

基于 Involution Prediction Head 的小目标检测算法

安鹤男¹, 邓武才¹, 管 聪², 姜邦彦²

(1. 深圳大学 电子与信息工程学院, 广东 深圳 518000; 2. 深圳大学 微纳光电子学研究院, 广东 深圳 518000)

摘 要: 针对通用目标检测算法在检测小目标时存在错检和漏检等问题, 提出了一种小目标检测算法 IPH (Involution Prediction Head), 将其运用在 YOLOv4 和 YOLOv5 的检测头部分, 在 VOC2007 数据集上的实验结果表明, 运用 IPH 后的 YOLOv4 小目标检测精度 APs (AP for small objects) 相比原始算法提升了 1.1%, 在 YOLOv5 上的 APs 更是提升了 5.9%。经智能交通检测数据集进一步检验, IPH 算法和去下采样能有效提升小目标检测精度, 减少错检和漏检的情况。

关键词: YOLOv4; IPH; 小目标检测; 特征提取; 注意力机制

中图分类号: TP391.4

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.223161

中文引用格式: 安鹤男, 邓武才, 管聪, 等. 基于 Involution Prediction Head 的小目标检测算法[J]. 电子技术应用, 2022, 48(11): 19-23.

英文引用格式: An Henan, Deng Wucui, Guan Cong, et al. Small object detection algorithm based on involution prediction head[J]. Application of Electronic Technique, 2022, 48(11): 19-23.

Small object detection algorithm based on involution prediction head

An Henan¹, Deng Wucui¹, Guan Cong², Jiang Bangyan²

(1. College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518000, China;

2. Institute of Microscale Optoelectronics, Shenzhen University, Shenzhen 518000, China)

Abstract: Aiming at the problems of false positive detection and low recall in the detection of small targets by the general target detection algorithm, a small target detection algorithm IPH (involution prediction head) is proposed, which is applied to the detection head of YOLOv4 and YOLOv5. The experimental results on the VOC2007 data set show that the detection accuracy APs (AP for small objects) of YOLOv4 after using IPH is improved by 1.1% compared with the original algorithm, and the APs on YOLOv5 is improved by 5.9%. Through further verification of the intelligent traffic detection data set, IPH algorithm and downsampling can effectively improve the accuracy of small object detection and reduce false positive detection and missed detection.

Key words: YOLOv4; involution prediction head; small object detection; feature extraction; attention module

0 引言

目标检测是计算机视觉领域的一项基本任务。随着硬件 GPU 以及卷积神经网络 (Convolution Neural Networks, CNN) 的不断发展, 目标检测取得了显著的发展, 已广泛应用于自动驾驶、视觉搜索、虚拟现实、增强现实等许多应用领域^[1]。目前大多数最先进的目标检测深度学习算法都是基于 CNN, 主要分为两大类: 两阶段 (Two-stage) 目标检测器和单阶段 (One-stage) 目标检测器。Two-stage 检测器的典型网络有 fast R-CNN^[2]、mask R-CNN^[3] 和 faster R-CNN^[4]。卷积神经网络将目标检测的过程分为两个步骤, 先使用区域生成网络 (Region Proposal Networks, RPN) 生成稀疏的候选锚框, 然后检测对象的位置和类别。而 One-stage 算法则是端到端的目标检测方法, 这类算法直接用一个 CNN 网络预测目标的位置和类别, 不需要 RPN 网络来生成锚框, 因而检测速度更快。One-

stage 检测器的典型网络有 SSD^[5] 和 YOLO^[6-9] 系列模型。

小目标检测广泛存在于目标密集图像和远距离成像目标图像中, 作为目标检测的一部分, 在现实中有着同样重要的需求。无论是车牌号检测、焊缝图像检测, 还是无人机航拍图像, 许多场景中都有小目标的存在。但由于小目标像素信息占比小、纹理特征不明显, 小目标的检测比大中型目标的检测更为困难。由此可知, 小目标检测是目标检测研究领域具有重要性和挑战性的研究方向。

