

基于 YOLOv5s 的轻量化朱鹮检测算法研究

张凡^{1,2}, 张鹏超^{1,2}, 王磊^{1,2}, 曹锐虎^{1,2}, 王晓鹏^{1,2}, 黄俊霖^{1,2}

(陕西理工大学机械工程学院, 723000, 陕西汉中; 2. 陕西省工业自动化重点实验室, 723000, 陕西汉中)

摘要: 针对当前朱鹮检测算法模型参数较多、计算量大的问题, 本文设计研究了一种基于 YOLOv5s 的高性能轻量化网络模型。首先, 结合 EfficientNet 网络中的 MBConvBlock 对原主干网络进行重构, 大幅降低网络参数; 同时在浅层网络中采用 Stem 模块, 提升浅层网络的特征提取能力; 然后改进卷积注意力模块(CBAM), 即将其中的通道注意力替换为高效通道注意力模块(ECA), 避免了降维操作, 有效提取了邻近通道间的信息, 且大幅降低了通道注意力的参数数量, 并将其嵌入特征融合网络路径聚合网络(PANet)中, 达到了引入微小参数数量而有效提升网络性能的目的, 并将其命名为高效卷积注意力模块(ECBAM)。最后, 在自建朱鹮数据集和公共数据集 PASCAL VOC、COCO 上进行实验, 实验结果表明, 与 YOLOv5s 算法相比, 本文算法模型参数数量降低了 52.37%, 计算次数降低了 54.55%, 在自建朱鹮数据集上 $P_{\text{mAP}}@0.5:0.95$ 仅降低了约 2 个百分点, 达到 0.666, 在公共数据集 PASCAL VOC 上 $P_{\text{mAP}}@0.5$ 达到 0.792, 在公共数据集 COCO 上 $P_{\text{mAP}}@0.5:0.95$ 达到 0.298, 证明了本文算法可在大幅降低网络计算量的同时保持较高的目标检测性能。

关键词: 目标检测; YOLOv5s; 轻量化; 注意力

中图分类号: TP391.41 **文献标志码:**

DOI: 10.7652/xjtuxb202301011 **文章编号:** 0253-987X(2023)01-0110-12

Research on Lightweight Crested Ibis Detection Algorithm Based on YOLOv5s

ZHANG Fan^{1,2}, ZHANG Pengchao^{1,2}, WANG Lei^{1,2}, CAO Ruihu^{1,2},
WANG Xiaopeng^{1,2}, HUANG Junlin^{1,2}

(1. School of Mechanical Engineering, Shaanxi University of Technology, Hanzhong, Shaanxi 723000, China;

2. Shaanxi Key Laboratory of Industrial Automation, Hanzhong, Shaanxi 723000, China)

Abstract: Aiming at the problems of excessive parameters and heavy calculation workload in the current Crested Ibis detection algorithm model, this paper studies and designs a high-performance lightweight network model based on YOLOv5s. First, the original backbone network is reconstructed by referring to MBConvblock in the EfficientNet network to greatly reduce the network parameters; at the same time, the Stem module is used in the shallow network to improve the feature extraction ability of the shallow network; then, the convolutional block attention module (CBAM) is improved, that is, the channel attention is replaced by the efficient channel attention

收稿日期: 2022-04-20。 作者简介: 张凡(1996—), 男, 硕士生; 张鹏超(通信作者), 男, 教授, 博士生导师。 基金项目: 国家自然科学基金资助项目(62176146)。

网络出版时间: 2022-08-01 网络出版地址: <https://kns.cnki.net/kcms/detail/61.1069.T.20220801.1516.004.html>

module (ECA), which avoids the dimensionality reduction operation, effectively extracts the information between adjacent channels, and greatly reduces the number of required parameters of channel attention; the ECA is embedded in the feature fusion network path aggregation network (PANet) for the purpose of introducing a small number of parameters to effectively improve the network performance, and is named as efficient convolutional block attention module (ECBAM). Finally, experiments are carried out on the self-built Crested Ibis data set and the public data sets Pascal VOC and COCO. The experimental results show that compared with YOLOv5s algorithm, the proposed algorithm reduces the number of required model parameters by 52.37% and the calculation workload by 54.55%; it decreases the $P_{mAP}@0.5:0.95$ by only about 2% to 0.666 on the self-built Crested Ibis data set, and reduces the $P_{mAP}@0.5:0.95$ to be 0.792 on the public data set Pascal VOC and 0.298 on the public data set COCO. Experiments show that the proposed algorithm can reduce the amount of network calculation workload while maintaining high detection performance.

Keywords: object detection; YOLOv5; lightweight; attention

朱鹮是世界上最为濒危的鸟类之一,在其保护过程中知悉,其分布情况是朱鹮保护这一任务的关键环节。现阶段探寻野外朱鹮踪迹均靠相关专业人员对其潜在出现区域进行科学考察^[1],由于朱鹮活动范围较大,且生活环境复杂,所以现有方法耗时耗力,效率较低。当前,计算机视觉发展日新月异,其广泛应用于工业、农业^[2-3]等领域,利用固定相机或无人机及计算机视觉技术代替人工巡查可对朱鹮出现区域进行长期监测,自动获得某一区域朱鹮的详细出现情况。因此,使用基于计算机视觉的方法来代替人工巡察具有重要的意义。

近年来,计算机视觉中涌现出许多令人瞩目的基于深度学习的目标检测算法,这些算法可有效改善基于传统机器学习算法在进行目标检测时存在的检测精度低、抗干扰能力差、无法适应复杂环境的问题,大致可以分为两类:一类是两阶段检测算法,在执行时需要分两步完成,首先获得候选区域,然后进行区域内目标分类,标志性算法如 R-CNN^[4] (regions with CNN features)、Fast-R-CNN^[5]、Faster-R-CNN^[6];另一类就是一阶段检测算法,通过目标检测网络直接预测出目标的类别和位置信息,这类基于回归的目标检测算法有 YOLO 系列(you only look once)^[7-11]与 SSD (single shot multibox detector)^[12]。这类算法的优点是检测速度较快、实时性好,但是相比于两阶段检测算法精度略低,以上经典网络都存在模型过大、运算量太大、难以生产部署并在移动端移植的问题。

2015 年 Han 等^[13]论述了一些模型压缩的方法,将裁剪、权值共享和量化编码等方法运用在模型压缩中。2016 年 Jin 等^[14]使用高稀疏度网络来降低网络模型的计算量,通过设计损失函数使得矩阵

