

基于改进 RT-DETR 的饼干包装外观缺陷快速检测

古莹奎 叶彪彪 郭明健 连增卫

(江西理工大学机电工程学院, 江西 赣州 341000)

摘要: [目的] 针对现有食品包装缺陷实时检测模型参数量庞大和计算复杂、在终端设备上部署困难等问题, 提出一种基于改进 RT-DETR 的轻量化模型 SGHS-DETR。[方法] 采用超轻量化网络 StarNet 作为特征提取主干以降低模型参数, 引入基于梯度路径规划的高效聚合模块 GELAN 进行特征融合并保留语义和细节特征; 采用基于小波分解的 HWD 轻量级下采样模块减少特征信息损失, 并替换损失函数为 ShapeIoU 进一步提升模型的检测精度。[结果] SGHS-DETR 模型在饼干包装数据集上的平均检测精度达到 92.6%, 较基准模型参数量和计算复杂度分别降低了 65.5% 和 72.1%, 同时检测速度提升了 74.4%。[结论] SGHS-DETR 模型能够快速有效地检测出饼干包装外观缺陷。

关键词: 食品包装; 缺陷检测; 轻量化; RT-DETR; StarNet

Rapid detection method of biscuit packaging appearance defects based on improved RT-DETR

GU Yingkui YE Biaobiao GUO Mingjian LIAN Zengwei

(School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China)

Abstract: [Objective] To address the problems of a large number of parameters and complex calculation of the current real-time detection models for food packaging defects and their difficult deployment on terminal equipment, this study proposes a lightweight model SGHS-DETR based on improved real-time detection transformer (RT-DETR). [Methods] To reduce model parameters, the ultra-lightweight network StarNet is employed as the feature extraction backbone. Additionally, GELAN, an efficient aggregation module based on gradient path planning, is introduced for feature fusion and preservation of semantic and detail features. Furthermore, the lightweight Haar wavelet downsampling (HWD) module based on wavelet decomposition is adopted to minimize information loss in features. Moreover, ShapeIoU replaces the loss function to further enhance detection accuracy. [Results] On the biscuit packaging dataset, the average detection accuracy of the SGHS-DETR model reaches 92.6%. Compared to the baseline model, this approach reduces the number of parameters and computational complexity by 65.5% and 72.1%, respectively and also increases detection speed by 74.4%. [Conclusion] The SGHS-DETR model can rapidly and effectively detect appearance defects in biscuit packaging.

Keywords: food packaging; defect detection; lightweight; real-time detection transformer; StarNet

食品包装作为食品工业的重要组成部分, 包装质量直接关系到食品的安全。受生产机械的影响, 食品包装不可避免地会出现破损、脏污等外观缺陷。较早期的人工检测, 基于机器视觉的包装外观缺陷检测已成为主流的解决方案^[1]。通过工业相机与图像处理算法, 机器视觉系统能够高效准确地识别出包装表面的各种缺陷, 可有效促进食品企业的自动化转型^[2]。

早期应用于食品外观缺陷检测的方法主要是传统的图像处理算法, 如自适应阈值分割算法、分水岭算法、模板匹配算法等^[3-5]。机器学习方法的引入克服了传统图像处理方法检测单一性的问题, 张宝胜等^[6]利用支持向量机分类器对提取的透明包装封口纹理特征进行缺陷分类, 检测结果与人工检测结果同一性为 97.5%。Xiong 等^[7]提出基于水平集映射的方法, 利用图像灰度值表示包

基金项目: 国家自然科学基金资助(编号: 61963018); 江西省自然科学基金重点基金资助(编号: 20212ACB202004)

通信作者: 古莹奎(1976—), 男, 江西理工大学教授, 博士。E-mail: guyingkui@163.com

收稿日期: 2024-07-28 **改回日期:** 2025-02-15

装轮廓和缺陷位置信息,实现了饼干缺陷包装的检测。但是,基于机器视觉的检测方法仍存在着识别种类有限、特征设计依赖人工和调试难度大等局限性。

随着深度学习的发展,基于卷积神经网络(CNN)架构的深度学习已经成为工业外观缺陷检测的主流方法^[8]。夏军勇等^[9]利用改进后的 Faster-R-CNN 对包装纸箱缺陷进行检测,试验准确率达到 93.9%,但是检测速度仅有 8.65 帧/s。为实现包装缺陷的快速识别,陈雪纯等^[10]提出一种改进 MobilenetV2 的轻量化网络快速检测方法,实现 18.87 帧/s 的识别速度。吴昊然等^[11]将 SE-Net 与 YOLOv5 的主干网络进行融合用于薯片包装上的缺陷检测,达到 94.8% 的检测精度并大幅提升了检测速度。尽管上述方法实现了较高的检测精度,但均针对尺度不变的缺陷进行检测,而对于跨尺度缺陷并未进行充分研究。此外,基于 CNN 架构的检测模型需要利用非极大值抑制(NMS)去除冗余的预测框,这种后处理操作一定程度上影响了模型的检测性能。

