

# 基于深度学习的图书破损检测方法研究

张志颖<sup>1†</sup>, 刘竞<sup>1,2</sup>, 许梦怡<sup>3</sup>, 郭剑明<sup>1</sup>

(1. 江苏大学科技信息研究所, 江苏 镇江 212013; 2. 江苏大学图书馆, 江苏 镇江 212013;

3. 海南大学, 海南 海口 570228)

**摘要:** 破损图书管理和维护是图书馆中的一项重要工作。在进行破损图书筛查时, 传统的人工筛查效率低下且易受主观因素影响, 难以满足现代图书馆的需求。因此, 提出了一种基于深度学习的图书破损检测模型。本文以YOLOv5为基础模型, 引入CBAM注意力机制、Meta-ACON激活函数和加权双向特征金字塔网络(BiFPN)对其进行改进。所提模型的mAP@0.5达到了94.89%, 其中破损图书类别的AP达到了94.59%, 与YOLOv5s相比提高了2.42%。该模型可以在不同背景下提供良好的检测性能, 有助于提高筛查破损图书的效率, 及时维护破损的图书。

**关键词:** 深度学习; 目标检测; YOLOv5; 破损图书检测; CBAM; Meta-ACON; BiFPN

**中图分类号:** TP399

**文献标识码:** A

## Research on Book Damage Detection Method Based on Deep Learning

ZHANG Zhiying<sup>1†</sup>, LIU Jing<sup>1,2</sup>, XU Mengyi<sup>3</sup>, GUO Jianming<sup>1</sup>

(1. Institute of Science and Technology Information, Jiangsu University, Zhenjiang, Jiangsu 212013, China;

2. Jiangsu University Library, Zhenjiang, Jiangsu 212013, China; 3. Hainan University, Haikou, Hainan 570228, China)

**Abstract:** The management and maintenance of damaged books has always been an important work in the library. In the detecting of damaged books, the traditional manual detecting is inefficient and susceptible to subjective factors, which makes it difficult to meet the needs of modern libraries. Therefore, we proposed a book damage detection model based on deep learning. We used YOLOv5 as a base model and improved it by introducing convolutional block attention module (CBAM), Meta-ACON activation function and bi-directional feature pyramid network (BiFPN). The mAP@0.5 of the proposed model reaches 94.89%. Specially, the AP of the damaged book reaches 94.59%, which is 2.42% higher than that of YOLOv5s. This model can provide good detection performance in different backgrounds, which helps to improve the efficiency of screening damaged books and timely maintenance of damaged books.

**Key words:** deep learning; object detection; YOLOv5; book damage detection; CBAM; Meta-ACON; BiFPN

随着图书馆藏书数量的不断增长, 对图书破损问题及时进行监控和维护愈发困难<sup>[1]</sup>。有效地检测和管理破损的书籍成为一个亟待解决的重要问题<sup>[2]</sup>。破损的图书不仅对读者的使用体验构成负面影响, 同时也损害了图书本身的文化和知识价

值。对于仍具备维护价值的破损图书, 及时的维护不仅可以降低维护成本, 还能延长使用寿命<sup>[3]</sup>。对于严重破损需要进行下架处理的图书, 如果未能及时处理, 不仅占用了宝贵的书架资源, 还会损害图书馆的形象, 不利于馆内空间的合理规划与资源优

化。但在面临庞大的馆藏时,维护人员往往需要耗费大量精力去发现破损图书。因此,利用深度学习技术自主检测破损图书具有重要意义。

近年来,深度学习在图书馆的应用场景逐渐扩大。Shi 等<sup>[4]</sup>提出了一种利用深度学习的 OCR 方法,以提高对图书条形码检测的准确性,从而更好地适用于图书分拣系统。Kalinina 等<sup>[5]</sup>针对书脊的识别进行研究,采用 YOLOv3 构建了书脊检测模型,通过识别图像中的书脊完成对图书的标识。Lei 等<sup>[6]</sup>提出了一种基于卷积神经网络的图书标签识别方法,他们通过图像分割技术提取图书标签上的字符,然后应用卷积神经网络训练分类器,以识别从图书图像中提取的书本信息。王红芳等<sup>[7]</sup>基于 YOLOv3-Tiny 实现对图书索书号区域的定位和识别,并根据索书号的排序判断是否错序。

