

基于改进 Faster R-CNN 的零食包装盒表面缺陷检测

巩雪¹, 孙雪刚¹, 褚洋洋², 崔功卓¹, 李欣妍¹

(1. 哈尔滨商业大学 轻工学院, 哈尔滨 150028;

2. 黑龙江林业职业技术学院 生态工程系, 黑龙江 牡丹江 157011)

摘要: **目的** 针对现有食品包装盒表面缺陷检测方法存在的复杂背景下小目标缺陷检测难、漏检率高、检测精度低等问题, 选择生活中常见的绿豆糕零食包装盒作为检测对象, 提出一种基于改进 Faster R-CNN 的绿豆糕包装盒表面缺陷检测方法。**方法** 以 Faster R-CNN 算法架构为基础, 以 Swin Transformer V2-T 为特征提取主干, 初步提高算法对包装盒缺陷特征的提取能力; 结合加权双向特征金字塔网络 (Bidirectional Feature Pyramid Network, BiFPN) 自适应调节每个尺度特征图的权重并对不同尺寸的特征进行多尺度融合, 以提高识别的准确率; 通过 ROIAlign 结合 ECA 注意力机制替换 ROI Pooling, 去除 2 次量化误差并进一步优化算法对包装盒缺陷的检测能力。**结果** 本检测方法可准确提取目标缺陷, 绿豆糕包装盒表面的 4 种缺陷的检测平均精确率 (Average Precision, AP) 较改进前分别提高 19.66%、12.96%、14.56%、18.86 个百分点, 同时平均精确率均值 (mean Average Precision, mAP) 在 IoU 为 0.5 上较改进前提高了 15.76 个百分点。**结论** 改进后的模型为 Faster R-CNN 在食品包装盒智能化生产上的应用提供了有益的参考和经验。

关键词: 零食包装盒; 缺陷检测; Faster R-CNN; 加权双向特征金字塔网络 (BiFPN); Swin Transformer V2
中图分类号: TB 487 **文献标志码:** A **文章编号:** 1001-3563(2024)23-0232-09
DOI: 10.19554/j.cnki.1001-3563.2024.23.025

Surface Defect Detection of Snack Packaging Box Based on Improved Faster R-CNN

GONG Xue¹, SUN Xuegang¹, CHU Yangyang², CUI Gongzhuo¹, LI Xinyan¹

(1. School of Light Industry, Harbin University of Commerce, Harbin 150028, China; 2. Department of Ecological Engineering, Heilongjiang Forestry Vocational and Technical College, Heilongjiang Mudanjiang 157011, China)

ABSTRACT: Aiming at the problems that the existing food packaging box surface defect detection methods have difficulty in small target defect detection under complex background, high missed detection rate and low detection accuracy, with the mung bean cake snack packaging box common in life as the detection object, the work aims to propose a mung bean cake packaging box surface defect detection method based on improved Faster R-CNN. Firstly, based on the Faster R-CNN algorithm architecture, Swin Transformer V2-T was used as the feature extraction backbone to preliminarily improve the ability of the algorithm to extract the features of the packaging box defect. Secondly, combined with the weighted bidirectional feature Pyramid Network (BiFPN), the weight of each scale feature map was adjusted and the multi-scale fusion was conducted on features of different sizes to improve the recognition accuracy. Finally, ROIAlign was combined with the ECA attention mechanism to replace ROI Pooling, removing two quantization errors and further optimizing the detection ability of the algorithm for packaging box defects. The detection method proposed could accurately extract the target defects, and the Average Precision (AP) of the four defects on the surface of the mung bean cake packaging box increased by 19.66%, 12.96%, 14.56%, and 18.86% respectively. At the same time, the mean average

precision (mAP) increased by 15.76% when the IOU was 0.5. The improved model provides useful reference and experience for the application of Faster R-CNN in the intelligent production of food packaging boxes.

KEY WORDS: snack packaging box; surface defect detection; Faster R-CNN; Bidirectional Feature Pyramid Network (BiFPN); Swin TransformerV2

在盒装食品生产加工过程中, 包装盒表面会因原材料印刷不足、机械振动、传送碰撞等非正常因素而出现污点、露白、折痕与划痕等缺陷; 同时包装盒作为产品的重要外观展示, 对于激发消费者的购买欲望以及增强企业的市场竞争力具有至关重要的作用, 所以良好的包装盒缺陷检测技术至关重要。而对于产品包装盒表面缺陷的检测, 目前大部分企业仍采用人工目测, 虽然灵活度较高, 但效率低、稳定性差、管理成本高, 难以满足现代企业的自动化生产需求^[1]; 同时传统的图像处理^[2-5]方法也因太过依赖预设参数和调整困难等问题而在包装缺陷检测中难以具有较高的稳定性和准确性。

相比之下, 近年来火热的基于深度学习的缺陷检测方法能够通过提前训练数据, 自动学习图像中深层语义特征, 实现对不同种类和形态的缺陷的准确识别^[6]; 同时在后期检测环境或需求变化时, 此类模型还可通过再次训练将标注好的缺陷图像进行自适应调整和优化, 提高模型的检测准确性和稳定性。Adem 等^[7]利用改进后的 Faster R-CNN 对无菌塑料包装的表面缺陷进行检测, 其通过调整原有模型中的卷积步长与填充大小来提高模型的检测速率; 同时通过增加合适的卷积核个数, 来增加特征图通道数, 以此解决因速率上升导致精度下降的问题。Kim 等^[8]利用 YOLOv4 并通过构建食品包装的 X 射线图像数据集, 实现了对意大利面、零食、开心果和红豆的包装缺陷检测, 同时通过分割数据集图像与去除分割图像中的边缘对象使算法的误检率进一步降低。Roslan 等^[9]通过 YOLOv5s 模型在自制数据集训练了 1 000 个 epoch 后, 实现了较高的塑型包装缺陷检测精度。方宇伦等^[10]通过把 VGG16 网络模型的 2 层全连接层分别替换为全局平均池化与 Dropout 正则化来减少网络的神经节点数量, 随后对模型进行剪枝操作, 使得改进后的模型在保持检测精度的同时缩小了模型大小。夏军勇等^[11]把 ResNet50^[12]作为特征提取主干并对接特征融合金字塔 (Feature Pyramid Network, FPN), 随后通过 K 均值聚类调整 anchor 的大小比例, 以此提高 Faster R-CNN 对食品包装盒缺陷的检测精度。叶宇星等^[13]通过添加 SE 注意力机制^[14]和卷积替换来改进 YOLOv5s, 同时融合空洞与深度可分离卷积来优化下采样机制, 使改进后的模型实现了对腌制蔬菜真空包装缺陷较高精度的检测。

