

文章编号:1007-2780(2024)06-0790-11

基于 YOLOv5 的液晶屏微弱特征缺陷检测算法

林峰, 石艳*, 陈顺龙, 廖映华, 赵练, 赵黎, 周泽民
(四川轻化工大学机械工程学院, 四川宜宾 644000)

摘要:针对液晶屏显示缺陷中微弱特征缺陷经多次卷积与背景纹理同化导致的检测精度低的问题,提出了一种基于 YOLOv5 的液晶屏微弱特征缺陷检测改进模型 YOLO-Mura。首先,在主干网络中引入 Involution 算子扩大感受野,增强在空间范围内的微弱特征缺陷信息,并降低模型的浮点运算次数。其次,采用 CARAFE 上采样算子优化上采样方式,加强对微弱特征缺陷的关注能力。然后,在颈部网络,通过嵌入 BiFormer 注意力模块,提升网络在强背景干扰下的特征提取能力。最后,采用 BiFPN 加权双向金字塔结构,提高不同层级的特征融合利用率。在自制液晶屏 Mura 缺陷数据集上的实验结果表明, YOLO-Mura 模型的精确率、召回率、mAP@0.5 分别提高了 2.2%、6.6%、2.7%,模型计算量降低了 66.5%。通过与主流目标检测算法进行比较,结果表明本文最终改进模型对于液晶屏微弱特征的 Mura 缺陷有较好的检测性能。

关键词:液晶屏; Mura 缺陷; YOLOv5 算法; 微弱特征检测

中图分类号: TP391.41 **文献标识码:** A **doi:** 10.37188/CJLCD.2023-0206

Weak feature defect detection method for LCD screens based on YOLOv5

LIN Feng, SHI Yan*, CHEN Shunlong, LIAO Yinghua, ZHAO Lian, ZHAO Li, ZHOU Zemin

(School of Mechanical Engineering, Sichuan University of Science & Engineering,
Yibin 644000, China)

Abstract: To address the problem of low detection accuracy of weak feature defects in LCD display defects caused by multiple convolution and background texture assimilation, an improved model YOLO-Mura for LCD weak feature defect detection based on YOLOv5 is proposed. Firstly, Involution operator is introduced in the backbone network to expand perceptual field, enhance the information of weak feature defects in spatial range, and reduce model FLOPs. Secondly, the CARAFE upsampling operator is used to optimize the upsampling method and enhance the ability to focus on weak feature defects. Then, in the neck network, the feature extraction ability of the network under strong background interference is enhanced by embedding the BiFormer attention module. Finally, the BiFPN weighted bidirectional pyramid structure is adopted to improve the feature fusion utilization at different levels. Experimental results on the homemade

收稿日期: 2023-06-06; **修订日期:** 2023-07-02.

基金项目: 宜宾市科技厅重点研发项目(No.2021GY0011); 宜宾市三江新区揭榜挂帅项目(No.2022JBGS001); 四川轻化工大学研究生创新基金(No.Y2022056)

Supported by R&D Project of Yibin Science and Technology Department (No.2021GY0011); Yibin Sanjiang New Area Unveiling Hanging Project (No.2022JBGS001); Sichuan University of Light Industry Graduate Innovation Fund (No.Y2022056)

*通信联系人, E-mail: sy71Email@163.com

LCD Mura defect dataset show that the accuracy, recall, and mAP@0.5 of YOLO-Mura model are improved by 2.2%, 6.6%, and 2.7%, respectively, and the model computation is reduced by 66.5%. In comparison with the mainstream target detection algorithms, the results show that the final improved model in this paper has better detection performance for Mura defects with weak features of LCDs.

Key words: LCD screen; Mura defects; YOLOv5 algorithm; weak feature detection

1 引言

液晶屏显示缺陷有频闪、色差、屏幕漏光及Mura缺陷等。液晶屏显示缺陷会极大地影响产品的视觉效果及用户体验。例如,在医疗领域,医用显示屏的不良显示缺陷可能会影响医生对患者病情的诊断。对于其他液晶屏应用领域,缺陷产品进入市场销售渠道会导致消费者的正常权益受到损害、企业召回维修的成本增加且品牌的信誉度下降。因此,液晶屏出厂前需要经过严格的质量检测。Mura缺陷表现出以下特征:低对比度、边缘模糊、形状各异以及整体亮度不均匀。在液晶屏显示缺陷中,Mura缺陷被认为是一种难以检测的缺陷之一^[1]。由于以上表现出的多种特性组合,Mura缺陷从中心到边缘的灰度过渡平缓,容易与背景混淆,导致检测困难。对于这种微弱特征的缺陷,一般人工难以设计出提取有明显区分性轮廓或结构信息的检测器^[2]。深度学习缺陷检测方法通过深度神经网络自动学习特征表示,可以从原始数据中学习更加抽象和高级的特征。

针对难以提取的微弱特征缺陷。2020年Zhang等人提出了一种新颖的弱特征增强网络(WFBNNet)^[3]。2021年Du等人提出基于帧间能量积累(IFEA)增强机制的端到端时空特征提取与目标检测框架^[4]。2021年Tao等人提出一种新颖的DeepScratchNet,通过聚合丰富的多维特征进行划痕表示,实现弱划痕自动检测^[5]。2023年汪西莉等人提出基于特征注意力金字塔的目标检测方法FAPNet,强化目标区域特征,弱化背景区域特征,进一步提升弱特征检测方法的性能^[6]。2023年周葳楠等人提出了一种增强弱特征表达的一阶段轻量级小目标检测算法SA-YOLO,充分提取弱特征,提高对弱特征检测的性能^[7]。针对微弱特征缺陷检测,基于深度学习的缺陷检测算法已成为主流。

本文为了提高对液晶屏微弱特征缺陷的检测能力,提出一种基于改进YOLOv5(You Only Look Once Version 5)的液晶屏微弱特征缺陷检测算法。本文的工作总结如下:

(1)使用Involution算子实现不同空间位置的自适应分配权重,在更广阔的空间中聚合上下文语义信息,大幅降低模型的浮点运算次数(FLOPs)^[8]。为了弥补网络性能损失,使用CARAFE上采样算子使网络更加关注目标区域,增强了模型的抗干扰能力。

(2)为了提高液晶屏微弱特征缺陷检测的性能,在颈部网络中加入了基于动态稀疏注意力机制BiFormer,以一种查询自适应的方式集中关注部分相关标记,而不会分散注意力于其他无关的标记,确保网络模型具备出色的性能和高效的计算效率^[9]。同时,将PANet特征融合模块替换为BiFPN,使不同尺度的特征图之间相互补充,更好地捕捉目标的多尺度特征,从而进一步提高微弱缺陷特征检测精度。

2 YOLO-Mura缺陷检测算法的整体结构

YOLO(You Only Look Once)系列算法是一种单阶段目标检测算法^[10]。该系列算法将目标检测任务转化为一个回归问题,通过一个深度卷积神经网络对输入图像进行处理,并直接输出预测物体的边界框和类别^[11]。YOLO算法的结构主要由4部分组成:输入端(Input),输入图像尺寸为640×640;主干网络(Backbone),用于提取输入图像的特征;颈部网络(Neck),用于融合提取的特征;头部网络(Head),负责将颈部网络提取的特征图转化为目标类别的预测和边界框的回归。使用100张液晶屏微弱特征缺陷样本对不同目标检测模型进行性能测试,YOLOv5n、YOLOv7-Tiny和YOLOv8s的漏检率分别为14%、

13%和11%,YOLOv5s的漏检率为9%。相比之下YOLOv5s的基础性能较好,因此本文选择YOLOv5s作为基础网络进行改进。

本文主要针对中心亮度与周围背景亮度差值小于5、形状边界不明显的微弱特征缺陷,进行基于YOLOv5s的液晶屏微弱特征缺陷检测模型改进。图1是改进后YOLOv5s模型的整体结构图。首先,在骨干网络中添加新型的卷积算子Involution,可以大幅降低模型运算量。然后,为了弥补网络性能损失,将YOLOv5s原始上采样nearest算子更换为CARAFE上采样算子,使采样点集中在目标区域而忽略背景部分,从而降低误检和漏检的情况。接着,在颈部网络添加基于动态稀疏注意力机制BiFormer,构建高效金字塔网络架构,有助于增强网络对于暗淡目标的检测性能。最后,将颈部网络的PANet特征融合模块替换成BiFPN模块,通过引入BiFPN加权双向金字塔结构增强对不同尺度目标的检测效果,从而提高网络整体性能。

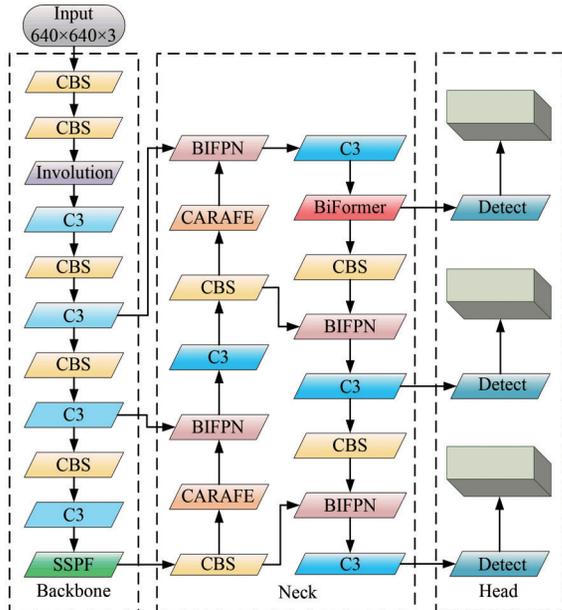


图1 YOLO-Mura网络模型

Fig. 1 YOLO-Mura network model

3 基于YOLOv5s的算法改进

3.1 骨干网络引入新型卷积算子

Involution是一种新型的卷积算子,它在改变网络整体架构的情况下,能够显著降低传统

卷积网络的FLOPs和参数量。与传统的卷积不同的是,Involution算子重新考虑卷积空间和通道范围的固有属性,在通道维度上共享参数,在空间维度上的参数各不相同,即通道不变性和空间特异性^[12]。Involution算子的通道共享属性使Involution可以使用更大的核,而不过多增加其参数量。Involution算子空间特异属性使相同网络不同的输入产生不同大小的特征图,自适应地提取更多的特征信息^[13]。综合以上两个因素,Involution算子在空间维度上具有广泛的覆盖范围,即更广的感受野,从而收集到更丰富的特征信息,提高主干网络从不同位置发掘不同特征的能力^[14]。图2展示了Involution算子的计算过程。首先,对于输入大小为 $H \times W \times C$ 特征图 F_{in} ,经过Involution内核 $\phi(F_{i,j})$,提取其中大小为 $1 \times 1 \times C$ 的张量做通道压缩,得到大小为 $1 \times 1 \times k^2$ 的张量 $F'_{i,j}$ (k 是Involution核大小)。再将 $F'_{i,j}$ 转换成大小为 $k \times k \times G$ (G 指将通道数 C 分为 G 组)的张量 $I_{i,j}$ 。然后,将输入特征图经过卷积核处理,输出为 $k \times k \times C$ 的特征图。最后,进行聚合运算输出特征图大小为 $1 \times 1 \times C$ 。

