

基于深度学习的滤光片检测与精准定位研究

张安^{a,b},伍世虔^b,黄建龙^b

(武汉科技大学 a. 机械自动化学院; b. 机器人与智能系统研究院,湖北 武汉 430081)

摘要:针对滤光片产品定位检测准确率低、实时性差的问题,提出改进的 YOLOv4 检测算法,并提出强定位非极大值抑制方法,改善 YOLOv4 算法中传统非极大值抑制法的预测框处理过程,解决预测框定位偏差的问题;设计一种静态模型滤波器融合方法,对静态 YOLOv4 主干特征提取网络中的卷积、池化等耗时运算进行线性融合,解决网络检测实时性差的问题。实验结果表明:改进后的 YOLOv4 定位准确率达到 99.41%,实时性比原始 YOLOv4 提高 14 帧/s。

关键词:深度学习;YOLOv4;精准定位;非极大值抑制;模型加速

中图分类号:TP183 **文献标志码:**A **文章编号:**1671-5276(2023)04-0012-04

Research on Filter Detection and Accurate Positioning Based on Deep Learning

ZHANG An^{a,b}, WU Shiqian^b, HUANG JianLong^b

(a. School of Mechanical Automation; b. Institute of Robotics and Intelligent Systems, Wuhan University of Science and Technology, Wuhan 430081, China)

Abstract: An improved YOLOv4 detection algorithm is proposed to deal with the low accuracy and poor real-time performance of filterproduct positioning and detection. A strong positioning non-maximum suppression method is initiated to improve the processing process of the prediction frame of the traditional non-maximum suppression method in the YOLOv4 algorithm and resolve the positioning deviation of the prediction frame. The design of static model filter fusion method is followed to linearly fuse the time-consuming operations like convolution and pooling in the static YOLOv4 backbone feature extraction network, thus solving the problem of poor real-time network detection. The experimental results show that the location accuracy of the improved YOLOv4 is 99.41%, and the real-time performance is 14 Frame/s, higher than that of the original YOLOv4.

Keywords: deep learning; YOLOv4; accurate positioning; non-maximum suppression; model acceleration

0 引言

智能工厂概念的出现,使制造业中自动化定位检测技术得到发展。滤光片产品在被生产出来后,需要对产品表面进行生产缺陷的检测,检测结束后要对滤光片利用工业机器人进行加工^[1]和码垛,这一系列的过程涉及到滤光片的定位工作,因此研究出一种定位准确率高、实时性好的算法对于滤光片生产很有意义。传统的算法会根据图像的具体特征进行目标检测与定位^[2-4],这些传统算法局限性明显,受光照强弱环境影响,鲁棒性无法满足实际的工程化要求。近些年深度学习的目标定位检测领域出现了一系列优秀的检测网络,主要分为两类,第一类是一阶检测器,比如 SSD^[5]、YOLOv4^[6]等,该类检测是通过在特征图网格上面直接回归生成出先验框,再通过非极大值抑制法筛选出最终的预测框;第二类是二阶检测器 Faster R-CNN^[7]、Mask R-CNN^[8]等,这类检测有专门 RPN 网络(region proposal network)来生成预选框,所以精度会比一阶检测高,但是 RPN 的处理过程繁琐,导致检测时间会比一阶慢很多。鉴于实际工程落地有高实

时性的要求,倪奕棋等^[9]在 SSD 网络里加入特征融合模块,使得在纱筒数据集上取得了良好的定位效果。LI M H 等^[10]专门设计一个目标框的检测回归网络,使得定位的结果令人满意。傅鹏等^[11]将卷积神经网络级联布局,以分层提取目标区域特征,逐步缩小搜索区域的方法,实现车牌的精准定位。但是以上研究者都是聚焦于神经网络结构的研究,而预测框的后处理策略也是值得研究的问题。

