

# 基于改进深度学习的食品外包装缺陷 在线智能检测方法

申 高<sup>1</sup> 王子剑<sup>1</sup> 王晓峰<sup>2</sup> 李瑞文<sup>3</sup>

(1. 安阳职业技术学院, 河南 安阳 455000; 2. 郑州轻工业大学, 河南 郑州 450002;  
3. 河南农业大学, 河南 郑州 450046)

**摘要:** [目的] 针对目前食品生产线外包装缺陷检测中人工检测效率低、漏检误检率高, 以及传统机器视觉算法对复杂纹理包装、微小缺陷识别能力不足的问题, 提出一种兼顾检测精度与速度的智能化检测方法。[方法] 以 Swin Transformer 为核心特征提取模块, 利用其对图像全局信息的建模能力和多尺度特征融合优势, 精准捕捉外包装表面的微小褶皱、印刷偏移等缺陷特征。同时引入 YOLOv12 的快速检测框架, 通过优化颈部网络与损失函数, 实现缺陷区域的快速定位与分类, 形成高精度特征提取—快速目标检测的一体化检测流程。[结果] 该方法对常见缺陷类型的平均检测精度均值 > 96.50%, 较优化前方法提升超过 10.00%。单张图像检测耗时 < 10 ms, 满足生产线每秒 30 帧的实时检测要求。此外, 在不同食品检测中, 该方法仍保持稳定性能, 鲁棒性显著优于对比方法。[结论] 通过融合 Swin Transformer 特征提取优势与 YOLOv12 快速检测能力, 解决了食品外包装缺陷检测中精度与速度难以兼顾的核心问题。

**关键词:** 食品生产线; 外包装缺陷检测; Swin Transformer; YOLOv12; 兼顾检测精度与速度

## Intelligent online detection method for food packaging defects based on improved deep learning

SHEN Gao<sup>1</sup> WANG Zijian<sup>1</sup> WANG Xiaofeng<sup>2</sup> LI Ruiwen<sup>3</sup>

(1. Anyang Vocational and Technical College, Anyang, Henan 455000, China; 2. Zhengzhou University of Light Industry, Zhengzhou, Henan 450002, China; 3. Henan Agricultural University, Zhengzhou, Henan 450046, China)

**Abstract:** [Objective] To propose an intelligent detection method that balances detection accuracy and speed for the problems of low manual detection efficiency, high missed and false detection rates, and insufficient recognition ability of traditional machine vision algorithms for complex texture packaging and small defects in packaging defect detection at current food production lines. [Methods] Using Swin Transformer as the core feature extraction module, this study utilizes its modeling ability for global image information and multi-scale feature fusion advantages to accurately capture defect features, such as small wrinkles and printing offsets on the packaging surface. Simultaneously, the YOLOv12 fast detection framework is introduced, which optimizes the neck network and loss function to achieve fast localization and classification of defect areas, forming an integrated detection process of high-precision feature extraction and fast object detection. [Results] The average detection accuracy of this method for common defect types is higher than 96.50%, improving by over 10.00% compared to the method before optimization. Single image detection takes less than 10 ms, meeting the real-time detection requirement of 30 frames per second for the production line. Additionally, this method still maintains stable performance in tests for different foods, demonstrating significantly better robustness than the comparative method. [Conclusion] By integrating the advantages of Swin Transformer feature extraction with the fast detection capability of YOLOv12, this study solves the core problem of balancing accuracy and

**基金项目:** 中国高等教育学会高等教育科学研究规划重点课题(编号: 22GDZY0311)

**通信作者:** 申高(1986—), 男, 安阳职业技术学院讲师, 硕士。E-mail: mbngfs@sina.com

**收稿日期:** 2025-08-12 **改回日期:** 2025-11-26

**引用格式:** 申高, 王子剑, 王晓峰, 等. 基于改进深度学习的食品外包装缺陷在线智能检测方法[J]. 食品与机械, 2025, 41(12): 236-244.

**Citation:** SHEN Gao, WANG Zijian, WANG Xiaofeng, et al. Intelligent online detection method for food packaging defects based on improved deep learning[J]. Food & Machinery, 2025, 41(12): 236-244.

speed in food packaging defect detection.

**Keywords:** food production line; packaging defect detection; Swin Transformer; YOLOv12; balancing detection accuracy and speed

食品包装作为食品进入市场前的最后一道防线,直接关系到消费者的健康与安全。在现代食品工业中,食品包装的质量与安全性已成为企业竞争力的重要组成部分,直接影响到产品的市场接受度和品牌形象。如食品包装存在微小破损,可能导致微生物侵入,引发食品变质,危害消费者健康<sup>[1-3]</sup>。印刷偏移或模糊的包装则可能误导消费者,影响产品的销售。因此,准确、高效的食品包装缺陷检测是保障食品安全、提升企业经济效益的关键环节<sup>[4]</sup>。

