

# 基于YOLOX的移动应用显示异常检测研究

李美静, 李 瑛\*, 孔 婧, 刘益玮

北华航天工业学院计算机学院, 河北 廊坊

收稿日期: 2023年11月22日; 录用日期: 2023年12月20日; 发布日期: 2023年12月28日

## 摘 要

针对现有移动应用中显示异常检测算法检测精度不高、检测性能低下的问题, 提出一种基于YOLOX的目标检测模型实现对四类常见显示异常的检测工作。针对小异常目标缺检、漏检问题, 扩展特征融合网络的特征层输入尺度并进行特征融合, 同时添加一个针对小目标的检测头, 提升对小异常目标检测能力。通过引入坐标注意力机制, 将位置信息和通道信息结合起来增强网络对显示异常的识别。最后通过GIoU定位损失, 实现高精度检测定位。实验结果表明, 改进的YOLOX算法最终mAP达到81.10%, 相比较YOLOX-s模型提高了2.25%, 同时改进后检测模型与其他主流模型相比拥有良好的检测精度与模型泛化能力。该模型为移动应用显示异常智能化检测奠定了基础。

## 关键词

目标检测, 异常检测, YOLOX, 移动测试

# Research on Display Anomaly Detection of Mobile Applications Based on YOLOX

Meijing Li, Ying Li, Jing Kong, Yiwei Liu

Department of Computer Science, North China Institute of Aerospace Engineering, Langfang Hebei

Received: Nov. 22<sup>nd</sup>, 2023; accepted: Dec. 20<sup>th</sup>, 2023; published: Dec. 28<sup>th</sup>, 2023

## Abstract

Aiming at the problems of low detection accuracy and low detection performance of existing display anomaly detection algorithms in mobile applications, an object detection model based on YOLOX was proposed to detect four common types of display anomalies. Aiming at the problem of missing detection and missing detection of small abnormal targets, the input scale of the feature

\*通讯作者。

layer of the feature fusion network is extended and the feature fusion is performed. At the same time, a detection head for small targets is added to improve the detection ability of small abnormal targets. By introducing the coordinate attention mechanism, the location information and channel information are combined to enhance the recognition of display anomalies. Finally, the GloU positioning loss was used to achieve high-precision detection and positioning. Experimental results show that the final mAP of the improved YOLOX algorithm reaches 81.10%, which is 2.25% higher than that of the YoloX-S model. At the same time, the improved detection model has good detection accuracy and model generalization ability compared with other mainstream models. This model lays the foundation for intelligent detection of mobile application display anomalies.

## Keywords

Object Detection, Anomaly Detection, YOLOX, Mobile Testing

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

用户与移动端应用的交互通过图形用户界面(Graphics User Interface, GUI)实现,界面的设计和行为直接影响用户对应用的满意度[1]。移动设备和操作系统的多样化导致在不同的设备上呈现应用程序 GUI 时会出现组件遮盖、文本重叠、显示空值、图片缺失等典型异常[2],如图 1 所示,从左往右依次为,组件遮盖类显示异常,此类异常表现为文本组件的部分内容被遮盖;文本重叠异常,即部分文字出现位置偏移并重叠的现象;显示空值异常,它表现为显示的文本由 null 值替代。图片缺失异常,即图片未显示的情况。手工测试此类显示异常耗费大量的时间和人力资源,测试人员通常需要使用不同的交互方式手动浏览应用程序的多个页面,还需考虑操作系统版本、分辨率以及屏幕大小等因素。同时,GUI 一些相对较小的异常目标,在手工测试时也容易被忽略。因此,高效实现移动应用显示异常的检测是提高移动应用测试效率的关键问题之一[3][4]。