综上所述,深度学习在图书管理领域取得了广泛应用,包括图书分拣、书脊识别、书名提取和错序检测等领域,并取得了一定的研究成果。但对于图书破损检测领域的研究还相对较少,目前该领域仍面临算法技术更新不及时的问题。破损图书检测对于图书馆维护与处理破损书籍具有重要意义,它可以接入图书馆的自助借还书系统中,在读者借还书的过程中对图书破损情况进行即时监控,还能对比借书和还书两次识别结果判断读者是否造成图书损坏。因此本文对破损图书检测展开研究,提出

了基于 YOLOv5 改进的图书破损检测模型,并引入了 CBAM 注意力机制<sup>[8]</sup>、Meta-ACON 激活函数<sup>[9]</sup>、BiFPN 加权双向特征金字塔网络<sup>[10]</sup>对模型进行优化。

## 1 材料和方法

### 1.1 数据集

为了构建一个高性能的图书破损检测模型,本研究在江苏大学图书馆、海南大学图书馆和安徽省图书馆进行图书样本采集,共获得 420 张破损图书样本和 400 张非破损图书样本。通过使用 Label-IMG 工具对这些样本进行标注,划分为破损图书(Damaged)和非破损图书(Undamaged)两个类别,建立了模型训练所需的数据集。为了提高模型的泛化性能,使其能够在复杂背景下准确识别出图书是否破损,本文对标注的数据集进行了数据增广。数据增广方式包括对图像进行放缩处理,以适应不同大小的尺寸;进行镜像和各种角度的旋转,以适应不同的拍摄角度;改变对比度来模拟不同亮度的环境;进行高斯模糊处理,以适应不同的分辨率。此外,我们还在图像中局部增加了阴影,以模拟图书局部被遮挡的情况。原始样本和增广样本效果如图 1 所示。



图 1 原始样本和增广样本

本研究通过上述数据增广的方法将数据集随机扩增了三次,最终获得 1680 张破损图书样本和 1600 张非破损图书样本。在训练过程中,本研究按照 7 : 2 : 1 的比例分配训练集、测试集和验证集,为了保证评价指标的严谨性,在分配数据集时确保了测试集和其他样本集的原始数据和增广数据均没有重叠,以保证模型得到充分训练和有效

验证。

### 1.2 图书破损检测模型

#### 1.2.1 YOLOv5

YOLO 是一种单阶段目标检测算法,在精度和速度上都具有良好的性能表现<sup>[11-13]</sup>。YOLOv5 的结构可分为输入、主干网络、颈部网络和预测层四个部分。YOLOv5 的主干网络为 CSPDarknet,

其作用是提取图片中的特征信息,由残差网络、Focus、CSP 和 SPP 等结构组成。YOLOv5 在颈部网络采用 FPN+PAN 结构,将低层的强定位信息与高层的强语义信息融合,得到更加有用的特征信息。最后在预测层得到三个不同尺度的特征图后,

通过计算输出图像中目标的位置、大小和类别等信息。本文以 YOLOv5 作为基础模型,引入 CBAM 注意力机制、Meta-ACON 激活函数和 BiFPN 双向特征融合对其进行改进,构成了本文所提出的图书破损检测模型,其网络结构如图 2 所示。