目前, 深度学习算法已经实现对部分食品包装缺陷的检测, 但其在缺陷背景复杂、缺陷微小等场景下的应用依旧困难, 而现代食品包装盒表面图案繁杂, 缺陷又多为不易察觉的划痕和细微的印刷露白等。针

对以上问题, 本文选择精度较高的二阶段算法 Faster R-CNN 作为改进对象, 提出一种从主干、特征融合和特征池化等方面改进 Faster R-CNN 的缺陷检测方法, 对 Faster R-CNN 在食品包装盒表面缺陷检测领域的应用具有积极意义。

1 Faster R-CNN 基本原理

Faster R-CNN 的基础模型结构流程如图 1 所示, 其主要由主干网络 VGG16、区域建议网络 (Region Proposal Network, RPN)、ROI 池化 (ROI Pooling) 以及最后的分类器组成。

1) 主干网络 VGG16: 生成特征图。待检测图像经过 VGG16 后输出下采样 16 倍的特征图, 但 VGG16 反复利用 3×3 的小型卷积和 2×2 的池化层, 对面积占比小的缺陷的识别能力差, 同时网络结构简单, 难以有效提取复杂背景下缺陷的特征。

2) RPN: 生成候选区域 (Proposal)。RPNHead 会接收主干输出的特征图并通过步长与填充均为 1、大小为 3×3 的滑动窗口对输入进行卷积和激活操作, 以此来生成特定向量并计算当前卷积覆盖区域的中心像素所对应的原图像素点, 并且以此像素为中心在原图上分别划出 128×128 、 256×256 、 512×512 像素大小, 比例对应为 $1:2$ 、 $1:1$ 、 $2:1$ 的 9 个 anchor; 接着通过特定向量预测每个 anchor 内含有缺陷的概率 (cls_logits) 和 anchor 中心点坐标以及大小的偏移量 (box_pred); 然后去除跨越边界的 anchor, 并剔除交并比 (IoU) > 0.7 的 anchor, 同时通过 box_pred 调整剩下的 anchor, 最终保留 2 000 个调整过的 anchor, 即 Proposal。

3) ROI 池化: 生成固定大小的特征矩阵。根据 cls_logits 筛选的一部分 Proposal 映射回主干输出的特征图上并分割为 7×7 个子区域, 随后进行 maxpooling 操作得到对应的特征矩阵。由于映射与分割时只能取整, 所以目标所在位置的计算会有所偏差, 从而影响模型对小目标缺陷位置的回归精度。

4) 分类器: 预测上一步中 Proposal 内缺陷的类别以及坐标回归信息。Two MLPHead 会通过 2 个全连接层输出 Proposal 对应特征矩阵的特征向量; 随后通过 FC3 层输出此 Proposal 中目标的共 $N+1$ 类别种概率 (cls_logits), 其中 N 为缺陷类别数、1 为背景的数目。经过 softmax 处理后的各类别概率的总和即为 1, 并确定其中输出概率最大的缺陷类别名称; 同时通过 FC4 输出的缺陷边界框回归参数 (box_pred) 来调整缺陷边界框位置。