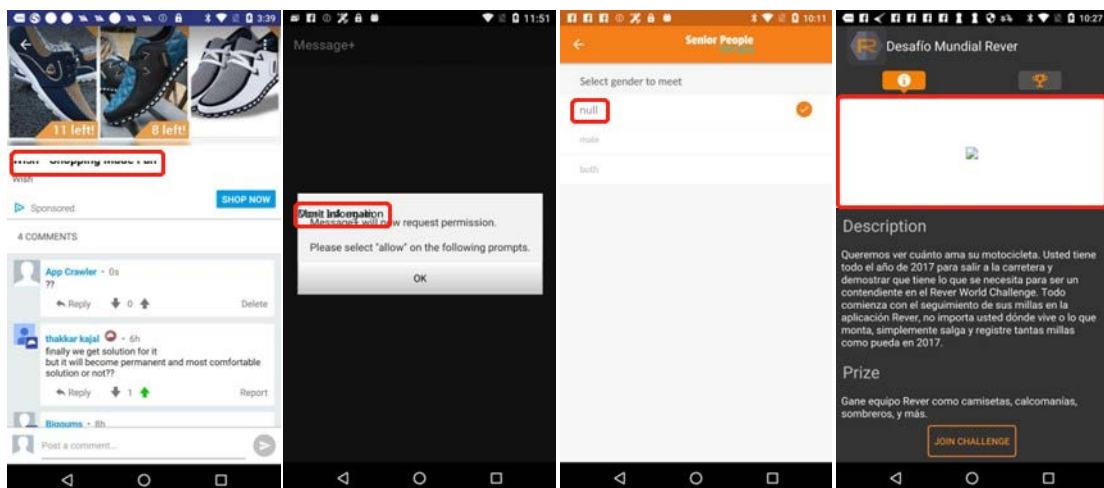


Figure 1. Example diagram of mobile application display exception

图 1. 移动应用显示异常示例图

目前目标检测算法主要可以分为传统检测算法以及基于深度学习的检测算法。传统目标检测算法主要基于手工提取特征,包括 Viola Jones (VJ)算法、Histogram of Oriented Gradients (HOG)检测算法、Deformable Parts Model (DPM)算法等。传统目标检测方法存在时间复杂度高、窗口冗余以及手工提取特征的特点,对于多样性特征变换没有体现出很好的鲁棒性。卷积神经网络兴起后基于深度学习的目标检测方法受到人们广泛关注。这种方法根据检测思想的不同可以分为两类,一类是基于候选区域的两阶段目标检测算法,如 R-CNN、SPPNet、FastR-CNN、Faster R-CNN 等;一类是基于回归的单阶段目标检测算法,如 SSD、YOLO 系列、RetinaNet 等。基于深度学习的检测算法在移动应用控件检测方面取得较好效果,但很少有研究者将基于深度学习的检测技术应用于移动应用显示异常领域。Liu 等人基于卷积和梯度加权类激活映射(Grad-CAM)实现了移动应用显示异常分类和定位。文献[5]采用 DenseNet 加 Non-local 自注意力模块替换传统的卷积神经网络,提升了显示异常网络检测的精度。文献 Nighthawk [6]则基于 Faster-RCNN 实现了移动应用显示异常目标的检测。

上述研究在移动应用显示异常检测上取得了不错的效果,但在小目标漏检、错检和模型检测精度上仍有待提高[7]。YOLOX [8]在 YOLOv3-SPP 基础上,引入新的特征提取机制和采样策略,同时优化网络结构和训练策略,在检测性能方面表现显著。其算法有着不同大小模型的发行版本如 YOLOX-x、YOLOX-l、YOLOX-m 以及 YOLOX-s。本文针对以上问题使用轻量级网络结构的 YOLOX-s 作为进行检测移动应用显示异常的基准模型并进行改进。

## 2. 相关工作

### 2.1. 原始 YOLOX 模型

YOLOX 算法基于 YOLO 系列算法的思想,融合了特征提取网络 CSPDarknet 架构、Ffocus 技术、马赛克数据增强和无锚框思想,提出解耦预测头以及 SimOTA 动态正样本匹配方法,提高检测速度和精度。其原理图如图 2 所示。

