统的目标检测算法。

针对上述现状, 考虑对算法速度、模型体积要求高的场景, 本文提出了一种改进的轻量级入侵检测算法。首先以自适应更新率的混合高斯前景提取算法定位运动目标, 针对镜头突变的情况, 基于前景图前后帧的信息熵差自适应地调整背景更新率。提取完成后再通过改进的残差压缩网络(R-SqueezeNet)对定位目标进行分类, 该分类网络借鉴了SqueezeNet的核心构件Fire Module并通过引入残差结构提升性能。实验结果表明, 本文算法较传统目标检测算法大幅优化了检测速度和模型体积。

1 改进的轻量级入侵检测算法

1.1 整体算法架构

入侵检测算法流程如图1所示。视频由非制冷型热像仪采集, 非制冷型热像仪的优点在于能够感知物体的温度, 在恶劣环境下也能探测出运动目标。

收稿日期: 2020-11-10
基金项目: 国家自然科学基金(61872143)
作者简介: 陈 涛(1996—), 男, 江苏泰州人, 硕士生, 主要研究方向为目标检测、深度学习。E-mail: ctzj1026@163.com
通信联系人: 万永菁, E-mail: wanyongjing@ecust.edu.cn

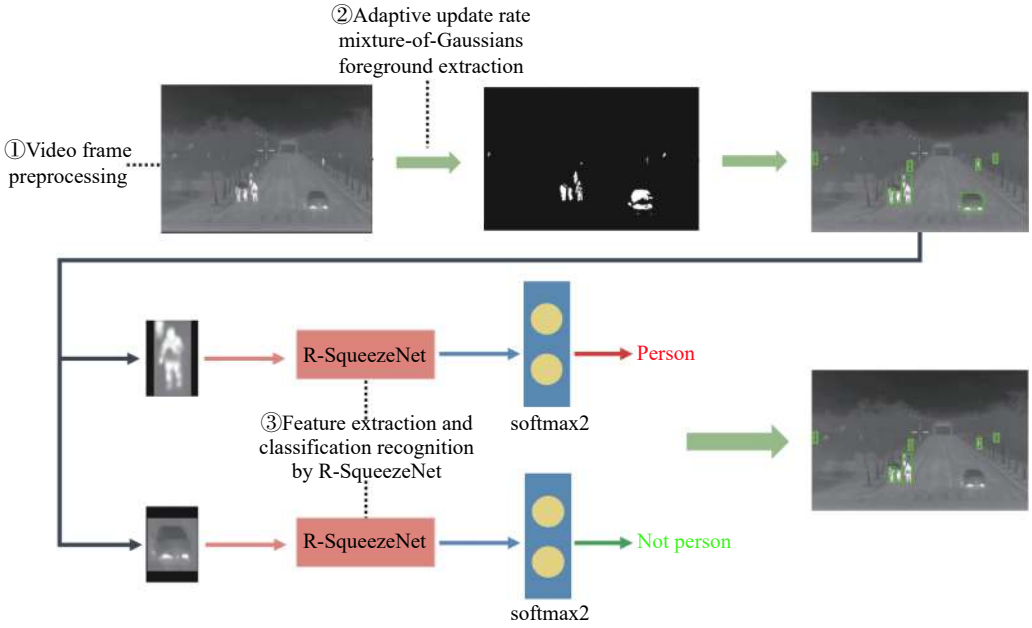


图 1 整体算法流程

Fig. 1 Overall algorithm flow

算法主要由三部分组成: 视频帧预处理、基于自适应更新率的混合高斯前景提取及基于改进的 R-SqueezeNet 卷积神经网络的特征提取和分类识别。具体步骤如下:

(1)实时读取监控视频,通过预处理获得像素值范围为 [0,255] 的灰度视频帧。预处理成灰度图的目的是为了降低整体网络的计算复杂度。

(2)基于自适应更新率的混合高斯前景提取算法提取出运动前景并等比例调整成统一大小 (不足部分以黑色背景填充)。等比例调整的目的是为了保留运动目标的形状轮廓信息,有利于后续的特征提取及分类识别。

