

基于 MYOLOv8 的目标检测方法

张正勃¹, 曹爱岷², 王兴盛³

(1. 西安交通工程学院 机械与电气工程学院, 西安 710300;

2. 北京大学深圳研究生院 新材料学院, 广东 深圳 518055;

3. 兰州交通大学 自动化与电气工程学院, 兰州 730070)

摘要: 针对当前的目标检测方法难以表征多尺度上下文特征的问题, 提出了 MYOLOv8 算法; 为了提高 YOLOv8 模型对于小、中、大型目标的检测能力, 提出了一种分层多尺度提取模块对空间特征进行分层特征聚合来捕获多尺度空间上下文信息; 为了进一步提高模型对于空间语义的提取能力, 提出了一种自适应的通道注意力机制, 该机制通过自适应地学习相邻通道之间的相互依赖关系来促进模型关注有用特征, 抑制无用特征; 为了提高模型对于边界困难样本的定位能力, 提出了一种 Slide Loss 来处理目标检测中的样本不平衡问题, 该方法采用对困难样本进行强加权的方式来促使模型着重优化难分样本; 在 MS COCO 数据集上的实验结果表明, 所提出的算法相比于 YOLOv8-n 和 YOLOv8-s, *mAP* 分别提升了 3.4% 和 1.4%, 同时具有相似的参数量和计算开销, 以及更快的推理速度。

关键词: 目标检测; 多尺度上下文; 分层特征聚合; 注意力机制; 样本不平衡

Object Detection Method Based on MYOLOv8

ZHANG Zhengbo¹, CAO Aimin², WANG Xingsheng³

(1. School of Mechanical and Electrical Engineering, Xi'an Traffic Engineering Institute, Xi'an 710300, China;

2. School of New Materials, Peking University Shenzhen Graduate School, Shenzhen 518055, China;

3. School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: Current object detection methods are difficult to characterize multi-scale context features, to address this problem, a MYOLOv8 algorithm is proposed. In order to improve the detection capability of YOLOv8 model for small, medium and large objects, a hierarchical multi-scale extraction module is proposed to perform hierarchical feature aggregation of spatial features to capture multi-scale spatial context information. In order to further improve the model's ability to extract spatial semantics, a self-adaptive channel attention mechanism is proposed, which promotes the model to focus on useful features and suppress useless features by adaptively learning the interdependencies between adjacent channels. In order to improve the model to locate difficult samples, a Slide Loss is proposed to deal with the sample imbalance problem in object detection, which employs a strong weighting for difficult samples to encourage the model to focus on optimizing difficult samples. Experimental results on the MS COCO dataset show that the proposed algorithm improves the *mAP* by 3.4% and 1.4% compared to YOLOv8-n and YOLOv8-s, respectively, while having the number of similar parameters and computational cost, as well as faster reasoning speed.

Keywords: object detection; multi-scale context; hierarchical feature aggregation; attention mechanism; sample imbalance

0 引言

目标检测是检测特定场景下物体的空间位置, 并且识别该物体所属的类别, 是环境感知的重要内容^[1], 其在智能监控、智能驾驶、智能机器人等领域均有着广泛的应用场景^[2-4]。尽管多阶段的目标检测算法取得了出色的性能, 但其复杂的网络结构往往阻碍了它们在现实生活中的应用^[5]。为了实现精度和速度的良好权衡, 学

术界在实时目标检测方面的研究付出了诸多努力^[6]。根据检测算法的原理不同, 主流的目标检测算法主要包括: Two-stage、one-stage 两类。其中, Two-stage 算法^[7-9]一般先采用一个网络生成候选区域, 然后使用 ROI-pooling (Region of Interest pooling) 层来调整 ROI 尺寸, 并将候选目标对应于特征图的相应位置, 最后由全连接层来整合其特征信息以得到目标分类和位置。虽

收稿日期: 2024-07-25; 修回日期: 2024-09-11。

基金项目: 西安交通工程学院中青年基金项目(2022KY-33); 陕西省教育厅科学研究计划项目(23JK0529)。

作者简介: 张正勃(1996-), 女, 讲师。

引用格式: 张正勃, 曹爱岷, 王兴盛. 基于 MYOLOv8 的目标检测方法[J]. 计算机测量与控制, 2025, 33(1): 93-98, 113.

