

基于改进 YOLOv5 的药用胶囊表面缺陷检测方法

武泽坤, 叶晓娴, 陈梦

(上海工程技术大学 艺术设计学院, 上海 201620)

摘要: **目的** 在质检过程中精确快速地检测到药用空心胶囊的表面缺陷。**方法** 基于 YOLOv5 算法, 针对模型网络参数量大和对长距离依赖关系的学习能力较弱的问题, 提出在主干网络部分引入 GhostNet 模块和坐标注意力机制, 使网络有效捕捉数据位置信息和通道信息的关系。**结果** 实验结果表明, 改进的网络结构能够在参数量下降为原来的 57% 的前提下, 对药用胶囊表面的破损、印刷错误、孔洞、划痕、凹陷等 5 类缺陷的平均检测精度达到 96.9%, 相较于 YOLOv5s 提高了 2.4 个百分点, 检测速度提升了 12 帧/s。**结论** 文中方法能够有效对药用胶囊表面缺陷进行分类和定位, 提高缺陷检测的准确率。

关键词: YOLOv5; 胶囊; 缺陷检测; 注意力机制; GhostNet

中图分类号: TB487 **文献标识码:** A **文章编号:** 1001-3563(2022)23-0297-08

DOI: 10.19554/j.cnki.1001-3563.2022.23.035

Surface Defect Detection Method for Pharmaceutical Capsules Based on Modified YOLOv5

WU Ze-kun, YE Xiao-xian, CHEN Meng

(School of Art and Design, Shanghai University of Engineering Science, Shanghai 201620, China)

ABSTRACT: The work aims to detect the surface defects of pharmaceutical hollow capsules in quality inspection accurately and quickly. Based on YOLOv5 algorithm and aiming at the problems of large amount of model network parameters and weak learning ability of long-distance dependence, GhostNet module and Coordinate attention mechanism were introduced into the backbone network to make the network effectively capture the relationship between data location information and channel information. The experimental results showed that the improved network structure could accurately detect five kinds of defects such as damage, printing error, hole, scratch and depression on the surface of pharmaceutical capsule on the premise of decreasing to 57% of the original parameters. The average accuracy of each defect was 96.9%, which was increased by 2.4 percentage points. The detection speed was increased by 12 FPS. The proposed method can effectively classify and locate the surface defects of pharmaceutical capsules, and improve the accuracy of defect detection.

KEY WORDS: YOLOv5; pharmaceutical capsules; defect detection; attention mechanism; GhostNet

在我国社会老龄化形势加剧的今天, 每年药品的生产量不断上升。这对药品质量检测精度和速度提出更高的要求。药用胶囊作为被广泛使用的药物之一, 随着生产规模的不断扩大, 其生产过程中存在的表面缺陷问题时有发生, 缺陷检测成为生产环节中的一大难点。

药用胶囊常见的表面缺陷有尺寸不正确、孔洞、凹陷、印刷缺失。目前, 对药物生产的缺陷检测工作主要使用人工方式进行, 但人工检测存在效率低、误检率高和漏检率高等问题, 难以满足日益扩大的药物生产规模需求。

为了在保障药品生产质量的同时降低人工成本,

收稿日期: 2022-04-24

作者简介: 武泽坤 (1997—), 男, 硕士生, 主攻计算机视觉。

通信作者: 叶晓娴 (1984—), 女, 博士, 教授, 主要研究方向为影像档案修复、视觉与媒体设计等。

主干网络由 focus 模块、CSP 模块组成, 对输入图像数据进行特征提取。focus 模块对输入图像数据进行切片操作, 将输入数据的高和宽缩减一半, 通道数增加为原始通道数的 4 倍, 将空间信息转变为通道信息, 减少浮点运算量。YOLOv5 提供 4 种不同大小的网络结构按照从小到大排序依次为 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x。其中 YOLOv5s 网络体量最小, 推理速度快, 但检测精度不高, 适用于运算能力较低的设备, 比如树莓派设备。在服务器设备上一般选用 YOLOv5l 网络, 其检测精度较高、推理速度适中。CSP 模块将基础层的特征映射分为两部分, 通过跨阶段层次结构将特征映射合并, 在减少计算量的前提下保证准确率。

