湖南大学学报(自然科学版)
Journal of Hunan University(Natural Sciences)
ISSN 1674-2974,CN 43-1061/N

《湖南大学学报(自然科学版)》网络首发论文

题目：基于改进 YOLOv5 的轻量化无人机检测算法
作者：彭艺，涂馨月，杨青青，李睿
收稿日期：2022-11-18
网络首发日期：2023-09-04
引用格式：彭艺，涂馨月，杨青青，李睿．基于改进 YOLOv5 的轻量化无人机检测算法[J/OL]．湖南大学学报(自然科学版).
https://link.cnki.net/urlid/43.1061.n.20230901.1403.002

基于改进 YOLOv5 的轻量化无人机检测算法

彭艺[1,2]，涂馨月[1]，杨青青[1,2†]，李睿[1]
（1. 昆明理工大学 信息工程与自动化学院，云南 昆明 650031；
2. 昆明理工大学 云南省计算机技术应用重点实验室，云南 昆明 650500）

摘　要：针对现有的无人机检测算法无法同时兼顾检测速度及检测精度的问题，本文提出了一种基于 YOLOv5s 的轻量化无人机检测算法 TDRD-YOLO，该算法首先以 YOLOv5s 的多尺度融合层和输出检测层分别作为颈部网络和头部网络，引入 MobileNetv3 轻量化网络对原骨干网络进行重构，并将骨干网络后的通道在原 YOLOv5s 的基础上进行压缩，减小网络模型大小；其次，将骨干网络中 Bneck 模块的注意力机制由 SE 修改为 CBAM，并在颈部网络引入 CBAM，使网络模型更加关注目标特征；最后修改颈部网络的激活函数为 h-swish，进一步提高模型精度。实验结果表明：本文提出的 TDRD-YOLO 算法平均检测精度达到 96.8%，与 YOLOv5s 相比，参数量减小 11 倍，检测速度提升 1.5 倍，模型大小压缩 8.5 倍.实验验证了本文算法可在大幅降低模型大小、提升检测速度的同时保持良好的检测性能.

关键词：无人机检测；YOLOv5；轻量化；注意力机制；深度学习

中图分类号：TP391　　**文献标志码：**A

Lightweight UAV detection algorithm based on improved YOLOv5

Peng Yi[1,2]　Tu Xinyue[1]　Yang Qingqing[1,2†]　Li Rui[1]
(1. School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650031, China;
2. Yunnan Key Laboratory of Computer Technologies Application, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: Aiming at the problem that existing UAV detection algorithms cannot simultaneously take into account detection speed and accuracy, a lightweight UAV detection algorithm TDRD-YOLO based on YOLOv5s is proposed in this paper. Firstly, the multi-scale fusion layer and output detection layer of YOLOv5s are used as neck network and head network respectively. MobileNetv3 lightweight network was introduced to reconstruct the original backbone network, and the channel behind the backbone network was compressed on the basis of the original YOLOv5s to reduce the size of the network model. Secondly, the attention mechanism of Bneck module in backbone network is modified from SE to CBAM, and CBAM is introduced in neck network to make the network model pay more attention to the target features. Finally, the activation function of the neck network was modified to h-swish to further improve the accuracy of the model. Experimental results show that the average detection accuracy of TDRD-YOLO algorithm proposed in this paper reaches 96.8%. Compared with YOLOv5s, the number of parameters is reduced by 11 times, the detection speed is increased by 1.5 times, and the model size is reduced by 8.5

收稿日期：2022-11-18
基金项目：国家自然科学基金(61761025)，National Natural Science Foundation of China(61761025)；云南计算机技术应用重点实验室开放基金项目资助(2021102)，Development Fund of Key Laboratory of Computer Technology Application in Yunnan Province(2021102)
作者简介：彭艺（1975-），女，云南人，昆明理工大学副教授
†通信联系人，E-mail：1814813449@qq.com

times. Experiments show that the proposed algorithm can greatly reduce the model size and improve the detection speed while maintaining good detection performance.

Key words:UAV detection; YOLOv5; Lightweight; Attention mechanism; Deep learning

小型无人机是一种小体积、低成本、低高度的飞行器，由于其便携性和易操作性，小型无人机已被广泛应用于军事和民用领域，无人机的普及在为社会带来便利的同时，也伴随着因非法使用无人机而引发的侵犯隐私、危害安全等诸多问题，因此针对飞行状态下无人机目标检测成为解决此问题的途径。以往的无人机检测主要依靠音频信号分析、雷达信号分析、射频信号分析的技术，文献[1]提出了一种基于音频信号的无人机检测方法，但在嘈杂环境中，无人机发出的音频信号容易受噪声干扰；文献[2]提出了利用雷达信号进行无人机检测，但无人机的尺寸和材料对检测结果影响较大；文献[3]提出通过利用 CNN 采集无人机与其控制器之间实时通信发射的射频进行无人机检测，但无法检测无无线连接的目标。

随着深度学习的发展，卷积神经网络在计算机视觉领域得到广泛应用，这为无人机检测提供了新思路。目标检测是计算机视觉中最核心的任务之一，其应用包括人脸检测、车辆检测、交通标志检测、医学影像分析等。目前，基于深度学习的目标检测主要分为两类：两阶段检测和单阶段检测[4]，代表性的两阶段检测算法包括 R-CNN[5]、FAST R-CNN[6]、FASTER R-CNN[7]、MASK R-CNN[8]等，单阶段检测算法包括 YOLO 系列[9-12] 、SSD[13]、DSSD[14]、RetinaNet[15]等。两阶段检测算法先生成候选区域再将候选区域放入分类器中进行分类和修正，单阶段检测直接对预测的目标进行回归，因此，两阶段检测具有良好的定位和物体识别精度，而单阶段检测具有良好的推理速度。YOLO 系列算法以其速度-精准度的均衡性[16]深受目标检测领域的青睐，文献[17]中利用 YOLOv3、YOLOv4、YOLOv5 进行无人机检测实验，文章中三种算法的检测精度分别为 92%、91%、95%，YOLOv5 在无人机检测中占优。然而以上算法都存在模型过大，计算复杂，难以生产部署与在移动端移植等问题。