然这种方式的检测精度较高,但其检测速度较慢。因此,此类方法不适用一些对实时性要求较高的场景。相反,One-stage 算法^[10-12]的工作流程较为简单,其不需要生成候选区域,只需要将图片送入网络便能够直接预测出所有包含类别概率和定位坐标的边界框。此算法能够完整地单次训练其共享特征,从而检测速度得到提升。其中,YOLO 系列是最典型的代表,例如:YOLOv5^[13]、YOLOv6^[15]、YOLOv7^[16]、YOLOv8^[17]等。

YOLOv5 是一种典型的一阶段基于 anchor 的目标检测算法,由 Glenn Jocher 等人在 2022 年首次提出^[13]。YOLOv5 的 backbone 沿用了之前 YOLO 系列采用的 CSP-DarkNet 结构,利用横向连接来缓解梯度消失问题,neck 则采用了 PANFPN 的结构^[14],利用自顶向下以及自底向上的结构来完成顶层和浅层的特征融合,head 采用了简单的 1×1 卷积来实现框的回归和类别的分类。YOLOv6^[15]采用了结构重参数化的方式来改造 backbone,通过多分支训练和单分支推理实现了精度和速度的权衡,head 部分则利用了解耦头的结构,通过引入两个额外的 3×3 卷积来提升性能。YOLOv7^[16]提出了 ELAN 的网络结构,通过将瓶颈块中各个层级的特征进行拼接,进一步缓解了梯度消失,并且提高了特征的利用效率。在正负样本分配这一环节,YOLOv7 采用了 YOLOv5 锚框匹配和动态样本分类结合的方式,有助于模型分配到高质量的样本,这进一步提升了检测的精度。YOLOv8^[17]摒弃了 anchor-base 思想,使用 anchor-free 的方式,通过回归中心点到四条边的距离,减小了边界框回归的难度。此外,YOLOv8 提出正常对齐的 anchor 应该预测高的分类得分,并且具有精确的定位,不对齐的 anchor 应当具有低的得分,并且在非极大值抑制环节被抑制,因此,YOLOv8 采用了 TOOD^[18]的正负样本分配方式,根据分类和回归分数的高阶加权来动态地引导模型关注高质量的 anchor,从而极大地提高了检测精度。为了保证算法在工业场景中的实时性应用,一般 YOLOv8 常用的模型尺寸为 YOLOv8-n (nano, 纳米版)和 YOLOv8-s (small, 小型版)。

尽管之前的 YOLO 算法取得了不错的检测性能,但是识别特定场景中不同尺度的目标仍然是一个具有挑战性的任务。基于此,文献 [19] 提出了一种多尺度轻量化目标检测模型,通过采用重构的多尺度融合方法对主干网络进行改进,同时增加卷积运算以及调整其输入输出通道数来提高对网络的特征融合能力,从而增强了模型对目标的检测能力。文献 [20] 提出了一种基于频率空间的目标检测算法,通过增加输出通道数近似无损变换的下采样策略,实现了完整的特征表达,同时引入固定组内通道的分组卷积,最大程度降低了模型的参数量。文献 [21] 提出了一种改进的 YOLOv8 目标检测算法,通过设计一种改进的 GAM-C2f 模块,以平衡模

型的计算效率和准确性,此外还设计一种 SPPFAPGC 模块,缓解了模型的局部信息丢失问题,提高了模型对全局特征的提取能力。

先前的方法只关注如何设计出一个相对复杂的神经网络架构来进行空间表征学习,以达到更好的检测效果,却往往忽略了如何高效地表征多尺度的空间上下文特征。因此,本文提出了 MYOLOv8 算法,具体工作包括:

1) 利用分层多尺度提取模块 (HMSE, hierarchy multi-scale extraction) 对输入特征进行分层多尺度聚合,来提取多尺度的空间上下文信息,显著地提高了 YOLOv8 的特征提取能力。