1 相关工作

小目标检测任务不仅要对象实例进行正确分类, 还需要准确定位, 这时局部细粒度信息起到十分重要的作用。一些工作显示了结合不同特征层的重要性, 而另一些研究显示了上下文信息是非常有用的。在 One-stage 检测器中 YOLOv4 和 YOLOv5 网络凭借着高精度和

检测速度快的优点而被广泛应用。YOLOv4 和 YOLOv5 是在 YOLOv3 的基础之上不断优化和改进的版本,它们结合了大量前人的研究,在主干特征提取网络、特征金字塔、数据增强和激活函数等方面进行了改进,通过适当组合这些技巧的消融研究,证明无论是速度还是精度上都有所提升,可以在传统的 GPU 上进行训练和测试,获得实时的高精度结果。

尽管 YOLOv4 和 YOLOv5 在许多应用中已经实现了对大中型物体的精准检测,但对于小目标检测效果不佳,仍有提升的空间。特征提取过程中多次进行下采样操作,图像分辨率会逐渐降低,导致小物体的消失和边缘信息变得粗糙,不利于小目标的检测。再加上小目标本身特征难以区分、分辨率低、背景复杂、上下文信息有限等原因,目前包括 YOLOv4 和 YOLOv5 在内的通用目标检测算法对小目标的检测效果都很差。

针对通用目标检测算法在检测小目标时检测精度低,存在错检和漏检等问题,本文在相关文献^[10-14]的基础上,基于 Involution^[14]算子和 SimAM^[13]注意力机制,提出了一种新型的小目标检测头 IPH(Involution Prediction Head)模型,并将其运用到 YOLOv4 和 YOLOv5 网络的检测头部分。为了防止下采样导致特征信息丢失,还对 PANet 上下文特征融合结构进行修改,去除部分下采样操作。在 VOC2007 数据集和智能交通检测数据集上进行了实验,证明本文所提方法均能提高小目标检测的 APs,有效减少 YOLOv4 和 YOLOv5 错检漏检的情况。

2 提出方法

本文基于 YOLOv4 检测头,融合了新型 Involution 算子和 SimAM 无参注意力机制,提出了一种新型的小目标检测头 IPH 模型。将本文提出的 IPH 算法分别运用到 YOLOv4 和 YOLOv5 网络的检测头部分,与现有的 YOLOv4、YOLOv5 以及 Retinanet^[15]的性能指标进行了对比。同时,为了防止下采样导致小目标边缘信息丢失,在 YOLOv4 网络的基础上,去除 PANet 网络的两层下采样操作。在 VOC 公共数据集与智能交通检测数据集上进行实验,验证所提算法的可行性。

2.1 新型小目标检测头 IPH 模型

目前大多数目标检测算法都是基于 CNN,用卷积来对图像进行特征提取。文献^[14]提出来的 Involution 新型算子参考卷积算子,设计一种与常规卷积核性质相反的算子,该算子的示意图如图 1 所示。跟卷积的空间权值共享和通道特异性的性质不同,Involution 具有通道权值共享和空间特异性的特点。在特征层进入检测头后,先用一层 Involution 操作进行加强特征提取再进行后面的卷积操作。Involution 层的操作如图 1 所示,利用特征图中当前位置的 $1 \times 1 \times C$ 向量与 K^2 个 $1 \times 1 \times C$ 的卷积核进行卷积操作,经过变换得到图 1 中 $K \times K$ 的 Involution 核的权重矩阵,每个像素点对应的 Involution 权重都是不

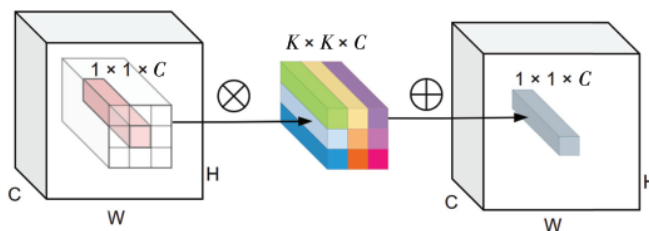


图 1 Involution 示意图

同的,接着使用这些 Involution 核对特征图进行卷积,经过 Involution 层后,特征图的尺寸保持不变。为了平衡精度和参数量,选择 K 为 7 的 Involution 核。

