

[自控·检测]

DOI:10.3969/j.issn.1005-2895.2022.02.009

# 基于改进 Yolo v5 的织物缺陷检测方法

王恩芝, 张团善, 刘 亚

(西安工程大学机电工程学院, 陕西 西安 710613)

**摘 要:**针对纺织品表面缺陷检测算法速度低、泛化性能差及鲁棒性弱等问题,课题组提出了一种基于改进 Yolo v5 的织物表面缺陷检测方法。在 Yolo v5 骨干网络的基础上增加了卷积注意力模块,增强目标检测网络对特征图中重要信息提取并弱化无关特征;针对网络特征融合阶段特征尺度不一致造成的冲突问题,引入自适应空间特征融合的方法;在训练过程中使用迁移学习加快训练速度。实验结果表明:与未改进的 Yolo v5 算法相比,所提出的检测框架能够有效提高网络精度达 98.8%,检测速度达 83 帧/s。该检测方法能满足实际工业要求。

**关键词:**缺陷检测;深度学习;目标识别;卷积注意力机制;自适应空间特征融合

中图分类号:TS107.8;TP 391.4 文献标志码:A 文章编号:1005-2895(2022)02-0054-07

## Fabric Defect Detection Method Based on Improved Yolo v5

WANG Enzhi, ZHANG Tuanshan, LIU Ya

(School of Mechanical and Electrical Engineering, Xi'an Polytechnic University, Xi'an 710613, China)

**Abstract:** A fabric defect detection method based on improved Yolo v5 was proposed in response to the problems of low speed, poor generalization performance and weak robustness of the textile surface defect algorithm. The convolutional attention module was added to the Yolo v5 backbone network to enhance the target detection network for extracting important information in the feature maps and weakening irrelevant features. For the conflict problem caused by inconsistent feature scales in the feature fusion stage of the network, an adaptive spatial feature fusion method was introduced. Transfer learning was used to accelerate the training speed during the training process. Experimental results show that the proposed detection framework can effectively improve the network accuracy up to 98.8% and detection speed up to 83 FPS compared with the unimproved Yolo v5 algorithm. The proposed detection method can meet the practical industrial requirements.

**Keywords:** defect detection; deep learning; object recognition; convolutional block attention; adaptive spatial feature fusion

织物表面缺陷检测是织物质量控制过程中一项关键步骤。由于织物样式不一、种类繁多,给织物检测算法研究带来困难。长期以来,纺织厂中织物表面存在的缺陷都是通过工人手工检测。据统计人工验布速度约为 12 m/min,由于各种干扰因素,验布员能专心验布时间一般只能达到 20~30 min<sup>[1]</sup>。为提高工厂生产效率,降低劳动成本,自动表面缺陷检测已经变得越来越普遍。随着计算机算力提高和软件技术的迅速发

展,卷积神经网络<sup>[2]</sup>也卷起了一阵热潮,计算机视觉中的深度学习的方法在近几年也达到了非常先进的水平,成为许多难以解决问题的一个端到端的解决方案。

针对纺织行业织物表面出现缺陷的问题,大量学者提出不同的织物缺陷检测算法,主要分为基于传统的图像处理与机器学习相结合的方法、深度学习的方法。基于传统的图像处理与机器学习结合的方法又可分为:基于频谱、模型和统计等方法<sup>[3]</sup>。传统缺陷检

