

## YOLOv3 网络在车标检测中的应用\*

王 林, 黄三丽

(西安理工大学 自动化与信息工程学院, 陕西 西安 710048)

**摘 要:** 针对车标检测存在检测时间长、检测率低、可识别类型少的问题, 提出一种采用 You Only Look Once (YOLOv3) 网络的方法。为了使该网络适用于小目标的车标检测, 将目标特征提取结构 Darknet-53 换成 Darknet-19, 并且将多尺度预测层数减少为两层以减少网络参数数量。同时, 为了增大车标在图像中所占比例, 让卷积神经网络能学习更多的车标特征, 采用将车辆从图像中裁剪后进行人工标注的方法, 构建了一个包含 46 类车标的数据集 (VLDS-46)。实验结果表明, 采用该模型进行车标检测时能在实现高检测率的同时达到实时性要求, 检测平均耗时为 9 ms。

**关键词:** 车标检测; 卷积神经网络; 数据集; Darknet-19

**中图分类号:** TN98

**文献标识码:** A

**DOI:** 10.16157/j.issn.0258-7998.200067

**中文引用格式:** 王林, 黄三丽. YOLOv3 网络在车标检测中的应用[J]. 电子技术应用, 2020, 46(9): 34-37, 42.

**英文引用格式:** Wang Lin, Huang Sanli. Application of YOLOv3 network in vehicle logo detection[J]. Application of Electronic Technique, 2020, 46(9): 34-37, 42.

## Application of YOLOv3 network in vehicle logo detection

Wang Lin, Huang Sanli

(School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China)

**Abstract:** Aiming at the problem that the vehicle logo detection has long detection time, low detection rate and few identifiable types, a method using You Only Look Once (YOLOv3) network is proposed. In order to make the network suitable for vehicle target detection of small targets, the target feature extraction structure Darknet-53 is replaced with Darknet-19 and the multi-scale prediction layer is reduced to two layers to reduce the number of network parameters. At the same time, in order to increase the proportion of the car logo in the image and let the convolutional neural network can learn more car logo features, this paper adopts a method of cutting the vehicle from the image and manually marking it, constructing a class of 46 vehicles Data set (VLDS-46). The experimental results show that when the model is used for vehicle logo detection, the real-time requirement can be achieved while achieving high detection rate, and the average detection time is 9 ms.

**Key words:** vehicle logo detection; convolutional neural network; data set; Darknet-19

## 0 引言

近年来,随着人工智能技术的不断发展,其被广泛用于公共安全、教育、人机交互等领域,例如:人脸检测<sup>[1]</sup>、图像分类<sup>[2]</sup>、图像分割等。同样的车标检测技术对于完善智能交通具有重要作用,能够辅助实现车辆信息的确认。21 世纪以来,随着国民经济快速发展,汽车数量在我国的增长速度越来越快。汽车数量过多带来交通压力的同时也会带来其他问题:套牌车、假车牌事件频频发生。这些事件的发生意味着仅仅通过车牌来确认车辆准确信息将受到阻碍。同车牌一样,车标也是车辆的一个重要属性,在结合车牌信息后,就能更准确地确认车辆信息。国内外学者对车标检测采用了不同的方法进

行研究。一般而言,分为传统机器学习方法和卷积神经网络方法。

采用传统机器学习方法在进行车标检测时繁琐,需要分为两个独立的步骤:首先进行车标定位,然后进行车标识别。文献[3]-[4]采用传统机器学习方法在进行车标定位和识别过程中都需要由专业人员人工设计特定特征,但是这些特征对于光照、阴影、车身倾斜等的变化所表现出来的鲁棒性不强,在应用于实际场景下会出现低检测率和识别率问题。针对传统机器学习方法的不足,卷积神经网络方法应运而生。

早在 20 世纪 80 年代末期,卷积神经网络<sup>[5]</sup>技术就已经应用于视觉任务中,但由于当时设备的运算能力不强和数据资源的不足,这一技术并没有得到研究人员的广泛研究和长足的发展。后来得益于 GPU (Graphics

