

一种改进的基于YOLOv5s的轻量化航拍目标检测模型

陈海燕, 毛利宏

引用本文

陈海燕, 毛利宏. 一种改进的基于YOLOv5s的轻量化航拍目标检测模型[J]. 计算机科学, 2024, 51(11A): 231100119-8.

CHEN Haiyan, MAO Lihong. Improved Lightweight Aerial Photography Object Detection Model Based on YOLOv5s [J]. Computer Science, 2024, 51(11A): 231100119-8.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于多模态对比学习的场景图生成方法](#)

Multimodal Contrastive Learning Based Scene Graph Generation

计算机科学, 2024, 51(11A): 231200185-5. <https://doi.org/10.11896/jsjcx.231200185>

[基于双重标签分配的遥感有向目标检测方法](#)

Remote Sensing Oriented Object Detection Method Based on Dual-label Assignment

计算机科学, 2024, 51(11A): 240100058-9. <https://doi.org/10.11896/jsjcx.240100058>

[PS-YOLOv8:增强电力线路检测中的小规模损坏检测](#)

PS YOLOv8:Enhancing Detection of Small-scale Damage in Power Lines Inspection

计算机科学, 2024, 51(11A): 240100003-6. <https://doi.org/10.11896/jsjcx.240100003>

[改进YOLOv5s的城市建筑环境下的烟火检测算法](#)

Improved YOLOv5s Algorithm for Detecting Fireworks in Urban Building Environments

计算机科学, 2024, 51(11A): 240100051-7. <https://doi.org/10.11896/jsjcx.240100051>

[基于改进Yolov8的敦煌壁画元素检测算法](#)

Dunhuang Mural Element Detection Algorithm Based on Improved Yolov8

计算机科学, 2024, 51(11A): 231000034-6. <https://doi.org/10.11896/jsjcx.231000034>

一种改进的基于 YOLOv5s 的轻量化航拍目标检测模型

陈海燕 毛利宏

兰州理工大学计算机与通信学院 兰州 730050

摘要 无人机航拍图像背景复杂、目标密集且小目标占比大,加大了目标检测的难度。基于深度学习的目标检测模型计算复杂度高,难以部署在无人机搭载的嵌入式设备上。针对此问题,提出了一种改进的基于 YOLOv5s 的轻量化航拍图像目标检测模型。首先将 YOLOv5s 主干网络的 C3 模块 BottleNeck 替换为轻量级的 ShuffleNetv2 网络,来降低模型的参数量和计算复杂度;其次在 ShuffleNetv2 网络中引入跨层信息交叉融合、SE 通道注意力机制以及残差连接,来缓解卷积操作导致的特征通道数减少、网络中间层特征图的信息利用不充分问题;再次在 YOLOv5s 多尺度特征融合网络中引入 SE 通道注意力机制,来提高网络对关键特征的捕捉和提取能力;最后对改进的目标检测模型采用通道剪枝的方法使模型进一步轻量化。实验结果表明:在 NWPU VHR-10 数据集上,改进后的模型与 YOLOv5s 模型相比,目标检测的准确率和平均精度均值分别提升了 3.5%,1.9%,模型的参数量和计算量降低了 76%,48.7%,模型大小压缩了 73.8%,检测速度提升了 48%。

关键词: 目标检测;轻量化网络;YOLOv5s;SE 通道注意力机制;通道剪枝

中图分类号 TP391.4

Improved Lightweight Aerial Photography Object Detection Model Based on YOLOv5s

CHEN Haiyan and MAO Lihong

School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China

Abstract The difficulty of target detection is increased by complex backgrounds, dense targets, and a high proportion of small objects in the unmanned aerial vehicle(UAV) aerial images. Deployment on embedded devices by drones is difficult due to the high computational complexity of the target detection model based on deep learning. Aiming at the above problems, an improved lightweight aerial image object detection model based on YOLOv5s is proposed. Firstly, the C3 module BottleNeck of the YOLOv5s backbone network is replaced with lightweight ShuffleNetv2 to reduce model parameters and computational complexity. Secondly, cross-layer information cross-fusion, SE channel attention mechanism, and residual connections are introduced in the ShuffleNetv2 network to alleviate the problem of reducing the number of feature channels caused by convolution operations and insufficient information utilization of feature maps in the middle layer of the network. Then, the SE channel attention mechanism is introduced into the YOLOv5s multi-scale feature fusion network, augmenting the network's ability to capture and extract key features. Finally, the proposed target detection model in this paper is further lightened by channel pruning. Experimental results on the NWPU VHR-10 dataset show that, compared with the YOLOv5s model, the proposed model is increased of 3.5% in precision and 1.9% in mean average precision. The number of parameters and computational workload is reduced by 76% and 48.7%, the model size is compressed by 73.8% and detection speed improved by 48%.

Keywords Object detection, Lightweight network, YOLOv5s, SE channel attention mechanism, Channel pruning

1 引言

随着无人机技术的迅速发展,无人机航拍已经被广泛应用于自然灾害监测、城市规划、环境监测等领域^[1-3]。精准实时的航拍图像目标检测可为灾情预警、城市建设、环境资源管理等领域提供实时高效、全面准确、安全可靠的数据,从而建立快速响应和决策支持系统,提升灾情预警的准确性和及时性,优化城市建设的规划和管理,提高环境资源的保护和可持续利用效率。然而,由于航拍图像具有背景复杂、目标密集、小目标占比大等特点,对其进行精准的目标检测难度较大;此外,无人机搭载的嵌入式设备资源有限,而现有目标检测模型参数量大、计算复杂,难以直接部署在无人机搭载的嵌入式设备上。因此,如何在保证航拍图像目标检测精度的同时使模型

轻量化成为目标检测技术在航拍领域实时应用的关键。

目标检测方法分为基于传统机器学习的目标检测方法和基于深度学习的目标检测方法。基于传统机器学习的目标检测方法主要包括支持向量机法^[4](Support Vector Machine, SVM)、随机森林法^[5](Random Forest)、尺度不变特征变换法^[6](Scale-Invariant Feature Transform, SIFT)等,此类方法复杂度低、易于解释和实现,但需要人工提取特征,模型的泛化能力弱^[7],因此对复杂的场景和目标变化较大情况下的目标检测精度不高。基于深度学习的目标检测方法因具有强大的非线性建模能力、大规模数据处理能力、端到端的学习能力,以及网络结构灵活、拓展性强等优点而被广泛应用于无人机航拍领域。如文献^[8]通过使用递归特征金字塔和交并比损失函数替换感兴趣区域池化的 L1 损失函数,提出了一种

基金项目:国家自然科学基金(62161019,62061024)

This work was supported by the National Natural Science Foundation of China(62161019,62061024).

