

基于 YOLOv8n 的轻量化卷烟种类目标检测方法研究

韩雨芯¹
HAN Yuxin

摘要

由于单纯依靠人工检测卷烟售卖情况存在着库存管理成本高、数据更新慢和监控效率低下的问题，为了实现卷烟在烟草店售卖情况的自动化检测，有效减少烟草行业零售终端操作不规范的现象，提出一种基于 YOLOv8n 模型的卷烟种类检测算法。首先，使用 PConv 卷积方式在主干网络进行有效特征提取，在模型获取全局上下文信息的同时有效地减少模型的参数量和计算量。然后，采用加权双向特征金字塔网络 BiFPN 的特征融合思想，使模型能够有效融合不同尺度的特征图的同时保证模型的检测精确率。最后，基于自建烟草数据集进行实验，所提出算法的模型参数量是原来的 20.16%，FLOPs 降至 8.4 GB，精确率提高了 3%。实验结果表明，所提出的算法有效地减少了模型在进行目标检测时的响应时间，实现了轻量化且精确率较高的卷烟种类的目标检测方法。

关键词

YOLOv8n；卷烟种类；PConv；BiFPN；轻量化

doi: 10.3969/j.issn.1672-9528.2024.08.014

0 引言

随着中国经济的不断发展，卷烟消费情况也在经历着巨大的变革。人们对卷烟的购买动机逐渐由攀比和炫耀转变为更加理性和健康的考量，这也导致了每年新增吸烟人口数量逐渐减少^[1]。在这种背景下，有效的投放策略可以帮助公司提高市场运营效率，降低成本，提高行业竞争力。因此，烟草公司需要更加注重产品规划，以确保投放的烟草品牌能够

更好地满足市场需求。

零售终端是烟草产业链中的关键一环，它直接影响着烟草产品的销售和市场表现。通过推进卷烟方面的目标检测，可以有效确保烟草专卖店销售的产品质量、规格和标准。然而，当前零售终端存在物资管理不规范、信息共享不畅难以实时了解物资使用，以及一线人员存在违规操作拆借行为而影响库存账实符合等情况，并且传统的识别方法主要依赖于人工检测和记录销售情况，这种方式效率低、成本高、准确性较低^[2]。这些现实问题都会造成烟草公司

1. 贵州财经大学信息学院 贵州贵阳 550025

- [10] 符占元, 专祥涛. 基于神经网络和遗传算法的 PID 参数自整定算法 [J]. 武汉大学学报(工学版), 2023, 56(3):379-386.
- [11] 杨旭红, 陈阳, 方剑峰, 等. 基于改进 PSO-PID 控制器的核电站汽轮机转速控制 [J]. 控制工程, 2022, 29(12):2177-2183.
- [12] 刘尚波, 丹泽升, 廉保旺, 等. 干涉式闭环光纤陀螺仪的 PSO-PID 控制优化方法 [J]. 红外与激光工程, 2024, 53(3):250-261.
- [13] LONG Z, JIANG Z, WANG C, et al. A novel approach to control of piezo-transducer in microelectronics packaging: PSO-PID and editing trajectory optimization [J]. IEEE transactions on components, packaging, and manufacturing technology, 2020, 10(5):795-805.

- [14] 李欢欢, 刘辉, 盖江涛, 等. 基于粒子群优化算法 PID 参数优化的双电机耦合驱动履带车辆转向控制 [J]. 工兵学报, 2024, 45(3):916-924.
- [15] 曾喆昭, 刘文珏. 一阶时滞系统的智慧 PI 控制 [J]. 控制理论与应用, 2019, 36(8):1322-1328.
- [16] 金鑫, 谭文, 李志军, 等. 典型工业过程鲁棒 PID 控制器的整定 [J]. 控制理论与应用, 2005(6):947-953.