为满足智能时代下对目标检测实时性及高效性的要求，研究者们提出了众多面向移动端的目标检测优化思路，主要包括设计轻量化网络和压缩模型。2016 年，文献[18]中提出一种轻量化模型 SqueezeNet，使用小核卷积代替原来的卷积对特征维度进行压缩，但增加了网络结构深度，并行能力差，推理时间被延长。文献[19-21] 中提出 MobileNet 系列的轻量化网络模型，通过引入深度可分离卷积、导致残差结构、线性瓶颈、SE 机制等来提升模型运算速度、减小参数量，但网络本身参数少，提取特征能力不足，识别精度低。2018 年，文献[22]中提出 ShuffleNetv1 网络模型，采用通道变换对组卷积得到的特征图进打乱，解决了分组导致的组间特征信息不通的问题，随后文献[23]中提出的 ShuffleNetv2 网络采用与输入特征通道数相同的卷积核来最小化占用，增加了数据并行度，但以上两种算法都存在输入输出的维度不一，通道变换需要大量指针跳转的问题。2020 年，文献[24]中提出 GhostNet 网络，借助少量卷积核线性生成 Ghost 特征图，显著减少模型的参数量，提高运算速度，但检测准确度低。2022 年，文献[25]中提出 MoCoViT 网络，结合了移动自注意力模块与移动前馈网络，能够很好地应用于移动设备，但在空间维度上的特征融合能力差。

本文针对现有算法无法同时兼顾检测精度及检测速度的问题，提出一种基于 YOLOv5s 的轻量化无人机检测算法 TDRD-YOLO，采用将 MobileNetv3 与 YOLOv5s 结合的框架作为初始网络结构，利用 MobileNetv3 网络替代 YOLOv5s 的骨干网络，并对原颈部和头部网络的通道进行压缩，减小模型大小。针对精度问题，首先通过在骨干网络的倒残差结构和颈部网络中引入注意力机制 CBAM，来提高模型对目标特征的关注程度，其次在颈部网络的卷积层引入激活函数 h-swish 来消除由 sigmoid 引起的潜在精度损失。最后对改进网络进行消融及对比实验，实验结果表明该算法在能够保持精度的同时大幅减小模型大小、提高检测速度。

1. 相关理论

1.1 YOLOv5算法

YOLO(You Only Look Once) 是由 Joseph Redmon 等在 2016 年提出的一种端到端目标检测算

法，其核心思想是把目标检测看作一个线性回归问题，与滑动窗口不同的是，YOLO 的基本思路是将输入图片划分成$S\times S$个网格，将图像分类及定位算法依次应用于每一个网格中，仅有存在目标中心点的网格会对目标进行下一步的预测。另一方面，YOLO 可以显式地输出边界框的坐标，因此神经网络输出的边界框坐标可以具有任意宽高比，并且能获取更精确的坐标。

目前 YOLO 系列算法包括的模型有 YOLOv3、YOLOv4、YOLOv5 等，与前几代模型相比，YOLOv5 在速度-精度方面有更好的均衡性，同时，与 YOLOv3 和 YOLOv4 不同，YOLOv5 具有四种不同量级的网络模型[26]：YOLOv5x、YOLOv5l、YOLOv5m、YOLOv5s，其中，YOLOv5s 模型的网络深度及宽度最小，更加适用于在移动端的目标检测，因此本文基于 YOLOv5s 模型进行网络框架的改进。YOLOv5 的基本网络模型如图 1 所示，模型由四个部分组成：输入层（Input）、骨干网络（Backbone）、颈部网络（Neck）、头部（Head）；骨干网络是模型的特征提取部分，颈部网络是模型的特征融合部分[27]。

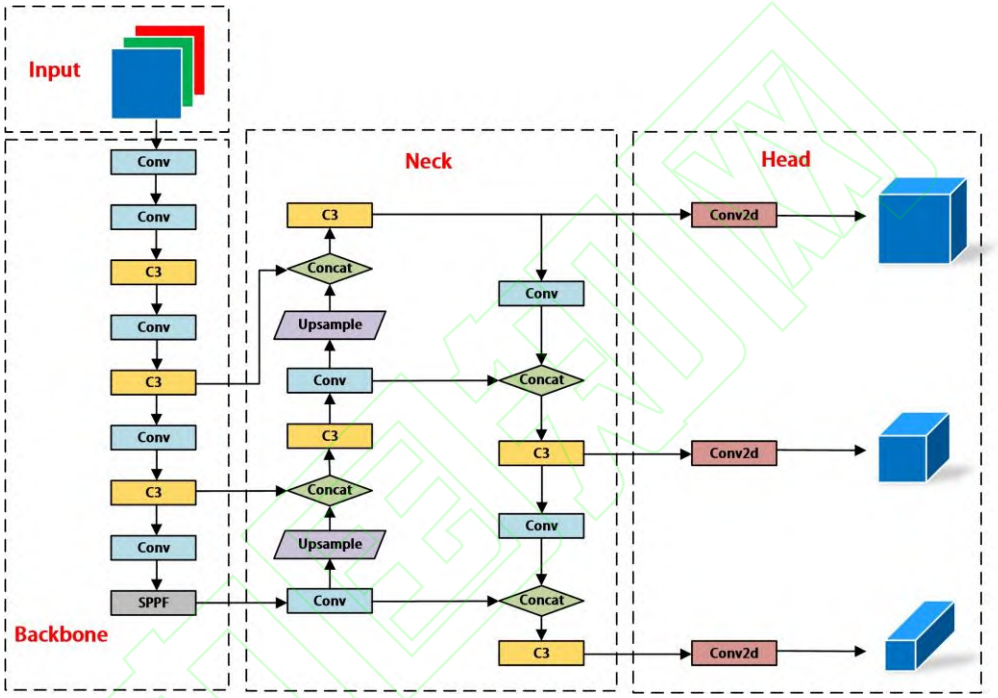

图 1 YOLOv5 网络框架结构

Fig. 1 YOLOv5 Network Structure

1.2 CBAM

CBAM（Convolutional Block Attention Module）[28]是一个卷积注意力机制模块，它融合了通道注意力（Channel Attention Module, CAM）和空间注意力(Spatial Attention Module, SAM）两个机制。CAM 将特征图输入两个并行的池化层分别进行最大池化和平均池化处理，处理后的特征图传入具有共享权重功能的多层感知器（MLP），MLP 输出的两个特征进行逐像素相加后通过 Sigmoid 激活函数生成通道特征权重，CAM 的结构如图 2 所示。SAM 将特征图输入全局最大池化层和全局平均池化层，将池化处理后的两个结果进行全连接，然后经过卷积降维后通过 Sigmoid 激活函数生成空间特征权重，SAM 的结构如图 3 所示。