除了使用 Involution 算子,还在检测头部分融合了注意力机制。现有的注意力机制往往是通过子网络来生成注意力权重,中山大学^[13]从神经科学理论出发,提出了一种概念简单且十分有效的注意力模块 SimAM。该模块根据空域抑制的原理,即激活的神经元通常会抑制周围神经元,通过一个能量函数就能为特征图推导出三维注意力权值,能量函数如式(1)所示。

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (1)$$

式中, $\hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i$, $\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2$, $\lambda = 0.0001$ 。

式(1)表明,当神经元与周围神经元的区别越大时能量越低,故每个神经元的重要程度可以通过 $1/e_t^*$ 来获得。该模块仅通过一个能量函数就能为三维的每一个像素点都产生一个对应的权值,属于轻量型的无参注意力模块。将这一注意力模块集成到小目标检测头第一层卷积之后,批量归一化(Batch Normalization, BN)层之前,形成最终的 IPH 小目标检测头模型,结构如图 2 所示。将激活函数换成了 SiLU, SiLU 函数如式(2)所示,曲线如图 3 所示。SiLU 是 Sigmoid 和 ReLU 的改进版,具有无上界有下界、平滑且非单调的特性。SiLU 可看做是平滑的 ReLU 激活函数,在深层模型上的效果要优于 ReLU。

$$f(x) = x \cdot \text{sigmoid}(x) \quad (2)$$

2.2 去下采样

以公共数据集 COCO 的物体定义为例,小目标是指小于 32×32 个像素点的物体。小目标相比于大目标分辨率低,难以提取到有鉴别力的特征,导致误检率高;小目标在图像中的位置过小,网络预测时偏移一个像素则对小目标影响巨大;小目标更容易出现聚集的情况,导致模型漏检严重。现

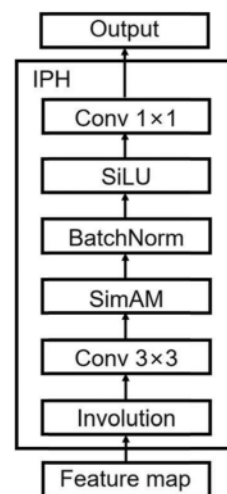


图 2 新型小目标检测头 IPH 结构图

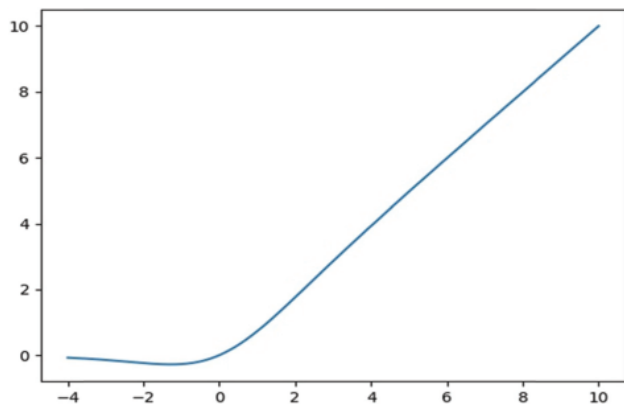


图3 SiLU 激活函数

在检测网络一般都是使用 CNN 作为特征提取工具,为了增大网络的感受野,特征图尺寸不断缩小,面积较小的信息在下采样过程中逐渐丢失,传递不到后面的检测器中。

由于小目标图像信息在整张图片中的占比低,纹理特征不明显,连续进行下采样容易丢失小物体的特征信息,且在之后的特征提取中无法找回。考虑到下采样操作对小目标检测的影响,去除了 YOLOv4 和 YOLOv5 中 PANet 网络部分的两层下采样操作和特征融合过程,并在 VOC2007 和智能交通检测数据集上进行实验,验证了该改进方案的有效性。从图 4 中可以看出,改进后的特征提取网络相较于 YOLOv4 原始网络少了 PANet 右边的两层下采样操作。