为提高模型的检测性能以及实现端到端的检测,Zhao 等^[12]提出了基于 Transformer 的实时目标检测模型(real-time detection transformer, RT-DETR),利用 Transformer 架构强大的全局特征表征学习能力,并行预测包含目标和背景在内的多个对象的类别、位置以及边框信息。该模型摒弃了 NMS 后处理操作,在收敛和推理速度方面更具优势。但是该模型仍然存在参数量庞大、硬件算力要求高、实际部署困难等问题。

针对上述问题,结合目前饼干包装外观缺陷尺度大导致检测精度、效率低等难题,研究拟提出一种基于改进 RT-DETR 的轻量化模型——SGHS-DETR,以期通过较少的计算资源实现较高的外观缺陷包装检测精度和准确率。

1 硬件设计和检测数据集制备

1.1 外观缺陷检测设备与检测流程

为实现饼干包装的智能检测,设计基于机器视觉的饼干包装外观缺陷检测装置如图 1 所示。检测设备包括运输系统、图像采集系统和次品剔除系统。图像采集系统包括工业相机、平面光源、光电传感器和工控机等。运输系统利用传送带运输小包饼干,并设计翻转装置将小包饼干翻转,以便拍摄饼干包装反面的图像。剔除系统利用高速气流将缺陷包装吹离传送带。

1.2 图像采集与数据集制作

根据企业生产标准,将外观缺陷分为脏污、跑袋、压边、皱边、异色、纵封变形、白边和纵封不良 8 种缺陷类型。其中,脏污缺陷表现为包装表面上直径 $> 1 \text{ mm}$ 的饼干碎屑等污染物;跑袋缺陷表现为白色锯齿切口处存在绿色

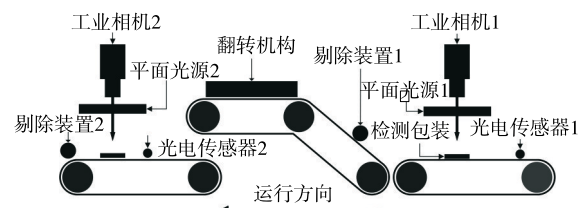


图 1 饼干包装外观缺陷智能检测设备

Figure 1 Intelligent detection equipment for biscuit packaging appearance defects

印刷。值得注意的是,同一个包装产品上可能同时出现多种缺陷,此外,同一类缺陷的外观和大小等特征也可能不同。

使用 130 万像素面阵工业相机,拍摄的样本图片分辨率为 1280×1024 。利用外观缺陷检测设备采集饼干包装的样本图片用于制作数据集,由于实际饼干包装生产线出现外观缺陷的概率很小,经过筛选得到 3 110 张样本图片。其中合格样本 1 200 张,外观缺陷图片共 1 910 张,外观缺陷标签数量 3 721 个,平均一个包装上约 2 个缺陷。

2 改进的 RT-DETR 缺陷检测方法

2.1 RT-DETR 基准模型

RT-DETR 是一款端到端的实时目标检测模型,通过 Vision Transformers (ViT) 对多尺度特征进行高效处理,从而在保持高精度的同时提供实时检测性能。RT-DETR 的 backbone 网络仍然采用 CNN 架构,如目前主流的 ResNet 系列。RT-DETR 采用由尺度内特征交互模块(AIFI)和跨尺度特征融合模块(CCFM)组成的高效的混合编码器对多尺度特征进行融合。同时,将基于自注意机制的 AIFI 模块应用于具有更丰富语义概念的高级特征,以捕获图像中概念实体之间的联系,以有助于后续模块检测和识别图像中的对象,促进图像中对象的检测和识别。RT-DETR 的解码器部分利用多层 Transformer 解码器,允许在推理期间灵活地选择不同数量的解码器层,从而自适应地调整推理速度而无需重新训练。RT-DETR 的网络结构如图 2 所示。

2.2 改进 RT-DETR 检测模型的搭建

尽管 RT-DETR 基准模型具有实时检测的性能,但在一些计算资源有限的终端检测设备中仍然存在参数量庞大,难以部署的问题。对 RT-DETR 基准模型做如下改进:通过替换超轻量化 StarNet^[13]为主干网络,引入融合多尺度特征的广义高效层聚合网络 GELAN^[14]和基于 Harr 小波的轻量级下采样 HWD^[15]模块,最后替换损失函数为 ShapeIoU^[16],降低模型的计算成本,便于部署在计算能力不高的终端检测设备中。