(3)采用改进的 R-SqueezeNet 分类网络提取运动前景的特征并识别,当识别为入侵目标时,实时标注并报警,其他目标不处理。

1.2 基于自适应更新率的混合高斯前景提取算法

混合高斯前景提取是一种基于混合高斯背景建模的算法,最早由 Stauffer 提出[11]。该算法对每个像素点的像素值复用多个单高斯模型,当像素值符合其中某个单高斯分布时,该像素点被判断为背景点,否则被判断为前景点。Zivkovic 等[12-13] 在此基础上对每个像素点的高斯分布数进行了自适应改进,减少了算法的计算量。

对于混合高斯前景提取算法,背景更新率 α 是一个重要指标, $\alpha \in (0,1]$。通常情况下, α 越大,背景更新速度越快,但检测效果较差,适用于急剧变化的场景。当 α=1 时,表现为逐帧更新背景。 α 越小,背景更新速度越慢,但检测效果较好,适用于稳定的场景[14-15]。当 α 趋近于 0 时,表现为不更新背景。在文献 [11] 中经过实验对比,综合考虑背景的更新速度和检测性能,更新率默认值 α_0 固定设为 0.001~0.005。

实验发现,当镜头近距离处出现突变情况,即有较大的运动目标突然侵入时,会影响摄像头白平衡、色温的处理机制[16],从而导致其他像素值发生变化,大量背景点被误判为前景点,且由于背景更新率较低,短时间内无法恢复。对于安防领域的入侵检测,镜头突变极易产生误报,因此实际检测时,固定的更新率并不适用。镜头突变前后的提取效果如图 2 所示。

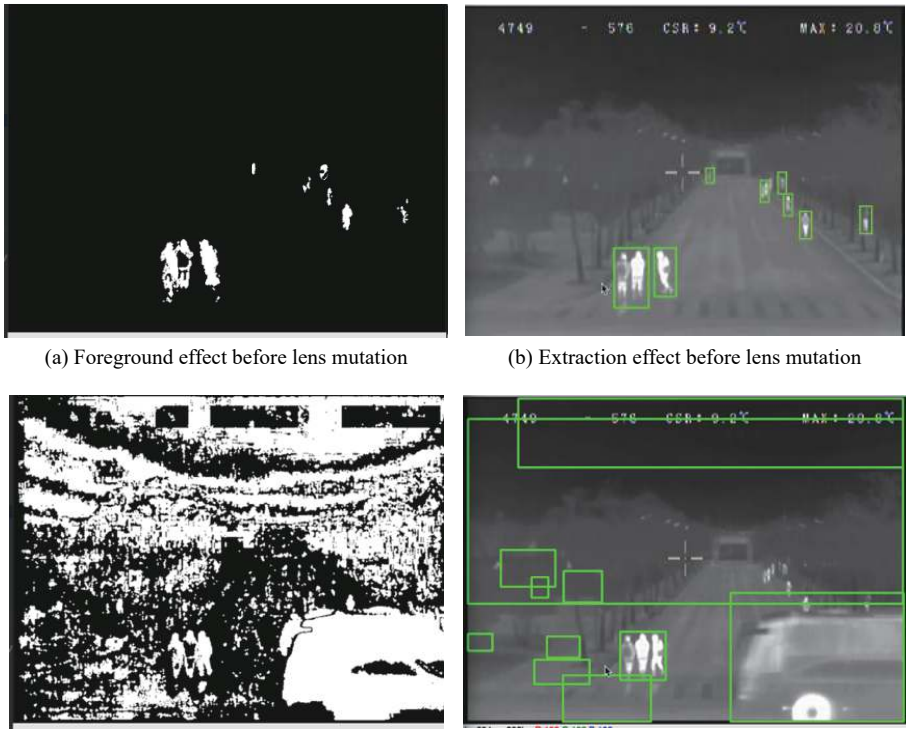
针对此类情况,对更新率 α 进行自适应改进,如式(1)所示:

$$\alpha = \begin{cases} \alpha_0, |\Delta H| < t_H \\ 1, |\Delta H| \geqslant t_H \end{cases} \tag{1}$$