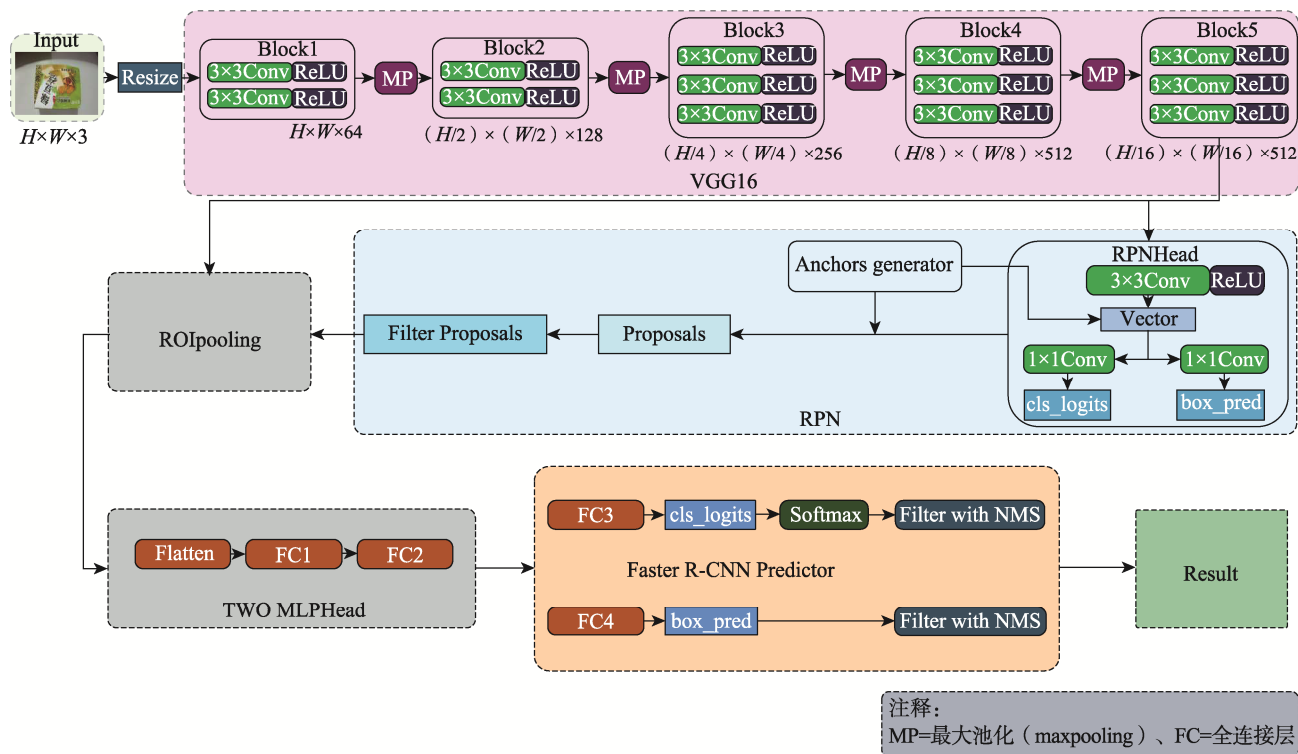


图1 Faster R-CNN 基础模型结构流程
Fig.1 Basic model structure flow of Faster R-CNN

2 改进的 Faster R-CNN 模型结构

改进的 Faster R-CNN 模型结构如图 2 所示, 改

进主要包含 3 个部分: 1) 更换主干网络为 Swin TransformerV2-T; 2) 融合 BiFPN; 3) 替换 ROI Pooling 为 ROI Align 并结合 ECA 注意力机制。

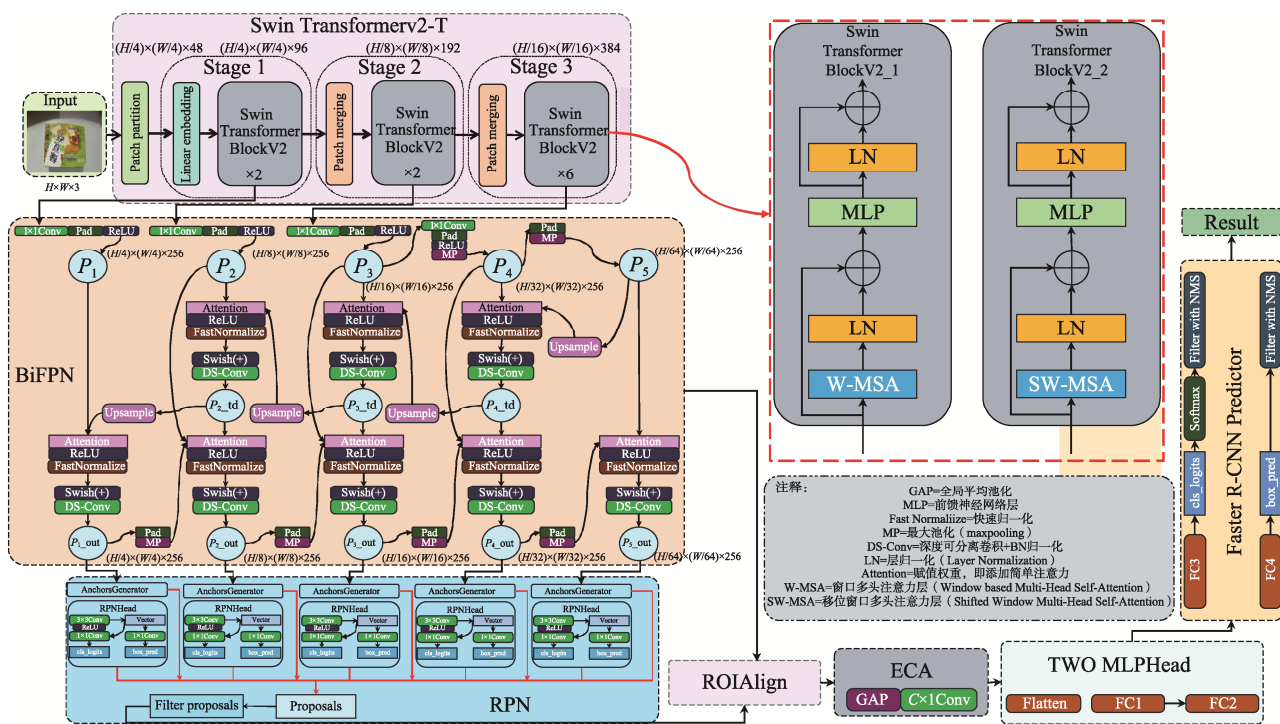


图2 改进的 Faster R-CNN 模型结构
Fig.2 Improved Faster R-CNN model structure

输入图像经过 Swin Transformer V2-T 后会输出 3 个尺寸不同的特征图进入 BiFPN, BiFPN 会在这 3 个特征图各自权重的影响下对其进行特征融合, 随后输出 5 个不同的特征图作为 RPN 的输入; 接着 RPN 会对这 5 个特征图进行 Proposal 的生成与筛选; 然后 ROIAlign 对筛选后的 Proposal 对应的 BiFPN 特征图上的区域输出无量化误差的特征矩阵并利用 ECA 注意力机制来增加特征矩阵中的重要通道对缺陷检测的贡献, 最后经过分类器输出最终的检测结果。