2.2.1 StarNet 超轻量化网络 使用 StarNet 超轻量化网

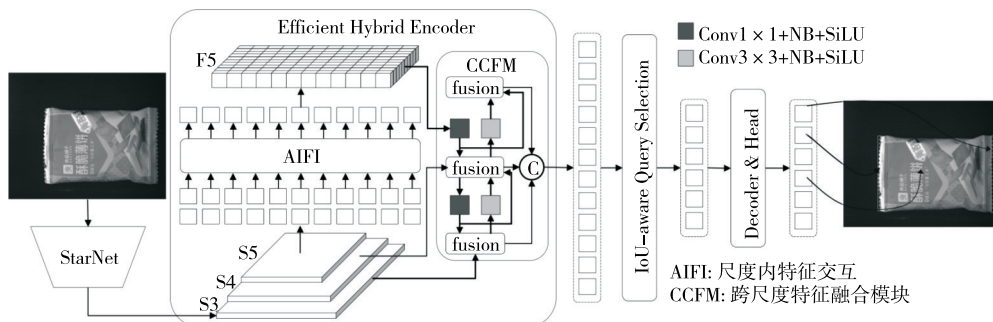


图 2 RT-DETR 的网络结构

Figure 2 Network architecture of RT-DETR

络作为 SGHS-DETR 模型的骨干网络来降低模型的复杂度,减轻计算成本。StarNet 采用 4 级分层架构,利用卷积层进行下采样,并通过 Star Blocks 进行特征提取。StarNet 网络采用的是基于逐元素乘法融合不同子空间特

征的学习范式(star operation, 星运算),没有任何复杂的结构和人为设计的超参数。尽管模型整体相对简单,但却表现出强大的特征提取能力。StarNet 的网络架构如图 3 所示。

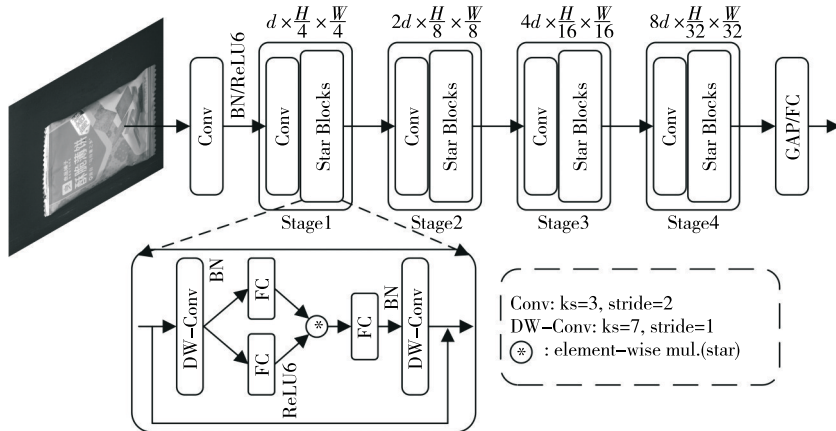


图 3 StarNet 网络整体架构

Figure 3 Network architecture of StarNet

Star Blocks 仅由简单的卷积层和全连接层组成,其核心操作 Star Operation 具有将输入映射到极高维的非线性特征空间的能力。在单层神经网络中,Star Operation 可表示为 $(W_1^T X) * (W_2^T X)$, 其中, $*$ 代表 Star Operation 操作, W_1 和 W_2 为模型权重, X 为输入特征,通过逐元素乘法融合两个线性变换的特征。以单输出通道转换和单元输入场景为例,定义 $w_1, w_2, x \in \mathbb{R}^{(d+1) \times 1}$, $\mathbb{R}^{(d+1) \times 1}$ 表示形状为 $(d+1) \times 1$ 的实数矩阵。这可以随时进行 $W_1, W_2 \in \mathbb{R}^{(d+1) \times (d+1)}$ 扩展以适应多个输出通道,以及处理 n 通道多元素输入 $X \in \mathbb{R}^{(d+1) \times n}$ 。

一般情况下,可以通过以下方式重新表述 Star Operation:

$$w_1^T x * w_2^T x = \left(\sum_{i=1}^{d+1} w_1^i x^i \right) * \left(\sum_{j=1}^{d+1} w_2^j x^j \right) = \sum_{i=1}^{d+1} \sum_{j=1}^{d+1} w_1^i w_2^j x^i x^j =$$

$$\underbrace{\alpha_{(1,1)} x^1 x^1 + \dots + \alpha_{(4,5)} x^4 x^5 + \dots + \alpha_{(d+1,d+1)} x^{d+1} x^{d+1}}_{(d+2)(d+1)/2 \text{ items}}, \quad (1)$$

式中:

i, j ——通道下标;

d ——输入通道数;

x ——输入矩阵;

w_1, w_2 ——单通道模型权重;

α ——子项的系数。

$$\alpha_{(i,j)} = \begin{cases} w_1^i w_2^j & i=j \\ w_1^i w_2^j + w_1^j w_2^i & i \neq j \end{cases} \quad (2)$$

重新表述的 Star Operation, 可以将其展开为 $(d+2)(d+1)/2$ 个不同子项的组合如式(1)所示。因此, 在 d 维度空间中使用 Star Operation, 可以得到

$\frac{(d+2)(d+1)}{2} \approx \left(\frac{d}{\sqrt{2}}\right)^2$ ($d \gg 2$) 的隐式维度特征空间。