式中, ΔH 为混合高斯前景图前后两帧的信息熵差。当镜头发生突变时,混合高斯前景图会急剧变化,可采用图像的信息熵差来判定是否发生突变[17-18]。当信息熵差的绝对值超过阈值 t_H ,表明发生镜头突变,置更新率 α 为 1,逐帧进行背景更新,从而保证被误判的背景点在最短时间内恢复。混合高斯前景图的信息熵如式(2)所示。

$$H = -\sum_{i=0}^{255} P(X_i) \times \log_2 P(X_i) \tag{2}$$

式中, $P(X_i)$ 为图像中像素值等于 i 的像素个数占总像素个数的比例。



(a) Foreground effect before lens mutation　(b) Extraction effect before lens mutation

(c) Foreground effect after lens mutation　(d) Extraction effect after lens mutation

图 2　镜头突变前后的提取效果对比

Fig. 2　Comparison of extraction effect before and after lens mutation

对于 t_H 的取值, 综合考虑算法的精度和提取前景点的数量, 基于多个场景包括白天室内、白天室外、夜晚室内、夜晚室外、雨天室外等多个包含镜头突变的视频进行实验, 不同 t_H 下算法的表现如图 3 所示。

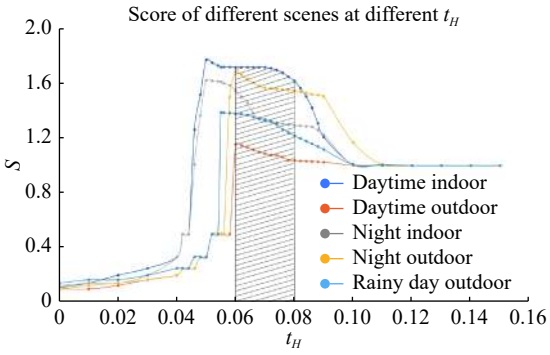


图 3　不同 t_H 下算法的表现

Fig. 3　Algorithm performance under different t_H

S 的计算公式如(3)所示。

$$S=\begin{cases}\dfrac{1}{m_{\alpha_0}-m+1}, m-m_{\alpha_0}<0\\ 1+\dfrac{n_{\alpha_0}-n}{n_{\alpha_0}}, m-m_{\alpha_0}=0\end{cases} \tag{3}$$

式中: m_{α_0}、n_{α_0} 分别为更新率固定为默认值 α_0 时, 检测出前景对象的个数和提取前景点的数量; m、n 分别为自适应更新率对应 t_H 下, 检测出前景对象的个数和提取前景点的数量。 $m-m_{\alpha_0}<0$ 表明当前对应 t_H 下发生了漏检, $m-m_{\alpha_0}=0$ 表明当前对应 t_H 下未发生漏检。

由图 3 可知, 由于每种场景的运动前景数不同, S 的极值不同。但对于图 3 中的几种场景, t_H 取 0.06~0.08 时算法的 S 达到相对较大的值。

图 4 示出了 t_H =0.07 时, 发生镜头突变后的相同时间段内, 基于固定更新率和自适应更新率算法的提取效果对比, 可以看出基于自适应更新率算法提取的无效前景对象大幅减少。

1.3　改进的轻量级卷积神经网络 R-SqueezeNet

以 3 个 Fire module 为例, 改进的 R-SqueezeNet 分类网络结构如图 5 所示。该网络借鉴了 SqueezeNet[10] 的核心构件 Fire module, 并通过引入残差结构提升网络性能。

Fire module 是模块化的卷积 (Conv) 层, 由 Squeeze 层和 Expand 层组成, 其计算流程如图 6 所示。其中 H、W、M 表示特征图的长、宽、通道数; k、c 表示卷积核的大小、个数; S_1 为 Squeeze 层中 1×1 卷积核的数量; E_1、E_3 为 Expand 层中 1×1、3×3 卷积核的数量。Fire module 的基本网络单元在保证特征信息不丢失的前提下, 减少 3×3 卷积核的通道数, 降低网络模型的参数量[10]。