特征融合网络采组成为 FPN (Feature Pyramid Networks) 和 PAN (Path Aggregation Network) 模块, 加强网络对不同尺度对象特征融合的能力。FPN 模块自上向下传递强语义特征, PAN 模块自底向上传递强定位特征, 两者集合加强网络对小目标对象的检测能力。

YOLOv5 的输出端相较于 YOLOv4 的做出改变, 采用 CIOU 函数作为边界框回归的损失函数, CIOU 函数将重叠面积、中心点距离、宽高比综合考虑, 使得检测框更加准确。

1.2 注意力机制

20 世纪 90 年代在视觉图像领域提出注意力机制, 这是一种模仿人类注意力所实现的信息处理资源的高效分配方式^[9]。

当人进入某一场景时, 视觉会优先停留在一些特殊的点, 剩余的场景信息会被暂时忽略。比如人在图像中在寻找具有某一特征的事物时, 会更多的将视线停留在符合这一特征的图像区域, 这即是注意力的有效分配。注意力作为人类大脑中一项必不可少的认知功能, 保证人类通过五感获取海量信息的同时, 筛选出部分有用的信息进行处理。在深度学习网络模型中, 注意力机制对重要信息添加高权重, 对次要信息添加低权重, 增强网络对关键信息的特征提取能力。

注意力机制的优点如下。

- 1) 结构简单, 适用于多种任务, 如机器翻译、目标检测、对话系统、数据监测等。
- 2) 通过建立循环来构建输入与输出间的依赖关系和并行化的方式提高运行速度。
- 3) 增强了网络对长距离依赖关系的学习能力, 避免为增强模型的表达能力而增加网络深度所造成的梯度爆炸现象。

2 胶囊缺陷检测模型

为简化模型的参数量和计算量, 强化其对于空间信息特征的提取能力。文中重点对 YOLOv5 的特征提取网络进行重新设计, 将原网络中的所有 Conv 模块和 BottleneckCSP 模块替换为 GhostConv 模块和 GhostBottleneck 模块, 并在主干部分的 GhostConv 模块后嵌入坐标注意力模块以提升网络的检测精度。改进后的网络结构见图 2。

2.1 坐标注意力

目标检测网络中常用的卷积运算对全局关系的捕获能力较弱, 当图像中出现连续缺陷的痕迹时, 缺陷长度越大, 其特征就越难表征, 为解决此问题, 将注意力机制引入检测模型中。注意力机制通过对人类视觉中的选择性特点进行模仿, 从输入中筛选出当前任务所需的关键信息, 通过对关键信息进行加权, 增强模型对目标的捕捉能力。注意力机制可有效提高模型在目标检测任务中的性能。

文中将注意力机制引入到 YOLOv5 网络中^[10], 增强模型网络捕捉位置信息和通道信息关系的能力, 提高缺陷检测的精确度。在选择注意力模块时必须考虑将其加入后对网络参数量和运算复杂度的影响, 对文中的研究对象来说, 结构复杂的注意力机制, 如自注意力 (Self-Attention, SA)^[11], 在大幅加深网络结构的基础上, 降低模型的推理速度, 不适合在工业缺陷检测模型中使用。轻量型的注意力模块在略微增加参数量和运算复杂度的前提下提高模型的检测精度, 适合在模型中嵌入使用。Hu 等^[12]提出通道注意力 (Squeeze-and-Excitation, SE) 模块, 针对通道关系进行建模, 使用全局池化, 避免普通卷积在通道关系特征提取上的信息丢失, 增强模型对输入数据通道信息的敏感度, 使模型在分类任务中有更好的表现, 但是在全局池化编码通道信息的同时, 空间信息被压缩为一维张量, 这将导致空间信息的丢失。CBAM (Convolution Block Attention Module)^[13]在 SE 注意力的基础上增加空间特征的获取, 该注意力的处理方式是对经过通道注意力模块加权后的特征图的所有通道进行最大池化和平均池化, 汇聚空间信息, 再通过 1×1 卷积和 Sigmoid 激活函数得到空间注意力特征, 但在进行卷积操作时, 也弱化了网络对长距离依赖关系的学习能力。

大部分轻量级注意力模块在构筑时, 仅对通道关系进行建模来重新权衡其重要性, 忽略了位置信息。坐标注意力^[14]模块由坐标信息嵌入和坐标注意力生成等 2 个部分组成, 坐标注意力结构见图 3。

