

面向机械臂抓取任务的 YOLOv8-VSC 目标检测网络

李丽婕 范江山 潘盛鑫

上海工程技术大学电子电气工程学院 上海 201620

摘要: 在计算机视觉领域, 目标检测技术对工业自动化和机器人导航等应用至关重要。YOLO 算法家族以其高效实时性备受关注, 但现有模型在复杂场景下存在复杂度高、小目标检测能力有限等问题。为此, 本文提出了一种优化的 YOLOv8-VSC 目标检测网络, 引入轻量化的 VanillaNet 模块、slim-neck 结构和 CBAM 注意力机制。该网络在保持检测效率和准确性的同时, 显著降低了模型复杂度, 提升了对小目标的检测能力。实验表明, YOLOv8-VSC 在多个标准数据集上表现出色, 尤其适合资源受限环境下的部署, 如机械臂抓取任务。

关键词: 目标检测; YOLOv8; VanillaNet; CBAM; 机械臂

引言

随着计算机视觉技术的发展, 目标检测成为最具挑战性和应用价值的研究方向之一。其核心是从图像或视频中快速、准确地识别目标对象的位置和类别, 广泛应用于工业自动化、机械臂操作和智能监控等领域。高效的实时目标检测对提高生产效率、保障安全和增强人机交互体验至关重要。

近年来, 深度学习和 CNN 推动了目标检测技术的发展。YOLO 系列将检测任务转化为回归问题, 避免了传统算法复杂的候选区域生成和特征提取, 实现了高效实时检测。但现有 YOLO 模型仍面临挑战: 复杂场景下模型复杂度高、小目标检测负担重; 多样化场景对实时性、准确性和轻量化要求更高; 小目标和复杂背景下的性能需提升。

为了克服这些挑战, 本文提出了一种改进的 YOLOv8-VSC 目标检测网络。该网络通过引入轻量化的 VanillaNet 模块、slim-neck 模块和 CBAM 注意力机制, 优化了模型的检测效率和鲁棒性。具体来说, VanillaNet 模块通过极简的网络设计显著降低了模型的计算复杂度; slim-neck 模块通过轻量化的卷积结构增强了对小目标的检测能力; 而 CBAM 注意力机制则通过通道和空间注意力的结合, 进一步提升了模型对关键特征的感知能力。这些改进不仅提高了模型的检测精度和实时性, 还显著降低了模型的复杂度, 使其更适合在资源受限的环境中部署 [1~3]。

1. 研究背景与理论基础

目标检测是计算机视觉的关键任务, 用于识别和定位图像或视频中的目标对象。近年来, 深度学习尤其是卷积

神经网络 (CNN) 推动了其发展。目标检测分为两阶段 (如 R-CNN) 和单阶段 (如 YOLO、SSD) 算法。前者精度高但速度慢, 后者速度快适合实时应用。YOLO 系列以高效实时性著称, 从 YOLOv3 到 YOLOv8 不断优化, 但模型复杂度仍较高, 限制了在资源受限设备上的应用。

1.1 YOLOv8n 网络

YOLOv8n 是 YOLOv8 系列中的轻量级版本, 专为资源受限的设备设计。它在保持 YOLOv8 高效检测能力的同时, 进一步优化了模型结构, 降低了计算复杂度。YOLOv8n 的主要特点包括:

1.1.1 轻量化的骨干网络: YOLOv8n 采用了更轻量化的骨干网络结构, 减少了模型参数和计算量, 同时保持了较高的特征提取能力。

1.1.2 高效的特征融合: 通过优化颈部网络 (Neck) 结构, YOLOv8n 在不同尺度特征的融合上表现出色, 进一步提升了对小目标的检测能力。

1.1.3 快速推理速度: YOLOv8n 在设计上注重推理速度的优化, 使其能够在嵌入式设备和移动设备上高效运行, 满足实时检测的需求。

尽管 YOLOv8n 在轻量化方面取得了显著进展, 但在一些复杂场景下, 尤其是在小目标检测和模型进一步优化方面, 仍有改进空间。为了进一步提升 YOLOv8n 的性能, 本文提出了一种改进的 YOLOv8-VSC 网络, 如图 1 所示, 通过引入以下三个关键模块来提升模型的效率和鲁棒性。