稀疏化。2016 年 Wen 等^[15]提出一种结构性稀疏的网络学习方式,通过模型自主训练学习来得到稀疏结构,从而降低模型大小。在随后的研究中,有人将轻量化网络应用于目标检测模块中,其可以很好地降低网络的参数与计算量。2017 年 Howard 提出的著名的轻量化网络 MobileNet^[16],其模型中提出的深度可分离卷积(depthwise separable convolution, DWSCConv)能够在参数大幅降低的情况下,有效保持原模型的预测精度。MobileNet-v2^[17]与 MobileNet-v3^[18]采用了倒残差结构与线性激活层,增强了网络的特征提取能力。2018 年 Zhang 等提出的 ShuffleNet 算法^[19]采用了通道混合机制,实现了不同通道间的信息交流。2020 年 Tan 等提出的 EfficientNet 算法^[20]采用倒残差结构与 SE 注意力模块(squeeze-and-excitation module)融合,有效提升了轻量化网络模型的预测精度。

本文针对现有算法存在网络参数量过大、检测效率低、难以实地部署并在移动端移植的问题,基于 YOLOv5s 目标检测算法提出了一种轻量化的朱鹮检测算法,在能够保持较高模型预测精度的同时大幅降低了模型的计算复杂度。本文的贡献可分为以下 3 个部分。

(1)本文从汉中朱鹮国家级自然保护区与互联网中采集朱鹮的图像资料,使用 LabelImg 对图像中的朱鹮进行标注,并对图像进行旋转、正切、随机裁剪、色彩空间变换等多种方式的数据增强,构建了一个数据量较为充足、能够全面展现生存在各种环境下的朱鹮的数据集。

(2)本文提出了一种基于 YOLOv5s 的轻量化朱鹮检测算法,其对 YOLOv5s 目标检测网络进行

了网络压缩与压缩后的精度提升。使用了 EfficientNet 中的 MBConvBlock 进行模型主干网络重构,实现了网络的轻量化设计,大幅降低了网络的参数量与计算量,在浅层网络中添加了 Stem 模块,明显增强了浅层特征提取能力,在不带入过多计算量的同时提升了网络检测精度。

(3) 改进了卷积注意力模块(convolutional block attention module,CBAM),并将其命名为高效卷积注意力模块(efficient convolutional block attention module,ECBAM),即在其中使用高效通道注意力模块(efficient channel attention module,ECA),然后将其添加在特征融合网络路径聚合网络(path aggregation network,PANet)^[21]中,可以

在略微提升网络计算量的情况下有效提升模型的检测精度,并通过实验进行了验证。

1 基于 YOLOv5s 的轻量化检测算法

1.1 YOLOv5s 目标检测算法

YOLO 目标检测算法属于单阶段目标检测算法,运用回归的思想直接处理整张输入图像,得到目标的类别和位置信息。YOLOv5s 目标检测网络结构可分为 3 个组成部分:主干网络、特征融合网络和头部输出层。YOLOv5s 网络结构如图 1 所示,其中不同颜色模块表示不同操作,其中 Focus、CSPBlock、空间金字塔池化(SPP)模块操作如图 1 所示, 1×1 或 3×3 表示对应模块中卷积操作的卷积核大小。

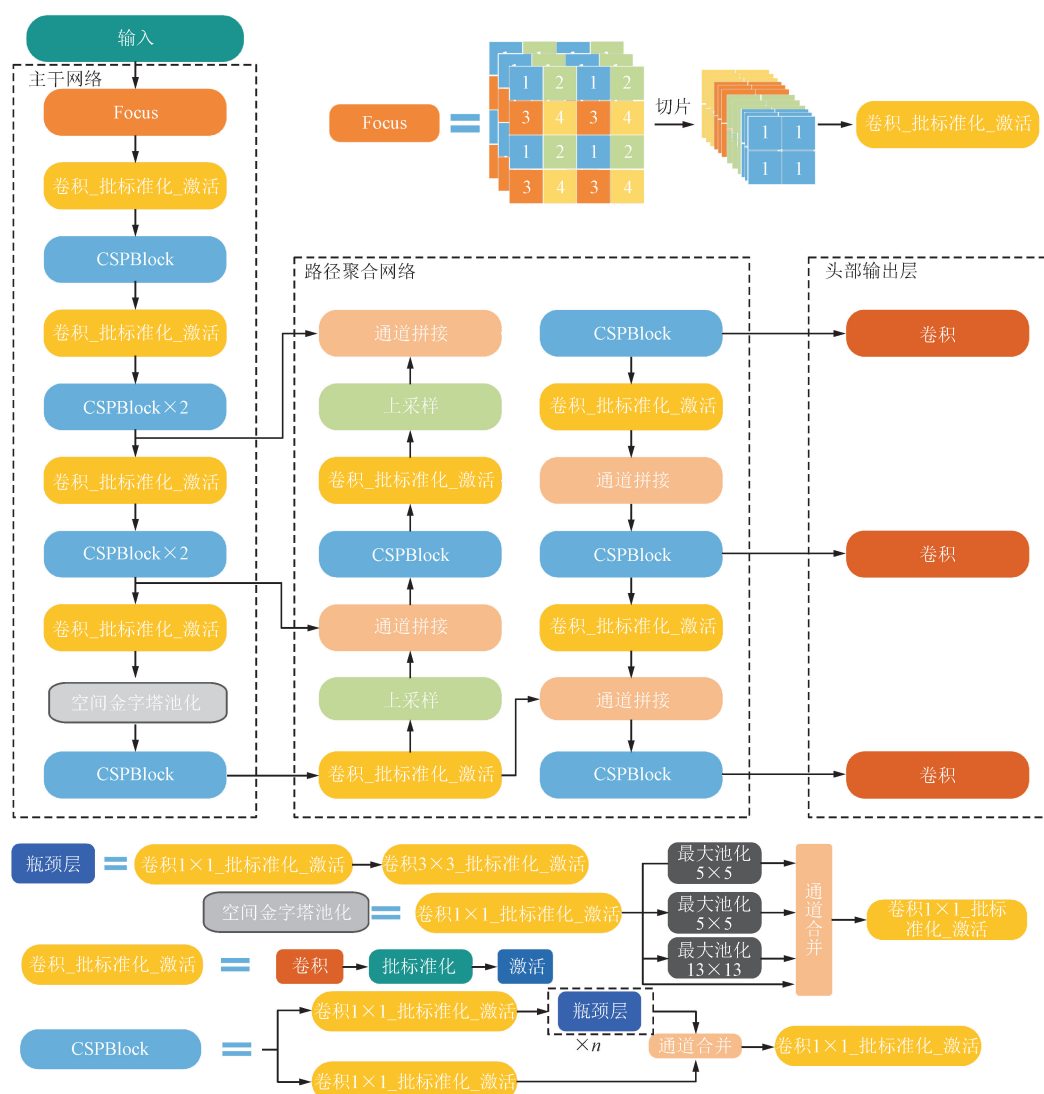


图 1 YOLOv5s 网络结构图

Fig. 1 YOLOv5s Network structure diagram

YOLOv5 的主干网络借鉴跨阶段局部网络(cross stage partial network^[22], CSPNet),沿用 YOLOv4 的 CSPDarkNet,在 YOLOv3 的 Darknet 的基础上,借鉴了跨阶段局部网络在 ResNet 中的应用。将 DarkNet 中的残差块在特征图的通道维度拆分成两部分,一部分正常经过卷积、瓶颈层(Bottleneck),与 ResBlock 的结构保持一致,另一部分经过少量卷积运算后与另一支路的输出在特征图的通道维度拼接。该种方式在减少计算量的同时还减轻了梯度消失等问题。