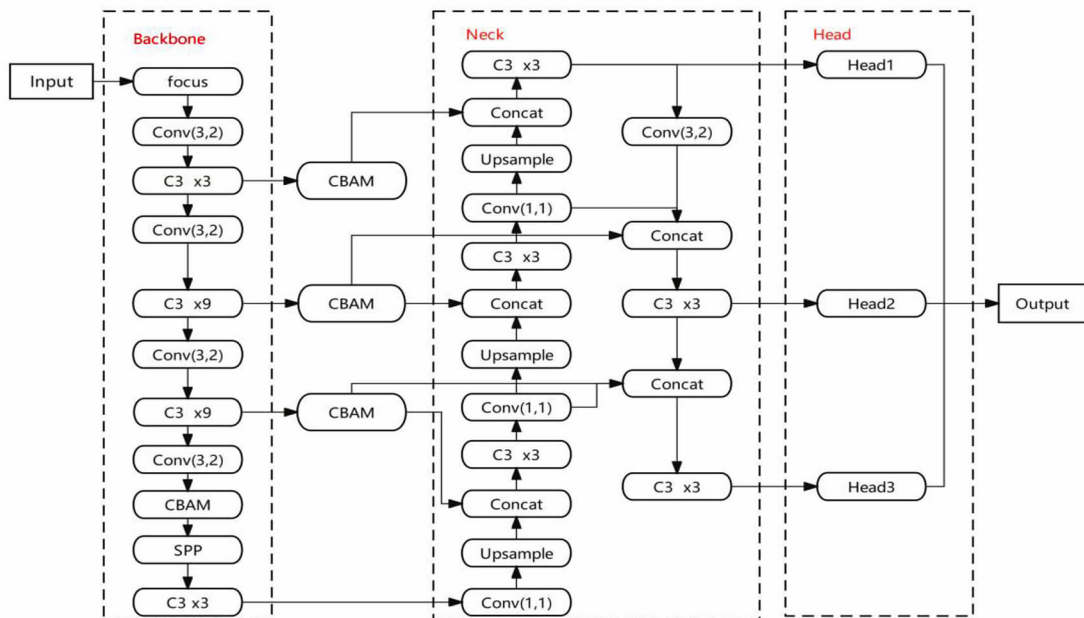


图 2 图书破损检测模型网络结构图

### 1.2.2 CBAM 注意力机制

在本研究中,通过引入 CBAM 注意力机制,在通道和空间两个维度依次推断注意力图,使模型聚焦于有价值的特征内容和位置信息。在通道注意力模块,输入的图书特征首先通过全局平均池化和最大池化在空间维度上进行压缩,再经过共享的全连接层计算输出通道权重向量,最后和输入的图书特征逐元素相乘输出通道注意力特征图。该模块计算了各个通道的重要性,让包含重要信息的通道获得更高的权重。通道注意力模块的计算过程如式(1)和式(2)所示。

$$M_c(F) = \sigma(\text{MLP}(F_{\text{avg}}^c) + \text{MLP}(F_{\text{max}}^c)) \quad (1)$$

$$F' = M_c(F) \otimes F \quad (2)$$

其中,  $F_{\text{avg}}^c$  和  $F_{\text{max}}^c$  分别表示池化操作得到的特征描述,MLP 表示全连接层处理,  $F$  为输入特征,  $\sigma$  为激活函数,  $F'$  为输出特征。在空间注意力模块,首先对由通道模块输入的特征图的特征点分别求平均值和最大值,并将这两个结果堆叠合并,再通过  $7 \times 7$  的卷积操作生成空间注意力图。最后和输入特征逐元素相乘输出 CBAM 注意力特征图。该模块更关注位置和形状信息,有助于准确地理解图像的内容。空间注意力模块的计算过程如式(3)

和式(4)所示。

$$M_s(F') = \sigma(f^{7 \times 7}([F_{\text{avg}}^s; F_{\text{max}}^s])) \quad (3)$$

$$F'' = M_s(F') \otimes F' \quad (4)$$

其中,  $F_{\text{avg}}^s$  和  $F_{\text{max}}^s$  分别表示求平均值和最大值的结果,  $f^{7 \times 7}$  表示  $7 \times 7$  的卷积层,  $F'$  为输入特征,  $F''$  为最终输出的特征。