收稿日期:2021-10-27;修回日期:2021-12-26

第一作者简介:王恩芝(1998),女,陕西蓝田人,硕士研究生,主要研究方向为数字图像处理与深度学习。E-mail:wangenzhi\_11@163.com

测的方法首先对图像进行预处理提高图像质量,再通过傅里叶变换或小波变换等方法人为设计特征或提取特征,最后利用机器学习的方法对提取出的特征进行分类、识别。朱浩等<sup>[4]</sup>提出了一种多纹理分级融合的方法,通过计算缺陷区域与正常图像块特征的相似度,得到 2 者相似图,将初步定位的缺陷区域的经纬向特征图与相似度特征图进行融合,检测缺陷存在区域。该方法只对具有较强周期性纹理的织物有效,当算法面对纹理差异性大、缺陷形态复杂、种类多样的情况,传统的机器学习检测算法容易出现漏检或错检等情况。张团善等<sup>[5]</sup>设计了一种较多噪声干扰的情况下的防羽布疵点检测算法,通过设计基矩阵并计算其差异矩阵,与疵点因子进行比较获得缺陷二值图,但此方法只适用无色差且密度较高的纯色防羽布,具有复杂图案的织物难以检测。基于深度学习的方法通常使用卷积神经网络进行特征提取,在应用分类器或者检测头对缺陷区域进行检测。安萌等<sup>[6]</sup>基于 Faster R-CNN 设计了一种新的网络框架,通过在候选框提取网络(region proposal network, RPN)中引入不同尺度的特征金字塔结构,增强网络提取细节特征的能力,但 Faster R-CNN 是一个二阶段检测网络,速度无法达到实时检测缺陷的效果,无法满足工业检测速度。针对织物缺陷检测过程速度慢的问题,周君等<sup>[7]</sup>提出了一种基于 S-YOLO v3 模型的织物实时缺陷检测算法,通过利用批归一化(BN)层的缩放因子评估每个卷积核对于神经网络的权重,对神经网络进行剪枝,设计一种轻量级的模型实现织物的实时检测,但该方法精度只能达到 94%,对于工厂该精度仍需提高。

针对以上算法存在的问题,课题组设计了基于改进 Yolo v5 的织物表面缺陷检测方法,可以在满足工厂检测精度要求的同时满足检测实时性要求,实现纺织行业检测自动化。

## 1 织物缺陷检测模型

目标检测算法分为 2 种框架:一种是一阶段检测算法(one-stage)直接设计一个端到端的提取特征、目标检测的框架,例如 Yolo<sup>[8-11]</sup>系列和 SSD<sup>[12]</sup>网络等;另一种是二阶段检测算法(two-stage),在进行目标识别过程中,需要首先对区域框进行检测,例如 R-

CNN<sup>[13-15]</sup>系列网络等。Yolo 系列网络是目前最主流且效果最好的基于锚框(anchor)的目标检测算法之一。最新的 Yolo v5<sup>[16]</sup>项目应用多种多样,也证明了 Yolo v5 算法的有效性。为提高 Yolo v5 网络的性能,课题组在网络的骨干网络(backbone)阶段引入卷积注意力机制(concolutional block attention module, CBAM)<sup>[17]</sup>通过对像素从语义和空间信息分别进行评分,抑制对检测不重要的特征。在金字塔注意力结构(pyramid attention network, PANet)之后引入自适应空间融合(adaptive spatial feature fusion, ASFF)<sup>[18]</sup>的方法,抑制梯度反传过程的不一致,增强网络检测精度。

### 1.1 Yolo v5 基础结构

Yolo v5 是目前 YOLO 系列算法的 SOTA(state-of-the-art),与其他的目标检测算法一样由输入、骨干网络(backbone)、特征融合(neck)、预测(prediction)结构组成。Yolo v5 包含 4 种不同宽度和深度的模型,分别为 YOLOv5s, YOLOv5m, YOLOv5l 和 YOLOv5x。借鉴 EfficientNet<sup>[19]</sup>的方法,通过一个常数平衡网络所有维度的宽度和深度。本研究的任务是织物缺陷检测,为满足工业上准确率的要求,在网络结构尽可能小的情况下,实现更高的精度和更快的检测速度,课题组以 YOLOv5l 为研究对象。Yolo v5 网络结构如图 1 所示。

Yolo v5 为增强网络模型的鲁棒性和泛化性能:在输入端引入 Mosaic 数据增强,自适应锚框计算,自适应图像缩放和可适用于各种尺寸的图像输入;在 backbone 阶段引入一种新的 Focus 结构。Focus 结构对图像进行切片,并在通道方向进行拼接,有效减少了信息丢失。Neck 部分使用特征金字塔网络(feature pyramid networks, FPN) + PANet, FPN 通过上采样将高层特征信息与底层特征进行融合,传递更强的语义特征,增强网络模型学习图像特征的能力,但可能丢失一些定位信息。PAN 对 FPN 获得的图像通过自下向上传递强的定位信息,将两者同时使用达到互补效果,增强模型的特征融合能力,提高了模型的鲁棒性。

### 1.2 改进的 Yolo v5 结构

#### 1.2.1 主干网络改进

为提高主干网络对目标特征提取能力,课题组在 backbone 阶段引入软注意力机制(soft-attention)中的卷积注意力机制模板(convolution block attention