【作者简介】

刘源(1990—),女,陕西西安人,硕士,助教,研究方向:系统工程。

(收稿日期: 2024-05-20)

无法实时掌握卷烟售卖情况，从而导致其无法及时进行烟草品牌生产方面和销售方面的调整。通过自动化检测卷烟在烟草店的销售情况，不仅可以有效减少人工成本，还可以快速实现卷烟市场的库存盘点和零售终端的信息收集，及时掌握市场动向。

1 相关工作

近年来，基于深度学习的目标检测算法在计算机视觉领域取得了显著进展，并已广泛应用于多个场景。基于深度学习的目标检测算法可以分为两阶段检测方法和单阶段检测方法这两种主要类型。两阶段检测方法（如 Faster R-CNN^[3] 和 Mask R-CNN^[4]）的基本思想是将目标检测分为两个阶段：首先生成候选框，然后对候选框进行分类和精细化调整。单阶段检测方法（如 YOLO^[5] 和 SSD^[6]）的基本思想是无需生成候选框，直接在单个神经网络中完成目标检测。

随着基于深度学习的目标检测算法在工业上的快速发展，其在烟草行业的卷烟生产和销售过程中也得到了广泛应用。例如，单宇翔等人^[7]利用 Mask R-CNN 构建了一种鲁棒性强、精确度高的卷烟烟盒二阶段分类算法，通过对烟盒陈列区域进行实例分割和正则化等操作，提高了检测效果。董佳等人^[8]基于 Faster R-CNN 模型进行优化设计，提出了一种二阶段分类算法，增强了烟草监管部门对烟草产品的准确识别和统计能力。张文静等人^[9]应用 Inception V3 网络通过迁移学习在复杂环境下识别烟草病害。钟宇等人^[10]将 Support Vector Machine (SVM) 支持向量机应用于卷烟包装的真伪鉴别，相比于人工的方法，可显著提高鉴别效率，解决主观性强的问题。虽然这些深度学习的方法使得烟草品牌监测效率和准确性得到了极大的提高，但在资源有限的环境下进行高精确率和快速响应的目标检测时仍有较大的提升空间。

2 改进 YOLOv8 算法

尽管两阶段检测算法具有更高的准确性和泛化能力，但该方法对比单阶段检测算法，计算复杂且检测速度慢，并不适合应用于对实时性要求较高的场景。随着一阶段目标检测算法的快速发展，相较于 SSD 模型先利用多尺度特征图预测边界框，接着计算预测边界框的相对偏移量和尺寸，并结合先验框的坐标和尺寸来计算真实的边界框的方法，YOLO 系列的算法计算成本更小。YOLOv8n 模型^[11] 使用解耦头（Anchor-free）的设计。相较于之前版本中的耦合头（Anchor-base）的方法，YOLOv8n 采用无锚模型，通过解耦头独立处理目标检测对象的分类和回归任务，使模型在目标检测过程中不仅有高精确率，还有高响应速度，因此 YOLO 算法能够更好地应用于工业实时检测任务。

为了将目标检测模型部署到工业自动化和智能监控等算力资源较为有限的边缘设备上，本文的算法要兼顾精确率和速度。本文针对传统的 YOLOv8n 模型中的算法进行了专门的轻量化改进，一方面通过优化特征提取过程显著降低了模型的计算量和参数量，另一方面通过优化特征融合方式使模型保持高精确率的检测效果。本文提出的 PB-YOLOv8n 模型的整体架构如图 1 所示。图 1 中展示了从输入图像到最终检测结果的整个处理流程，包括特征提取、特征融合以及检测输出等关键步骤。本文提出的 PB-YOLOv8n 模型通过这种结构设计，不仅优化了计算过程，还提高了目标检测的实时性，这对于实时监控和需要快速响应的工业应用场景至关重要。