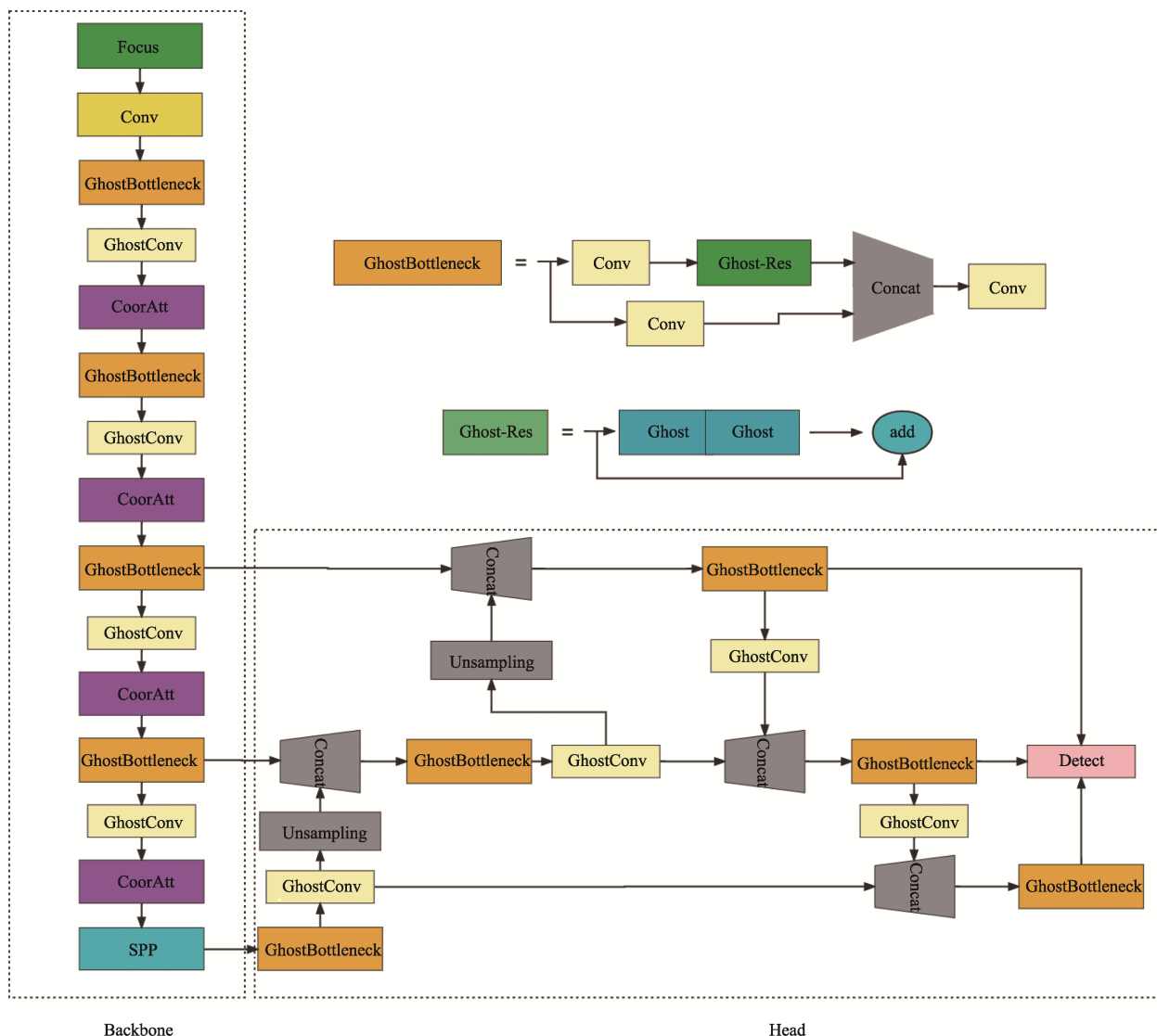


图2 YOLOv5-CA-Ghost 网络结构
Fig.2 Architecture diagram of YOLOv5-CA-Ghost

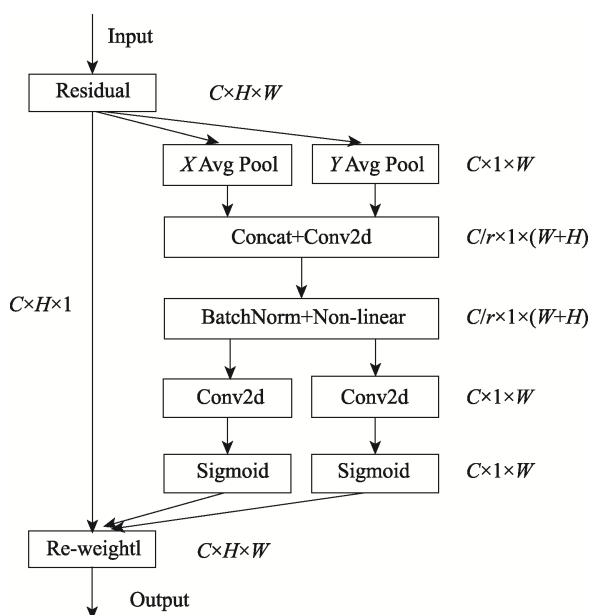


图3 坐标注意力结构
Fig.3 Coordinate attention structure

在通道注意力中,全局池化用于对全局空间信息进行编码,同时空间信息也被压缩到通道信息中,位置信息难以保存,在目标检测任务中位置信息至关重要。坐标注意力模块分解全局池化转化为2个方向上的特征编码操作。有输入特征图,使用池化核尺寸为 $(H,1)$ 、 $(1,W)$,分别沿着水平方向和垂直方向对各通道进行编码,则第 c 通道在高度 h 上的输出为:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq j \leq W} X_c(h, j) \quad (1)$$

式中: $z_c^h(h)$ 为第 c 通道在高度 h 处的输出; W 为特征图的宽度; $X_c(h, j)$ 为输入特征图 X_c 在 (h, j) 坐标上的数值。

同样的,第 c 通道在宽度 w 上的输出为:

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq i \leq H} X_c(i, w) \quad (2)$$

式中: $z_c^w(w)$ 为第 c 通道在宽度 w 处的输出; H

为特征图的高度; $X_c(i, w)$ 为输入特征图 X_c 在 (i, w) 坐标上的数值。

以上 2 种变换沿着水平方向和垂直方向聚合特征, 得到一对方向感知的特征图。这种转换方式与 SE 模块通过全局池化获取单一特征向量不同。具体而言, 坐标注意力模块更容易捕捉到沿着一个方向的长期依赖关系, 并保存另一方向上的位置信息, 这正是缺陷检测任务所需的。