目前,国内外关于食品生产线外包装缺陷检测进行了较为深入的研究,主要集中在人工检测和机器视觉检测两种方法上<sup>[5-7]</sup>。人工检测虽然具有一定的灵活性,但存在效率低下、主观性强等问题。机器视觉检测方法主要集中在传统算法(基于图像处理和模式识别技术等)和深度学习算法(CNN和YOLO系列等)上<sup>[8-9]</sup>。王静蕾等<sup>[10]</sup>针对盒装食品生产线中包装缺陷检测方法存在效率低、漏检率高、难以适配高速生产节奏的问题,提出一种改进YOLOv10模型用于盒装食品包装缺陷检测方案。该方法平均精度均值显著高于传统YOLOv10或其他对比算法,模型推理速度满足盒装食品生产线的实时检测需求。但其在复杂生产环境下的检测稳定性较差,实验室结果与实际生产场景存在差距。向硕等<sup>[11]</sup>针对酒品包装检测中现有方法精度低、模型体积大、参数量多等问题,提出一种改进的YOLOv8n模型用于食品包装缺陷检测。改进后的轻量化YOLOv8n模型参数量较原YOLOv8n减少30%以上,可成功部署在边缘设备上,降低硬件投入成本。模型在酒品包装缺陷检测中的平均精度均值仍保持较高水平,漏检率低于5%,同时推理速度达30帧/s以上,满足高速生产线需求。但其轻量化设计牺牲了部分微小缺陷(如酒瓶表面微小划痕)的检测精度,小众缺陷覆盖能力不足。古莹奎等<sup>[12]</sup>针对饼干包装生产线速度快、外观缺陷检测需“快且准”,现有算法难以平衡检测速度与精度的问题,提出一种改进RT-DETR的饼干包装外观缺陷快速检测方法。改进后的RT-DETR模型推理速度显著优于传统RT-DETR及YOLO系列模型。模型对饼干包装的褶皱、印刷缺陷识别准确率高,漏检率低于3%。但其模型较为复杂,难以部署在算力有限的小型生产线设备上。付赫等<sup>[13]</sup>针对饮料包装生产线中现有缺陷检测算法模型体积大、对硬件要求高等问题。提出一种改进YOLOv8用于食品包装缺陷检测。该方法参数量较传统算法减少50%以上,硬件部署成本降低40%左右。模型在主流饮料包装缺陷(如瓶身破损、标签偏移)

的检测准确率达90%以上。但对饮料包装的微小缺陷(如易拉罐拉环微小裂纹)检测精度不足,存在抓大放小的问题。在深度学习的众多模型中,Swin Transformer和YOLOv12因其独特的特性,在图像检测领域备受关注,为实现高精度、高效率的食品包装缺陷检测提供了有力的技术支持。

基于此,研究拟提出一种融合Swin Transformer强大特征提取与建模能力和改进YOLOv12快速检测能力的食品生产线食品外包装缺陷智能化检测方法,旨在为食品自动化生产线的智能化质检提供可靠技术支撑。

## 1 食品生产线外包装缺陷智能化检测系统

食品生产线外包装缺陷智能化检测系统结构如图1所示,主要由主控平台、数据采集模块、传送机构和执行机构等组成,主控平台作为整个检测系统的智慧大脑,采用高性能处理器,具备强大的运算和决策能力,负责协调系统中各个模块的协同工作,确保整个检测流程有条不紊地进行。数据采集模块作为系统的神经末梢,利用工业相机和传感器来完成图像采集和食品检测。工业相机具有高分辨率、高速拍摄的特点,能够在极短的时间内捕捉到食品包装的清晰图像。传送装置是整个系统中的运输纽带,它以稳定可靠的运行,持续不断地将食品包装输送到指定的检测位置。六轴机械臂作为系统中的精准执行结构,每个关节都由精密的减速器、伺服电机和编码器构成,实现高精度的运动控制<sup>[14]</sup>。

## 2 食品生产线外包装缺陷检测方法

试验提出一种面向食品生产线的实时高精度外包装缺陷智能化检测方法。该方法针对传统检测高精度建模

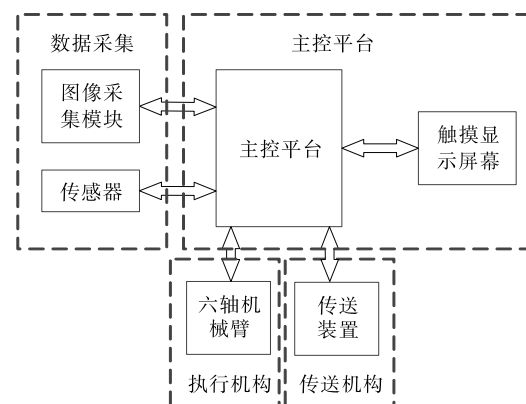


图1 食品生产线外包装缺陷智能化检测系统

Figure 1 Intelligent detection system for packaging defects in food production lines

与快速检测难以兼顾的痛点,创新性地融合 Swin Transformer 的强特征提取能力与改进 YOLOv12 的高效检测优势,采用 Swin Transformer 直接替换 YOLOv12 原有的特征提取网络,旨在同时满足食品包装缺陷(如皱包缺陷、未封边缺陷等)的精准识别需求,同时优化颈部网络与损失函数实现缺陷区域的快速定位与分类。

### 2.1 图像预处理方法

在实际的食品生产环境中,由于光线的不均匀、生产设备的振动以及采集设备本身的噪声等因素,采集到的图像往往会包含各种噪声和干扰,这些噪声不仅会降低图像的清晰度,还对后续的特征提取和缺陷识别造成误导。试验采用中值滤波器—均值滤波器对图像数据进行预处理。通过中值滤波对脉冲噪声的强抑制能力,剔除孤立的亮暗点(如食品表面的粉尘反光),避免均值滤波对这类噪声的平滑扩散(导致缺陷边缘模糊)。通过均值滤波器(高斯核)进行二次优化:在中值滤波基础上,通过高斯核对图像进行平滑处理,消除像素值的随机波动(如食品表面的纹理噪声)<sup>[15-16]</sup>。