从而在显著放大特征维度的同时,不会在单层内产生任何额外的计算开销。通过堆叠多个 Star Operation,可以将隐式维度以指数方式增加到几乎无限。对于宽度为 d 的初始网络层,应用一次 Star Operation

$\sum_{i=1}^{d+1} \sum_{j=1}^{d+1} w_1^i w_2^j x^i x^j$, 可以得到 $\mathbb{R}^{\left(\frac{d}{\sqrt{2}}\right)^2}$ 的隐式特征空间。

当堆叠多个 Star Operation 时,利用 O_l 表示第 l 个 Star Operation 的输出,可得:

$$\begin{cases} O_1 = \sum_{i=1}^{d+1} \sum_{j=1}^{d+1} w_{(1,1)}^i w_{(1,2)}^j x^i x^j \\ O_2 = W_{2,1}^T O_1 * W_{2,2}^T O_1 \\ O_3 = W_{3,1}^T O_2 * W_{3,2}^T O_2 \\ \vdots \\ O_l = W_{l,1}^T O_{l-1} * W_{l,2}^T O_{l-1} \end{cases} \quad (3)$$

式中:

$w_{(1,1)}^i, w_{(1,2)}^j$ ——第一层网络的权重参数,分别作用 x^i 和 x^j ;

$W_{l,1}^T, W_{l,2}^T$ ——第 l 层的权重矩阵。

通过堆叠 l 层可以获得 $\mathbb{R}^{\left(\frac{d}{\sqrt{2}}\right)^2}$ 维的特征空间,通过堆叠多个层,Star Operation 可以以指数方式大幅放大隐式维度。鉴于 Star Operation 在低维空间中计算的同时产生高维特征的独特优势,运用 Star Operation 方法建立的 StarNet 主干网络拥有极少的参数量,可实现较高的检测精度。

2.2.2 GELAN 模块 引入一种基于梯度路径规划的广义高效层聚合网络(GELAN)架构,进一步增强模型的特征融合能力。GELAN 是由跨阶段部分网络^[17](cross stage partial network, CSPNet)和高效层聚合网络^[18](efficient layer aggregation network, ELAN)组合设计的网络框架,可以使用任何计算模块,将 ELAN 的能力进行泛化,使网络更加灵活。GELAN 通过利用特征空间维度层在不同抽象层次提取特征的能力,既可以获得高级语义信息,也可以获得基本细节信息。

RepNCSPELAN4(Rep-Net with CSP and ELAN)是一种基于 GELAN 架构的特征融合模块,其网络结构如图 4 所示。它将输入分成两个部分,用多个 RepNCSP 模块分别处理每个部分,将输出连接起来,然后将它们传递给最终的卷积层。跨阶段部分连接 CSP 促进了网络不同阶段之间的信息流动,ELAN 通过注意机制增强特征表征。RepNCSPELAN4 通过将高效的特征提取与注意机制相结合来增强特征表示,它使网络能够捕捉和强调与目标检测任务相关的重要特征。这些架构组件旨在有效地管理

输入属性,同时保留准确目标检测所需的空间和语义信息。在不同的网络层应用上采样技术来提高特征图的空间分辨率。通过将上采样特征图与前一层的特征图连接起来,便于多尺度数据集成,这使得模型可以保留细粒度特征和空间连接,从而提高目标定位和识别。

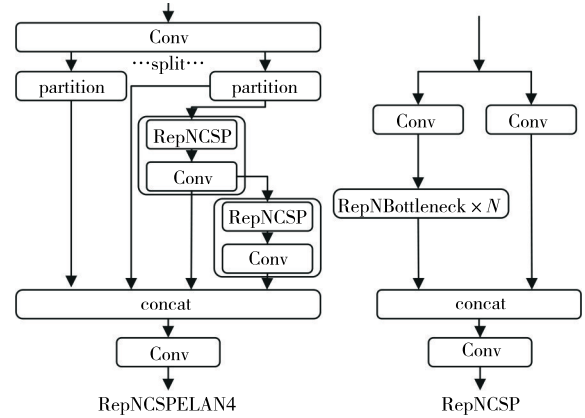


图 4 RepNCSPELAN4 网络结构

Figure 4 Network structure of RepNCSPELAN4

2.2.3 小波下采样(haar wavelet downsampling, HWD)模块 为了能尽可能多地保留特征图的信息,在模型中引入 HWD 下采样模块替换 Neck 中的普通卷积下采样模块。HWD 的核心思想是利用 Haar 小波变换在尽可能多地保留特征信息的同时降低特征图的空间分辨率,HWD 结构如图 5 所示。