2) 为了进一步提取到有用信息,并且抑制无用信息,本文将自适应通道注意力模块 (SCA, self-adaption channel attention) 整合到了分层多尺度提取模块中,其通过一个自适应的一维卷积窗口对通道的相关性进行特征学习,避免了传统通道注意力机制的通道降维操作对特征学习的影响。

3) 为了缓解目标检测中的样本不平衡问题,本文提出了利用 Slide 函数对位于边界的困难样本进行强化加权,对简单样本以及负样本进行弱化加权,促使模型在训练的过程中显式地关注边界样本,并且,Slide 损失函数的设计没有引入过多的超参数,这大大降低了模型优化的难度。

最后,将改进算法在 COCO 数据集上进行实验,结果表明,相比于 YOLOv8-n 和 YOLOv8-s,本文所提出算法的类别平均精度 mAP 分别提升了 3.4% 和 1.4%,同时具有相似的参数量和计算成本,延迟也分别减少了 1.1 ms 以及 0.6 ms,表明了所提出算法的有效性。

1 MYOLOv8 算法

本节主要描述了 MYOLOv8 算法的设计细节。首先,针对不同尺寸的目标检测设计了一种分层多尺度特征提取模块 (HMSE, hierarchy multi-scale extraction)。然后,分别对通道注意力模块和损失函数的设计进行了技术改进,并且讨论了改进前后的优缺点。最后,提供了用于多种目标检测的 MYOLOv8 算法的整体结构。

1.1 分层多尺度提取模块

为了有效地检测不同尺度的物体,先前的检测方法一般是通过融合 neck 中不同层级的中间特征来提取多尺度特征。例如:YOLOv8 采用了 FPN 和 PAFPN 结构,通过自顶向下和自底向上两种方式来促进浅层和深层信息交互,但是这些主流的检测结构往往在 backbone 部分忽略了多尺度特征表示的重要性,而仅仅只专注于如何提高效率以及加快收敛速度,如 CSP^[22]、ELAN 等结构。

为了提高 YOLOv8 的多尺度建模能力,本文提出了一种分层多尺度特征提取模块 (HMSE),来聚合不

同层次的特征以增强多尺度特征表示。HMSE 和 YOLOv7 的 ELAN 模块、YOLOv8 的 C2f 模块的原理图如图 1 所示。

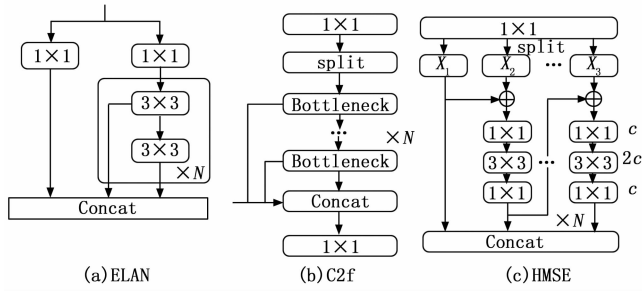


图 1 HMSE 与 ELAN、C2f 的对比图

首先, 采用一个 1×1 的卷积, 将输入特征 $X \in \mathbb{R}^{N \times C \times H \times W}$ 的通道数扩大到原来的三倍, 其中 N 表示批量的大小, C 表示通道数, H 和 W 分别表示输入特征的高和宽。然后, 将扩大通道后的特征再分成三组, 即 $X_i \in \mathbb{R}^{N \times C \times H \times W}$, $i = 1, 2, 3$, 除 X_1 外, 其他特征会通过 3 个级联的二维卷积组成的瓶颈块结构来提取特征。值得注意的是, X_2 和 X_3 在经过瓶颈块结构之前, 会与上一组的输出特征进行元素式加法操作。通过这种方式, 每组输出特征具有了本层和上层两种不同的空间感受野, 即不同大小的空间范围。最后, 将三组输出特征进行拼接, 并通过一个 1×1 的卷积来完成信息交互, 这种分层结构促进了不同层级下的特征融合, 有助于模型提取到多尺度的空间上下文信息, HMSE 的函数表达式如公式 (1):

$$Y_i = \begin{cases} X_i & i = 1 \\ \text{Bottleneck}(Y_{i-1} + X_i) & i > 1 \end{cases} \quad (1)$$