作为一阶目标检测器 YOLOv4 定位精度、实时性无法满足实际工程的高要求。所以深入分析滤光片定位准确率不高的问题,在传统非极大值抑制方法(non-maximum suppression, NMS)上提出改进,本文所提出的强定位 NMS 方法能够充分发挥回归出的不同尺寸预测框的作用,忽略传统 NMS 利用的分类置信度指标,以预测框坐标为衡量指标计算出合适的最终输出框,进而解决预测框定位不准的问题。针对算法实时性不佳的问题,提出一种静态模型滤波器融合方法,该方法可以在网络完成训练后,对静态模型中的卷积、池化等耗时运算进行线性融合,实现模型压缩加速运算的效果,进而解决算法实时性不佳的问题。

基金项目:国家自然科学基金项目(61775172);湖北省技术创新专项重大项目(ZDCX2019000025)

第一作者简介:张安(1997—),男,湖北黄冈人,硕士研究生,研究方向为机器视觉和智能制造,17671340939@163.com。

1 滤光片检测与定位方法

1.1 YOLOv4 网络结构

本文以 YOLOv4 作为算法框架。YOLOv4 主要包含 3 个网络部分,第一部分是主干特征提取网络,本文借鉴了 Inception^[12]的思想,将一个多级感受野的模块作为 Backbone 的基本组成单元,由 1×1 卷积、3×3 卷积、平均池化和残差连接组成,如图 1 所示。第二部分是空间金字塔池化(spatial pyramid pooling)结构,该结构是由 5×5,9×9,13×13 3 个不同尺寸的最大池化组成,实现特征图的下采样,可以减少参数计算量,提取主要特征。第三部分是加强特征提取结构(path augmentation network),该结构是一种特征金字塔结构,它融合了 Backbone 提取的深语义信息和浅语义信息,通过反卷积上采样将小尺寸的特征图恢复到跟大尺寸特征图一样的尺寸,这相当于将不同语义信息的特征进行了融合,最后获得 13×13、26×26 以及 52×52 尺寸的检测特征图,如图 2 所示。

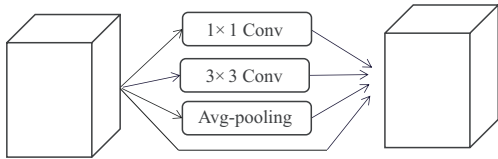


图 1 Inception 基础单元

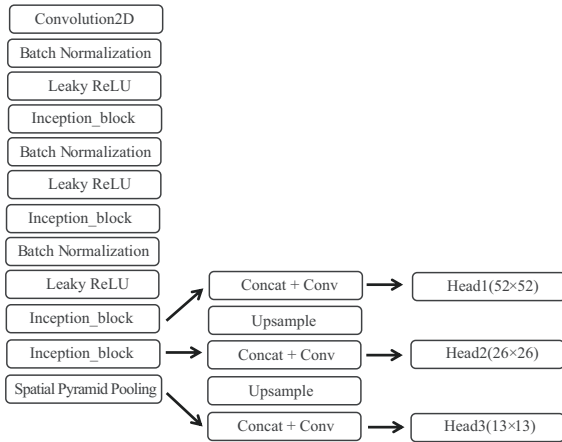


图 2 YOLOv4 的网络结构

1.2 非极大值抑制的改进

1) 传统非极大值抑制问题分析

非极大值抑制就是抑制不是极大值的元素,NMS 属于 YOLOv4 最后的预测框后处理阶段,在 YOLOv4 里 NMS 的计算流程是先利用所有预测框的置信度由大到小的排序,选择置信度最大的框作为基准框,用基准框跟其他框去计算交并比(intersection over union, IOU)。传统 NMS 是设定阈值 IOU 为 0.5,如果预测框与基准框计算所得的 IOU 大于 0.5,就将该框删除,经过反复迭代,最终就保留最后一个框作为预测框。针对滤光片产品定位的分析,传

统 NMS 是基于分类得分来排序预测框,只有置信度最高的框才能被保留,但预测的置信度高低与定位准确率的高低不是强相关,如图 3 所示。置信度高的预测框,定位的准确率反而不佳,置信度低的框,定位反而满足要求,这种现象就是传统 NMS 是以分类置信度来筛选预测框导致的,所以传统 NMS 不适用于滤光片定位。