2.1 Swin Transformer V2-T

要想提升复杂背景下对露白、污点等占图像面积 $< 0.5\%$ 的小目标缺陷的检测精度, 主干网络的选择至关重要。Swin Transformer 在 Vision Transformer 的基础上融合了 ResNet 的思想, 继承了 CNN 与 Transformer 的优点, Swin Transformer V2 又在一代的基础上优化了 Swin Transformer Block 的结构和内部数据处理的方式, 具有极高的特征提取能力。为了兼顾模型的运算速度, 本文选择 Swin Transformer V2 系列中的 Tiny 版本作为改进的 Faster R-CNN 算法的主

干网络; 同时为了减少计算量, 只保留 Swin Transformer V2-T 的前 3 个 stage 结构, 精简后的主干结构如图 3 所示。

1) Patch Partition 层将输入图像进行缩小, 其把输入图像按 4×4 大小的区域分割为一个 patch, 并在深度方向上将每个 patch 展平。

2) Linear Embedding 层用于扩大输入通道数, 以此模拟卷积网络提取特征的同时减少计算量。

3) Patch Merging 层的主要作用与 Patch Partition 层和 Linear Embedding 层相同, 即缩小特征图与扩展通道数, 原理如图 4 所示。

Patch Merging 层把传递过来的 (图中假设为 6×6 大小) 特征图上每 2×2 区域内的像素拼接为一个 patch, 再根据位置信息把原特征图分为 $(M/2) \times (N/2)$ 个特征矩阵, 随后在深度方向排列所有特征矩阵, 最后通过 LayerNorm 层归一化并利用全连接层使特征图通道数由输入特征图的 4 倍变为 2 倍。

4) Swin Transformer Block V2 模块是 Swin Transformer V2 区别于一代的关键之处, 其原理结构如图 5 所示。

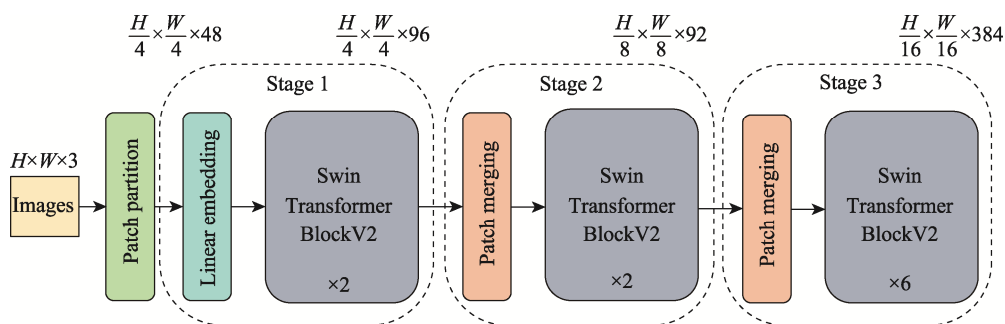


图 3 Swin Transformer V2-T 网络的主要结构
Fig.3 Main structure of Swin Transformer V2-T network

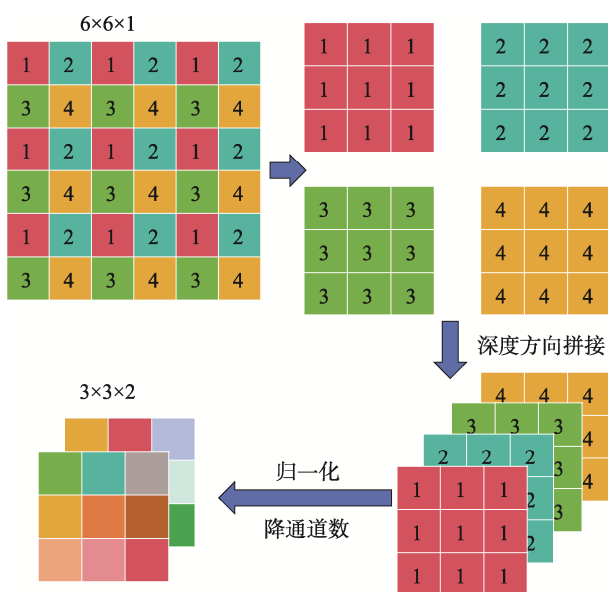


图 4 Patch Merging 原理
Fig.4 Principle of Patch Merging

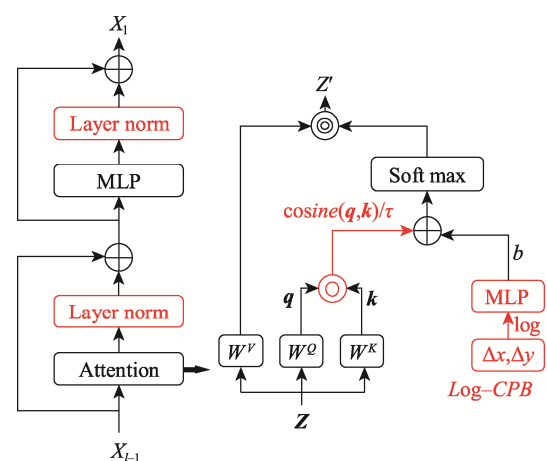


图 5 Swin Transformer Block V2 结构
Fig.5 Structure of Swin Transformer Block V2

为了稳定深层和浅层的输出, 进而加强特征提取能力与加速训练收敛。Swin Transformer Block V2 把一代 Block 中的查询向量 q 与关键向量 k 的点积注意

力替换为余弦注意力,利用余弦注意力的计算与输入振幅无关的特点来稳定输出并减少输出值极端化的问题,原理如式(1)所示。

$$\text{Sim}(\mathbf{q}_i, \mathbf{k}_j) = \cos(\mathbf{q}_i, \mathbf{k}_j) / \tau + B_{ij} \quad (1)$$