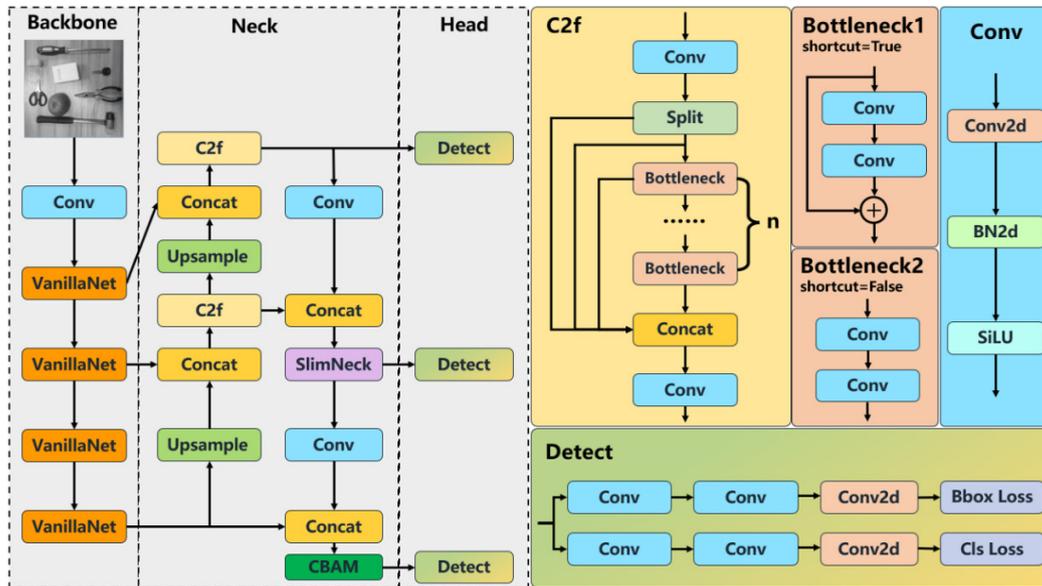


图 1 YOLOv8-VSC 网络结构图

(1) VanillaNet 模块: VanillaNet 是一种极简的神经网络架构, 摒弃了复杂的残差连接和注意力机制, 仅使用基本的卷积操作。这种设计显著降低了模型的计算复杂度, 同时保持了较高的特征提取能力, 适合在资源受限的环境中部署。

(2) Slim-neck 模块: Slim-neck 基于轻量化的 GSConv 卷积, 替代了 YOLOv8 中的 C2f 模块。GSConv 结合了分组卷积和深度可分离卷积, 降低了计算成本, 同时增强了对小目标的检测能力, 优化了特征融合过程。

(3) CBAM 注意力机制: CBAM 是一种轻量化的注意力机制模块, 通过通道注意力和空间注意力的结合, 显著提升了模型对关键特征的感知能力。该模块可以无缝集成到 CNN 架构中, 以较小的计算开销实现显著的性能提升。

这些改进不仅提高了模型的检测精度和实时性, 还显著

降低了模型的复杂度, 使其更适合在资源受限的环境中部署。

2. 实验验证与性能分析

2.1 数据集

为了评估 YOLOv8-VSC 在机械臂抓取任务中的性能, 本研究选择了 Cornell 和 Jacquard 两个数据集。Cornell 数据集包含 240 个物体的 885 张 RGB-D 图像, 标注了 8019 个抓取点 (正向 5110 个, 负向 2909 个)。Jacquard 数据集通过 CAD 模型构建, 包含 11619 个物体的 54485 张 RGB-D 图像, 分辨率为 1024×1024 , 涵盖丰富的物体类型和抓取场景, 有助于提升模型的泛化能力。为更全面验证 YOLOv8-VSC 性能, 我们将两个数据集合并, 共 55370 张图像, 覆盖多种物体和场景。



图 2 数据集

2.2 评估指标与实验环境

为评估 YOLOv8-VSC 性能, 采用以下指标: mAP@0.5 (IoU=0.5 时的平均精度)、mAP@0.5:0.95 (IoU 从 0.5 到 0.95 的平均精度)、GFLOPs (计算复杂度)、参数量、模型大小和推理速度 (单位: ms)。实验环境: Ubuntu 22.04 系统, PyTorch 1.12 框架, Python 3.9, CUDA 11.8, 硬件为 NVIDIA GeForce RTX 4090 GPU (24GB 显存), CPU 为 Intel Xeon Platinum 8375C。

2.3 实验结果分析

表 1 不同模块的影响

模型配置	mAP@0.5 (%)	mAP@0.5:0.95 (%)	GFLOPs	Parameters (M)	Model Size (M)	Speed (ms)
YOLOv8n	91.6	58.1	8.20	3.01	5.95	1.492
YOLOv8n + VanillaNet	90.6	57.3	5.07	1.73	3.48	1.187
YOLOv8n + VanillaNet + slim-neck	93.8	57.6	4.90	1.72	3.46	1.414
YOLOv8-VSC	93.9	58.2	4.54	1.26	2.59	1.227

1. VanillaNet 模块: 仅使用 VanillaNet 替换 YOLOv8n 的骨干网络后, 模型的参数量和计算复杂度显著降低, 但 mAP@0.5 和 mAP@0.5:0.95 略有下降, 表明 VanillaNet 在简化模型的同时牺牲了一定的检测精度。

2. slim-neck 模块: 加入 slim-neck 模块后, 模型的 mAP@0.5 和 mAP@0.5:0.95 显著提升, 表明 slim-neck 模块有效增强了模型对小目标的检测能力。

3. YOLOv8-VSC: 在引入 CBAM 注意力机制后, 模型的

为了深入分析 YOLOv8-VSC 网络的性能提升原因, 我们进行了消融实验, 重点研究了不同模块对模型性能的影响。实验数据与之前提到的合并后的 Cornell 和 Jacquard 数据集保持一致。