### 2.2 融合 Swin Transformer 和 YOLOv12 的缺陷检测方法

食品生产线外包装缺陷(如微小褶皱等)具有特征尺度小、分布零散、易与背景混淆的特点。传统 YOLOv12 依赖卷积层进行局部特征提取,对全局上下文关联的建模能力较弱,易漏检微小缺陷;而 Swin Transformer 通过窗口注意力机制,既能捕捉局部细节,又能建模全局像素关联,可精准识别微小缺陷,但单独使用时检测速度难以满足生产线实时性要求。因此,试验融合 Swin Transformer 特征提取和改进 YOLOv12 模型的协同架构,前者负责将原始图像转化为高辨识度缺陷特征图,后者负责将特征图快速转化为缺陷位置+类别的检测结果,实现高精度特征感知与高速度检测输出的平衡<sup>[17-20]</sup>。

Swin Transformer 模型结构如图 2 所示。

**Patch Partition 模块:**对输入进行分块,然后在 channel 方向展平。Linear Embedding 层:对输入做线性变换,将维度由 48 变成  $C$ ,图像 shape 由  $(H/4, W/4, 48)$  变成  $(H/4, W/4, C)$ 。

**Stage1:**先通过 Linear Embedding 层,输出大小为  $(H/4) \times (W/4)$ ,通道数为  $C$  的特征图。

**Stage2:**串联多个 Swin Transformer Block,这些 Block 成对出现,第 1 个 Block 使用 W-MSA (windows multi-head self-attention),第 2 个 Block 使用 SW-MSA (shifted windows multi-head self-attention),交替使用,输出特征图大小不变,仍为  $(H/4) \times (W/4) \times C$ 。

**Stage3:**先通过一个 Patch Merging 层进行下采样,它将特征图中每  $2 \times 2$  相邻的 4 个图块拼接在一起,这样通道数变为  $4C$ ,再通过一个线性层将通道数降为  $2C$ ,特征图

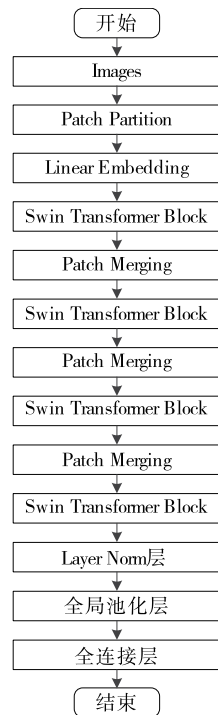


图 2 Swin Transformer 模型

Figure 2 Swin Transformer model

分辨率减半  $(H/8) \times (W/8)$ 。然后再次串联多个成对的 (W-MSA, SW-MSA) Block,对新的特征图进行处理,输出  $(H/8) \times (W/8) \times 2C$  的特征图。

**Stage4:**重复 Patch Merging+Swin Transformer Blocks 的组合,最终输出  $(H/16) \times (W/16) \times 4C$  的特征图。

**分类头:**通过 Layer Norm 层、全局池化层以及全连接层输出最终结果。

YOLOv12 模型网络结构延续了 YOLO 系列特征提取—特征融合—检测头的三段式架构,核心优化集中在轻量化特征提取和高效特征融合,整体兼顾检测速度与精度,尤其适配实时检测场景。YOLOv12 模型网络结构如图 3 所示。

(1) 骨干网络:采用改进的 CSPNeXt-20 作为骨干网络,替代了前代的 CSPDarknet,在减少计算量的同时提升特征提取能力。将输入图像(如  $640 \times 640$ )逐步缩小至  $40 \times 40$ 、 $20 \times 20$ 、 $10 \times 10$ ,对应输出 3 个不同尺度的特征层。A2C2F 模块结合区域注意力与 CSP 瓶颈层,增强特征的空间与上下文理解,提升多尺度目标检测精度。区域注意力(A2):将输入特征图划分为若干区域(如按垂直或水平方向均分),各区域独立计算自注意力,减少计算量并保持大感受野;常采用可分离卷积(如  $7 \times 7$  深度可分离卷积)辅助位置信息感知。C2F 部分:由 2 个 CSP 瓶颈层组成,含  $1 \times 1$  卷积降维、中间卷积块与  $1 \times 1$  卷积升维,并通过残差连接稳定训练。先降维再卷积,可有效融合