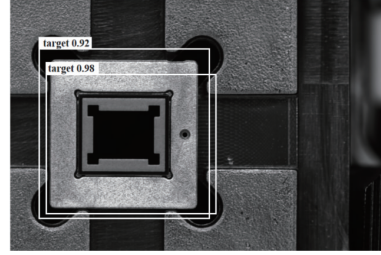


图 3 传统 NMS 中定位准确性与分类置信度不是强相关

2) 强定位非极大值抑制的提出

针对传统 NMS 在滤光片定位任务不适用的问题,本文弱化置信度的衡量指标,提出强定位 NMS,以多个预测框的坐标值为驱动力。YOLOv4 有 3 个尺寸的特征图输出,分别检测小目标、中目标和大目标,每个特征图中的每个特征点会有 3 批检测小、中和大目标不同尺寸的预测框输出,且每批 3 个预测框,如图 4 所示。9 个预测框会等比例映射到原图上,再通过强定位 NMS 去处理 9 个框。首先利用所有有效的输出预测框计算各个框中心点间的欧式距离,两两预测框中心点之间的欧式距离公式为

$$d_{ist,i} = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \quad (1)$$

利用所有中心点欧式距离均值作为阈值筛选掉那些有极端欧式距离的候选框,利用剩余的框计算出均值坐标矩阵作为最终的预测框,以确保能够充分利用所有预测框的坐标,进而得到能够完全包含滤光片产品的最终预测框。

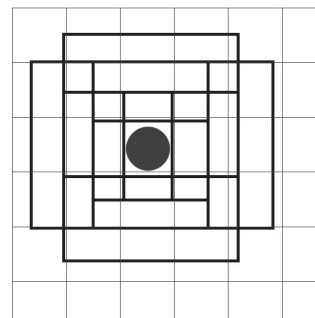


图 4 特征点回归出 9 个不同尺寸的预测框

1.3 模型压缩加速运算的策略

模型需要进行前向传播和反向传播来更新滤波器的参数,使得模型达到收敛的条件,训练阶段模型的参数在动态变化,此时的模型为动态模型。神经网络在完成训练后,滤波器内的参数就会停止更新,此时的模型为静态模型。卷积可加性的融合等式如式(2)所示。

$$M \times K1 + M \times K2 = M \times (K1 + K2) \quad (2)$$

式中: M 为输入特征矩阵; $K1$ 、 $K2$ 是卷积核矩阵。这个等式表达的意思就是对于输入特征图 M ,先进行 $K1$ 卷积、 $K2$ 卷积后再对结果进行相加,与先进行 $K1$ 和 $K2$ 的逐点相加后再进行卷积得到的结果是一致的,而且参数量也不变,这样卷积融合的好处是在参数量不变和运算结果不变的情况下,减少了运算次数,实现模型容量压缩的效果,以提升算法的推理速度。

将Inception模块进行静态模型融合,对于 1×1 的Convolution,将静态滤波器周围填充0,让其成为 3×3 的Convolution;对于平均池化运算,为让其变为 3×3 的Convolution,将滤波器内的权重设为 $1/9$,进而得到 3×3 的Convolution;对于残差连接,相当于将滤波器中点位置的值为1,周围的值为0,即可将残差连接转为 3×3 的Convolution,最后将3个变换得到的 3×3 的Convolution与Inception原有的 3×3 的Convolution进行加法运算,就得到1个合并后的 3×3 的Convolution。这种静态滤波器融合运算,可以将网络进行预测阶段的4次推理卷积运算变为一次卷积运算,这样可以大大缩短网络的推理时间,使得模型的实时性更加良好,融合过程如图5所示。

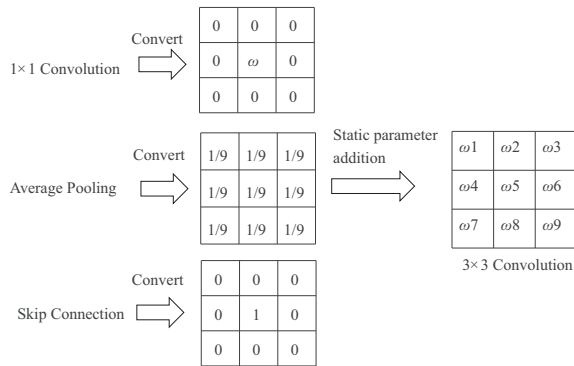


图5 静态滤波器融合过程

2 实验结果分析

2.1 实验环境和数据集

本实验的运行硬件平台是Windows系统,GPU型号为Nvidia GeForce RTX 2060独立显卡,其显存为6GB;CPU型号为英特尔Core i5-10200H @ 2.40 GHz 4核,其内存为16GB。软件运行平台为Anaconda3的Python3.6的Pycharm解释器和Pytorch深度学习框架。本文数据集选用滤光片产品,该图像是在环形光源照射下,利用CCD工业相机拍摄所得。

