DOI: 10.13382/j.jemi.B2306184

改进 RetinaNet 的工艺流程检测算法*

李 玮 高 林
(青岛科技大学自动化与电子工程学院 青岛 266061)

摘 要:现阶段,图像深度学习算法无法检测时序性的工艺流程问题。本文针对针织机械山板总成的人为装配工艺进行研究,提出 MS-RetinaNet 目标检测算法。借鉴自然语言处理的思想,引入 Swin-Transformer 结构,保留了 CNN 结构的层次性,弥补了 CNN 结构对于高层语义信息融合不足的问题,增强了全局与细节学习能力;使用改进的 GIoU Loss,增加判定因子式,缓解损失计算退化的影响,优化边界框回归效果;根据多尺度目标参数,采用最佳锚框比,提高了召回率和检测精度;设计时序检测头,使算法具有判别目标先后顺序和逻辑关系的能力。实验结果表明,算法 AP 可达 90.3%,高于当前主流算法 2%以上,单张图片检测速度约 46 ms,满足了工艺流程的时序检测要求,综合性能优越。

关键词: 工艺流程;MS-RetinaNet;Swin-Transformer;判定因子式;检测头
中图分类号: TP391.4;TN05 **文献标识码**: A **国家标准学科分类代码**: 520.6040

Improved RetinaNet process flow detection algorithm

Li Wei Gao Lin
(School of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: At this stage, the image deep learning algorithm cannot detect the chronological process problem. In this paper, the artificial assembly process of the mountain board assembly of knitting machinery is studied, and the MS-RetinaNet object detection algorithm is proposed. Using the idea of natural language processing for reference, the Swing-Transformer structure is introduced to retain the hierarchy of CNN structure, make up for the lack of high-level semantic information fusion in CNN structure, and enhance the ability to learn overall and details. The improved GIoU Loss is used to increase the judgment factor formula, mitigate the impact of loss calculation degradation, and optimize the regression effect of the bounding box. According to the multi-scale target parameters, the best anchor frame ratio is adopted to improve the recall rate and detection accuracy. The chronological detector is designed to enable the algorithm to distinguish the sequence and logical relationship of the target. The experimental results show that the algorithm AP can reach 90.3%, which is more than 2% higher than the current mainstream algorithm. The detection speed of a single image is about 46 ms, meeting the chronological detection requirements of the process flow, and the overall performance is superior.

Keywords: technological process; MS-RetinaNet; Swin-Transformer; judgment factor formula; detector head

0 引 言

近年来,随着新一代信息技术的推广应用,我国大批针织[1]机械公司也逐步开始进行智能化改造,将人工智能[2]技术应用于生产过程中。目前大规模量化生产智能化应用较多,但针对生产工艺的应用研究较少,工艺过程多数还依赖人工操作。人工操作易出现操作不规范、错误操作等问题,这种问题在针织机械山板总成系统的装配工艺上尤为明显。由于核心系统皆为进口,我国尚无明确的标准工艺规范,装配方式优劣存在争议,易出现工艺流程混乱和操作失误等问题,导致生产效率低、增加生产成本。当前检测手段局限于人工经验,智能化水平低下,因此,山板总成的装配工艺检测成为针织机械领域的

收稿日期: 2023-01-18 Received Date: 2023-01-18
*基金项目:国家自然科学基金(61971253)、山东省自然科学基金(ZR2021MFZ23)项目资助

一个研究重点。随着计算机视觉技术[3]的发展,基于深度学习[4]的目标检测技术在工业生产线上的应用越来越成熟。当前主流图像目标检测算法分为以 YOLO[5]系列为代表的一阶段算法和 RCNN[6]系列为代表的二阶段算法。针对图像算法存在的静态、无时序参数等问题,一些学者利用视频目标检测算法进行时序性问题的研究。参考文献[7]提出了一种改进的区域 3 D 卷积神经网络,使用时域反卷积网络增加特征图长度,提高时域上行为的定位精度;参考文献[8]提出基于金字塔结构的无锚框时序行为检测方法,设计了嵌入自注意力模块,建模多尺度类激活热力图,有效提升了检测效果。Transformer[9]的应用进一步推动了时序行为检测的发展,参考文献[10]提出了一个新颖的双分支检测框架,双分支协作机制利用行为类别标签和行为帧之间的互补信息,获得更精确的检测结果;参考文献[11]提出了一种基于端到端 Transformer 的 TAD 方法,加入时间上可变形的注意力模块,实现了低计算、高性能。