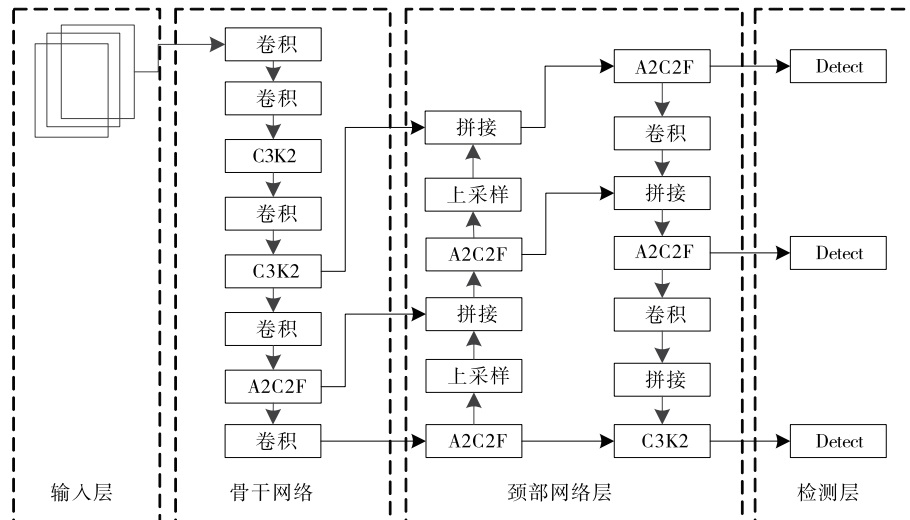


图3 YOLOv12网络结构

Figure 3 YOLOv12 network structure

跨通道与跨区域信息。C3K2模块基于CSP(跨阶段局部)架构,通过分组卷积与深度可分离卷积,在减少参数数量的同时增强特征提取与融合能力,提升模型效率与精度。CSP架构:将输入特征分为两部分,一部分直接跨层连接,另一部分经多组卷积(含深度可分离卷积)处理后再与直连部分拼接。

(2) 颈部网络:采用改进型FPN+PAN结构,解决“小目标特征易丢失”的问题,实现多尺度特征的深度融合。自顶向下融合(FPN):将高层语义特征(C5)通过上采样,与中层特征(C4)、低层特征(C3)依次融合,补充小目标所需的细节特征。自底向上融合(PAN):将融合后的低层特征(C3融合后)通过下采样,反向传递至中层、高层,增强大目标的位置定位精度。输出特征层:最终输出3个用于检测的特征层(记为P3、P4、P5),尺寸分别为 $80 \times 80$ 、 $40 \times 40$ 、 $20 \times 20$ ,对应检测“小目标、中目标、大目标”。

(3) 检测层:采用轻量化检测头,移除前代的锚框,改用无锚框设计,减少先验框参数带来的计算冗余。核心操作:对Neck输出的P3、P4、P5特征层,分别通过 $1 \times 1$ 卷积压缩通道数,再用 $3 \times 3$ 卷积输出检测结果。输出维度:每个特征层的每个像素点输出4个坐标值( $x, y, w, h$ ) + 1个置信度 +  $N$ 个类别概率( $N$ 为检测类别数),最终通过非极大值抑制(NMS)筛选出最终的目标框。

YOLOv12的颈部采用常规卷积( $3 \times 3$ 等固定核)进行特征融合,其感受野固定且各向同性,难以适应食品外包装中不规则缺陷(如非对称褶皱、弯曲划痕、边缘变形)的特征分布。试验引入可变卷积,可变卷积(deformable convolution)作为一种创新的卷积方式,为解决不规则缺陷特征捕捉问题带来了新的希望。其核心原理是引入额外的偏移量,使卷积核的采样点能够根据输入特征图的

内容进行动态调整,从而更加灵活地适应各种复杂的形状和变化。将颈部网络原有的 $3 \times 3$ 常规卷积替换为Deformable ConvV2。当检测图像中的不规则缺陷时,它可以根据缺陷的形状和位置,动态调整采样点,使得融合后的特征能够更准确地反映缺陷的特征,为后续的检测提供更有力的支持。

YOLOv12采用CIoU损失作为定位损失,虽考虑了边界框的重叠度、中心点距离和长宽比,但对缺陷边缘的细微偏移(如褶皱的边缘定位偏差1~2像素)不敏感。同时,分类损失与定位损失的权重固定,当缺陷与背景相似度高(如浅色包装上的浅划痕)时,分类置信度易受定位误差干扰,导致误分类为背景。试验引入EIoU损失函数、边界距离损失函数和动态权重Focal损失函数,如式(1)所示。

$$L_{\text{ZIOU}} = \alpha L_{\text{EIoU}} + \beta L_{\text{edge}} + \gamma L_{\text{cls}}, \quad (1)$$

式中:

$L_{\text{EIoU}}$ ——EIoU损失函数;

$L_{\text{edge}}$ ——边界距离损失函数;

$L_{\text{cls}}$ ——动态权重Focal损失函数;

$\alpha, \beta, \gamma$ ——函数权重系数。

融合Swin Transformer和YOLOv12的模型结构如图4所示。

### 3 食品生产线外包装缺陷智能化检测试验分析

#### 3.1 试验参数

为验证所提食品生产线外包装缺陷智能化检测方法的优越性,搭建了如图5所示的缺陷检测试验平台,该平台可实现对小包装食品外包装缺陷的自动化采集与检测。试验选取小包装食品为研究对象,涵盖6类样本类