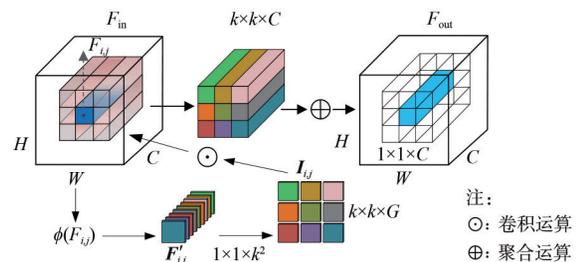


图2 Involution算子的前馈计算过程

Fig. 2 Feedforward calculation process of Involution operator

3.2 上采样方法优化

在对液晶屏进行检测时,液晶屏的微弱特征缺陷容易与背景混淆,造成误检和漏检的情况。YOLOv5s网络采用nearest算子,处理低对比度缺陷图像时会存在特征丢失的风险^[15]。而CARAFE算子能增大上采样的有效感受野,将采样点集中在目标区域而忽略背景部分,从而降低误检和漏检的情况^[16]。如图3所示,CARAFE算子由上采样核预测模块和特征重组模块组成^[17]。在上采样核预测模块中,使用 1×1 卷积将输入特征图的通道数压缩到 C_m 。对于压缩后的特征图通道

C_m , 采用 $k_{\text{encoder}} \times k_{\text{encoder}}$ 的卷积层预测上采样核, 得到输出通道数为 $\sigma^2 k_{\text{up}}^2$, 上采样核大小为 $\sigma H \times \sigma W \times k_{\text{up}}^2$ 。最后, 使用 softmax 激活函数对上采样核进行归一化处理。在特征重组模块中, 输出特征图中的每个特征点被映射回输入特征图。使用以该点为中心的 $k_{\text{up}} \times k_{\text{up}}$ 区域和其预测的上采样核做点积运算, 计算最终的上采样结果。

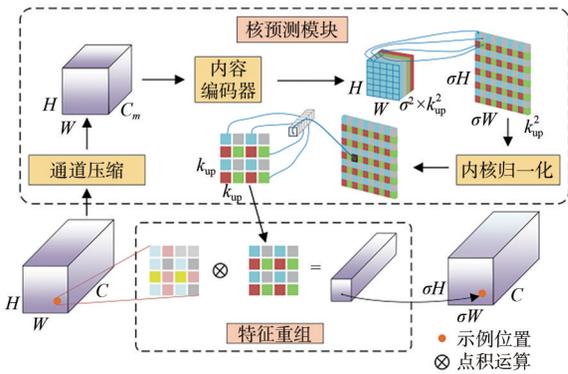


图 3 CARAFE 上采样结构

Fig. 3 CARAFE upsampling structure

3.3 颈部网络引入注意力模块

BiFormer 是一种通过双层路由的动态稀疏注意力模块, 能实现具有内容感知的更灵活的计算分配, 更好地捕捉液晶屏缺陷图像中的全局和局部信息^[18]。BiFormer 使用 BRA (Bi-Level Routing Attention) (图 4(a)) 作为基本构建块。BRA 对于给定输入二维特征图 $X \in \mathbb{R}^{H \times W \times C}$, 首先将输入特征图划分为 $S \times S$ 个非重叠区域, 其中每个区域包括 $(H \times W)/S^2$ 个特征向量, 这一步骤是通过将 X 重塑为 $X^r \in \mathbb{R}^{(S^2 \times HW \times C)/S^2}$ 来完成的。然后, 通过线性映射获得 $Q = X^r W^q, K = X^r W^k, V = X^r W^v$, 其中 $Q, K, V \in \mathbb{R}^{(S^2 \times HW \times C)/S^2}$, $W^q, W^k, W^v \in \mathbb{R}^{C \times C}$ 分别是 query、key、value 的投影权重。

BiFormer 使用四级金字塔结构^[19], 如图 4(b) 所示。在每个阶段采用不同的方法处理输入数据。对于第一阶段 ($i=1$), 使用重叠的 patch embedding 降低输入空间的分辨率, 并增加通道数量。而对于第 2、第 3 和第 4 阶段 ($i=2, 3, 4$), 重复使用 patch merging 模块实现相同的效果。

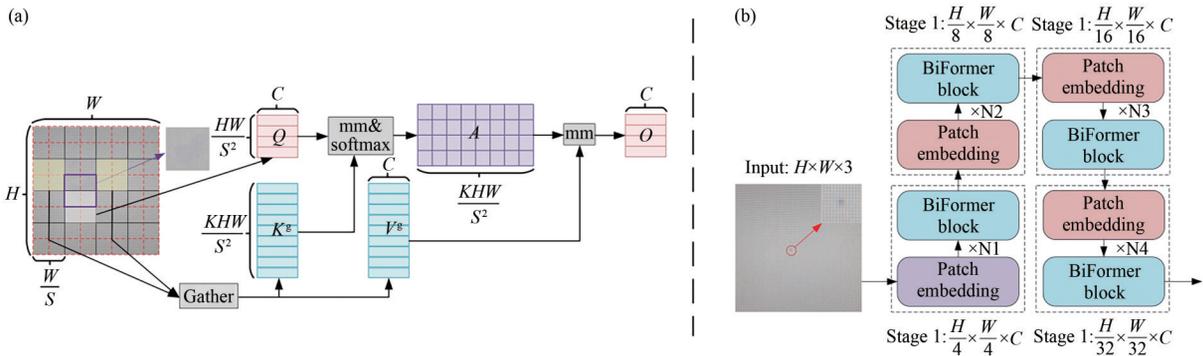


图 4 (a) 双层路由注意 (BRA) 模块; (b) BiFormer 的整体架构。

Fig. 4 (a) Double layer routing attention (BRA) module; (b) Overall architecture of BiFormer.