式中: τ 为缩放余弦的系数; B_{ij} 为像素 i 和 j 之间的相对位置偏差。

Swin Transformer Block V2 中还将 LayerNorm 层移到 Attention 层与前馈神经网络 (Multilayer Perceptron, MLP) 之后,从而对 Attention 与 MLP 的输出进行归一化,进一步稳定输出结果,优化训练过程。同时为了解决下游任务训练中图像分辨率不适配的问题, Swin Transformer Block V2 利用对数间隔的连续位置偏差 (Log-CPB) 替代原来相对位置编码的线性变换和利用一个小型的元网络自适应生成相对位置编码来解决上述问题,其中对数间隔的连续位置偏差公式如式(2)所示。

$$\begin{aligned} \widehat{\Delta x} &= \text{sign}(x) \cdot \log(1 + |\Delta x|) \\ \widehat{\Delta y} &= \text{sign}(y) \cdot \log(1 + |\Delta y|) \end{aligned} \quad (2)$$

式中: $\widehat{\Delta x}, \widehat{\Delta y}$ 分别是对数间隔坐标; $\Delta x, \Delta y$ 为线性标度坐标。

2.2 BiFPN 模块

为了有效检测出图像中折痕等客观形态特征不定的缺陷,必须增强模型对底层特征的提取能力,并兼顾图像的高层语义信息。传统的方法是在主干网络后串接 FPN 或是路径聚合网络 (Path Aggregation

Network, PANet), 2 者对特征的融合方式简单直接,无论待融合的特征重要与否,2 者都会对其进行相加,从而影响模型的检测精度;而 BiFPN 内的节点融合其他多个节点的输出,并且这些节点都经过简单的自注意机制来突出重要特征,具有更强的特征提取能力。所以本文利用 BiFPN 作为主干后的特征融合网络,其详细的结构如图 6 所示。

当主干 Swin Transformer V2-T 的 1~3 stage 的特征图输入 BiFPN 后会经过 1×1 卷积、填充以及 ReLU 函数来调整通道数至 256;随后通过相同操作以及 maxpooling 使 P_3 输出 P_4 ;接着 P_4 通过下采样得到 P_5 ;随后对 $P_1 \sim P_5$ 添加简单的注意力,即生成可训练的权重参数,并对权重进行 ReLU 操作和快速归一化 (Fast Normalized Fusion) 来确保参数 > 0 和减少参数量。相比于 softmax 归一化,快速归一化计算更加简单高效,2 者定义如式(3)和式(4)所示。

$$O = \sum_i \frac{e^{w_i}}{\sum_j e^{w_j}} \times I_i \quad (3)$$

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

式中: $\varepsilon = 0.0001$; I_i 为第 i 个特征。

式(4)通过把式(3)的指数级运算替换为简单的加和来降低计算量, ε 的目的为稳定输出结果。

随后大小不同的各个特征图进行相互融合,以图 6 中的 P_3 列为例,具体原理如式(5)、式(6)所示。首先 P_3 、 $P_4_{\text{td_up}}$ 会与各自对应权重相乘,接着 2 者结

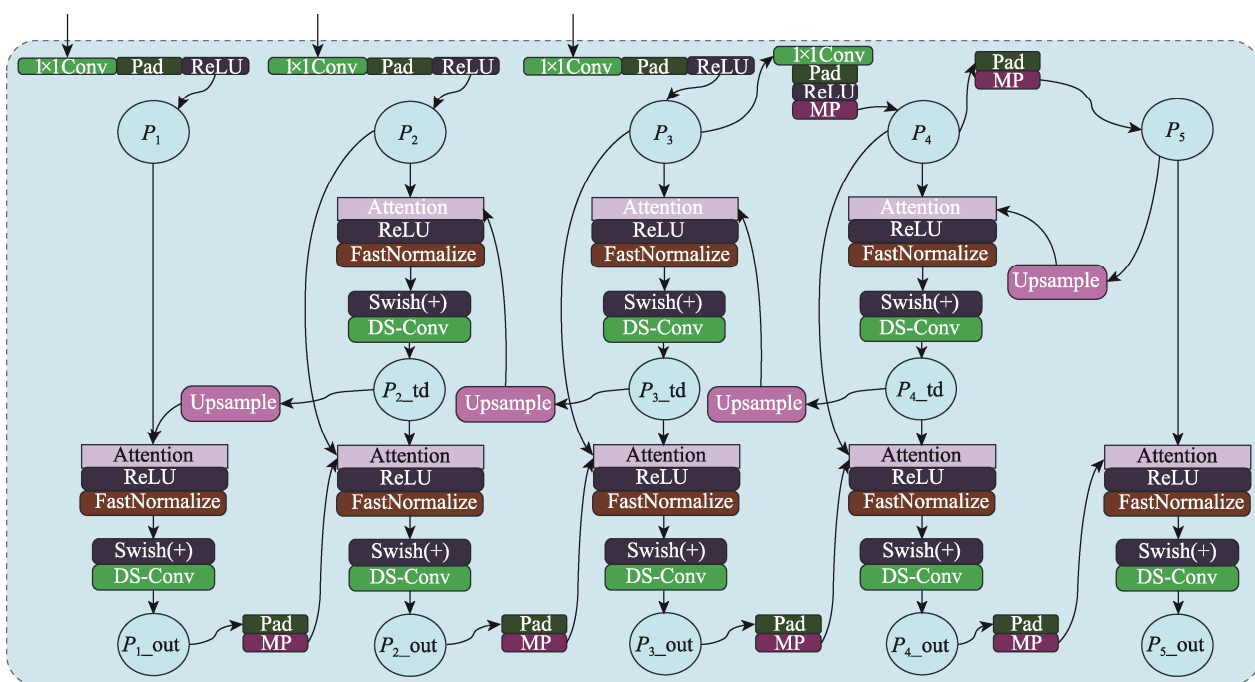


图 6 BiFPN 模型结构
Fig.6 BiFPN model structure

果相加并除以 2 者权重与 ε 之和, 再经过 DS-Conv' 操作得到 P_{3_td} , 随后 P_{3_td} 会与 P_3 、 $P_{2_out_down}$ 进行特征融合得到 P_{3_out} 。