残差网络[19](ResNet) 中残差结构的提出有效解决了深度神经网络的退化问题[20-21], 残差结构如图 7

(a) Effect of regular algorithm (b) Effect of adaptive algorithm

图 4 提取效果对比

Fig. 4 Comparison of extraction effect

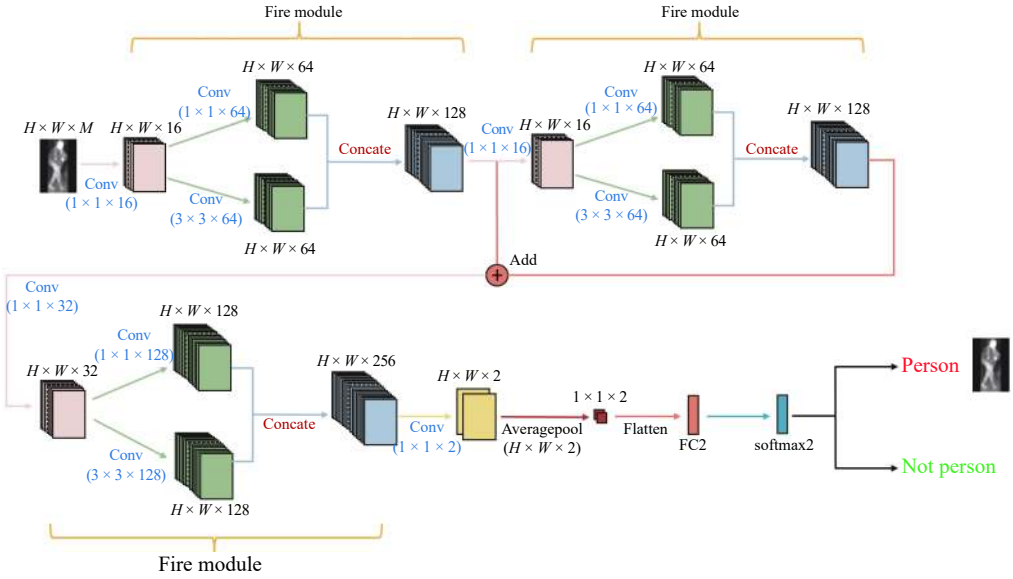


图 5 R-SqueezeNet 结构

Fig. 5 Structure of R-SqueezeNet

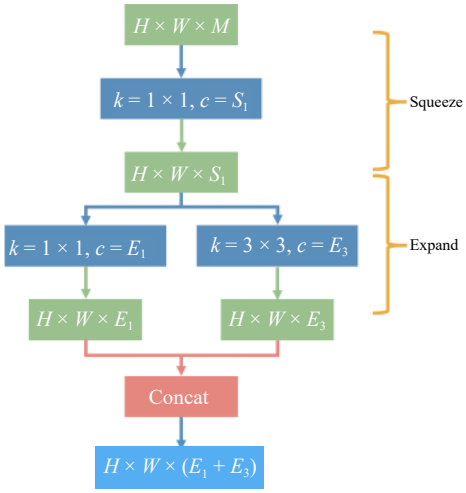


图 6 Fire module 计算流程

Fig. 6 Calculation process of fire module

所示。R-SqueezeNet 通过引入残差结构，在不增加参数的前提下，提升了网络模型的准确率。

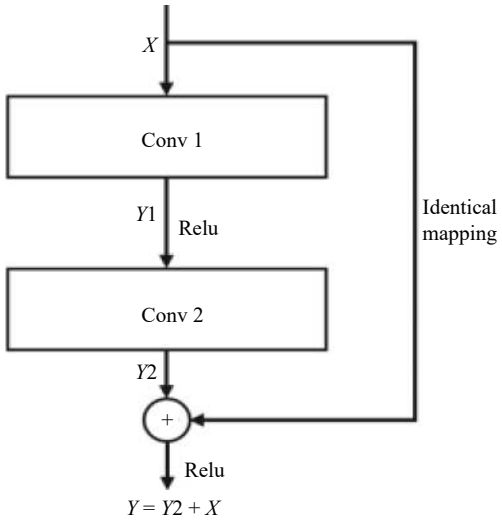


图 7 残差结构

Fig. 7 Residual structure

2　实验与结果分析

实验的软件环境为 OpenCV 及 Keras 深度学习框架, 硬件环境为 Intel i5-8250U 1.6 GHz 低压处理器。以行人为入侵对象进行检测实验。共获取 17 493 个样本进行训练, 其中正样本 8 693 个, 负样本 8 800 个; 共获取 4 374 个样本进行测试, 其中正样本 2 174 个, 负样本 2 200 个。正负样本在白天、夜晚、雨天、晴天等多个环境下采集, 包含行人、机动车、非机动车、动物、植被等。