2.2 实验结果

a) 评价指标

1) 交并比(Iou):这项指标是衡量预测框与真实框重合程度,预测框为 A ,真实框为 B ,如式(3)所示。

$$I_{ou} = \frac{A \cap B}{|A| + |B|} \quad (3)$$

2) 准确率:预测框完全能够包含目标的为定位正确,

目标没有完全被包含在预测框内,则为定位错误。准确率为定位正确的图像数量除以测试数据集的总数量。

b) 强定位 NMS 最终框计算方式的实验

对于改进的强定位 NMS,在进行欧式距离阈值筛选出最佳预测框后,为了充分利用筛选出的高质量框,本文对剩余的框利用最大外接矩阵、预测框的中值计算、预测框的均值计算做了对比实验。通过实验得到,使用预测框均值作为最终输出框,能够获取最大平均交并比,说明均值能够充分利用最佳的几个预测框,最终获得最佳的最终预测框,实验结果如表1所示。

表1 候选框中值、均值、最大值实验结果

网络结构	最终框计算方式	IOU 值
YOLOv4	最大值	0.706
YOLOv4	中值	0.915
YOLOv4	均值	0.916

c) 传统 NMS 与强定位 NMS 的实验

利用传统的非极大值抑制方法,预测框有大量定位不准的现象,是因为传统非极大值抑制利用的是置信度作为重要的评判标准,定位准但置信度不高的预测框被筛选出去,定位不准置信度高的框反而得到保留。对于改进的强定位非极大值抑制,能够以所有预测框的定位坐标为评判依据。实验结果表明,强定位非极大值抑制有提升效果,如表2所示。

表2 强定位 NMS 与传统 NMS 实验结果

网络结构	NMS 类型	IOU 值
YOLOv4	传统 NMS	0.878
YOLOv4	强定位 NMS	0.916

d) 传统 NMS 与强定位 NMS 效果图

通过框与框中心点的欧式距离和剩余框的均值作为评判条件,充分发挥了所有预测框的坐标,进而使得改进的强定位 NMS 的预测框能够完全包含滤光片,而传统 NMS 的部分预测框有偏移的现象,检测定位结果如图6所示。

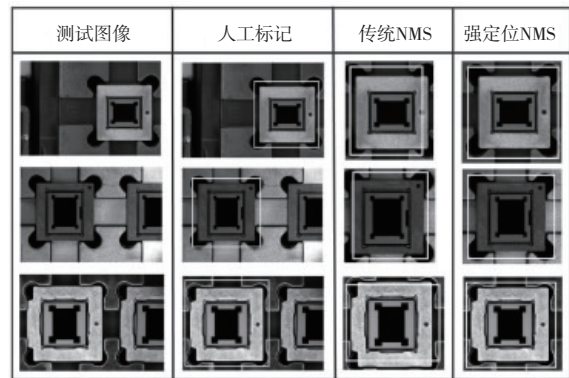


图6 人工标记图、传统 NMS 定位局部图和强定位 NMS 局部图

e) 静态模型融合实验结果

通过实验得到,静态滤波器在进行融合前后的预测平

均交并比是完全一样的,说明融合后的滤波器不会产生定位精度损耗,验证了静态滤波器的可加性原理。与此同时,静态融合后的模型大小也略有减小,且算法的FPS (frames per second)得到了明显的提升,测试的FPS为58帧/s,比未融合前FPS提升14帧/s。实时性满足工业检测的要求,这是一种低成本,提高实时性的有效方法,如表3所示。

