

DOI: 10.19650/j.cnki.cjsi.J2311860

基于形变卷积和深层聚合网络的水下文物检测*

周道先¹, 张吟龙^{2,3,4}, 徐高飞⁵, 杨雨沱^{2,3,4}, 梁 炜^{2,3,4}

(1. 沈阳理工大学自动化与电气工程学院 沈阳 110168; 2. 中国科学院沈阳自动化研究所机器人学国家重点实验室 沈阳 110016; 3. 中国科学院网络化控制系统重点实验室 沈阳 110016; 4. 中国科学院机器人与智能制造创新研究院 沈阳 110169; 5. 中国科学院深海科学与工程研究所 三亚 572000)

摘要:搭载有视觉检测系统的自主水下航行器(AUV)具有水下文物探测功能,对深海考古有着重要意义。水下文物所处环境复杂多变,目标存在破损、堆叠和泥沙掩埋等情况,导致判别特征提取困难,使得AUV视觉检测系统无法可靠、准确地实现水下文物的检测。针对上述问题,提出一种基于可形变深层聚合网络模型的水下文物检测算法。为了充分提取复杂环境下水下文物目标特征信息,设计了具有可形变卷积层的多尺度深层聚合网络。在此基础上,引入SimAM注意力模型进行特征优化,来增强文物目标潜在特征信息并削弱背景干扰。最后,通过不同尺度的特征融合实现水下文物检测。在采集的水下文物数据集上进行大量验证和分析,算法的精确率、召回率和平均精度均值(mAP)分别达到了92.7%、90.5%和92.2%。此外,算法已部署到AUV系统中。在实际深海测试场景中,视觉检测系统的文物检测帧率达到19 fps,可满足实时检测的任务需求。

关键词: 自主水下航行器;水下目标检测;可形变卷积;多尺度深层聚合;注意力机制

中图分类号: TP391.4 TH701 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Underwater cultural artifact detection based on deformable convolution and deep layer aggregation network

Zhou Daoxian¹, Zhang Yinlong^{2,3,4}, Xu Gaofer⁵, Yang Yutuo^{2,3,4}, Liang Wei^{2,3,4}

(1. School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang 110168, China; 2. State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China; 3. Key Laboratory of Networked Control System, Chinese Academy of Sciences, Shenyang 110016, China; 4. Institute of Robotics and Intelligent Manufacturing Innovation, Chinese Academy of Sciences, Shenyang 110169, China; 5. Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya 572000, China)

Abstract: Autonomous underwater vehicles (AUVs) equipped with visual detection systems are capable to detect underwater artifacts, which is of great significance to deep-sea archaeology. Underwater cultural artifacts are situated within a dynamic and intricate environment, where the target objects are often fragmented, layered and concealed under sediment. These conditions pose significant challenges to the effective extraction of discernible features, thereby impeding the capability of AUV for reliable and accurate detection of underwater artifacts. In order to solve these problems, this paper proposes an underwater cultural artifact detection algorithm based on the deformable deep aggregation network model. To fully extract the target feature information of underwater cultural artifact in complex environments, this paper designs a multi-scale deep aggregation network with deformable convolutional layers. Besides, the SimAM attention module is designed for the feature optimization, which enhances the potential feature information of cultural artifact target and weakens the background interference information simultaneously. Finally, the prediction of cultural artifact is achieved through fusing feature at different scales. The proposed algorithm has been extensively validated and analyzed on the collected underwater artifact datasets, and the precision, recall, and mAP of algorithm are 92.7%, 90.5%, and 92.2%, respectively. Additionally, the proposed

收稿日期: 2023-09-01 Received Date: 2023-09-01

* 基金项目: 中国科学院对外合作重点项目(173321KYSB20200002)、国家自然科学基金(62273332)、中国科学院青年创新促进会会员(2022201)、广东省基础与应用基础研究基金(2023A1515011363)项目资助

algorithm has been deployed to the AUV system, specifically the artifact detection frame rate of visual detection system reaches 19 fps in the actual deep-sea test scenario and this can satisfy the real-time detection task.

Keywords: autonomous underwater vehicle; underwater target detection; deformable convolution; multi-scale deep layer aggregation; attention mechanism

0 引 言

海洋是地球表面最广阔的区域,对人类文明传播和发展至关重要。在漫长的历史演进过程中,由于航海技术和海洋气象条件的限制,许多船只在海上沉没,船舶和货物散落海底,沉睡千年^[1]。这些海底文物蕴含着丰富的历史、文化和技术信息,对研究人类文明和促进社会发展具有重要意义,因此海底文物探测成为国内外关注的焦点^[2]。

自主水下航行器 (autonomous underwater vehicle, AUV) 是海底探测中常用的装备,相比载人潜水器 (human occupied vehicle, HOV) 和遥控操作水下机器人 (remotely operated vehicle, ROV) 具有作业范围广、探测效率高和作业灵活等优势^[3]。AUV 在海底探测任务中携带两种主要作业载荷,侧扫声呐和水下摄像机^[4]。侧扫声呐可用于大范围快速搜索,但其分辨率较低,获取的目标信息有限^[5]。水下摄像机虽探测距离较小,却能提供目标形状、颜色和纹理等丰富信息,适用于近距离精细探测,尤其在水下小尺度目标的探测中应用广泛^[6]。在水下探测作业过程中,如果 AUV 能自主检测所拍摄视频图像中的目标,则能够根据发现目标的位置对航行路径进行重新规划,围绕感兴趣的目標进行更多的拍摄,便于后续海底文物的分析判断^[7]。因此,基于视觉的目标检测方法已成为 AUV 水下文物 (underwater cultural artifacts, UCA) 检测的常用手段。

基于视觉的目标检测算法可分为两类,基于模型的目标检测和基于数据的目标检测^[8]。基于模型的目标检测算法通过提取特定模型特征来完成检测任务。王慧斌等^[9]提出基于区域显著度的水下目标检测算法,通过对显著区域内背景与目标间的纹理差异计算,分割得到目标。Cutter 等^[10]利用 Haar-like 特征和多个级联的分类器实现鱼类目标的检测。Rizzini 等^[11]根据水下图像颜色的均匀性和轮廓的锐度信息检测水下目标。但传统基于模型的目标检测方法主要提取图像底层和中间层次特征,很难捕捉到具有代表性的语义信息,因此算法鲁棒性差,无法应用在真实水下场景中。