$$P_{3_td} = \text{DS-Conv}' \left(\frac{\omega_1 \cdot P_3 + \omega_2 \cdot P_{4_td_up}}{\omega_1 + \omega_2 + \varepsilon} \right) \quad (5)$$

$$P_{3_out} = \text{DS-Conv}' \left(\frac{\omega'_1 \cdot P_3 + \omega'_2 \cdot P_{3_td} + \omega'_3 \cdot P_{2_out_down}}{\omega'_1 + \omega'_2 + \omega'_3 + \varepsilon} \right) \quad (6)$$

式中: $\varepsilon=0.0001$; ω_i 为各自节点特征的权重; $P_{4_td_up}$ 为 P_{4_td} 上采样结果; $P_{2_out_down}$ 为 P_{2_out} 下采样结果; DS-Conv' 为 Swish 函数结合 DS-Conv (深度可分离卷积和快速归一化)。

经过数据集训练, 那些重要特征的权重就会较大, 对输出的影响也会增强, 从而提高模型对包装盒缺陷的检测能力。

2.3 ROIAlign 结合 ECA 注意力机制

2.3.1 ROIAlign 替换 ROI Pooling

为解决 ROI Pooling 2 次量化导致缺陷定位出现误差的问题, 本文利用 ROIAlign 代替 ROI Pooling, 其原理如图 7 所示。

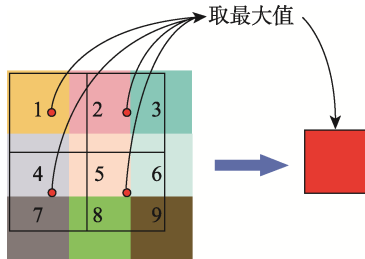


图 7 ROIAlign 原理
Fig.7 Principle of ROIAlign

当 Proposal 投影回 BiFPN 输出的特征图上时, 由于缩放原因, 部分 Proposal 的角点坐标不是整数, 此时并不量化取整; 接着将 Proposal 的投影结果平均划分为 7×7 的子区域, 并在各区域中心设定 4 个点; 同时利用双线性差值计算中心点像素值并取最大像素值为结果, 从而有效防止因随意量化而引起的缺陷定位出现误差的问题。

2.3.2 添加 ECA 注意力机制

仅依靠 ROIAlign 并不能充分捕捉图像中的关键信息, 尤其是在处理包装盒表面缺陷背景复杂的情况时。ECA 注意力机制作为 SENet 的改良版, 不同于需要很多计算资源的 CA 以及 CBAM 注意力机制, 同时其摒弃了 SENet 通过全连接实现通道之间信息传递的方式, 利用计算简单的一维卷积使模型快速发现重要通道, 从而实现对关键特征的增强和突出。ECA 注意力机制原理如图 8 所示。

通过给定的通道数 C , 可自适应地确定卷积核 K 的大小, 随后利用全局平均池化把 $H \times W$ 的平面转为

1×1 大小, 接着利用大小为 $C \times 1 \times 1$ 的卷积对重要通道赋予较大权重并进行 sigmoid 激活和调整各个通道的权重, 最后再扩充回原本 $H \times W$ 大小的维度。其中 K 的计算方法如式 (7) 所示。

$$K = \psi(C) = \left\lceil \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right\rceil_{\text{odd}} \quad (7)$$

式中: $\gamma=2$; $b=1$; $|A|_{\text{odd}}$ 为最接近 A 的奇数。

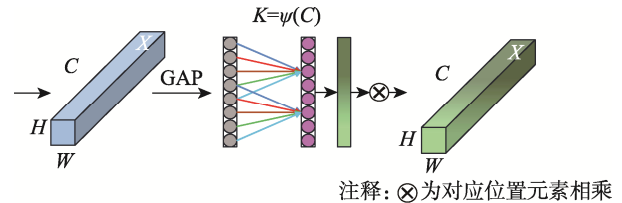


图 8 ECA 注意力机制原理
Fig.8 Principle of ECA attention mechanism

3 实例验证

3.1 数据集构建

自动装盒设备包装过程主要分为开盒、推料装盒、折小耳、折下页、喷胶、折上页以及装盒成型等, 这个过程中原材料印刷不足、机械振动、传送碰撞等非正常因素会导致产品表面出现多种缺陷。本文选择其中 4 类缺陷作为检测目标, 分别为原材料印刷问题造成的污点与露白 2 种缺陷, 开盒到装盒成型过程以及产品运输中出现的划痕与折痕 2 种缺陷。图 9 展示了部分不同包装盒上的缺陷种类。

通过 Labelimg 标注所采集的绿豆糕包装盒缺陷图像中的缺陷并划分缺陷种类, 随后每张图像都会生成各自的 xml 文件, 其包含了图像内缺陷的种类, 边框左上角及右下角坐标和边框的高度、宽度等关键信息; 随后通过离线增强增加图像数量。数据集共有 4 980 张图像, 大小为 640 像素 $\times$ 480 像素, 其中各类缺陷图像数量大体相同; 数据集内 80% 的图像作为训练集, 剩余 20% 均分为验证集和测试集。