\* 基金项目:陕西省科技计划重点项目(2017ZDCXL-GY-05-03)

Processing Unit)强大的计算能力和互联网上丰富的数据资源,卷积神经网络技术得到了进一步的发展,而且也取得了突破性的进展。卷积神经网络一般包含卷积层、池化层、全连接层等。借助于卷积神经网络的强表达能力,不需要人工设计特征,在网络学习车标特征后就能自动进行车标的定位以及分类。2014年,GIRSHICK R等<sup>[6]</sup>人第一次将卷积神经网络应用于目标检测上并且获得了比只使用方向梯度直方图<sup>[7]</sup>(Histogram of Oriented Gradient, HOG)特征更高的准确率,该网络称为区域卷积神经网络(Region Convolutional Network, R-CNN)。但该网络存在耗时久的弊端,伴随着对该网络框架的不断改进,其目标检测的精度不断提高,运行时间也不断减少。2015年,HE K<sup>[8]</sup>等人通过改进R-CNN网络缩短其检测时间,但其最大帧率也只能达到5FPS。这对于帧率最小为25FPS的监控视频来说是远远不够的。为了在准确率和时间上得到均衡,YOLO网络<sup>[9]</sup>得到了应用,它将物体检测作为回归问题来求解,在输入图像后经过一次前馈就能得到图像中所有待检测物体的位置和其所属的类别以及相应的置信度,进行目标检测时耗时少。

目前采用卷积神经网络进行车标检测时遇到两个问题:(1)公开的车标数据集少且包含的车标种类少<sup>[10]</sup>;(2)采用卷积神经网络方法时参数多,运行耗时。因此,本文提出的采用改进的YOLOv3<sup>[11]</sup>网络进行车标检测主要有以下两个贡献:(1)替换YOLOv3中的Darknet-53为层数少的Darknet-19<sup>[12]</sup>结构,减少运算量进而减少车标检测时间;(2)构建了一个包含46类车标的数据集。

## 1 网络结构设计

YOLOv3网络可以看成由以下两部分组成:(1)用于提取图像特征的Darknet-53网络结构;(2)多尺度预测目标坐标、置信度、类别等的检测网络。Darknet-53由53层卷积层构成,由于车标是小目标,如果直接将网络应用于车标检测,将无法学习到车标的准确特征。YOLOv3的检测网络部分由3层多尺度预测层构成,对于检测小目标而言,特征提取网络层数较少,因而可以减少多尺度预测层的层数,进而减少网络参数量,提高检测速度。

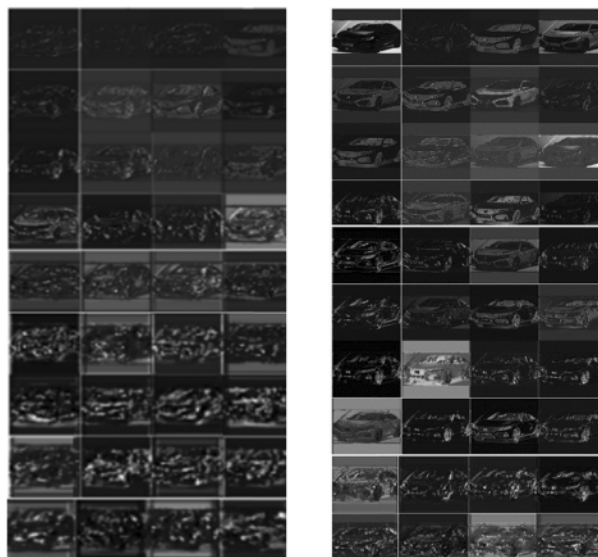
### 1.1 特征提取网络

卷积网络提取的图像特征的好坏决定最终目标检测结果优劣。在应用Darknet-53进行图像分类时,分类准确率优于Vgg16<sup>[13]</sup>和Resnet18<sup>[14]</sup>网络,而且在FLOPs(Floating-Point Operations Per Second)上,前者更小,这就得出Darknet-53网络在运行速度更快的同时也能保证高准确率这一结论。对比Resnet101和Resnet155,Darknet-53能获得差不多的Top-1准确率,但是Darknet-53网络结构的运算量却更小。由此可以得出,采用Darknet-53网络结构在提取到高质量目标特征的同时还能减少运算量,使目标检测实时性成为可能。