近年来随着 GPU 等算力资源的提高,以及深度神经网络模型的快速发展,基于数据的目标检测算法取得了令人鼓舞的检测性能,许多学者也将其应用到水下目标检测场景中。Lei 等^[12]将 Swin Transformer 引入 YOLOv5 的骨干网络,增强了对水下目标的特征提取,使网络具有

了低质量水下图像目标的检测能力。Yan 等^[13]在一阶段目标检测模型中加入卷积块注意力模块 (convolutional block attention module, CBAM),使网络更加专注于目标特征信息,提高了检测精度。杨玉娥等^[14]将主动定位技术应用于水下识别,提高了目标在不同姿态、不同位置下的识别效果。但上述几种方法仍不能充分利用复杂水下环境中目标的特征,当不同尺度水下目标存在遮挡、重叠问题时检测精度不佳,存在漏检和误检的问题。Song 等^[15]提出了 Boosting 区域卷积神经网络 (region convolutional neural networks, R-CNN) 的两阶段水下目标检测算法,通过对水下目标进行不确定性建模和困难样本挖掘,提高了对被遮挡目标的检测效果。Zeng 等^[16]基于生成对抗的思想提出了 Faster R-CNN-AON 网络,通过引入 AON 对抗网络,防止检测网络过拟合,有效提升了整体检测性能。Guo 等^[17]构建了水下图像处理融合算法,并引入到水下目标检测任务中,实现了对水下沉船遗迹的检测。张有波等^[18]对神经网络模型进行微调,通过迁移学习提高了小样本下水下文物的检测精度。上述研究虽然在性能上有了较大提高,但算法参数量大,无法满足 AUV 端水下实时检测的任务要求。

目前水下文物目标检测仍然面临着如下挑战:

1) 水下目标图像成像质量差。由于水对不同波长光线吸收存在差异以及水下光线的散射,水下图像会出现颜色偏差和能见度低的问题^[19]。此外,AUV 水下照明条件不足和 CMOS 成像水平受限,水下图像的成像质量偏低^[20]。

2) 复杂的水下环境导致目标识别困难。水下文物形态各异,并且容易堆积,这使得它们很容易被遗漏和错误检测^[21]。此外,由于水下文物年代久远,文物往往会出现沉积物覆盖、海洋生物附着以及破损和形态残缺等问题,导致视觉检测过程中文物判别性特征提取困难,对文物检测造成严重干扰。

3) 水下目标样本获取困难。不同于大气光学图像,受水下复杂环境的影响和成像设备的限制,目标检测算法的前期研究中难以获得足够多具有相关特征的样本。

针对上述问题,本文提出一种基于可形变深层聚合网络模型的水下文物目标检测算法 (underwater cultural artifacts detection network, UCADN)。设计了专门用于水下文物检测的特征提取网络,通过可形变卷积层和更多跳跃连接的深层聚合结构,增强网络在复杂场景下文物目标特征提取能力。引入了 SimAM 注意力机制,通过特

征优化来增强水下文物的特征并削弱背景冗余信息,提高模型抗干扰能力,并保持模型参数量。搭建一套 AUV 水下视觉检测系统。通过采集大量水下文物目标图像,建立了水下文物数据集,验证了水下文物目标检测算法的准确性和实时性。

1 AUV 水下视觉检测系统

本文构建了一套 AUV 水下视觉检测系统。如图 1 所示,该系统用于采集水下文物目标数据并验证所部署的算法效果。本文算法被集成到嵌入式计算平台中,并在 AUV 上进行部署,将水下摄像机捕获的图像输入到检测系统中进行自主识别和分析,并能输出检测结果。为了便于对海底目标进行拍摄,将摄像机放置在 AUV 的底部位置。

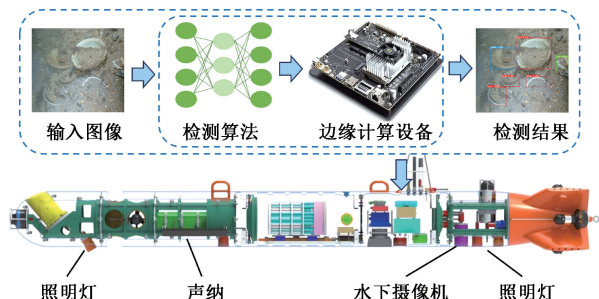


图 1 AUV 水下视觉检测系统

Fig. 1 AUV underwater visual detection system

因此,本文的研究重点在于利用水下摄像机获取光学图像以实现水下文物目标的检测,并在真实的水下环境中对算法性能进行测试。

2 算法框架

本文提出了一种高效轻量的 UCADN,用于考古 AUV 的水下文物目标检测任务。UCADN 结合了可形变卷积模块和注意力机制,以提高复杂水下环境中文物目标检测性能。如图 2 所示,UCADN 由 3 部分构成,即特征提取网络、特征优化网络和特征融合网络。首先,特征提取网络采用了深层聚合结构,融合了可形变卷积层和多跳跃连接。其次,特征优化网络通过引入 SimAM 注意力机制,针对水下文物的关键特征进行增强,减弱无效背景信息。最后,特征融合网络对不同尺度的特征进行融合,进一步增强算法对目标的表示能力。

2.1 特征提取网络

在水下文物目标检测的过程中,普遍存在目标的类型、尺寸、形状和纹理特征的多样性,这增加了检测的难度。传统深度学习模型中,卷积操作具有固定的结构,限制了网络感受野的范围,网络在特征提取过程中只能捕捉局部信息。