3.2 实验参数配置及评价指标

3.2.1 实验参数配置

改进模型及实验对比模型均在 AutoDL 的云端服务器上训练, 中央处理器 (CPU) 为 Intel Xeon(R) Gold 6330, 2 GHz, 显卡为 Nvidia Geforce RTX 3090(24 GB), 框架镜像及语言版本包括 PyTorch 2.0.0、Cuda 11.8、Python 3.8。

模型的训练参数: batchsize 为 2, 初始学习率为 0.003, 权重衰减系数为 0.0001, 动量系数为 0.9, epoch 数为 40; 预处理 resize 时限制的最小与最大尺寸为 800 像素与 1 333 像素; RPN 进行 nms 处理的 IoU 阈值为 0.7, RPN 在 nms 处理后保留的候选区域个数为



注：WD 为污点，ZH 为折痕，HH 为划痕，LB 为露白。

图 9 绿豆糕包装盒缺陷种类

Fig.9 Defect types of mung bean cake packaging box

2 000; RPN 训练及计算损失时筛选的 Proposal 样本个数为 256, 正样本占总样本的比例为 0.5; Faster R-CNN 计算损失时采样的样本数为 512, 正样本占有所有样本的比例为 0.25。

在训练本文改进模型时, 首先利用 Pytorch 官方提供的 Swin Transformer V2-T 在大型数据集 ImageNet 1k 上的预训练权重进行迁移训练; 具体为先冻结主干的权重, 然后对改进后模型在 COCO2017 数据集进行预训练, 得到 BiFPN、RPN、ECA 以及其他模块的权重, 随后解冻主干的权重, 再把整体模型放入本文的绿豆糕缺陷数据集中进行训练。

3.2.2 评价指标

本文选择准确率 (Precision, P)、召回率 (Recall, R)、平均精确率 (Average Precision, AP) 以及平均精确率均值 (mean Average Precision, mAP) 作为算法的评价指标。其中, mAP 选择 $mAP^{0.5}$ 、 $mAP^{0.75}$ 和 $mAP^{0.5:0.95}$ 作为评价指标; $mAP^{0.5}$ 、 $mAP^{0.75}$ 为预测结果边界与真实边界的 IoU 阈值达到 0.5 或 0.75 的 mAP ; $mAP^{0.5:0.95}$ 为预测结果边界与真实边界的 IoU 阈值达到 0.5:0.95 的平均 mAP , 即在 0.5~0.95, 每 0.05 一个档位上的所有 mAP 的平均值; $mAP^{0.75}$ 和 $mAP^{0.5:0.95}$ 通常用来体现模型对小目标的检测能力。

3.3 结果与分析

3.3.1 消融实验

为深入探究改进后的 Faster R-CNN 模型中不同改进模块对模型整体性能的影响, 在试验环境等因素相同的情况下, 对每个模块单独测试, 表 1 展示了不同改进模块对 Faster R-CNN 改进的具体贡献。表 1 中, 模型 1 为原始 Faster R-CNN, 模型 2 为 Faster R-CNN+Swin Transformer V2-T, 模型 3 为 Faster R-CNN+BiFPN (取 VGG16 的 3~5 Block 的特征图作为 BiFPN 的输入), 模型 4 为 Faster R-CNN+ROIAlign+ECA 注意力机制, 模型 5 为本文整体改进模型。

对比前 4 种模型, 添加 BiFPN 后对模型综合检测性能提升最大, 同时折痕的 AP 较原模型提高约 7.88 百分点; 替换主干为 Swin Transformer V2-T 时对露白缺陷的检测能力提升最高, AP 比原模型高约 10.30 百分点; ROIAlign 结合 ECA 注意力机制对模型在划痕、污点、露白 3 种小缺陷的检测效果上均有提升作用; 虽然检测速率略有降低, 但整体改进模型对 4 种缺陷的 AP 较改进前显著提高 19.66、12.96、14.56 和 18.86 百分点, 同时 $mAP^{0.5}$ 也提高 15.76 百分点。

表 1 消融实验对比

Tab.1 Comparison of ablation experiments

模型	$AP/\%$				$P/\%$	$R/\%$	$mAP^{0.5}/\%$	检测速度/(帧·s ⁻¹)
	划痕	折痕	污点	露白				
1	74.98	82.61	79.86	72.92	76.12	74.54	77.59	19.47
2	86.72	87.82	85.68	83.22	84.94	83.18	85.86	17.26
3	90.63	90.49	86.45	81.27	87.07	85.89	87.21	14.84
4	77.34	81.37	82.25	74.02	77.66	75.28	78.75	19.78
5	94.64	95.57	94.42	91.78	92.45	91.98	93.35	13.54

3.3.2 不同模型测试对比

为了进一步体现本文改进模型的具体效果, 在所有预训练与训练数据集及环境配置相同的情况下, 将本文训练好的检测模型与 YOLO V4-tiny 和 YOLOv5s 模型进行对比, 结果如图 10 与表 2 所示。图 10 与表 2 中, 模型 1 为 YOLOv4-tiny, 模型 2 为 YOLOv5s, 模型 3 为本文改进模型。

由图 10 和表 2 可以得出, 本文改进模型的检测效果最佳, YOLOv4-tiny 整体效果最差。其中本文改进模型对绿豆糕包装盒缺陷的 $mAP^{0.5:0.95}$ 和 $mAP^{0.75}$ 比 YOLOv5s 高约 11.49、11.17 百分点, 可见其对小目标缺陷的检测能力更强, 同时检测速度可达每秒 13 张缺陷图像, 满足工业生产需求。

为使对比结果更加直观, 图 11 从测试集中挑选了涵盖多种缺陷的图像进行测试对比。

对于第 1 张图像中的缺陷, YOLOv4-tiny 漏检 2 个缺陷、YOLOv5s 漏检 1 个缺陷、本文改进模型无漏检与误检;对于第 2 张图像中的缺陷, YOLOv4-tiny