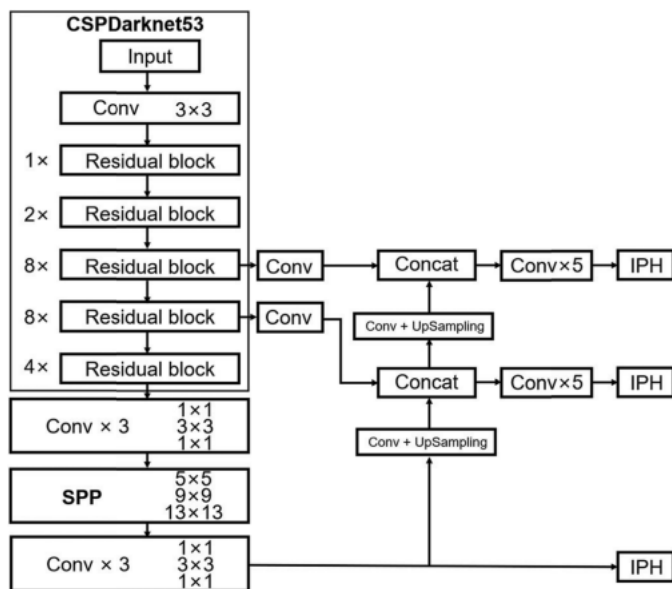


图4 改进后的 YOLOv4 网络框架图

3 实验分析

3.1 训练配置

为了评估网络性能,本文将使用新型 IPH 检测头的网络和原始网络在同一数据集上进行训练,评价指标包

括交并比(Intersection of Union, IoU)为 0.5~0.95、0.5 和 0.75 时的精度均值(Average Precision, AP)以及大、中、小物体的精度均值 AP_L、AP_M、AP_S。训练所使用的服务器配置如表 1 所示,实验中所有模型都从头开始训练, Batch size 设置为 8,初始学习率设为 0.001,优化器使用 Adam,总共进行 300 轮迭代。

表1 服务器配置

服务器配置	
CPU	Intel Xeon E5-2620 v4
GPU	GTX 1080 Ti×4
操作系统	CentOS Linux release 7.6.1810(Core)
操作框架	PyTorch 1.6, Python 3.7, CUDA 10.1

3.2 训练数据集

VOC2007 数据集训练集总共 5 011 张图片,测试集 4 952 张图片,总共有 20 种常见的类别,无论是从目标的类别上还是数据量的规模上都能满足大部分模型的训练要求。智能交通电子提示装置检测数据集是“先导杯”智能交通检测竞赛中新提出的目标检测数据集。该数据集分为三个类,分别为 rect_eye、sphere_eye 和 box_eye 三种类型的摄像头装置,大部分都小于 32×32 像素,信息占比低,且特征不明显,适合用于小目标检测的研究。

3.3 实验结果

3.3.1 VOC2007 数据集结果

使用本文提出的 IPH 网络结构对 YOLOv4 以及 YOLOv5 的检测头进行了更改,在 VOC2007 数据集上进行了实验,不使用预训练模型,从头开始训练 300 轮,最终的结果如表 2 所示,训练过程损失曲线如图 5 所示。由表 2 的结果可知,本文提出的 IPH 网络结构可以直接替换传统的目标检测头,提升通用目标检测网络的小目标检测精度。现有的 YOLOv4 和 Retinanet 网络小目标检测精度 AP_S 只有 7% 左右,而加了 IPH 的 YOLOv4 小目标检测精度比原网络提升了 1.1%。YOLOv5 使用了 Anchor Free 等技巧,对小物体比较友好,原网络 AP_S 达到了 13.4%,在 YOLOv5 上使用 IPH 后 AP_S 达到了 19.3%,相较于原始网络的 AP_S 提升了 5.9%,可知本文提出的 IPH 结构适用于小目标检测,可以有效提升小目标检测的精度。由损失曲线可知,IPH 模型并没有影响网络的收敛速度, Loss 值反而比原模型更小,说明 IPH 模型具有更好的鲁棒性与识别精度,更少出现误检和漏检的情况。对 YOLOv4 和 YOLOv5 进行了去下采样操作,在 VOC2007 数据集上进行了消融实验,从表 2 中的结果可知,去下采样操作同样可以有效提升小目标检测的精度。