### 1.2.3 Meta-ACON 激活函数

本研究在 YOLOv5 中引入了 Meta-ACON 激活函数,Meta-ACON 通过两个可学习的参数动态调整激活函数一阶导数的上下限。此外,为了避免对不必要的神经元进行激活,Meta-ACON 通过显式地学习是否激活神经元从而动态的控制激活函数的线性或非线性,有效提升了模型传输效率和泛化性能。Meta-ACON 激活函数的计算如式(5)所示。

$$f(x) = (p_1 - p_2)x \cdot \sigma[\beta(p_1 - p_2)x] + p_2x \quad (5)$$

其中,  $p_1$  和  $p_2$  是可学习的参数,用于计算一阶导数的最大值和最小值;  $\beta$  是转换因子,调整函数的激活状态,动态调整函数是线性或非线性。

### 1.2.4 BiFPN 加权双向特征金字塔网络

YOLOv5 传统的特征融合方法是采用 FPN+

PAN 结构,将尺度不同的特征图以相同权重进行加权,但是当输入的特征图分辨率不同时,该方法会导致输出的特征图不平等。BiFPN 通过引入双向信息流,根据不同输入特征的重要性设置不同的权重,同时反复采用这种结构来加强特征融合,有效地促进了高层和低层特征之间的交互,有助于提高目标检测的精度和鲁棒性。其具体流程如图 3 所示。

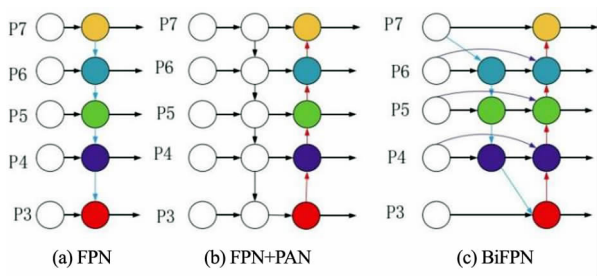


图 3 BiFPN 示意图

以 P6 为例,其输入、输出如式(6)和(7)所示。

$$P_6^{\text{id}} = \text{Conv}\left(\frac{\omega_1 \cdot P_6^{\text{in}} + \omega_2 \cdot \text{Resize}(P_7^{\text{in}})}{\omega_1 + \omega_2 + \epsilon}\right)$$

(6)

$$P_6^{\text{out}} = \text{Conv}\left(\frac{\omega'_1 \cdot P_6^{\text{in}} + \omega'_2 \cdot P_6^{\text{id}} + \omega'_3 \cdot \text{Resize}(P_5^{\text{out}})}{\omega'_1 + \omega'_2 + \omega'_3 + \epsilon}\right)$$

(7)

其中,  $P_6^{\text{id}}$  为自顶向下第 6 级的中间特征,  $P_6^{\text{out}}$  为自底向上第 6 级的输出特征。 $\omega_1, \omega_2, \epsilon$  等是不同层的权重。

2 实验和分析

2.1 评价指标

本实验采用精确率( $P_r$ )、召回率( $R_e$ )、平均精度(AP)和各类别平均精度 mAP 作为性能评价标准。其计算如式(8)一式(11)所示。

$$P_r = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

(8)

$$R_e = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

(9)

$$\text{AP} = \int_0^1 p(r) dr$$

(10)

$$\text{mAP} = \frac{1}{n} \sum_{i=1}^n \text{AP}_i$$

(11)

其中,TP 表示被正确分类的正样本,FP 表示被错误分类为正样本的负样本,FN 表示被错误分类为负样本的正样本,  $p(r)$  表示在不同召回率下的精确率值。

2.2 注意力机制对比实验

为了验证在图书检测模型中添加注意力机制的效果,本文实验了 SE<sup>[14]</sup>、ECA<sup>[15]</sup> 和 CBAM 三个被广泛使用的模型,实验结果如表 1 所示。