漏检与误检各 1 个、YOLOv5s 漏检与误检各 1 个、本文改进模型漏检 1 个。综上可知, 本文的改进算法较其他 2 种模型可更加准确无误地检测到绿豆糕包装盒上的缺陷。

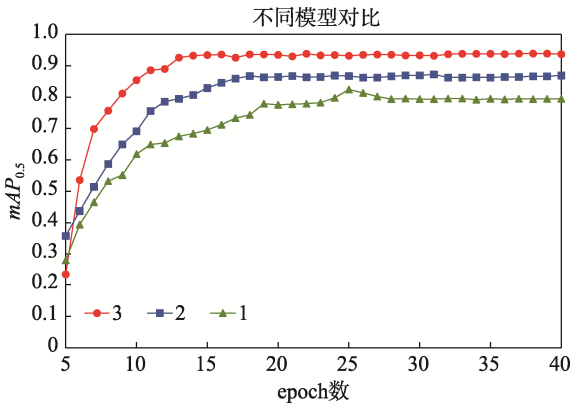


图 10 不同模型测试效果对比
Fig.10 Comparison of testing effects of different models

表 2 不同模型性能对比
Tab.2 Performance comparison of different models

模型	AP/%				$mAP^{0.5}/\%$	$mAP^{0.5:0.95}/\%$	$mAP^{0.75}/\%$	检测速度/(帧·s ⁻¹)
	划痕	折痕	污点	露白				
1	83.46	84.27	79.47	72.32	79.88	47.32	45.56	41.49
2	90.92	88.36	85.34	82.97	86.90	50.48	48.18	50.24
3	94.64	95.57	94.42	91.78	93.35	61.97	59.35	13.54



图 11 不同模型测试效果对比
Fig.11 Comparison of test effects of different models

4 结语

在迫切需要新型食品包装盒表面缺陷检测方法的背景下, 针对包装盒表面缺陷的背景复杂和缺陷纹理细小难以检测等问题, 本文改进后的 Faster R-CNN 对所选绿豆糕包装盒的 4 种缺陷具有极高的检测准确性; 不足的是, 尽管改进后的模型的检测速率可满足工业生产的需要, 但相较于其他算法, 该模型在检测速度方面尚有优化的空间, 后续有待研究稳定缺陷检测准确率的同时确保缺陷检测速率的提升。

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.
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.
[2] 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, China. IEEE, 2022: 1-7.
- [3] 马志刚, 赵志强. 基于机器视觉的包装品质检测系统设计[J]. 包装工程, 2022, 43(21): 193-197.
- MA Z G, ZHAO Z Q. Design of Packaging Quality Inspection System Based on Machine Vision[J]. Packaging Engineering, 2022, 43(21): 193-197.
- [4] LIU X Y, LI Y, GUO Y Y, et al. Printing Defect Detection Based on Scale-Adaptive Template Matching and Image Alignment[J]. Sensors, 2023, 23(9): 4414.
- [5] GAO C R. Design of Tourism Package with Paper and the Detection and Recognition of Surface Defects - taking the Paper Package of Red Wine as an Example[J]. Journal of Intelligent Systems, 2021, 30(1): 720-727.
- [6] 刘玉洪, 吴一全. 基于机器视觉的太阳能电池片缺陷检测算法综述[J]. 光学精密工程, 2024, 32(6): 868-900.
- LIU Y Q, WU Y Q. Review of Defect Detection Algorithms for Solar Cells Based on Machine Vision[J]. Optics and Precision Engineering, 2024, 32(6): 868-900.
- [7] ADEM K, KÖZKURT C. Defect Detection of Seals in Multilayer Aseptic Packages Using Deep Learning[J]. Turkish Journal of Electrical Engineering & Computer Sciences, 2019, 27(6): 4220-4230.
- [8] KIM K, KIM H, CHUN J, et al. Real-Time Anomaly Detection in Packaged Food X-Ray Images Using Supervised Learning[J]. Computers, Materials & Continua, 2021, 67(2): 2547-2568.
- [9] ROSLAN M I B, IBRAHIM Z, AZIZ Z A. Real-Time Plastic Surface Defect Detection Using Deep Learning[C]//2022 IEEE 12th Symposium on Computer Applications & Industrial Electronics (ISCAIE). Penang, Malaysia. IEEE, 2022: 111-116.
- [10] 方宇伦, 陈雪纯, 杜世昌, 等. 基于轻量化深度学习 VGG16 网络模型的表面缺陷检测方法[J]. 机械设计与研究, 2023, 39(2): 143-147.
- FANG Y L, CHEN X C, DU S C, et al. Surface Defect Detection Method Based on Lightweight Deep Learning VGG16Net[J]. Machine Design & Research, 2023, 39(2): 143-147.
- [11] 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.
- [12] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [13] 叶宇星, 孙志锋, 马风力, 等. 基于改进 YOLOv5s 的腌制蔬菜真空包装缺陷检测[J]. 包装工程, 2023, 44(9): 45-53.
- YE Y X, SUN Z F, MA F L, et al. Vacuum Packaging Defect Detection of Pickled Vegetables Based on Improved YOLOv5s[J]. Packaging Engineering, 2023, 44(9): 45-53.
- [14] HU J, SHEN L, SUN G. Squeeze-and-Excitation Networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. IEEE, 2018: 7132-7141.
- [15] LIU Z, HU H, LIN Y T, et al. Swin Transformer V2: Scaling up Capacity and Resolution[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA. IEEE, 2022: 11999-12009.
- [16] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and Efficient Object Detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 10778-10787.
- [17] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE, 2017: 2980-2988.
- [18] REN S Q, HE K M, 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(6): 1137-1149.
- [19] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 11531-11539.
- [20] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada. IEEE, 2021: 9992-10002.