3.4 特征融合优化

加权双向特征金字塔网络 (BiFPN) 能够实现简单快速的多尺度特征融合, 在提高目标检测性能的同时减少计算量和资源占用^[20]。BiFPN 主要思想有两点: 一是高效的双向跨尺度连接, 二是加权特征融合^[21]。FPN 结构如图 5(a) 所示, 红色路径为自顶向下进行特征融合的通路。用融合后的具有更高语义信息的特征图进行预测, 可以提高一定的检测精度^[22]。PANet 结构如图 5(b) 所示, 橙色路径为在 FPN 结构的基础上再建立的一条自底向上通路, 将位置信息传到预测特征图

中, 使得预测特征图同时具备较高的语义信息和位置信息, 从而提高目标检测任务精度^[23]。BiFPN 结构如图 5(c) 所示, 在 PANet 结构的基础上, 为同一层在输入节点和输入节点间新加蓝色路径, 在不增加太多成本的情况下融合更多的特性^[24]。同时, 使用 NAS 技术重复使用 BiFPN 结构以实现更高层次的特征融合。公式 (1)、(2) 为使用快速归一化带权特征融合方法对不同输入特征有区分的融合:

$$P_6^{\text{id}} = \text{Conv} \left(\frac{w_1 \cdot P_6^{\text{in}} + w_2 \cdot \text{Resize}(P_7^{\text{in}})}{w_1 + w_1 + \epsilon} \right), \quad (1)$$

$$P_6^{\text{out}} = \text{Conv} \left(\frac{\omega'_1 \cdot P_6^{\text{in}} + \omega'_2 \cdot P_6^{\text{td}} + \omega'_3 \cdot \text{Resize}(P_5^{\text{out}})}{\omega'_1 + \omega'_2 + \omega'_3 + \epsilon} \right), \quad (2)$$

其中:Resize为下采样或上采样操作, ω 是学习到的权重参数。

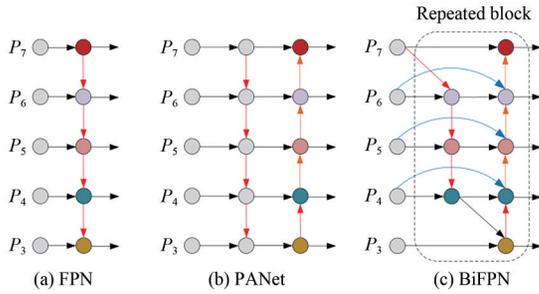


图5 特征融合结构

Fig. 5 Feature fusion structure

4 实验结果与分析

4.1 实验平台和数据集

本试验平台的硬件配置为 NVIDIA RTX 2060 Supers 8 GB (GPU)、Intel (R) Core (TM) i5-12400F 2.50 GHz (CPU);软件环境为 CUDA 11.6、cuDNN 8.2 和 Python 3.8,并采用 Pytorch 1.10.0 深度学习框架。本文使用 LabelImg 标注工具对液晶屏缺陷数据集进行标注。将液晶屏微弱特征的 Mura 缺陷按形状划分为 3 类,如图 6 所示,分别为点状 Mura、线状 Mura 以及区域状 Mura。

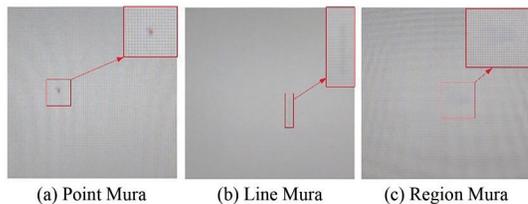


图6 液晶屏 Mura 缺陷样本图片

Fig. 6 Sample pictures of LCD Mura defect

如表 1 所示,总计有 2 674 张图片作为实验数据。将标记好的数据集随机打乱后,按照 8:1:1 的比例划分为训练集、验证集以及测试集。模型在训练时使用高斯模糊、亮度变换等图像增强技术,模拟屏摄图像的模糊、亮度不匀等情况,增加

表 1 液晶屏 Mura 缺陷数据集

Tab. 1 LCD Mura defect data set

液晶屏 Mura 缺陷类型	图片数量
点状 Mura	855
线状 Mura	867
区域状 Mura	952

样本的多样性,有助于提高模型的泛化能力。

4.2 实验细节和评估指标

4.2.1 实验细节

根据硬件条件使用如下超参数进行训练:动量设置为 0.937,初始学习率为 0.01,使用 SGD 优化器进行优化,权重衰减为 0.000 5,并使用 Warmup 减缓模型训练过程中高初始学习率导致的模型振荡^[25]。通过使用高斯模糊与旋转变换等数据增强方法在线增强数据,令输入网络的图像保持多样性。模型输入图像尺寸为 $640 \times 640 \times 3$,Batch size 设置为 16,网络训练次数设置为 500 epoch。

4.2.2 评估指标

本实验通过精确率(Precision)、召回率(Recall)、浮点运算次数(FLOPs)和全类平均精度均值(mean Average Precision, mAP)等指标衡量算法性能。全类平均精度均值与精确率和召回率相关,如公式(3)所示。精确率的计算为正确预测的正样本个数与总预测为正样本的个数之比,如公式(4)所示。召回率的计算为正确预测的正样本个数与总正样本个数之比,如公式(5)所示。

$$\text{mAP} = \frac{\sum_{i=1}^N \int_0^1 \text{PdR}}{N}, \quad (3)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (4)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (5)$$

在上述公式中,mAP 的值由各类别缺陷的 AP 计算得出并取其平均值。其中 N 表示网络检测的总类别数,TP (True Positives)、FP (False Positives)和 FN (False Negatives)分别代表正确检测框、误检框和漏检框的数量。较高的 mAP 值表示该模型在给定数据集上具有更佳的性能。

4.3 实验及分析

4.3.1 骨干网络改进实验及分析

本文通过实验探究 Involution 算子对网络性能的影响,将 Involution 模块引入 YOLOv5s 模型的主干网络中,实现对算力的重新调配。在 YOLOv5s 模型的基础上,将 Involution 算子添加到基础模型的骨干网络中,进行 6 组模型对比,实验结果如表 2 所示。从实验 2、3、4 中可以看出,Involution 层嵌入网络中的位置越靠后,降低网络 FLOPs 的效果越差,网络精确率、mAP@0.5 值降低越明显。从实验 1、2、5、6 中可知,虽然引入多个 Involution 层可以大幅度降低网络 FLOPs,但会令 mAP@0.5 值明显降低,且模型收敛速度减慢,其中实验 6 引入 3 个 Involution 层对模型性能影响最大。与实验 1 YOLOv5s 模型相比,其

