计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications [doi: 10.15888/j.cnki.csa.009695]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于变形卷积和多重注意力的零售商品检测①

王　添[1], 刘立波[1,2]

[1](宁夏大学 信息工程学院, 银川 750021)
[2](宁夏"东数西算"人工智能与信息安全重点实验室, 银川 750021)
通信作者: 刘立波, E-mail: liulib@163.com

摘　要: 针对零售商品旋转和变形导致难以准确提取全局特征及无关特征干扰的问题, 提出一种基于改进 YOLOv8s 的零售商品检测算法. 首先, 利用归一化可变形卷积替代部分标准卷积, 通过充分捕获长距离依赖关系以及突出通道关键特征, 增强对全局特征的提取能力; 其次, 使用改进的动态检测头, 使用基于空间感知、尺度感知和任务感知的多重注意力机制来捕获更具区分性的商品局部特征, 以抑制无关特征干扰; 最后, 采用 InnerEIoU 损失函数替换 CIoU, 以降低商品漏检率. 实验结果表明, 所提算法在 RPC 零售商品数据集上的 *mAP*@0.5:0.95 达到 93.3%, 较原始算法提升了 1.5%, 并优于其他主流检测算法; 同时模型参数量和计算量分别下降了 10.0% 和 6.5%, 能够在存储和计算资源受限的实际场景中, 准确地进行零售商品检测.

关键词: 零售商品检测; YOLOv8s; 可变形卷积; 轻量级; 注意力机制

引用格式: 王添,刘立波.基于变形卷积和多重注意力的零售商品检测.计算机系统应用. http://www.c-s-a.org.cn/1003-3254/9695.html

Retail Commodity Detection Based on Deformable Convolution and Multiple Attention

WANG Tian[1], LIU Li-Bo[1,2]

[1](School of Information Engineering, Ningxia University, Yinchuan 750021, China)
[2](Ningxia Key Laboratory of Artificial Intelligence and Information Security for Channeling Computing Resources from the East to the West, Yinchuan 750021, China)

Abstract: A retail commodity detection algorithm based on improved YOLOv8s is proposed in response to the difficulty in accurately extracting global features and irrelevant feature interference caused by retail commodity rotation and deformation. Firstly, using normalized deformable convolutions to replace some standard convolutions enhances the ability to extract global features by fully capturing long-range dependencies and highlighting key channel features. Secondly, using an improved dynamic detection head and a multi-attention mechanism based on spatial perception, scale perception, and task perception captures more discriminative local features of goods to suppress irrelevant feature interference. Finally, the InnerEIoU loss function is used to replace CIoU to reduce the missed detection rate of goods. Experimental results show that the proposed algorithm achieves an *mAP*@0.5:0.95 of 93.3% on the RPC retail commodity dataset, which is 1.5% higher than the original algorithm and better than other mainstream detection algorithms. At the same time, the number of model parameters and the amount of computation decrease by 10.0% and 6.5% respectively, enabling accurate retail commodity detection in practical scenarios with limited storage and computing resources.

Key words: retail commodity detection; YOLOv8s; deformable convolution; lightweight; attention mechanism

① 基金项目: 国家自然科学基金 (62262053); 宁夏科技创新领军人才计划 (2022GKLRLX03)
收稿时间: 2024-04-25; 修改时间: 2024-06-17; 采用时间: 2024-06-28; csa 在线出版时间: 2024-09-24

随着信息技术的快速发展, 利用前沿科技赋能传统业务已成为推动零售业发展的新趋势[1]. 人工收银存在耗时长、速度慢、成本高等缺点, 严重影响着商品结算效率. 自助收银[2]作为新兴的收银结算方式, 不仅能够有效弥补人工收银的局限性, 而且还可以提升用户的购物体验, 具有重要的研究价值和实际意义.

传统的自助结算方式包括自助扫码和无线射频识别[3], 但存在条码易污染或信号干扰等问题. 随着计算机视觉技术的发展, 基于商品检测的无人自助结算方式逐渐兴起, 按照检测方法可分为两类: 一类是基于人工特征[4]的视觉方法, 通过提取商品的颜色、纹理等特征来进行检测, 然而提取过程较复杂, 可用特征较少, 难以优化检测性能. 另一类是基于深度学习的视觉方法, 其不仅特征提取方式简单, 而且能够捕获更丰富、抽象的语义特征, 逐渐成为当前主流的商品检测模型. 然而, 基于深度学习的商品检测仍然面临着诸多挑战, 在实际结算过程中摆放的商品经常出现不同程度的杂乱现象, 其中商品的旋转和包装变形问题尤为突出[5]. 吕晓华等[6]引入轻量级多层感知机以捕获变形商品的全局特征, 进而感知商品的整体外观, 但难以理解不同像素间的空间关系, 限制了对全局特征的捕获. Li 等[7]使用三维注意力机制 SimAM 来增强变形商品的特征表示以聚合关键特征, 但其因特征空间分布变化而易聚焦于背景信息, 使模型被无关特征干扰. Hu 等[8]提出图卷积网络以有效捕获旋转不变特征, 提高对旋转商品的检测精度, 但过于依赖图结构且计算成本较大, 同时无法应对商品变形现象. YOLOv8s 算法通过先进的卷积神经网络架构保留了不同像素的空间关系以增强全局特征提取能力. 同时其密集残差结构提取的更抽象特征表示能够减少对特征分布变化的依赖以聚焦商品特征. 然而, 模型自身也存在一定的局限性: 卷积层的规则网格采样方式限制了网络感受野, 难以更准确地提取商品全局特征; 而且 PANet[9]模块通过特征图直接缩放并相加来融合特征, 会混合不同层级的有效特征与背景特征, 导致模型易受到无关特征干扰; 同时 CIoU 损失函数在计算损失时使用宽高比, 存在失效问题, 增加了商品漏检的风险.