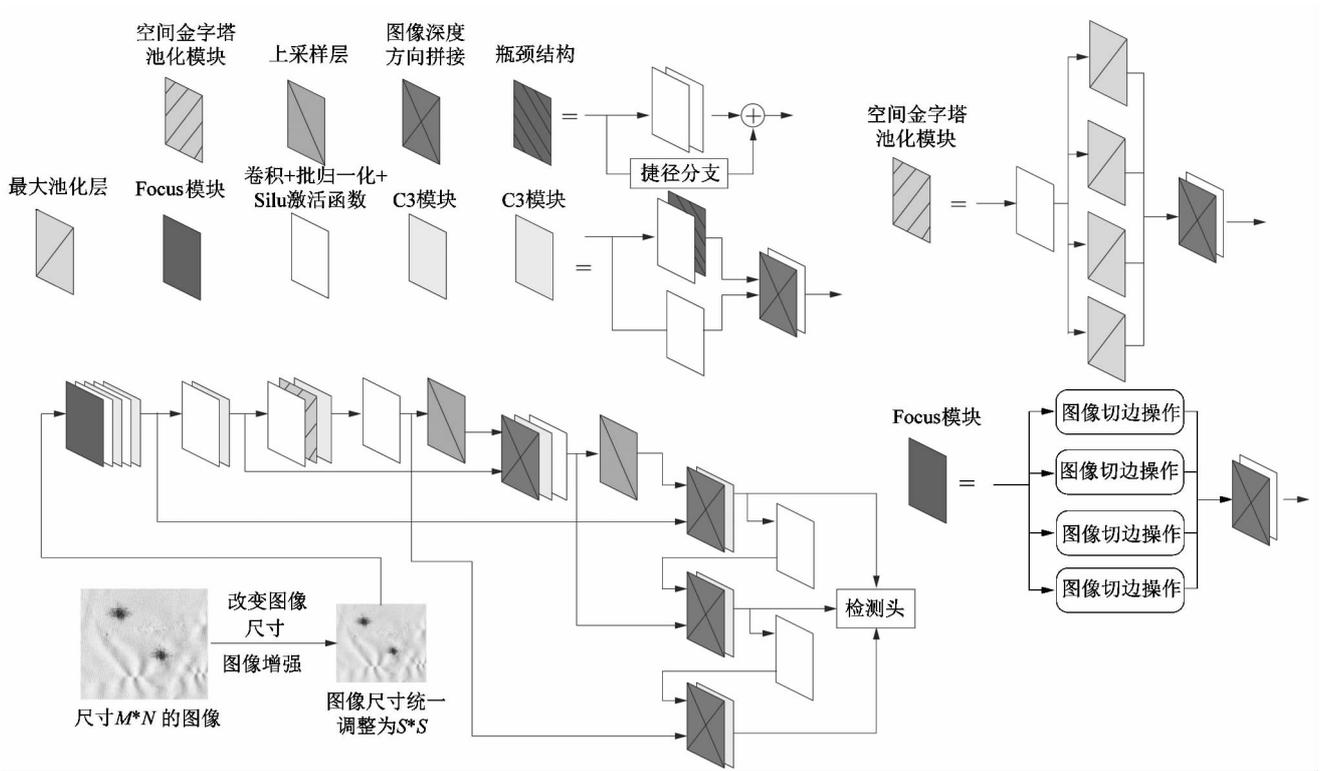


图1 Yolo v5 网络结构

Figure 1 Yolo v5 network structure

module, CBAM), 该模块可以在空间维和通道维分别推断出注意力的权重,使网络能够关注到图像中关于目标的重点信息。如图2所示,新的主干网络在每一个跨阶段连接网络(cross stage partial, CSP)结构之后引入一个CBAM模块。

CBAM分为空间注意力机制(spatial attention)和通道注意力机制(channel attention),分别从空间维和通道维增强对目标的关注程度。通道注意力模块对输入特征图分别进行全局最大池化和全局平均池化聚合图像的空间信息,经过2个多层感知机结构构建模型