YOLOv5 的特征融合网络使用了空间金字塔池化^[23]与路径聚合网络,其输入先经过空间金字塔池化后进入路径聚合网络,空间金字塔池化将不同大小池化核生成的特征图在通道维度拼接在一起,与普通单次池化相比,实现了不同尺度的信息融合。路径聚合网络结构如图 1 所示,YOLOv5s 的路径聚合网络结构沿用 YOLOv4 的结构,在原 YOLOv3 的特征金字塔网络(feature pyramid networks, FPN)只有自上而下的特征金字塔的基础上引入了

自底向上的路径,使得底层信息传到顶部,实现不同尺度的特征融合,提升了网络预测的准确性。

头部输出层实现对最终预测信息的输出,3 个不同尺度的特征图分别有一个头部输出层。头部输出层输入对应特征层的 $S \times S$ 的网格(S 为网格边长),每个网格输出共 $3(4+1+n)$ 个信息,其中 3 为每个网格输出的可能目标数,4 为位置信息(t_x, t_y, t_w, t_h),1 为置信度(Confidence), n 为边界框中的物体分别为每个预设类别的概率。

在模型预测中,将每个边界框预测出的 N 个类别的置信度进行高低排序并设定阈值,用非极大值抑制算法剔除多余边框,从而得到预测目标的类别与位置信息。

1.2 基于 YOLOv5s 的轻量化朱鹮检测算法

本文针对野外朱鹮检测,在结合 YOLOv5s 与 MBConvBlock 的基础上提出一种轻量化的 YOLOv5s 网络模型,其网络结构如图 2 所示,主要工作有以下 3 个方面。

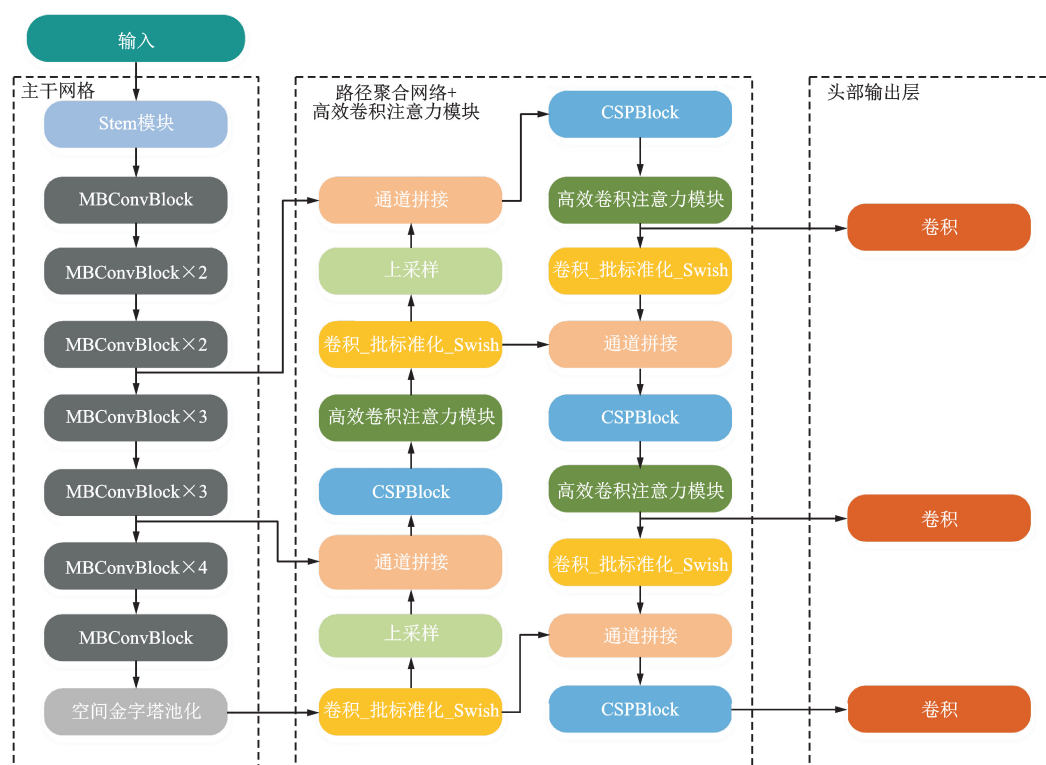


图 2 基于 YOLOv5s 的轻量化朱鹮检测算法网络结构图

Fig. 2 Network structure diagram of lightweight Crested Ibis detection algorithm based on yolov5s

1.2.1 轻量化主干网络

采用含有深度可分离卷积的 MBConvBlock 进行主干网络重构,MBConvBlock 源自于高性能轻量

化分类网络 EfficientNet,替代原主干网络所采用的 CSPBlock 来大幅降低网络模型中的计算量。MBConvBlock 使用深度可分离卷积代替部分普通卷

积,其在深度可分离卷积模块中加入了 SE 注意力模块,在第 2 个 1×1 卷积后又加入了随机失活 (Dropout) 层,减少训练过程中产生的过拟合现象。图 3 所示为 MBConvBlock 结构图,当输入、输出特征图长宽相等时存在残差边。