HWD 模块包括两个子模块:无损特征编码模块和特征学习模块。无损特征编码模块利用 Haar 小波变换层来有效地降低特征图的空间分辨率,Haar 小波变换可以将空间维度的部分信息编码到信道维度,而不会造成任何信息损失。特征学习块由标准卷积层、批量归一化和 ReLU 激活层组成,用于调整通道数和过滤冗余信息。以高为 H ,宽为 W 和通道数为 C 的特征图作为输入,输入尺寸为 $H \times W \times C$,输出尺寸 $\frac{H}{2} \times \frac{W}{2} \times C$,相比传统 3×3 下采样卷积,HWD 参数量更少。此外,当通道数 $C > 1$ 时,HWD 的计算量小于下采样卷积计算量,两种下采样方法的参数量和计算量对比如表 1 所示。由此可知,HWD 下采样方法在一定程度上加快了模型的推理速度,同时保留更多细节提升了模型的检测性能。

2.2.4 ShapeIoU 损失函数 边界框回归损失函数在目标检测的定位中发挥着重要作用,RT-DETR 使用的是 GIoU 损失函数,通过引入最小封闭框对缺陷进行定位。尽管这能很好地反映边界的相交情况,却忽略了边界框本身的形状和尺度也会对边界框回归产生影响。由于饼干包装外观缺陷标注框的形状和尺寸跨度较大,为了进一步

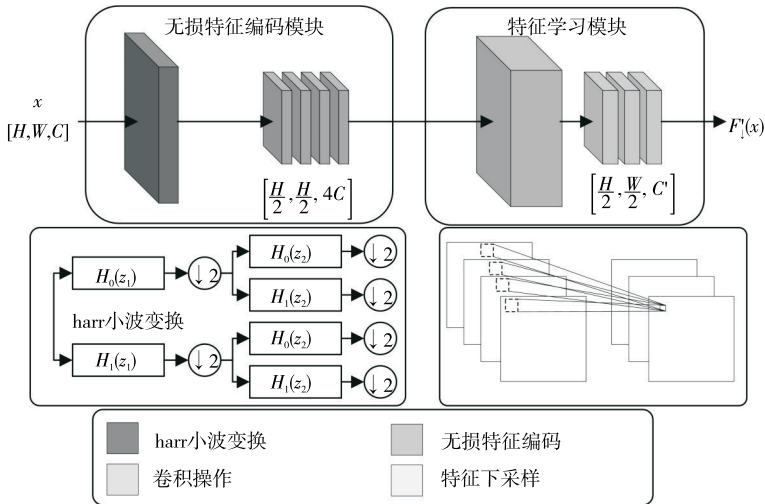


图 5 HWD 结构示意图

Figure 5 Structure diagram of HWD

表 1 对比两种下采样方法的参数量和计算量

Table 1 Comparison of parameters and giga floating-point operations per second (GFLOPs) between two downsampling methods

模块	参数量	计算量
HWD	$4C^2$	$2HWC^2 + 3.75HWC$
下采样卷积	$9C^2$	$4HWC^2 - 0.25HWC$

提高回归的精度,引入考虑边界框回归样本自身的形状和尺寸对边界框回归影响的 ShapeIoU 损失函数替换 RT-DETR 原有的损失函数。Shape-IoU 损失函数为:

$$L_{\text{ShapeIoU}} = 1 - \text{IoU} + D_{\text{shape}} + 0.5 \times V_{\text{shape}}, \quad (4)$$

式中:

IoU ——交并比;

D_{shape} ——形状距离系数;

V_{shape} ——形状差异系数。

D_{shape} 系数用于衡量预测框和真实框的中心点距离:

$$D_{\text{shape}} = hh \times (x_c - x_c^{gt})^2 / c^2 + ww \times (y_c - y_c^{gt})^2 / c^2, \quad (5)$$

$$\begin{cases} ww = \frac{2 \times (w^{gt})^s}{(w^{gt})^s + (h^{gt})^s} \\ hh = \frac{2 \times (h^{gt})^s}{(w^{gt})^s + (h^{gt})^s} \end{cases}, \quad (6)$$

式中:

ww, hh ——水平方向和垂直方向上的权重系数(与真实标注框形状有关);

w^{gt}, h^{gt} ——真实标注框的像素宽和高;

s ——缩放因子(与数据集中目标大小相关);

c ——真实标注框和预测框之间最小检测框的像素斜边长;

$(x_c^{gt}, y_c^{gt}), (x_c, y_c)$ ——真实标注框和预测框重心点的坐标。

V_{shape} 系数用于衡量预测框与真实框在宽度和高度上的差异:

$$V_{\text{shape}} = \sum_{t=w, h} (1 - e^{-\omega_t})^\theta, \theta = 4, \quad (7)$$

式中:

ω_w, ω_h ——预测框宽度和高度的损失权重系数;

θ ——损失增长指数。

$$\begin{cases} \omega_w = hh \times \frac{|w - w^{gt}|}{\max(w, w^{gt})} \\ \omega_h = ww \times \frac{|h - h^{gt}|}{\max(h, h^{gt})} \end{cases}, \quad (8)$$