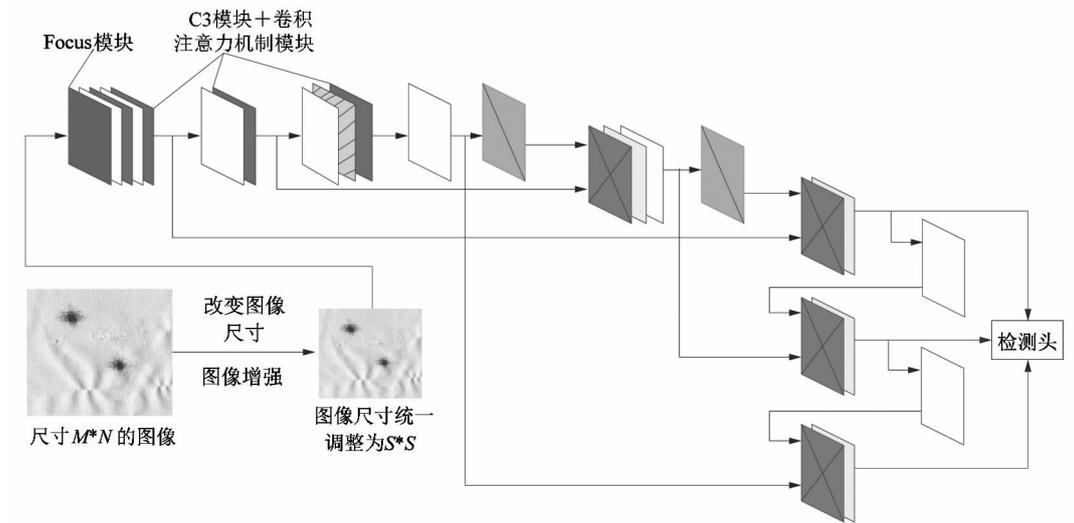


图2 改进 backbone 的 Yolo v5 结构

Figure 2 Improvement to backbone's Yolo v5 structure

之间的相关性,最后经过 Sigmoid 激活函数获得每一个通道的权重,通道注意力机制如图 3 所示,表示形式:

$$M_c(\mathbf{F}) = \sigma[W_1(W_0(\mathbf{F}_{avg}^c)) + W_1(W_0(\mathbf{F}_{max}^c))]$$

式中: $\sigma$  表示对所有数据进行 Sigmoid 激活操作; $\mathbf{F}_{avg}^c$

表示对输入特征矩阵进行全局平均池化; $\mathbf{F}_{max}^c$  表示对输入矩阵进行全局最大池化; $W_0$  表示第 1 个全连接层的权重; $W_1$  表示第 2 个全连接层的权重; $M_c(\mathbf{F})$  表示通道注意力模块的输出特征图。

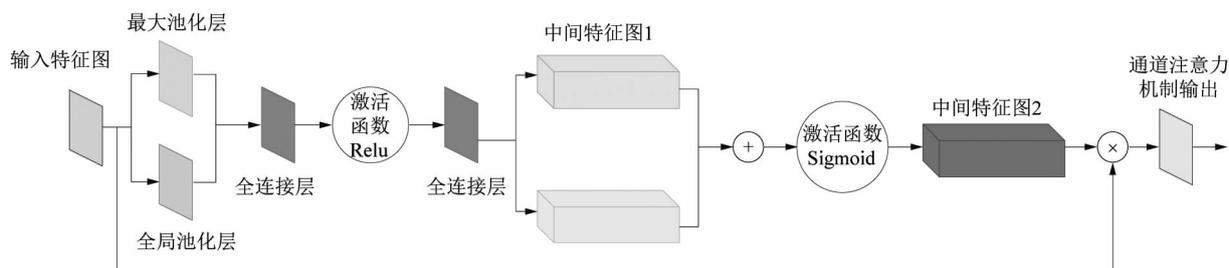


图 3 通道注意力机制模块

Figure 3 Channel attention block

空间注意力模块将通道注意力模块中的输出作为本模块的输入特征图。首先对输入特征图做一个基于通道的全局最大池化和全局平均池化;然后将 2 个单通道特征图进行通道间拼接,而后通过一个分辨率为  $7 \times 7$  的卷积对特征图进行降维获得一个单通道特征图,便于最后特征融合;最后经过 Sigmoid 激活函数获得图像空间位置的每一个特征权重。空间注意力机制

如图 4 所示,表示形式为:

$$M_s(\mathbf{F}) = \sigma(f^{7 \times 7}([\mathbf{F}_{avg}^s; \mathbf{F}_{max}^s]))$$