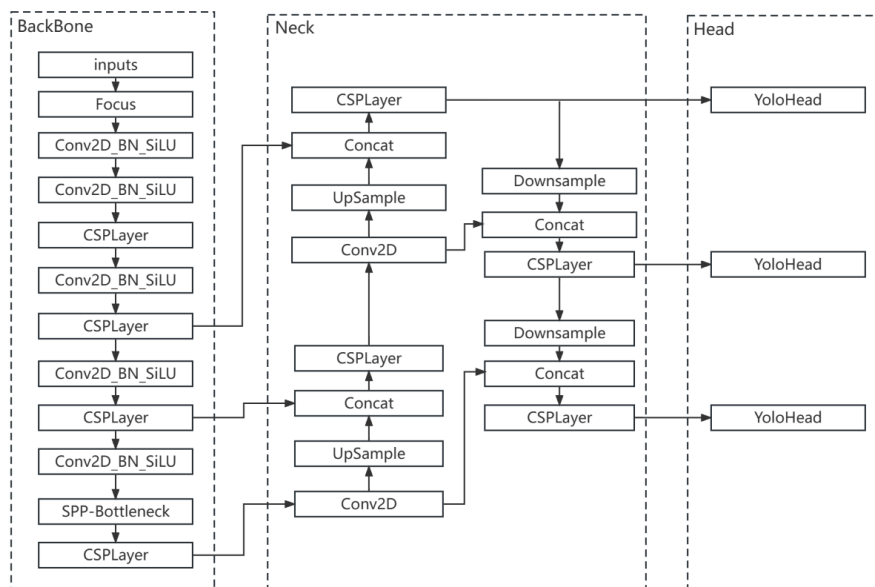


Figure 2. YOLOX-s network structure diagram

图 2. YOLOX-s 网络结构图

YOLOX 算法 Backbone 采用 CSPDarknet 结构,由 Focus、CSPNet 和 SPP 网络组成。Focus 模块对图

片进行等间隔采样切片,然后将得到的采样图像进行堆叠操作。CSPNet 将残差块分为两部分:主干部分和残差边部分,主干部分继续进行原来的残差块的堆叠,用于提取高层次的特征表示。而残差边部分则通过少量处理直接连接到下一层,起到信息传递和维持低层次特征的作用。SPP 网络是一种空间金字塔池化技术,用于获得多尺度的上下文信息。通过在不同尺度上进行池化操作,并将池化结果拼接在一起,可以增强网络对不同尺度目标的感知能力。

Neck 部分采用特征金字塔网络和路径聚合网络(FPN + PAN)的结构,主干网络进行特征提取得到 3 个不同尺度的特征图信息会送入 PAFPN, PAFPN 模块通过上采样和下采样操作,融合多尺度的特征信息,生成三个有效的特征图。上采样操作会增加特征图的分辨率,使得特征图能够捕捉更加细粒度的目标信息。而下采样操作则会减小特征图的分辨率,但同时增加特征图的感受野,使其能够捕捉更大尺度的目标信息。通过上采样和下采样的操作,PAFPN 模块能够有效地融合来自不同尺度的特征信息,提供更全面、丰富的多尺度特征表示。

Head 部分引入 Anchor-free 机制和 SimOTA 方法,采用准确率更高、收敛速度更快的解耦头。经过 PANFPN 模块处理后的特征图会被送入 YOLOHead 模块,解耦头将分类任务和定位任务分开处理,根据任务的不同,分别设置分类任务通道和定位任务通道。每个通道都通过一系列卷积层对特征进行处理,以提取与任务相关的特征表示。最后得到最终的分类和定位结果。

## 2.2. CA 注意力机制

为了更精确的检测出显示异常,减少背景的干扰,本文对引入注意力机制进行研究。区别于传统只考虑通道间信息的编码,丢失了位置信息的 Squeeze-and-Excitation (SE)注意力机制或者引入位置信息,但是位置信息不精细的 Convolutional Block Attention Module (CBAM)注意力机制。Coordinate Attention (CA)注意力机制通过将位置信息嵌入到特征层通道中,使得轻量级网络能够在更大的区域上更加关注目标区域,同时避免了产生大量的计算开销,该结构将空间的详细位置信息和通道注意力结合,能帮助模型更加精准地定位和识别目标。其模型结构如图 3 所示。

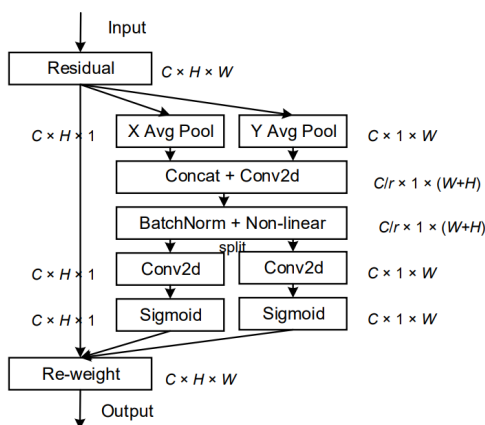


Figure 3. CA attention mechanism structure diagram

图 3. CA 注意力机制结构图

为了获取图像宽度和高度上的注意力并对精确位置信息进行编码,首先将输入特征图分为宽度和高度两个方向分别进行全局池化,得到两个方向的特征图。然后将两个方向的特征图拼接,将其维度降低为原来的  $C/r$ ,并在批量归一化后经过激活函数得到形如  $1 \times (W + H) \times C/r$  的特征图,再将特征图按照原来的高度和宽度进行卷积核为  $1 \times 1$  的卷积还原为原始的通道数,经过激活函数后得到特征图在两个方向