上述研究针对视频时序行为检测算法提出了一些改进策略,应用效果优异,但时序性问题的研究未涉及图像检测领域。本文针对山板总成的装配工艺流程问题,采用基于深度学习的图像目标检测技术对其进行研究。算法框架使用 RetinaNet[12]结构,骨干网络采用 Swin-Transformer,缓解卷积操作的局限,增加算法的全局泛化能力;使用滑动窗口模式,减少 Transformer 结构中注意力机制的计算量;改进了 GIoU Loss[13]策略,以判定因子式弥补损失退化的不足,优化模型训练方向;使用最佳锚框比,增加特殊尺度目标的关注度,提升样本召回效果;创新设计了时序检测头,在不增加算法冗余的同时,使算法具备良好的时序检测功效。此方法通过图像流数据实现对山板总成装配工艺流程的精准检测,有效降低了误操作率,为时序工艺检测提供了新思路。

1　MS-RetinaNet 网络结构

1.1　整体网络与检测头结构

RetinaNet 是一阶段目标检测算法,整体框架包含 ResNet-FPN[14]结构和 2 个全连接网络,因其提出了 Focal Loss[15]分类损失函数,使其精度可以高于骨干网络为 VGG-16[16]的 Faster RCNN 算法。

改进的网络以 Swin-Transformer 作为特征提取,融合 FPN 结构,在模型测试端的全连接分类和回归网络后,增加具有时序功效的检测头。使模型在目标检测的基础上,对各类别目标的出现顺序和逻辑关系进行判定,实现目标时序检测。如图 1 所示,检测头依次设置两层结构和内置编码。首先是类别层,在此层进行类别的处理,对算法模型训练出的各类别进行排序和选择,使检测头能按照预设的时序将选定类别输出至下一层;第 2 层是框选分数层,对输入类别的预测框进行阈值判别,若预测框概率分数符合阈值判别标准,则视为通过,下一次检测,类别层按预设选定下一个类别输出至框选分数层;若预测框概率分数未达到阈值判别标准,后续检测,类别层继续输出当前类别,框选分数层继续执行其阈值判别,直到符合阈值标准,类别层才执行下一个类别的输出。同理,直到第 n 个类别输出完成,返回至预设的第 1 个类别,以此循环检测判别,形成时序性。其中阈值设定需要符合实际应用场景,为了保证模型的实际检测精确度,本文设定检测头阈值为 0.9。

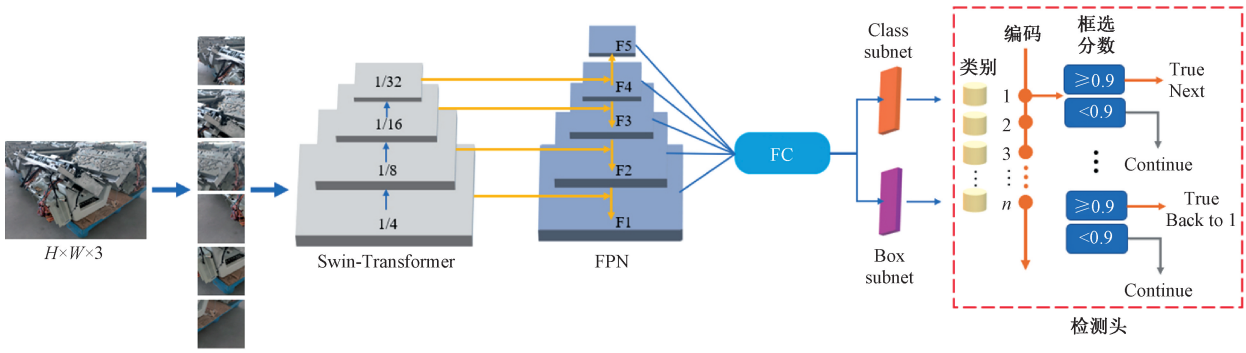


图 1　网络整体结构

Fig. 1　Overall network structure

逻辑关系的检测采用内置编码设定,在类别层固定的时序中插入编码,使下一轮检测一个或多个类别满足某种逻辑关系,才能继续执行。假设在第 3 个类别检测后,设置编码第 2、3 类别需要同时存在且满足框选分数层的要求,若检测符合标准,则下一次检测执行第 4 类别的判别;反之将继续执行此逻辑检测,直至执行通过。