通信作者:陈海燕(chenhaiyan@sina.com)

改进的基于 Faster R-CNN^[9] 的航拍静态图像目标检测模型,该模型结合区域建议网络和卷积神经网络,通过使用区域建议网络预测图像中潜在物体的目标度分数和边界框坐标,减少了全面搜索目标位置的计算开销,而卷积神经网络通过学习图像特征来提高目标的分类和定位精度,以实现高效准确的航拍图像目标检测。文献[10]通过采用稠密特征提取网络替换 SSD(Single Shot multibox Detector, SSD)^[11] 主干网络、引入分组卷积和焦点损失函数(Focal Loss),提出了一种改进的基于 SSD 算法的航拍图像目标检测模型,该模型通过优化特征提取网络结构,提升了对不同尺度目标的特征提取能力,实验结果表明改进后的模型与原模型的平均检测精度均值分别为 29.3% 和 24.0%。文献[12]通过对进空间-通道注意力机制进行改进,并在主干网络中引入 Transformer 模块,提出了一种面向无人机航拍的实时密集小目标检测模型,该模型能够更好地捕获航拍图像中复杂背景下密集小目标的特征,相比原模型,检测精度提高了 6.4%。文献[13]通过增加检测分支和使用解耦检测头以及采用 Alpha-IoU 优化损失函数的方式,提出了一种针对航拍图像中小目标检测的 Drone-YOLO 模型,该模型能够有效提升航拍图像中小目标的检测精度,在 VisDrone 数据集上检测精度提升了 4.91%,但该模型在提升检测精度的同时未考虑模型的复杂性。以上文献中基于深度学习的航拍图像目标检测模型尽管在一定程度上提高了对航拍图像目标检测的精度,但这些模型仅考虑了航拍图像目标密集、尺度变化、目标遮挡等特点,而没有考虑模型网络结构的复杂性,参数量大小以及在实际应用中的硬件部署问题。为了使航拍图像目标检测模型能够在无人机搭载的嵌入式设备上部署,需要对模型进行轻量化。目前目标检测模型轻量化的方法主要有两种:一是采用参数量较小的轻量级网络结构。这种方法通过设计参数量较小的网络结构,能够在检测性能损失较小的情况下显著减少模型的参数量。一些经典的轻量级网络结构有 MobileNet^[14-16] 系列、ShuffleNet^[17-18] 系列、GhostNet^[19] 等。例如文献[20]通过引入通道分割和跨层信息融合的方法,提出了一种改进的基于 Shuf-

fleNetv2 的轻量级目标检测模型,该模型通过手动调节通道分割率来控制计算开销和冗余,并利用跨层信息融合将模型中间层输出通道的信息和保留模块的信息进行交换,从而使模型具有了更高的计算效率和泛化能力。目标检测模型轻量化的第二种方法是采用稀疏训练、通道剪枝等方法优化目标检测模型,从而使其轻量化。例如文献[21]将 YOLOv5 的主干网络替换为轻量级 MobileNetv2 网络,并结合稀疏训练和通道剪枝技术,在降低模型参数量的同时保证了模型具有较高的检测精度,实验结果表明其推理速度提升了近 1.7 倍,然而,这一速度提升是以模型的 mAP 值下降 6.74% 为代价的。

以上研究表明:基于深度学习的目标检测模型具有较强的特征提取能力,从而对航拍图像具有较高的检测精度,但由于航拍图像存在背景复杂、目标密集且小目标占比较大的特点,目标易被复杂的背景干扰而造成特征提取困难;其次,现有的基于深度学习的目标检测模型网络规模大,参数量多,限制了模型在无人机搭载的嵌入式设备上的部署。

针对以上问题,本文以 YOLOv5s 为基础网络,提出了一种改进的基于 YOLOv5s 的轻量化航拍图像目标检测模型。首先,采用轻量级 ShuffleNetv2 网络替换 YOLOv5s 主干网络中 C3(Cross Stage Partial Network with 3 Convolutions, C3)模块的 BottleNeck,来降低模型的参数量和计算复杂度;其次,在 ShuffleNetv2 网络中引入跨层信息交叉融合、SE(Squeeze-and-Excitation)通道注意力机制^[22] 以及残差连接来对 ShuffleNetv2 网络进行改进,增强模型对小目标特征的提取能力。再次,在 YOLOv5s 多尺度特征融合网络中引入 SE 通道注意力机制,来提高网络对不同尺度关键特征的捕捉和提取能力;最后,对改进的目标检测模型采用通道剪枝的压缩方法,通过裁剪对模型检测性能贡献较小的通道来进一步减少模型的参数量和计算量。