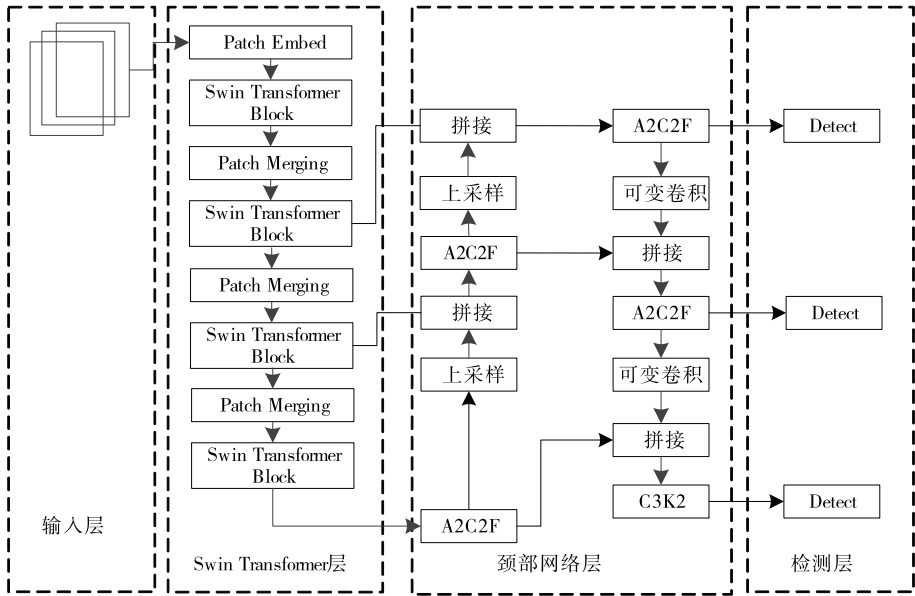


图 4 融合 Swin Transformer 和 YOLOv12

Figure 4 Integrating Swin Transformer and YOLOv12

型,包括包装正常样本及 5 类典型缺陷样本(皱包缺陷、未封边缺陷、封边不严缺陷、连包缺陷、压线缺陷),部分试验样本如图 6 所示。其中,正常样本 1 000 包,皱包缺陷、未封边缺陷、封边不严缺陷、连包缺陷、压线缺陷样本各 200 包,数据集总规模为 2 000 包。为保证试验的科学性与泛化性,数据集按照 7:1:2 的比例划分为训练集(1 400 包)、验证集(200 包)和测试集(400 包),分别用于模型训练、参数优化与性能评估。系统参数见表 1,算法参数见表 2。参数确定方法:经验初始化、网格搜索优化(在经验初始值附近构建搜索空间,遍历组合验证性能)、自适应动态调整(针对不同缺陷类型动态调整权重)、正则化约束(防止过拟合)。

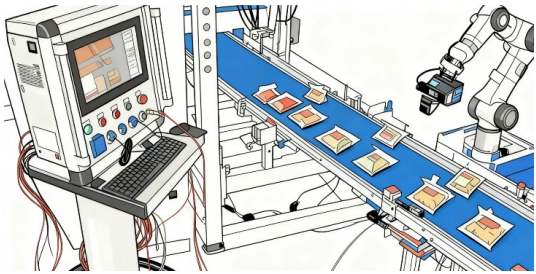


图 5 缺陷检测试验平台

Figure 5 Defect detection test platform

3.2 缺陷智能化检测试验

为深入验证所提融合模型(Swin Transformer+改进 YOLOv12)的训练稳定性与特征学习效率,图 7 对比了该模型与优化前 3 种对比模型(YOLOv12 原始模型、

YOLOv12+可变卷积、YOLOv12+损失函数)在训练过程中总损失值的动态变化趋势。总损失值综合反映了模型对缺陷定位精度(边界框回归误差)与分类准确性(类别判断误差)的拟合效果,其数值越低、收敛速度越快,表明模型对食品包装缺陷特征的学习能力越强。在同一环境下,通过试验数据集进行训练。不同方法随迭代变化损失值如图 7 所示。

由图 7 可以看出,YOLOv12 模型在 70 轮后趋于稳定,最终损失值维持在 1.3 左右。这是由于其依赖常规卷积进行特征提取,对微小缺陷和不规则缺陷的特征捕捉能力有限,且 CIoU 损失对边缘定位误差敏感,导致模型难以快速拟合复杂缺陷特征。YOLOv12+可变卷积模型:收敛速度略有提升,60 轮后稳定,最终损失值降至 0.90。可变卷积通过动态调整采样点增强了对不规则缺陷的适应性,但其损失函数仍采用原始 CIoU+交叉熵组合,对难分类样本关注度不足,导致损失值下降幅度受限。YOLOv12+损失函数模型:10 轮后下降速度加快,60 轮后稳定在 0.85。改进的 EIou 损失强化了边界框整体定位精度,动态 Focal 损失通过降低背景样本权重聚焦缺陷特征,弥补了原始损失函数的短板,但常规卷积的固定感受野仍限制了对全局-局部特征关联的建模,损失值难以进一步降低。试验所提融合模型:初始损失值最低 4.5,收敛速度最快,20 轮后基本稳定,最终损失值降至 0.45,显著低于其他模型。这得益于 Swin Transformer 的窗口注意力机制,既能通过局部窗口捕捉微小缺陷细节,又能通过移位窗口建模全局关联,配合改进 YOLOv12 的可变卷积与优化损失函数,形成了强特征提取-精准特



Figure 6 Some experimental samples

表 1 系统参数  
Table 1 System parameters

配置	参数
CPU	Intel (R) Core (TM) i7-13700 处理器
GPU	NVIDIA GeForce RTX 3070
Python 环境	Python 3.7.7
工业相机	Basler acA2500-14gm GigE
光源	奥普特 OPT-R4020
触发装置	E3Z-D61 光电传感器
传送驱动系统台达	ASD-A2-0721-M 伺服驱动器与 750 W 伺服电机驱动
分拣机器人	六轴工业机器人
数据采集卡	PCI-1716 数据采集卡
操作系统	Windows 11
深度学习框架	Pytorch2.0

