

基于改进 MTCNN 算法的低功耗边缘人脸检测跟踪系统*

祁星晨, 卓旭升

(武汉工程大学 电气信息学院, 湖北 武汉 430205)

摘要: 边缘设备的快速发展和深度学习的落地应用越来越多, 两者结合的趋势越发明显。而针对低功耗边缘设备 AI 应用的潜力还未完全开发出来, 大量设备隐藏着大量计算能力, 释放其潜力所带来的社会效益和经济效益是非常明显的。因此, 以目标检测任务中较为常见的人脸检测为例, 将 MTCNN 人脸检测算法改进并移植到资源极其紧张的低功耗嵌入式平台, 在一定环境条件下, 最终成功地检测到人脸, 并绘制出人脸候选框, 结合舵机云台具备了一定的人脸跟踪能力。

关键词: 低功耗边缘设备; 目标检测; 人脸检测跟踪; 级联卷积神经网络

中图分类号: TP391

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.201100

中文引用格式: 祁星晨, 卓旭升. 基于改进 MTCNN 算法的低功耗边缘人脸检测跟踪系统[J]. 电子技术应用, 2021, 47(5): 40-44.

英文引用格式: Qi Xingchen, Zhuo Xusheng. Low-power edge AI face detection and tracking system based on improved MTCNN algorithm[J]. Application of Electronic Technique, 2021, 47(5): 40-44.

Low-power edge AI face detection and tracking system
based on improved MTCNN algorithm

Qi Xingchen, Zhuo Xusheng

(School of Information and Electrical Engineering, Wuhan Institute of Technology, Wuhan 430205, China)

Abstract: The rapid development of edge devices and the application of deep learning are increasing, the trend of combining the two is becoming more and more obvious. The potential of AI applications for low-power edge devices has not yet been fully developed. A large number of devices hide a lot of computing power. The social and economic benefits brought by the release of its potential are very obvious. Therefore, taking the more common face detection in objective detection tasks as an example, the MTCNN face detection algorithm is improved and transplanted to a low-power embedded platform with extremely limited resources. Under certain environmental conditions, the face is finally successfully detected, and the face candidate boundingbox is drawn, it has face tracking function combined with the servo.

Key words: low-power edge devices; object detection; face detection and tracking; cascaded convolutional neural network

0 引言

近年来, 边缘设备等爆炸式增长, 百亿数量级的边缘设备接入互联网。传统的 AI 计算架构主要是依靠云计算, 虽然云计算能够提供足够的计算能力和可靠的计算结果, 但其不断地消耗大量电力, 且边缘设备也需要消耗能量收集数据并传输到云端, 传输过程存在着延迟。而边缘设备与 AI 的结合能够降低能源的消耗以及降低延迟, 使得原本在云端完成的任务可在边缘设备完成, 降低了云端的负担, 发掘了边缘设备的计算能力^[1-3]。

目标检测是 AI 领域的热点研究任务, 广泛应用于机器人导航、工业检测和航空航天等诸多领域, 其目的是找出图像中所有感兴趣的物体, 包含物体定位和物体分

类两个子任务。而人脸检测属于目标检测任务的子集, 在人脸检测基础上的应用有人脸识别、表情分析等。关于目标检测任务的研究有许多, 产出了许多成果, 主要有 Viola-Jones^[4]、Faster-RCNN^[5]、YOLO^[6]、SSD^[7]、Cascade R-CNN^[8]、MTCNN^[9]等, 这些研究在目标检测任务上精度和速度的要求不断提高。对于边缘设备, 特别是低功耗边缘设备, 资源极其有限, 目标检测任务往往需要较大的算力, 因而难以得到应用。文献[10]和文献[11]均利用深度学习的方法对人脸进行检测, 在精度上有着较好的表现, 但是其应用主要是依靠 PC 系统上强大的 GPU, 难以实际应用于低功耗的边缘设备。文献[12]、文献[13]和文献[14]在 FPGA 平台和 NVIDIA JETSON TX2 嵌入式平台上实现了目标检测和识别的应用, 相比较与 PC 系统的功耗, FPGA 平台和 NVIDIA JETSON TX2 嵌入式平

* 基金项目: 湖北省自然科学基金(2016CFC757)