综上, 本文提出一种基于改进 YOLOv8s 的零售商品检测算法 (YOLOv8-NDI). 主要贡献如下: 1) 采用归一化可变形卷积 (NDC) 模块代替部分标准卷积, 在突出通道显著特征的同时捕获空间长距离依赖关系, 扩大模型感受野, 以增强对全局特征的提取能力; 2) 引入改进的动态头 (DH) 模块, 利用多重注意力捕获更具区分性的商品局部特征, 以抑制无关特征干扰; 3) 引入 InnerEIoU 损失函数, 利用辅助边界框的宽高值重新计算宽高损失, 以降低商品漏检率. 在 RPC 数据集上通过与原模型及其余主流模型进行对比实验, 结果表明所提算法具有较好的检测结果.

1 所提算法

1.1 YOLOv8-NDI 整体结构

YOLOv8-NDI 的整体架构如图 1 所示, 主要分为 3 个部分: 主干、颈部和头部. 其中, 主干端用于特征提取. 利用 CBS 模块的卷积、批标准化 (BN)[10]、SiLU 激活函数等操作, 增强模型非线性表达能力. 同时使用轻量级 C2f 模块, 通过密集残差结构增强特征表达能力. 此外, 引入 SPPF[11]模块有效捕获多层级的空间特征; 颈部端采用基于 FPN[12]和 PAN 的特征金字塔结构, 融合主干端输出特征; 头部端采用解耦头结构, 将回归分支和预测分支分离, 加速模型收敛[13]. 同时考虑到 YOLOv8s 算法的局限性, 设计了 3 部分改进. 首先, 针对所有 C2f 模块中的 Bottleneck 部分, 使用归一化可变形卷积模块替换原先的第 2 个卷积层, 以增强对全局特征的准确提取; 其次, 引入改进的动态头模块, 将其嵌入原先检测头的首部, 以抑制无关特征干扰; 最后, 采用 InnerEIoU 损失函数替换 CIoU[14], 降低商品漏检率.

1.2 引入归一化可变形卷积

在实际结算场景中, 模型对商品特征的提取能力极大地受到其摆放杂乱程度的影响. 具体表现在商品自身旋转和包装变形等因素. 标准卷积采用规则网格提取特征, 无法有效捕获旋转和变形商品的部分边缘细节特征, 导致其全局外观特征提取不充分, 难以应对不同杂乱程度下的准确检测, 计算公式如下:

$$\boldsymbol{Y}(p_0)=\sum_{p_n\in R}\boldsymbol{W}(p_n)\cdot\boldsymbol{X}(p_0+p_n) \tag{1}$$

其中, $\boldsymbol{Y}$为输出特征图, p_0为当前像素位置, p_n为以p_0为中心的网格各点偏移量, $\boldsymbol{R}$为偏移量集合$\{(-1,-1)(-1,0),\cdots,(1,1)\}$, $\boldsymbol{W}$为卷积核, $\boldsymbol{X}$为输入特征图.

为此, 本文引入 Zhu 等[15]提出的可变形卷积网络 (DCNv2), 通过偏移量自适应改变采样位置, 使模型能

够关注到旋转和变形商品的边缘细节特征, 同时根据调制标量区分感兴趣区域, 将感受野锁定在目标范围内. 可变形卷积 (DC) 计算公式如下:

$$\boldsymbol{Y}(p)=\sum_{n=1}^{N}\boldsymbol{W}_n\cdot\boldsymbol{X}(p_0+p_n+\Delta p_n)\cdot\Delta m_n \tag{2}$$

其中, Δp_n为偏移量, Δm_n为调制标量, $p=p_0+p_n+\Delta p_n$, 使用双线性插值法 (BI) 计算加入浮点数偏移量后的新像素值, 可变形卷积示意图如图 2 所示.

可变形卷积虽然有效捕获了空间维度的商品特征, 但其学习的偏移量等参数会被分配到通道维度的所有特征图中. 由于各通道特征本身存在较大质量差异且分布不均衡, 对所有通道特征进行空间维度的参数学习, 会使其因背景干扰而出现偏差, 导致无法精准定位商品空间位置. 因此, 本研究引入归一化注意力机制 (NAM)[16]在无参数量增加的同时提升通道局部特征的权重分配, 以充分捕获商品判别性特征且减少背景干扰, 并结合该机制形成 NDC 模块. 通过将 NDC 替换基准网络中 Bottleneck 模块的第 2 个卷积层, 从空间和通道维度中充分捕获长距离依赖关系, 并利用自适应调整后的采样方式, 更精准地对齐到商品的边缘和轮廓, 促进商品全局外观特征的有效捕获, 从而解决不同杂乱程度下商品的误检等问题, NDC 模块结构如图 3 所示.