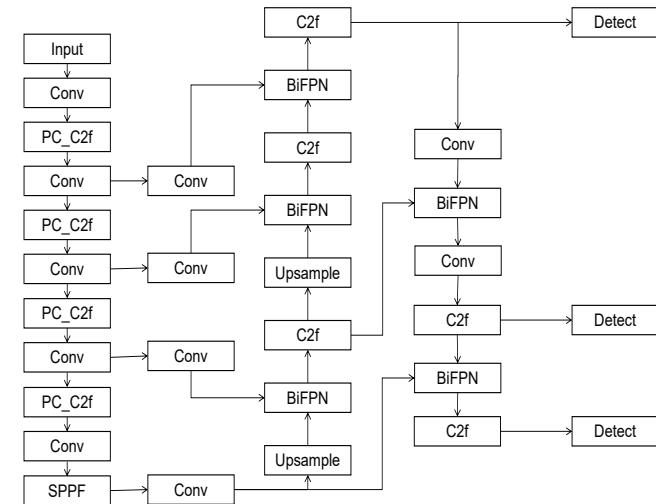


图 1 PB-YOLOv8 架构图

本文首先将主干网络里的传统卷积替换为部分卷积 PConv^[12]，通过对输入特征图的部分通道进行卷积操作，有效降低了模型的参数量和计算量，既保证了输出通道大小不变，又减少了计算成本。其次，以往的 YOLO 算法经常会采用 FPN（特征金字塔网络）^[13] 和 PANet（路径聚合网络）^[14] 的特征融合方式。但是，FPN 在融合特征的过程中只有自顶向下的过程，很难把低层信息传递到最后一层。虽然 PANet 相比于 FPN 多了一个自底向上的传递，让图像的低层信息也能得到充分的利用，但是 PANet 中的单节点并没有融合多层的特征，能够获取的信息并不多，却增加了模型的参数。

为了克服以上方式在特征融合上的缺点，本文在颈部网络中使用更加灵活和高效的加权双向特征金字塔网络 BiFPN^[15]，通过在不同尺度的特征图之间建立起双向连接的方式，充分利用细节信息和语义信息，让模型能够充分学习目标特征，从而实现更加精准的定位和分类功能，提高模型的准确性和鲁棒性。

在烟盒种类的目标检测任务中，特征图的通道冗余是导

致计算成本高的一个主要因素。由于模型在卷积过程中产生的特征图在不同通道之间具有高度的相似性，如果这部分冗余不处理，将直接影响模型的效率和速度。在不影响模型性能的前提下，为了有效降低这种冗余并减少模型的参数量，本文通过在模型中使用 PConv 卷积的方式代替标准卷积来减少参与计算的通道数，降低特征图的冗余度。PConv 卷积方式是通过选取前段或后段连续 c_p 个通道来代表整个特征图，不仅可以保持输出通道的大小不变，同时保证了内存的连续性，还减少了计算成本。普通卷积的计算量如公式（1）所示，PConv 的计算量如公式（2）所示。当 $c_p/c=1/4$ 时，PConv 的计算量仅为普通卷积计算量的 $1/16$ ，因此本文改进后的算法在有效提取目标特征的同时不仅可以减少模型的计算量，还可以有效提高模型的计算速度。

$$h \times w \times k^2 \times c^2 \quad (1)$$

$$h \times w \times k^2 \times c_p^2 \quad (2)$$

每一层的 PConv 的运算可以用式（3）表达：

$$x' = \begin{cases} W^T(X \odot M) \frac{\text{sum}(1)}{\text{sum}(M)} + b, & \text{if } \text{sum}(M) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

式中： W 为卷积核； X 为输入的 feature map， M 为输入的 mask（0-1 分布）， \odot 表示逐元素相乘，且 X 和 M 是当前运算区域的像素； x' 是单个像素，即卷积核 W 的中心位置像素； b 为卷积运算的偏置； $\text{sum}(1)$ 是指与卷积核相同大小（例如 3×3 ），元素全部为 1 的矩阵。