上的注意力权重,最后在原始特征图上通过乘法加权计算获得具有两个方向上注意力权重的特征图。

本文在YOLOX-s的Neck层的上采样和下采样操作之后引入CA注意力机制,在下采样操作之后引入CA注意力机制可以帮助模型更好地捕捉到低级别特征图中的重要信息。通常下采样操作会导致信息损失,但引入CA注意力机制可以增强低级别特征图的有用信息,有助于提高模型的感知能力。在上采样操作之后引入CA注意力机制可以帮助放大上采样后的特征图中的有用信息,使其更容易与下游层的特征图进行融合。这有助于提高高分辨率特征的表征能力,有助于检测小目标或更精细的目标特征。

### 3. 基于改进YOLOX的移动应用显示异常检测

为了提升模型对于移动应用显示异常目标样本的检测效果。本文在原有YOLOX-s模型基础上对整体网络进行改进。采用多尺度特征融合,通过主干网络提取特征,颈部网络实现特征融合,最后借助解耦头对目标进行分析预测。主要改进为以下三个方面。

(1) 针对部分小异常目标检测效率偏低问题,扩展特征融合网络的特征层输入尺度,从主干特征提取网络中提取浅层特征层并进行特征融合。同时将检测层中原3个解耦检测头提升至4个,提升小目标样本检测性能。

(2) 为解决模型在复杂背景下异常目标检测效果不佳问题,在Neck层上采样和下采样之后引入CA注意力机制,增强了网络的特征表达能力。

(3) 为进一步实现模型高精度定位,使用GIoU损失替换原始的IoU损失。

#### 3.1. 扩展特征层尺度

针对原模型对移动应用部分小异常目标检测效果不佳问题,扩展特征融合网络的特征层输入尺度。对于 $640 \times 640$ 输入图片,原网络结构Neck层输入特征层尺度为3个,分别为 $80 \times 80$ 、 $40 \times 40$ 以及 $20 \times 20$ 。本文将主干网络 $160 \times 160$ 浅层特征层提取出来,提高网络对特征图浅层信息的泛化能力。针对 $160 \times 160$ 尺度的特征层新增一个解耦预测头,新增的解耦头对浅层高分辨率特征更敏感,可以更准确地定位和识别小目标。通过对特征融合网络改进,网络模型可以更加全面地捕获多种尺度的特征信息,提升对移动应用显示异常中小目标的检测的性能[9]。

#### 3.2. 引入CA注意力机制

本文在YOLOX-s的Neck层的上采样和下采样操作之后引入CA注意力机制,在下采样操作之后引入CA注意力机制可以帮助模型更好地捕捉到低级别特征图中的重要信息。通常下采样操作会导致信息损失,但引入CA注意力机制可以增强低级别特征图的有用信息,有助于提高模型的感知能力。在上采样操作之后引入CA注意力机制可以帮助放大上采样后的特征图中的有用信息,使其更容易与下游层的特征图进行融合。这有助于提高高分辨率特征的表征能力,有助于检测小目标或更精细的目标特征[10]。

#### 3.3. 引入GIoU损失函数

YOLOX-s模型采用IoU (Intersection over Union)交并比损失来计算预测框和真实框的位置损失。IoU与损失计算公式如下:

$$\text{IoU} = \frac{A \cap B}{A \cup B} \quad (1)$$

$$\text{LossIoU} = 1 - \text{IoU} \quad (2)$$

其中,A代表预测框,B代表真实框, $A \cap B$ 代表A与B交集的面积, $A \cup B$ 代表A与B并集的面积。

虽然 LossIoU 具有良好的尺度不变性，但是也存在着当预测框与真实框都没有交集的时候，无法反映预测框的实际距离大小问题。在 IoU 的基础上，GIoU 引入了预测框和真实框的最小闭包区。GIoU 和损失函数计算公式如下：

$$GIoU = IoU - \frac{A_c - U}{A_c} \tag{3}$$

$$LossGIoU = 1 - GIoU \tag{4}$$

其中， $A_c$  为最小闭包区， $U$  为预测框与真实框的并集。与 IoU 只关注重叠区域不同，GIoU 不仅关注重叠区域，还关注其他的非重合区域，LossGIoU 最小化使得  $A_c$  最小化，通过这种方式能很好的度量预测框与真实框的距离。本文采用 GIoU 替换 IoU 以实现进一步高性能检测定位[11]。