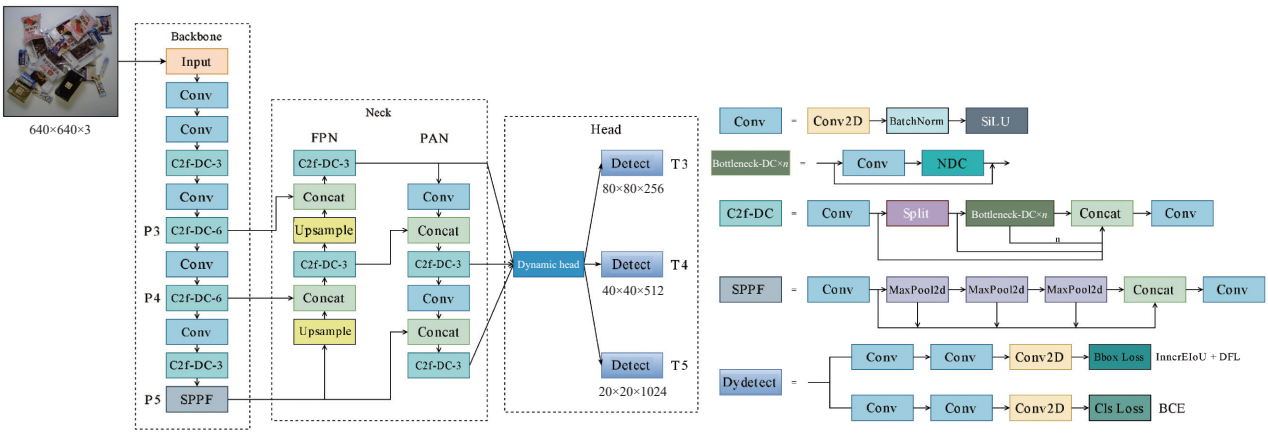


图 1 YOLOv8-NDI 的整体架构

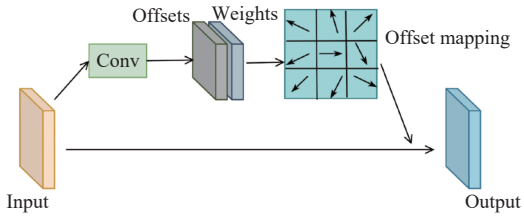


图 2 可变形卷积模块示意图

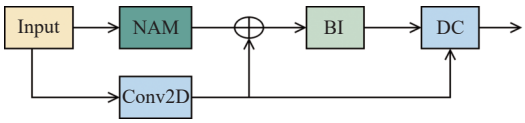


图 3 归一化可变形卷积模块

首先, 将输入特征图$\boldsymbol{F}\in\mathbb{R}^{H\times W\times C}$使用 BN 层, 计算公式如下:

$$\mathrm{BN}(\boldsymbol{x})=\gamma\left((\boldsymbol{x}-\mu_B)/\sqrt{\sigma_B^2+\varepsilon}\right)+\beta \tag{3}$$

其中, μ_B 和σ_B分别是批量B的均值和标准差, γ和β是仿射变换参数, $H\times W$为图像分辨率, C为通道数. 再根据缩放因子γ计算各通道权重, 最后利用$Sigmoid$函数提升特征的泛化性, 得到特征$\overline{\boldsymbol{F}}\in\mathbb{R}^{H\times W\times C}$, 计算过程如下:

$$\overline{\boldsymbol{F}}=Sigmoid\left(W_\gamma(\mathrm{BN}(\boldsymbol{F}))\right) \tag{4}$$

其中, $\mathrm{BN}(\cdot)$表示批量归一层, $W_\gamma=\gamma_i/\sum_{j=0}\gamma_j$为通道权重.

其次, 将F通过标准卷积层计算可学习参数, 接着将参数与上述优化特征$\overline{\boldsymbol{F}}$相加并利用双线性插值求得新特征图, 最后对新特征图使用可变形卷积层, 计算公式如下:

$$\mathrm{NDC}(\boldsymbol{F})=\mathrm{DC}\left(\mathrm{BI}\left(\mathrm{Conv}(\boldsymbol{F})\oplus\overline{\boldsymbol{F}}\right)\right) \tag{5}$$

其中, $\mathrm{Conv}(\cdot)$表示标准卷积层, $\mathrm{BI}(\cdot)$表示双线性插值, $\mathrm{DC}(\cdot)$表示可变形卷积层.

1.3 引入改进的动态检测头

针对杂乱场景下旋转和变形商品的特征图出现空间位置不连续及错位扭曲现象, 造成商品与背景特征之间混合, 使模型被无关特征干扰的问题, 本文引入基于多重注意力机制的动态检测头 (dynamic head)[17], 通过在空间位置、层级尺度和输出通道中分别运用空间感知、尺度感知和任务感知注意力, 以捕获更精细且更具区分性的局部细节商品特征, 准确区分杂乱摆放商品自身与背景之间的特征差异, 并增强模型对有用特征的利用能力, 以抑制无关特征干扰, 从而有效利用局部判别性特征对杂乱摆放商品进行精准识别.

在输入特征到 dynamic head 模块之前, 需要调整 Neck 端输出的不同层级特征尺度. 通过上、下采样方式得到三维张量$\boldsymbol{F}\in\mathbb{R}^{L\times S\times C}$, 其中$L$表示输出层级数, S表示特征图尺寸, C表示通道数. 接着将特征图依次输入空间感知、尺度感知和任务感知的注意力模块中. 文中设置 Neck 端输出的 3 层特征图通道数均为 128.