1.2　Swin-Transformer 结构

Swin-Transformer 网络整体结构具有层次性,自底向

上的下采样[17]倍数成2倍增加,能够有效提取出具有层次性的特征图。相对于纯卷积结构,更加聚焦于图像中的感受野,应用于多目标动态特征的目标检测效果优异。首先,图像送入Patch Partition[18]层,这一步相当于卷积的下采样操作。将图像划分出相同大小的块,每个块由4×4 pixels组成,嵌入向量后,在通道方向上进行展开平铺。一般采用的是RGB三通道图像,且每个块有16个像素,所以展开平铺后通道48。故通过Patch Partition后图像宽高变为原来的1/4,深度变为原来的16倍,如图2所示。

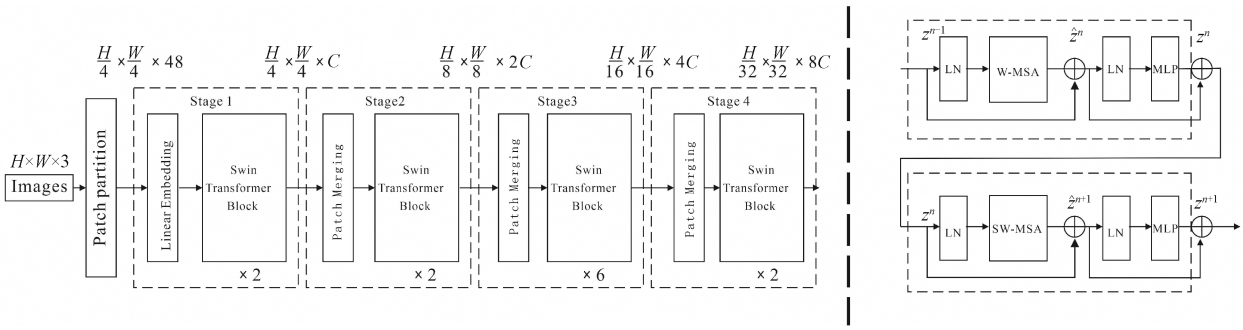


图2　Swin-Transformer结构

Fig. 2　Swin-Transformer structure

接着经过4个stage,stage1中的Linear Embedding层对送入图像的每个像素通道做深度变换,深度由48变为C,再送入Swin Transformer Block;剩余的3个stage都是重复经过Patch Merging层下采样和Swin Transformer Block计算,类似于ResNet-FPN的操作;但是不同点在于ResNet-FPN采用卷积和池化进行倍数下采样,而经过Patch Merging层操作,就会将特征图分为多个2×2的块,把每个块相同位置的像素特征取出并组合在一起,在图像深度方向进行拼接,这样就不存在量化误差。此时特征图的高和宽就会减半,深度就会变为原来的4倍。最后经过一个全连接层,特征图的高和宽不变,深度翻倍。每一个stage设置的Swin Transformer Block都是偶数次,内部先经过一个窗口多头自注意力结构(windows multi-head self-attention[19],W-MSA),再经过一个滑动窗口多头自注意力结构(shifted windows multi-head self-attention,SW-MSA)。最后通过MLP[20]层,实际相当于一个全连接层和激活函数,进行分类操作。

1.3　窗口多头自注意力结构

W-MSA是在多头自注意力机制(multi-head self-attention,MSA)的基础上做了改进。如图3所示,假设第n层的W-MSA将特征图划分为相同大小的4个窗口,每个窗口为4×4,分别对等分的窗口区域内做MSA,增加区域特征信息的精确性,且每个窗口内的计算互不影响。在第n+1层使用SW-MSA,将W-MSA特征图进行特定像素的平移,可以等效为向右下方沿对角线平移2个像素块,使第n层W-MSA划分的相邻窗口区域之间保持一定的信息交互,得到9个大小不一的窗口。

再将新划分的9个窗口进行平移,如图4所示,将A、B和C平移到底部,接着将D、G和A平移到最右侧,

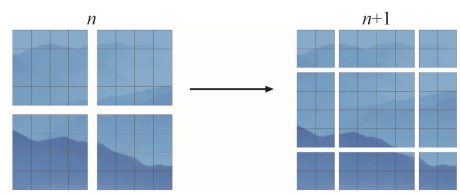


图3　W-MSA和SW-MSA

Fig. 3　W-MSA and SW-MSA

E作为一个窗口,F和D组成一个窗口,H和B组成一个窗口,I、G、C、A组成一个窗口,这样便拼接成和第n层W-MSA中4个大小相同的窗口,然后分别在4个拼接的窗口内进行MSA计算,目的是为了避免9个窗口计算额外增加了训练的复杂度。