式中: $f^{7 \times 7}$  表示对图像进行  $7 \times 7$  的卷积; $\mathbf{F}_{avg}^s$  表示对输入特征矩阵进行全局平均池化; $\mathbf{F}_{max}^s$  表示对输入矩阵进行全局最大池化; $M_s(\mathbf{F})$  表示空间注意力模块的输出特征图。

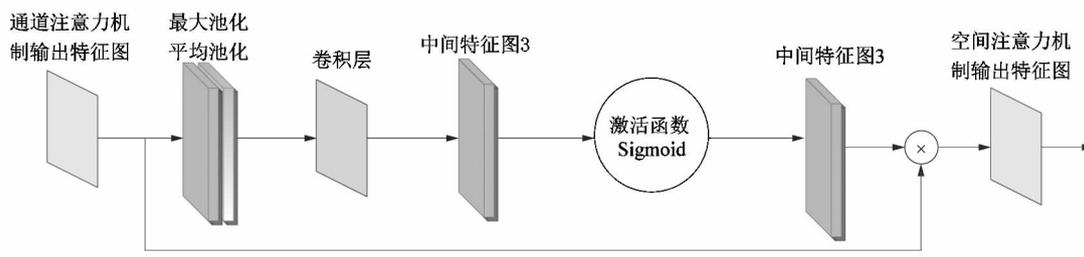


图 4 空间注意力机制

Figure 4 Spatial attention module

### 1.2.2 特征融合阶段改进

为了更加充分利用网络中高层特征的语义信息和底层特征的空间信息,课题组在 Yolo v5 的 Neck 阶段引入自适应空间特征融合 (adaptive spatial feature fusion, ASFF), 分别对 PANet 输出的 3 个水平特征图进行加权融合,通过增加可学习的参数抑制了梯度反传过程的不一致,充分利用不同尺度的特征。修改后的自适应空间特征融合的 Neck 如图 5 所示。

改进后的 Yolo v5 模型如图 6 所示,输入图像首先

经过由 CBAM 增强的 backbone 提取特征,然后对特征图应用 FPN + PANet + ASFF 进行特征融合,充分利用特征图中高层语义信息和底层空间信息,最大限度获得目标信息,而后将获得的特征图送入检测头中,检测头分别输出 3 个尺度  $8 \times 8$ ,  $16 \times 16$  和  $32 \times 32$  的输出检测层,分别用于检测大、中和小 3 种目标。