台的功耗稍低,能效高。文献[15]则将深度可分离卷积于表情识别,文献[16]将 MobileNet-SSD 的目标检测任务部署于 NVIDIA JETSON TX2 平台上,文献[17]将 YOLOv3 的目标检测任务部署于 NVIDIA JETSON TX1 平台上,文献[18]将改进的 R-DA 和 KCF 算法部署于 NVIDIA JETSON TX1 平台。相比较而言,上述平台对于大部分低功耗边缘设备来说,算力较强,因此将精简和优化后的目标检测模型部署于低功耗边缘设备是十分必要的。

基于上述背景和已有的研究,本文以人脸检测为例的目标检测任务,针对 Cortex-M7 为内核的低功耗上的实现和应用,选择了 MTCNN 这一灵活的算法作为本文主要实现的人脸检测功能的主干,在 ST 公司的 STM32H747-DISCO 双核平台上,实现了人脸检测和跟踪,解决了一部分目标检测任务在低功耗边缘设备应用问题。

1 MTCNN 及其改进

1.1 MTCNN

MTCNN 人脸检测和对齐算法是利用多任务级联卷积神经网络的结构完成的,主要任务是寻找图像中人脸所在的位置区域,以及人脸对齐,即找出特征点的位置,如鼻子左侧、鼻孔下侧、瞳孔位置、上嘴唇下侧等点,其主要结构如图 1 所示。

工作过程主要是建立图像金字塔后,由 P-Net 这一建议网络对图像中可能存在人脸的位置提出建议,生成大量裁剪的分辨率为 24×24 的 RGB 图片,输入到 R-Net (也就是精调网络),对 P-Net 提出的建议经过仔细判断,并再次提出经过修改和判断的意见,从原图像中裁剪出意见中对应的、分辨率为 48×48 的图片,最后将这些图片输入到 O-Net 也就是输出网络,得到最终的人脸分类概率、人脸检测框和人脸特征点的位置。图 2 展示了这一三层、三级串联的卷积神经网络结构的工作过程。

1.2 改进 MTCNN

MTCNN 算法具有一定的改进空间,本文主要从三层级联卷积神经网络、极大值抑制(NMS)和输出结构三方面进行速度优化和改进,虽然损失了精度,但是换来了能够在低功耗边缘设备上运行的可能性。

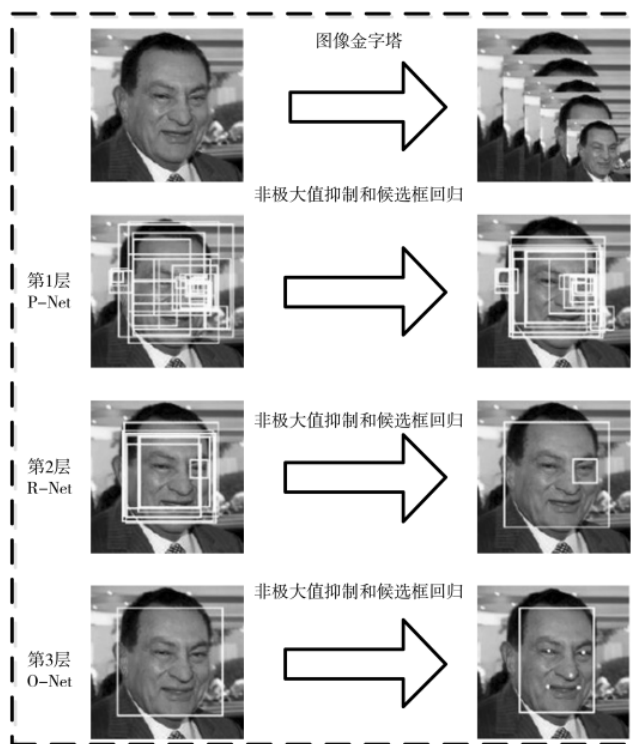


图 2 MTCNN 工作流水线

1.2.1 深度可分离卷积

深度可分离卷积^[19]是谷歌公司于 2017 年在 Mobilenets 中提出的适用于移动设备的轻量级卷积核。

如图 3 所示,标准的卷积核大小为 $D_k \times D_k \times M \times N$,其中 D_k 为核的空间尺寸, M 是输入通道, N 是输出通道。如图 4(a)和图 4(b)所示,深度可分离卷积在同样输入和特征图输出情况下,其大小为 $D_k \times D_k \times M + M \times N$,很明显地可以看出深度可分离卷积的大小较小,若 D_F 为输出特征图大小,那么标准卷积的计算消耗为 $D_k \times D_k \times M \times N \times D_F \times D_F$,深度可分离卷积的计算消耗为 $D_k \times D_k \times M \times D_F \times D_F + M \times N \times D_F \times D_F$,因此深度可分离卷积不仅在大小上而且在速度上比标准的卷积核有优势,故而本文将使用深度可分离卷积替代 MTCNN 中的部分卷积核,并适当减小卷积核的大小。MTCNN 和改进后 3 个网络的对比见表 1。