式中, Bottleneck 表示 1×1 、 3×3 和 1×1 组成的瓶颈块结构。

1.2 自适应通道注意力模块

为了捕获更加具有区分性的特征, 先前的方法一般采用通道注意力机制来促使模型关注特征“是什么”, 这些方法致力于开发复杂的注意力模块以获取更好的性能, 这不可避免地增加模型的计算开销。为了降低模型的复杂性, 先前的方法利用两层级联的多层感知机结构, 并且通过通道的降维和升维来完成跨通道特征交互, 但是最新研究表明, 特征降维会影响模型的学习通道特征的能力。

为了有效地解决上述问题, 本文提出一种自适应的通道注意力机制 (SCA, self-adaption channel attention) 模块, 如图 2 所示。其核心是利用一个内核为 K 的一维卷积来捕获 K 个领域通道之间的相互依赖关系, 从而避免了通道的降维操作。同时, 为了避免手动设置一维卷积的内核大小, 本文提出将卷积的大小设置成与通道维度成正比, 以一种自适应的方式来完成跨通道交互。

首先, SCA 对输入特征为 $X \in \mathbb{R}^{N \times C \times H \times W}$ 执行全局

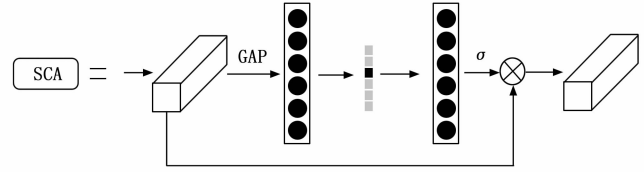


图 2 SCA 模块的原理图

空间平均池化 (GAP) 操作来获取全局的空间特征 $F \in \mathbb{R}^{N \times C \times 1 \times 1}$ 。然后, 为了减少计算开销, 同时学习通道之间的相关性, 将中间特征图 F 的尺寸重塑成 $\mathbb{R}^{N \times 1 \times C}$, 并通过一个和通道维度成正比关系, 大小为 K 的一维卷积来捕获相邻 K 个通道之间的局部上下文特征。最后, 通过 Sigmoid 激活函数生成通道注意力掩码 $M^{N \times 1 \times C}$, 并将输入特征与通道注意力掩码进行点乘来得到最终被激活的输出特征。SCA 的函数表达式以及一维核大小的计算方式分别如公式 (2) 和 (3) 所示, 其中 $|t|_{\text{odd}}$ 表示离 t 最近的奇数:

$$Y = \sigma(\text{Conv1D}_k[\text{GAP}(X)]) \quad (2)$$

$$k = \psi(C) = \left\lfloor \frac{\log_2(C)}{2} + \frac{1}{2} \right\rfloor_{\text{odd}} \quad (3)$$

本文将 SCA 整合到了 HMSE 模块的最后一个卷积层之后, 来促进模型关注更加具有区分性的通道特征。SCA 通过一个一维窗口直接对通道特征进行滑动处理, 而不是借助多层感知机 (MLP, multilayer perceptron) 对通道进行降维、升维操作, 因此, 没有引入显著的参数量和计算成本。

1.3 损失函数改进

目标检测的有效性高度依赖损失函数的设计, 损失函数度量了模型的预测值与真实值之间的差异。目标检测任务中的损失函数主要包括了分类损失和回归损失两部分。

对于分类损失, 先前的方法一般采用的是二元交叉熵损失函数、Focal loss^[23]、VFL loss 等, 但是上述方法不能很好地处理目标检测中的样本不平衡问题, 即简单样本多, 困难样本少, 简单样本的损失主导分类损失, 导致模型不能有效地更新困难样本的梯度, 从而出现错误的分类。尽管 Focal loss 一定程度上优化了样本不平衡问题, 但其超参数较多, 针对不同的数据集均需要进行特定的适配, 优化相对困难。

简单样本和困难样本的区别是基于预测框和真实框的交并比 (IOU, intersection over union), 由于分类的不确定性, 位于边界的样本往往具有较大的损失, 因此, 本文提出采用 Slide 损失函数来优化样本不平衡问题, 其核心是为困难样本分配更高的损失权重, 从而让模型去着重优化此类样本, Slide 损失函数表达式见公式 (4):

$$f(x) = \begin{cases} 1 & x \leq \mu - 0.1 \\ e^{1-x} & \mu - 1 < x < \mu \\ e^{1-x} & x \geq \mu \end{cases} \quad (4)$$