式中:

w, h ——预测框的像素宽和高。

3 结果与讨论

3.1 试验设置

试验设备如图 1 所示,所提方法基于 Python3.8.16 和 Pytorch1.13.1 框架实现。输入图像来自于工业相机拍摄像素尺寸 1280×1024 的 RGB 图片,经过模型 resize 为大小 $640 \times 640 \times 3$,在单张高性能 NVIDIA GeForce RTX 4060Ti 显卡上训练。采用 Adamw 优化器进行优化,模型的训练周期(epoch)为 100,批量大小(batch size)为 8、初始学习率为 0.000 1,使用余弦退火策略动态调整学习率。

采用平均精度均值(mean average precision, mAP)评估模型在实际检测中的平均检测精度;使用参数量(parameters)、计算量(GFLOPs)和模型权重(weight)衡量

模型的轻量化效果;使用帧率(frames per second,FPS)计算模型的实时检测性能。

3.2 消融试验

为评估各改进模块对模型检测性能的影响,通过消融试验对比改进前后模型的检测性能,将改进模块逐个添加至基准 RT-DETR 模型中进行训练,结果如表 2 所示。

表 2 消融试验
Table 2 Ablation test

模型	mAP/ %	参数 量/M	计算 量/G	模型权 重/MB	速度/ s ⁻¹
RT-DETR-r18	91.1	19.88	57.0	40.5	88.0
RT-DETR-StarNet	91.6	8.93	24.6	18.5	148.2
RT-DETR-StarNet- GELAN	92.1	7.54	17.3	15.4	136.6
RT-DETR-StarNet- GELAN-HWD	92.3	6.88	15.9	14.1	145.3
SGHS-DETR	92.6	6.88	15.9	14.1	153.5

由表 2 可知,以 ResNet18 网络为主干的 RT-DETR-r18 基准模型的 mAP 为 91.1%,采用 StarNet 超轻量化主干网络之后,RT-DETR-StarNet 模型的 mAP 提升了 0.5%,同时模型的参数量和计算量都下降了 50% 以上,模型权重大小下降了 62%。原因在于 StarNet 采用的 Block 中运用 Star Operation 来融合不同子空间的特征,Star Operation 具有将输入映射到极高维度的非线性特征空间的能力,StarNet 仅利用极少的参数量和计算量便可以实现较好的特征提取效果。在 RT-DETR-StarNet-GELAN 模型中引

入 GELAN 模块后,mAP 提高了 0.5%,参数量和计算量分别减少了 1.39 M 和 7.3 G,模型权重减少了 3.1 MB。这是因为 GELAN 模块通过精心设计的梯度路径,优化网络结构,网络可以更有效地聚合来自不同层级的特征信息,同时减少模块的计算量和计算参数。进一步引入 HWD 轻量化下采样模块后,模型的 mAP 提高了 0.2%,参数量和计算量分别减少了 0.66 M 和 1.4 MB,模型权重减少了 1.3 MB。原因在于 HWD 下采样模块利用 Haar 变换在降低特征图的空间分辨率的同时尽可能多地保留特征图信息,并且使用 1×1 卷积层调整通道数,相比于传统的 3×3 卷积层减少了计算量。在 RT-DETR-StarNet-GELAN-HWD 模型基础上更改损失函数为 ShapeIoU 后得到 SGHS-DETR,模型的 mAP 提高了 0.3%,未引入额外参数。原因在于 ShapeIoU 通过关注边界框自身形状和尺寸来计算损失,使边界框回归更准确,可以有效提高检测性能。

模型改进前后对各类缺陷的检测精度如表 3 所示。尽管 RT-DETR-r18 基准模型的整体检测精度较高,但是对于脏污的检测效果不理想。分析原因可能是由于 RT-DETR 的特殊结构,考虑检测的实时性能,RT-DETR 仅使用了一个 Transformer 的 Encoder 来处理主干提取的最后一层特征。这意味着 Transformer 的输入只接受来自网络中最底层的级语义信息,而忽略了浅层特征中的局部细节信息,这些信息对于小目标的检测尤为重要。由于脏污像素尺寸较小,RT-DETR 模型对于此类目标的检测显得不尽人意。SGHS-DETR 改进模型对于脏污缺陷的检测精度提升至 66.0%,并在 70 个脏污缺陷验证图片中,改进模型的检出率由基准模型的 74% 提升至 76%。

表 3 改进前后模型对各类缺陷检测精度对比

Table 3 Comparison of detection accuracy of various defects before and after model improvement

模型	AP50/%							
	脏污	错位	压边	皱边	异色	纵封变形	白边	纵封不良
RT-DETR-r18	57.6	88.4	89.3	97.6	99.5	98.6	98.1	99.5
SGHS-DETR	66.0	90.0	93.1	98.5	99.5	98.5	95.8	99.5

SGH-DETR 改进模型的参数量、计算量和模型权重分别为基准模型的 34.6%,27.9%,34.8%,轻量化效果明显,而且 mAP 提升了 1.5%,对于脏污的检测精度提高了 8.4%。同时,SGHS-DETR 改进模型的检测速度达到 153.5 帧/s,比基准模型提速了 74.4%,完全满足实际生产线的检测速度需求。