### 3.4. 改进后的 YOLOX 模型

针对原模型对移动应用部分小异常目标检测效果不佳问题，扩展特征融合网络的特征层输入尺度并进行特征融合，增添针对小目标的检测头。为解决模型在复杂背景下异常目标检测效果不佳问题，引入 CA 注意力机制。为实现模型高精度定位，使用 GIoU 损失替换原始的 IoU 损失。改进后的模型结构如图 4 所示。

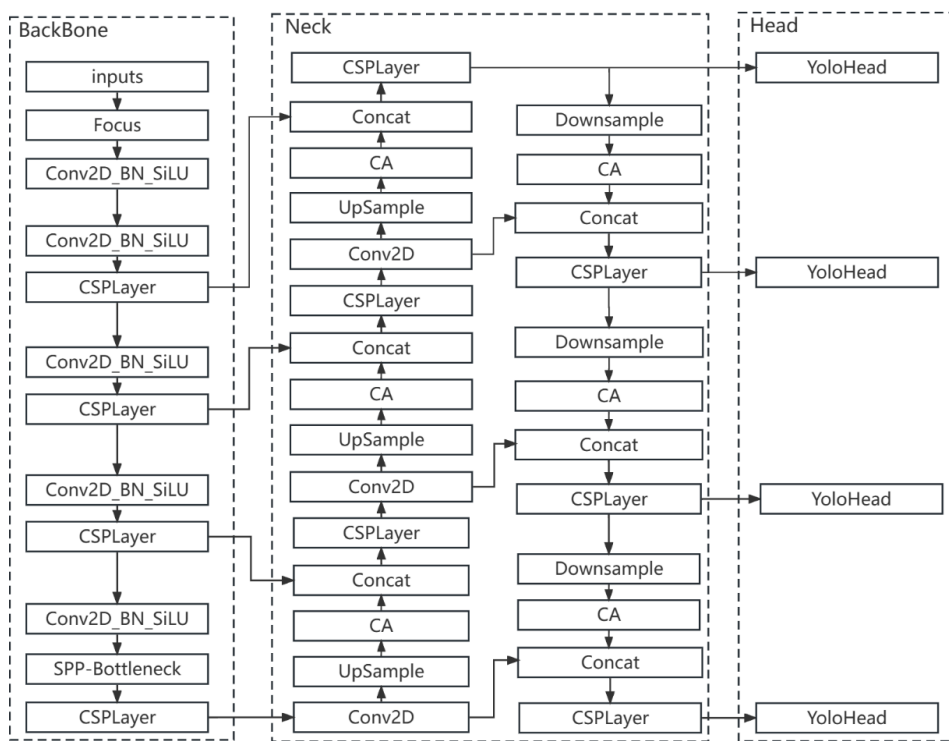


Figure 4. Improved YOLOX network structure diagram  
图 4. 改进后 YOLOX 网络结构图

## 4. 实验结果与分析

### 4.1. 数据集与实验环境

为实现移动应用显示异常类型检测任务，本文基于开源 RICO 数据集，构建了包括：显示空值、组

件遮盖、文本重叠、图片缺失的四类异常数据集。总计 3159 张，单类数据最多 995 张，最少 611 张。每类详细数据如图 5 所示。

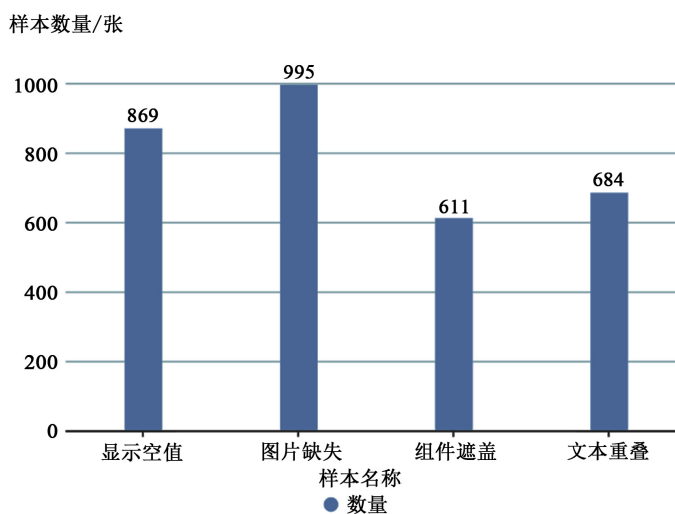


Figure 5. Sample distribution diagram of data set  
图 5. 数据集样本分布图

本次实验在 Python3.8, CUDA10.1, PyTorch1.8.1 环境下进行。所有模型均在 NVIDIA Tesla V100 GPU 上进行训练和测试。数据集按照 9:1 划分为训练集和验证集，batch size 设置为 8，使用 SGD 优化器，学习率设置为 0.001，动量因子为 0.95，权重衰减系数为  $5e-4$ 。总共进行 100 次迭代训练并在最后 30 次迭代训练关闭 Mixup 和 Mosaic。