2 轻量化航拍目标检测模型的建立

本文轻量化航拍图像目标检测模型整体架构如图 1 所示。

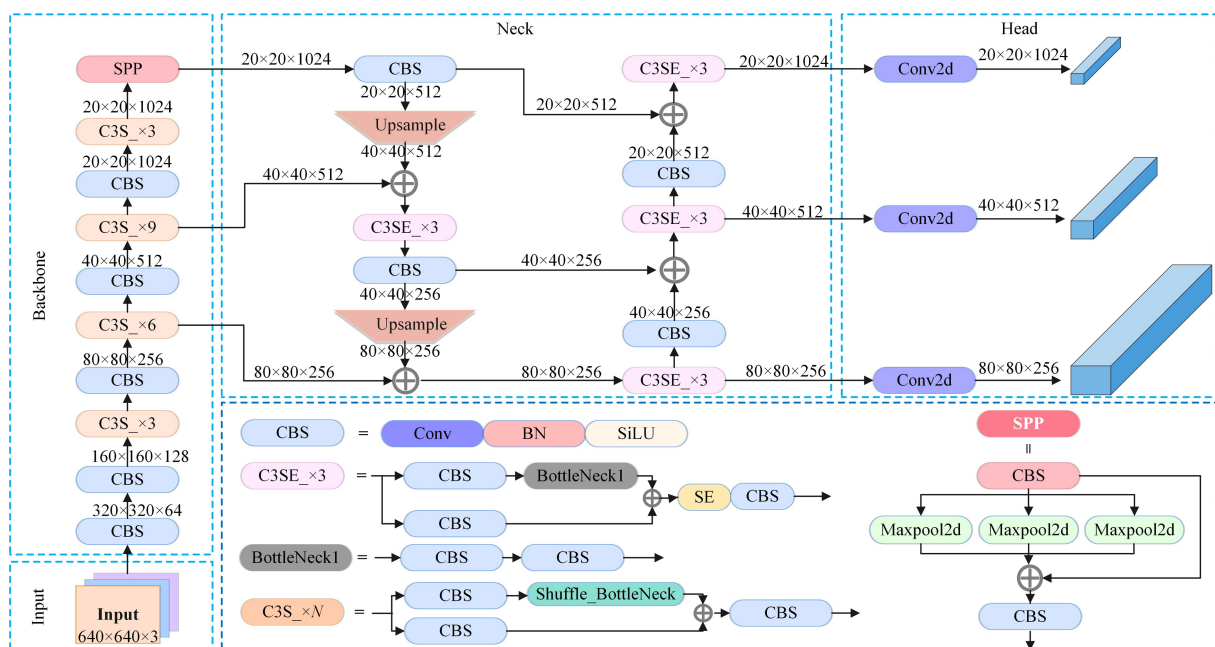


图 1 轻量化航拍目标检测模型整体架构

Fig. 1 Overall architecture of lightweight aerial photography object detection model

本文所提出的轻量化航拍图像目标检测模型以YOLOv5s目标检测模型为基础模型,分别对YOLOv5s目标检测模型的主干网络和多尺度特征融合网络进行优化改进。首先将YOLOv5s主干网络中C3模块的BottleNeck替换为轻量级的ShuffleNetv2网络;其次在ShuffleNetv2网络中引入跨层信息交叉融合、SE通道注意力机制以及残差连接对其进行改进;最后在YOLOv5s的多尺度特征融合网络引入SE通道注意力,挖掘不同尺度特征之间的关联信息,来扩大模型的感受野并提升对目标的检测能力。该模型主要由主干网络(Backbone)、多尺度特征融合网络(Neck)和头部检测器

(Head)组成。其中Backbone主要包括CBS模块和改进的C3S模块,该主干网络负责对输入的图像进行特征提取;Neck网络主要包括增加了SE通道注意力机制的PANet^[23]模块,以此来实现在不同尺度的特征图上使用网格的锚框来预测目标的位置和类别。

2.1 YOLOv5s 主干网络优化

YOLOv5s主干网络结构如图2所示,主要由卷积、批归一化层、激活函数共同组成的CBS模块,C3模块以及SPP模块组成。

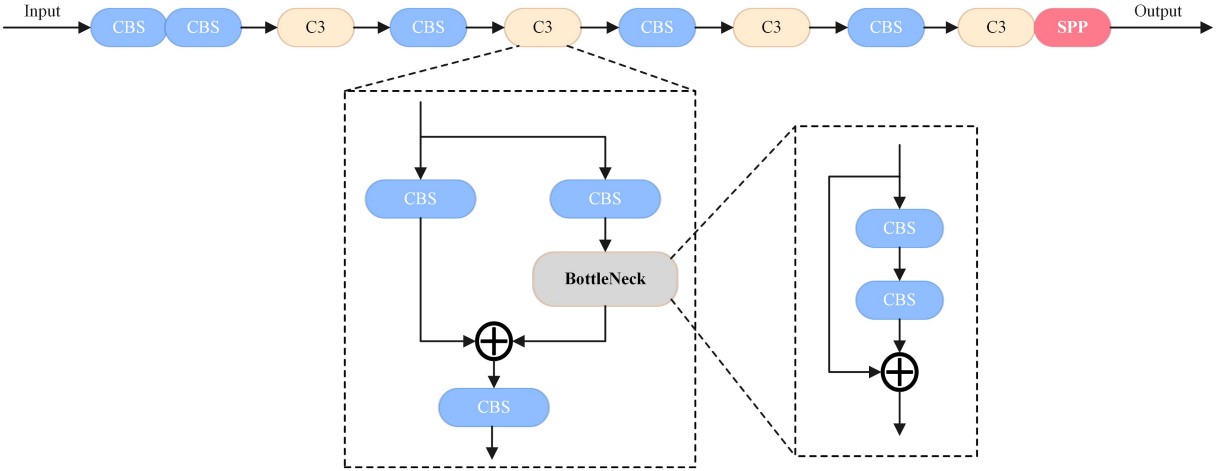


图2 YOLOv5s 主干网络结构

Fig. 2 Backbone network structure of YOLOv5s

YOLOv5s 为了提高网络的特征表达能力和检测性能,通常会在主干网络叠加多个C3模块,这也增加了模型的参数数量和计算复杂度。针对这一问题,本文将YOLOv5s主干网络中C3模块的BottleNeck替换为轻量级的ShuffleNetv2网络来降低模型的参数数量和计算复杂度。ShuffleNetv2网络的基本模块构成如图3所示,其主要由 1×1 普通卷积、 3×3 深度卷积和通道混洗操作组成。

ShuffleNetv2网络的基本模块中的左右分支的空间和通道信息相对独立,无法有效地交互和共享中间层的特征信息,因此,为了加强ShuffleNetv2网络下采样模块中间层特征的信息利用,增强模型对目标的特征提取能力,本文对ShuffleNetv2下采样模块进行改进。