## 2 实验及结果分析

### 2.1 软硬件环境

实验中采用的硬件配置:CPU 为 Intel (R) Core

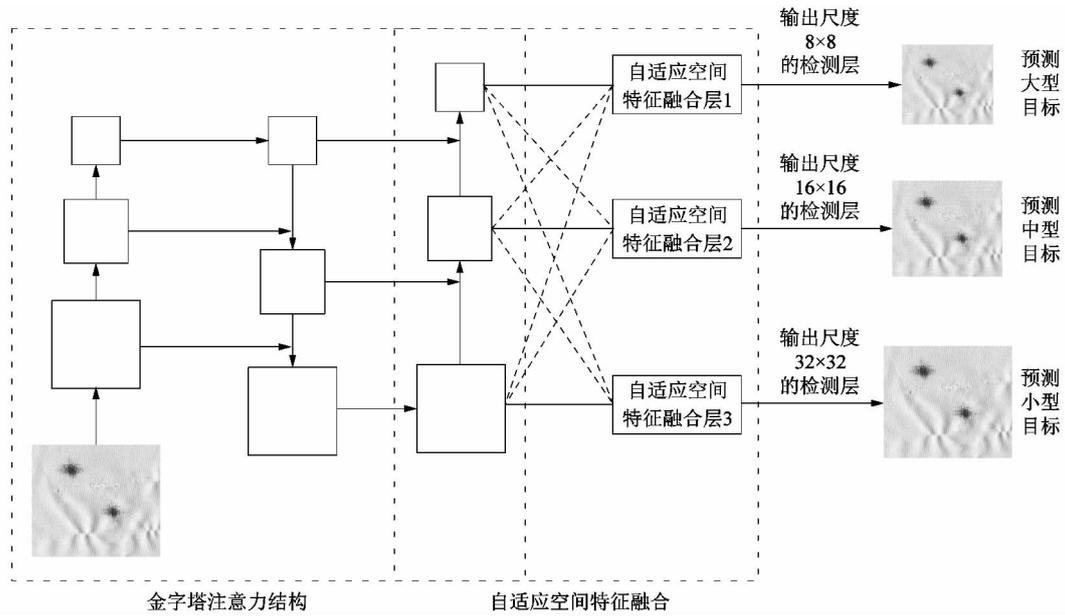


图 5 自适应空间特征融合 Neck

Figure 5 Adaptive spatial feature fusion Neck

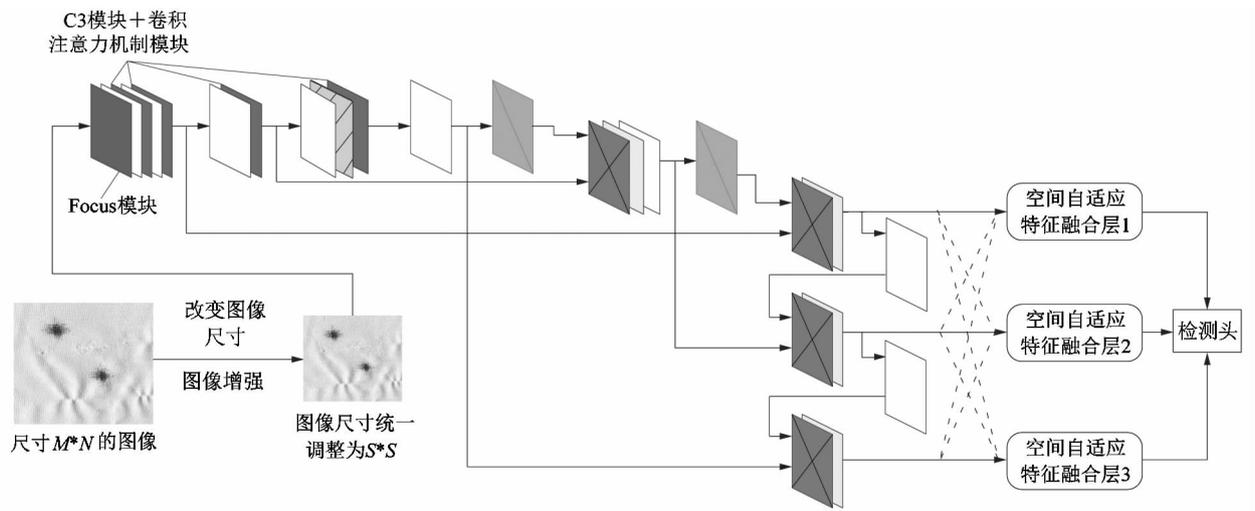


图 6 改进后的 Yolo v5 模型

Figure 6 Improved Yolo V5 model

(TM) i9-9900K, 主频为 3.60 GHz, 内存为 32 GiB, 显卡为 NVIDIA; 软件配置为 Win10 操作系统, 显卡驱动为 NVIDIA GeForce RTX 2080Ti 和 CUDA10.1, 深度学习框架采用 PyTorch-1.7.1, 使用 LableMe 作为标注工具, 编程语言为 Python3.7。

## 2.2 网络训练

本算法的验证数据集使用工业 CCD 相机获取, 手动采集 3 000 张图片, 图像包含复杂纹理的织物和简单纹理的织物, 且样本表面至少存在一种缺陷, 手动裁

剪为尺寸 256 像素 × 256 像素, 部分具有缺陷的织物样本如图 7 所示。深度学习网络模型的训练数据都是基于大量数据驱动, 为增强模型的鲁棒性和泛化能力, 对采集到的图像随机进行旋转、缩放, 及改变样本的亮度、色度和饱和度, 将原来的 3 000 张图片扩充为 10 000 张图片, 随机选取 7 000 张图片作为训练集, 2 000 张图片作为验证集, 1 000 张图片作为测试集。数据集中分别包含毛边、孔洞 (包括擦洞和破洞)、黄渍、断经纬、滴墨和破损 6 种缺陷。