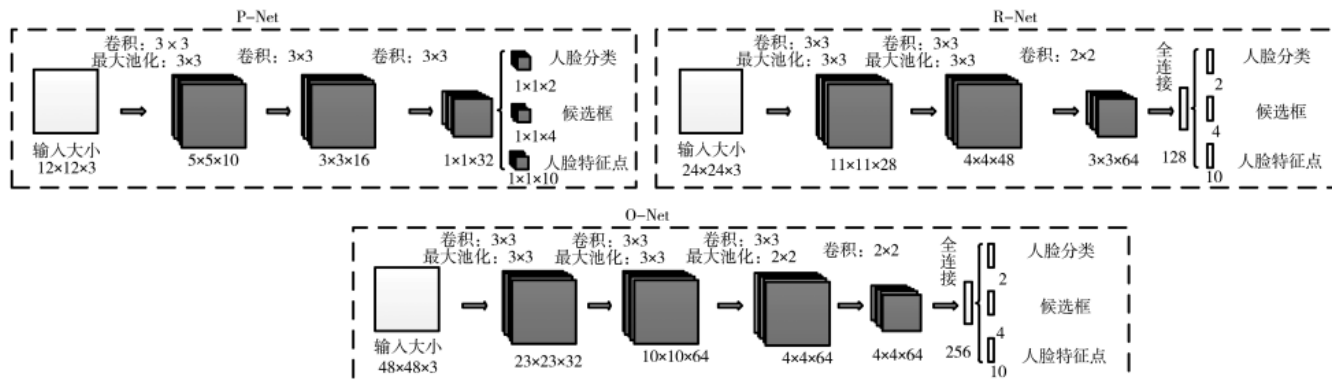


图 1 MTCNN 中 P-Net、R-Net 和 O-Net 的结构

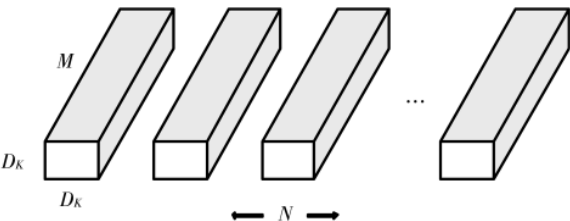
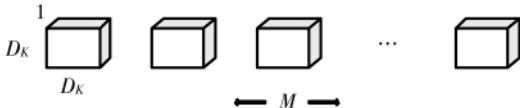
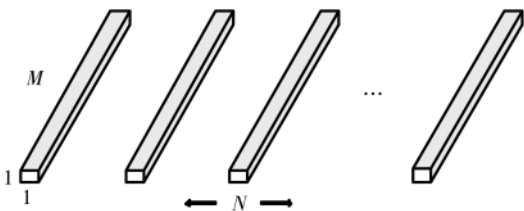


图3 标准卷积的大小



(a)深度卷积



(b)点向卷积

图4 深度可分离卷积构成

1.2.2 非极大值抑制算法(NMS)

MTCNN 中使用的非极大值抑制有效地将 P-Net、R-Net 和 O-Net 输出的人脸候选框重复部分较大的去除,进而使 R-Net 和 O-Net 的计算量减少。对于低功耗边缘设备需要更高效率的非极大值抑制,并可根据应用场景和经验调整相应的阈值,以便适应不同的环境,第一个阈值为候选框控制是人脸的最低概率,第二个阈值为控制多个候选框最大的 Iou 值,具体计算过程可见以下算法。

- (1) $B=\{b_1, \cdots, b_n\}$ 为 n 个候选框
- (2) 对应人脸概率 $P=\{p_1, \cdots, p_n\}$
- (3) $D=\{\}$;
- (4) 取 P 中大于阈值 P_f 组成新的 P, B ;