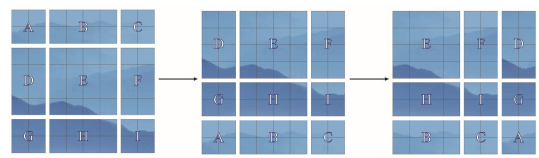


图4　SW-MSA原理

Fig. 4　SW-MSA principle

同时,由于平移后的窗口图像特征乱序,为了防止拼接窗口在计算时,其不同小窗口发生信息错乱,采用mask MSA的方法;如图5所示,对区域F做MSA,区域F内的所有像素与区域D内的像素依次做$\boldsymbol{q}$、$\boldsymbol{k}$计算,$\boldsymbol{q}$是索引图像特征点的查询向量,$\boldsymbol{k}$是指示图像特征重要程度的键向量;计算结果减去100,得到一个相对小的值,经过Softmax后权重变为0,这样进行MSA操作时,非当前计算窗口D的像素值权重都为0,等效于只对当前计

算 F 的每个像素进行 MSA。

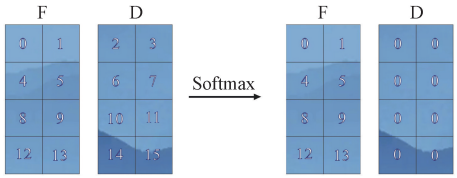


图 5　SW-MSA 计算

Fig. 5　SW-MSA calculation

所有窗口 MSA 计算完成后，需要将窗口平移回移动前的位置，包括计算得到的数据结果，这样的计算方式不存在量化误差，比卷积更有优势。最后，通过一个 Encoder 全连接层结构，找到需要的编码，将该编码对应的特征向量取出，经过 MLP 结构生成最终结果。

1.4　改进的 GIoU Loss

在原始 RetinaNet 算法中，采用 Smooth L1 Loss 作为边界框回归损失，但它与实际检测结果的好坏关联度不够高，存在 Smooth L1 Loss 值相同而实际预测框和真实框交并不同的问题。IoU Loss 能有效解决此问题，但其缺点较多，对此，引入 GIoU Loss，如式(1)所示：

$$L_{GIoU} = 1 - IoU + \frac{|C \backslash (A \cup B)|}{|C|} \tag{1}$$

式中：A 为真实框面积；B 为预测框面积；C 为包含 A 和 B 的最小矩形框面积；IoU 是 A 和 B 的交并比；GIoU Loss 先计算 C 中除 $A \cup B$ 以外的面积和 C 的比值，再加上 1 - IoU，这样更关注非重合区域。当 A 和 B 没有交集时，IoU 值为 0，但 GIoU Loss 仍可优化训练；当 A 和 B 的 IoU 相同，两个框重合方向或角度不同时，C 的值不同，GIoU Loss 也不同，有一定的区分度。但当预测框与真实框出现水平重合、竖直重合、包含和被包含等情况时，如图 6 所示，实线框分别代表 A 和 B，C 为虚线框，此时 C 与 $A \cup B$ 值相同，导致 GIoU Loss 退化为 IoU Loss。

针对 GIoU Loss 存在的问题，本文提出改进策略，在损失函数中增加判定因子式，如式(2)、(3)、(4)所示：

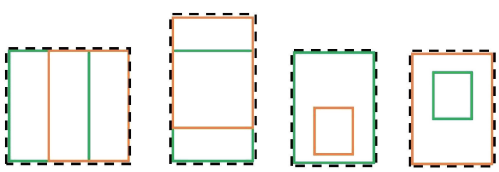


图 6　GIoU Loss 失效的情况

Fig. 6　Failure of GIoU Loss

$$S_i = |C - (A \cup B)| \tag{2}$$

$$\alpha = \begin{cases} 1, S_i = 0 \\ 0, S_i \neq 0 \end{cases} \tag{3}$$

$$L_p = \alpha \left| \frac{(A \cup B) - (A \cap B)}{A \cup B} \right| \tag{4}$$

式中：S_i 表示 C 和 $A \cup B$ 的面积差值；α 是判定因子；Lp 是判定因子式，表示 A 和 B 并集中非交集区域与 A 和 B 并集的比值。改进后的 GIoU Loss 函数如式(5)所示：

$$L_{GIoU} = 1 - IoU + \frac{|C \backslash (A \cup B)|}{|C|} + \alpha \left| \frac{(A \cup B) - (A \cap B)}{A \cup B} \right| \tag{5}$$