式中, μ 表示所有预测框和真实框之间的平均 IOU, 当小于平均 IOU 时为负样本, 具有较小的权重 1, 当位于边界 ($\mu - 0.1, \mu$) 时, 具有相同的权重 $e^{1-\mu}$, 当大于平均 IOU 时, 权重呈指数衰减。这表明约靠近边界时, 权重越大, 从而促使模型更好地学习困难样本。Slide Loss 的原理图如图 3 所示。

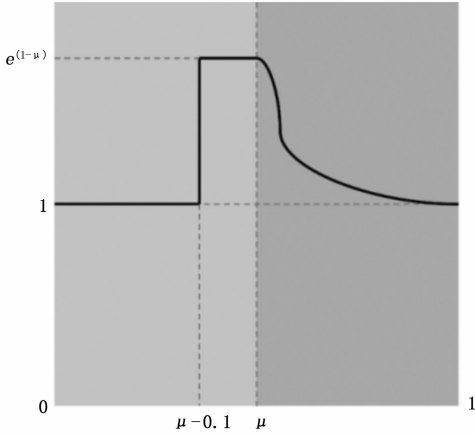


图 3 Slide Loss 的原理图

1.4 网络的整体结构

提出的 MYOLOv8 的整体网络结构如图 4 所示。MYOLOv8 的结构和 YOLOv8 类似, 不同点在于, MYOLOv8 在 backbone 和 neck 中采用了 HMSE 模块进行特征提取, 而不是 C2f 模块。同时, MYOLOv8 三层特征图的输出特征通道个数分别为 320、640 和 1 280, 而不是 YOLOv8 的 256、512 和 1 024。此外, MYOLOv8 在 backbone 的 stage1 到 stage4 中, HMSE 模块的数量都是 3, 而 C2f 分别是 3、6、6、3。最后, 和之前的 YOLO 算法一样, 本文对 HMSE 模块的数量和通道数分别乘以深度和宽度因子, 从而构建出不同大小的网络。

2 实验过程与分析

本节首先描述了基准数据集、实验细节和评价指标。然后, 在 COCO 数据集上进行了广泛的消融实验,

以验证所提出的 MYOLOv8 算法的有效性以及寻求相关参数的最佳设置。最后, 将本文所提出的 MYOLOv8 算法与 YOLOv5、YOLOv7、YOLOv8 等经典算法以及 DCT-YOLOv8 等改进算法进行了对比。

2.1 实验细节

本文在 pytorch 框架上实现了所提出的改进 YOLOv8 模型。所有的实验均在两块 V100 GPU 上进行, 单卡的 batch size 为 32。与 YOLOv8 训练方式不同之处有: 学习率更新方面, 将网络的初始学习率设为 0.005, 并且采用了线性 warm-up 和余弦退火的学习率更新方式; 数据增强方面, 增加了 5% 概率的 mixup 数据增强来进一步提升模型的泛化性能; 训练策略方面, 在最后 10 个 epoch 不关闭 mosaic 数据增强, 并且增加类别平衡策略, 以此来改善稀少样本的检出效果。为了进行公平的比较, 所有的模型均在目标数据集上从头训练 500 个 epoch, 没有使用额外的预训练模型。

2.2 数据集和评价指标

为了有效地评估所提出算法的有效性, 本文在广泛使用的 MS COCO 数据集^[24]上进行实验。COCO 数据集包含 115 k 个训练样本和 5 k 个测试样本, 总共 80 种类别。评价指标则采用了标准的 COCO 度量方式, 即将所有类别的平均精度 mAP 作为主要的度量方式, IOU 阈值为 0.5 和 0.75 的 mAP 以及小目标、中目标、大目标物体的平均精度 AP 作为辅助的度量方式。在计算成本方面, 本文用参数量基于 640×640 的输入尺寸计算的浮点运算总量 (FLOPs) 来衡量模型的复杂程度, 此外, 使用全精度浮点 (FP32) 的模型在 Nvidia3090 GPU 上测量的 FPS 来衡量模型的推理速度。