因此，本文将 Pconv 替换模型的传统卷积，构建 PC_C2f 模块，如图 2 所示。

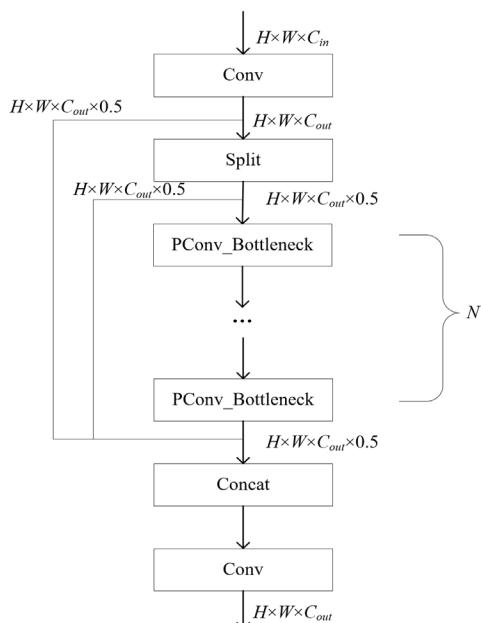


图 2 PC_C2f 结构图

在烟盒目标检测领域，烟盒目标经常被遮挡、反光等因素影响，摄像头捕捉到的图像特征会有较大差异，因此要对主干网络提取到的特征进行有效的特征融合。

BiFPN 结构保留了 PANet 中的自上而下和自下而上的特征融合路径，同时对深层与浅层特征进行了跨尺度融合。此外，BiFPN 还引入了可学习的权重，用于对不同的输入特征进行学习，其计算公式为：

$$O = \sum_i \frac{w_i}{\varepsilon + \sum_j w_j} \quad (4)$$

式中： w_i 表示每个输入项的权重， ε 表示一个很小的正数， $\sum_j w_j$ 表示所有权重的总和。

这种设计方式在增加少量计算量的前提下提高了模型在不同网络层级之间传递特征图的信息传递效率，有效改善了由于模型在卷积过程中造成的图像特征细节信息丢失的问题。本文考虑到浅层特征图保留了目标的更多细节信息，而这些信息对于精确定位和识别小目标至关重要，因此，本文提出在颈部网络中使用 P2、P3、P4、P5 这四个层级的特征图，通过增强的跨尺度连接和自顶向下、自底向上的信息流进行有效的特征融合，以利用不同尺度的特征信息。本文使用四层 BiFPN 结构如图 3 所示。

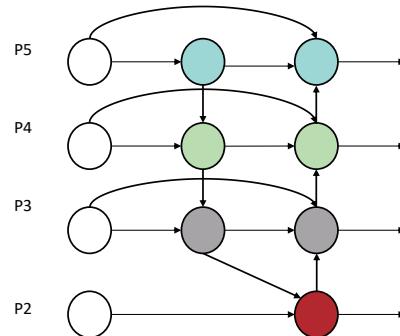


图 3 四层 BiFPN 结构图

3 实验

3.1 实验平台及参数设置

本文的所有实验都是基于 Linux 操作系统，CPU 型号是 Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20 GHz，配置两个 NVIDIA Tesla P100 显卡，显存共 24 GB，深度学习框架为 PyTorch1.10.1+cu111。模型所用参数如表 1 所示。

表 1 模型参数信息

参数	值
Image_size	640×640
Batch_size	16
Epochs	120
Optimizer	SGD
Num_workers	8
Num_devices	2

3.2 实验数据及评价标准

本文所使用的数据集为自建数据集，包含了 146 种盒烟照片，共计 19 484 张图片。其中，训练集有 13 703 张，验证集有 5781 张。在标注过程中，使用的是 LabelImg 标注工具，样例数据集如图 4 所示。



图 4 数据集示例

基于 YOLOv8n 的目标检测的评价指标主要有精确率 (precision, P)、召回率 (recall, R)、平均精确率 (average precision, AP)、全类平均精确率 (mean average precision, mAP)。其中， P 表示模型正确检测到的目标数量占所有检测到的目标数量的比例，即模型区分负样本的能力，如公式 (5) 所示； R 表示模型在实际为正的样本中被预测为正样本的概率，即模型识别正样本的能力，如公式 (6) 所示；AP 是通过积分 P-R 曲线得到的面积值，如公式 (7) 所示，P-R 曲线是以精确率为纵坐标、 R 为纵坐标的样本点组成的曲线；mAP 是 AP 的平均值，如公式 (8) 所示。此外，参数量 (Params) 等指标用于衡量模型的计算复杂度，FPS 用于衡量模型的检测速度。