空间感知模块用于聚焦不同空间位置的判别区域. 首先使用 3×3 卷积获取特征图的偏移量和调制量, 接着利用可变形卷积调整杂乱摆放商品的特征空间分布以增强不同杂乱商品间的可区分性, 最后采用自适应加权融合方式跨层级聚合特征以提高模型对商品杂乱场景的理解能力, 其计算公式如下:

$$\pi_S(\boldsymbol{F})\cdot\boldsymbol{F}=\frac{1}{L}\cdot\sum_{l=1}^{L}\sum_{k=1}^{K}\omega_{l,k}\cdot\boldsymbol{F}(l;p_k+\Delta p_k;c)\cdot\Delta m_k \tag{6}$$

其中, π_S表示空间感知注意力, K表示采样位置个数, $p_k+\Delta p_k$表示可学习的偏移位置, Δm_k表示具体采样点的调制标量, L表示层级数量, $\omega_{l,k}$表示卷积核某点的特征值. 为进一步使卷积在训练过程中获取更全面的特征信息, 以增加商品特征与背景特征的区分度, 将可变形卷积部分替换为由 Wang 等提出的 DCNv3 模块[18], 相较于原来有 3 处改进. 首先, 使用深度可分离卷积和全连接层来获取偏移量和调制量参数, 以降低模型计算量; 其次, 将空间聚集部分进行分组, 使每组具有相互独立的学习参数, 以减少不同杂乱商品间的特征干扰; 最后, 利用基于所有采样点的 Softmax 归一化替代原来的逐采样点的 *Sigmoid* 归一化, 以调整对不同采样点的关注程度, 提取更关键的商品特征.

尺度感知模块用于对不同尺度特征进行融合, 帮助模型理解和定位不同摆放杂乱商品的关系和位置. 首先进行全局平均池化, 以实现特征降维, 接着使用 1×1 卷积, 然后利用 Hard-Sigmoid 函数激活, 最后将得到的张量与输入特征相乘, 其计算公式如下:

$$\pi_L(\boldsymbol{F})\cdot\boldsymbol{F}=\sigma\left(f\left(\frac{1}{SC}\sum_{S,C}\boldsymbol{F}\right)\right)\cdot\boldsymbol{F} \tag{7}$$

其中, π_L表示尺度感知注意力, $\sigma(x)=\max(0,\min(1,(x+1)/2))$为 Hard-Sigmoid 激活函数, $f(\cdot)$表示 1×1 卷积, $\frac{1}{SC}\sum_{S,C}\boldsymbol{F}$表示单个维度下的平均池化.

任务感知模块用于适配各种视觉任务. 利用动态线性整流函数 (DyReLU)[19], 根据不同任务来调整特征通道数, 其计算公式如下:

$$\pi_C(\boldsymbol{F})\cdot\boldsymbol{F}=\max(\alpha_1(\boldsymbol{F})\cdot\boldsymbol{F}_C+\beta_1(\boldsymbol{F}),\alpha_2(\boldsymbol{F})\cdot\boldsymbol{F}_C+\beta_2(\boldsymbol{F})) \tag{8}$$

其中, π_C表示任务感知注意力, $[\alpha_1,\alpha_2,\beta_1,\beta_2]^{\mathrm{T}}=\theta(\cdot)$是一种控制激活阈值的超函数. 通过全局平均池化、两个全连接层、一个归一化层和 Shifted Sigmoid 函数, 将参数映射到 [−1, 1].

本研究将 3 个注意力机制模块按照空间、尺度和任务感知的顺序, 串联构成 dynamic block 模块, 并对其重复堆叠 3 次, 形成全新的 dynamic head 检测头, 其结构如图 4 所示.

1.4 InnerEIoU 损失函数

YOLOv8s 使用 CIoU 作为边界框损失函数, 计算公式如下:

$$\begin{cases} L_{\mathrm{IoU}}=1-\dfrac{|b\cap b^{\mathrm{gt}}|}{|b\cup b^{\mathrm{gt}}|} \\ v=\dfrac{4}{\pi^2}\left(\arctan\dfrac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}}-\arctan\dfrac{w}{h}\right)^2 \\ a=\dfrac{v}{L_{IoU}+v} \\ L_{\mathrm{CIoU}}=L_{\mathrm{IoU}}+\dfrac{\sigma(b,b^{\mathrm{gt}})}{(w^c)^2+(h^c)^2}+av \end{cases} \tag{9}$$

其中, b^{gt}, w^{gt}, h^{gt}分别为真实框中心点、宽和高; b, w, h分别为预测框中心点、宽和高; $\sigma(b,b^{\mathrm{gt}})$表示真实框与预测框之间的欧氏距离; w^c和h^c分别为真实框与预测框的最小外接矩形的宽和高; v为纵横比参数; a为协调比例参数.

CIoU 在设计宽高的损失项时, 使用宽和高的相对比例而非具体值. 根据v的定义, 当预测框宽和高的值满足$\{(w=kw^{\mathrm{gt}},h=kh^{\mathrm{gt}})\mid k\in\mathbb{R}^+\}$时, 该损失项则不再起

作用[20]. 同时计算相对比例意味着它对宽高比的敏感度是固定的, 而摆放杂乱的零售商品往往因旋转或变形导致其形状多变, 所以固定的宽高比可能不足以准确描述其多样性. 为此, 本文将 CIoU 损失函数替换为 EIoU[21], 不仅兼顾重叠面积、中心点距离和宽高 3 方面的损失, 还将纵横比v的影响因子拆开, 分别计算宽高损失, 克服 CIoU 损失的局限性, 以适应不同形状的商品并提升边界框的回归精度和收敛速度, EIoU 计算公式如下:

$$L_{\mathrm{EIoU}} = L_{\mathrm{IoU}} + \frac{\sigma^2(b,b^{\mathrm{gt}})}{c^2} + \frac{\sigma^2(w,w^{\mathrm{gt}})}{w^c} + \frac{\sigma^2(h,h^{\mathrm{gt}})}{h^c} \tag{10}$$

其中, c表示预测框与真实框最小外接矩形的对角线长度.