然而,由于水下文物出现破损和掩埋等现象,其呈现出不规则的特征。这种情况下,传统卷积操作难以充分提取水下文物特征导致检测失效。为了增强卷积神经网络

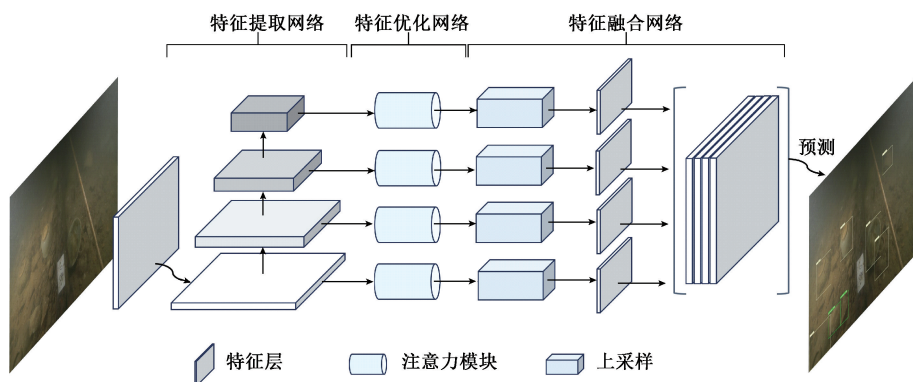


图 2 算法框架

Fig. 2 Algorithm framework

络对水下文物的检测能力,可通过扩大网络的感受野来更好地捕捉长程空间关系,并构建隐式空间模型。在复杂的水下环境中,传统的标准卷积只能进行固定大小采样。相比之下,可形变卷积能够通过引入可学习的偏移量,使其能够动态调整采样位置,更好地适应破损掩埋等目标的形状,从而更好地学习目标特征^[22]。如图 3 所示,可形变卷积模块在传统标准卷积的

基础上为卷积核中的每个采样点增加了一个二维偏移量 $\{\Delta p_n | n = 1, \dots, N\}$, $N = |R|$, 数学定义如下:

$$Y(p_0) = \sum_{p_n \in R} w(p_n) \cdot X(p_0 + p_n + \Delta p_n) \quad (1)$$

式中: X 为输入特征图; R 为 3×3 的卷积核; p_n 为卷积核中的第 n 个点; $w(p_n)$ 为点 p_n 对应权重; p_0 为输入输出特征图上的 p_0 点; Δp_n 为可形变卷积采样点的二维偏移量; Y 为输出特征图。

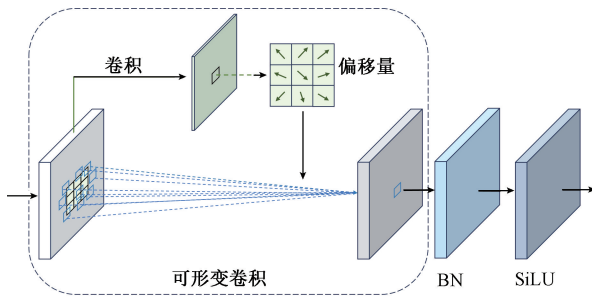


图3 可形变卷积模块

Fig. 3 Deformable convolution module

深层聚合网络 (deep layer aggregation, DLA) 作为一种紧凑而高效的特征提取主干, 在目标检测和语义分割等计算机视觉任务中得到了广泛应用^[23]。DLA 网络以迭代的方式合并分层特征图, 从而实现了在保持较少参数的前提下获得更高精度的目标特征表示。为了适应水下文物环境中多样的目标尺寸和形状特征, 本文对 DLA 网络结构进行了相应的设计, 使其能够输出具有 4 个不同尺度特征层的特征图。在这一基础上, 针对水下环境中存在的不规则文物形态, 本文引入了可形变卷积来取代传统的卷积操作, 以增强网络对于不规则目标的特征提取能力。将特征提取网络命名为基于可形变卷积的多尺度深层聚合网络 (multi-scale deep layer aggregation with deformable convolution network, MDLA-DCN)。该网络在复杂的水下环境中展现出较强的能力, 显著增强了对于具有复杂形态的水下文物目标特征的提取能力。

MDLA-DCN 网络结构如图 4 所示, 主体为 4 个不同分辨率的并行子网络。每个子网络由一系列可形变卷积模块组成, 同一子网络特征图分辨率不随网络深度的变化而改变, 而并行子网络的特征图分辨率依次降低 1/2, 同时通道数量增大 2 倍。MDLA-DCN 网络内部通过上采

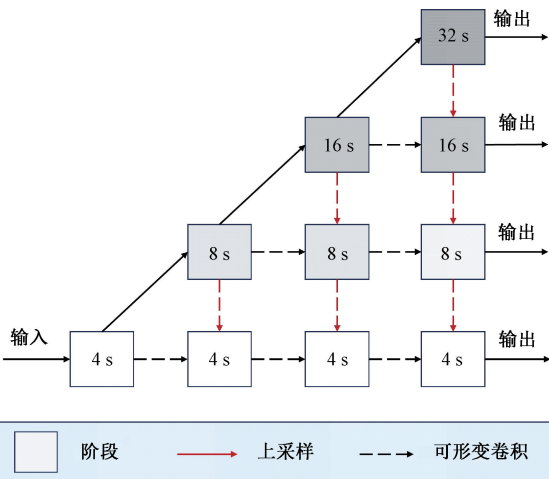


图4 MDLA-DCN 网络

Fig. 4 MDLA-DCN network

样进行跨并行子网络的信息交换, 使每个子网络重复地从其他并行子网络接收信息。网络中多跳跃连接聚合了不同分辨率的特征得到增强的水下文物特征, 该特征在空间信息和语义信息上更加准确。本文为充分利用多尺度特征信息, 将并行子网络产生的 4、8、16、32 倍下采样特征图作为输出。

2.2 特征优化网络

本文的特征优化网络中引入 SimAM 的无参数注意力机制, 提升模型的抗干扰能力^[24]。SimAM 注意力机制具有两个显著的优势。首先, 相较于通道注意力和空间注意力, SimAM 注意力机制能够同时聚焦于通道和空间位置特征, 通过分析特征映射来推断注意力权重。其次, 与基于坐标或三重结构的注意力机制不同, SimAM 注意力机制主要基于能量函数的定义选择进行操作, 避免了过度重组^[25]。这种机制不需要引入额外的参数, 能够更全面、更实用地评估特征权重, 增强水下文物目标的特征表达, 减弱背景冗余信息, 从而提升模型的抗干扰能力。