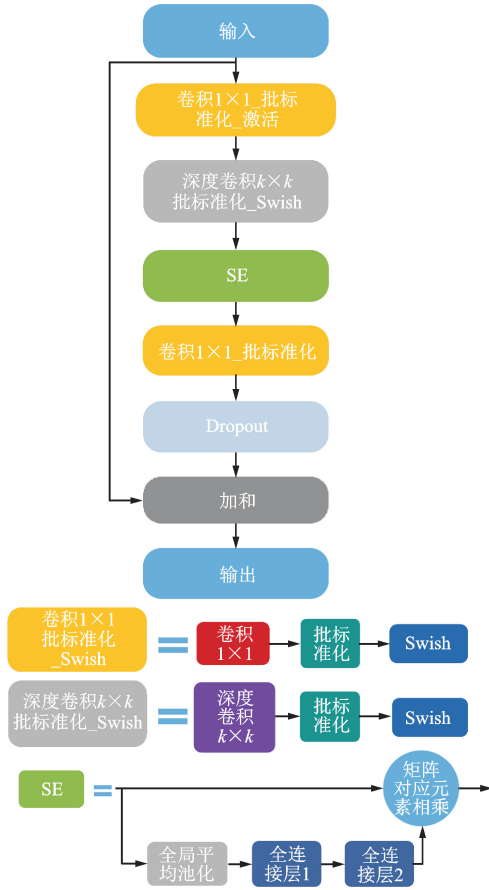


图 3 EfficientNet 中的 MBConvBlock

Fig. 3 MBConvBlock in EfficientNet

普通卷积与深度卷积如图 4 所示。每个普通卷积的通道数与输入特征图的通道数相同,每个卷积核的卷积计算在输入特征图的所有通道上进行,输出特征图数等于卷积核数,即 $N_1 = C_1$ 。 D_K 代表卷积核大小, D_F 代表输入特征图大小, M_1 代表输入特征图通道数,当 N_1 个卷积核对输入特征图做普通卷积计算时计算量(记为 F_1)由下式表示

$$F_1 = D_K D_K M_1 N_1 D_F D_F \quad (1)$$

深度可分离卷积包括深度卷积与点卷积,深度卷积的每个卷积核仅对输入特征图的一个通道进行卷积计算,卷积核数与输入特征图通道数相等,输出特征图数等于卷积核数,即 $M_1 = N_1 = C_1$,点卷积即卷积核大小为 1 的普通卷积。

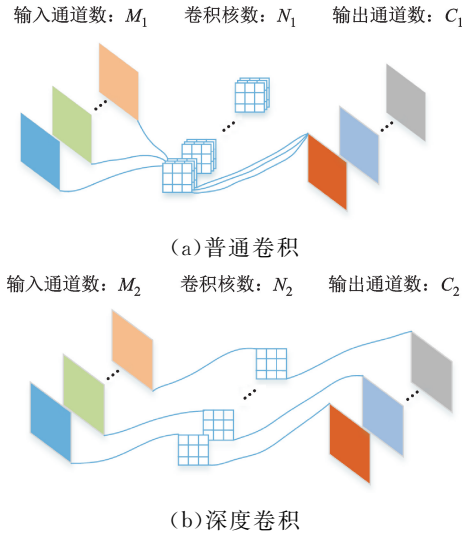


图 4 普通卷积与深度卷积

Fig. 4 Depth convolution and ordinary convolution

假定输入、输出特征图通道数相同时,即 $M_1 = M_2, N_1 = N_2$,使用深度可分离卷积代替普通卷积后的计算量(记为 F_2)由下式表示

$$F_2 = D_K D_K M_2 D_F D_F + M_2 N_2 D_F D_F \quad (2)$$

两者相比为

$$\frac{F_2}{F_1} = \frac{D_K D_K M_2 D_F D_F + M_2 N_2 D_F D_F}{D_K D_K M_1 N_1 D_F D_F} = \frac{1}{N_1} + \frac{1}{D_K^2} \quad (3)$$

经过将普通卷积替换为深度可分离卷积后,卷积核大小为 3 时,网络模型的计算量可缩小到原计算量的 $1/8 \sim 1/9$,大大缩减了网络模型的计算量。

本文使用的 MBConvBlock 参数设置见表 1,其中倍率表示模块中第一个 1×1 卷积的卷积核数是输入特征图通道数的倍数;SE 表示模块第一次堆叠时是否采用 SE,后续均不采用;步距表示模块第一次堆叠时其中深度卷积的步距。其中第一个 MBConvBlock 中不含模块内部的第一个 1×1 卷积;SE 中激活函数设置为 Relu 与 Sigmoid,其余均采用 Swish。

1.2.2 高效卷积注意力模块(ECBAM)

在特征融合网络路径聚合网络中,卷积层的通道数本身已经多达上百层,再经过多次拼接后可达上千层,特征间地位相同,会导致重要特征不明显,难以充分彰显。为了兼顾网络模型的轻量化与模型精度问题,本文在原路径聚合网络中的 CSPBlock 后加入本文设计的高效卷积注意力模块,学习不同通道的重要程度、与目标的空间位置信息和类别信息。图 5 所示为本文所使用的高效卷积注意力模块。

表 1 MBConvBlock 参数设置
Table 1 MBConvBlock parameter settings

模块	卷积核大小	通道数	倍率	堆叠次数	SE	步距	失活率
MBConvBlock	3×3	16	1	2	不采用	1	0,0.028
MBConvBlock	3×3	24	6	1	不采用	2	0.057
MBConvBlock	5×5	32	6	2	不采用	2	0.085,0.114
MBConvBlock	3×3	64	6	3	采用	2	0.142,0.171,0.200
MBConvBlock	5×5	88	6	3	采用	1	0.228,0.257,0.285
MBConvBlock	5×5	144	6	4	采用	2	0.314,0.342,0.371,0.400
MBConvBlock	5×5	240	6	1	采用	1	0.428

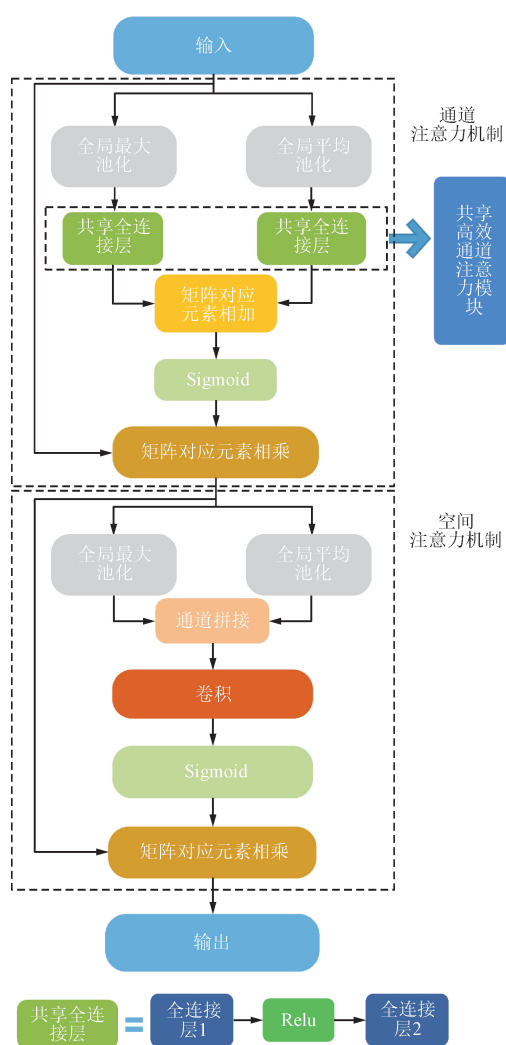


图 5 高效卷积注意力模块 (ECBAM)