FLOPs 降低了 81%、mAP@0.5 降低了 4.6%。从表 2 的对比实验中可以看出,仅添加 1 个 Involution 算子且添加位置靠前的方案能够较好地平衡 FLOPs 和检测精度。实验 2 相比实验 1 的网络精确率仅降低 2.9%,mAP@0.5 值降低 2.3%,FLOPs 降低 66.5%。因此,为了平衡模型的复杂程度与性能,选择方案 2 为初步改进结果,并将其模型命名为 YOLO-I。图 7 展示了不同大小感受野对微弱特征提取的能力,图 7(a)、(b) 两组特征图可视化展示了传统卷积与 Involution 算子对不同大小微弱缺陷特征的提取能力。蓝色区域表示背景区域,黄色区域表示网络对缺陷的关注区域。传统卷积限制了卷积的感受野,对微弱特征图像提取能力较差。而 Involution 算子在空间维度上具有广泛的覆盖范围,能够较好地提取微弱特征。

表 2 在主干网络不同位置中引入 Involution 层的对比实验

Tab. 2 Comparison experiments of introducing Involution layers in different locations of the backbone network

实验	第二层	第四层	第六层	Precision/%	Recall/%	FLOPs/G	mAP@0.5/%
1				96.4	88.9	15.8	94.5
2	✓			93.5	91.5	5.3	92.1
3		✓		92.5	89.0	6.8	90.5
4			✓	91.8	90.3	16.0	89.6
5	✓	✓		87.1	86.3	3.0	90.5
6	✓	✓	✓	87.0	85.5	3.0	89.9

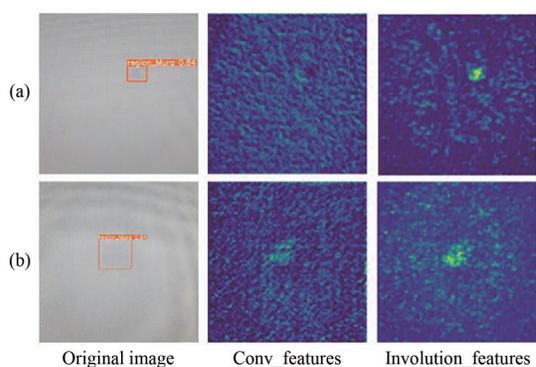


图 7 卷积特征图的可视化

Fig. 7 Visualization of feature map

4.3.2 优化上采样方式实验及分析

为了弥补模型性能损失,将原始上采样 nearest 替换为 CARAFE,并将添加 CARAFE 模块的模型命名为 YOLO-IC。CARAFE 算子旨在从大的感受野中捕获环境特征信息,将采样点集中在目

标区域以抵消模型带来的性能下降。相应的结果如表 3 所示,精确度提高了 1.1%、召回率提高了 1.5%、mAP@0.5 提高了 1.7%。

表 3 改进的上采样方式

Tab. 3 Improved upsampling method

Model	Precision/%	Recall/%	mAP@0.5/%
YOLO-I	93.5	91.5	92.1
YOLO-IC	94.6	93.0	93.8

为了进一步展示改进上采样方式的作用,将特征图的所有通道对应像素点的值取平均得出可视化结果。如图 8 所示,添加 Involution 算子的 YOLO-I 上采样相比于 YOLOv5s 的上采样,对缺陷区域有一定的聚合作用。添加 CARAFE 算子的 YOLO-IC 上采样相比于 YOLO-I 上采样,能较好地 将采样点集中在目标区域。

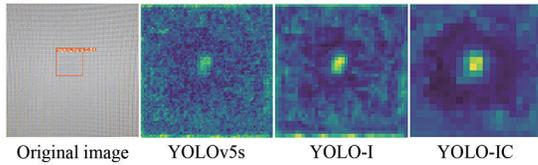


图 8 上采样特征图的可视化

Fig. 8 Visualization of upsampled feature maps

4.3.3 添加注意力模块实验及分析

为了进一步增强模型对微弱特征 Mura 缺陷的检测精度,在 YOLO-IC 模型的基础上,分别添加 Efficient Channel Attention (ECA)、Coordinate Attention (CA)、Shuffle Attention (SA)、BiFormer 注意力模块到主干网络中,进行相应的实验。对比 4 种不同注意力模块对模型性能的影响,结果如表 4 所示。从精确率、召回率、mAP@0.5 方面综合考虑,添加 BiFormer 模块在注意力模块实验中获得了最优的效果,并将其命名为 YOLO-ICB。与 YOLO-IC 模型相比,其精确率提高了 1.4%,召回率和 mAP@0.5 提高了 2.7%,进一步增强了模型性能。

表 4 添加不同注意力模块的模型性能

Tab. 4 Performance of model with different attention modules

注意力机制	Precision/%	Recall/%	mAP@0.5/%
YOLO-IC	94.6	93.0	93.8
ECA	95.2	92.7	95.8
CA	93.4	94.5	94.2
SA	93.5	93.8	94.7
BiFormer	96.0	95.7	96.5

4.3.4 优化特征融合模块实验及分析

特征融合是目标检测中的一个重要环节,能够融合高低不同层级的特征信息,传统的 Concat 只是单纯地将不同层级的特征在通道层面相连^[26]。为了提升模型对微弱缺陷的检测精度,本文使用 BiFPN 模块替换原始 PANet 加强网络特征的融合,丰富语义信息。BiFPN 通过自底向上 (PAN) 和自顶向下 (FPN) 的双向通道以及横向连接的方式,有效地融合了不同尺度的特征信息,解决了主干网络中特征信息丢失的问题,提高了目标检测算法的性能和精确度。表 5 实验数据表明了 BiFPN 模块的有效性,经过特征融合网络对特