征融合—高效损失优化的协同机制,使模型能快速且稳定地拟合各类缺陷特征。

将训练后模型用于测试集测试,优化前后模型检测效果见表 3。

由表 3 可知,所提融合 Swin Transformer 和改进 YOLOv12 的试验方法在食品包装缺陷检测任务中表现出显著优越性。试验方法的准确率达到 98.50%,相较于 YOLOv12 模型的 89.25% 提升 10.36%,相较于仅优化可变卷积模型的 93.50% 提升 5.35%,相较于仅优化损失函数模型的 94.75% 提升 3.96%。这表明融合 Swin

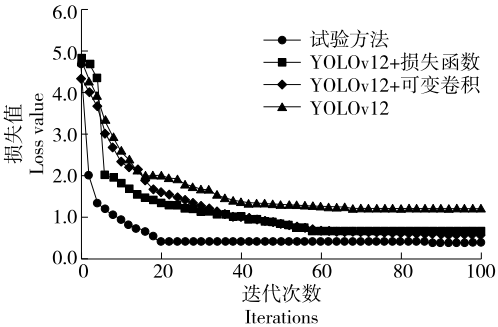


图 7 不同方法随迭代变化损失值  
Figure 7 Loss values with iteration in different methods

Transformer 的全局—局部特征提取能力与改进 YOLOv12 的结构优化后,模型对各类缺陷(尤其是微小缺陷如压线、封边不严等)的识别准确性大幅提高。试验方法的平均精度均值为 96.75%,较 YOLOv12 模型的 85.60% 提升 13.03%,较 YOLOv12+可变卷积的 90.15% 提升 7.32%,较 YOLOv12+损失函数的 91.80% 提升 5.39%。平均精度均值的显著提升证明该方法对缺陷的定位精度和类别判断一致性更优,尤其对形态不规则的缺陷(如非对称皱包等)的检测鲁棒性更强。试验方法的平均检测时间为 8.85 ms,虽略高于其他对比模型(8.00~8.45 ms),但仍远低于食品生产线实时检测的阈值(通常要求单帧检测时间<30 ms),可满足高速生产线(30 包/s 以上)的检测需求。这一结果验证了融合架构在精度提升的同时,通过轻量化设计与协同优化有效控制了计算开销,实现了精度与速度的平衡。

表 2 算法参数  
Table 2 Algorithm parameters

参数	数值	参数	数值
原图尺寸	640 像素×640 像素	批量大小	8
图像分块	4×4	迭代次数	100
线性嵌入维度	96	优化器	AdamW
注意力 dropout 概率	0.1	损失函数权重	0.5, 0.3, 0.2
前馈网络 dropout 概率	0.1	置信度阈值	0.5
初始学习率	0.001	NMS 阈值	0.45
学习率衰减策略	余弦退火		

表 3 优化前后模型检测效果

Table 3 Model detection performance before and after optimization

方法	准确率/%	平均精度 均值/%	平均检测 时间/ms
YOLOv12	89.25	85.60	8.00
YOLOv12+可变卷积	93.50	90.15	8.45
YOLOv12+损失函数	94.75	91.80	8.12
试验方法	98.50	96.75	8.85

为量化试验方法(融合 Swin Transformer 与改进 YOLOv12)对不同类别缺陷的识别精度,明确模型在类别区分中的优势与短板,基于测试集 400 个样本的检测结果构建混淆矩阵。矩阵中行代表真实缺陷类别,列代表模型预测类别,对角线数值为各类别缺陷的正确识别数量,非对角线数值为类别混淆数量,试验方法的缺陷检测混淆矩阵见表 4。

表 4 试验方法的缺陷检测混淆矩阵

Table 4 Confusion matrix for defect detection in experimental methods

真实/预测	正常	皱包	未封边	封边不严	连包	压线
正常	200	0	0	0	0	0
皱包	0	39	0	1	0	0
未封边	0	0	40	0	0	0
封边不严	0	1	1	38	0	0
连包	0	0	1	0	39	0
压线	0	1	0	1	0	38

由表 4 可以看出,试验方法对 6 类样本的整体识别准确率(对角线总和/总样本数)达 98.75%(395/400 包),其中 4 类样本(正常、未封边、皱包、连包)召回率 $\geq 97.5\%$ ,2 类样本(压线、封边不严)召回率 $\geq 95\%$ ,无任何类别出现高混淆率问题。完全满足食品生产线多类别缺陷高精度、低误判的检测需求。

为验证所提食品生产线外包装缺陷智能化检测方法的优越性与泛化能力,选取改进 YOLOv10 模型<sup>[10]</sup>和改进 RT-DETR 模型<sup>[12]</sup>作为对比方法,开展跨品类小包装食品缺陷检测对比试验。试验样本覆盖 4 类典型袋装食品:袋装榨菜、袋装薯片、袋装雪饼、袋装饼干,每类样本的测试集规模均为 400 包,涵盖正常样本与 5 类典型缺陷样本(比例与原始数据集一致),以全面评估不同方法在多样化包装材质、缺陷类型及背景干扰下的检测性能。不同方法在不同样本中的检测效果见表 5。

由表 5 可以看出,所提融合 Swin Transformer 和改进 YOLOv12 的检测方法,在不同类型小包装食品的缺陷检

表 5 不同方法在不同样本中的检测效果

Table 5 Detection performance of different methods on different samples

方法	样本	数量	检测准 确数	检测准 确率/%	平均精度 均值/%	平均检测 时间/ms
改进 YO- LOv10 模型	榨菜	400	362	88.60	85.60	15.0
	薯片	400	356			
	雪饼	400	348			
	饼干	400	352			
改进 RT- DETR 模型	榨菜	400	384	93.80	91.70	12.0
	薯片	400	374			
	雪饼	400	368			
	饼干	400	372			
试验方法	榨菜	400	392	97.80	97.00	9.0
	薯片	400	390			
	雪饼	400	388			
	饼干	400	394			