Fig. 5 ECBAM Attention module

卷积注意力模块可分为两个部分,通道注意力模块(channel attention module,CAM)和空间注意

力模块(spatial attention module,SAM),分别在特征图的通道维度与空间维度实现注意力机制。

通道注意力模块首先将输入的特征图分别经过全局最大池化和全局平均池化,得到两个 $1 \times 1 \times C$ (C 为输入特征图通道数)的特征图,再将它们送入共享的含有两层神经网络的全连接层,其激活函数分别设置为 Relu 与 Sigmoid,而后将共享的全连接层输出的两个特征向量进行对应元素相加与 Sigmoid 激活,生成的特征向量即为原输入特征图通道维度的注意力权重,将其与输入特征图进行对应元素相乘操作重新生成特征图。两个全连接层的设计是为了提取跨通道交互的非线性映射,增强其特征提取能力,但其中也包括了通道数降维来降低模型的计算量,文献[24]研究结果表明,通道注意力中的第一个全连接层对特征通道数的降维操作不利于通道注意力的权重预测,特征的非线性映射范围拓展在所有通道之间是低效的,也是不必要的。同时,考虑到原通道注意力模块中绝大部分计算量在其两层全连接层中,因此本文使用高效通道注意力模块来替换原通道注意力模块,该模块避免了降维操作,使用 1 维卷积来实现临近通道间的特征提取,有效捕获了跨通道交互的信息,并且参数量大幅降低,图 6 为嵌入卷积注意力模块的高效通道注意力模块。

空间注意力模块将通道注意力模块的输出特征图作为输入特征图,首先将输入特征图做全局最大池化和全局平均池化,得到两个 $H \times W \times 1$ 的特征图,将其在通道维度拼接后用卷积进行特征提取,再经过 Sigmoid 函数生成空间注意力权重,与该模块的输入特征图进行对应元素相乘操作重新生成特征图。

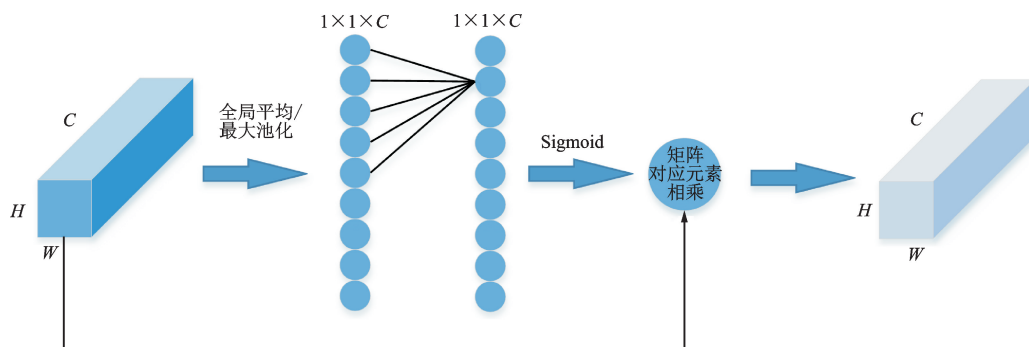


图 6 轻量化通道注意力模块(ECA)

Fig. 6 Lightweight channel attention module (ECA)

1.2.3 Stem 模块

原始 EfficientNet 网络的第一层由普通卷积层、批标准化层和非线性激活层构成。本文为加强浅层特征提取,将原网络第一层替换为 Stem 模块,其起源于 Inception v4 网络^[25],结合 MBConvBlock 来重构 YOLOv5s 主干网络。如图 7 所示,Stem 模块不采用深度可分离卷积以增强浅层特征的提取能力,并且通过两个分支网络进行特征提取,在不引入过多模型参数与计算量的情况下提升网络的特征提取能力,并通过实验验证了其有效性。图 7 中 Stem 模块中卷积核的个数从上至下分别设置为 16、8、16、16。

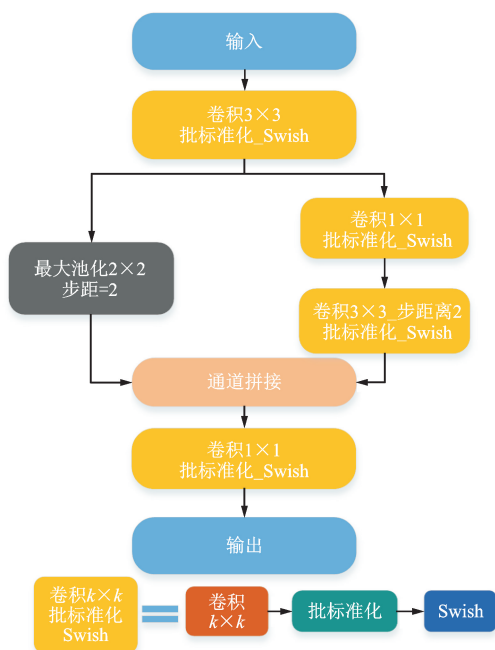


图 7 Inception v4 中的 Stem 模块

Fig. 7 Stem module in Inception v4

2 实验结果与分析

2.1 实验数据准备

本文研究以野外朱鹮为研究对象,因其为濒危野生动物,并且活动范围不固定,数据较少且不易获取。一个好的数据样本与其生存环境、生活习性等诸多因素有关,因此本文在实地拍摄的基础上还通过网络渠道获取了部分数据。本实验共采集原始图片 2 000 张,共包括目标对象 3 215 个。采用 8:2 的比例划分数据集,最终选择 1 600 张作为训练集,400 张作为测试集,采用 LabelImg 标注工具对每幅图像进行标注。

2.2 实验训练环境与训练方法

本实验训练环境采用 Ubuntu 20.04 LTS 操作系统,硬件采用 NVIDIA TITAN Xp、i7-6700、32 GB 内存,采用 Pytorch 深度学习框架。

在模型训练过程中,为减小模型陷入局部最优的可能性,使用随机梯度下降(stochastic gradient descent, SGD)优化器,将网络训练的初始学习率设置为 0.01,动量因子设为 0.937,权重衰减设置为 0.000 5,批次大小为 16,共训练 300 轮。为增强数据集的多样性与模型自身的鲁棒性,YOLOv5 对输入的图片进行 Mosaic 数据增强处理。对输入图像进行随机缩放、随机裁剪、随机排布的方式拼接,再输入网络训练。图 8 所示为采用 Mosaic 数据增强处理后的图像示例。