$$P = \frac{TP}{TP + FP} \quad (5)$$

$$R = \frac{TP}{TP + FN} \quad (6)$$

$$AP = \int P(R)dR \quad (7)$$

$$mAP = \frac{1}{C} \sum_i^C AP_i \quad (8)$$

式中：TP (true positive) 是预测为正例，真实值也为正例的个数；FP (false positive) 是预测为正例，实际为负例的个数；FN (false negative) 是预测为负例，实际是正例的个数；本文的 mAP 表示在 IoU 阈值在 50% ~ 95% 的范围内的 mAP 平均值；IoU 表示预测的边界框和真实边界框之间的重叠程度。

3.3 消融实验性能对比

为了进一步验证 PB-YOLOv8n 中不同改进部分的集体效果，本文对每个模块进行了消融实验，实验结果如表 2 所示。

表 2 消融实验对比

实验	$P/\%$	mAP/%	Params	FLOPs/GB	FPS
1	80.2	57.1	3 354 662	9.7	24.6
2	81.9	57.3	2 901 350	8.4	25.7
3	78.3	55.9	3 137 242	9.8	22.4
4	83.3	56.9	2 678 322	8.4	23.4

实验 1 是未改进的 YOLOv8n 模型；实验 2 是在主干网络中的 C2f 模块替换为本文提出的 PC_C2f 模块，在精确率和检测速度上都有所提高；实验 3 是在颈部网络中构建 BiFPN 的跨级连接方式；实验 4 是本文提出的 PB_YOLOv8n 模型，通过融合 PC_C2f 模块和 BiFPN 网络结构，模型参数量是原来的 20.16%，FLOPs 降至 8.4 GB，精确率提高了 3%。虽然 BiFPN 中的跨级连接方式会增加模型的计算量，但是由于本文所引入的 PConv 的参数共享特性可以与 BiFPN 结构的信息聚合相互配合，使得模型在引入 BiFPN 后的模型的参数量和计算量得以减少，并且能够有效提高模型的精确率和检测速度。由图 5 可知，本文的改进方法是有效的，改进后的 PB_YOLOv8n 结构可以在模型轻量化的同时保证正确分类和定位烟盒种类，满足针对烟盒种类进行实时目标检测的工业需求。

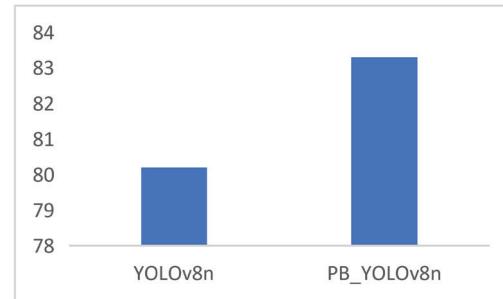


图 5 模型修改前后精度值的对比

3.4 对比实验性能对比

为了更进一步验证本文提出的改进算法的有效性，将改进后的 PB-YOLOv8n 与流行的两阶段检测算法 Faster-RCNN、一阶段检测算法 SSD，以及 YOLO 系列的代表性算法如 YOLOv5n、YOLOv7-tiny 进行了比较，具体对比结果如表 3 所示。

表 3 不同模型对比实验

模型	$P/\%$	$R/\%$	mAP/%	FPS
Faster-RCNN	44.4	74.3	68.7	8.8
SSD	66.5	15.9	47.4	26.2
YOLOv5n	75.5	76.9	50.4	27.0
YOLOv7-tiny	78.1	83.3	53.9	17.9
YOLOv8n	80.2	80.3	57.1	24.6
PB-YOLOv8n	83.3	80.5	56.9	23.4

由表3可以看出，本文提出的PB-YOLOv8n模型相对于其他主流目标检测模型精确率更高，计算量更小，并且响应速度较快，更适合在监控这种资源有限的环境下进行应用和推广。PB-YOLOv8n采用了轻量级的网络架构设计和高效的特征融合策略，在保持较低计算量的同时，通过优化卷积核和采用BiFPN的跨级连接方式有效提升了主干网络中特征提取和颈部网络特征融合的能力，这使得模型在多尺度和复杂背景下都能保持较高的检测精确率，能够显著提高模型的运行效率。