测任务中,均表现出更高的检测精度、更强的泛化能力和更快的检测速度。试验方法的平均检测准确率 97.8%,较改进 RT-DETR 模型的 93.80% 提升 4.26%,较改进 YOLOv10 模型的 88.60% 提升 10.38%,平均精度均值 97.00%,较改进 RT-DETR 模型的 91.70% 提升 5.78%,较改进 YOLOv10 模型的 85.60% 提升 13.32%。表明该方法对不同品类食品包装的缺陷检测具有更强的泛化能力,尤其在处理复杂背景和微小缺陷时表现更优。试验方法的平均检测时间优于改进 YOLOv10 模型和改进 RT-DETR 模型。在保证精度的同时,满足了多品类食品生产线的高速检测需求(单包检测时间 $<30\text{ ms}$ )。这得益于试验方法融合的 Swin Transformer 能自适应捕捉不同食品包装的材质特征,配合改进 YOLOv12 的可变卷积技术,有效解决了不同包装表面缺陷的差异化检测难题。

为验证试验方法在实际生产线分拣场景中的实用性,基于搭建的缺陷检测试验平台开展动态分拣试验,选取袋装榨菜、袋装薯片、袋装雪饼、袋装饼干为分拣目标,设置传送带速度为低速(100 mm/s)和高速(200 mm/s)两个梯度,模拟不同生产节奏下的分拣需求。分拣过程通过六轴机械臂完成,系统根据试验方法的检测结果(缺陷类型及位置)控制机械臂执行抓取动作,不同目标分拣效果见表 6。

由表 6 可以看出,试验方法在实际分拣场景中表现出高成功率、高效率 and 强适应性,即使在传送带速度提升至 200 mm/s 的高速生产节奏下,仍能满足食品生产线的分拣需求。在 100 mm/s 低速场景下,4 类目标的平均分拣成功率达 98.75%。当传送带速度提升至 200 mm/s 时,平均

表 6 不同目标分拣效果  
Table 6 Sorting effects for different targets

目标	100 mm/s			200 mm/s		
	食品数/个	分拣成功率/%	平均分拣时间/s	分拣食品数/个	分拣成功率/%	平均分拣时间/s
榨菜	400	99.00	0.75	400	97.50	0.62
薯片	400	98.50	0.78	400	96.75	0.65
雪饼	400	98.00	0.80	400	96.00	0.68
饼干	400	99.50	0.72	400	98.00	0.60

成功率降至 97.06%,但仍维持在较高水平。低速场景下,4 类目标的平均分拣时间为 0.76 s,高速场景下缩短至 0.64 s,均满足生产线节拍要求(单包处理时间<1 s)。其中,饼干的分拣时间最短(低速 0.72 s,高速 0.60 s),雪饼因包装表面图案复杂导致检测耗时略长,进而影响整体分拣效率,但仍在合理范围内。这得益于试验方法的快速检测能力,为机械臂预留了充足的动作响应时间,进而为食品生产工业化应用提供了有力支撑。

4 结论

试验针对食品生产线外包装缺陷检测中人工检测效率低、传统机器视觉算法对微小及不规则缺陷识别能力不足、精度与速度难以兼顾的问题,提出融合 Swin Transformer 与改进 YOLOv12 的智能化检测方法,并搭建配套检测系统。结果表明,试验方法对食品包装缺陷的检测准确率达 98.50%、平均精度均值达 96.75%,较 YOLOv12 原始模型分别提升 10.36% 和 13.30%,同时单帧检测时间仅 8.85 ms,满足生产线每秒 30 帧的实时性要求,实现精度与速度的平衡。在不同材质包装的跨品类检测中,该方法平均准确率达 97.8%,比改进 YOLOv10 模型<sup>[10]</sup>和改进 RT-DETR 模型<sup>[12]</sup>分别提升了 10.38% 和 4.26%,且检测时间(9 ms)更短,证明其对不同包装特征与背景干扰具有适应性。动态分拣试验中,传送带速度 100~200 mm/s 时,平均分拣成功率维持 97.06%~98.75%,平均分拣时间 0.64~0.76 s,满足工业化生产节拍需求。后续可进一步探索模型轻量化设计,以适配算力有限的边缘检测设备,扩大应用场景。