SimAM 注意力基于视觉神经科学理论, 即拥有更多信息的神经元比其邻近神经元更显著, 并对邻近神经元产生空间抑制。在处理与视觉相关的任务时, 赋予这些承载更关键信息的神经元更高的权重。在水下文物所处场景中, 这些网络神经元负责提取需要被增强的关键文物目标特征。SimAM 注意力机制第 i 个神经元的最小能量函数如下:

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

式中: t 为输入特征图在单个通道上的目标神经元; λ 为

正则项; $\hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i$ 为所有神经元在单个通道上的均值;

$\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2$ 为所有神经元在单个通道上的方差。其中, M 每个通道神经元个数, $M = H \times W$; i 为神经元索引值; x_i 为输入特征图在单个通道上的其他神经元。

式(2)表明能量越低, 水下文物特征图中目标神经元与周围神经元的区别越明显, 这些神经元往往携带更多有效信息并且重要性更高。最终输出的特征图如下:

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{E}\right) \odot X \quad (3)$$

式中: X 表示输入特征; E 表示通道和空间维度上的所有特征。

本文 SimAM 整体结构如图 5 所示, 输入特征图在经过 SimAM 注意力机制后, 通过 sigmoid 函数进行了权值归一化, 将所求的神经元权值和原始特征图的特征相乘, 从而获得最终输出特征图。SimAM 注意力机制使得模型可以更好地聚焦主体目标, 达到更好的检测效果。

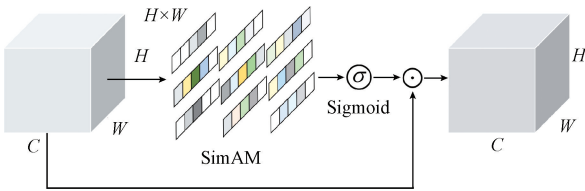


图 5 SimAM 注意力机制

Fig. 5 SimAM attention mechanism

2.3 特征融合网络

经过特征优化网络的处理,获得了不同尺度的特征图,这些特征图用于有效表示水下文物目标的关键特征。为了实现对水下文物的多层次特征提取与融合,本文设计了一种融合网络,用于将深层和浅层特征相互结合。

特征融合网络如图 6 所示。首先,对于由特征优化网络生成的 4 个不同源特征图,本文采用 3×3 卷积操作来进行通道降维,以确保通道数量的一致性,并减轻网络内部的计算负担。随后,对生成的低分辨率特征图进行上采样,使其分辨率与高分辨率特征图保持一致。在上采样过程中,由于转置卷积可以提供网络可学习的参数,有助于提升网络性能,因此本文选择转置卷积进行上采样操作。最后,经过调整后的 4 个特征图通过 Concat 拼接操作,从而产生一个包含多样信息的特征图,并用于最终预测。通过这种多尺度特征融合的方法,能够有效减少小尺度目标特征的丢失,并克服了深层网络中浅层特征在空间位置上被低估利用的问题,从而保障了不同尺度水下文物目标特征的稳健性与可靠性。

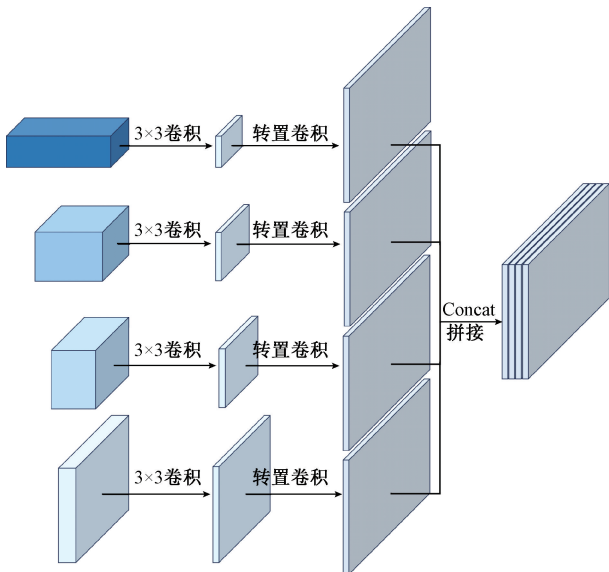


图 6 特征融合网络

Fig. 6 Feature fusion network

3 实验结果与分析

为验证本文所提算法对水下文物目标检测的性能,构建了水下文物数据集,并使用该数据集对算法进行训练和测试。此外,本算法与其他主流检测算法进行了对比,验证本文提出的算法在复杂水下环境对水下文物目标表现出良好的检测性能。

3.1 数据集构建

本文所采用的数据集是通过考古 AUV 搭载的高清水下摄像机拍摄真实海底文物遗址获得。鉴于水下环境的复杂性,该数据集涵盖了多种情景,包括低光照、目标堆叠、目标掩埋和目标破损等。在经过人工筛选、去重和质量评估后,构建了 UCA 水下文物数据集。该数据集共包含 10 714 张图像,涵盖了 5 类目标,分别为瓷盘、碗、罐、香炉和瓷片,这 5 种瓷器常见于水下考古遗址。为了确保数据集的合理性,本文将 UCA 数据集按照 6:2:2 的比例进行划分,分为训练集、验证集和测试集。部分图像示例如图 7 所示。



图 7 不同场景水下文物目标图像

Fig. 7 Target images of underwater cultural artifacts in different scenes

3.2 实验环境和训练参数

本文实验平台硬件环境为高性能服务器,配置如下: Inter Core i7 处理器,主频 3.6 GHz; 16 GB 内存; 搭载 4 张显存为 11 GB 的 Nvidia Geforce GTX 1080Ti 显卡。软件环境为 Ubuntu18.04 的操作系统, Python 3.7, CUDA11.0 等。

算法相关训练参数如下:用于更新卷积核参数的梯度下降优化器为 Adam,优化器 Momentum 为 0.937,训练过程中的学习率更新方式为 step,最大学习率为 0.001,训练 batch size 为 16,权重衰减系数为 0.000 5,训练迭代周期 epoch 为 300。