- (5) P 小于 P_f 有 s 个,那么 $n=n-s$;
- (6) While $B!=\{\}$ do
- (7) 找出最大概率的索引 $m=\operatorname{argmax}(P)$
- (8) 取出该索引的候选框 $M=b_m$
- (9) $D=D \cup M, B=B-M$
- (10) for b_i in B do
- (11) if $\operatorname{Iou}(M, b_i)>N_f$
- (12) Iou 为两人脸候选框的重复面积
- (13) $B=B-b_i; P=P-p_i$
- (14) end
- (15) end
- (16) end
- (17) return D, P

1.2.3 训练和移植改进 MTCNN

本文使用 widerface 数据集,深度学习框架为 Tensor-Flow 2.1.0, GTX-1050 训练改进 MTCNN 的 P-Net 和 R-Net。训练完毕后,将 P-Net 输入为固定大小的图像金字塔,由于边缘设备的图像采集设备往分辨率是固定的,并不需要处理各种分辨率场景的图像,因此固定数量和分辨率的图像金字塔能够更灵活地适应不同计算能力的边缘设备。本文图像金字塔中图像数量为 2,分辨率分别为 80×60 和 40×30 。

1.3 云台跟踪单张人脸

人脸跟踪是在三维空间一个点上旋转摄像头以保持人脸在图像中心,因此本文使用两轴舵机云台作为机动装置,控制舵机角度使用 PWM 占空比的方式,如图 5 方框中所示。

假设人脸候选框已给出,人脸在单帧图像中的位置,候选框的左上角坐标为 (x_0, y_0) ,右下角坐标 (x_1, y_1) ,那么人脸候选框中心点为: $(x_c, y_c) = ([(x_1-x_0)/2], [(y_1-y_0)/2])$,其中 $[]$ 为取整函数。

摄像头获取的单帧图像中心点为 (x_i, y_i) ,采用增量

表1 MTCNN 和改进后的 MTCNN 结构对比

网络	算法	
	MTCNN	改进 MTCNN
P-Net	Conv1	CNN, 10 个滤波器
	Conv2	CNN, 24 个滤波器
	Conv3	CNN, 32 个滤波器
	Conv_cls	CNN
	Conv_bbox	CNN
	参数量/个	6 072
R-Net	Conv1	CNN, 28 个滤波器
	Conv2	CNN, 48 个滤波器
	Conv3	CNN, 128 个滤波器
	全连接	128 个滤波器
	参数量/个	100 178
	省略(可见图 1)	
O-Net	省略(可见图 1)	移除 O-Net
单个人脸耗时(Core i5-4520, TensorFlow 2.1)	600~700 ms	10~16 ms

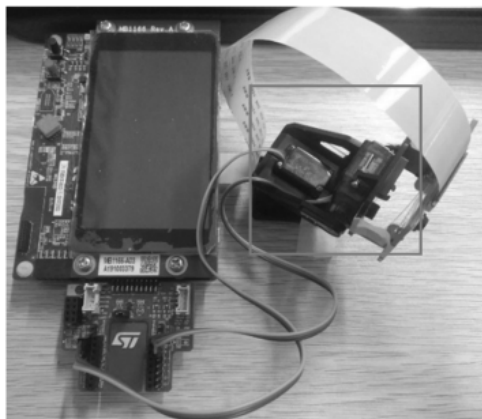


图5 本文平台实物图

式PI算法,对横向轴舵机的控制量则有 $S_x = k_p \cdot (x_c - x_i) + k_i \cdot \sum (x_c - x_i)$, 其中, k_p 、 k_i 为增益; S_y 同理。

2 测试结果

利用ST公司的Cube.AI工具箱对改进MTCNN的P-Net和R-Net进行转换,以便移植到边缘设备,最终可将改进MTCNN移植到STM32H747-Disco平台,并进行测试,平台如图5所示。

本文边缘平台的速度可达600~800 ms一张人脸,多张人脸的计算速度会增加约每张人脸200 ms,由于算法的可调阈值,故而对于人脸个数可控,因此速度也可控,单张人脸检测结果如图6和图7所示,图中的框为人脸检测框。