为有效利用坐标注意力嵌入中的式 (1) 和 (2) 得到的全局感受野与其位置信息的表征, 设计坐标注意力生成的过程。在完成信息嵌入后, 该过程将得到的 2 个方向上的输出进行拼接, 将拼接后的张量送入 1×1 的卷积函数 F_1 中进行计算:

$$f = \delta(F_1[z^h, z^w]) \quad (3)$$

式中: f 为对水平方向和垂直方向进行编码的中间特征映射, $f \in R^{c/r \times (H+W)}$; r 为控制模块大小的超参数, 缩小 f 的通道数, 降低模型的计算复杂度和参数量; δ 为激活函数; $[\cdot, \cdot]$ 为张量在空间维度上的拼接操作。将中间特征映射 f 分解为 2 个单独张量 $f^h \in R^{c/r \times H}$ 和 $f^w \in R^{c/r \times W}$, 并利用 2 个 1×1 卷积函数 F_h 和 F_w , 将 f^h 和 f^w 转换为通道数相同的张量输出到 X , 见式 (4) 和式 (5)。

$$g^h = \sigma[F_h(f^h)] \quad (4)$$

$$g^w = \sigma[F_w(f^w)] \quad (5)$$

其中, σ 为 simoid 激活函数, 扩展输出 g^h 和 g^w 作为坐标注意力权重, 则在第 c 通道上坐标 (i, j) 的注意力权重 Y 为:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (6)$$

2.2 GhostNet

在缺陷检测任务中, 部分特征图包含较多的背景信息, 这使得常规卷积操作中产生大量重复的特征图。冗余的中间特征映射对缺陷检测的精度提升有

限, 同时增大对计算资源的消耗。

冗余的特征图一般由常规卷积产生, 针对缺陷数据集的特点, 使用 Ghostnet 模块^[15], 减少模型参数量和计算复杂度, 加快检测速度。Ghostnet 模块结构如图 4 所示, Ghostnet 卷积过程为以下 3 步。