3.3 评价指标

本文使用4个主要指标来测试模型的性能。精确率(Precision, P)表示模型认为是正类的正类比例,计算公式如式(4)。召回率(Recall, R)表示模型划分的正类占总正类的比例,计算公式如式(5)所示。 $F1$ 是精确率和召回率的调和均值,由于评估模型分类性能,计算公式如式(6)所示。平均精度(average precision, AP)是指每个类由精确率和召回率取不同阈值组成的曲线下的面积,值越大,该类的识别精度越好,计算公式如式(7)所示。平均精度均值(mAP)表示所有类的平均精度,该值越大,模型识别目标的准确性越好,计算公式如式(8)所示。

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (6)$$

$$AP = \int_0^1 P(R) dR \quad (7)$$

$$mAP = \frac{1}{N} \sum_{n=1}^N AP_n \quad (8)$$

式中: TP 表示模型正确预测的正样本数量; FP 表示模型预测的正样本实际上是负样本的数量; FN 表示模型预测

为负的正样本数量; N 表示所有类别的数量; AP_n 表示第 n 类的平均精度。

3.4 对比试验

为了验证本文提出的水下目标检测算法的有效性,将其与Faster R-CNN^[26]、SSD^[27]、YOLOv5-l^[28]、YOLOv7^[29]等主流目标检测算法进行对比实验。采用本文构建的真实水下文物UCA数据集上进行检测效果和性能对比。

为确保实验的可比性,参考了各个对比算法已公开的代码,并采用原始参数设置。所有对比算法均经过相同的训练流程,总共训练300 epoch,对所得模型进行定性和定量分析,以评估各个算法的性能。

本文通过不同模型的检测效果来定性分析算法性能,Faster R-CNN、SSD、YOLOv5-l、YOLOv7与本文所提算法检测效果如图8所示。可以看出,SSD算法在本文场景中检测效果最差,由于其对浅层特征的代表能力不够强,出现较多误检和漏检情况。Faster R-CNN算法和YOLOv5算法检测效果相当,YOLOv7算法效果较好,但是3种算法在目标出现掩埋或堆叠时仍存在漏检现象。与上述3种方法相比,本文所提算法取得了更好的识别效果,得益于特征提取网络的优化和SimAM注意力机制的引入,使其能够有效提取复杂环境下目标的特征信息,提高了算法鲁棒性。

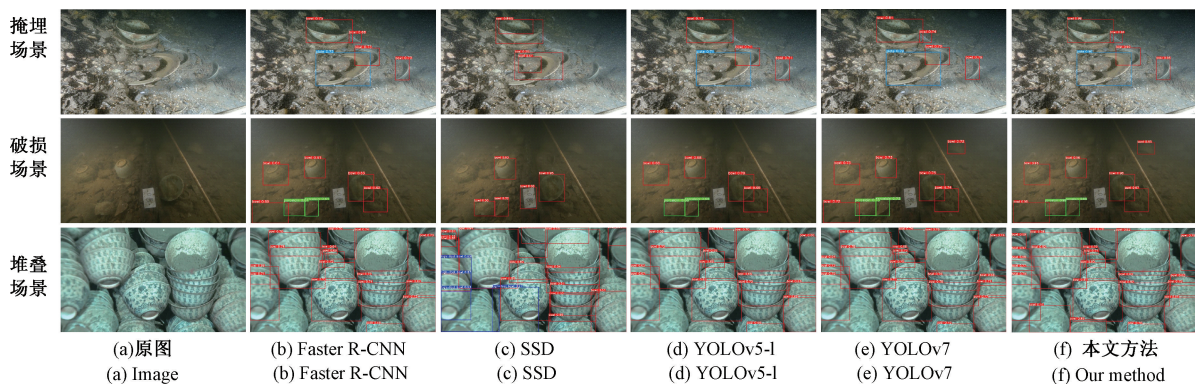


图8 不同算法检测效果

Fig. 8 Detection effect of different algorithms

为了更好的验证本文所提算法的优越性,使用UCA数据测试集进行量化分析。此外,选择上述主流目标检测算法和ROIMIX^[30]、SWIPENet^[31]、Boosting R-CNN等3种典型水下目标检测算法同本文算法进行对比实验。同时引入精确率、召回率、 $F1$ 和 mAP 等4个通用指标用以量化评价算法性能。各算法性能指标对比如表1所示。

实验结果表明,本文算法在各项指标中均优于其他算法,精确率、召回率、 $F1$ 和 mAP 分别为92.7%、

90.5%、91.5%和92.2%。同ROIMIX、SWIPENet和Boosting R-CNN等经典水下目标检测算法相比,本文算法依然取得了较好结果, mAP 分别高出三者2.1%、1.4%和0.5%。本文算法为水下文物目标设计了基于可形变卷积的多尺度深层聚合网络用以进行特征提取,通过更好地融合语义信息和空间信息对复杂环境下的目标进行识别和定位。可形变卷积扩大了检测网络的感受野,能够对破损的不规则文物特征进行有效提取,多尺度深层聚合网络减少了目标特征的上下文信息损失,更好地

表 1 不同算法性能对比

Table 1 Performance comparison of different algorithms

%				
算法	精确率	召回率	F1	mAP
Faster R-CNN ^[26]	90.2	88.5	89.3	89.4
SSD ^[27]	81.9	82.2	82.0	82.8
YOLOv5-l ^[28]	88.3	87.8	88.1	88.7
YOLOV7 ^[29]	90.1	88.4	89.3	89.9
ROIMIX ^[30]	89.3	88.2	88.8	90.1
SWIPENet ^[31]	90.9	89.8	90.3	90.8
Boosting R-CNN ^[15]	91.8	90.4	91.1	91.7
本文	92.7	90.5	91.5	92.2

注:加粗字体为各列最优结果,下同。

捕捉文物目标的全局信息。引入 SimAM 注意力模块进行特征优化,有效削减了背景冗余信息,使网络专注于目标特征信息。最后通过多尺度特征融合模块实现不同网络层的渐进特征融合。通过这些有针对性的网络结构和模块设计,水下文物目标的固有特征被保留在网络的深层,增强了网络对复杂环境下的文物目标特征表示能力,从而显著提高了检测性能。