但由于车标尺寸小,而卷积神经网络在进行卷积层

操作后通常会进行池化,而池化会使特征图像尺寸变小。对于小目标而言,如果特征提取网络层数过多,最终目标尺寸过小会使得特征提取变得困难,将影响模型的准确率。同时由于网络层数的增加,参数数目也会随之增加,使得训练过程中会出现以下问题:(1)训练更耗时;(2)检测识别过程更耗时;(3)实际使用时对设备要求更高。

考虑到上述原因,本文将把YOLOv3网络中的Darknet-53替换成Darknet-19以提高卷积网络学习车标特征的能力。Darknet-19是YOLOv2中所采用的特征提取网络,它由19层卷积层构成,虽然它的卷积层数更少,对于车标这种小目标也能提取到高质量特征。图1分别为Darknet-19各卷积层特征提取图和Darknet-53中部分卷积层特征提取图。



(a) Darknet-19 特征图

(b) Darknet-53 特征图

图1 特征提取图

由图1可以看出,Darknet-53在图像尺寸大时所提取的特征比Darknet-19更加细致,但随着网络层数增加,Darknet-53所提取的车标特征变少,特征图变得模糊,在第10层卷积层之后的特征图基本无法分辨车标特征,以至于之后的卷积层对模型准确率的提高产生不了作用。虽然Darknet-19网络学习到的特征没有Darknet-53多,但其网络层数少,在模型训练和检测识别时耗时更少,同时也能保证一定的准确率。综合考虑网络参数数量和学习特征能力,本文采用Darknet-19网络结构作为主体特征提取网络。当输入图像尺寸为480×480时,经过最后一个池化层后,特征图大小变成15×15。

### 1.2 检测网络

YOLOv3采用Achoor Boxes定位目标的思想来源于Faster R-CNN,但并不是直接通过网络预测目标的4个点的坐标,而是通过与人为确定的Anchor Boxes的偏移

量调整位置,这种方法可以简化目标定位问题进而让模型在训练时能更快收敛并且训练过程中更稳定。对比YOLOv2只有一层多尺度预测层,YOLOv3采用了3层多尺度预测层来提高模型对小目标的检测率。因为本文中采用的特征提取网络是Darknet-19,其卷积层数不多,为了减少网络的模型的大小来加快检测速度,本文将检测网络中的多尺度预测层减少为两层。实验结果表明,虽然减少了预测层的数量,但是模型的平均Intersection Over Union(IOU)值并没有受到多大影响,但检测时间却有所缩短。同时,为了适应车标检测这个任务,通过对VLDS-46这个数据集分别进行 $k=6$ 和 $k=9$ 的K-Means聚类,当 $k=6$ 时,平均IOU的值最大。其聚类结果为(28,40)、(54,51)、(81,53)、(85,79)、(121,106)、(133,64)。

## 2 实验与分析

该部分主要由数据集构建和处理、模型训练和测试两部分组成。其中为了提高车标检测率,对数据集进行了特殊处理,同时本文自建了一个包含46类车标的数据集。

### 2.1 数据集

#### 2.1.1 数据处理

YOLOv3网络在训练模型时会将图像按照原图长宽比缩放到特定尺寸大小,例如 $480 \times 480$ ,然后输入到卷积神经网络中学习。而车标在整个图像中所占比例小,为了让卷积网络能学习到车标更多特征,如图2所示,本文采用先对图像进行裁剪,然后把包含单一车标的车辆图像输入网络中训练的方法。这样处理使得车标所占比例值有所提高,进而提高最终训练模型的准确率。