2.1　等比例调整运动前景

采用自适应更新率的混合高斯前景提取算法获得运动前景后, 等比例调整运动前景的大小以保留运动前景的形状轮廓信息, 有助于卷积网络的特征提取和分类识别。图 8 示出了基于 LeNet[22]、AlexNet[23]、ZFNet[24] 及 R-SqueezeNet 等比例调整及非等比例调整运动前景的分类精度 (Accuracy) 对比, 。

其中, R-SqueezeNet 采用 3 个 Fire module。由图 8 可知, 在多个卷积网络模型下, 以等比例调整的运动前景作为样本的分类精度均优于以非等比例调整的运动前景作为样本的分类精度。

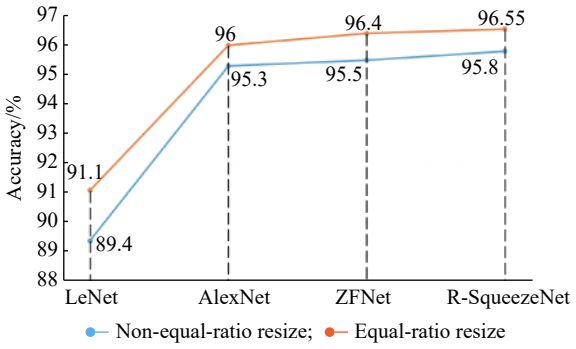


图 8　不同前景调整方式的分类精度对比

Fig. 8　Comparison of different foreground resizing method

2.2　R-SqueezeNet

为确定 Fire module 的个数和残差结构是否有效, 在不同条件下进行对比实验。表 1 为本文采集样本、公开的 Kaggle 猫狗数据集及 cifar-10 随机两类数据集下统一大小后分类网络的检测精度、单张推理时间及模型体积对比。

表 1　分类网络模型对比

Table 1　Comparison of classification network models

Model	Number of Fire module	Accuracy/%			Inference time/ms	Size/MB
		This paper	cats vs dogs	cifar-10		
SqueezeNet	1	90.56	89.9	91.6	1.34	0.16
	2	94.35	90.61	93.85	1.7	0.33
	3	96.32	92.96	95.74	2.3	0.89
	4	96.58	93.17	96.13	3.78	1.5
	5	96.67	93.39	96.37	4.04	2.8
R-SqueezeNet	**3**	**96.55**	**93.2**	**96.1**	**2.88**	**0.89**
	4	96.73	93.32	96.35	3.92	1.5
	5	96.8	93.49	96.55	4.56	2.8
	6	97.01	93.56	96.61	7.02	4.1
	7	97.2	93.71	96.69	7.68	6.4
	8	97.38	93.87	96.73	9.3	8.9

由表 1 可知, 残差结构在对网络模型体积、单张推理时间影响不大的前提下, 一定程度地提升了分类模型的准确率。在 Fire module 增加至 3 个后, 准确率提升不明显。以采用 3 个 Fire module 并引入残差结构的网络 (R-SqueezeNet3) 和采用 4 个 Fire module 并引入残差结构的网络 (R-SqueezeNet4) 对比, R-SqueezeNet4 的模型体积增加约一倍, 分类准确率平均仅提升约 0.18%。因此综合模型体积、分类准确率及单张推理时间, 建议采用 3 个 Fire module。

2.3　入侵检测算法

对于监控视频的入侵检测, 采用误检率和漏检率判断算法的检测精度。考虑到视频帧具有连续性, 定义当检测区域内出现非入侵对象被检测为入侵对象时, 即发生误检(False detection, FD); 当入侵

对象从进入检测区域到离开期间被正确检测的帧数占期间总视频帧数的比例低于50%时，即发生漏检(Missed detection, MD)。

实验在相同的软硬件条件下进行，以检测入侵的行人为例。图9示出了不同 t_H 下基于不同特征提取网络的算法性能对比。表2为基于自适应更新率前景提取和非自适应更新率前景提取的算法性能对比。