征块权重的调整后,模型对融合后的特征块利用率更高。由表 5 可知,PANet 特征融合相比单路自顶向下的模型的精确率、召回率、mAP@0.5 分别提高了 3.3%、12.4% 和 4.4%。PANet 特征融合相比单路自底向上的模型的精确率、召回率、mAP@0.5 分别提高了 1.2%、12.3% 和 3.1%。在 YOLO-ICB 基础上添加 BiFPN 模块后,模型精确率提升了 2.6%、召回率提升了 0.3%、mAP@0.5 提升了 0.7%。结果表明,BiFPN 特征融合能够提高多尺度特征的融合效率,从而提高模型的检测精度,本文将该最终模型称为 YOLO-Mura。

表 5 改进的特征融合

Tab. 5 Improved feature fusion

特征融合方式	Precision/%	Recall/%	mAP@0.5/%
自顶向下	92.7	83.3	92.1
自底向上	94.8	83.4	93.4
PANet	96.0	95.7	96.5
ASFF	98.4	95.9	97.0
BiFPN	98.6	96.0	97.2

4.4 消融实验

为了验证改进方法对液晶屏微弱特征缺陷检测模型性能的影响,本文对改进模型进行了消融实验。通过向基线 (Baseline) 模型 YOLOv5s 中加入 Involution 算子、CARAFE 算子、BiFormer 模块及 BiFPN 模块进行模型的改进,其他训练参数设置与本文最终提出的模型一致。其中“√”代表在网络中使用了对应改进策略。

表 6 实验数据表明,B 组在 YOLOv5s 主干网络中单独引入 Involution 层,使 FLOPs 减少 66.5%、精确率降低了 2.9%、召回率提升 0.6%、mAP@0.5 降低 2.4%。由此可知,引入 Involution 算子能在损失少量模型性能的同时有效降低了模型的复杂程度。C 组单独引入了 CARAFE 上采样算子,与 A 组 YOLOv5s 模型相比,其召回率、mAP@0.5 分别增加了 2.8%、0.3%。D 组在 B 组 YOLO-I 的基础上添加上采样算子,其精确率、召回率、mAP@0.5 分别增加了 1.1%、1.5% 和 2.7%。由此可见,更换 CARAFE 上采样能弥补因降低模型复杂程度带来的性能损失。为了进一步提升模型性能,在 D 组 YOLO-IC 模型的基础上,G 组添加了 BiFormer 注意力模块,其精

表6 消融实验结果

Tab. 6 Ablation experimental results

序号	Involution	CARAFE	BiFormer	BiFPN	FLOPs/G	Precision/%	Recall/%	mAP@0.5/%	FPS
A					15.8	96.4	88.9	94.5	99.0
B	✓				5.3	93.5	91.5	92.1	131.6
C		✓			16.5	94.0	92.7	94.8	116.3
D	✓	✓			5.9	94.6	93.0	93.8	122.4
E	✓		✓		6.0	94.2	93.1	92.6	119.0
F		✓	✓		19.1	94.5	90	93.2	112.5
G	✓	✓	✓		6.1	96.0	95.7	96.5	109.8
H	✓	✓		✓	5.8	94.8	93.5	94.3	113.5
J	✓		✓	✓	5.6	96.2	95.4	94.6	110.3
I	✓	✓	✓	✓	6.3	98.6	96.5	97.2	103.1

准确率、召回率、mAP@0.5、FLOPs分别增加了1.4%、2.7%、2.3%和0.2%。H组添加了BiFPN特征融合模块,其精确率、召回率、mAP@0.5分别增加了0.2%、0.5%和0.5%,FLOPs降低了0.1%。由此可见,添加BiFormer注意力模块和添加BiFPN特征融合模块能提升模型性能,且对模型复杂程度只造成较小的影响。I组为最终改进的YOLO-Mura模型,相比A组基线YOLOv5s模型,其精确率、召回率、mAP@0.5、FPS分别增加了2.2%、6.6%、2.7%和4.1。以上结果表明,最终改进的YOLO-Mura模型对于微弱特征Mura缺陷具有较高的检测精度,同时具有较快的检测速度。综合上述研究,确定最终改进后的模型YOLO-Mura达到了综合最佳的改进效果。

4.5 热力图分析

深度CNN的特征图的最后一层包含许多高级语义特征和详细的空间信息。特征图的最后一层的可视化说明了图像的哪一部分对模型的最终输出贡献更大^[27]。在图9中,红色突出显示的区域表示模型集中注意力的感兴趣区域,它是评估模型检测液晶屏缺陷特征能力的主要依据。如图9(b)、(c)所示,YOLO-I模型具有更集中的感兴趣区域,但缺陷区域的覆盖面积较小,表明Involution算子的结合使骨干网络能够在更大的空间区域上聚合上下文信息。然而,通道维度共享降低了模型的特征提取能力。如图9(c)、(d)所示,YOLO-IC模型的感兴趣区域更好地代表了Mura缺陷的特征区域,表明执行基于内容的上采样的CARAFE算子提高了模型捕获详细特征的

能力。如图9(d)、(e)所示,YOLO-ICB模型的感兴趣区域更好地聚焦在缺陷区域,提高了模型的特征提取能力并抑制了非关键信息。图9(e)、(f)的比较表明,YOLO-Mura模型的感兴趣区域进一步聚焦在缺陷区域,并准确地描绘了缺陷特征的感兴趣区域。