表3 静态融合实验结果

网络结构	NMS类型	模型	IOU值	FPS值/ (帧/s)	模型大小/MB
YOLOv4	强定位	无融合	0.916	44	68.9
YOLOv4	强定位	有融合	0.916	58	66.6

f) 本文方法与其他方法测试对比结果

利用1378张图像模拟实际工程的量产测试,从定位准确率、FPS的角度去作对比,二阶算法Faster R-CNN相比于一阶算法,定位的准确率高些,但是在实时性上,Faster R-CNN的FPS无法满足实际工程的要求。对于SSD一阶检测器准确率虽然能够达到95.13%,但是它的FPS没有达到50帧/s,FPS为50帧/s以上时,才能满足实际工程的实时性要求,YOLOv4在没有对传统NMS改进之前,其定位的准确率不佳,实时性也无法达到50帧/s,在对传统NMS进行改进后,强定位NMS发挥出良好的定位结果,定位准确率达到99.41%,同时利用静态模型融合后,FPS值达到了58帧/s,满足实际工程的高实时性的要求,如表4所示。

表4 与其他算法模拟量产测试

算法模型	总数量	正确数量	正确率/%	FPS值/(帧/s)
Faster R-CNN	1378	1353	98.18	8
SSD	1378	1311	95.13	37
YOLOv3	1378	1093	79.31	34
YOLOv4	1378	842	61.10	44
YOLOv4+融合+ 强定位NMS	1378	1370	99.41	58

3 结语

针对滤光片检测定位不佳问题,本文对YOLOv4检测算法提出改进工作。主要对YOLOv4做了两点改进:1)本文针对YOLOv4算法中预选框后处理的NMS方法提出改进工作,提出强定位NMS方法,该方法利用欧式距离去筛选掉分布极端的预测框,然后充分利用所有预测框的坐标信息代替导致定位偏移的分类置信度评判标准去计算出最终的预测框坐标。实验证明,改进的NMS能够让

YOLOv4算法预测框定位不准的问题得到改善;2)本文提出的静态模型滤波器融合法对YOLOv4的主干特征提取网络进行融合,使算法的实时性达到了58帧/s,进而使得算法的实时性达到了实际工程的要求。因此,本文提出的算法为智能制造中自动化的检测技术提供了较大的参考价值。

但是,本文所提出的压缩方法只能对YOLOv4的主干特征提取的部分结构进行压缩,暂时不能将YOLOv4的所有网络结构进行压缩,下一步研究工作是在保证精度不受影响的前提下,尝试探索对YOLOv4的所有网络结构进行线性融合,以进一步加快算法的检测速度。

参考文献:

- [1] 王昊,张得礼,王宏涛. 基于遗传算法的机器人加工工位姿优化研究[J]. 机械制造与自动化,2021,50(6):174-177.
- [2] 张震,张照崎,朱留存,等. 基于Harris-改进LBP的特征匹配及目标定位算法[J]. 吉林大学学报(理学版),2021,59(3):568-576.
- [3] 张毅,张瀚,韩晓园. 融合MeanShift和改进SURF算法的目标定位策略[J]. 控制工程,2020,27(4):629-634.
- [4] LOU H C, LU M, CUI H H, et al. High-precision location for occluded reference hole based on robust extraction algorithm[J]. Measurement Science and Technology,2021,32(3):035021.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [M]//Computer Vision - ECCV 2016. Cham:Springer International Publishing,2016:21-37.
- [6] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [C]//Computer Vision and Pattern Recognition, Online: IEEE,2020.
- [7] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,39(6):1137-1149.
- [8] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision. Venice, Italy. IEEE,2017:2980-2988.
- [9] 倪奕棋,管声启,管宇灿,等. 基于改进的SSD深度学习算法的双目视觉纱筒识别定位[J]. 纺织高校基础科学学报,2021,34(2):59-66.
- [10] LI M H, HU T J. Deep learning enabled localization for UAV autoland[J]. Chinese Journal of Aeronautics,2021,34(5):585-600.
- [11] 傅鹏,谢世朋. 基于级联卷积神经网络的车牌定位[J]. 计算机技术与发展,2018,28(1):134-137.
- [12] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE,2016:2818-2826.

收稿日期:2022-02-14