1) 常规卷积。首先使用普通卷积得到本征特征图。卷积操作和运算量表示为:

$$Y = X \times f + b \quad (7)$$

$$F_1 = n \times h' \times w' \times c \times k \times k \quad (8)$$

式中: $X \in R^{c \times h \times w}$ 为卷积中通道为 c 、高度为 h 、宽度为 w 的输入特征图; $Y \in R^{h' \times w' \times n}$ 为 n 个高度为 h' 、宽度为 w' 的输出特征图; $f \in R^{c \times k \times k \times n}$ 为 $k \times k$ 大小的卷积核; b 为偏置项。式 (8) 为计算量。

2) Ghost 卷积生成。将通过常规卷积操作得到的本征特征图使用线性变换 Φ_i 生成相似特征图:

$$Y_{ij} = \Phi_{ij}(Y_i'), i \in [1, m], j \in [1, s] \quad (9)$$

$$n = m \times s \quad (10)$$

式 (9) 中: Y_i' 为本征特征图的第 i 个特征图; Φ_{ij} 为第 i 个特征图进行的第 j 个线性变换的线性变换函数; Y_{ij} 为 $d \times d$ 大小的操作和经线性变换 Φ_{ij} 生成的第 j 个特征图。式 (10) 中: n 为输出特征图的数量; m 为本征特征图的数量; s 为变换数。

由于 Ghost 模块中存在一个恒等映射过程, 则线性运算的数量和计算量分别为:

$$m \times (s-1) = \frac{n}{s} \times (s-1) \quad (11)$$

$$F_2 = m \times h' \times w' \times c \times k \times k + (s-1) \times \frac{n}{s} \times h' \times w' \times d \times d \quad (12)$$

3) 特征图拼接。将本征特征图和本征特征图线性变换得到的特征图拼接。

$d \times d$ 与 $k \times k$ 大小相似, 且 $s \ll c$, 则普通卷积与 Ghost 卷积的计算量比值为:

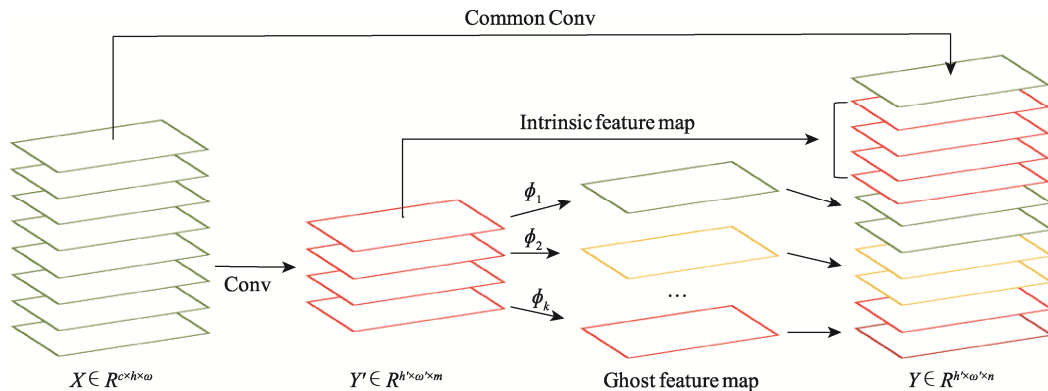


图 4 GhostNet 模块原理
Fig.4 Principle of GhostNet module

$$r_s = \frac{n \times h' \times w' \times c \times k \times k}{m \times h' \times w' \times c \times k \times k + (s-1) \times \frac{n}{s} \times h' \times w' \times d \times d} = \frac{n \times c \times c \times k \times k}{\frac{1}{s} \times c \times k \times k + (s-1) \times \frac{s-1}{s} \times d \times d} \approx \frac{s \times c}{s+c-1} \quad (13)$$

使用 Ghostnet 模块作为 Backbone 网络部分的卷积层, 简化模型网络的结构, 减少模型的参数量和计算量, 利于模型部署。

3 实验结果与分析

3.1 实验环境配置

实验环境为 Ubuntu16.04 LTS64 位操作系统、Intel®Core™i9-7700HQCPU@3.7 GHz×10 核 CPU, GPU 为 QuadroRTX500016GB, 所使用深度学习框架为 PyTorch 1.7, cuda 10.2 加速。