3.3 不同算法性能对比试验

为了进一步测试模型的性能,将常见的目标检测模型与改进的 RT-DETR 模型进行对比,结果如表 4 所示。

由表 4 可知,SGHS-DETR 模型在 mAP、参数量、计算

表 4 不同模型饼干包装缺陷检测结果对比

Table 4 Comparison of detection results of biscuit packaging defects among different models

模型	mAP/%	参数量/M	计算量/G	模型权重/MB
YOLOv3-tiny	85.7	12.10	18.9	22.4
YOLOv5s	89.2	9.12	23.8	18.5
YOLOv8s	90.9	11.12	28.5	22.5
RT-DETR	91.1	19.88	57.0	40.5
SGHS-DETR	92.6	6.88	15.9	14.1

量和模型权重上均实现了最优性能。SGHS-DETR 模型的精度为 92.6%, 对比其他模型, 分别高出 1.5%, 6.9%, 3.4%, 1.7%, 说明改进模型在饼干包装外观缺陷检测方面具有较高的准确性和可靠性。同时, 在参数量方面, SGHS-DETR 模型实现了 6.88 M 的参数量, 比基准模型减小了 13.00 M 的参数量; 在计算量和模型权重方面也都远小于其他模型, 表明 SGHS-DETR 模型在拥有较高精度的前提下具有体积小、计算成本低的优势, 适合部署在计算资源有限的饼干包装缺陷检测设备上。如图 6 所示, 相较于 YOLO 系列的模型, SGHS-DETR 模型收敛速度更快、精度更高。

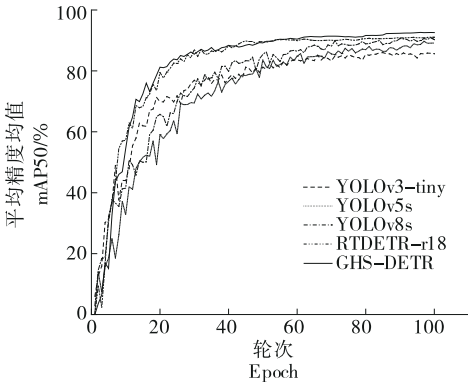


图 6 SGHS-DETR 模型与常见模型的 mAP 对比

Figure 6 Comparison of mAP between the proposed model and common models

3.4 其他轻量化网络对比

以 RT-DETR-r18 目标检测网络为基础网络, 替换主流的轻量化特征提取网络, 如 EfficientViT^[19]、FasterNet^[20]、MobileNetV4^[21]、GhostNetV2^[22], 在保持除主干网络外其他参数一致的基础上, 对比不同主干网络对于目标训练效果, 结果见表 5。

表 5 不同轻量化特征提取对比

Table 5 Comparison of feature extraction among different lightweight backbones

主干网络	mAP/%	参数量/M	计算量/G	模型权重/MB
EfficientViT	91.4	10.71	27.3	22.7
FasterNet	91.3	10.82	28.5	22.3
MobileNetV4	91.8	11.32	39.5	23.4
GhostNetV2	90.7	9.78	24.9	20.4
StarNet	91.6	8.93	24.6	18.5

由表 5 可知, 使用 StarNet 主干网络相较 EfficientViT、FasterNet、MobileNetV4、GhostNetV2, 模型的平均精度虽然相差较小, 但是模型的参数量、计算量和网络权重大小

都远大于 StarNet 主干网络。在特征提取能力相似的情况下, 选择更为轻量化的 StarNet 作为改进模型的主干网络。

4 结论

对比基于 Transformer 的实时目标检测基准模型, 基于改进 RT-DETR 的轻量化模型对饼干包装外观缺陷检测的平均精度提升了 1.5%, 对于脏污这类小目标的检测精度提高了 8.4%。模型的参数量、计算量和模型权重分别降低至基准模型的 34.6%, 27.9%, 34.8%, 检测速度提高了 74.4%; 对比目前主流的目标检测模型, 基于改进 RT-DETR 的轻量化模型在检测精度、检测速度、模型轻量化程度上均处于领先状态。研究提出的基于改进 RT-DETR 的轻量化模型对饼干包装外观缺陷具有较好的检测结果, 可以满足实际应用。但该模型对于小尺寸的脏污检测效果仍然不佳, 检测精度较低, 后续将围绕提高小目标检测精度继续对模型进行优化。