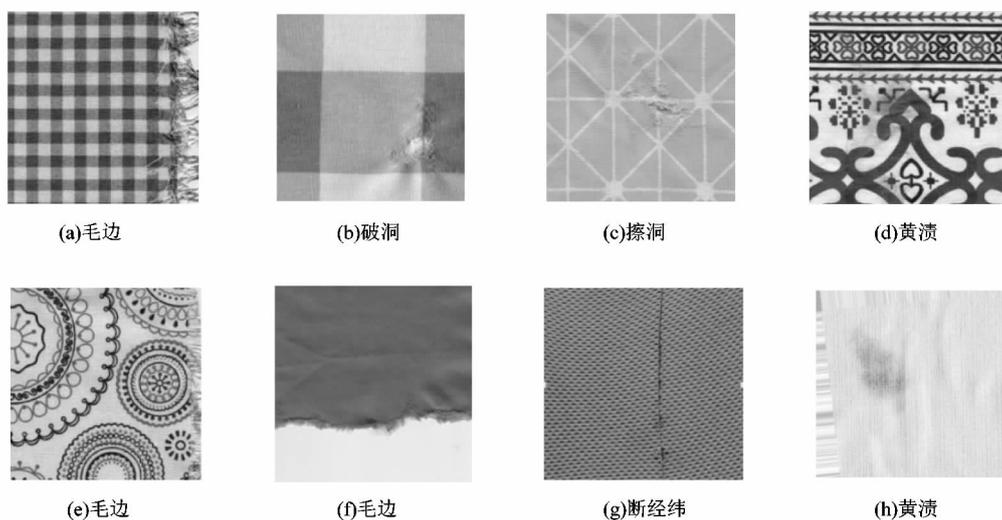


图7 部分含有缺陷的织物样本

Figure 7 Some defective fabric samples

经过多次实验,初始学习率设为 0.001,迭代 300 代之后学习率调为 0.000 1,批大小为 16,权重衰减系数为 0.000 5,动量为 0.937,优化器使用随机梯度下降,总训练次数为 1 000。

### 2.3 算法改进前后对比

考虑到工厂中的实际应用,课题组采用准确率 (precision) 和均值平均精度 (mean average precision, mAP) 作为模型的评价指标。Precision 能够准确、合理地评价算法的定位和目标检测能力;使用  $P_{mA}$  来衡量网络的算法性能,适用于本研究的多标签缺陷图像分类任务,并将改进后的网络与常用的 SSD 网络进行结果比较,公式如下:

$$P = \frac{T_p}{T_p + F_N};$$

$$P_{mA} = \frac{\sum_{n=1}^N P(n) \Delta P(n)}{C}。$$

式中: $T_p$  表示本该是正样本的样本被算法认为是正样本,预测正确; $F_N$  表示本该是正样本的样本被算法认为是负样本,预测错误; $N$  为测试集中样本总数; $P(n)$  是在同时识别  $n$  个样本时的大小; $\Delta P(n)$  表示召回率在检测样本的个数从  $n-1$  变为  $n$  时的变化情况; $C$  为多分类任务中类别个数。

算法改进前后结果如表 1 所示。

表 1 算法改进前后结果比较

Table 1 Comparison of results of before and after improvement algorithm

算法	准确率 $P/\%$	均值平均 精度 $P_{mA}/\%$	检测速度/ (帧 · s <sup>-1</sup> )
Yolo v5	0.937	0.955	100
Yolo v5 + CBAM + ASFF	0.988	0.935	83
SSD	0.784	0.786	48

图 8 所示为改进之后的网络模型检测结果,从效果图看,改进的算法对毛边、孔洞、黄渍、断经纬、滴墨、破损 6 种缺陷检测精度较高,经测算达 98.8%。

### 3 结论

课题组针对织物表面存在缺陷,传统方法检测率低、泛化性能差且实时性低的问题提出一种基于改进 Yolo v5 算法的织物缺陷检测算法。实验结果表明:在 Yolo v5 基础网络中增加卷积注意力机制和自适应空间特征融合可以提高模型对缺陷目标的关注程度,并增强语义信息与空间信息的融合程度,在检测速度仅降低 17 帧/s 的情况下将疵点检测的准确率提高至 98.8%。课题组设计的织物缺陷检测算法实时性高、且准确率高,能够满足实际工厂要求。该算法仍然严重依赖大量具有多样性的样本,设计一种基于小样本的织物表面缺陷检测算法是下一步的研究重点。