模型训练使用 Adam 优化器, 输出图片尺寸设置为 640 像素×640 像素, Batchsize 设置为 16, 总训练批次 (epoch) 为 300, 初始学习率为 0.001。数据集中共 2 276 张图片, 包括 5 种缺陷和合格产品。将 80% 的数据集作为训练集, 10% 的数据集作为验证集, 余下 10% 作为测试集。

3.2 数据预处理

模型训练使用的数据来自于 MVTEC AD 数据集^[16], 该数据集用于对常见的生活用品进行缺陷检测, 它包含超过 5 000 张高分辨率图像, 分为 15 个不同的对象和纹理类别, 文中选择其中的胶囊缺陷数据集。图片的分辨率为 1 000 像素×1 000 像素, 表面缺陷分别为印刷错误、孔洞、划痕、凹陷、裂纹。

原始数据集中的缺陷产品数据量为 350 张, 合格产品数据量为 219 张, 各类缺陷数据在训练过程中容易被正样本稀释, 影响模型对合格和缺陷 2 类数据的检测精度, 因此, 在训练前使用 Opencv 库将缺陷数据进行数据增强, 分别对缺陷数据进行旋转、翻转、对比度增强。经过数据增强后的图像见图 5。模型训练前对原始数据进行数据增强, 是增强模型泛化能力, 提高鲁棒性的有效方法。

3.3 评价指标

实验选用平均精度 (Average Precision, AP) 和平均精度均值 (Mean Average Precision, mAP) 作为评价指标, 具体的公式为:

$$P = \frac{T_p}{T_p + F_p} \quad (14)$$

$$R = \frac{T_p}{T_p + F_N} \quad (15)$$

$$A_{mAP} = \frac{\sum A_p}{N} \quad (16)$$

式中: T_p 为被检测为正类的正类; F_p 为被检测为正类的负类; F_N 为被检测为负类的正类。平均精度由 P-R 曲线下的面积表示, 平均精度均值由平均精度和总类别数的比值得到。



图 5 数据增强

Fig.5 Data augmentation

3.4 实验结果对比

文中在 YOLOv5s 网络中加入坐标注意力机制和 Ghost 模块对其进行改进, 经过 300 轮训练, 各模型的平均精度均值见图 9, 可以看到文中模型的收敛速度要快于其他 3 个模型, 通过消融实验对 2 种模块在药用胶囊表面缺陷数据集上进行验证^[15], 分别对比两模块对模型网络的影响。

表 1 2 种方法的消融实验分析

Tab.1 Analysis of ablation experiments of two methods

方法	参数量/ 10^6	平均精度均值/%
YOLOv5s	7.0	95.3
YOLOv5s-CA	7.4	95.9
YOLOv5s-Ghost	3.7	95.2
YOLOv5s-CA-Ghost	4.0	96.9

如表 1 所示, 嵌入坐标注意力模块和替换普通卷积为 Ghost 卷积后, 参数量大幅下降的同时, 检测精度提升。在嵌入坐标注意力模块时, 平均精度均值提高了 0.6%; 替换常规卷积为 Ghost 卷积时, 参数量下降了 52.8%, 平均精度均值仅下降了 0.1%。通过嵌

入坐标注意力模块, 增强了 YOLOv5 模型网络对位置特征的获取能力, 通过替换普通卷积为 Ghost 卷积, 减少了重复的特征信息, 模型参数量减少为原先的 57%。实验数据表明, 同时使用 2 种模块, 平均精度均值提升了 1.6%, 参数量下降了 42.8%, 达到轻量化

模型和提升检测精度的目的。
为更好地验证改进后 YOLOv5 网络性能, 在实验配置不变的条件下, 分别在 YOLOv4、Faster RCNN、SSD 模型对相同数据集进行训练与推理, 表 2 为各模型在相同数据集下的表现。