## 4.2. 实验结果分析

为了验证提出的模型的有效性，将本文提出的改进算法与 YOLOX-s, YOLOv5-s 算法进行对比实验。在显示异常数据集上实验结果如表 1 所示，在显示异常数据集上改进后的 YOLOX 模型检测性能整体提升，mAP 相较于 YOLOX-s 和 YOLOv5-s 分别提高了 2.25% 和 4.37%。

Table 1. Comparison experiment results table

表 1. 对比实验结果表

模型	null AP (%)	cover AP (%)	overlap AP (%)	missing AP (%)	mAP (%)	Parameter (M)
YOLOv5-s	80.70	71.40	65.90	88.95	76.74	7.02
YOLOX-s	97.46	68.58	61.80	87.60	78.86	8.94
本文模型	96.79	77.96	64.08	85.59	81.11	9.69

各类算法在显示空值类和图片缺失类异常的 AP 值相对较高，而在组件遮盖和文本重叠类异常 AP 值较低，其中重要的原因是空值和图片缺失类异常数据集相对单一，而组件遮盖类和文本重叠类异常数据集比较丰富，且这两类数据集样本数量相对较少。由实验结果可知，本文提出的改进模型综合表现更好。

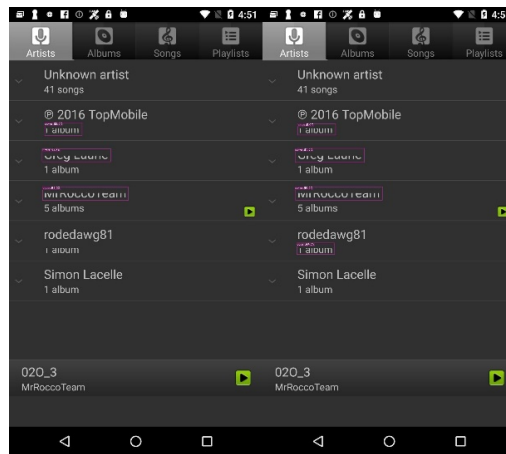
为了进一步验证本文所采用的各个改进策略的有效性，将各个改进策略分别加入原模型，进行消融实验。试验结果如表 2 所示。实验结果显示，原始 YOLOX 模型 mAP 为 78.86%，参数量为 9.84M。改进后的模型参数量提升了 0.75 M，但 mAP 提升了 2.25%，逐步引入改进策略后，mAP 值总体稳步提升，多种改进方式组合在一起时能明显提升模型检测性能，验证了每一步改进策略的有效性[12]。