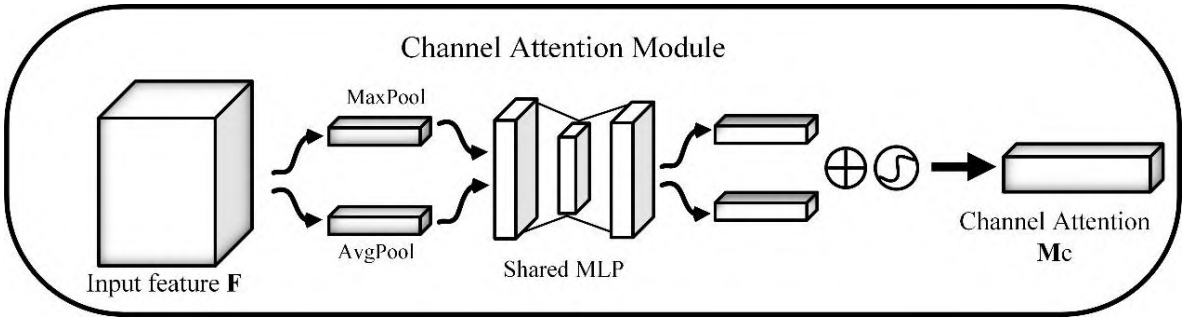

图 2 通道注意力模型结构

Fig. 2 Channel Attention Model Structure

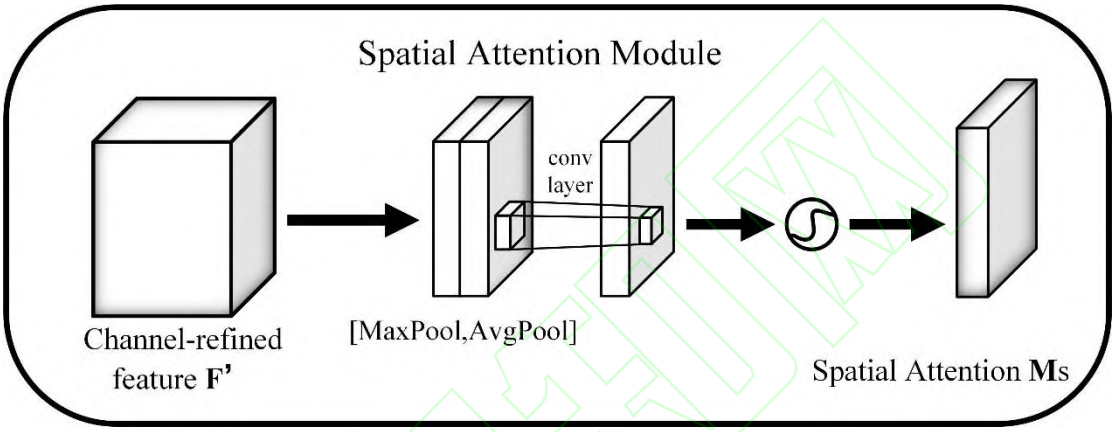

图 3 空间注意力模型结构

Fig. 3 Spatial Attention Model Structure

按 CAM 在前、SAM 在后的顺序将两个模块组合在一起形成了 CBAM，原始特征图首先输入 CAM 生成通道特征权重，该权重与原始特征图相乘，对特征进行自适应调整，然后继续通过 SAM，进行与 CAM 自适应调整过程同样的操作，最后输出的特征图更加关注目标的特征，从而达到提高训练准确度的目的。CBAM 的结构如图 4 所示，其中 CAM 和 SAM 在获取特征权重的时候均用到最大池化和平均池化，2 种池化的结合有利于模型提取更加丰富的特征，另外，CBAM 作为一个轻量级即插即用的通用注意力模块，可以以较小的代价添加到任意网络的卷积层后面。

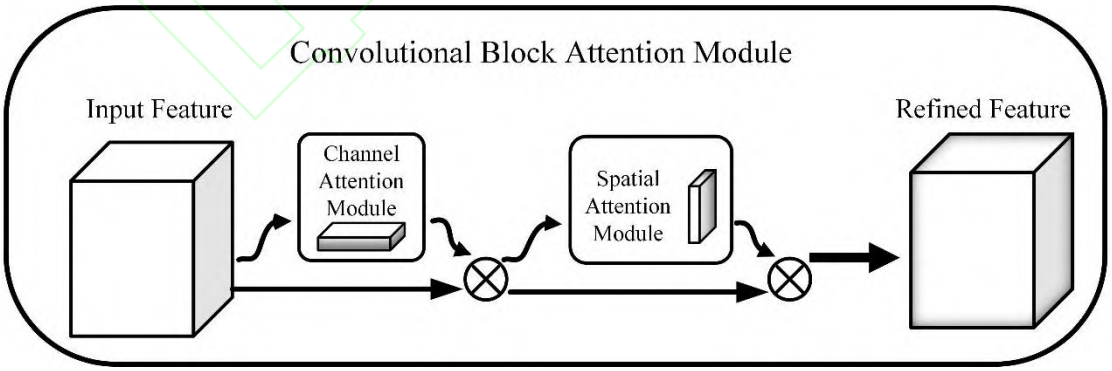

图 4 CBAM 模型结构

Fig. 4 CBAM Model Structure

1.3 MobileNet v3算法

MobileNetv1 是由 Google 团队于 2017 年提出的一种轻量级卷积神经网络，其核心思想是利用深度可分离卷积（Depthwise Separable Conv，DSC）[29]大大减小计算量和模型的大小；MobileNetv2[30]在 MobileNetv1 的基础上加入了倒残差结构（Inverted Residuals）和线性瓶颈结构（Linear Bottlenecks），使得模型参数量降低、精确度提高；MobileNetv3[31]引入了 SE（Squeeze-and-Excitation）通道注意力机制，使用 h-swish(x) 激活函数来代替 ReLu 6，移除了

MobileNetv2 输出端由一个3×3卷积和一个1×1卷积构成的瓶颈层连接，并将平均池化层前移，进一步降低网络参数、减少计算量。

深度可分离卷积分为逐通道分离（Depthwise Conv，DW）和逐点分离（Pointwise Conv，PW）两个过程[32]。DW 的卷积核数与输入图像道数相同，通道与卷积核是一一对应的关系，所以 DW 输出的特征图数与输入图像的通道数相同；PW 将 DW 输出的特征图输入1×1的卷积核进行卷积，使输出的每个特征图均包含输入层所有特征图的信息。深度可分离卷积的原理图如图 5 所示。