2.3 性能评价指标

本文采用目标检测常用性能评价指标,即测试样本的平均精度均值(mean average-precision, mAP, 记作 P_{mAP})、模型的参数数量或模型大小、网络计算量,对所提算法的有效性进行评价。模型平均精度越大,模型性能越好;参数数量与计算量越大,模型运行时代价越大。

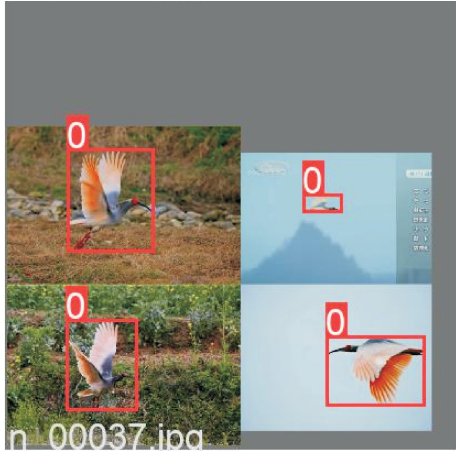


图 8 采用 Mosaic 数据增强后的图像

Fig. 8 Image augment with mosaic

精度均值(average-precision, AP)记作 S , P 为准确度(accuracy), R 为召回率(Recall)。以召回率为横轴、准确度为纵轴绘制 P - R 曲线并对其积分求出曲线下的面积 S , 表达式如下

$$P = \frac{T_P}{T_P + F_P} \quad (4)$$

$$R = \frac{T_P}{T_P + F_N} \quad (5)$$

$$S = \int_0^1 P(R) dR \quad (6)$$

式中:判断目标是否检测正确使用重叠度(intersection over union, IOU)阈值为 0.5~0.95 中等间隔取的 10 个值,分别使用每个阈值得到 10 个平均精度值,再做平均得到评价指标,记为 $P_{mAP}@0.5:0.95$,重叠度阈值取 0.5 时得到的平均精度记为 $P_{mAP}@0.5$ 。评价指标 P_{mAP} 涉及的变量见表 2。

表 2 评价指标 P_{mAP} 涉及的变量

Table 2 Variables involved in the evaluation index map

真实情况	预测结果	
	正样本	负样本
正样本	T_P	F_N
	(True Positive)	(False Negative)
负样本	F_P	T_N
	(False Positive)	(True Negative)

2.4 实验结果与对比分析

本文算法各模块参数设置除上文所提外,其余模块参数设置均沿用 Ultralytics 发布的 YOLOv5-V5.0^[11]版 YOLOv5s 设置。在图 2 中,去除水平方向箭头,以其余箭头首尾相连作为正方向,从“空间金字塔池化”至右下角“卷积_批标准化_Swish”的

输出特征图数分别设置为 240、144、144、88、88、88、144、144、240,选择 Swish 作为激活函数。高效卷积注意力模块中高效通道注意力模块卷积核大小设置为 5,空间注意力模块中的卷积核大小设置为 7。从网络模型的参数量、计算量、 $P_{mAP}@0.5:0.95$ 或 $P_{mAP}@0.5$ 方面进行对比。

2.4.1 横向对比实验

目前尚未有文献对野外复杂环境下的朱鹮进行相关研究,为验证本文算法的优势,将本文提出的算法与目前多种高性能目标检测模型在自建数据集与公共数据集上进行训练与测试。

表 3 为本文算法在自建朱鹮数据集上的测试结果,从表 3 可以看出,与 YOLOv5 系列中最小的模型 YOLOv5s 相比,本文所提出的网络模型在模型大小与检测精度之间取得了更佳平衡,本文算法相比 YOLOv5s 的参数量降低了 52.37%,计算量降低了 54.55%,但是 $P_{mAP}@0.5:0.95$ 仅降低了约 2 个百分点, $P_{mAP}@0.5$ 基本保持不变,并且具有较高水平,因为在本文研究结果中的识别类别准确度更有意义,所以朱鹮的预测边界框与实际边界框的重合程度略低,并不过多影响实际检测效果。

为了验证本文算法在目标检测公共领域的测试结果,将本文所提出的网络模型在公共数据集 PASCAL VOC 与 COCO 上进行了实验。PASCAL VOC 数据集是计算机视觉领域的重要数据集,包含 20 类常见物体,包括人、动物、交通工具等,采用其流行训练方式,即 PASCAL VOC_2007 与 PASCAL VOC_2012 的训练集与验证集共同作为训练集,共包含 16 551 张照片,测试集采用 PASCAL VOC_2007 的测试集,包含 4 952 张照片。COCO 数据集是一个大型的、具有丰富物体的数据集,其主要从复杂的日常场景中截取,本文采用其 2017 版本,包含 80 类常见物体,训练集、验证集分别有 118 287、5 000 张照片。

表 4、5 分别为本文算法与一些主流网络在公共数据集 PASCAL VOC、COCO 上的测试结果,本文网络算法参数和训练方法设置与前文保持不变。表 4、5 中缺省项表示未在公开数据中找到对应数据。从表 4、表 5 中可以看出,与当前诸多先进网络相比,本文算法的实验结果在模型大小与计算量方面具有一定的竞争力。例如在表 4 中,与针对嵌入式设备设计的更轻的网络 YOLO-Nano^[26]与轻量化网络 ThunderNet_MM^[27]相比,本文算法在检测精度方面有明显提升,并且模型参数数量与计算量相近,与

庞大的 PP-YOLO 算法^[28]相比,检测精度仍有一定差距。例如在表 5 中,本文算法相比 YOLOv4-tiny^[10]算法参数数量更少,计算量更低,但达到了更高的检测精度,相比近年旷世提出的 YOLOX-Tiny^[29]在检测精度上接近,但本文算法参数数量与计算量大幅下降,对硬件的计算能力与存储的要求进一步降低,更适合在嵌入式设备部署。

与 YOLOv5s^[11]相比,在两个公共数据集上本文算法的测试结果检测精度已经十分接近,但仍有差距,部分原因来自于将 YOLOv5s 的主干网络替换为轻量化网络,即在融合两种网络模型时,平衡其超参数选择不够恰当,导致网络模型两部分的特征提取能力未在总参数数量或计算量的限制下“协作”达到最优。

表 3 在自建朱鹮数据集上的横向对比实验

Table 3 Comparison with the other state-of-the-art object detection system on self-built Crested Ibis data set