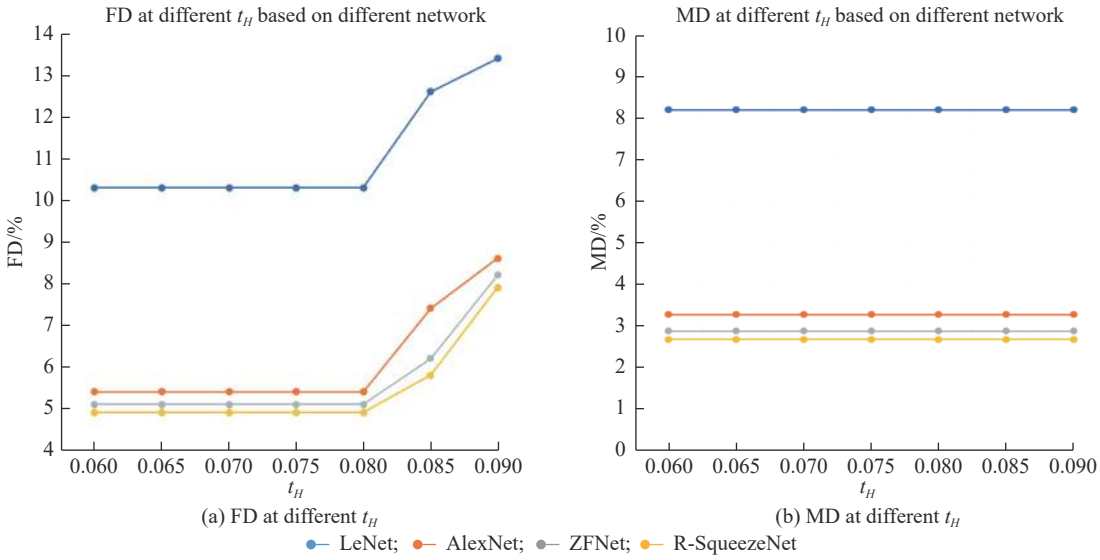


图9 不同 t_H 下的算法性能对比

Fig. 9 Performance comparison of algorithms at different t_H

表2 基于自适应和非自适应前景提取的算法对比

Table 2 Comparison of algorithms based on adaptive and non-adaptive foreground extraction

Based	Backbone	FD/%	MD/%
Non-adaptive extraction	LeNet	13.8	8.2
	AlexNet	9.1	3.3
	ZFNet	8.7	2.9
	R-SqueezeNet	8.6	2.7
Adaptive extraction	LeNet	10.3	8.2
	AlexNet	5.4	3.3
	ZFNet	5.1	2.9
	R-SqueezeNet	4.9	2.7

由图9可知，当 t_H 取0.06~0.08时，对于不同特征提取网络，入侵检测算法对应的误检率和漏检率保持不变。由图3可知，白天室外及夜晚室外场景下，t_H 取0.06时 S 取得极大值，且 t_H <0.06时 S 值急速降低，出现漏报现象，因此综合考虑入侵检测算法性能和其他可能出现的特殊情况，建议 t_H 取0.07。

由表2可知，基于自适应前景提取的算法在不影响漏检率的前提下，降低了算法的误检率。

本文算法和传统目标检测算法的性能对比如表3所示。

由表3可知，在相同软硬件条件下，本文算法和传统目标检测算法相比，在检测精度相近的前提下，检测速度较传统目标检测算法平均提升约30倍。模型体积低于1 MB，缩减至YOLOv3-tiny的1/40。

表3 本文算法和传统目标检测算法对比

Table 3 Comparison with traditional object detection algorithm

Algorithm	Backbone	Size/MB	FD/%	MD/%	FPS
SSD[6]	VGG16[25]	95.7	5.3	2.9	1
RetinaNet[7]	ResNet50[19]	146.1	4.2	2.4	<1
YOLOv2[8]	Darknet19[8]	194	4.5	2.5	<1
YOLOv3[9]	Darknet53[9]	246.9	4.2	2.3	1
YOLOv3-tiny[9]	Darknet13[9]	35.6	7.6	3.7	5
This paper	**R-SqueezeNet**	**0.89**	**4.9**	**2.7**	**44**

3 结 论

目标检测包括定位和分类两大任务，传统深度学习算法用同一个主干网络进行位置回归和目标分类，算法模型较为复杂，无法应用于对检测速度、模型体积要求高的场景。针对此现状，本文提出一种改进的轻量级入侵检测算法：先通过自适应更新率的混合高斯前景提取算法完成入侵检测的定位任

务, 再基于 R-SqueezeNet 网络对定位的运动目标进行分类判别。本文算法通过前景提取代替基于主干网络的位置回归及分类网络的优化, 整体的检测速度、模型体积均优于传统目标检测算法。

参考文献:

[1] 赵潇. 基于人类视觉系统的监控视频目标提取技术研究[D]. 重庆: 重庆邮电大学, 2018.