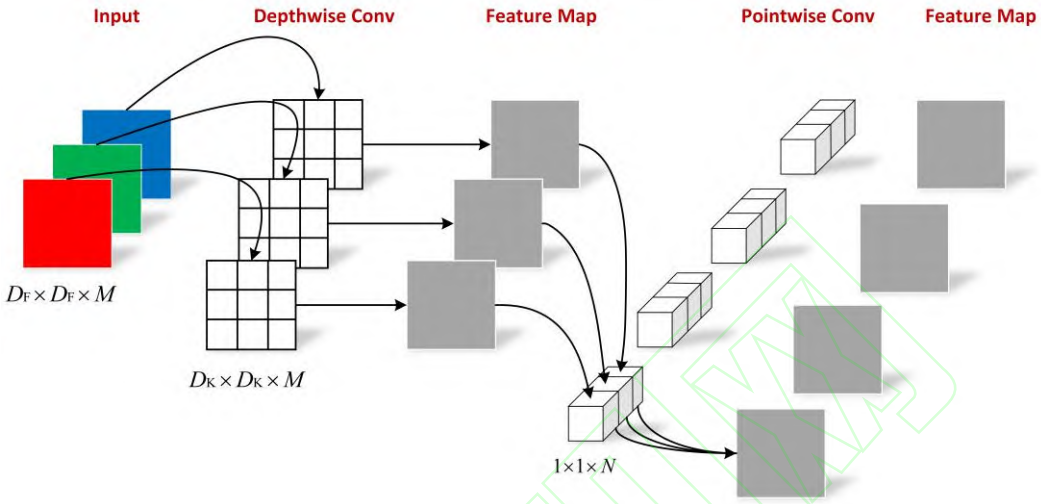

图 5 深度可分离卷积原理图

Fig. 5 Depth-separable Convolution Schematic

图中，令输入图像大小为 $D_F \times D_F \times M$，与大小为 $D_K \times D_K \times M$ 的卷积核进行卷积得到 M 通道特征图，接着 M 通道特征图输入大小为 $1 \times 1 \times N$ 的卷积核，得到 N 通道特征图。整个过程的计算量如下：

$$D_F \cdot D_F \cdot M \cdot D_K \cdot D_K + D_F \cdot D_F \cdot M \cdot N \quad (1)$$

若将大小为 $D_F \times D_F \times M$ 的输入图像通过大小为 $D_K \times D_K \times N$ 的普通卷积得到与上述过程相同的特征图，计算量如下所示：

$$D_K \cdot D_K \cdot D_F \cdot D_F \cdot M \cdot N \quad (2)$$

（1）式与（2）式的比值为：

$$\frac{1}{D_K^2} + \frac{1}{N} \quad (3)$$

在特征提取时一般选用的卷积核大小为 3×3，即理论上深度可分离卷积较普通卷积计算量减小了 8-9 倍。

MobileNetv3 网络的核心模块为 Bottleneck（Bneck），该模块沿用了 MobileNetv2 的倒残差结构，将输入特征图首先通过1×1的卷积进行升维，使得网络可以提取更多特征，然后进行3×3的 DW 卷积。与 MobileNetv2 不同的是，这里将 DW 卷积后生成的特征图经过 SE 模块，调整每个通道的权重，提高模型的精度。最后通过1×1卷积降维，当输入和输出特征的层数相同时，Bneck 会采用跳跃连接（Shortcut），Bneck 网络结构如图 6 所示。

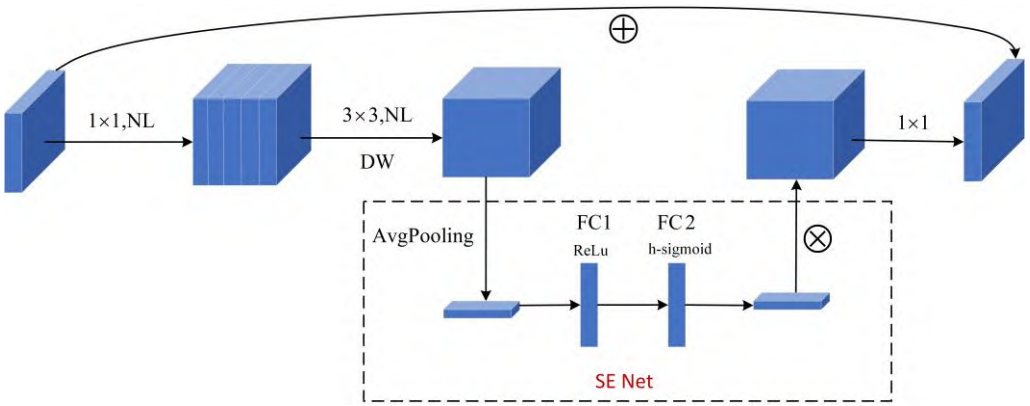

图 6 MobileNetv3 的 Bneck 结构

Fig. 6 Bneck of MobileNetv3 Structure

2. 改进的 MobileNetv3

2.1 改进的MobileNetv3

MobileNetv3 在 Bneck 结构中采用了 SE 注意力机制，SE 分为压缩（Squeeze）和激发（Excitation）两个过程，特征图首先通过池化层进行压缩提取全局特征，再将全局特征通过两个全连接层对各通道分配权重系数，使模型更专注于权重高的通道，提高训练精确度。

本文基于 MolieNetv3s 的特征提取网络结构，将其 Bneck 中的注意力机制改进为 CBAM。与 SE 相比，CBAM 将注意力同时运用于通道和空间两个维度上，提高模型精确度；另一方面，改进后的网络模型参数量得到了进一步降低。改进后的特征提取网络结构如表 1 所示。

表 1 改进的 MobileNetv3 特征提取网络结构

Tab. 1 Improved feature extraction network structure of MobileNetv3

Input	Operator	Kernel	Stride	exp size	CBAM	NL
$640^2\times3$	Conv	3	2	-	-	HS
$320^2\times16$	Bneck	3	2	16	√	RE
$160^2\times16$	Bneck	3	2	72	-	RE
$80^2\times24$	Bneck	3	1	88	-	RE
$80^2\times24$	Bneck	5	2	96	√	HS
$40^2\times40$	Bneck	5	1	240	√	HS
$40^2\times40$	Bneck	5	1	240	√	HS
$40^2\times40$	Bneck	5	1	120	√	HS
$40^2\times48$	Bneck	5	1	144	√	HS
$40^2\times48$	Bneck	5	2	288	√	HS
$20^2\times96$	Bneck	5	1	576	√	HS
$20^2\times96$	Bneck	5	1	576	√	HS
$20^2\times96$	Conv	1	1	-	√	HS