参考文献

[1] 赵敏, 范英, 高思伟, 等. 结合 Swin-Transformer 的改进 YOLOv5s 包装盒缺陷检测算法[J]. 制造业自动化, 2024, 46(12): 34-40.  
ZHAO M, FAN Y, GAO S W, et al. Improved YOLOv5s packaging defect detection algorithm combined with swin-Transformer[J]. Manufacturing Automation, 2024, 46(12): 34-40.  
[2] CHEN C, WU B, ZHANG H Y. An image recognition

technology based on deformable and CBAM convolution Resnet50[J]. IAENG International Journal of Computer Science, 2023, 50(1): 274-281.  
[3] 蔡家一, 刘世伟, 单龙祥, 等. 基于机器视觉与 YOLOv5 的裂纹蛋分拣机器人设计与试验[J]. 智能化农业装备学报(中英文), 2025, 6(1): 41-50.  
CAI J Y, LIU S W, DAN L X, et al. Design and experiment of cracked egg sorting robot based on machine vision and YOLOv5 [J]. Journal of Intelligent Agricultural Mechanization, 2025, 6(1): 41-50.  
[4] 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, 151.  
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, 151.  
[5] HUANG J, REN L F, ZHOU X K, et al. An improved neural network based on SENet for sleep stage classification[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(10): 4 948-4 956.  
[6] 施利春, 边可可, 王松伟, 等. 基于改进 U-Net 和 IWOA-LSSVM 的番茄综合品质检测方法研究[J]. 食品与机械, 2025, 41(8): 109-117.  
SHI L C, BIAN K K, WANG W S, et al. Research on tomato comprehensive quality detection method based on improved U-Net and IWOA-LSSVM[J]. Food & Machinery, 2025, 41(8): 109-117.  
[7] 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.  
[8] 徐杰, 刘畅. 基于改进 ELM 和计算机视觉的核桃缺陷检测[J]. 食品与机械, 2024, 40(5): 122-127.  
XU J, LIU C. Walnut defect detection based on improved ELM and computer vision[J]. Food & Machinery, 2024, 40(5): 122-127.  
[9] 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.  
[10] 王静蕾, 赵英杰, 李峰, 等. 基于改进 YOLOv10 的盒装食品生产线包装缺陷检测方法[J]. 食品与机械, 2025, 41(4): 236-241.



- WANG J L, ZHAO Y J, LI F, et al. Packaging defect detection method for boxed food production line based on improved YOLOv10[J]. Food & Machinery, 2025, 41(4): 236-241.
- [11] 向硕, 曾水玲, 贺刚健, 等. 基于 YOLOv8n 改进的轻量化酒品包装缺陷检测算法[J]. 包装与食品机械, 2025, 43(4): 1-12.
- XIANG S, ZENG S L, HE G J, et al. Improved lightweight algorithm for liquor packaging defect detection based on YOLOv8n[J]. Packaging and Food Machinery, 2025, 43(4): 1-12.
- [12] 古莹奎, 叶彪彪, 郭明健, 等. 基于改进 RT-DETR 的饼干包装外观缺陷快速检测[J]. 食品与机械, 2025, 41(2): 234-241.
- GU Y K, YE B B, GUO M J, et al. Rapid detection method of biscuit packaging appearance defects based on improved RT-DETR[J]. Food & Machinery, 2025, 41(2): 234-241.
- [13] 付赫, 王桂英. 饮料包装缺陷检测的轻量化算法研究[J]. 包装与食品机械, 2025, 43(1): 32-39.
- FU H, WANG G Y. Research on lightweight algorithms for beverage packaging defect detection[J]. Packaging and Food Machinery, 2025, 43(1): 32-39.
- [14] 韩宇, 齐康康, 郑纪业, 等. 基于改进 YOLOv11 的轻量化肉牛面部识别方法[J]. 智慧农业(中英文), 2025, 7(3): 173-184.
- HAN Y, QI K K, ZHENG J Y, et al. Lightweight cattle facial recognition method based on improved YOLOv11[J]. Smart Agriculture, 2025, 7(3): 173-184.
- [15] 刘达, 朱兆优, 叶海鹏, 等. 基于改进 YOLOv5 的酒瓶缺陷检测方法研究[J]. 机电工程技术, 2024, 53(3): 211-215.
- LIU D, ZHU Z Y, YE H P, et al. Research on defect detection method for wine bottles based on improved YOLOv5[J]. Mechanical & Electrical Engineering Technology, 2024, 53(3): 211-215.
- [16] 陈卫东, 刘超, 王莹, 等. 基于机器视觉的食品瓶罐包装缺陷检测研究进展[J]. 粮油食品科技, 2024, 32(4): 185-191.
- CHEN W D, LIU C, WANG Y, et al. Research progress on inspecting food bottle and can packaging shortcomings based on machine vision[J]. Science and Technology of Cereals, Oils and Foods, 2024, 32(4): 185-191.
- [17] 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.
- [18] 张润梅, 贾振楠, 李佳祥, 等. 基于多感受野特征增强的改进 EfficientDet 遥感目标检测算法[J]. 电光与控制, 2024, 31(7): 53-60, 96.
- ZHANG R M, JIA Z N, LI J X, et al. An improved EfficientDet remote sensing target detection algorithm based on multi-sensory field feature enhancement[J]. Electronics Optics & Control, 2024, 31(7): 53-60, 96.
- [19] 刘雪, 沈长盈, 吕学泽, 等. 基于改进 MobileNetV3-Large 的鸡蛋新鲜度识别模型[J]. 农业工程学报, 2022, 38(17): 196-204.
- LIU X, SHEN C Y, LYU X Z, et al. Recognizing egg freshness using an improved MobileNetV3-Large[J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(17): 196-204.
- [20] 文韬, 代兴勇, 李浪, 等. 基于机器视觉与光谱融合的柑橘品质无损检测分级系统设计与试验[J]. 江苏大学学报(自然科学版), 2024, 45(1): 38-45.
- WEN T, DAI X Y, LI L, et al. Design and experiment of non-destructive testing and grading system for citrus quality based on machine vision and spectral fusion[J]. Journal of Jiangsu University (Natural Science Edition), 2024, 45(1): 38-45.