首先,将图3(b)中的下采样模块左右分支中 3×3 深度卷积输出的特征进行信息交叉融合,融合过程如下:假设输入ShuffleNetv2下采样模块的特征图为 $X \in R^{C \times H \times W}$,其中 C 为特征图通道数, H 为特征图高度, W 为特征图宽度。经过图3(b)左侧分支的 3×3 深度卷积、右侧分支的 1×1 卷积和 3×3 深度卷积之后得到的特征分别为 $F_1 \in R^{C \times H/2 \times W/2}$ 和 $F_2 \in R^{C \times H/2 \times W/2}$,然后将特征 F_1 在通道方向上以 $1/2$ 的比例进行分割得到特征 $F_{1,1} \in R^{C/2 \times H/2 \times W/2}$ 和 $F_{1,2} \in R^{C/2 \times H/2 \times W/2}$,将特征 F_2 在通道方向上以 $1/2$ 的比例进行分割得到特征 $F_{2,1} \in R^{C/2 \times H/2 \times W/2}$ 和 $F_{2,2} \in R^{C/2 \times H/2 \times W/2}$ 。将分割后得到不同分支不同层的特征 $F_{1,1}, F_{1,2}, F_{2,1}, F_{2,2}$ 输出的通道数用 $C = [c_1, c_2, \dots, c_{c/2}, \dots, c_c, \dots, c_i, \dots, c_{2c}]$ 一维数组表示,其中 $c_i (1 \leq i \leq 2c)$ 表示分割后不同特征图的通道数。将一维数组通过重排的方法转换成二维数组 $(g, n) = [c_1, \dots, c_{c/2}; \dots; c_{(3c/2)+1}, \dots, c_{2c}]$,其中 $g=4, n=C/2$ 。

其次,将二维数组进行转置得到 $(n, g) = [c_1, c_{(c/2)+1}, c_{c+1}, c_{(3c/2)+1}; \dots; c_{c/2}, c_{c+1}, c_{3c/2}, c_{2c}]$,再将转置后的二维数组展开成一维数组 $n \times g = [c_1, c_{(c/2)+1}, c_{c+1}, \dots, c_c, c_{3c/2}, c_{2c}]$,实现在不同层不同通道信息交叉融合,本文称之为跨层信息交叉融合。

然后,借鉴轻量级MobileNetv3网络结构中使用SE通道

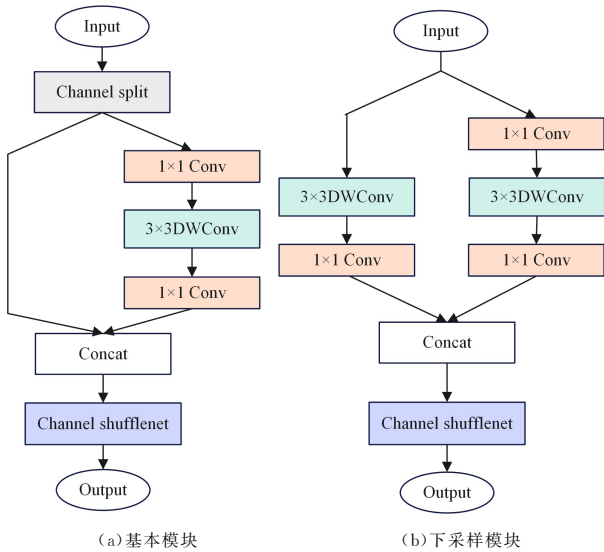


图3 ShuffleNetv2 网络模块

Fig. 3 ShuffleNetv2 network module

ShuffleNetv2下采样模块(见图3(b))中的 3×3 深度卷积用于捕获空间特征,而 1×1 卷积用于融合通道信息。但

注意力机制来对不同通道特征图加权的思想,本文在图 3(b)中的 ShuffleNetv2 下采样模块右分支经过跨层信息交叉融合后,引入 SE 通道注意力机制,动态地调整不同特征通道的权重,使模型更加关注当前任务所需的通道信息。引入 SE 通道注意力机制的过程如下:假设在图 3(b)的 ShuffleNetv2 下采样模块右分支经过跨层信息交叉融合后输出特征图为 $X' \in R^{C \times H \times W}$,引入 SE 通道注意力机制后的特征图 \tilde{X} 可由式(1)得到。

$$\tilde{X} = X' \cdot \sigma(W_2 \cdot \delta(W_1 \cdot AvgPool(X))) \quad (1)$$