当 S_i 不为 0 时，α 始终为 0，则 L_p 为 0，不影响原始 GIoU Loss 的计算；当出现图 6 问题时，即 S_i 为 0，α 为 1，L_p 不为 0，防止 GIoU Loss 退化为 IoU Loss，使模型更关注非交集区域，反应真实框与预测框重合度的效果更好，提升边界框的回归效果，对模型训练优化的走向起着至关重要的作用。

2　实验结果与分析

2.1　实验条件与数据集

本实验数据由海康彩色工业相机采集，配备 400×10^4 pixels 的摄像头，相机可设置连续抓拍；相机连接服务器，服务器基于 Ubuntu 系统，搭载 GeForce RTX 2080Ti 显卡，48 G 显存；采用 Python 编程语言，算法环境由 Pytorch 深度学习框架构建。实验测试运行时，相机连续抓拍设置为 1 s/次；由于彩色工业相机抓拍图片的质量较大，为了保证时效性和传输稳定性，采用本地服务器多点多线程传输，连续抓拍的图片保存在预设的本地磁盘中，通过服务器自动检测并给出工艺流程判别结果。

实验采用 coco 数据集格式，样式如图 7 所示，图片总数为 7 460，除去背景分为 8 类，其中各类别标注框的数量分别为三角电机(Stepper motor1，S-m1)4244、推针电机(Stepper motor2，S-m2)4239、度目电机(Stepper motor3，S-m3)4378、调试螺丝刀(Commissioning screwdriver，C-s)844、千分表(dial indicator，D-i)1974、润滑铁片(lubricating iron sheet，L-i-s)848、选针器(needle selector，N-s)4024、螺纹锁固剂(thread locking agent，T-l-a)1153，总计 21 704 个样本标注框。

2.2　改进的锚框比

由于 RetinaNet 属于 Anchor-based 算法，是根据初始的锚框参数来训练模型。RetinaNet 算法初始的锚框长宽比率为 0.5、1 和 2，对应 3 种固定的尺度和锚框基础边长。这样做只能满足大部分样本，而小部分样本可能会受影响，导致召回率变差，回归效果受损，影响检测精度。本文提出改进固定锚框比的方法，如图 8 所示，数据集中小部分样本采用初始锚框比无法适用，需要增加合适的锚框比；根据锚框长宽比选取原则，不能过于极端，且比

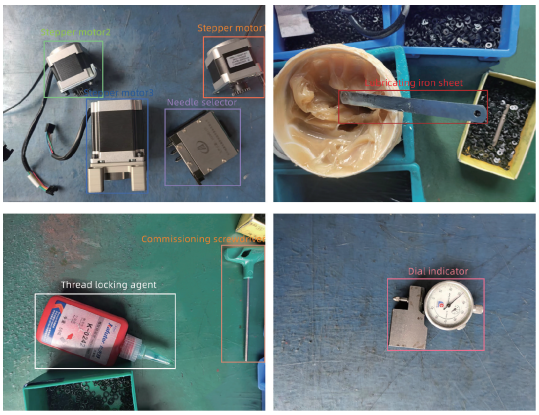


图7　数据集样本

Fig. 7　Data set sample

率乘积基本保持为1;故除原有比率以外,选择并添加比率5和0.2,改进固有的比率,优化模型的训练方向,提升对小部分样本的召回和边界框回归效果;而样本真实框基础边长与原有锚框基本无差,故不变更。

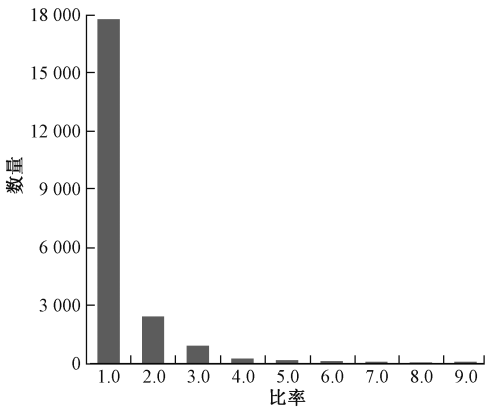


图8　各锚框比数量

Fig. 8　Ratio and quantity of each anchor box

2.3　山板总成的工艺检测流程

研究对象山板总成如图9所示,主要由山板、三角电机、推针电机、度目电机和选针器组成。

装配工艺大体分为6个关键点步骤,其中省略重复步骤:

1)装配2个三角电机,先进行润滑操作,再安装至山板上。

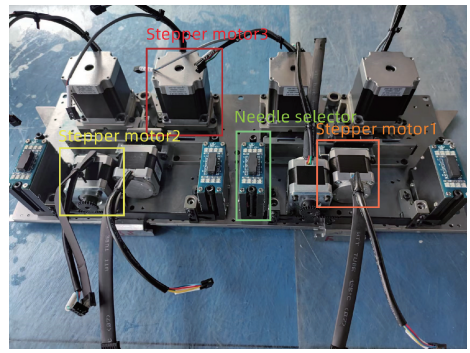


图9　山板总成

Fig. 9　Mountain plate assembly

2)用调试螺丝刀检查三角电机组装的齿轮连杆是否滑动顺畅,必须要调节一下吻合度。

3)装配2个推针电机,先进行润滑和螺纹锁固操作,再安装至山板上。

4)装配4个度目电机。

5)装配4个选针器。

6)检查山板总成的中山高度差是否合适。

每一步装配工艺必须严格按照规范工艺流程来执行,否则可能导致山板总成的使用出现问题。实际视觉算法检测设计时,根据检测操作器件及其逻辑关系来确认当前工艺流程是否完成;配合改进的时序检测头,设置6个编码;1)插入编码1,检测S-m1和L-i-s同时存在,再插入编码2,检测2个S-m1并存;2)检测是否存在C-s;3)插入编码3,检测S-m2和L-i-s同时存在,接着检测T-l-a,再插入编码4,检测2个S-m2并存;4)插入编码5,检测4个S-m3并存;5)插入编码6,检测4个N-s并存;6)检测是否存在D-i。

2.4　训练与测试结果

本文算法使用的Swin-Transformer有Tiny、Small、Base、Large结构,模型选择时需综合考虑网络参数带来的复杂度和精度平衡问题。以RetinaNet为模型框架,进行多版本Swin-Transformer的骨干网络参数消融实验,实验结果如表1所示,其中输入尺寸384×384的Swin_Large模型在实验软硬件基础上存在内存溢出,无法训练与测试。

表1　骨干网络消融实验

Table 1　Core network ablation experiment

算法框架	骨干网络	Input_size	Windows	C	AP	Params	GFLOPs	FPS
RetinaNet	Swin_Tiny	224×224	7×7	96	88.3	37.13	212.95	21.6
	Swin_Small	224×224	7×7	96	88.8	58.28	302.94	15.5
	Swin_Base	224×224	7×7	128	89.6	96.89	445.35	8.9
	Swin_Base	384×384	12×12	128	89.6	97.03	457.06	8.8
	Swin_Large	224×224	7×7	192	89.9	206.56	851.55	4.1
	Swin_Large	384×384	12×12	192				

由表 1 可以看出,随着骨干网络参数的增加,整体模型的 AP、参数量、GFLOPs 逐步增加,FPS 逐步减小,但 AP 的提升幅度远不及 FPS 降低幅度带来的影响。综合考虑模型部署后存在的检测精度与速度偏差问题,选择 Swin_Tiny 作为本文算法的骨干网络。

基于此,本文选择包括 RetinaNet 算法、small 版本的 YOLOX 算法(YOLOX_s,YOLOX)、ResNet-FPN 和 ROI Align 结构的 Faster RCNN 算法(Faster RCNN,F-RCNN)、使用 Smooth L1 Loss 的 Swin-Transformer-Tiny 结构 RetinaNet 算法(L1-Swin-RetinaNet,L1-S-t)、使用 GIoU Loss 和最佳锚框比的 Swin-Transformer-Tiny 结构 RetinaNet 算法(G-Swin-RetinaNet,G-S-t)、使用改进 GIoU Loss 和最佳锚框比的 Swin-Transformer-Tiny 结构 RetinaNet 算法(MS-RetinaNet,MS-R-t)进行对比实验,各模型边界框损失和分类损失曲线如图 10 所示,由于 YOLOX 算法训练的独特性,迭代次数与其他算法不一致,图示部分未达到收敛。

由图 10(a)可以看出,得益于 RPN 结构,Faster RCNN 训练损失收敛最快,YOLOX 收敛相对最慢,损失值最大;由于 GIoU Loss 计算特性,训练前期收敛较慢,本文 MS-RetinaNet 最终收敛的损失值要小于 G-Swin-RetinaNet,表明改进 GIoU Loss 的有效性。由图 10(b)看出,YOLOX 收敛速度和损失值表现最差;Faster RCNN 依旧收敛最快,分类损失值和剩余算法收敛后几乎相同;使用 Swin-Transformer 结构的 RetinaNet 算法收敛速度都要快于 RetinaNet 算法,表明 Swin-Transformer 作特征提取的优越性。