网络模型	输入图像尺寸/(像素×像素)	模型参数数量/ 10^6	计算次数/ 10^9	$P_{mAP}@0.5;0.95$	$P_{mAP}@0.5$
YOLOv5s ^[11]	608×608	7.0	14.3	0.689	0.893
YOLOv5l ^[11]	608×608	46.5	97.3	0.746	0.918
YOLOv3-SPP ^[11]	608×608	62.6	140.5	0.744	0.919
YOLOv3-Tiny ^[11]	608×608	8.7	11.5	0.629	0.896
本文	608×608	3.3	6.5	0.666	0.916

表 4 在 PASCAL VOC 数据集上的横向对比实验

Table 4 Comparison with the other state-of-the-art lightweight object detection system on PASCAL VOC

网络模型	输入图像尺寸	模型参数数量/ 10^6	计算次数/ 10^9	$P_{mAP}@0.5$
YOLOv3-Tiny ^[9]	416×416	8.7	5.6	0.571
YOLO-Nano ^[26]	416×416		4.6	0.691
ThunderNet_MM ^[27]	416×416		0.8	0.738
YOLOv5s ^[11]	416×416	7.0	6.7	0.821
PP-YOLO ^[28]	416×416	44.9		0.843
本文	416×416	3.3	3.1	0.792

表 5 在 COCO 数据集上的横向对比实验

Table 5 Comparison with the other state-of-the-art lightweight object detection system on COCO

网络模型	输入图像尺寸	模型参数数量/ 10^6	计算次数/ 10^9	$P_{mAP}@0.5;0.95$	$P_{mAP}@0.5$
YOLOv4-Tiny ^[10]	416×416	6.1	6.9	0.217	0.402
PP-YOLO-Tiny ^[28]	416×416	4.2		0.227	
CSL-YOLO ^[30]	416×416	3.2	1.5	0.245	0.44
YOLOX-Nano ^[29]	416×416	0.9	1.1	0.253	
YOLOX-Tiny ^[29]	416×416	5.1	6.5	0.317	
YOLOv5s ^[11]	416×416	7.0	7.0	0.322	0.516
PP-YOLO ^[28]	416×416	44.9		0.425	0.628
本文	416×416	3.3	3.3	0.298	0.508

2.4.2 纵向对比实验

本文在自建数据集上进行了纵向对比实验,结果见表 6。首先使 MBConvBlock 进行 YOLOv5s 的主干网络重构并进行实验,再添加 Stem 模块进行

模型训练与测试,最后添加卷积注意力模块或高效通道注意力模块单独进行实验。由表 6 可以看出,使用了 MBConvBlock 进行模型主干网络重构和改变网络的超参数选择后,大量减少了网络模型参数数量

与计算量。网络模型参数 P_{mAP} 降低了 52.41%, 计算量降低了 59.44%, 但 $P_{mAP}@0.5:0.95$ 也降低了约 5.5 个百分点, 而在浅层网络中使用 Stem 模块可以在增加极少参数与约 10% 计算量的情况下达到大约 2.2 个百分点的性能提升。

在加入了 Stem 模块的基础上, 本文还进行了在 YOLOv5s 特征融合网络路径聚合网络中仅使用通道注意力模块 (CAM) 或空间注意力模块 (SAM) 的实验, 在仅使用通道注意力模块时能够达到约 0.6 个百分点的性能提升, 而在仅使用空间注意力模块时仅能够达到约 0.3 个百分点的性能提升。同时使用通道与空间注意力模块时, 相比于 YOLOv5s, 模型参数约增加了 0.38%, 但是 $P_{mAP}@0.5:0.95$ 增加了约 1.3%。同时, 将原通道注意力模块中的两层全连接层替换为高效通道注意力模块, 实验证明其与含有两个全连接层的原通道注意力模块在自建朱鹮数据集上能达到相近的 P_{mAP} , 证明了其有效捕获了跨通道交互的信息, 并且捕获所有通道之间的非线性特征映射是低效的。在实验中发现当卷积注意力模块中的通道注意力模块第一层全连接层神经元数取输入特征层的 1/8、1/16 时, 文中取 1/8, P_{mAP} 相比不加卷积

注意力模块时会有提升, 但当其个数取输入特征层的 1/4、1/2 时, 多次训练及测试结果中的 P_{mAP} 有时会下降, 猜测可能是网络参数数量变大, 但数据集有限, 变得更加难以训练, 网络难以收敛, 而当其为 1/8、1/16 时, 网络收敛较好。当将两层全连接层替换为高效通道注意力模块时, 高效通道注意力模块参数十分少, 所以推测其训练较为充分, 但相反其特征提取与特征表达能力也远不及两层全连接层, 故其中存在一个临界值, 即复杂模型在某种程度的不充分训练下的测试结果等于简单模型在较充分训练下的测试结果, 所以卷积注意力模块存在某种训练情况使得其测试结果超过本文所提出的高效卷积注意力模块的测试结果。本文推测卷积注意力模块适合大样本大规模数据训练, 不适合小样本训练。

实验结果表明, 将 EfficientNet 的第一个普通卷积层替换为 Stem 模块, 同时使用轻量化注意力模块与空间注意力的情况下, 模型 $P_{mAP}@0.5:0.95$ 较高, 达到 0.666, 略微增加了模型参数数量和约 10% 的计算量, 但 $P_{mAP}@0.5:0.95$ 显著提升了约 3.2 个百分点。

表 6 在自建朱鹮数据集上的纵向对比实验结果

Table 6 Ablation Experiment of self-built Crested Ibis data set

网络模型	输入图像尺寸 /(像素×像素)	模型参数数量	计算次数 /10 ⁹	P_{mAP} @0.5:0.95	P_{mAP} @0.5
YOLOv5s	608×608	7 022 326	14.3	0.689	0.893
MBConvBlock+YOLOv5s	608×608	3 341 858	5.8	0.634	0.882
Stem+MBConvBlock+YOLOv5s	608×608	3 343 730	6.5	0.656	0.906
CBAM+MBConvBlock+YOLOv5s	608×608	3 354 456	5.9	0.648	0.884
CAM+Stem+MBConvBlock+YOLOv5s	608×608	3 356 034	6.5	0.662	0.914
SAM+Stem+MBConvBlock+YOLOv5s	608×608	3 344 024	6.5	0.659	0.908
CBAM+Stem+MBConvBlock+YOLOv5s	608×608	3 356 328	6.5	0.669	0.915
ECA+Stem+MBConvBlock+YOLOv5s	608×608	3 343 745	6.5	0.659	0.904
ECBAM+Stem+MBConvBlock+YOLOv5s	608×608	3 344 039	6.5	0.666	0.916

3 结 论

经过以上数据集准备、算法搭建与实验, 得出以下结论。

(1) 本文提出了一种基于 YOLOv5s 的轻量化朱鹮检测算法, 其对 YOLOv5s 目标检测网络进行了网络压缩与压缩后的精度提升。使用 EfficientNet 中的 MBConvBlock 进行了模型主干网络重构,