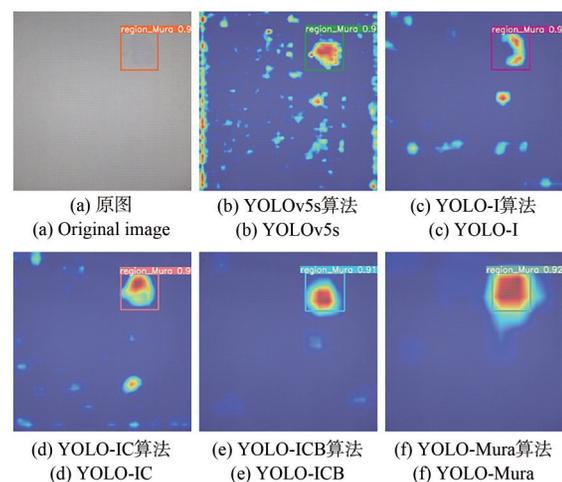


图9 改进算法的热力图

Fig. 9 Heat map of the improved algorithm

4.6 对比度分析

对比度指的是一幅图像中最亮区域和最高暗区域之间的差异范围,差异范围越大表示对比度越大,差异范围越小表示对比度越小。对应图像的灰度值为0~255,数值越大表示该区域越亮。测量背景区域的平均灰度值和缺陷区域的平均灰度值,即可对模型能检测出微弱缺陷的最低对比度进行量化分析。使用500张液晶屏微弱缺陷

图片进行测试,测试结果如表6所示。YOLOv5s模型和最终改进模型YOLO-Mura的漏检率分别为9%、1.2%,漏检样本背景区域与缺陷区域的平均灰度值差值分别为4.82、2.62。结果表明,最终改进模型能够较好地减少漏检率以及提高低对比度缺陷的检测能力。

表7 漏检缺陷的灰度值对比

Tab. 7 Comparison of grayscale values of missed defects

模型	漏检数量	背景平均 灰度值	缺陷平均 灰度值	差值
YOLOv5s	45	184.57	179.75	4.82
YOLO-Mura	6	184.74	182.12	2.62

4.7 不同目标检测算法对比

在实验数据集、实验软硬件设备以及相关参数一致的情况下,通过与YOLOv3-Tiny、YOLOv4s-Mish、YOLO5n、YOLOv5s等算法进行对比实验,评估所提出的模型与主流的目标检测算法之间的差距,结果如表8所示。在综合考虑下,YOLOv5s的基础较好,所以本文选择YOLOv5s作为基础模型进行改进。由表8可见,本文最终改进的YOLO-

表8 不同算法性能对比

Tab. 8 Performance comparison of different algorithms

Model	FLOPs/		Precision/	Recall/	mAP@0.5
	G		%	%	
YOLOv3-Tiny	12.9		84.9	88.2	90.1
YOLOv4s-Mish	20.6		91.6	93.5	95.4
YOLOv5n	4.2		93.6	90.1	93.2
YOLOv5s	15.8		96.4	88.9	94.5
YOLOv7-Tiny	13.0		87.1	86.3	90.5
YOLOv8n	8.0		92.8	92.2	94.8
YOLOv8s	28.4		93.4	92.9	95.3
YOLO-Mura	6.3		98.6	96.5	97.2

参 考 文 献:

- [1] 钱基德,陈斌,钱基业,等. 基于机器视觉的液晶屏Mura缺陷检测方法[J]. 计算机科学,2018,45(6):296-300,313. QIAN J D, CHEN B, QIAN J Y, *et al.* Machine vision based inspection method of Mura defect for LCD [J]. *Computer Science*, 2018, 45(6): 296-300, 313. (in Chinese)
- [2] LIN G M, KONG L F, LIU T J, *et al.* An antagonistic training algorithm for TFT-LCD module Mura defect detection [J]. *Signal Processing: Image Communication*, 2022, 107: 116791.
- [3] ZHANG C H, GAO S S, PAN X, *et al.* Coarse to fine: weak feature boosting network for salient object detection [J].

Mura模型的精确率、召回率、mAP@0.5最高。针对液晶屏微弱特征Mura缺陷构建的检测算法,通过增加网络层数并替换不同网络模块,使算法的网络构架在满足高精度、低FLOPs的同时进一步减少了资源消耗。

5 结 论

本文针对微弱特征的Mura缺陷经多次卷积与背景纹理同化导致的检测精度低的问题,提出了一种基于改进YOLOv5s的液晶屏微弱特征Mura缺陷检测算法YOLO-Mura。首先,通过引入Involution算子构建新骨干网络,缓解了卷积神经网络有效感受野受限、特征通道冗余的问题,使FLOPs降低了66.5%,在大幅降低模型复杂程度的同时仅对模型性能造成轻度损失。为了补偿降低模型复杂程度带来的性能损失,通过引入CARAFE模块改进上采样方式,对Mura缺陷的关键信息进行加权强化,抑制背景干扰,提高了模型检测性能。其次,为了提升模型对微弱特征Mura缺陷的检测性能,在颈部网络嵌入BiFormer注意力模块,以查询自适应的方式增强相关标记的注意力。为了进一步优化模型性能,引入BiFPN加权双向特征金字塔网络,该网络可以实现简单且快速的多尺度特征融合,利用低层级的空间位置信息并将其引入高层级特征中,从而增强了特征的表达能力和语义信息。本文最终改进的模型YOLO-Mura与基线模型YOLOv5s相比,其精确率、召回率、mAP@0.5、FPS分别提升了2.2%、6.6%、2.7%和4.1,模型复杂程度FLOPs降低了66.5%。将YOLO-Mura模型与主流的目标检测算法进行了充分的对比,结果表明本文最终改进模型在不影响推理速度的前提下,降低了模型复杂程度,提升了算法的检测性能。