表中，Kernel 表示卷积核大小，Stride 表示卷积步距，exp size 表示 Bneck 模块的第一个 1×1 卷积的输出维度，CBAM 表示是否使用注意力机制，NL 表示使用的激活函数种类，其中，RE 表示激活函数为 ReLu6，HS 表示激活函数为 h-swish 。

2.2 TDRD-YOLO

虽然 YOLOv5s 模型的网络深度及宽度在整个 YOLOv5 体系中最小，但其在 Backbone 层仍然存在大量的卷积过程，卷积操作庞大的运算量使得模型繁杂、参数大，从而导致生成的训练权重文件体积大、检测耗时长。针对这一问题，本文设计了一种更加轻量化的网络模型 TDRD-YOLO（Tiny Drone Real-time Detection-YOLO），以适用于移动端的无人机检测。

TDRD-YOLO 的基本框架是将 YOLOv5s 的骨干网络替换成改进后的 MobileNetv3 特征提取网络，形成一种新的网络模型 MobileNetv3-YOLOv5s，大幅减小模型大小和参数量。TDRD-YOLO 的网络结构如图 7 所示。

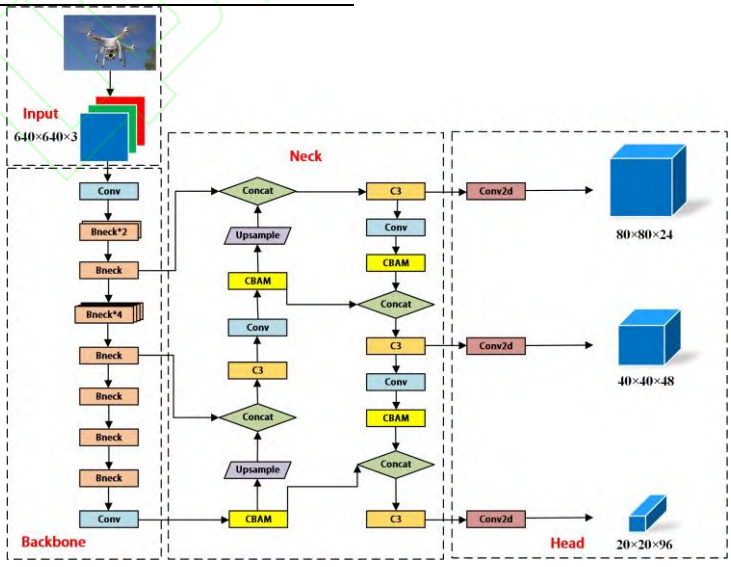

图 7 TDRD-YOLO 网络框架结构

Fig. 7 TDRD-YOLO Network Framework

骨干网络采用 Bneck 模块进行特征提取，Bneck 在保持输入和输出能够紧凑链接的情况下，将内部通道进行升维再降维，以提高非线性全通道变换的表达能力提取丰富的特征，又有效减小计算量。

颈部网络部分采用 C3 结构，通过残差块来提高网络的特征融合能力，使模型在保持丰富特征信息的同时，减少计算量。在 Conv 层将激活函数改进为 h-swish，h-swish 的表达式如式（4）所示。

$$\text{h-swish}[x]=x\frac{\text{ReLU}6(x+3)}{6} \tag{4}$$

虽然用非线性的 swish 函数代替 ReLU 可以显著提高神经网络的准确性，但因为其计算量大，不适合嵌入移动设备，而 h-swish 使用 sigmoid 函数的分段线性硬模拟（piece-wise linear hard analog）$\frac{\text{ReLU}6(x+3)}{6}$ 来近似代替 swish 中的 sigmoid，解决了 swish 函数计算量太大难以嵌入移动端的问题，另一方面，ReLU6 几乎可以在所有软件和硬件框架上使用，并且在量化的时候，它消除了由 sigmoid 引起的潜在精度损失。此外，TDRD-YOLO 采用多尺度预测的方法，将输入图像大小规整为 640×640，经骨干网络处理后输出大小为 20×20×48 的特征图，在第 14、第 19 层引入上采样层，将特张图大小分别扩张为 40×40 和 80×80，在第 20、24、28 层引入全连接层，将特征图通道分别调整至 24、48、96，从而生成三个不同尺寸的特征图输入检测层进行预测，分别为 20×20×96、40×40×48、80×80×24，解决了多尺度目标检测的问题；模型还在第 12、17、22、26 层加入 CBAM；尺寸为 $H\times W\times C$ 的特征图首先输入通道注意力模块，进行平均池化和最大池化操作，得到两个 $1\times1\times C$ 的特征图，接着将这两个特征图分别输入 MLP，MLP 输出的两个特征图进行逐像素相加后再通过 sigmoid 函数获得通道特征权重，将该权重与输入的特征图相乘得到通道注意力特征图。为获得空间维度的注意力特征权重，同样地，我们对尺寸为 $H\times W\times C$ 的特征图进行平均池化和最大池化操作，得到两个 $H\times W\times1$ 的特征图，将这两个特征图在通道维度上拼接起来，此时特征图的尺寸为 $H\times W\times2$，接着经过卷积核为 7×7 的卷积使其通道数降为 1，再通过 sigmoid 函数生成空间特征权重。通过这样的操作，我们将特征图在通道和空间两个维度上的权重进行分配，增加有用特征权重，抑制无效特征权重，使模型更加关注包含重要信息的目标区域，抑制无关信息，提高网络关注目标特征的能力。

头部网络通过计算损失函数和非极大抑制（NMS）来进行预测，损失函数包括定位损失（L_{box}）、目标置信度损失（L_{conf}）、类别损失（L_{cls}）三个部分构成，本文在定位损失部分采用 CIoU Loss 作为损失函数，CIoU Loss 的表达式如下：

$$L_{\text{CIoU}}=1-\left(\text{IoU}-\frac{d_o^{\,2}}{d_c^{\,2}}-\frac{v^2}{1-\text{IoU}+v}\right) \tag{5}$$

$$v=\frac{4}{\pi^2}\left(\arctan\frac{w^t}{h^t}-\arctan\frac{w^p}{h^p}\right)^2 \tag{6}$$