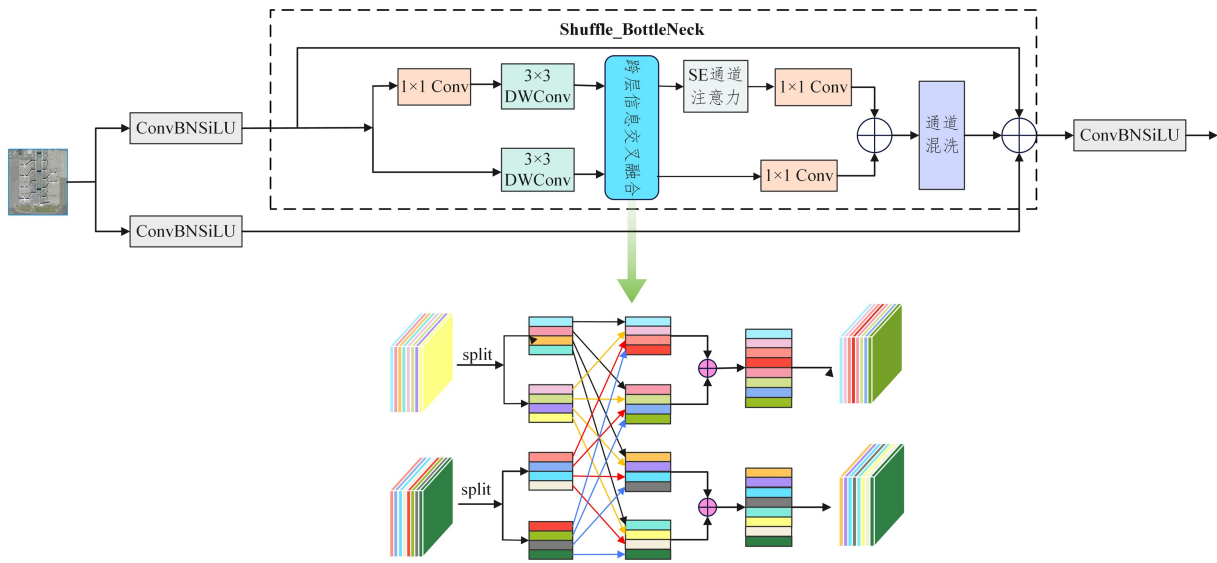


图 4 改进的 C3(C3S)模块

Fig. 4 Improved C3(C3S) module

2.2 基于 SE 通道注意力机制的多尺度特征融合网络优化

YOLOv5s 多尺度特征融合网络采用 PANet 模块实现自顶向下和自底向上的双向融合,如图 5 所示。

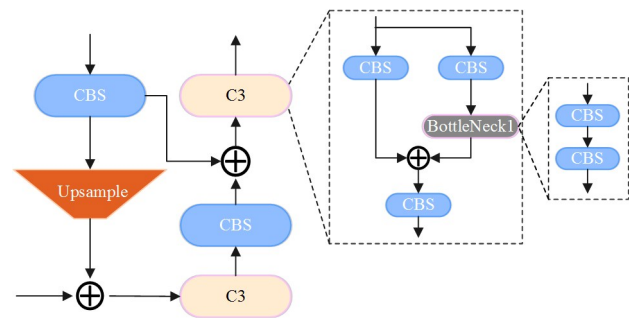


图 5 特征融合示意图

Fig. 5 Feature fusion diagram

图 5 所示的这种融合方式只是使用了简单的通道拼接操作,没有考虑不同特征图之间的关联和权重分配,因此,为解决 YOLOv5s 多尺度特征融合网络中不同层级特征不能动态地提取语义信息和空间信息的问题,本文在 PANet 模块的上采样和下采样之后的 C3 模块中引入 SE 通道注意力机制,如图 6 所示。通过式(2)一式(4)来对不同通道特征按贡献度大小分配相应的权重,减少冗余信息,提升模型对不同尺寸目标的感知能力。在图 6 所示的 PANet 模块的上采样和下采样之后的 C3 模块中引入 SE 通道注意力机制的过程如下:假设在 C3 模块通道拼接后输出的特征图为 $X \in R^{C \times H \times W}$,将其作为 SE 通道注意力机制的输入,首先由式(2)得到每个通道的压

其中, $W_1 \in R^{(C/r) \times C}$ 和 $W_2 \in R^{C \times (C/r)}$ 为卷积权重, r 为超参数; $AvgPool$ 为全局平均池化; δ 为 ReLU 激活函数; σ 为 sigmoid 激活函数。

最后,将 YOLOv5s 主干网络结构中的 C3 模块中卷积操作与改进后的 ShuffleNetv2 下采样模块中的通道混洗输出进行残差连接,使 YOLOv5s 网络可以在保留原始特征的基础上学习到更复杂和抽象的特征,更好地利用通道间的信息来增强特征在通道维度上的表达能力。本文改进后的 C3 模块(C3S)如图 4 所示。

缩特征 $S_q(X) \in R^C$,其次由式(3)得到每个通道权重 $E_x(X) \in R^C$,最后由式(4)将通道权重 $E_x(X)$ 和原始的特征图 X 进行加权得到 SE 通道注意力特征图 $\tilde{X} \in R^{C \times H \times W}$ 。

$$S_q(X) = AvgPool(X) \quad (2)$$

$$E_x(X) = \sigma \cdot (FC(S_q(X))) \quad (3)$$

$$\tilde{X} = E_x(X) \cdot X \quad (4)$$

其中, $AvgPool$ 为全局平均池化; FC 为全连接层操作; σ 为 sigmoid 激活函数。

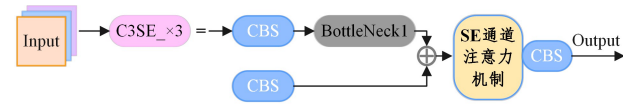


图 6 引入注意力机制的 C3 模块

Fig. 6 C3 module with introduced attention mechanism

2.3 基于批归一化层稀疏训练的模型通道剪枝方法

为进一步减少模型的参数量和计算量,提高模型的推理速度,对本文图 1 中的目标检测模型进行通道剪枝。首先在模型训练过程中引入稀疏性约束,在损失函数中添加稀疏正则项(L1 正则化),通过约束参数的绝对值之和来促使部分通道的权重值为 0,如式(5)所示:

$$L = \sum_{(x,y)} l(f(x,W), y) + \lambda \sum_{\gamma \in \Gamma} g(\gamma) \quad (5)$$

其中, x 为模型的输入数据, y 为数据标签, W 为模型的可训练权重, $f(x,W)$ 为模型的预测输出, λ 为平衡因子, γ 为缩放因子。

其次通过控制超参数 λ 的大小,使模型获得合适的稀疏

度,优化器通过计算梯度并更新缩放参数,促使模型对通道的重要性进行自适应的学习;通过不断迭代训练,使得贡献度较小的通道的缩放参数逐渐减小甚至趋近于零,实现对通道的剪枝。最后将剪枝后的模型重新训练后进行微调来进一步提高模型的精度和泛化能力。

3 实验与结果分析

3.1 实验环境与配置

本文实验所用的硬件和软件环境配置如表 1 所列。数据集采用的是 NWPU VHR-10 和 DroneVehicle,两个数据集都具有背景复杂、目标密集且小目标占比大的特点。选择 NWPU VHR-10 数据集评估模型的检测性能,并在 DroneVehicle 数据集上验证模型的鲁棒性。NWPU VHR-10 包含飞机、舰船、油罐、棒球场、网球场、篮球场、田径场、港口、桥梁、汽车。共有 800 张图像,其中 650 张含有目标,实验中仅选取含有目标的图像,按照 2:2:6 的比例随机划分成训练集、验证集、测试集。在训练过程中,初始学习率设置为 0.001,余弦退火率为 0.1,学习率动量为 0.937,权重衰减系数为 0.0005,最大迭代次数为 600,批处理(batch size)设置为 4,采用 Adam 作为优化函数。

表 1 实验环境配置

Table 1 Experimental environment configuration

环境配置	主要信息
实验平台	Ubuntu18.04 操作系统
GPU 型号	NVIDIA Titan V
编程语言	Python3.8
CUDA 版本	11.4
深度学习框架	Pytorch1.12.0

3.2 评价指标

为了评价本文模型的性能,采用平均精度均值(mean Average Precision, mAP)、准确率(Precision, P)、计算量(Floating Point of Operations, FLOPs)、参数量(Params)、模型体积、帧率(Frames Per Second, FPS)作为评价指标。准确率和召回率表达式如式(6)、式(7)所示,其中 TP, FP, FN 分别是真正例、假正例、假负例;AP 的表达式如式(8)所示,表示准确率(Precision, P)和召回率(Recall, R)分别为横纵坐标所围成曲线的面积;mAP 表示对每个类别的平均精度(Average Precision, AP)求均值,表达式如式(9)所示,其中 C 表示类别数。