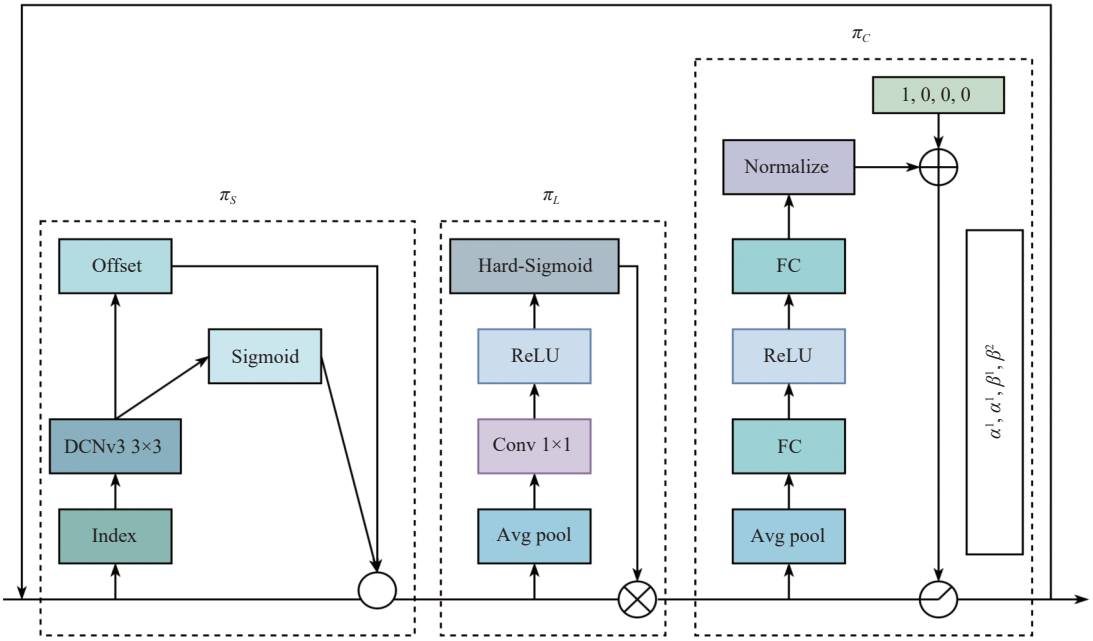


图 4 改进动态头模块

此外, 本文还借助辅助框进一步提升定位边界框的准确率. InnerEIoU 损失在原边界框相对位置基础上, 根据预测框和真实框尺度大小, 将两者宽高调整为原来的 ratio 倍, 以生成辅助边框, 之后借助辅助边框重新计算 EIoU 损失, 所提方法能够更精确地捕捉到摆放杂乱商品的实际尺寸, 有效降低了商品漏检率.

2 实验与结果分析

2.1 实验环境及参数设置

实验使用 Ubuntu 16.04 操作系统, 基于开源深度学习框架 PyTorch 1.12.1 构建网络模型, 硬件配置为 NVIDIA Geforce RTX 3090 GPU、24 GB 显存、Intel Xeon 银牌 4210R CPU, 软件环境为 Python 3.9.12 和 Cuda 11.3.

实验过程采用控制变量法, 确保所有参数保持一致. 优化算法选择随机梯度下降 (SGD), 共进行 300 轮迭代, 训练批尺寸设置为 64, 动量因子为 0.937, 初始学习率为 0.01, 权重衰减为 0.000 5, 训练时输入图像大小为 640×640. 在训练过程中使用早停机制, 设置 patience 为 50, 并在最后 10 轮迭代关闭了 Mosaic 增强, 确保模型在自然数据分布下完成收敛.

2.2 数据集

实验数据来自旷世南京研究所发布的大型零售商品数据集 RPC[22], 其包含 200 个商品类别, 总共 83 739 张图片, 分为单品图和结算图两种形式. 根据零售环境中商品摆放的杂乱程度将 30 000 张结算图分为 3 个级别, 以模拟真实的不同结算场景. 其中, 商品类别为 3–5 种、数量 3–10 个、摆放相对分散、发生轻微或无旋转变形且遮挡较小的图片为简单级别; 商品类别为 5–8 种、数量 10–15 个、摆放相对集中、发生中度旋转变形且存在一定遮挡的图片为中等级别; 商品类别为 8–10 种、数量 15–20 个、摆放非常集中、发生各种重度旋转变形且遮挡普遍的图片为困难级别,

各级别的图片数量占比为 1:1:1. 商品不同摆放杂乱级别示例如图 5 所示. 本研究选取所有结算图作为实验对象, 并按照 8:1:1 的比例, 随机将其划分为训练集, 验证集和测试集.

(a) 简单级别 (b) 中等级别 (c) 困难级别

图 5 商品不同摆放杂乱级别示例

2.3 评价指标

为了评估所提算法的检测效果, 使用均值平均精度 (*mAP*), 参数量 (Params), 计算量 (GFLOPs) 作为定量评价指标. *mAP*@0.5:0.95 表示 IoU 从 0.5 到 0.95, 步长为 0.05 时的平均精确度, 简称为 *mAP*, 计算公式如下:

$$\begin{cases} Precision = \dfrac{TP}{TP+FP} \\ Recall = \dfrac{TP}{TP+FN} \\ AP = \displaystyle\int_0^1 P(R)\,\mathrm{d}R \\ mAP = \dfrac{\displaystyle\sum_{i=1}^{N} AP_i}{N} \end{cases} \tag{11}$$

其中, $Precision$表示查准率 (简称P); $Recall$表示查全率 (简称R); TP表示预测正确的正样本; FP表示负样本预测为正样本; FN表示正样本为负样本; AP表示PR曲线的面积; N表示待检测样本的类别总数.