[2] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2005: 886-893.

[3] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2001: 511-518.

[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. USA: IEEE, 2014: 580-587.

[5] REDMON J, DIVVALA S, GIRSHICK R, *et al*. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2016: 779-788.

[6] LIU W, ANGUELOV D, ERHAN D, *et al*. SSD: Single shot multibox detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[7] LIN T Y, GOYAL P, GIRSHICK R, *et al*. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Italy: IEEE, 2017: 2980-2988.

[8] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2017: 7263-7271.

[9] REDMON J, FARHADI A. Yolov3: An incremental improvement[EB/OL]. arxiv. org, (2018-04-10)[2020-11-01]. https://arxiv.org/pdf/1804.02767.pdf.

[10] IANDOLA F N, HAN S, MOSKEWICZ M W, *et al*. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size[EB/OL]. arxiv. org, (2016-03-20)[2020-11-01]. https://arxiv.org/pdf/1602.07360v3.pdf.

[11] STAUFFER C, GRIMSON W E L. Learning patterns of activity using real-time tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 747-757.

[12] ZIVKOVIC Z. Improved adaptive Gaussian mixture model for background subtraction[C]//Proceedings of the 17th International Conference on Pattern Recognition. UK: IEEE, 2004: 28-31.

[13] ZIVKOVIC Z, VAN DER HEIJDEN F. Efficient adaptive density estimation per image pixel for the task of background subtraction[J]. Pattern Recognition Letters, 2006, 27(7): 773-780.

[14] LEE D S. Effective Gaussian mixture learning for video background subtraction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(5): 827-832.

[15] 龚安, 牛博, 史海涛. 基于分块的帧差法和混合高斯算法的油田作业区入侵检测[J]. 计算机与数字工程, 2019, 47(12): 3041-3044.

[16] 刘馨. 监控视频中的图像颜色评价与优化[D]. 杭州: 浙江大学, 2015.

[17] 李均, 王志诚, 吴雨轩, 等. 熵概念的延拓——从热熵到信息熵[J]. 大学物理, 2020, 39(10): 29-33.

[18] 王林, 王超凡. 差分信息熵在拼接图像质量评估中的应用[J]. 计算机仿真, 2020, 37(4): 265-268, 273.

[19] HE K, ZHANG X, REN S, *et al*. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2016: 770-778.

[20] 汪斌, 陈宁. 基于残差注意力U-Net结构的端到端歌声分离模型[J]. 华东理工大学学报(自然科学版). doi: 10.14135/j.cnki.1006-3080.20200903001.

[21] 高磊, 范冰冰, 黄穗. 基于残差的改进卷积神经网络图像分类算法[J]. 计算机系统应用, 2019, 28(7): 139-144.

[22] 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.

[23] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[24] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//European Conference on Computer Vision. Cham: Springer, 2014: 818-833.

[25] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. arxiv. org, (2014-10-22)[2020-11-01]. https://arxiv.org/abs/1409.1556.

An Improved Lightweight Intrusion Detection Algorithm and Its Application

CHEN Tao[1], CHEN Tianyu[1], WAN Yongjing[1], WANG Rong[1], SUN Jing[2]

（1. School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China; 2. Department of Research and Development, Shanghai Joosee Smart Technology Company Limited, Shanghai 201611, China）

Abstract: With the development of target detection algorithms, intrusion detection based on surveillance video has attracted more and more attention. At present, the traditional target detection algorithm is more complex, and can not be detected in real time in the scene of limited computing power and storage space.Aiming at this pain point, a lightweight intrusion detection algorithm is proposed: firstly extract the preliminary screening target through the adaptive update rate of the mixed Gaussian foreground extraction algorithm, and then identify the preliminary screening target based on the improved residual squeeze network (R-SqueezeNet) classification. Experimental results show that, without reducing the detection accuracy, the algorithm can increase the detection speed by an average of 30 times compared with the traditional algorithm, and the model size is reduced to 1/40 of YOLOv3-tiny.

Key words: surveillance video**;** intrusion detection**;** lightweight**;** adaptive update rate**;** R-SqueezeNet