为了进一步验证本文算法性能的优越性,结合 L1-Swin-RetinaNet、G-Swin-RetinaNet、RetinaNet、YOLOX 和 Faster RCNN 算法,在测试集上对其最优的训练模型进行测试,以多种指标验证,结果如表 2 所示。由于增加检测头对模型检测速度在 ms 级上几乎无影响,FPS 可忽略检测头影响;本实验中无小目标样本,故表 2 无 AP_S 和 AR_S 的指标。

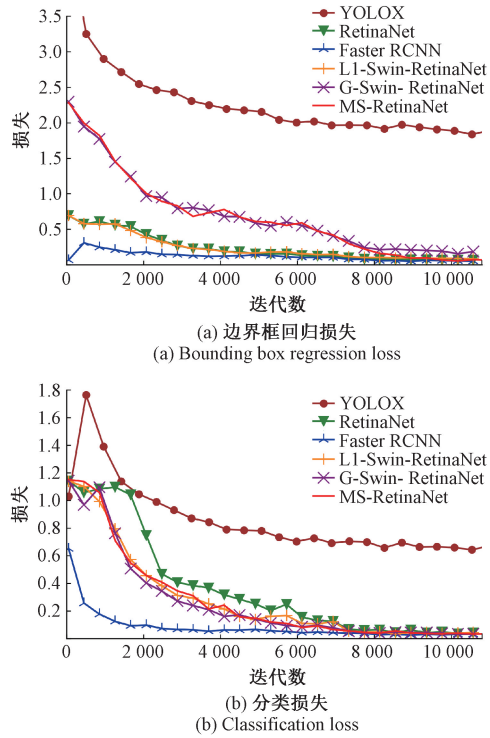


图 10　训练损失曲线

Fig. 10　Training loss curve

表 2　各算法模型性能对比

Table 2　Performance comparison of algorithm models

模型	AP/%	AP_{50}/%	AP_{75}/%	AP_S/%	AP_M/%	AP_L/%	AR/%	AR_S/%	AR_M/%	AR_L/%	FPS	GFLOPs
YOLOX	87.0	**98.2**	94.9		62.5	87.8	90.9		72.4	89.9	77.6	33.32
RetinaNet	83.8	94.3	91.2		58.3	85.1	88.4		62.4	89.5	22.2	207.46
Faster RCNN	87.8	97.8	**96.4**		70.5	88.7	90.6		70.5	91.4	16.8	206.70
L1-Swin-RetinaNet	88.3	97.0	94.6		69.9	89.2	91.0		69.8	91.8	21.6	212.95
G-Swin-RetinaNet	89.1	97.1	95.3		70.3	90.1	91.6		75.5	92.4	21.8	212.95
MS-RetinaNet	**90.3**	97.2	**96.4**		**78.0**	**91.3**	**92.7**		**79.9**	**93.5**	21.8	216.48

从表 2 可以看出,L1-Swin-RetinaNet 的 AP 和 AR 表现强于原始 RetinaNet,说明 Swin-Transformer 结构的特征提取更全面,对检测精度有一定的提升;G-Swin-RetinaNet 的指标相对优于 L1-Swin-RetinaNet,表明 GIoU Loss 搭配最佳锚框比使算法精度得到进一步优化;采用改进的 GIoU Loss 和最佳锚框比,提升了边界框回归效果,使 MS-RetinaNet 的检测精度最高,AP 达到了 90.3%,除 AP_{50} 略逊于 YOLOX 的 98.2%,其他 AP 指标高于实验所有模型,AP_M 甚至高出第 2 名 7.5%,各项召回率也取得最佳,表明本文的改进策略优越,较大的提升了模型的性能;YOLOX 的检测速度是绝对的优势,而 MS-RetinaNet 位于各模型第 3,FPS 达到了 21.8,快于 Faster RCNN,但 Transformer 结构计算复杂度高,GFLOPs 稍高于 RetinaNet 算法。