2.4 消融实验

为验证所提算法的有效性, 将各改进模块依次加入基准网络中, 在 RPC 数据集上展开消融实验分析. 其中, “√”表示引入该模块. NDC 表示归一化可变形卷积模块, DH 表示改进的动态头模块, IE 表示 InnerEIoU 损失函数. 使用 Params、GFLOPs 和 *mAP* 为定量评价指标, 实验结果如表 1 所示.

由表 1 实验结果可知, 基准模型 YOLOv8s 的 *mAP* 为 91.8%, 参数量为 10.69M, 计算量为 29.1G; 由实验 2 可知, 引入归一化可变形卷积模块后, *mAP* 提升 0.7%, 计算量减少 3.2G, 说明该模块在降低计算量的同时能够有效捕获旋转和变形商品的全局外观特征; 由实验 3 可知, 使用改进的动态头模块后, *mAP* 提升 1.0%, 模型参数量降低 12.3%, 说明该模块能在促进模型轻量化的同时有效捕获更具区分性的商品局部特征, 以抑制无关特征干扰; 由实验 4 可知, 模型引入 InnerEIoU 损失函数替换 CIoU 后, *mAP* 提升 0.4%, 说明借助辅助边界框宽和高的具体值来计算损失, 能够降低商品的漏检率; 由实验 5 可知, 当模型融合 3 种改进模块后, *mAP* 提升 1.5%, 参数量下降 1.07M, 计算量减少 1.9G, 说明三模块融合不仅能够轻量化模型并降低计算量, 还可以较好的提高模型检测精度, 验证了改进方法的有效性. 因此, 选用实验 5 为最终模型, 以下简称 YOLOv8-NDI (ours).

表 1 各改进模块的消融实验分析

实验	NDC	DH	IE	Params (M)	GFLOPs	*mAP* (%)
1	—	—	—	10.69	29.1	91.8
2	√	—	—	10.90	**25.9**	92.5
3	—	√	—	**9.38**	29.7	92.8
4	—	—	√	10.69	29.1	92.2
5	√	√	√	9.62	27.2	**93.3**

2.5 对比实验

(1) 主流检测模型效果对比

为进一步验证改进模型的优越性, 选取 YOLOv3-Tiny[23], YOLOv3, YOLOv6s[24], YOLOv7-Tiny[25], YOLOXs[26], 和 YOLOv9[27]等主流目标检测模型在 RPC 数据集上进行对比实验, 使用 Params、GFLOPs 和 *mAP* 为评价指标进行定量分析, 实验结果如表 2 所示.

表 2 主流检测模型的实验对比

模型	Params (M)	GFLOPs	*mAP* (%)
YOLO3-Tiny	8.71	13.7	86.5
YOLOv3	59.70	158.7	90.9
YOLOv5s	7.21	17.7	90.3
YOLOv6s	18.59	45.5	90.6
YOLOv7-Tiny	6.25	14.9	89.4
YOLOX-Tiny	**5.09**	**6.6**	88.2
YOLOXs	9.01	27.2	89.5
YOLOv8s	10.69	29.1	91.8
YOLOv9	7.93	31.3	92.5
YOLOv8-NDI (ours)	9.62	27.2	**93.3**

由表 2 实验结果可知, 改进模型相比于轻量化的 YOLOv3-Tiny、YOLOv7-Tiny、YOLOX-Tiny 模型检测精度分别提升 6.8%、3.9%、5.1%, 相比于 YOLOv3、YOLOv5s、YOLOv6s、YOLOXs、YOLOv9 的检测精度分别提升 2.4%、3.0%、2.7%、3.8%, 0.8%, 说明所提模型与主流模型相比具有一定的检测精度优势; 改进模型的参数量相比于 YOLOv3s 和 YOLOv6s, 分别减少 83.9% 和 48.3%, 但高于其余模型, 分析原因为引入了更复杂的卷积方式和网络结构; 所提模型的计算量相比于 YOLOv3、YOLOv6s 和 YOLOv9 分别下降 131.5G、18.3G 和 4.1G. 综上, 所提模型综合性能较优于其余主流检测算法, 证明了 YOLOv8-NDI (ours) 在实际应用中的优越性.

(2) 零售商品不同杂乱程度下模型效果对比

为了检验所提模型在实际结算场景下的实用性, 将改进模型与基线模型在包含简单 (easy)、中等 (medium)、困难 (hard) 这 3 种摆放杂乱程度下的 RPC 数据集中分别进行对比实验, 使用 *mAP* 为评价指标, 实验结果如表 3 所示.

表 3 零售商品不同杂乱程度的模型效果对比 (%)

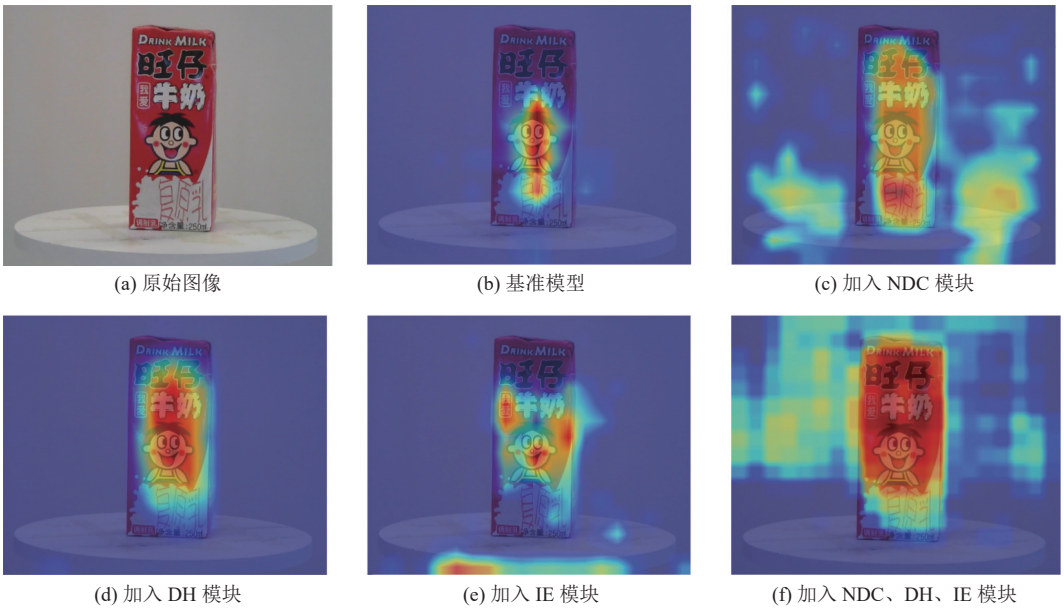
方法	Easy	Medium	Hard
YOLOv8s	94.9	90.0	88.8
YOLOv8-NDI (ours)	**95.9**	**91.2**	**90.5**