2.3 消融实验

2.3.1 通道扩展因子

首先, 本文探讨了 HMSE 模块中瓶颈块的通道扩展因子的大小对模型性能的影响, 实验结果见表 1 所示。实验结果表明, 通道扩张率显著影响了检测模型的性能, 当 $\gamma = 2$ 时, 模型实现了最佳的精度和性能的权

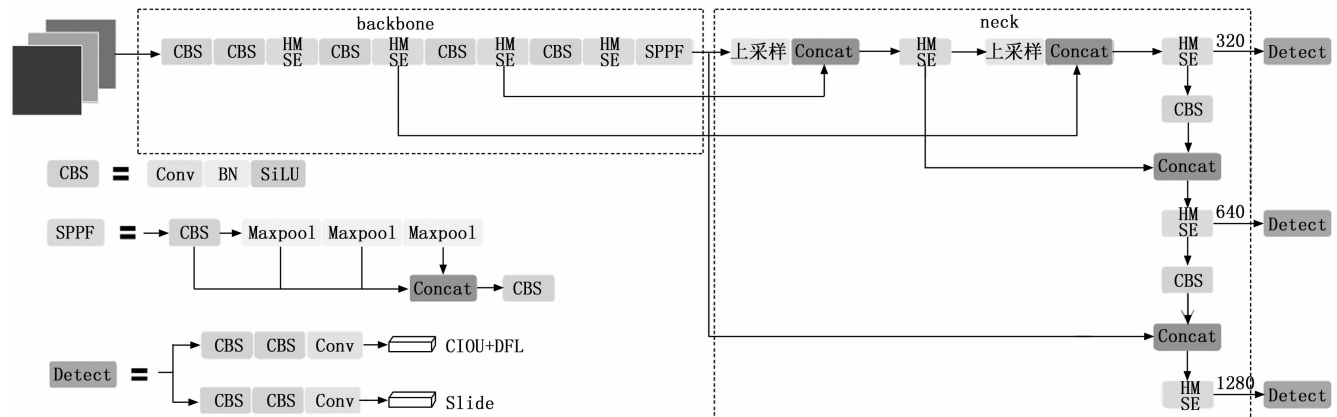


图 4 MYOLOv8 的整体网络结构图

衡, 它比 $\gamma = 1$ 时 mAP 指标高 1.4%, 这表明在瓶颈块中适当增加通道的维度, 更有利于建模多尺度的特征表示。尽管 $\gamma = 3$ 比 $\gamma = 2$ 的 mAP 指标高 0.4%, 但其计算成本高, 为了使整体网络保持较高的计算效率, 本文在后续实验中将一直采用 $\gamma = 2$ 作为默认设置。

表 1 通道扩展因子对模型性能的影响

γ	Params/ ($\times 10^6$)	FLOPs/ ($\times 10^9$)	$mAP/\%$
1	2.4	7.6	39.2
2	2.9	9.0	40.6
3	3.3	10.3	41.0

2.3.2 特征融合策略

其次, 本文还探讨了 HMSE 模块中的特征融合策略对模型性能的影响, 具体来说, HMSE 是通过来自相邻分支的特征图执行元素式的加法来完成特征融合的, 实验结果见表 2 所示。实验结果表明, 在执行特征融合后, 模型的 mAP 提升了 1.6%, 具体而言, 模型对于小目标、中目标和大目标的平均精度 (AP, average precision) 分别提升了 2.4%、1.5% 和 2.0%, 这表明 HMSE 模块对中间特征执行分层特征融合能够有效地聚合不同感受野下的特征表示, 从而显著地提升模型对于小目标和大目标的检测性能。

表 2 特征融合对模型性能的影响

融合	Params/ ($\times 10^6$)	FLOPs/ ($\times 10^9$)	$mAP/\%$	$AP_s/\%$	$AP_m/\%$	$AP_l/\%$
No	2.9	9.0	39.0	20.1	41.2	50.1
Yes	2.9	9.0	40.6	22.5	42.7	52.1