3.5 消融实验

为了证明本文算法所设计的各模块的有效性,对每个改进点进行渐进的性能测试。以下消融实验在 UCA 数据集上进行,不同变体算法的消融结果如表 2 所示。

表 2 消融实验
Table 2 Ablation experiment

%				
方法	精确率	召回率	F1	mAP
DLA ^[23]	88.4	87.3	87.6	88.8
MDLA	89.6	89.2	89.4	90.5
MDLA+DCN	90.9	90.1	90.9	91.4
MDLA+DCN+SimAM	92.7	90.5	91.5	92.2

注:MDLA 为本文使用普通卷积的提取网络

从表 2 可以看出,与原始的 DLA 网络相比,本文设计的多尺度深层聚合网络(MDLA)mAP 提高了 1.7%,精确率提高了 1.2%,有效增强了不同尺度目标的检测能力。使用 DCN 可形变卷积替代普通卷积后有效提升了 MDLA 网络对不规则目标的特征提取能力,mAP 进一步提高 0.9%。由于 DCN 扩大了检测网络的感受野,使得网络增强了聚合能力,可以捕获目标更全面的特征信息。引入 SimAM 注意力模块,F1 上升了 0.6%,mAP 上升了 0.8%,因为注意力模块增强了目标的潜在信息,减弱了冗余信息的影响,使得各个指标进一步提高,算法有了更高的检测精度。实验证明,在复杂环境水下文物检测任

务中,加入可形变卷积和注意力模块是合理的,能有效提高算法的适应性和准确性。

3.6 AUV 视觉检测系统性能测试

为了更好的测试本文视觉检测系统的效果,将该系统部署到 AUV 中,并在真实水下环境下测试性能。

1) AUV 实验平台

AUV 是海底探测中的常用装备,在水下相关研究中具有重要作用^[32],参与实验的 AUV 如图 9 所示,其主要参数如表 3 所示。

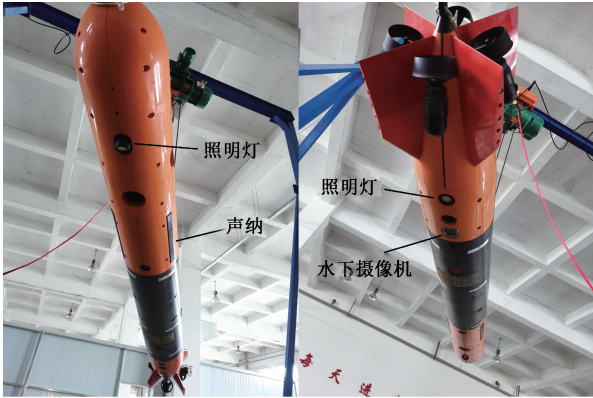


图 9 AUV 平台
Fig.9 AUV platform

表 3 AUV 主要参数
Table 3 Main parameters of the AUV

参数	数值
最大作业深度	1 000 m
巡航速度	2 kn(kn≈0.514 m/s)
最大速度	5 kn
直径	350 mm
长度	约 3.6 m
重量	约 250 kg

由于大功率、高负载的计算平台受到空间和功率的限制难以在 AUV 中应用。根据实际需求,选择 Nvidia Jetson TX2 图像边缘计算设备作为 AUV 嵌入式计算平台。原因如下:(1)该嵌入式平台尺寸为 50 mm×87 mm,常规负载下功耗仅为 7.5 W,满足 AUV 的功耗和尺寸要求;(2)CPU 采用 ARM Cortex-A57,GPU 采用 Nvidia Pascal GPU,拥有 256 个 CUDA 核心,满足算法运行要求。

2) 性能对比试验

本文的视觉检测算法集成到 Nvidia Jetson TX2 并部署到 AUV 上,用于水下考古遗址上获取的图像进行性能测试。实验在元代沉船遗址上进行,该遗址位于东南沿海,水下深度为 30 m。沉船的长度为 13.07 m,宽度为

3.7 m。所使用的 AUV 摄像头检查了一个面积为 48 m² 的区域。该遗址包含了一系列文物,包括瓷盘、瓷碗和香炉,这些是本次测试的主要目标。本次实验的结果如图 10 所示,取得了有效的检测结果。

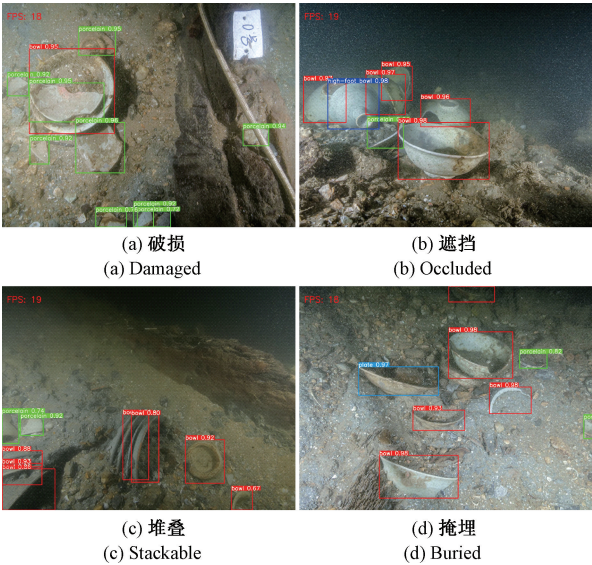


图 10 AUV 水下视觉检测系统测试结果

Fig. 10 Test results of AUV underwater visual detection system

为了评估所提出的目标检测算法的实时性能,本文选择经典的一阶段检测算法进行对比分析。同时,引入了两个性能指标——每秒帧数(frames per second, FPS)和模型参数(Params),以便对算法进行定量评估。系统性指标如表 4 所示。本文算法在实际测试中检测帧率和参数量优于 SSD 算法、YOLOv5-l 算法和 YOLOv7 算法。分析原因如下:(1) 本文算法设计 MDLA 作为基础特征提取网络,通过不同尺度深层聚合的方式将不同层级的特征进行有效地融合,从而提高了特征的利用效率。MDLA 在降低模型参数数量的同时保证了检测精度。(2) 设计的注意力特征优化模块在增强目标特征信息同时未增加模型参数量。本文算法在分辨率为 640×640 的图像上实现了 19 fps 的检测速度,基本满足实时检测的要求。