由表 3 可知, YOLOv8-NDI (ours) 针对简单、中等、困难 3 种杂乱程度的检测精度相较于基线模型分别提升 1.0%, 1.2%, 1.7%. 在不同杂乱程度下的检测效果均能得到有效提升, 表明模型在实际零售商品结算场景中具有较好的实用性.

2.6 模型可视化分析

(1) 各改进模块的热力图分析

为进一步验证改进方法的有效性, 将所提方法逐个加入模型中并借助热力图可视化出模型预测的判别性区域位置, 如图 6 所示.



(a) 原始图像 (b) 基准模型 (c) 加入 NDC 模块

(d) 加入 DH 模块 (e) 加入 IE 模块 (f) 加入 NDC、DH、IE 模块

图 6 嵌入各改进模块的热力图对比

热力图中红色高亮部分代表模型重点关注的预测类别和定位相关的区域, 该区域覆盖于商品判别性区域的面积越多, 越能捕获到区分商品的关键视觉信息, 对商品摆放杂乱环境的抗干扰力越强, 模型对商品的检测精度越高. 当加入归一化可变形卷积 (NDC) 模块后, 模型关注到商品的整体外观, 判别性区域覆盖更广,

说明充分捕获长距离依赖关系, 能够有效提取商品的全局特征; 当加入改进的动态头 (DH) 模块后, 模型更关注于商品本身, 同时背景信息被大量忽视, 说明有效提取更具区分性的商品判别性特征, 能够显著抑制无关特征干扰. 当采用 InnerEIoU(IE) 损失函数后, 红色高亮区域占据了商品边缘两侧以及中部区域, 锁定了商品的边界位置, 证明该函数能够增强对商品边框的定位能力; 在融合 3 种改进方法后, 高亮部分基本全面覆盖商品判别性区域, 验证了所提方法的有效性.

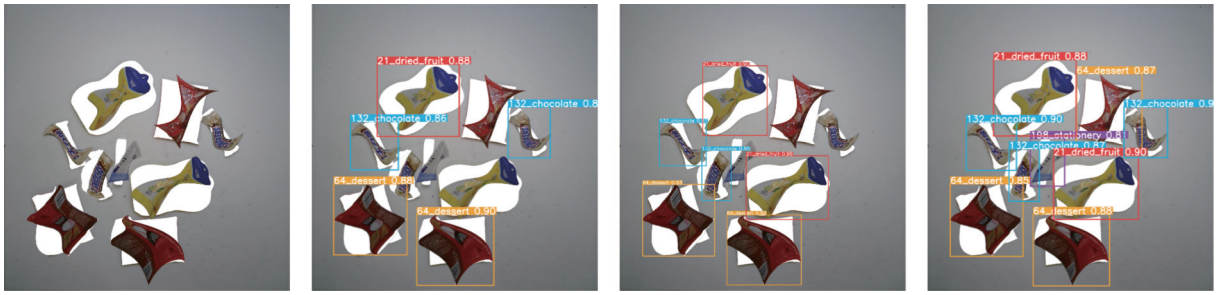
(2) 模型检测结果可视化分析

为了直观地展示模型对商品的检测结果, 将主流检测模型 YOLOX-Tiny、YOLOv5s、YOLOv8s、YOLOv9 与改进模型在简单、中等和困难 3 种难度级别的结算图中进行对比, 检测结果可视化如图 7 所示, 在简单级别中, 各模型均能检测出所有商品, 改进模型相较于基线模型对旋转商品 75_drink 和 181_tussye 的精度提升分别为 5% 和 3%; 在中等级别中, YOLOX-Tiny 模型出现两个商品漏检. 模型改进前后对商品 69_dessert 发生旋转或变形时的精度均提升 2%; 在困难级别中, YOLOX-Tiny 和 YOLOv5-s 模型均出现两个商品漏检, 模型改进前后对旋转商品 116_conned_food 和 117_conned_food 精度分别提升 4% 和 6%. 结果表明, 改进模型相较于除基准模型和 YOLOv9 以外的其余主流模型能够有效降低商品的漏检率, 与基准模型相比, 对旋转和变形商品具有更高的检测精度.

最后, 为进一步体现改进模型的优势, 将基准模型、YOLOv9 和改进模型在旋转角度更大且变形程度更严重的商品结算图中进行了对比, 检测效果如图 8 所示. 结果显示: 基准模型出现 4 个商品漏检, YOLOv9 出现 3 个商品漏检, 改进模型无漏检现象, 表明所提模型在显著降低商品漏检率的同时能够兼顾商品的旋转和变形现象, 具有明显的检测优势.