2.3.1 不同模块的影响 (Impact of Different Modules)

为了验证 VanillaNet、slim-neck 和 CBAM 模块对 YOLOv8-VSC 性能的提升效果, 我们在合并后的数据集上进行了消融实验。实验结果如表 1 所示。

检测精度进一步提升, 同时保持了较低的计算复杂度和参数量, 推理速度也有所提高。

2.3.2 不同骨干网络的影响 (Impact of Backbone Networks)

为了验证 VanillaNet 作为骨干网络的性能优势, 我们将其与其他几种流行的轻量化骨干网络进行了对比, 包括 ShuffleNetV2、EfficientNetV2、GhostConv、和 Swin Transformer。实验结果如表 2 所示。

表 2 不同骨干网络的影响

骨干网络	mAP@0.5 (%)	mAP@0.5:0.95 (%)	GFLOPs	Parameters (M)	Model Size (M)	Speed (ms)
YOLOv8n	91.6	58.1	8.20	3.01	5.95	1.492
YOLOv8n + ShuffleNetV2	85.2	57.1	6.50	2.39	5.40	1.322
YOLOv8n + EfficientNetV2	85.4	57.4	2.57	2.13	4.29	1.425
YOLOv8n + GhostConv	87.8	56.7	7.76	2.82	5.60	1.339
YOLOv8n + Swin Transformer	91.8	79.8	9.08	3.00	6.02	1.690
YOLOv8n + VanillaNet	90.6	57.3	5.07	1.73	3.48	1.187

1. VanillaNet: 在保持较高检测精度的同时, 显著降低了模型的计算复杂度和参数量, 推理速度也得到了提升。

2. Swin Transformer: 虽然在检测精度上表现最佳, 但计算复杂度和推理速度较高, 不适合在资源受限的环境中部署。

3. ShuffleNetV2 和 EfficientNetV2: 虽然在参数量和模型大小上有所减少, 但在检测精度上略低于 VanillaNet。

2.3.3 不同注意力机制的影响 (Impact of Attention Mechanisms)

为了验证 CBAM 注意力机制在 YOLOv8-VSC 中的有效性, 我们将 CBAM 与其他几种注意力机制进行了对比, 包括 SE (Squeeze-and-Excitation)、EMA (Exponential Moving Average) 和 CA (Channel Attention)。实验结果如表 3 所示。

表 3 不同注意力机制的影响

注意力机制	mAP@0.5 (%)	mAP@0.5:0.95 (%)	GFLOPs	Parameters (M)	Model Size (M)	Speed (ms)
YOLOv8n	91.6	58.1	8.20	3.01	5.95	1.492
YOLOv8n + SE	90.2	56.9	8.21	3.02	5.97	1.391

注意力机制	mAP@0.5 (%)	mAP@0.5:0.95 (%)	GFLOPs	Parameters (M)	Model Size (M)	Speed (ms)
YOLOv8n + EMA	91.7	57.9	4.58	1.26	2.65	1.439
YOLOv8n + CBAM	92.4	58.3	4.55	1.36	2.62	1.382

1.CBAM: 在提升检测精度的同时,显著降低了模型的计算复杂度和参数量,推理速度也得到了提升。

2.SE 和 EMA: 虽然在检测精度上有所提升,但在模型复杂度和推理速度上不如 CBAM。

通过消融实验,我们验证了 YOLOv8-VSC 中各个模块的有效性。VanillaNet 模块显著降低了模型的计算复杂度和参数量; slim-neck 模块增强了对小目标的检测能力; CBAM 注意力机制进一步提升了模型的检测精度和鲁棒性。这些改进使得 YOLOv8-VSC 在保持高检测精度的同时,更适合在资源受限的环境中部署。结论:

本研究提出的 YOLOv8-VSC 网络通过引入 VanillaNet 模块、slim-neck 模块和 CBAM 注意力机制,在模型复杂度、检测精度和推理速度方面显著优于 YOLOv8n。VanillaNet 实现轻量化,降低参数量和计算复杂度; slim-neck 增强小目

标检测能力; CBAM 提升检测精度和鲁棒性。实验表明, YOLOv8-VSC 在 Cornell 和 Jacquard 数据集上表现优异,更适合资源受限环境下的机械臂抓取任务。未来将优化模型结构、探索更高效的轻量化技术,并拓展应用场景以验证其泛化能力。

参考文献:

- [1] 方斌,刘飞.基于改进 YOLOv8s 的目标检测与抓取[J].制造技术与机床,2024,(12):164-171.
- [2] 郭忠峰,王健鹏,杨钧麟,等.基于改进 YOLOv8-Pose 的码垛快速识别与抓取点检测[J].组合机床与自动化加工技术,2024,(11):125-129.
- [3] 张新月,胡广锐,李浦航,等.基于改进 YOLOv8n 的轻量化红花识别方法[J].农业工程学报,2024,40(13):163-170.