图2 样本处理

在图2中原始图像尺寸为 $800 \times 600$ ,而车标尺寸仅为 $28 \times 28$ ,占整体图像比例为0.00163。如果将整个图像输入网络训练模型,会导致网络无法学习到更多车标的特征。而裁剪后的图像尺寸为 $560 \times 269$ ,车标尺寸占图像比例为0.0052,为原始图像时的3.19倍,虽然从该扩大倍数上看没有得到很大变化,但如果车辆占图像比很小,那这种做法就会带来明显的益处。

#### 2.1.2 数据获取

对于车标,公开的数据集少或者所包含的车标类别不多。所以本文从现有的公开数据集(如COCO数据集)中将车辆图像筛选出来。同时,通过网络爬虫技术从互联网

爬取大量的车辆图像,考虑到所爬取的图像大都为高分辨率图像,为了让数据集中的图像尺寸等多样化,从视频监控下的图像中选取了部分图像来进行数据集的扩充。首先使用YOLOv3现有的模型进行车辆检测,将图像中的车辆单独检测并且裁剪好以获得车辆数据集。然后使用标签标注应用LabelImg对裁剪好的车辆进行车标位置的标注,获得车标标签集。构建的VLDS-46数据集总共包含46种车标合计62920张图像。将其分为训练集和测试集,其中训练集51615张,测试集11305张。表1为每种车标类别在训练集和测试集中所包含的数目。

表1 训练集和测试集中不同车标数量

类别	训练集	测试集	类别	训练集	测试集
奥迪	1123	240	捷豹	890	240
本田	1200	240	江淮	1200	245
奔驰	1134	245	凯迪拉克	1240	240
宝马	1220	250	雷克萨斯	1100	240
别克	1200	240	路虎	1200	260
比亚迪	1100	240	铃木	1200	250
宝骏	1090	260	马自达	1100	250
保时捷	980	250	名爵	1240	245
标致	1128	240	MINI	1000	250
奔腾	1100	250	起亚	1120	260
长安汽车	1200	240	奇瑞	1200	240
长安商用	1110	240	启辰	980	240
大众	1380	250	日产	1320	250
东风	1200	240	荣威	1000	260
丰田	1120	240	三菱	1200	240
福特	1100	240	斯巴鲁	1100	260
广汽传祺	1000	250	Smart	980	240
哈弗	1000	240	特斯拉	900	240
海马	1090	240	五菱宏光	1230	240
吉利	1200	240	沃尔沃	1200	260
雪佛兰	1100	240	雪铁龙	1200	240
Jeep	960	240	现代	1200	240

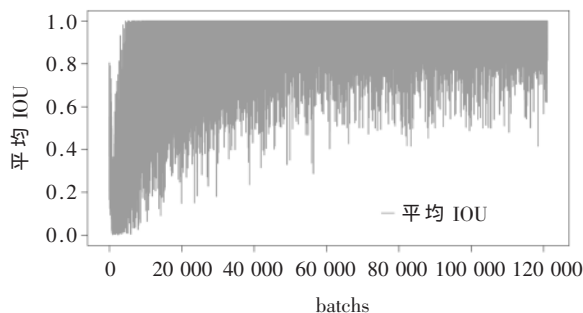
## 2.2 实验结果及分析

### 2.2.1 模型训练

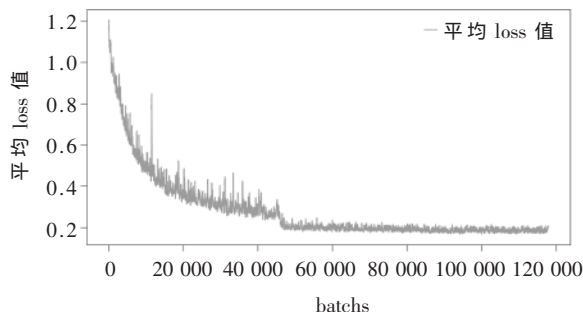
采用Pytorch深度学习框架搭建车标检测网络。在操作系统为Centos7.0、有4块gtx1080Ti GPU的工作站上进行模型的训练与测试。通过不断调整网络超参数,如图3和图4所示,分别为采用Darknet-19和Darknet-53作为特征提取网络,图像输入尺寸为 $480 \times 480$ 时训练过程中损失函数值图和平均IOU值图。