表 1 注意力机制对比实验

| 模型           | mAP    | AP      |           | Precision |           | Recall  |           |
|--------------|--------|---------|-----------|-----------|-----------|---------|-----------|
|              |        | damaged | undamaged | damaged   | undamaged | damaged | undamaged |
| YOLOv5s      | 92.18% | 92.17%  | 92.19%    | 98.30%    | 98.39%    | 89.25%  | 89.58%    |
| YOLOv5s+SE   | 92.53% | 92.29%  | 92.76%    | 97.74%    | 98.11%    | 89.55%  | 90.72%    |
| YOLOv5s+ECA  | 92.75% | 92.56%  | 92.93%    | 97.76%    | 98.11%    | 89.72%  | 90.88%    |
| YOLOv5s+CBAM | 93.47% | 93.34%  | 93.59%    | 98.75%    | 99.13%    | 91.69%  | 91.91%    |

由表 1 可以看出,CBAM 在对破损和非破损图书进行检测时均具有更好的表现,其 mAP 达到了 93.47%,分别比 SE 和 ECA 高了 0.94% 和 0.72%。CBAM 首先对输入的特征图进行权重的计算,使模型关注有用的通道信息,捕捉到重要特征,然后又对特征图的不同空间位置进行加权处理,使模型关注重要的空间信息,更加准确地定位目标位置。结合实验结果可以看出,添加 CBAM 使模型更加关注重要的特征内容和位置信息,从而提高了模型的表征能力和检测精度。

2.3 消融实验

为了验证本文所提的各方法对模型的提升效果,本研究进行了消融实验,实验结果如表 2 所示。

由表 2 可以看出,模型 2 的 mAP 相比 YOLOv5s 提高了 0.81%,达到了 92.99%,其中未破损图书类别的 AP 增加了 1.23%,这表明 Meta-ACON 有效提高了模型的性能。由于 Meta-ACON 的上下界是动态调整的,并且其通过调整函数是线性或非线性来决定激活函数的激活状态,这使得该模型能极大地提高传输效率和泛化性能。

表 2 消融实验

| 模型      | CBAM | Meta-ACON | BiFPN | mAP    | AP      |           | Precision |           | Recall  |           |
|---------|------|-----------|-------|--------|---------|-----------|-----------|-----------|---------|-----------|
|         |      |           |       |        | damaged | undamaged | damaged   | undamaged | damaged | undamaged |
| YOLOv5s | —    | —         | —     | 92.18% | 92.17%  | 92.19%    | 98.30%    | 98.39%    | 89.25%  | 89.58%    |
| 模型 1    | ✓    | —         | —     | 93.47% | 93.34%  | 93.59%    | 98.75%    | 99.13%    | 91.69%  | 91.91%    |
| 模型 2    | —    | ✓         | —     | 92.99% | 92.55%  | 93.42%    | 97.62%    | 98.25%    | 89.60%  | 91.37%    |
| 模型 3    | —    | —         | ✓     | 93.04% | 92.74%  | 93.34%    | 98.25%    | 98.12%    | 91.69%  | 91.56%    |
| 本文模型    | ✓    | ✓         | ✓     | 94.89% | 94.59%  | 95.18%    | 98.93%    | 99.15%    | 92.02%  | 91.92%    |

通过模型 3 可以看出添加了 BiFPN 后的 mAP 相比 YOLOv5s 提高了 0.86%，达到了 93.04%，其中破损图书类别的 AP 提高了 0.57%。BiFPN 通过双向连接更好地融合了不同层次的特征,使网络更具有适应性。还通过自适应地捕获图像特征更好地适应不同尺度和复杂度的目标,有助于提高网络的效率和性能。此外,BiFPN 有效减少了信息的丢失,提升了模型对全局上下文的感知能力,有助于更准确地定位和识别目标。

最后,通过对比本文模型可以看出,结合所有改进的模型的 mAP 提高了 2.71%，达到了 94.89%。其中,破损图书的 AP 为 94.59%，增加了 2.42%，非破损图书的 AP 为 95.18%，增加了 2.99%。结果表明,本文所提方法有效提高了图书破损检测模型的性能。图 4 展示了该模型的训练