2.3.3 SCA 模块的有效性

本文通过将 SCA 模块替换成 SE^[25]、BAM^[26]、CBAM^[27] 模块, 从而探讨了 SCA 模块的有效性, 实验结果见表 3 所示。实验结果表明, SCA 模块与相比于纯通道注意力机制 SE 以及纯空间注意力机制 BAM 的结构, 性能提升最为明显, mAP 分别提升了 1.2% 和 1.0%, 同时具有更低的参数量、FLOPs, 并且 FPS 达到了最高的 140, 这充分表明了所提出的 SCA 模块能够更加有效地激活来自输出特征的有效特征, 这更有利于后续的分类和回归, 同时具备更低的计算成本和更快的推理速度。此外, SCA 模块与同时包括通道注意力和空间注意力机制的 CBAM 模块相比, 性能仍有 0.4% 的增益, 这得益于 SCA 模块中的自适应核选取策略, 能

表 3 注意力机制对模型性能的影响

注意力模块	Params/ ($\times 10^6$)	FLOPs/ ($\times 10^9$)	FPS	$mAP/\%$
SE	2.9	9.0	122	39.4
BAM	2.9	9.0	130	39.6
CBAM	3.0	9.2	120	40.2
SCA(Ours)	2.9	9.0	140	40.6

够根据输出通道的维度自适应地学习相邻通道之间的依赖关系, 同时避免了特征降维对性能造成地负面影响。

2.3.4 损失函数对比

最后, 本文替换了网络的分类损失为传统的 BCE 损失、Focal 损失, 来探讨了 Slide 分类损失的有效性, 实验结果见表 4 所示。实验结果表明, 在使用 Slide 分类损失时, 模型的 mAP 比标准的 BCE 损失要高出 1.3%, 这表明对简单样本和困难样本进行损失加权能提升模型的鲁棒性。此外, 与 Focal 损失相比, 模型的 mAP 和 AP50 仍获得了 0.5% 和 0.4% 的性能增益, 这表明 Slide 损失通过对困难样本进行强加权, 能够促使模型在训练的过程中能够显式地关注困难样本, 从而提高模型对于难分样本的分类指标, 并且 Slide 损失没有引入额外的超参数调优, 因此, 本文的方案相对更通用。

表 4 分类损失函数的类型对模型性能的影响

Loss	$mAP/\%$	$AP_{50}/\%$	$AP_{75}/\%$
BCE	39.3	55.6	40.7
Focal Loss	40.1	56.4	41.5
Slide Loss(Ours)	40.6	56.8	42.2

2.4 与其他目标检测方法的对比

本文比较了常见的目标检测算法与所提出的 MYOLOv8 算法, 实验结果见表 5 所示, 其中, tiny (微小版)、nano (纳米版) 和 small (小型版) 分别表示 3 种不同大小的模型。实验结果表明, 所提出的 MYOLOv8-n 在 MS COCO 数据集上取得了 40.6% 的 mAP , 比 YOLOv8-n 和 DCT-YOLOv8-n 分别高出 3.4% 和 2.2%, 同时只有 2.9 M 的参数量和 9.0 G 的 FLOPs, 甚至比更大的 YOLOv5-s 的 mAP 高 2.9%, 这说明了所提出的 HMSE 模块具有更好的参数利用效率, 同时, AP_s 和 AP_l 高 0.8% 和 3.5%, 这进一步说明了所提出的 HMSE 模块能够有效地聚合具有不同感受野的特征, 从而提升模型的多尺度特征表示能力。此外, 与之前的 YOLO 方法类似, 本文通过扩大网络的宽度因子和深度因子, 进一步构建了 MYOLOv8-s, 其在 MS COCO 数据集上取得了 46.3% 的 mAP , 这比之前的基线模型都要高, 同时, MYOLOv8-s 也只有 8 ms 的延迟, 在拥有更高精度的同时, 也推理速度方面也具有出色的性能表现。

本文提供了四组 YOLOv8 和 MYOLOv8 的检测结果对比图, 如图 5 所示。前两组实验结果比较了两种算法对于困难样本的检测结果, 由图可知, 所提出的 MYOLOv8 算法对于遮挡、模糊目标, 具有更高的检测置信度, 这进一步证明了所提出的 Slide Loss 能够提高模型对于困难样本的定位能力。后两组实验比较了两种