- Computer Graphics Forum*, 2020, 39(7): 411-420.
- [4] DU J M, LU H Z, ZHANG L P, *et al.* A spatial-temporal feature-based detection framework for infrared dim small target [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 3000412.
- [5] TAO X, ZHANG D P, HOU W, *et al.* Industrial weak scratches inspection based on multifeature fusion network [J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 5000514.
- [6] 汪西莉,梁正印,刘涛. 基于特征注意力金字塔的遥感图像目标检测方法[J]. *遥感学报*, 2023, 27(2): 492-501.
WANG X L, LIANG Z Y, LIU T. Feature attention pyramid-based remote sensing image object detection method [J]. *National Remote Sensing Bulletin*, 2023, 27(2): 492-501. (in Chinese)
- [7] 周葳楠,吴治海,张正道,等. 基于弱特征增强的轻量化小目标检测方法[J]. *控制与决策*, 2024, 39(2): 381-390.
ZHOU W N, WU Z H, ZHANG Z D, *et al.* Lightweight small target detection method based on weak feature enhancement [J]. *Control and Decision*, 2024, 39(2): 381-390. (in Chinese)
- [8] LI D, HU J, WANG C H, *et al.* Involution: inverting the inherence of convolution for visual recognition [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021: 12316-12325.
- [9] YANG F, WANG T, WANG X F. Student classroom behavior detection based on YOLOv7-BRA and multi-model fusion [J/OL]. *arXiv*, 2023: 2305.07825.
- [10] BIE M L, LIU Y Y, LI G N, *et al.* Real-time vehicle detection algorithm based on a lightweight You-Only-Look-Once (YOLOv5n-L) approach [J]. *Expert Systems with Applications*, 2023, 213: 119108.
- [11] 周森森,汤全武,石甜甜,等. 基于改进YOLOv5s的轨轨表面裂纹检测算法[J]. *液晶与显示*, 2023, 38(5): 666-679.
ZHOU M S, TANG Q W, SHI T T, *et al.* Rail surface crack detection algorithm based on improved YOLOv5s [J]. *Chinese Journal of Liquid Crystals and Displays*, 2023, 38(5): 666-679. (in Chinese)
- [12] 杨洪刚,陈洁洁,徐梦飞. 双线性内卷神经网络用于眼底疾病图像分类[J]. *计算机应用*, 2023, 43(1): 259-264.
YANG H G, CHEN J J, XU M F. Bilinear involution neural network for image classification of fundus diseases [J]. *Journal of Computer Applications*, 2023, 43(1): 259-264. (in Chinese)
- [13] 安鹤男,邓武才,管聪,等. 基于Involution Prediction Head的小目标检测算法[J]. *电子技术应用*, 2022, 48(11): 19-23.
AN H N, DENG W C, GUAN C, *et al.* Small object detection algorithm based on Involution Prediction Head [J]. *Application of Electronic Technique*, 2022, 48(11): 19-23. (in Chinese)
- [14] DING X H, ZHANG X Y, HAN J G, *et al.* Scaling up your kernels to 31×31 : revisiting large kernel design in CNNs [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans: IEEE, 2022: 11953-11965.
- [15] 杨鹤,柏正尧. CoT-TransUNet:轻量化的上下文Transformer医学图像分割网络[J]. *计算机工程与应用*, 2023, 59(3): 218-225.
YANG H, BAI Z Y. CoT-TransUNet: lightweight context Transformer medical image segmentation network [J]. *Computer Engineering and Applications*, 2023, 59(3): 218-225. (in Chinese)
- [16] WANG J Q, CHEN K, XU R, *et al.* CARAFE++: unified content-aware ReAssembly of FEatures [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(9): 4674-4687.
- [17] WANG J Q, CHEN K, XU R, *et al.* CARAFE: content-aware ReAssembly of FEatures [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019: 3007-3016.
- [18] ZHU L, WANG X J, KE Z H, *et al.* BiFormer: vision transformer with Bi-level routing attention [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE, 2023: 10323-10333.
- [19] PARK J, KIM J, KIM C S. BiFormer: learning bilateral motion estimation via bilateral transformer for 4K video frame interpolation [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE, 2023: 1568-1577.
- [20] SYAZWANY N S, NAM J H, LEE S C. MM-BiFPN: multi-modality fusion network with Bi-FPN for MRI brain tumor segmentation [J]. *IEEE Access*, 2021, 9: 160708-160720.
- [21] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection [C]//*Proceedings of the*

- IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 10778-10787.
- [22] LIN T Y, DOLLÁR P, GIRSHICK R, *et al.* Feature pyramid networks for object detection [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 936-944.
- [23] LIU S, QIL, QIN H F, *et al.* Path aggregation network for instance segmentation [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 8759-8768.
- [24] LI T G, ZHANG Y Z, LI Q Q, *et al.* AB-DLM: an improved deep learning model based on attention mechanism and BiFPN for driver distraction behavior detection [J]. *IEEE Access*, 2022, 10: 83138-83151.
- [25] XIONG R B, YANG Y C, HE D, *et al.* On layer normalization in the transformer architecture [C]//*Proceedings of the 37th International Conference on Machine Learning*. Virtual: PMLR, 2020: 10524-10533.
- [26] 董小伟, 韩悦, 张正, 等. 基于多尺度加权特征融合网络的地铁行人目标检测算法[J]. *电子与信息学报*, 2021, 43(7): 2113-2120.
- DONG X W, HAN Y, ZHANG Z, *et al.* Metro pedestrian detection algorithm based on multi-scale weighted feature fusion network [J]. *Journal of Electronics & Information Technology*, 2021, 43(7): 2113-2120. (in Chinese)
- [27] SELVARAJU R R, COGSWELL M, DAS A, *et al.* Grad-CAM: visual explanations from deep networks via gradient-based localization [C]//*Proceedings of the IEEE International Conference on Computer Vision*. Venice: IEEE, 2017: 618-626.

作者简介:



林 峰,男,硕士研究生,2020年于湖南涉外经济学院获得学士学位,主要从事图像处理、目标检测、机器学习等方面的研究。E-mail:1067691765@qq.com



石 艳,女,博士研究生,教授,2008年于四川大学获得硕士学位,主要从事过程装备优化设计理论、智能装备设计与应用技术等方面的研究。E-mail:sy71Email@163.com