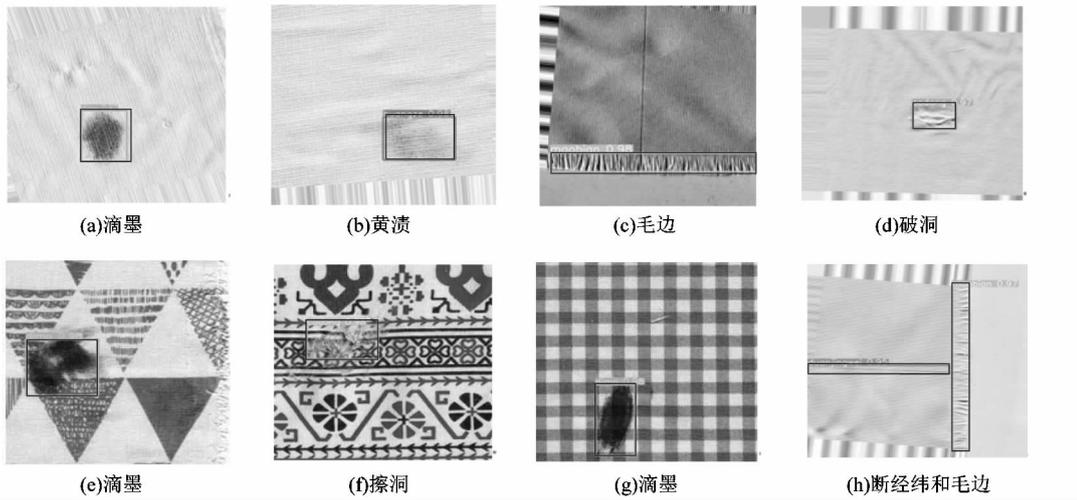


图 8 改进模型检测结果

Figure 8 Improved model detection results

参考文献:

[1] 刘纪,张团善,徐杰.应用对抗网络和霍夫变换的织物点状缺陷检测研究[J].毛纺科技,2020,48(12):62-66.

[2] HUANG Y H, QIU C Y, WANG X N, et al. A compact convolutional neural network for surface defect inspection [J]. Sensors, 2020, 20(7):1974.

[3] 刘纪,张团善,李秀昊.基于L0梯度最小化和K-Means聚类的织物缺陷检测研究[J].轻工机械,2021,39(1):67-71.

[4] 朱浩,丁辉,尚媛园,等.多纹理分级融合的织物缺陷检测算法[J].纺织学报,2019,40(6):118-125.

[5] 张团善,石玮.噪声干扰下的防羽布疵点检测算法[J].西安工程大学学报,2020,34(1):14-19.

[6] 安萌,郑颺默,王诗宇,等.一种改进Faster R-CNN的面料疵点检测方法[J].小型微型计算机系统,2021,42(5):1029-1033.

[7] 周君,景军锋,张缓缓,等.基于S-YOLOv3模型的织物实时缺陷检测算法[J].激光与光电子学进展,2020,57(16):55-63.

[8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779-788.

[9] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]// IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI: IEEE, 2017: 6517-6525.

[10] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J/OL]. (2018-04-08) [2021-10-10]. <https://arxiv.org/abs/1804.02767>.

[11] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J/OL]. (2020-04-23) [2021-10-10]. <https://doi.org/10.48550/arxiv.2004.10934>.

[12] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: single shot multibox detector [C]//European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[13] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 580-587.

[14] GIRSHICK R. Fast R-CNN [C]//Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 1440-1448.

[15] REN S Q, HE K M, GIRSHICK R, et al. Faster r-cnn: towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28: 91-99.

[16] 彭成,张乔虹,唐朝晖,等.基于YOLOv5增强模型的口罩佩戴检测方法研究[J/OL].计算机工程,2021:1-12 [2021-10-06]. <https://doi.org/10.19678/j.issn.1000-3428.0061502>.

[17] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer, 2018: 3-19.

[18] LIU S T, HUANG D, WANG Y H. Learning spatial fusion for single-shot object detection [EB/OL]. (2019-11-25) [2021-10-05]. <https://arxiv.org/pdf/1911.09516/pdf>.

[19] TAN M X, LE Q V. Efficientnet: rethinking model scaling for convolutional neural networks [C]//Proceedings of the International Conference on Machine Learning. Long Beach: PMLR, 2019: 6105-6114.