其中，d_o 表示实际框中心点和预测框中心点之间的距离，d_c 表示实际框和预测框交集部分的对角线长度，v 为实际框与预测框的宽高比相似度，w^t 和 h^t 表示实际框的宽和高，w^p 和 h^p 表示预测框的宽和高；目标置信度损失和类别损失使用二进制交叉熵（BCE）作为损失函数，目标置信度损失的表达式如（7）式所示，类别损失的表达式如（8）式所示；总损失为三部分损失之的加权和，总损失函数的表达式如（9）式所示。

$$L_{\text{conf}}=-\sum_{\text{i}=0}^{S\times S}\sum_{j=0}^{A}I_{i,j}^{\text{obj}}\left[\hat{c}_i\ln(c_i)+(1-\hat{c}_i)\ln(1-c_i)\right]-\sum_{\text{i}=0}^{S\times S}\sum_{j=0}^{A}I_{i,j}^{\text{nobj}}\left[\hat{c}_i\ln(c_i)+(1-\hat{c}_i)\ln(1-c_i)\right] \tag{7}$$

$$L_{\text{cls}}=-\sum_{\text{i}=0}^{S\times S}\sum_{j=0}^{A}I_{i,j}^{\text{obj}}\sum_{c\in classes}\left[\hat{p}_i\ln(p_i)+(1-\hat{p}_i)\ln(1-p_i)\right] \tag{8}$$

$$Loss=aL_{\text{CIoU}}+bL_{\text{conf}}+cL_{\text{cls}} \tag{9}$$

其中，$S\times S$ 表示划分特征图的网格尺寸，A 表示候选边框的数量，$I_{i,j}^{\text{obj}}$、$I_{i,j}^{\text{nobj}}$ 分别表示第 i 个网格的第 j 个候选边框是否存在正样本和负样本，$\hat{c}_i$、c_i 表示预测类别和实际类别，$\hat{p}_i$、p_i 表示预测的置信度和真实的置信度，a、b、c 分别表示三种损失对应的权重。

3. 实验及结果分析

3.1 数据集及实验环境

针对无人机检测任务，目前尚未有标准的专用数据集，因此，本文中采用自制数据集，从网络上收集了 4424 张无人机图像，格式均为 JPG 格式，使用 labelimg 以 YOLO 格式对图像进行注释后对应生成 4424 份 txt 标签文件，标签数据按照 8∶1∶1 的

比例分为训练集、验证集和测试集，数据集的存储格式结构基于 VOC 2007 建立，部分无人机目标数据如图 8 所示。

本文使用 Win11 系统，实验硬件配置为 Intel(R) Core(TM) i5-12400F 2.50 GHz 处理器，模型在 NVIDIA GeForce RTX 3060 GPU 上运行，实验软件环境为 python3.8.3，pytorch1.10.1，cuda11.3，编译器为 pycharm2020.1，模型训练参数为 epochs=200、batch size=4。

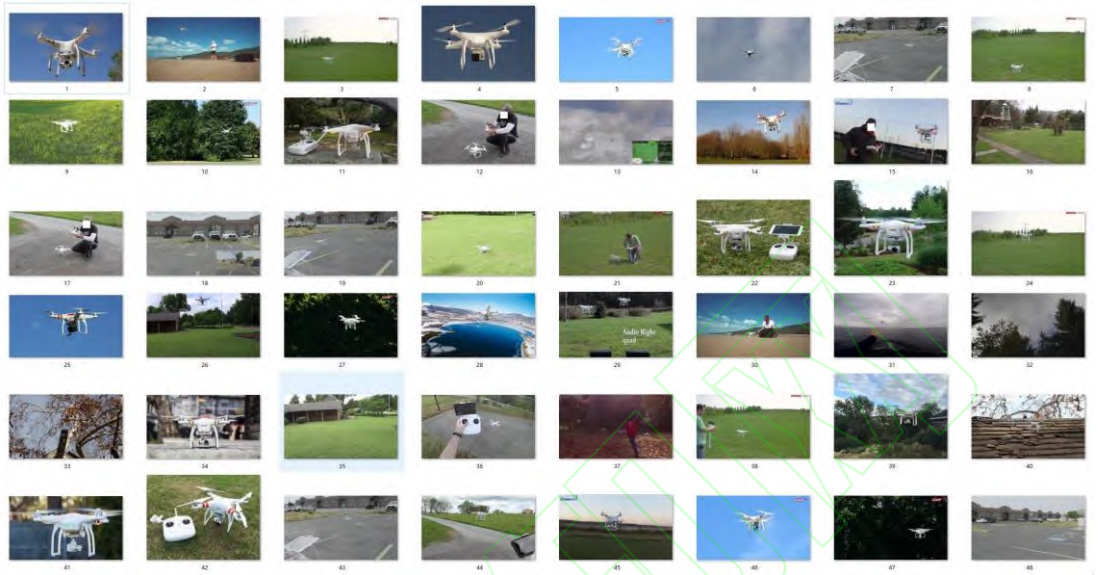

图 8 部分无人机数据集

Fig. 8 Part of The UAV Data Set

3.2 评估指标

本文引用精度（Precision）、召回率（Recall）、平均精度（AP）、平均精度均值（mAP）来评估模型的性能，其中，精度表示预测为正的样本中确实为正的样本所占比例，表达式如式（10）所示；召回率（Recall）表示总的正样本中成功被预测为正的样本所占比例，表达式如式（11）所示；平均精度是对不同召回率点上的准确率求平均，AP 值越大，代表模型的平均准确率越高；平均精度均值表示模型检测到的所有类别的 AP 值的平均值，由于本文仅针对无人机进行检测，故 AP 与 mAP 数值相等。本文使用 mAP@0.5 作为评估模型精度的指标，mAP@0.5 表示 IoU 为 0.5 时的 mAP 值，相较于 mAP@0.5:0.95（IoU 在 0.5 至 0.95 时的 mAP 值），mAP@0.5 只考虑单一阈值因此计算相对简单，更加适合快速检测，此外，对于本文检测的目标对象而言，无人机多为小目标，其边界框较小故通常具有较低的 IoU 值，使用较低 IoU 阈值（0.5）可以提高小目标检测的召回率。

$$precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{10}$$

$$recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{11}$$

其中，TP 表示预测值为正且真实值为正的样本，FP 表示预测值为正但真实值为负的样本，FN 表示预测值为负但真实值为正的样本。

3.3 消融实验

为验证每个改进模块的优化效果，本文设计并进行了消融实验，进一步评估改进技术对 YOLOv5s 算法的影响，实验结果如表 2 所示。

表 2 消融实验结果

Tab. 2 Results of ablation experiment

实验	CBAM(mb)	CBAM(yolo)	h-swish	压缩通道	参数量	mAP@0.5	模型大小
实验一	×	×	×	×	3200784	94.7%	6.8M
实验二	×	×	×	√	627010	92%	1.7M
实验三	√	×	×	√	628074	94.1%	1.7M
实验四	×	√	×	√	628043	94.6%	1.7M