表 4 不同算法推理性能

Table 4 Inference performance of different algorithms				
方法	mAP/%	模型参数/M	分辨率	FPS/fps
SSD ^[26]	82.8	24.5	640×640	15
YOLOv5-l ^[27]	88.7	46.5	640×640	11
YOLOv5-s ^[27]	86.5	14.1	640×640	21
YOLOv7 ^[28]	89.9	74.4	640×640	10
YOLOv7-tiny ^[28]	86.4	13.2	640×640	22
本文	92.2	18.3	640×640	19

由于 YOLOv5-s 算法和 YOLOv7-tiny 算法较多的减少了网络模型的深度,本文算法在检测帧率上略低于二者,但是 mAP 分别高出二者 5.7% 和 5.8%,能够很好的弥补在时间性能的劣势,证明了本文所提出的算法具有实际应用意义。

4 结 论

本文提出了一种基于可形变深层聚合网络模型的水下文物检测算法。为了充分捕捉目标的特征信息,本文设计了 MDLA-DCN 特征提取网络,其中嵌入了可形变卷积,以确保在复杂场景下水下目标特征信息的高效利用。在此基础上,本文引入 SimAM 注意力模块进行特征优化,以增强目标潜在特征信息的同时减弱背景干扰信息。最后,通过多尺度特征融合实现不同尺度目标预测。该算法具有轻量级特性,适用于部署在图像边缘计算设备上。为了验证所提算法的有效性,本文构建了 UCA 数据集,并对算法进行了训练和测试。实验结果表明,该算法在精确率、召回率、F1 值和 mAP 等指标上分别取得了 92.7%、90.5%、91.5% 和 92.2% 的分数。本文算法部署到 AUV 上在实际场景测试中达到了 19 fps 的检测速度,满足实时检测任务需求。

本文提出的算法具有较高的检测准确性和计算效率,能够满足在水下环境中检测文物目标的任务需求。该算法的创新思想也可以应用于其他水下目标检测任务。尽管本文算法取得了较好的检测效果,但仍然存在一些局限性,例如对于目标表面附着海洋生物时检测效果不佳。为了进一步提高算法模型的泛化能力,在未来的研究中将进一步丰富数据集,同时重点解决目标附着海洋生物时出现检测失效的问题。

参考文献

[1] JING Y. Protection of underwater cultural heritage in China: New developments [J]. International Journal of Cultural Policy, 2019, 25(6) : 756-764.

[2] 魏峻. 水下文化遗产保护与管理的中国实践 [J]. 中国文化遗产, 2020(6) : 67-73.

WEI J. Chinese practices for the protection and management of underwater cultural heritage [J]. China Cultural Heritage, 2020(6) : 67-73.

[3] MANLEY J E. Unmanned maritime vehicles, 20 years of commercial and technical evolution [C]. OCEANS Conference, 2016.

[4] 吴有生, 赵羿羽, 郎舒妍, 等. 智能无人潜水器技术

- 发展研究[J]. 中国工程科学, 2020, 22(6): 26-31.
- WU Y SH, ZHAO Y Y, LANG SH Y, et al. Development of autonomous underwater vehicles technology[J]. Strategic Study of CAE, 2020, 22(6): 26-31.
- [5] 库安邦, 周兴华, 彭聪. 侧扫声纳探测技术的研究现状及发展[J]. 海洋测绘, 2018, 38(1): 50-54.
- KU AN B, ZHOU X H, PENG C. The research on development of side-scan sonar detection technology[J]. Hydrographic Surveying and Charting, 2018, 38(1): 50-54.
- [6] 宋保维, 潘光, 张立川, 等. 自主水下航行器发展趋势及关键技术[J]. 中国舰船研究, 2022, 17(5): 27-44.
- SONG B W, PAN G, ZHANG L CH, et al. Development trend and key technologies of autonomous underwater vehicles[J]. Chinese Journal of Ship Research, 2022, 17(5): 27-44.
- [7] 林森, 赵颖. 水下光学图像中目标探测关键技术研究综述[J]. 激光与光电子学进展, 2020, 57(6): 060002.
- LIN S, ZHAO Y. Review on key technologies of target exploration in underwater optical images[J]. Laser & Optoelectronics Progress, 2020, 57(6): 060002.
- [8] FAYAZ S, PARAH S A, QURESHI G J. Underwater object detection: Architectures and algorithms-a comprehensive review[J]. Multimedia Tools and Applications, 2022, 81(15): 20871-20916.
- [9] 王慧斌, 张倩, 王鑫, 等. 基于区域显著度与水下光学先验的目标检测[J]. 仪器仪表学报, 2014, 35(2): 387-397.
- WANG H B, ZHANG Q, WANG X, et al. Object detection based on regional saliency and underwater optical prior knowledge[J]. Chinese Journal of Scientific Instrument, 2014, 35(2): 387-397.
- [10] CUTTER G, STIERHOFF K, ZENG J M. Automated detection of rockfish in unconstrained underwater videos using Haar cascades and a new image dataset: Labeled fishes in the wild[C]. Proceedings of the IEEE Winter Applications and Computer Vision Workshops, 2015: 57-62.
- [11] RIZZINI D L, KALLASI F, OLEARI F, et al. Investigation of vision-based underwater object detection with multiple datasets[J]. International Journal of Advanced Robotic Systems, 2015, 12(6): 77.
- [12] LEI F, TANG F, LI S. Underwater target detection algorithm based on improved YOLOv5[J]. Journal of Marine Science and Engineering, 2022, 10(3): 310.
- [13] YAN J, ZHOU Z, ZHOU D, et al. Underwater object detection algorithm based on attention mechanism and cross-stage partial fast spatial pyramidal pooling[J]. Frontiers in Marine Science, 2022(9): 1056300.
- [14] 杨玉娥, 杜文豪, 刘鲁宁. 基于主动电场定位的海参水下识别技术[J]. 仪器仪表学报, 2023, 44(8): 210-217.
- YANG Y E, DU W H, LIU L N. Underwater sea-cucumber identification technology based on active electrolocation[J]. Chinese Journal of Scientific Instrument, 2023, 44(8): 210-217.
- [15] SONG P, LI P, DAI L, et al. Boosting R-CNN: Reweighting R-CNN samples by RPN's error for underwater object detection[J]. Neurocomputing, 2023, 530: 150-164.
- [16] ZENG L, SUN B, ZHU D. Underwater target detection based on Faster R-CNN and adversarial occlusion network[J]. Engineering Applications of Artificial Intelligence, 2021, 100: 104190.
- [17] GUO W, ZHANG Y, ZHOU Y, et al. Underwater real-time target detection based on key frame and model compression[C]. Journal of Physics: Conference Series. IOP Publishing, 2021, 1800(1): 012001.
- [18] 张有波, 郭威, 周悦, 等. 基于多粒度剪枝的水下遗迹实时目标检测[J]. 激光与光电子学进展, 2021, 58(14): 1410019.
- ZHANG Y B, GUO W, ZHOU Y, et al. Real-Time target detection of underwater relics based on multi-granularity pruning[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1410019.
- [19] ZHANG W, ZHUANG P, SUN H H, et al. Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement[J]. IEEE Transactions on Image Processing, 2022, 31: 3997-4010.
- [20] 强伟, 贺昱曜, 郭玉锦, 等. 基于改进 SSD 的水下目标检测算法研究[J]. 西北工业大学学报, 2020, 38(4): 747-754.