(a) 原始图像 (b) YOLOX-Tiny (c) YOLOv5s (d) YOLOv8s (e) YOLOv9 (f) YOLOv8-NDI

图 7 检测结果可视化对比图



(a) 增强后图像 (b) YOLOv8s (c) YOLOv9 (d) YOLOv8-NDI

图 8 旋转和变形程度更加严重的效果对比图

3 结论与展望

本文提出了基于变形卷积和多重注意力的零售商品检测模型 YOLOv8-NDI, 以解决模型因商品旋转和变形导致难以准确提取全局特征和无关特征干扰的问题. 引入归一化可变形卷积替代常规卷积, 通过精准捕获长距离依赖关系及突出通道显著特征, 增强对全局特征的提取能力; 引入改进的动态头, 通过提取更具区分性的商品局部特征, 以抑制无关特征干扰; 使用损失函数 InnerEIoU 以降低商品漏检率. 实验结果表明, 所提模型不仅能够显著提高对旋转和变形商品的检测精度, 还实现了模型轻量化, 并有效减少了计算量. 但针对遮挡程度较大和商品分布较密集的复杂场景, 模型适应能力不足, 将在后续研究中解决, 建立具有更强泛化能力和实时性的零售商品检测模型.

参考文献

1 Har LL, Rashid UK, Te Chuan L, *et al*. Revolution of retail industry: From perspective of retail 1.0 to 4.0. Procedia Computer Science, 2022, 200: 1615–1625.

2 Pham LH, Tran DNN, Nguyen HH, *et al*. DeepACO: A robust deep learning-based automatic checkout system. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans: IEEE, 2022. 3106–3113.

3 Juels A. RFID security and privacy: A research survey. IEEE Journal on Selected Areas in Communications, 2006, 24(2): 381–394. [doi: 10.1109/JSAC.2005.861395]

4 Franco A, Maltoni D, Papi S. Grocery product detection and recognition. Expert Systems with Applications, 2017, 81: 163–176. [doi: 10.1016/j.eswa.2017.02.050]

5 Wei YC, Tran S, Xu SX, *et al*. Deep learning for retail product recognition: Challenges and techniques. Computational Intelligence and Neuroscience, 2020, 2020: 8875910.

6 吕晓华, 魏铭辰, 刘立波. 基于位置可学习视觉中心机制的零售商品检测方法. 物联网学报, 2023, 7(4): 142–152. [doi: 10.11959/j.issn.2096-3750.2023.00366]

7 Li JX, Tang FQ, Zhu C, *et al*. BP-YOLO: A real-time product detection and shopping behaviors recognition model for intelligent unmanned vending machine. IEEE Access, 2024, 12: 21038–21051. [doi: 10.1109/ACCESS.2024.3361675]

8 Hu ZM, Zeng XP, Xie K, *et al*. Efficient defect detection of rotating goods under the background of intelligent retail. Sensors, 2024, 24(2): 467. [doi: 10.3390/s24020467]

9 Liu S, Qi L, Qin HF, *et al*. Path aggregation network for instance segmentation. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8759–8768.

10 Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning. Lille: PMLR, 2015. 448–456.

11 Liu ST, Huang D, Wang YH. Learning spatial fusion for single-shot object detection. arXiv:1911.09516, 2019.

12 Lin TY, Dollár P, Girshick R, *et al*. Feature pyramid networks for object detection. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 936–944.

13 张瑶, 陈姚节. 改进 YOLOv8 的水面小目标检测算法. 计算机系统应用, 2024, 33(4): 152–161. [doi: 10.15888/j.cnki.csa.009445]

14 Zheng ZH, Wang P, Ren DW, *et al*. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. IEEE Transactions on Cybernetics, 2022, 52(8): 8574–8586. [doi: 10.1109/TCYB.2021.3095305]

15 Zhu XZ, Hu H, Lin S, *et al*. Deformable ConvNets v2: More deformable, better results. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 9300–9308.

16 Liu YC, Shao ZR, Teng YY, *et al*. NAM: Normalization-based attention module. arXiv:2111.12419, 2021.

17 Dai XY, Chen YP, Xiao B, *et al*. Dynamic head: Unifying object detection heads with attentions. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 7369–7378.

18 Wang WH, Dai JF, Chen Z, *et al*. Internimage: Exploring large-scale vision foundation models with deformable convolutions. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 14408–14419.

19 Chen YP, Dai XY, Liu MC, *et al*. Dynamic ReLU. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 351–367.

20 郭伟, 王珠颖, 金海波. 高阶深度可分离无人机图像小目标检测算法. 计算机系统应用, 2024, 33(5): 144–153. [doi: 10.15888/j.cnki.csa.009471]

21 Zhang YF, Ren WQ, Zhang Z, *et al*. Focal and efficient IoU loss for accurate bounding box regression. Neurocomputing,

2022, 506: 146–157. [doi: 10.1016/j.neucom.2022.07.042]

22 Wei XS, Cui Q, Yang L, *et al*. RPC: A large-scale retail product checkout dataset. arXiv:1901.07249, 2019.

23 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.

24 Li CY, Li LL, Jiang HL, *et al*. YOLOv6: A single-stage object detection framework for industrial applications. arXiv: 2209.02976, 2022.

25 Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 7464–7475.

26 Ge Z, Liu ST, Wang F, *et al*. YOLOX: Exceeding YOLO series in 2021. arXiv:2107.08430, 2021.

27 Wang CY, Yeh IH, Liao HYM. YOLOv9: Learning what you want to learn using programmable gradient information. arXiv:2402.13616, 2024.

(校对责编: 孙君艳)