**Table 2.** Results of ablation test  
**表 2.** 消融实验结果表

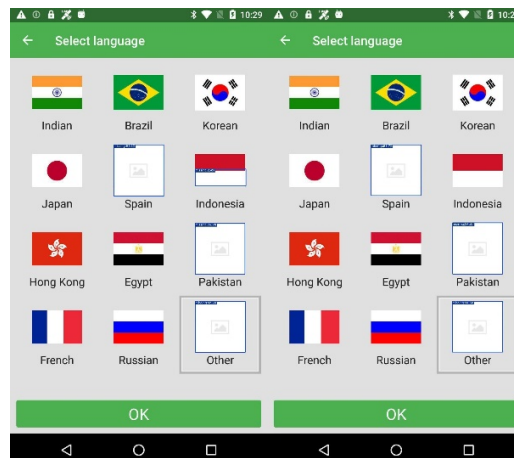
模型	多尺度融合	CA 注意力	GIoU 损失	mAP (%)	Parameter (M)
YOLOX-s	-	-	-	78.86	8.94
改进模型 1	✓	-	-	79.33	9.68
改进模型 2	✓	✓	-	79.59	9.68
改进模型 3	✓	-	✓	79.99	9.69
本文模型	✓	✓	✓	81.11	9.69

### 4.3. 检测可视化对比

为了更好地验证检测效果，本文使用基础 YOLOX 模型与改进的 YOLOX 模型对测试集图片进行可视化对比。左侧为使用基础 YOLOX 网络的结果，右侧为改进的 YOLOX 网络的检测结果。根据检测结果可知，图 6 为组件遮盖类异常检测对比，改进前算法漏掉一个遮盖类异常，改进后的算法对目标漏检问题进行改善；图 7 为图片缺失类异常检测对比，改进后算法解决错检问题；图 8 为文本重叠类异常，改进后算法对显示异常的定位更加准确；图 9 显示空值异常检测，改进后提升了异常检测准确度。