- QIANG W, HE Y Y, GUO Y J, et al. Exploring underwater target detection algorithm based on improved SSD [J]. Journal of Northwestern Polytechnical University, 2020, 38(4): 747-754.
- [21] CHEN X Q, XIA K, HU W, et al. Extraction of underwater fragile artifacts: Research status and prospect[J]. Heritage Science, 2022, 10(1): 9.
- [22] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 764-773.
- [23] YU F, WANG D, SHELHAMER E, et al. Deep layer aggregation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2403-2412.
- [24] YANG L, ZHANG R Y, LI L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks [C]. Proceedings of the International Conference on Machine Learning, 2021: 11863-11874.
- [25] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [26] REN S Q, 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, DOI: 10.1109/TPAMI.2016.2577031.
- [27] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Proceedings of the European Conference on Computer Vision, 2016: 21-37.
- [28] JOCHER G, CHAURASIA A, STOKEN A, et al. ultralytics/YOLOv5:v7.0-YOLOv5 sota realtime instance segmentation [J]. Zenodo, 2022, DOI: 10.5281/zenodo.7347926.
- [29] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [30] LIN W H, ZHONG J X, LIU S, et al. Roimix: Proposal-fusion among multiple images for underwater

object detection [C]. 2020 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2020: 2588-2592.

- [31] CHEN L, ZHOU F, WANG S, et al. SWIPENET: Object detection in noisy underwater scenes[J]. Pattern Recognition, 2022, 132: 108926.
- [32] XU G, LIU K, ZHAO Y, et al. Research on the modeling and simulation technology of underwater vehicle[C]. OCEANS Conference. IEEE, 2016: 1-6.

作者简介



周道先, 2020 年于华北电力大学科技学院获得学士学位, 现为沈阳理工大学硕士研究生, 主要研究方向为机器视觉、人工智能。
E-mail: zdx20110908@163.com

Zhou Daoxian received his B. Sc. degree from North China Electric Power University Science and Technology College in 2020. He is currently a M. Sc. candidate at Shenyang Ligong University. His main research interests include machine vision and artificial intelligence.



张吟龙(通信作者), 2010 年于沈阳建筑大学获得学士学位, 2013 年于中国科学院大学获得硕士学位, 2019 年于中国科学院大学获得博士学位, 现为中国科学院沈阳自动化研究所副研究员, 主要研究方向为工业视觉、多源信息融合。

E-mail: zhangyinlong@sia.cn

Zhang Yinlong (Corresponding author) received his B. Sc. degree from Shenyang Jianzhu University in 2010, M. Sc. degree and Ph. D. degree both from University of Chinese Academy of Sciences in 2013 and 2019, respectively. He is currently an associate researcher at the Shenyang Institute of Automation, Chinese Academy of Sciences. His main research interests include industrial vision and multi-source information fusion.



徐高飞, 2011 年于吉林大学获得学士学位, 2019 年于中国科学院大学获得博士学位, 现为中国科学院深海科学与工程研究所工程师, 主要研究方向为深海装备研发及其智能控制技术研究。

E-mail: xugf@idsse.ac.cn

Xu Gaofei received his B. Sc. degree from the Jilin University in 2011, Ph. D. degree from the University of Chinese Academy of Sciences in 2019. He is currently an engineer at Institute of

Deep-sea Science and Engineering, Chinese Academy of Sciences. His main research interests include development and intelligent control of deep-sea equipment.



杨雨沱, 2010 年于扬州大学获得学士学位, 2013 年于中国科学院大学获得硕士学位, 现为中国科学院沈阳自动化研究所副研究员, 主要研究方向为工业无线传感器网络和工业物联网。

E-mail: yangyutuo@sia.cn

Yang Yutuo received his B.Sc. degree from Yangzhou University in 2010, M.Sc. degree from University of Chinese Academy of Sciences in 2013. He is currently an associate researcher at Shenyang Institute of Automation, Chinese Academy of Sciences. His main research interests include industrial wireless networks and internet of things.



梁伟, 1996 年于沈阳建筑大学获得学士学位, 1999 年于中国科学院大学获得硕士学位, 2002 年于中国科学院大学获得博士学位, 现为中国科学院沈阳自动化研究所研究员, 主要研究方向为工业无线网络和无线网络安全。

E-mail: weiliang@sia.cn

Liang Wei received her B.Sc. degree from Shenyang Jianzhu University in 1996, M.Sc. degree and Ph.D. degree both from University of Chinese Academy of Sciences in 1999 and 2002, respectively. She is currently a professor at the Shenyang Institute of Automation, Chinese Academy of Sciences. Her main research interests include industrial wireless sensor networks and wireless network security.