3.3.2 智能交通检测数据集结果

表 3 列出了在智能交通电子提示装置检测数据集中 IPH 和去下采样改进对精度均值影响的消融研究结果。通过表 3 的实验结果表明,本文提出的新型 YOLO

表 2 不同网络在 VOC2007 测试集上的结果

网络结构	AP@[.5:.95]	AP@0.5	AP@0.75	APs	APm	APl
Retinanet	36.6%	66.0%	36.6%	7.0%	23.6%	44.7%
YOLOv4	21.9%	47.6%	15.1%	7.1%	14.8%	24.6%
YOLOv4-IPH	22.5%	48.5%	15.7%	8.2%	15.9%	25.3%
YOLOv4-去下采样	20.9%	46.0%	16.2%	8.3%	13.5%	22.4%
YOLOv5	24.7%	45.4%	23.2%	13.4%	22.9%	26.2%
YOLOv5-IPH	27.6%	51.2%	25.9%	19.3%	23.4%	29.8%
YOLOv5-去下采样	26.0%	50.9%	23.1%	15.5%	23.0%	27.9%

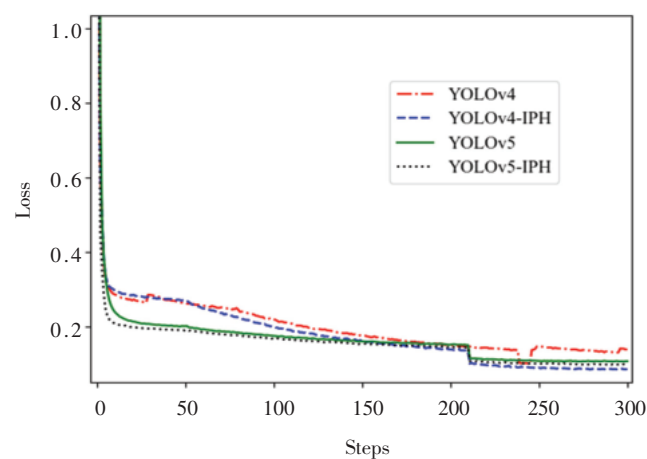


图 5 VOC 数据集训练损失曲线

小目标检测头在 YOLOv4 原网络的基础上提升了小目标检测的精度。在使用去下采样操作后,小目标检测的精度再一次得到提升,最终网络为 YOLOv4+IPH+去下采样,APs 值比原始 YOLOv4 提升了 2.8%。

分别使用最终改进的网络和原始 YOLOv4 算法对数据集图片进行检测,对比结果如图 6 所示。第一排为 YOLOv4 算法的检测结果,第二排为本文最终改进算法的检测结果,可以看到本文提出的网络找出了许多 YOLOv4 原方法漏检的目标。从两组对比图可知,YOLOv4 无法检测出远处的摄像头目标,而本文提出的网络则能将原网络漏框的目标检测出来,且检测框的置信度较高。对于第二组图片左上角的 rect_eye 摄像头,原方法将摄像头的支架错误地检测成了 sphere_eye 摄像头,而改进的方法则没有出现这种情况。由图 6 的检测结果可知,改进