过程,可以看出在迭代 250 次的时候,分类、置信度和边界框的损失函数均已收敛,说明模型得到了充分的训练。

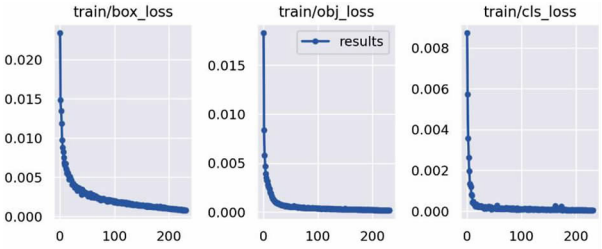


图 4 本文所提模型训练图

2.4 对比实验

为了进一步验证所提模型的性能,本文对比了几个主流目标检测算法,结果如表 3 所示。

表 3 对比实验

| 模型            | mAP    | AP      |           | Precision |           | Recall  |           |
|---------------|--------|---------|-----------|-----------|-----------|---------|-----------|
|               |        | damaged | undamaged | damaged   | undamaged | damaged | undamaged |
| YOLOv7        | 92.30% | 92.28%  | 92.32%    | 98.74%    | 98.40%    | 89.58%  | 90.39%    |
| SSD           | 88.81% | 87.83%  | 89.79%    | 95.03%    | 95.33%    | 80.55%  | 84.04%    |
| Dynamic R-CNN | 90.68% | 90.53%  | 90.82%    | 87.74%    | 97.26%    | 88.44%  | 86.48%    |
| 本文模型          | 94.89% | 94.59%  | 95.18%    | 98.93%    | 99.15%    | 92.02%  | 91.92%    |

由表 3 可以看出,本文提出的模型在性能上相较于 SSD<sup>[16]</sup> 和 Dynamic R-CNN<sup>[17]</sup> 都取得了显著的提升。鉴于本文所使用的数据集相对较小,与 YOLOv7<sup>[18]</sup> 相比本文模型因其更少的参数量也具有 一定的优势,既提升了速度也提高了准确率。这表明本文模型在破损图书检测任务的实际应用中具有更可靠的性能表现。

在训练出图书破损检测模型后,本文采集了一些图书进行测试,图 5 中展示了部分经过模型识别的输出图。可以看出,本文所提模型能够准确地判断图书是否存在破损,所建立的检测模型具有实用性。

3 结 论

考虑到图书馆在破损图书筛查中所面临的人力成本和时效性问题,本文提出了一种基于深度学习的图书破损检测模型。通过在 YOLOv5 中引入 CBAM 注意力机制、Meta-ACON 激活函数以及 BiFPN 加权双向特征金字塔网络,显著提升了模型性能,使其能够准确地识别图书是否存在破损情况。这项工作在图书馆进行破损图书管理方面具有重要意义。它不仅有助于提高图书馆的工作效率,还可以考虑整合至图书馆的自助借还书设备中,通过在借还书过程中对图书是否破损进行实时判断,并对比借书和还书的识别结果以判断读者是否造成了图