参考文献

[1] 戈明辉, 张俊, 陆慧娟. 基于机器视觉的食品外包装缺陷检测算法研究进展[J]. 食品与机械, 2023, 39(9): 95-102, 116.
GE M H, ZHANG J, LU H J. Research progress of food packaging defect detection based on machine vision[J]. Food & Machinery, 2023, 39(9): 95-102, 116.
[2] XIAO Z F, WANG J L, HAN L, et al. Application of machine vision system in food detection[J]. Frontiers in Nutrition, 2022, 9: 888245.
[3] 李丹, 白国君, 金媛媛, 等. 基于机器视觉的包装袋缺陷检测算法研究与应用[J]. 激光与光电子学进展, 2019, 56(9): 188-194.
LI D, BAI G J, JIN Y Y, et al. Machine-vision based defect detection algorithm for packaging bags[J]. Laser & Optoelectronics Progress, 2019, 56(9): 188-194.
[4] HE Y F, FAN B J, SUN L, et al. Rapid appearance quality of rice based on machine vision and convolutional neural network research on automatic detection system[J]. Frontiers in Plant Science, 2023, 14: 1190591.
[5] 肖斌, 杨涛, 梅艳莹. 基于分块特征匹配的纸箱缺陷检测方法[J]. 传感器与微系统, 2020, 39(9): 138-141.
XIAO B, YANG T, MEI Y Y. Defect detection method of carton based on block feature matching[J]. Transducer and Microsystem Technologies, 2020, 39(9): 138-141.
[6] 张宝胜, 周聪玲, 王永强. 基于机器视觉的透明包装袋真空封口纹理缺陷检测方法[J]. 食品与机械, 2023, 39(7): 111-118.
ZHANG B S, ZHOU C L, WANG Y Q. Method for detecting texture defect of vacuum seal of transparent packaging bag based on machine vision[J]. Food & Machinery, 2023, 39(7): 111-118.

- [7] XIONG K K, LI W S, DONG S, et al. Defect detection of biscuit packaging based on level set map[C]// 2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). Beijing: IEEE, 2022: 1-7.
- [8] LIU Y, ZHANG C S, DONG X J. A survey of real-time surface defect inspection methods based on deep learning[J]. Artificial Intelligence Review, 2023, 56(10): 12 131-12 170.
- [9] 夏军勇, 王康宇, 周宏娣. 基于改进 Faster R-CNN 的食品包装缺陷检测[J]. 食品与机械, 2023, 39(11): 131-136.
XIA J Y, WANG K Y, ZHOU H D. Food packaging defect detection by improved network model of Faster R-CNN[J]. Food & Machinery, 2023, 39(11): 131-136.
- [10] 陈雪纯, 方宇伦, 杜世昌, 等. 基于深度学习的包装缺陷快速检测方法[J]. 机械设计与研究, 2021, 37(6): 165-169.
CHEN X C, FANG Y L, DU S C, et al. Rapid packaging defect detection method based on deep learning[J]. Machine Design & Research, 2021, 37(6): 165-169.
- [11] 吴昊然, 陈晓星, 高傲. 基于深度卷积神经网络的食品包装缺陷检测算法研究[J]. 智能计算机与应用, 2023, 13(3): 10-15.
WU H R, CHEN X X, GAO A. Research on food packaging defect detection algorithm based on deep convolutional neural network[J]. Intelligent Computer and Applications, 2023, 13(3): 10-15.
- [12] ZHAO Y A, LV W Y, XU S L, et al. Detrs beat yolos on real-time object detection[C]// 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 16 965-16 974.
- [13] MA X, DAI X Y, BAI Y, et al. Rewrite the Stars[C]// 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 5 694-5 703.
- [14] WANG C Y, YE H I H, LIAO H M. YOLOv9: learning what you want to learn using programmable gradient information[J/OL]. arXiv. (2024-02-21) [2024-07-31]. <https://arxiv.org/abs/2402.13616>.
- [15] XU G P, LIAO W T, ZHANG X, et al. Haar wavelet downsampling: a simple but effective downsampling module for semantic segmentation[J]. Pattern Recognition, 2023, 143: 109819.
- [16] ZHANG H, ZHANG S J. Shape-IoU: more accurate metric considering bounding box shape and scale[J/OL]. ArXiv. (2024-01-12) [2024-07-31]. <https://arxiv.org/abs/2312.17663>.
- [17] CHEN X Q, YANG C Z, MO J, et al. CSPNeXt: a new efficient token hybrid backbone[J]. Engineering Applications of Artificial Intelligence, 2024, 132: 107886.
- [18] WANG C Y, LIAO H M, YE H I H. Designing network design strategies through gradient path analysis[J]. Journal of Information Science and Engineering, 2023, 39(2): 975-995.
- [19] LIU X Y, PENG H W, ZHENG N X, et al. EfficientViT: memory efficient vision transformer with cascaded group attention[C]// 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023: 14 420-14 430.
- [20] CHEN J R, KAO S H, 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). Vancouver: IEEE, 2023: 12 021-12 031.
- [21] QIN D F, LEICHTNER C, DELAKIS M, et al. MobileNetV4: universal models for the mobile ecosystem[C]// European Conference on Computer Vision. Shanghai: Springer, 2025: 78-96.
- [22] TANG Y H, HAN K, GUO J Y. GhostNetV2: enhance cheap operation with long-range attention[J]. Advances in Neural Information Processing Systems, 2022, 35: 9 969-9 982.