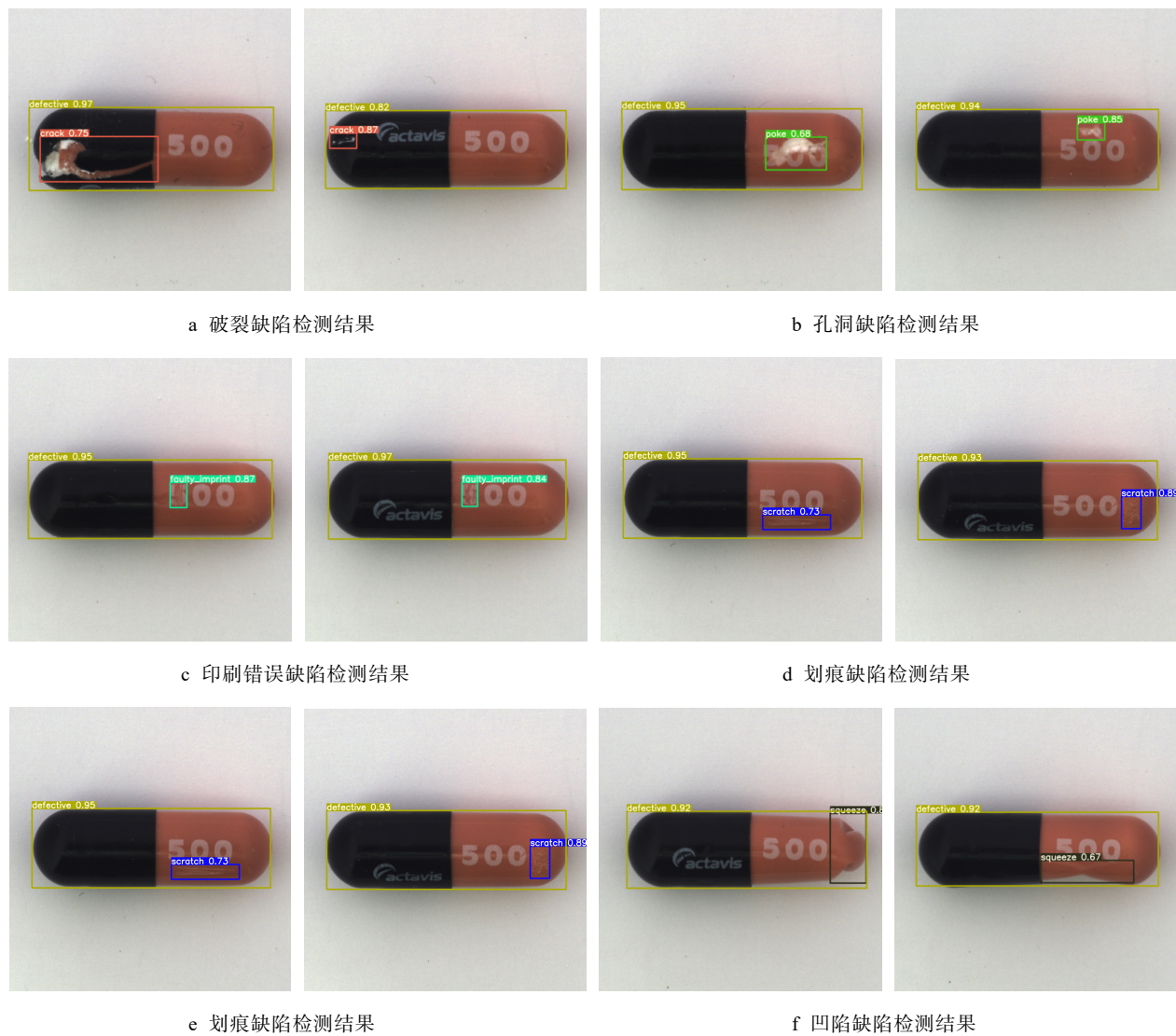


图 6 检测效果
Fig.6 Detection effect

表 2 各模型在数据集上的结果
Tab.2 Results of each model on dataset

模型	AP 值/%							mAP 值/%	速度/(帧·s ⁻¹)
	合格	缺陷	破裂	印刷错误	孔洞	划痕	挤压		
YOLOv4	98.5	95.4	90.2	97.6	96.5	91.3	94.8	94.9	53
Faster RCNN	93.1	90.2	80.8	84.3	89.3	81.2	86.6	86.5	47
SSD	90.2	87.6	79.6	78.5	82.3	76.5	82.8	82.5	39
文中模型	99.5	96.0	90.4	99.5	99.5	99.5	93.9	96.9	80