表 5 与其他算法的性能对比

算法	$mAP/\%$	$AP_{50}/\%$	$AP_{75}/\%$	$AP_s/\%$	$AP_m/\%$	$AP_l/\%$	延迟/ms	参数量/ $(\times 10^6)$	FLOPs/ $(\times 10^9)$
YOLOv5-n ^[15]	28.0	45.9	29.4	14.0	31.8	36.6	4.8	1.9	2.3
YOLOX-tiny ^[28]	32.8	50.3	34.8	14.0	35.5	48.3	4.7	5.1	7.6
YOLOv6-n ^[15]	36.2	51.6	39.1	16.8	40.2	52.6	7.5	4.3	5.5
YOLOv7-tiny ^[16]	37.5	55.8	40.2	19.9	41.1	50.8	5.2	6.2	6.9
YOLOv8-n ^[17]	37.2	52.7	40.3	18.9	40.5	52.5	8.2	3.2	8.9
DCT-YOLOv8-n ^[20]	38.4	53.9	41.7	18.3	41.5	52.2	7.8	3.0	8.8
MYOLOv8-n(Ours)	40.6	56.8	42.2	22.5	42.7	52.3	7.1	2.9	9.0
YOLOv5-s ^[13]	37.7	57.1	41.0	21.7	42.5	48.8	5.4	7.2	8.3
YOLOX-s ^[28]	40.7	59.6	44.3	23.9	45.2	53.8	5.8	9.0	13.4
YOLOv6-s ^[15]	43.7	60.8	47.0	23.6	48.7	59.8	8.9	17.2	21.9
YOLOv8-s ^[17]	44.9	61.5	48.5	24.3	50.0	61.2	8.6	11.2	28.6
DCT-YOLOv8-s ^[20]	46.2	62.5	50.2	24.8	50.2	60.8	8.3	10.6	27.9
MYOLOv8-s(Ours)	46.3	62.7	50.4	25.4	51.2	62.8	8.0	9.0	22.6



图 5 YOLOv8 与 MYOLOv8 的检测结果对比图

算法对于多尺度目标的定位能力，由实验结果可知，所提出的 MYOLOv8 算法对于小尺度的样本具有更好的检测能力，同时对于大型目标也实现了更高的检测置信度，这得益于分层多尺度提取模块的设计，能够更加有效地编码多尺度的空间上下文特征。

4 结束语

为了提高 YOLOv8 的多尺度特征提取能力，本文提出了一种 MYOLOv8 算法。首先，提出了一种分层

特征提取模块来提取输入特征中的多尺度上下文信息，从而提高模型对于小、中、大型目标的检测能力。其次，提出了一种自适应的通道注意力模块，通过一个自适应的一维卷积来学习相邻通道之间的相关性，避免了特征降维对特征相关性学习的影响，并且没有引入过多的参数量的计算成本。最后，提出了一种 Slide 损失函数对困难样本进行加权，促使模型在训练过程中重点关注困难样本，提高模型的边界定位能力。大量的消融实验和验证实验表明，所提出的 MYOLOv8 算法在 MS COCO 数据集上实现了更高的精度，同时具有更少的参数量、计算成本以及更快的推理速度。

参考文献:

- [1] 李悦, 马志勇, 俞建杰, 等. 融合 YOLOV5 + DSST + KCF 的目标检测跟踪算法在激光通信中的应用 [J]. 计算机测量与控制, 2023, 31 (10): 49-53.
- [2] XIAOYAN W, PEI C, RUI L, et al. Research on intelligent analysis technology of power monitoring video data based on convolutional neural network [C] // 2020 5th International Conference on Mechanical, Control and Computer Engineering. IEEE, 2020: 2461-2464.
- [3] ADAM R, JANCIAUSKAS P, EBEL T, et al. Synthetic training data generation and domain randomization for object detection in the formula student driverless framework [C] // 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering. IEEE, 2022: 1-6.
- [4] CHEN Y, LI W, SAKARIDIS C, et al. Domain adaptive faster R-CNN for object detection in the wild [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA, 2018: 3339-3348.

(下转第 113 页)