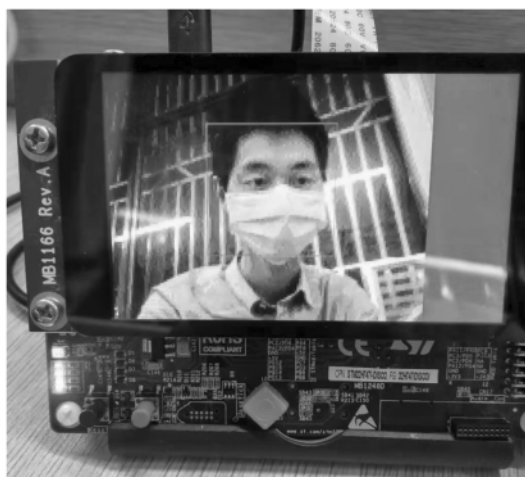


图6 单张人脸检测结果(戴口罩)

本文平台运行在全速(即480 MHz),并且打开了绝大部分外设,根据ST数据手册、文献[13]、文献[14]中提

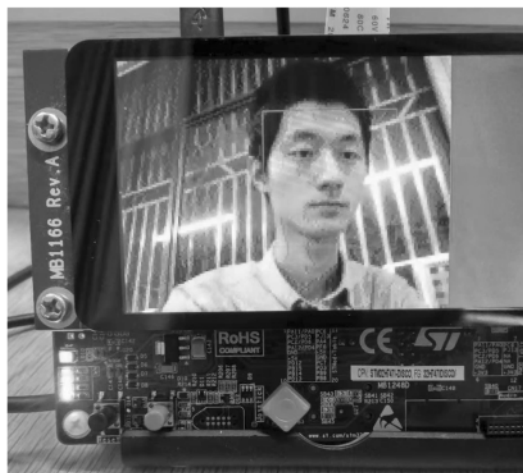


图7 单张人脸检测结果(不戴口罩,背景有强烈白光)

到的平台,表2仅列出CPU的峰值功耗,其余外设不计,并对比了边缘平台所拥有的各种资源,可见低功耗边缘平台的资源是非常紧凑的,而且计算能力远不如大型平台。

3 结论

本文完成了目标检测任务在资源极其有限的计算平台的实现。基于MTCNN算法进行改进,可应用于固定场景、人脸个数不多的条件下,非常适合电池供电的低功耗边缘设备对目标检测功能的需要。测试结果表明,深度学习的方法鲁棒性较高,不仅是人脸检测任务,本文方法可为其他目标检测在低功耗边缘应用作为参考,可用于多种场景。本文方法不足之处和未来需要的工作:

(1)本文目标检测任务在资源极其有限的平台运行,检测速度也极其有限,因此下一步可通过量化、稀疏化等方法对卷积神经网络模型进行优化,提高检测速度。

(2)人脸检测仅有一种物体的检测,因此未来可继续研究如何在低功耗边缘设备上提供两种或两种以上物体的检测,以此充分利用边缘设备的计算能力。

参考文献

- [1] WANG F, ZHANG M, WANG X, et al. Deep learning for edge computing applications: a state-of-the-art survey[J]. IEEE Access, 2020, 8: 58322-58336.
- [2] JOHN A, PANICKER R C, CARDIFF B, et al. Binary classifiers for data integrity detection in wearable IoT edge devices[J]. IEEE Open Journal of Circuits and Systems, 2020, 1: 88-99.
- [3] DE DONNO M, TANGE K, DRAGONI N. Foundations and evolution of modern computing paradigms: cloud, IoT, edge, and fog[J]. IEEE Access, 2019, 7: 150936-150948.

表2 多种边缘平台功耗对比

参数	Xilinx ZYNQ7020	Nvidia TX2	STM32H747
峰值功耗/W	约4	约7.5	0.531
频率/MHz	666(CPU)+150(PL)	2000(CPU)+900(GPU)	480
RAM	512 MB(DDR3)	4 GB(LPDDR4)	1 MB(片上)+32 MB(SDRAM)