书损坏,起到一定的监督作用。本研究未来仍有一些可深入探讨的方向,可以进一步细化图书破损程度以提供更详尽的信息,还可以对图书的破损类别

进行研究,如卷边、书脊脱落、污渍等类别的识别,这些研究将更全面地满足图书馆管理的需要。

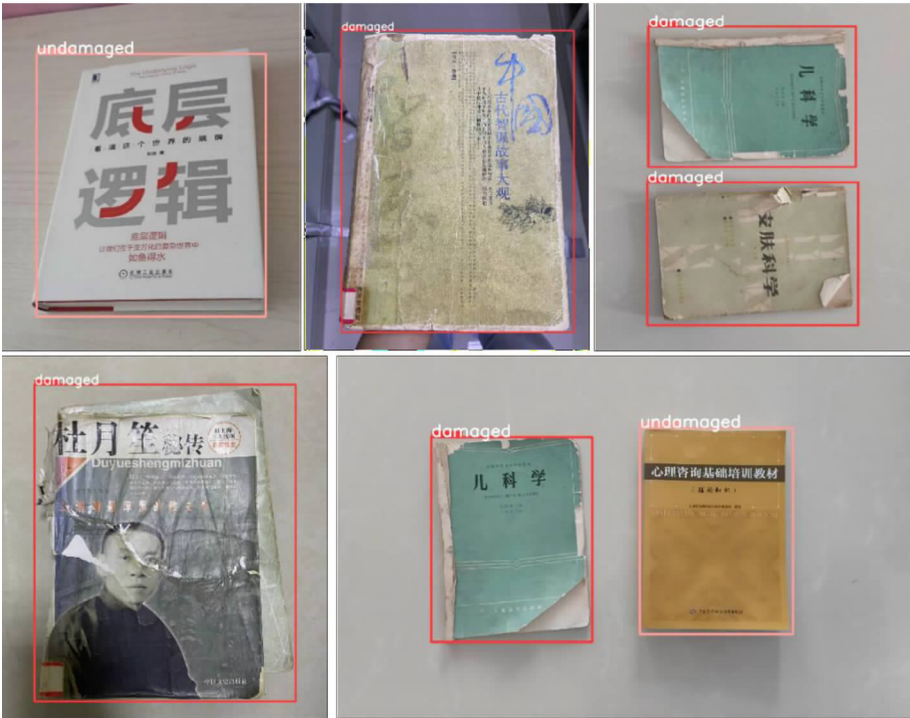


图 5 本文所提模型预测结果图

参考文献

[1] OSUNRINDE A A, AGBETUYI P A. Maintenance culture in academic libraries of universities in south west, nigeria [J]. International Journal of Library and Information Science, 2018, 4(4): 10—22.

[2] MARIEN S, MUNDT A. Let’s get technical-what to do with all those damaged books? [J]. Against the Grain, 2015, 27(3): 38.

[3] 江华. 高校图书馆书库盘点工作提升流通服务实践—以新疆大学为例[J]. 科技资讯, 2019, 17(6): 196—197.

[4] SHI X, TANG K, LU H. Smart library book sorting application with intelligence computer vision technology[J]. Library Hi Tech, 2020, 39(1): 220—232.

[5] KALININA M, NIKOLAEV P. Research of YOLO architecture models in book detection[C]//8th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2020). Atlantis Press, 2020: 218—221.

[6] LEI L, TANG J, WANG Z. Book title recognition for smart library with deep learning[C]// Mobile Multimedia/Image Processing, Security, and Applications 2018. Orlando, United States; SPIE, 2018: 8.

[7] 王红芳, 刘泽远, 李英健, 等. 基于改进的 YOLOv3-Tiny 深度网络在架图书错序检测方法[J]. 现代电子技术, 2022, 45(22): 164—170.

[8] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV), 2018: 3—19.

[9] MA N, ZHANG X, LIU M, et al. Activate or not: learning customized activation [C]//Proceedings of the IEEE/CVF

Conference on Computer Vision and Pattern Recognition, 2021: 8032—8042.

[10] TAN M, PANG R, LE Q V. Efficientdet: scalable and efficient object detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781—10790.

[11] PEI H, SUN Y, HUANG H, et al. Weed detection in maize fields by UAV images based on crop row preprocessing and improved YOLOv4[J]. Agriculture, 2022, 12(7): 975.

[12] 吴杰, 施磊, 张志安. 基于深度学习的害虫图像识别与分类方法研究[J]. 计算技术与自动化, 2023, 42(1): 166—173.

[13] THUAN D. Evolution of Yolo algorithm and YOLOv5: the state-of-the-art object detection algorithm[D]. Oulu :Oulu University of Applied Sciences, 2021.

[14] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132—7141.

[15] WANG Q, WU B, ZHU P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11534—11542.

[16] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[M]// Cham; Springer International Publishing, 2016: 21—37.

[17] ZHANG H, CHANG H, MA B, et al. Dynamic R-CNN: towards high quality object detection via dynamic training [C]// Computer Vision-ECCV 2020. Cham; Springer International Publishing, 2020: 260—275.

[18] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464—7475.