$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

$$AP = \int P(R) dR \quad (9)$$

$$mAP = \frac{1}{C} \sum_{i=1}^C AP_i \quad (10)$$

3.3 改进的 BottleNeck 模块对模型性能的影响

为验证改进的 BottleNeck 模块对航拍图像目标检测模型性能的影响,本文在保持 YOLOv5s 多尺度特征融合网络和头部检测器不变的基础上对改进的主干网络 C3S 模块中的 BottleNeck 进行性能分析,设计消融实验以验证改进的 BottleNeck 模块的有效性。实验结果如表 2 所列。

表 2 改进的 BottleNeck 模块性能分析

Table 2 Performance analysis of improved BottleNeck module

跨层信息交叉融合	SE 通道注意力	残差连接	mAP/%	Params	FLOPs
—	—	—	83.4	6 735 367	14.8×10^9
✓	—	—	84.1	6 861 271	14.8×10^9
✓	✓	—	85.6	6 924 223	14.8×10^9
✓	✓	✓	85.9	6 924 223	14.8×10^9

注“—”和“✓”分别表示未使用和使用此项。

由表 2 可以看出,当 BottleNeck 模块采用跨层信息交叉融合后,mAP 值从 83.4% 提高到 84.1%,这表明在 ShuffleNetv2 网络下采样模块中使用跨层信息交叉融合能够有效地利用中间层特征的信息,增强模型对目标的特征提取能力。在跨层信息交叉融合的基础上引入 SE 通道注意力机制时,模型的 mAP 值进一步提升到 85.6%,相较于仅使用跨层信息交叉融合,引入 SE 通道注意力机制能使模型更加关注与当前任务相关的通道信息,模型的 mAP 值提升了 1.8%。在使用跨层信息交叉融合和 SE 通道注意力机制的基础上引入残差连接,模型的 mAP 值达到了 85.9%,与未使用改进的 BottleNeck 模块的检测模型相比,mAP 值提升了 3.0%。模型的参数量从 6 735 367 增加到 6 924 223。值得注意的是,计算量在不同实验条件下都保持在 14.8×10^9 FLOPs 至 14.9×10^9 FLOPs 之间,表明改进后的模型在计算复杂度上没有显著的增加。

3.4 改进的各模块对模型检测性能的影响分析

为分析本文所提出的轻量化航拍图像目标检测模型各模块对目标检测性能的影响,本文设计了多组消融实验。表 3 是改进的各模块对模型检测性能的影响分析。YOLOv5s 是基础检测模型,YOLOv5s_SV2 是将 YOLOv5s 主干网络中 C3 模块的 BottleNeck 替换为 ShuffleNetv2 下采样模块后的检测模型,YOLOv5s_NSE 是在多尺度特征融合网络引入 SE 通道注意力机制之后的检测模型,YOLOv5s_C3S 是将 YOLOv5s 主干网络 C3 模块改进后的网络模型,YOLOv5s_BSN 是在 YOLOv5s 中引入改进后的 C3S 模块以及多尺度特征融合网络中引入 SE 通道注意力机制后的检测模型。

表 3 改进各模块对检测性能的影响

Table 3 Influence of improved modules on detection results

模型	mAP/%	Params	FLOPs
YOLOv5s	84.6	7 037 095	15.8×10^9
YOLOv5s_SV2	83.4	6 735 367	14.8×10^9
YOLOv5s_NSE	84.5	6 787 287	14.9×10^9
YOLOv5s_C3S	85.9	6 924 223	14.9×10^9
YOLOv5s_BSN	86.6	6 976 143	14.9×10^9

从表 3 可以看出,将 ShuffleNetv2 下采样模块直接替换 YOLOv5s 主干网络 C3 模块的 BottleNeck 进行实验(表 3 中第 2 行),模型的 mAP 值出现了下降,其参数量和计算量均有所降低。用 ShuffleNetv2 下采样模块替换 YOLOv5s 主干网络中 C3 模块的 BottleNeck,同时在多尺度特征融合网络中引入 SE 通道注意力机制进行实验(表 3 中第 3 行),模型在参数量和计算量有所增加的情况下 mAP 值有所提升,但不明显。因此,在 ShuffleNetv2 网络中引入跨层信息交叉融合、SE 通道注意力以及残差连接后,将其引入 YOLOv5s 网络进行实验(表 3 第 4 行),相比只将 ShuffleNetv2 的下采样模块替换 YOLOv5s 主干网络中 C3 模块的 BottleNeck 检测模型(表 3

第 2 行), mAP 值提升了 2.9%; 与 YOLOv5s 基础检测模型相比(表 3 第 1 行), mAP 值提升了 1.5%, 且计算量和参数量也有所降低。YOLOv5s_BSN(表 3 第 5 行)是在 YOLOv5s 网络模型中引入改进后的 C3S 模块以及在多尺度特征融合网络中引入 SE 通道注意力机制后的检测模型, 其模型 mAP 值有大幅度提升, 相比 YOLOv5s 基础网络模型 mAP 值提升了 2.4%, 参数量和计算量分别下降了 0.9% 和 5.7%。

3.5 不同裁剪比例下的模型检测性能分析

表 4 为不同裁剪比例下的模型检测性能指标值, 其中参数量(Params)和计算量(FLOPs)用于衡量模型的复杂度。在本文方法中, 模型裁剪前的参数量为 6976143, 计算量为 14.9×10^9 FLOPs。从表 4 可知, 在不同裁剪比例下, 模型的计算量和浮点数都在明显降低, 随着裁剪比例增大, 模型的计算量和参数量呈递减趋势且所占用的存储空间变小。当裁剪比例取 0.75 时模型的准确率 P 值和平均精度均值 mAP 最大。图 7 为裁剪比例取 0.75 时通道剪枝前后网络各卷积层输出的通道数统计。

表 4 不同裁剪比例下模型性能指标值

Table 4 Model performance metrics values at different cropping ratios

裁剪率	$P/\%$	mAP/ $\%$	Params	FLOPs	模型大小/MB
0.60	85.9	84.7	2243724	9.6×10^9	4.9
0.65	86.1	85.4	2035275	9.3×10^9	4.5
0.70	86.9	85.3	1848501	8.9×10^9	4.1
0.75	88.4	86.2	1686764	8.1×10^9	3.8
0.80	85.6	85.3	1561683	7.5×10^9	3.6