为了对比卷积神经网络层数对车标检测模型训练过程中的影响,分别采用Darknet-19和Darknet-53结构进行训练。loss代表模型在训练过程中损失函数值,batches代表整个模型在训练过程中输入样本的批量数。两种模型在训练过程中均对训练集进行了190 epoch,训练过程中主要的超参数设置为batch=128(batch代表每次训练时向网络输入样本数目),subdivision=32;为了



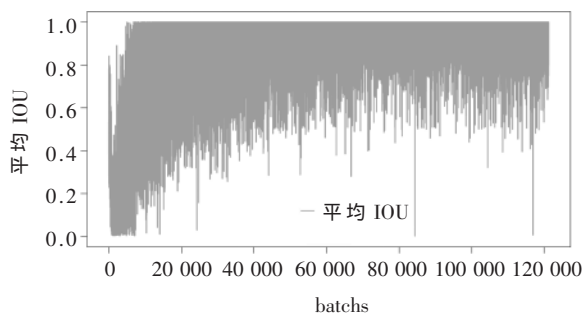


(a) 平均 IOU 曲线

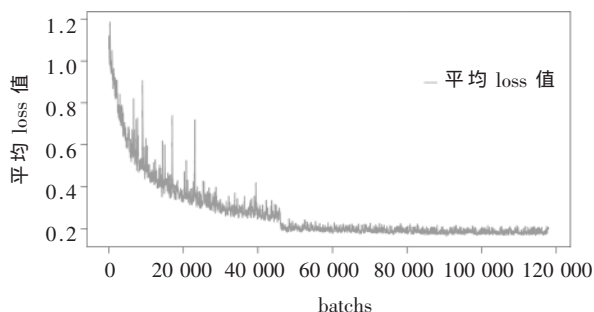


(b) 平均 loss 值曲线

图 3 基于 Darknet-19 的 IOU 和 loss 值图像



(a) 平均 IOU 曲线



(b) 平均 loss 值曲线

图 4 基于 Darknet-53 的 IOU 和 loss 值图像

加快模型收敛速度,将初始 Learning\_rate 设置为 0.02。对比两者的最终平均 loss 值发现 Darknet-19 网络略低,两者分别为 0.254 和 0.265。由两个网络的 IOU 图像可以看出,因为 Darknet-19 网络卷积层数更少,模型在训练学习过程中更稳定,表现为训练时平均 IOU 值更高,说

明定位更加准确。

### 2.2.2 模型测试

模型的测试在 VLDS-46 中进行,对于 Darknet-19 网络下的模型,对图像进行车标检测所需要的时间约为 9 ms;而对于 Darknet-53 网络下的模型,检测时间约为 27 ms。基于 Darknet-19 网络的模型在测试集中测试的最终平均测试准确率可达到 97.81%。表 2 为部分车标在测试集中的测试结果。

表 2 部分车标实验结果

车标	正确个数	错误个数	检测率/%
本田	235	5	97.91
现代	234	6	97.5
大众	250	0	100
Jeep	237	3	98.75
雪佛兰	236	4	98.33
路虎	250	10	96.15
宝马	250	0	100
海马	232	8	96.67
马自达	243	7	97.2
广汽传祺	248	2	99.2

由于车标之间的相似性以及训练样本数量分布不均,部分车标的检测准确率略低于平均检测准确率。由表 2 可以看出,在实际检测时由于车身所处的角度不同,本田和现代这两种车标会出现误检,例如会将本田错误识别为现代;路虎车分为多个系列,每个系列有不同的车标图案且区别较大,由于数据集中将多个系列的车标归类于同一种,因此测试时错误个数偏多,检测率低;对于海马和马自达两种车标,相似度高,所以检测时两者会出现互相误识别的情况,导致错误个数偏多;对于大众、宝马、广汽传祺等车标的检测率较高。