实现了网络的轻量化设计, 大幅降低了网络的参数数量与计算量, 在浅层网络中添加的 Stem 模块与在路径聚合网络中添加的高效卷积注意力模块可以在略微提升计算量的情况下有效提升模型的检测精度。

(2) 在自建数据集与公共数据集上验证了所提算法可以在参数量大幅降低的同时保持较高的检测精度, 本文算法相比 YOLOv5s 的参数数量降低了

52.37%, 计算量降低了 54.55%, 但是在自建数据集上 $P_{mAP}@0.5:0.95$ 仅降低了约 2 个百分点, 达到 0.666, 在公共数据集 PASCAL VOC 上 $P_{mAP}@0.5$ 达到 0.792, 在公共数据集 COCO 上 $P_{mAP}@0.5:0.95$ 达到 0.298, 达到与 YOLOv5s 相近的检测结果。

进一步的研究将探讨在网络中使用轻量化模块带来损失的原因, 使得模型在进一步轻量化的同时保持网络自身的检测精度, 以及融合不同网络模型时, 在不同网络模型的超参数之间使用更恰当的权重。

参考文献:

- [1] 闫鲁, 黄治学. 35 年来朱鹮的保护经验与发展思路 [J]. 野生动物学报, 2021, 42(3): 890-896.
YAN Lu, HUANG Zhixue. Thirty-five years conservation experience of crested ibis (*Nipponia nippon*) and future perspectives [J]. Chinese Journal of Wildlife, 2021, 42(3): 890-896.
- [2] 陈保家, 陈学力, 沈保明, 等. CNN-LSTM 深度神经网络在滚动轴承故障诊断中的应用 [J]. 西安交通大学学报, 2021, 55(6): 28-36.
CHEN Baojia, CHEN Xueli, SHEN Baoming, et al. An application of convolution neural network and long short-term memory in rolling bearing fault diagnosis [J]. Journal of Xi'an Jiaotong University, 2021, 55(6): 28-36.
- [3] 闫彬, 樊攀, 王美茸, 等. 基于改进 YOLOv5m 的采摘机器人苹果采摘方式实时识别 [J/OL]. 农业机械学报 [2022-07-21]. <http://kns.cnki.net/kcms/detail/11.1964.S.20220719.1604.008.html>.
YAN Bin, FAN Pan, WANG Meirong, et al. Real time recognition of apple picking mode of picking robot based on improved YOLOv5m [J/OL]. Transactions of the Chinese Society for Agricultural Machinery [2022-07-21]. <http://kns.cnki.net/kcms/detail/11.1964.S.20220719.1604.008.html>.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA IEEE, 2014: 580-587.
- [5] GIRSHICK R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2015: 1440-1448.
- [6] REN Shaoqing, HE Kaiming, 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.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 779-788.
- [8] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 6517-6525.
- [9] REDMON J, FARHADI A. YOLOv3: an incremental improvement [J/OL]. arXiv, 2018 [2020-03-16]. DOI: 10.48550/arXiv.1804.02767.
- [10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [J/OL]. arXiv, 2020 [2020-03-16]. DOI: 10.48550/arXiv.2004.10934.
- [11] GitHub, Inc. YOLOv5 [EB/OL]. [2021-06-15]. <https://github.com/ultralytics/yolov5>.
- [12] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [C]//Computer Vision - ECCV 2016. Cham, Switzerland: Springer International Publishing, 2016: 21-37.
- [13] HAN S, MAO H, DALLY W J. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding [J]. Fiber, 2015, 56(4): 3-7.
- [14] JIN Xiaojie, YUAN Xiaotong, FENG Jiashi, et al. Training skinny deep neural networks with iterative hard thresholding methods [J/OL]. arXiv, 2016 [2021-11-13]. DOI: 10.48550/arXiv.1607.05423.
- [15] WEN Wei, WU Chunpeng, WANG Yandan, et al. Learning structured sparsity in deep neural networks [C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2016: 2082-2090.
- [16] HOWARD A G, ZHU Menglong, CHEN Bo, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [J/OL]. arXiv, 2017 [2021-11-16]. DOI: 10.48550/arXiv.1704.04861.
- [17] SANDLER M, HOWARD A, ZHU Menglong, et al. MobileNetV2: inverted residuals and linear bottlenecks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 4510-4520.
- [18] HOWARD A, SANDLER M, CHEN Bo, et al. Searching for MobileNetV3 [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV).

- Piscataway, NJ, USA: IEEE, 2019: 1314-1324.
- [19] ZHANG Xiangyu, ZHOU Xinyu, LIN Mengxiao, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 6848-6856.
- [20] TAN Mingxing, LE Q. EfficientNet: rethinking model scaling for convolutional neural networks [C]//Proceedings of the 36th International Conference on Machine Learning, Chia Laguna Resort, Sardinia, Italy: PMLR, 2019: 6105-6114.
- [21] LIU Shu, QI Lu, QIN Haifang, et al. Path aggregation network for instance segmentation [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 8759-8768.
- [22] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2020: 1571-1580.
- [23] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [24] WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 11531-11539.
- [25] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, Inception-ResNet and the impact of residual connections on learning [J/OL]. arXiv, 2016 [2021-10-14]. DOI: 10.48550/arXiv.1602.07261.
- [26] WONG A, FAMUORI M, SHAFIEE M J, et al. YOLO nano: a highly compact you only look once convolutional neural network for object detection [C]//2019 15th Workshop on Energy Efficient Machine Learning and Cognitive Computing-NeurIPS Edition (EMC2-NIPS). Piscataway, NJ, USA: IEEE, 2019: 22-25.
- [27] QIN Zheng, LI Zeming, ZHANG Zhaoning, et al. ThunderNet: towards real-time generic object detection on mobile devices [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2019: 6717-6726.
- [28] LONG Xiang, DENG Kaipeng, WANG Guanzhong, et al. PP-YOLO: an effective and efficient implementation of object detector [J/OL]. arXiv, 2020 [2021-12-16]. DOI: 10.48550/arXiv.2007.12099.
- [29] GE Zheng, LIU Songtao, WANG Feng, et al. YOLOX: exceeding YOLO series in 2021 [J/OL]. arXiv, 2021 [2022-03-04]. DOI: 10.48550/arXiv.2107.08430.
- [30] ZHANG Yuming, LEE C C, HSIEH J W, et al. CSL-YOLO: a new lightweight object detection system for edge computing [J/OL]. arXiv, 2021 [2022-01-06]. DOI: 10.48550/arXiv.2107.04829.

(编辑 武红江 杜秀杰)