由表 2 可知,轻量化模型后的平均精度均值达到 96.9%, 对比 YOLOv4、Faster RCNN 和 SSD 提升了 2.0%、10.1%、14.4%, 且改进后模型在推理速度上高达 80 帧/s, 也优于其他 3 个模型, 较其他 3 个模型分别提升了 27、33、41 帧/s, 实验证明轻量化后的 YOLOv5 网络优于 YOLOv4、Faster RCNN 和 SSD 等 3 种网络。

4 结 语

文中针对药用胶囊表面缺陷在生产过程的自动化检测, 提出了一种 YOLOv5-GhostNet-CA 轻量级网络模型。其中, 通过融合注意力机制, 结合 GhostNet 模块, 使特征图中的空间位置特征更容易被获取。在数据集上的实验结果证明, 文中的模型在参数量大幅下降的基础上, 提高了缺陷检测的准确率, 可达 96.9%; 缩短模型推理时间, 在检测的实时性上可满足工业缺陷检测的需求; 模型尺寸减小, 更容易部署至边缘计算设备。提高检测精度、降低模型的参数量、增强网络的泛化能力依旧是未来工业缺陷检测的重点。

参考文献:

- [1] KEKRE H B, MISHRA D, DESAI V. Detection of Defective Pharmaceutical Capsules and Its Types of Defect Using Image Processing Techniques[C]// 2014 International Conference on Circuits, Power and Computing Technologies, Nagercoil, India, 2014: 1190-1195.
- [2] 赖大虎, 黄宴委. 基于极端学习机的胶囊缺陷检测[J]. 福州大学学报(自然科学版), 2012, 40(4): 489-494.
LAI Da-hu, HUANG Yan-wei. Inspection for Defected Capsules Based on Extreme Learning Machine[J]. Journal of Fuzhou University (Natural Science Edition), 2012, 40(4): 489-494.
- [3] QI Dan-yang, JIANG Zheng. Capsule Defects Classification Based on Hierarchical Support Vector Machines[J]. Advanced Materials Research, 2014, 926-930: 3373-3378.
- [4] ZHOU Jun-lin, HE Jiao, LI Guo-li, et al. Identifying Capsule Defect Based on an Improved Convolutional Neural Network[J]. Shock and Vibration, 2020, 2020: 8887723.
- [5] LIU R, GU Q, WANG X, et al. Region-Convolutional Neural Network for Detecting Capsule Surface Defects[J]. Boletin Tecnico/Technical Bulletin, 2017, 55(3): 92-100.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, IEEE Recognition, 2016: 779-788.
- [7] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA, 2017: 6517-6525.
- [8] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[R/OL]. (2020)[2021-03-14]. <http://arxiv.org/pdf/2004.10934>.
- [9] TSOTSOS J K, CULHANE S M, KEI WAI W Y, et al. Modeling Visual Attention via Selective Tuning[J]. Artificial Intelligence, 1995, 78(1/2): 507-545.
- [10] 孙刘杰, 张煜森, 王文举, 等. 基于注意力机制的轻量级 RGB-D 图像语义分割网络[J]. 包装工程, 2022, 43(3): 264-273.
SUN Liu-jie, ZHANG Yu-sen, WANG Wen-ju, et al. Lightweight Semantic Segmentation Network for RGB-D Image Based on Attention Mechanism[J]. Packaging Engineering, 2022, 43(3): 264-273.
- [11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all You Need[C]// Advances in Neural Information Processing System, 2017: 5998-6008
- [12] HU Jie, SHEN Li, ALBANIE S, et al. Squeeze-and-Excitation Networks[C]//IEEE Transactions on Pattern Analysis and Machine Intelligence: 2011-2023.
- [13] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block Attention Module[C]// Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [14] HOU Qi-bin, ZHOU Da-quan, FENG Jia-shi. Coordinate attention for efficient mobile network design[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA. IEEE, 2021: 13708-13717.
- [15] HAN Kai, WANG Yun-he, TIAN Qi, et al. GhostNet: More Features from Cheap Operations[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 1577-1586.
- [16] BERGMANN P, BATZNER K, FAUSER M, et al. The MVTEC Anomaly Detection Dataset: A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection[J]. International Journal of Computer Vision, 2021, 129(4): 1038-1059.

责任编辑: 曾钰婵