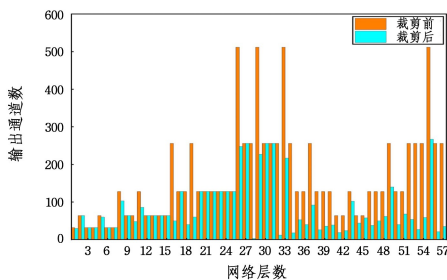


图 7 卷积层输出通道数

Fig. 7 Number of output channels of convolution layer

3.6 NWPU VHR-10 数据集上目标检测性能分析

为了对比分析本文所提出的轻量化目标检测模型相对于其他轻量化目标检测模型的检测性能, 本文分析了目前几种主流的轻量化目标检测模型在 NWPU VHR-10 数据集上的检测性能, 结果如表 5 所列。

表 5 NWPU VHR-10 数据集上不同目标检测模型性能对比

Table 5 Comparison of detection performance of different objects on NWPU VHR-10 dataset

模型	$P/\%$	mAP/ $\%$	Params	FLOPs	模型大小/MB	FPS
YOLOv5s	85.4	84.6	7037095	15.8×10^9	14.5	150
YOLOv3-tiny	86.6	82.0	8687482	12.9×10^9	16.6	123
YOLOv5-lite	86.7	80.7	5309095	15.2×10^9	10.7	156
YOLOv7-tiny	83.8	75.6	6039342	13.3×10^9	12.3	127
YOLOv5s-BSN	88.5	86.6	6976143	14.9×10^9	14.5	130
本文模型	88.4	86.2	1686764	8.1×10^9	3.8	222

由表 5 可知, YOLOv5s-BSN 模型的 P 值、mAP 值、参数量、计算量、模型大小、检测速度分别为 88.5%, 86.6%,

6976143 , 14.9×10^9 , 14.5 MB, 130 FPS; 与基础网络 YOLOv5s 相比, 在模型大小保持不变和检测速度降低 20FPS 的基础上, 模型的 P 值提升了 3.6%, mAP 值提升了 2.4%, 计算量和参数量分别降低了 0.9%, 5.7%。本文模型是通过对比 YOLOv5s-BSN 模型进行通道剪枝得到的, 与剪枝前的 YOLOv5s-BSN 相比较, 在 P 值和 mAP 值分别降低了 0.1% 和 0.4% 的基础上, 模型计算量和参数量分别降低了 75.8%, 45.6%, 模型大小压缩了 73.8%, 检测速度提升了 70.8%。与基础网络 YOLOv5s 相比, 模型的 P 值和 mAP 值分别提升了 3.5%, 1.9%, 模型参数量和计算量降低了 76%, 48.7%, 节省了 10.7 MB 的内存空间, 模型大小压缩了 73.8%, 检测速度提升了 48%。相较于轻量级网络 YOLOv3-tiny, YOLOv5-lite, YOLOv7-tiny, 本文模型在准确率 P 、均值平均精度 mAP、参数量、计算量、模型大小、检测速度这 6 个指标中均有明显优势。在 NWPU VHR-10 数据集上的目标检测部分结果如图 8 所示。可以看出, YOLOv3-tiny 和 YOLOv5-lite 以及 YOLOv7-tiny 这些轻量化目标检测模型对具有复杂背景、目标密集以及小目标占比大等特点的航拍图像存在误检和漏检的情况。然而, 本文模型通过对 YOLOv5s 主干网络和多尺度特征融合网络的优化改进, 在保证目标检测模型进一步轻量化的同时, 使模型更加关注图像中目标特征的提取, 从而提升了对航拍图像中密集目标和小目标的检测能力。

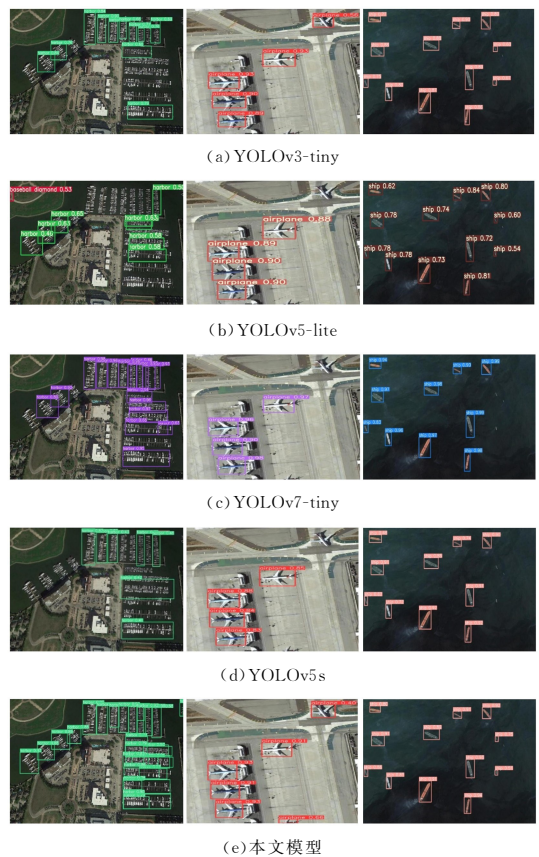


图 8 NWPU VHR-10 数据集上不同模型检测结果对比图
Fig. 8 Comparison of detection results of different models on NWPU VHR-10 dataset

3.7 DroneVehicle 数据集上目标检测性能分析

为了进一步验证本文轻量化目标检测模型对航拍图像目标检测的性能, 在 DroneVehicle 数据集上对模型检测性能进

行评估,实验结果如表 6 所列。

表 6 DroneVehicle 数据集上不同目标检测模型性能对比
Table 6 Comparison of detection performance of different objects on DroneVehicle dataset

模型	P/%	mAP/%	Params	FLOPs	模型大小/ MB	FPS
YOLOv5s	72.0	69.3	7 037 095	15.8×10^9	14.5	207
YOLOv3-tiny	68.4	61.6	8 675 932	12.9×10^9	17.4	232
YOLOv5-lite	71.0	68.3	5 303 290	15.1×10^9	11.3	300
YOLOv7-tiny	71.4	70.0	6 018 420	13.1×10^9	12.3	347
YOLOv5s-BSN	72.7	70.8	6 962 658	14.9×10^9	14.5	102
本文模型	71.3	69.7	1 679 354	8.1×10^9	3.8	328