后的算法可以有效地改善小目标漏检以及错检的情况。

4 结论

在基于深度学习的目标检测任务中,小目标由于像素信息占比小、特征不明显,很难被检测到。为了提高通用目标检测器在小目标检测上的精度,本文基于 YOLOv4 模型,融合 Involution 算子和 SimAM 注意力机制,提出了一种新型的小目标检测头 IPH 模型,并去除 PANet 网络的下采样操作,在 VOC2007 数据集上的结果表明本文算法对比目前主流的目标检测算法能有效提升小目标检测精度,比传统检测头更适合用于小目标检测。通过在智能交通检测数据集上的进一步消融实验,证明了 IPH 算法和去下采样两种改进方案均能提升小目标检测的精度,解决原方法错检及漏检严重的问题。

参考文献

[1] LIU Y, SUN P, WERGELES N, et al. A survey and performance evaluation of deep learning methods for small object detection[J]. Expert Systems with Applications, 2021, 172(4): 114602.

[2] GIRSHICK R. Fast R-CNN[J]. arXiv: 1504.08083, 2015.

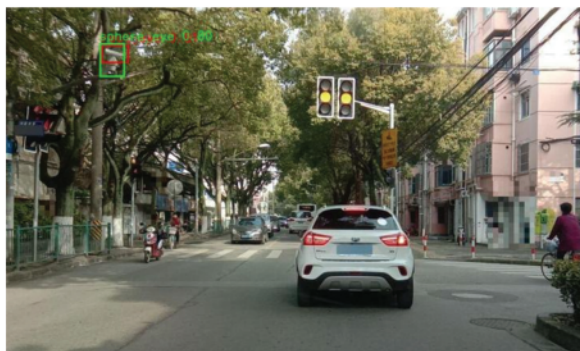
[3] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386-397.

[4] REN S Q, HE K M, 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, 2017, 39(6): 1137-1149.

[5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]// European Conference on Computer Vision, 2016, 9905: 21-37.

表 3 不同网络在智能交通检测数据集上的结果

网络结构	AP@[.5:.95]	AP@0.5	AP@0.75	APs	APm	APl
YOLOv5	12.1%	36.1%	4.9%	11.6%	31.2%	-
YOLOv5-IPH	12.4%	38.1%	4.2%	12.4%	35.3%	-
YOLOv5-去下采样	13.9%	38.6%	5.0%	12.8%	32.6%	-
YOLOv4	10.2%	30.6%	4.1%	10.1%	28.8%	-
YOLOv4-IPH	11.6%	34.1%	4.4%	11.2%	35.1%	-
YOLOv4-去下采样	11.8%	35.0%	4.5%	12.4%	31.6%	-
最终网络	13.2%	36.6%	5.3%	12.9%	39.3%	-



(a) 原方法检测结果



(b) 改进方法检测结果

图 6 不同方法检测效果对比

- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [7] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 6517-6525.
- [8] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv: 1804.02767, 2018.
- [9] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: optimal speed and accuracy of object detection[C]//CVPR, 2020.
- [10] 程松, 戴金涛, 杨洪刚, 等. 基于改进型 YOLOv4 的焊缝图像检测与识别[J]. 激光与光电子学进展, 2022, 59(16): 1610002.
- [11] CHENG X B, QIU G H, JIANG Y, et al. An improved small object detection method based on Yolo V3[J]. Pattern Analysis and Applications, 2021, 24: 1347-1355.
- [12] DAI X, CHEN Y, XIAO B, et al. Dynamic head: unifying object detection heads with attentions[C]//CVPR, 2021.

- [13] YANG L, ZHANG R Y, LI L, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks[C]//International Conference on Machine Learning. PMLR, 2021.
- [14] LI D, HU J, WANG C, et al. Involution: inverting the inheritance of convolution for visual recognition[C]//CVPR, 2021.
- [15] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42(2): 318-327.

(收稿日期: 2022-07-07)

作者简介:

安鹤男(1963-), 男, 副教授, 硕士生导师, 主要研究方向: 计算机视觉、图像处理。

邓武才(1996-), 通信作者, 男, 硕士, 主要研究方向: 计算机视觉、图像处理, E-mail: 2721569802@qq.com。

管聪(1998-), 男, 硕士, 主要研究方向: 计算机视觉、图像处理。



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