- [4] VIOLA P A, JONES M J. Robust real-time face detection[C]. ICCV 2001. Proceedings of Eighth IEEE International Conference on Computer Vision, 2001.
- [5] 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 and Machine Intelligence, 2015, 39(6): 1137-1149.
- [6] REDMON J, FARHADI A. YOLOv3: an incremental improvement[R]. arXiv e-Prints, 2018.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]. European Conference on Computer Vision, 2016.
- [8] CAI Z, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 6154-6162.
- [9] ZHANG K, ZHANG Z, LI Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.
- [10] 孙贵华, 陈淑荣. 一种改进的 RefineDet 多尺度人脸检测方法[J]. 电子技术应用, 2019, 45(8): 34-39.
- [11] 王静波, 孟令军. 卷积神经网络人脸检测算法[J]. 电子技术应用, 2020, 46(1): 34-38.
- [12] 林志文, 林志贤, 郭太良, 等. 基于 FPGA 加速的卷积神经网络识别系统[J]. 电子技术应用, 2020, 46(2): 24-27.
- [13] 张雷, 王越. 嵌入式平台下的车辆跟踪系统设计[J]. 电子技术应用, 2019, 45(11): 13-16.
- [14] 陈辰, 严伟, 夏珺, 等. 基于 FPGA 的深度学习目标检测系统的设计与实现[J]. 电子技术应用, 2019, 45(8): 40-43, 47.
- [15] 柳永翔, 付晓峰, 付晓鹏, 等. 深度可分离 CNN 在表情识别中的应用研究[J]. 工业控制计算机, 2020, 33(10): 71-73, 76.
- [16] 童星, 张激. 基于 SSD-MobileNet 模型的 ROS 平台目标检测[J]. 计算机系统应用, 2019, 28(1): 94-99.
- [17] 葛雯, 张雯婷, 孙旭泽. 基于 Jetson TX1 的目标检测系统[J]. 沈阳工业大学学报, 2019, 41(5): 539-543.
- [18] 许喜斌. 结合 R-DAD 和 KCF 的行人目标跟踪改进算法[J]. 智能计算机与应用, 2019, 9(4): 263-266, 270.
- [19] HOWARD A, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. ArXiv abs/1704.04861, 2017.

(收稿日期: 2020-11-11)

作者简介:

祁星晨(1995-), 男, 硕士, 主要研究方向: 深度学习、目标检测。

卓旭升(1967-), 男, 博士, 副教授, 主要研究方向: 智能感知、智能控制等。



扫码下载电子文档

(上接第 39 页)

- [9] 邓勇杰, 文志诚, 姜旭炜. 基于灰色理论和 BP 神经网络安全态势预测[J]. 微型机与应用, 2015, 34(20): 1-3, 8.
- [10] 朱晨飞. 基于神经网络的网络安全态势评估与预测方法研究[D]. 北京: 中国人民公安大学, 2019.
- [11] 陈维鹏, 敖志刚, 郭杰, 等. 基于改进的 BP 神经网络的网络空间态势感知系统安全评估[J]. 计算机科学, 2018, 45(S2): 335-337, 341.
- [12] 郭文忠, 林宗明, 陈国龙. 基于粒子群优化的网络安全态势要素获取[J]. 厦门大学学报(自然科学版), 2009, 48(2): 202-206.
- [13] 李天骐. 基于神经网络的网络安全态势评估与预测技术研究[D]. 北京: 华北电力大学(北京), 2016.
- [14] 朱江, 明月, 王森. 基于深度自编码网络的安全态势要素获取机制[J]. 计算机应用, 2017, 37(3): 771-776.
- [15] 周长建, 司震宇, 邢金阁, 等. 基于 Deep Learning 网络态势感知建模方法研究[J]. 东北农业大学学报, 2013, 44(5): 144-149.
- [16] 俞中华, 杨晓东. 基于深度自编码网络的网络安全态势感知与预警机制[J]. 广播电视网络, 2020, 27(6): 63-65.
- [17] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [18] 乔俊飞, 潘广源, 韩红桂. 一种连续型深度信念网的设计与应用[J]. 自动化学报, 2015, 41(12): 2138-2146.
- [19] LI F C, JIANG Y X, ZHOU D Q. The building model of decision on the core competitive capacity of enterprises and evaluating demonstration[J]. Business Economics and Administration, 2006(6): 42-46.
- [20] 中国国家标准化管理委员会. GB/T20984-2007 信息安全技术信息安全风险评估规范[S]. 北京: 中国标准出版社, 2007.
- [21] 王娟, 张凤荔, 傅翀, 等. 网络态势感知中的指标体系研究[J]. 计算机应用, 2007(8): 1907-1909, 1912.

(收稿日期: 2020-09-11)

作者简介:

熊中浩(1994-), 男, 硕士研究生, 助理工程师, 主要研究方向: 智能算法、优化控制。

张伟(1976-), 男, 硕士研究生, 高级工程师, 主要研究方向: 网络安全。

杨国玉(1980-), 男, 硕士研究生, 高级经济师, 主要研究方向: 信息化与网络安全管理。



扫码下载电子文档

版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所