实验	CBAM(mb)	CBAM(yolo)	h-swish	压缩通道	参数量	mAP@0.5	模型大小
实验五	√	√	×	√	626992	94.8%	1.7M
实验六	×	×	√	√	628897	95%	1.7M
实验七	√	√	√	√	628056	96.8%	1.7M

实验一为原始的 MobileNetv3-YOLOv5s 算法，训练参数量为 3200784，mAP@0.5 为 0.947，训练模型的权重大小为 6.8M；实验二为压缩通道后的 MobileNetv3-YOLOv5s 算法，参数量为 627010，mAP@0.5 为 0.92，训练模型的权重大小为 1.7M，可见在压缩通道后训练的参数量和权重文件大幅减小，但 mAP@0.5 的值也随之减小；实验三、实验四、实验五分别将主干网络中 Bneck 层的注意力机制改进为 CBAM，在颈部网络中引入了 CBAM，将主干网络中 Bneck 层的注意力机制改进为 CBAM 的同时在颈部网络中引入了 CBAM，三次实验都在参数量和模型大小基本不变的情况下，平均精度得到了提高；实验六修改颈部网络的激活函数为 h-swish，模型的平均精度较实验二提高 3%；实验七为本文所设计的 TDRD-YOLO 算法，较实验一训练参数量和模型大小都明显减小，且平均准确度有了 2.1%的提升，证明了 TDRD-YOLO 算法的有效性和优越性。

3.4 模型检测结果分析

为进一步验证本文所提方法的有效性，将 YOLOv5s 与 MobileNetv3-YOLOv5s 及本文设计的 TDRD-YOLO 用统一数据集进行检测，不同场景下三种模型的检测效果如图 9、图 10、图 11 所示，其中，（a）、（b）、（c）分别代表 YOLOv5s、MobileNetv3-YOLOv5s 及 TDRD-YOLO 的检测结果。图 9 为简单背景下的检测效果，TDRD-YOLO 与 YOLOv5s 检测的精确度相同且最高，但 YOLOv5s 有错检的情况，TDRD-YOLO 无错检漏检；图 10 为复杂背景下的检测效果，YOLOv5s、MobileNetv3-YOLOv5s 均出现错检的情况，TDRD-YOLO 在精确度较 YOLOv5s 提高 0.03 的同时无错检漏检；图 11 为小目标检测的效果，MobileNetv3-YOLOv5s 的检测精确度较 YOLOv5s 大幅下降，TDRD-YOLO 在大幅度压缩模型的情况下，精确度较 YOLOv5s 仅下降了 0.01。TDRD-YOLO 算法在三种场景下均无错检，且能和 YOLOv5s 保持基本相同的检测精度。

(a) (b) (c)

图 9 简单背景下的检测效果图

Fig. 9 Detection effect in simple background

(a) (b) (c)

图 10 复杂背景下的检测效果图

Fig. 10 Detection effect in complex background

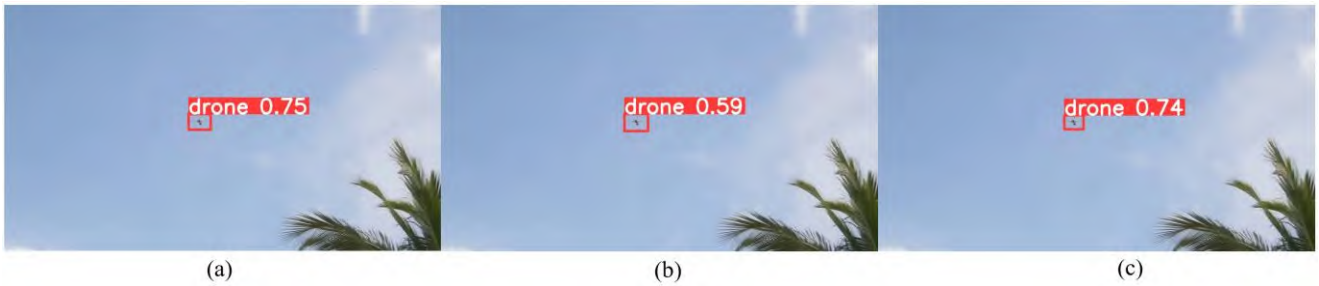

图 11 小目标检测效果图

Fig. 11 Small target detection effect

表 3 显示了三种模型在平均精度(mAP@0.5)、参数量、检测速度及模型大小方面的比较，可以看出，YOLOv5s 在无人机数据集上的检测平均精度明显优于其他模型，但其参数量庞大、检测速度较慢，对比以 YOLOv5s 为基础改进的 MobileNetv3-YOLOv5s 和本文所提出的 TDRD-YOLO 可以看出，虽然 MobileNetv3-YOLOv5s 的参数量、检测速度和模型大小均大幅减小，但 MobileNetv3-YOLOv5s 的检测平均精度下降了 6.1%，而经过本文所设计的方法对其进行改进后，TDRD-YOLO 的检测精度达到了 96.8%，平均精度较 YOLOv5s 而言仅下降 1.3%，检测速度达到 114.64 FPS，已接近检测速度最快的 MobileNetv3-YOLOv5s，实验结果表明 TDRD-YOLO 能够在保持精度的同时达到快速检测，在精度-速度方面均衡性强，由此进一步证明了本文所设计的算法改进的有效性。