**Figure 6.** Example diagram of component covering inspection effect comparison  
**图 6.** 组件遮盖检测效果对比示例图



**Figure 7.** Example picture of image missing detection effect comparison  
**图 7.** 图片缺失检测效果对比示例图



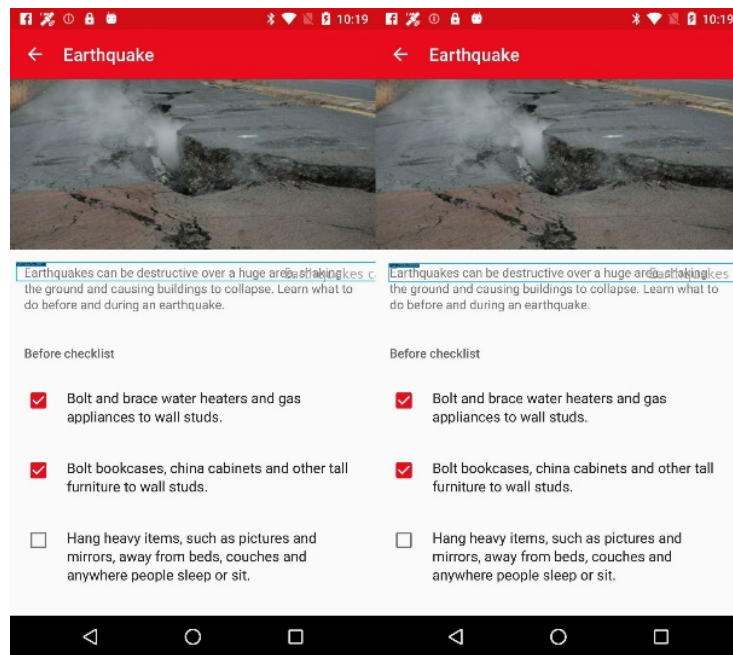


Figure 8. Example of text overlap detection effect comparison

图 8. 文本重叠检测效果对比示例图

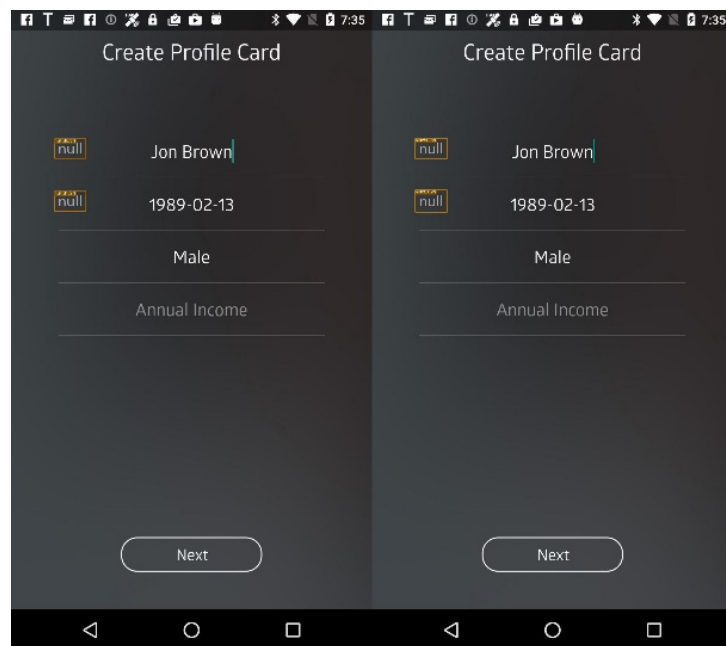


Figure 9. Display an example of null value detection effect comparison

图 9. 显示空值检测效果对比示例图

通过上述对比实验以及结果分析可知，本文改进后的算法在各种复杂场景下保持良好检测效果。

## 5. 结论

本文提出一种基于 YOLOX 的改进算法，算法针对移动应用显示异常检测精度不高、漏检错检以及小异常目标检测检测效率低下问题，提出增加一个针对小目标的检测尺度和对应的检测头，并且通过优

化损失函数以及引入 CA 注意力机制, 综合提升检测效果。通过实验可知, 本文研究为移动应用显示异常检测提供有益参考, 在今后的研究中将继续尝试对模型进行优化并结合自动化测试平台, 进一步提高移动应用自动化测试效率[13] [14]。

## 基金项目

校级研究生创新资助项目(项目编号: YKY-2022-34); 校级研究生创新资助项目(项目编号: YKY-2022-36); 校级研究生创新资助项目(项目编号: YKY-2023-36); 廊坊市科技支持计划项目(项目编号: 2021011029)。

## 参考文献

- [1] 郝琳. 面向 GUI 控件识别的深度学习图像匹配算法研究[D]: [硕士学位论文]. 廊坊: 北华航天工业学院, 2023.
- [2] Liu, Z., Chen, C., Wang, J., et al. (2020) Owl Eyes: Spotting UI display Issues via Visual Understanding. *Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering*, New York, 21-25 December 2020, 398-409. <https://doi.org/10.1145/3324884.3416547>
- [3] 王寅, 黄河基, 李佳芮. 基于 YOLO 的移动应用自动测试[J]. 现代信息科技, 2023, 7(18): 106-109.
- [4] 郑炜, 唐辉, 陈翔, 等. 安卓移动应用兼容性测试综述[J]. 计算机研究与发展, 2022, 59(6): 1371-1384.
- [5] 戴业伟. 基于深度学习的移动应用界面显示问题检测[D]: [硕士学位论文]. 成都: 四川大学, 2021.
- [6] Liu, Z., Chen, C., Wang, J., et al. (2022) Nighthawk: Fully Automated Localizing UI Display Issues via Visual Understanding. *IEEE Transactions on Software Engineering*, **49**, 403-418. <https://doi.org/10.1109/TSE.2022.3150876>
- [7] 刘洋, 战荫伟. 基于深度学习的小目标检测算法综述[J]. 计算机工程与应用, 2021, 57(2): 37-48.
- [8] Ge, Z., Liu, S., Wang, F., et al. (2021) YOLOX: Exceeding YOLO Series in 2021. *arXiv*, **1**, 1-7.
- [9] 鞠默然, 罗江宁, 王仲博, 等. 融合注意力机制的多尺度目标检测算法[J]. 光学学报, 2020, 40(13): 132-140.
- [10] 廖延娜, 姚亮. 改进 YOLOX 的桥梁病害检测识别[J]. 应用光学, 2023, 44(4): 794-800.
- [11] 谷长江, 高法钦. 改进 YOLOX-S 的金属零件缺陷检测算法研究[J]. 计算机时代, 2023(7): 29-33.
- [12] 刘健. 基于 YOLOX 的输电线路异物检测算法研究及软件设计[D]: [硕士学位论文]. 徐州: 中国矿业大学, 2022.
- [13] 薛峰, 武君胜, 张涛, 等. 面向移动应用自动化测试的同构用户界面视觉判断方法[J]. 西北工业大学学报, 2022, 40(4): 805-810.
- [14] Li, Y.C., et al. (2017) DroidBot: A Lightweight UI-Guided Test Input Generator for Android. *2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C)*, Buenos Aires, 20-28 May 2017, 23-26. <https://doi.org/10.1109/ICSE-C.2017.8>