由表 6 可知,YOLOv5s-BSN 的 P 值与 mAP 值比基础模型 YOLOv5s 高出了 0.7% 和 1.5%。本文模型是对 YOLOv5s-BSN 模型进行通道剪枝得到的,与剪枝前的 YOLOv5s-BSN 相比,P 值和 mAP 值分别降低 1.4% 和 1.1%,模型计算量和参数量分别降低了 75.8%,45.6%,模型大小压缩了 73.8%,检测速度提升了 221%。为了直观地评价本文模型的检测性能,在 DroneVehicle 数据集上对目标检测结果进行了可视化,如图 9 所示。从图 9 可以看出,本文模型相对于 YOLOv3-tiny, YOLOv5-lite 以及 YOLOv7-tiny 这些轻量化目标检测模型,更加关注图像中车辆目标特征的提取,从而提升了航拍图像中密集小目标的检测性能。

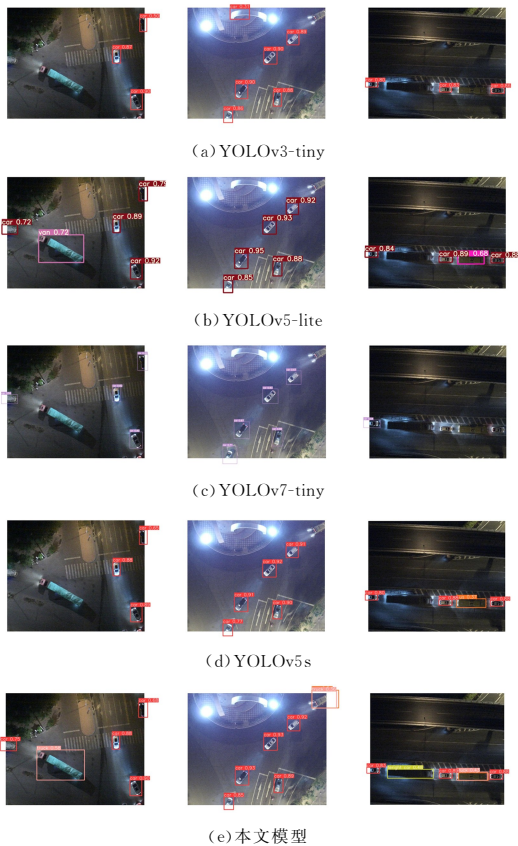


图 9 DroneVehicle 数据集上不同模型检测结果对比图
Fig. 9 Comparison of detection results of different models on DroneVehicle dataset

结束语 针对航拍图像目标检测模型不易在无人机搭载的嵌入式设备上部署的问题,本文提出了一种轻量化航拍图像目标检测模型。以 YOLOv5s 作为基础模型,首先使用改进的 ShuffleNetv2 网络替换 YOLOv5s 主干网络 C3 模块中

的 BottleNeck,减少主干网络层数,降低模型参数数量和计算复杂度;其次在 YOLOv5s 多尺度特征融合网络中引入 SE 通道注意力机制,以提高模型对不同尺度关键特征的捕捉和提取能力;最后采用通道剪枝的方法对改进后的模型进行通道剪枝,裁剪掉模型中不重要的通道,进一步使得模型轻量化,从而减少了模型的运算量,提高了模型的推理速度。但当前的工作仅在理论上完成对航拍图像的目标检测,后续考虑将模型部署于无人机搭载的嵌入式设备进行实验验证。

参考文献

- [1] JUNOS M H, KHAIRUDDIN A S M, DAHARI M. Automated object detection on aerial images for limited capacity embedded device using a lightweight CNN model [J]. Alexandria Engineering Journal, 2022, 61(8): 6023-6041.
- [2] LIU S, D LIU Y H, SUN Y M, et al. Small object detection in UAV aerial images based on inverted residual attention [J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(3): 514-524.
- [3] DENG L, BI L, LI H, et al. Lightweight aerial image object detection algorithm based on improved YOLOv5s [J]. Scientific Reports, 2023, 13(1): 7817.
- [4] CORTES C, VAPNIK V N. Support Vector Networks [J]. Machine Learning, 1995, 20(3): 273-297.
- [5] STATISTICS L B, BREIMAN L. Random Forests [J]. Machine Learning, 2001, 45(1): 5-32.
- [6] LOWE D G. Distinctive Image Features from Scale-Invariant Keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [7] GUO L, WANG Q L, XUE W, et al. A Small Object Detection Algorithm Based on Improved YOLOv5 [J]. Journal of University of Electronic Science and Technology of China, 2022, 51(2): 251-258.
- [8] LIU P, LI K, ZHOU X, et al. RFP-based Faster R-CNN in Aerial Image Detection [C] // Proceedings of the 2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE). IEEE, 2020: 325-329.
- [9] 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 & Machine Intelligence, 2017, 39(6): 1137-1149.
- [10] HUANG Y C, LI L H, MA J J, et al. Research on aerial photography target detection algorithm based on improved SSD algorithm [J]. Journal of Telemetry, Tracking and Command, 2022, 43(3): 79-85.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C] // proceedings of the Computer Vision—ECCV 2016; 14th European Conference, Amsterdam, The Netherlands, Part I 14. Springer, 2016: 21-37.
- [12] FENG Z Q, XIE Z J, BAO Z W, et al. Real-time dense small object detection algorithm for UAV based on improved YOLOv5 [J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(7): 251-265.
- [13] XIE C H, WU J M, XU H Y. Small Object Detection Algorithm

Based on Improved YOLOv5 in UAV Image[J]. Computer Engineering and Applications, 2023, 59(9):198-206.

- [14] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [J]. arXiv:1704.04861, 2017.
- [15] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:4510-4520.
- [16] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3 [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:1314-1324.
- [17] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:6848-6856.
- [18] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design [C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018:116-131.
- [19] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More features

from cheap operations [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:1580-1589.

- [20] YE L. AugShuffleNet: Communicate More, Compute Less [J]. arXiv:2203.06589, 2022.
- [21] ZHANG L, CAI J. Target detection system based on lightweight Yolov5 algorithm [J]. Computer Technology and Development, 2022, 32(11):134-139.
- [22] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:7132-7141.
- [23] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:8759-8768.



CHEN Haiyan, born in 1978, Ph.D, associate professor. Her main research interests include image processing and pattern recognition.