表 3 不同模型的检测结果对比

Tab. 3 Comparison of detection results of different models

模型	平均精度(%)	参数量	检测速度（FPS）	模型大小（MB）
Faster R-CNN	93.8	12018832	65.62	18.9
SSD	90.5	5982440	83.71	9.4
MobileNetv3-YOLOv3	87.6	710892	103.52	2.1
MobileNetv3-YOLOv4	90.3	739334	97.81	2.6
YOLOv5s	98.1	7022326	76.92	14.5
MobileNetv3-YOLOv5s	92	627010	116.27	1.7
TDRD-YOLO	96.8	628056	114.63	1.7

结论

本文提出了一种轻量化的无人机检测算法 TDRD-YOLO，该算法通过改进网络结构实现了检测模型的较低的复杂度和较高的检测精确度，使其具备较强的精度-速度均衡性。实验验证，在自制无人机数据集上，TDRD-YOLO 算法的模型参数量比 YOLOv5s 减小近 11 倍，检测速度提升近 1.5 倍，模型大小压缩近 8.5 倍，检测的平均精度仅降低了 1.3%，在有效压缩了模型、提升检测速度的基础上，获得与 YOLOv5s 相当的检测精确度，该算法相较于 YOLOv5s 和 MobileNetv3-YOLOv5s 有更好的精度-速度均衡性；同时，实验表明本文所提算法更适合复杂场景下的无人机检测。因此，该算法满足面向移动端的目标检测的轻量化及时效性要求，为无人机检测的生产部署与移动端移植提供了一定的理论依据。虽然 TDRD-YOLO 在速度-精度方面有较好的均衡性，但较 YOLOv5s 而言检测平均精度有所下降，另一方面，虽然改进后的 TDRD-YOLO 在参数量和模型大小上取得了显著的缩小，但为保证模型检测的准确度，网络结构中仍然存在一系列的卷积和池化操作，这些操作仍然会对检测速度的提升产生限制，后续可作进一步优化。

参考文献

[1] Wang, W. , An, T. F. , Ou, J. P. (2018) Research on passive audio detection and recognition technology of UAV. J. Acoustic technology. , 37(1):89-93.

[2] Zeng H, Zhang H, Chen J, et al. UAV target detection algorithm using GNSS-based bistatic radar[C]//IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019: 2167-2170.

[3] Al-Emadi S, Al-Senaid F. Drone detection approach based on radio-frequency using convolutional neural network[C]//2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT). IEEE, 2020: 29-34.

[4] Jiao L, Zhang F, Liu F, et al. A survey of deep learning-based object detection[J]. IEEE access, 2019, 7: 128837-128868.

[5] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[6] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[7] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.

[8] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.

[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] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[11] Redmon J, Farhadi A.Yolov3: An incremental improvement[J].arXiv preprint arXiv:1804.02767, 2018.

[12] Bochkovskiy A, Wang C Y, Liao H Y M.Yolov4: Optimal speed and accuracy of object detection[J].arXiv preprint arXiv:2004.10934, 2020.

[13] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot Multibox Detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.

[14] Fu C Y, Liu W, Ranga A, et al.Dssd: Deconvolutional single shot detector[J].arXiv preprint arXiv:1701.06659, 2017.

[15] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[16] 邵延华, 张铎, 楚红雨, 等. 基于深度学习的 YOLO 目标检测综述[J]. 电子与信息学报, 2022, 44(10):3697-3708.
SHAO Yanhua, ZHANG Duo, CHU Hongyu, et al. A Review of YOLO Object Detection Based on Deep Learning[J]. Journal of Electronics & Information Technology, 2022, 44(10):3697-3708.

[17] Al-Qubaydhi N, Alenezi A, Alanazi T, et al. Unauthorized Unmanned Aerial Vehicle Detection using YOLOv5 and Transfer Learning[J]. 2022.

[18] Iandola F N, Han S, Moskewicz M W, et al.SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size[J].arXiv preprint arXiv:1602.07360, 2016.

[19] Howard A G, Zhu M, Chen B, et al.Mobilenets: Efficient convolutional neural networks for mobile vision applications[J].arXiv preprint arXiv:1704.04861, 2017.

[20] SANDLER M, HOWARD A, ZHU M, et al. MobileNet V2: inverted residuals and linear bottlenecks[C]//Computer Vision and Pattern Recognition. IEEE, 2018:4510-4520.

[21] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]//International Conference on Computer Vision. IEEE, 2019:1314-1324.

[22] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//Computer Vision and Pattern Recognition. IEEE, 2018:6848-6856.

[23] MA N, ZHANG X, ZHENG H T, et al. ShuffleNet V2: Practical guidelines for efficient CNN architecture design[C]//European Conference on Computer Vision. Springer, 2018:116-131.

[24] HAN K, WANG Y, TIAN Q, et al. GhostNet: More Features From Cheap Operations[C]//Computer Vision and Pattern Recognition. IEEE, 2020:1580-1589.

[25] MA H, XIA X, WANG X, et al.MoCoViT: Mobile Convolutional Vision Transformer[J].arXiv preprint arXiv:2205.12635, 2022.

[26] 王文亮,李延祥,张一帆,等.MPANet-YOLOv5:多路径聚合网络复杂海域目标检测[J].湖南大学学报(自然科学版),2022,49(10):69-76.DOI:10.16339/j.cnki.hdxbzkb.2022360.
WANG Wen-liang,LI Yan-xiang,ZHANG Yi-fan,et al.MAPNet-YOLOv5:Multi-path Aggregation Network for Complex Sea Object Detection[J]. Journal of Hunan University(Natural Sciences), 2022,49(10):69-76.DOI:10.16339/j.cnki.hdxbzkb.2022360.

[27] 余加勇,刘宝麟,尹东等.基于 YOLOv5 和 U-Net3+的桥梁裂缝智能识别与测量[J].湖南大学学报(自然科学版),2023,50(5):65-73.DOI:10.16339/j.cnki.hdxbzkb.2023056.Journal of Hunan University(Natural Sciences), 2023,50(5):65-73. DOI:10.16339/j.cnki.hdxbzkb.2023056.

[28] YU Jia-yong,Liu Bao-lin,YIN Dong,et al. Intelligent Identification and Measurement of Bridge Cracks Based on YOLOv5 and U-Net3+[J].

[29] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[30] Kaiser L, Gomez A N, Chollet F.Depthwise separable convolutions for neural machine translation[J].arXiv preprint arXiv:1706.03059, 2017.

[31] 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.

[32] 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.

[33] Qian S, Ning C, Hu Y. MobileNetV3 for image classification[C]//2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE). IEEE, 2021: 490-497.