4 结论

针对传统烟盒识别过程中人工记录费时、检测效率低和实时检测速度慢的问题，本文提出一种基于改进YOLOv8n的轻量化模型PB-YOLOv8n。首先，本文为了在主干网络中更高效地提取特征，使用PC_C2f模块替换原始网络中的C2f模块，目的是通过一种更高效的部分卷积方式，显著减少模型的参数量和计算量，同时保持特征提取的高效性和准确性。接着，为了更好地融合在主干网络部分提取到的目标特征，本文改进了颈部网络的特征融合结构。本文采用的BiFPN跨级连接方式，实现了主干网络提取的特征图与多个层级之间的特征信息融合。这种方法增强了网络不同层级间的信息传递，有效地减少了因特征图分辨率变化导致的图像特征信息丢失。此外，通过BiFPN加权融合的方式，本文的算法能够自适应地调整主干网络中得到的各层特征图的权重，从而实现更精准的目标检测。实验结果表明，本文提出的PB-YOLOv8n算法比原算法的参数量降低了20.16%，单张检测速度为12.4 ms，提高了27.91%，精确率为83.3%，优于Faster-RCNN、SSD、YOLOv5n、YOLOv7-tiny、YOLOv8n等主流算法，并且本文提出的算法满足烟盒检测精确率以及部署在算力有限环境中的要求。未来将进一步收集更多环境下的烟盒图像，在模型轻量化的基础上提高模型的精确率和泛化能力，减少目标物体误检、漏检等问题的发生。

参考文献：

- [1] 高淳. 经济新常态下卷烟品牌营销策略分析 [J]. 活力, 2023(14):172-174.
- [2] 刘璐璐. 企业数字化审计路径探索——以烟草行业零售终端物资管理为例 [J]. 财经界, 2023(31):156-158.
- [3] 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 and machine intelligence, 2017, 39:1137-1149.
- [4] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[J].

IEEE transactions on pattern analysis and machine intelligence, 2020, 42:386-397.

- [5] REDMON J, DIVVALA S K, GIRSHICK R B, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016:779-788.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[EB/OL].(2015-12-08)[2024-04-23].<https://arxiv.org/abs/1512.02325>.
- [7] 单宇翔, 龙涛, 楼卫东, 等. 基于深度学习的复杂场景中卷烟盒检测与识别方法 [J]. 中国烟草学报, 2021, 27(5):71-80.
- [8] 董佳, 陈杜勇, 刘凯平, 等. 融合深度学习的烟草品牌识别模型 [J]. 福建电脑, 2024, 40(3):34-38.
- [9] 张文静, 孙秀朋, 乔永亮, 等. 基于InceptionV3的烟草病害识别 [J]. 中国烟草学报, 2021, 27(5):61-70.
- [10] 钟宇, 徐燕, 刘德祥, 等. 基于计算机视觉和机器学习的真伪卷烟包装鉴别 [J]. 烟草科技, 2020, 53(5):83-92.
- [11] KHARE O, GANDHI S, RAHALKAR A, et al. YOLOv8-based visual detection of road hazards: potholes, sewer covers, and manholes[C]//2023 IEEE Pune Section International Conference (PuneCon). Piscataway: IEEE, 2023:10449999.
- [12] CHEN J, KAO S, HE H, et al. Run, don't walk: chasing higher flops for faster neural networks[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023:12021-12031.
- [13] LIN T, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . Piscataway: IEEE, 2017:10716717.
- [14] 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. Piscataway: IEEE, 2018:8759-8768.
- [15] TAN M, PANG R, LE Q V. Efficientdet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Piscataway: IEEE, 2020: 10781-10790.

【作者简介】

韩雨芯（1999—），女，山东菏泽人，硕士研究生，研究方向：计算机视觉、计算机系统与结构。

（收稿日期：2024-05-18）