### 2.2.3 部分结果图

图 5 所示为采用改进后的 YOLOv3 网络进行车标检测的结果图。由结果可以看出,对于不同角度、不同尺寸大小的图像,该模型都能够实现准确检测。

## 3 结束语

本文提出了一种采用 Darknet-19 为特征提取网络的车标检测方法并且自建了一个 VLDS-46 数据集,为了让网络学习到更多特征,采用将车辆单独输入网络中训练的方法,这可以减少背景的影响,该方法在大大缩减时间的同时保证了较高检测率。由于本文中的车标种类多,因此对于车标相似度高的情况下容易出现误检。通过分析模型训练过程中的日志发现整体的训练 loss 值偏高、平均 IOU 偏低。接下来的工作将通过优化网络解决高相似度车标之间的错误识别问题并且提高 IOU 值,同时会结合一个耗时少的车辆检测网络在实现车辆检测的同时完成车标检测,并且达到实时性。

(下转第 42 页)

使用基于注意力机制的特征增强模块时,实验(3)获得了3.8 mAP的性能提升。当模型额外使用自定义 anchor 配置时,实验(4)获得了最高性能 90.5 mAP。这说明本文提出的模型能够很好地满足煤矿场景下的安全帽佩戴检测需求。

### 3 结论

本文提出了一种基于 Darknet53 的安全帽佩戴检测模型,引入了基于注意力机制的特征提取模块用于增强模型的信息传播。此外,本文制定了安全帽佩戴预训练数据集和实际矿井场景数据集,用于训练模型和测试模型在不同场景下的有效性。实验结果表明提出模型在实际矿井场景下有着优异的预测性能。

(下转第 46 页)

(上接第 37 页)



图 5 车标检测结果

### 参考文献

- [1] 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.
- [2] 张璘, 杨丰琦. 基于深度学习的图像分类搜索系统[J]. 电子技术应用, 2019, 45(12): 51-55.
- [3] 李侠, 李文举, 孙娟红, 等. 基于模板匹配和边缘检测的车标定位方法[J]. 计算机工程与设计, 2011, 32(10): 3453-3456.
- [4] WANG Y, LIU Z, XIAO F. A fast coarse-to-fine vehicle logo detection and re-cognition method[C]. 2007 IEEE International Conference on Robotics and Biomimetics(ROBIO). IEEE, 2007: 691-696.
- [5] FUKUSHIMA K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. Biological-Cybernetics, 1980, 36(4): 193-202.
- [6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic

### 参考文献

- [1] RAWAT W, WANG Z. Deep convolutional neural networks for image classification: a comprehensive review[J]. Neural Computation, 2017, 29(9): 2352-2449.
- [2] 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.
- [3] 张杰, 隋阳, 李强, 等. 基于卷积神经网络的火灾视频图像检测[J]. 电子技术应用, 2019, 45(4): 34-38, 44.
- [4] 黄海新, 梁志旭, 张东. 基于深度学习的图像风格化算法研究综述[J]. 电子技术应用, 2019, 45(7): 27-31.
- [5] ZHAO Z Q, ZHENG P, XU S T, et al. Object detection segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE Computer Society, 2014: 580-587.
- [7] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]. 2005 IEEE Computer Society Conference on CVPR, 2005.
- [8] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]. Advances in Neural Information Processing Systems, 2015: 91-99.
- [9] 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, 2016: 779-788.
- [10] 张力, 张洞明, 郑宏. 基于联合层特征的卷积神经网络在车标识别中的应用[J]. 计算机应用, 2016, 36(2): 444-448.
- [11] REDMON J, FARHADI A. Yolov3: an incremental improvement[J]. arXiv Preprint arXiv: 1804.02767, 2018.
- [12] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [13] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv: 1409.1556, 2014.
- [14] 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, 2016: 770-778.

(收稿日期: 2020-02-05)

### 作者简介:

王林(1963-), 男, 博士, 教授, 主要研究方向: 复杂网络、社团分析、深度学习、数据挖掘等。

黄三丽(1993-), 女, 硕士研究生, 主要研究方向: 深度学习。

## 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