实际模型评估中,各算法模型对不同类别的检测效果不一。为了更好的评估各模型的优劣,本文对测试集中各类别 AP 进行评估,如图 11 所示。

由图 11 可以看出,YOLOX 的各类别 AP 较为平均,

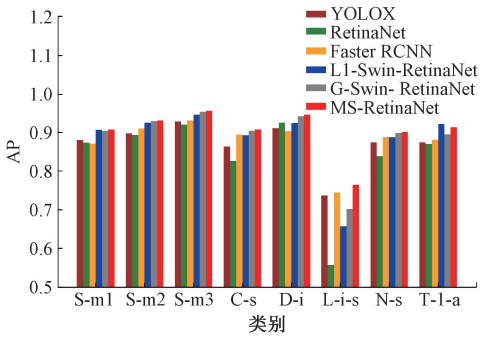
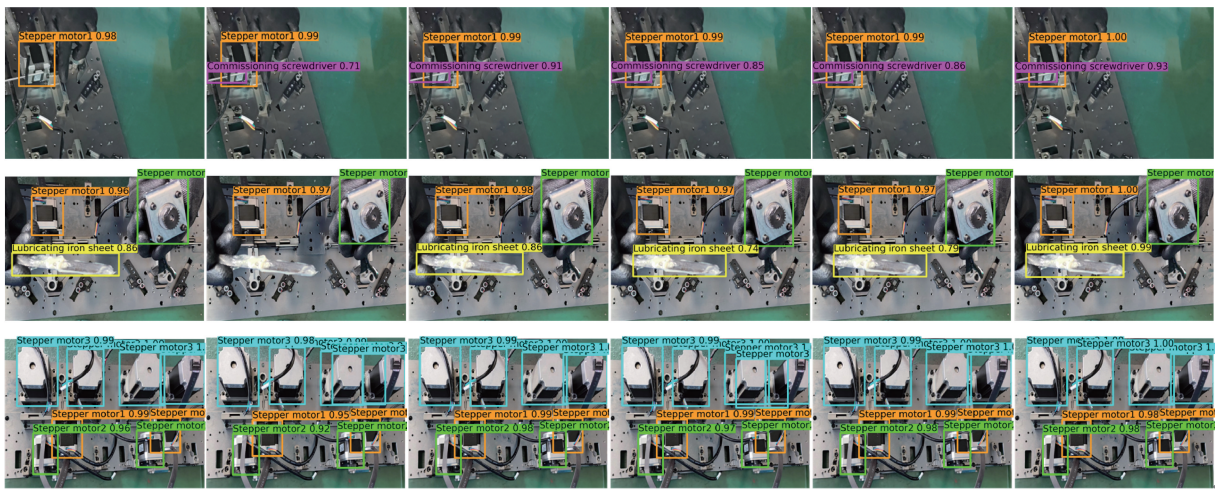


图 11　各模型各类别 AP 对比

Fig. 11　Comparison of various models and categories of AP

优于 RetinaNet，略逊于 Faster RCNN；MS-RetinaNet 各类别 AP 值表现优越，除 T-l-a 类别外，均取得最佳；尤其是 L-i-s，此类样本标注框相对较少，且长宽比偏离固定的训练锚框比，G-Swin-RetinaNet 使用改进的锚框比，精度稍强于 L1-Swin-RetinaNet 和 RetinaNet，但相比 Faster RCNN 和 YOLOX 差距明显。原因在于 YOLOX 拥有优化的锚框生成策略，而 Faster RCNN 存在 RPN 结构，两次边界框的回归校正，使得模型受此类问题影响较小；MS-RetinaNet 采用改进的 GIoU Loss 和最佳锚框比，增强对多尺度样本的关注与优化，实现了对该类算法缺点的弥补，AP 达到 76.4%，相对最优。

为了展示本文算法的实际检测效果，对比各类算法，依次选择 C-s 特征不全、宽高比差距较大和静态多目标的抓拍图像进行测试，结果如图 12 所示。



(a) YOLOX模型　(b) RetinaNet模型　(c) F-RCNN模型　(d) L1-S-t模型　(e) G-S-t模型　(f) MS-RetinaNet模型

(a) YOLOX model　(b) RetinaNet model　(c) F-RCNN model　(d) L1-S-t model　(e) G-S-t model　(f) MS-RetinaNet model

图 12　检测结果

Fig. 12　Test result

由图 12 可以看出，YOLOX 的实际检测精度欠佳，未能检测出 C-s，导致工艺流程误缺失影响时序检测效果；对 C-s 的检测结果只有 MS-RetinaNet 和 Faster RCNN 符合时序检测头的阈值要求，且定位效果较好；Retinanet 漏检出了 L-i-s，检测效果差，其中 L-i-s 概率分数大于阈值 0.9 的仅有 MS-RetinaNet；L1-Swin-RetinaNet 和 RetinaNet 出现了相同的问题，将 S-m3 误检测，算法检测出的结果多于实际情况，造成时序检测混乱；Faster RCNN 和 G-Swin-RetinaNet 对目标定位不够精准，若出现实际工艺抓拍到不同角度目标，可能检测出错，且存在多数检测未达到阈值标准的问题；MS-RetinaNet 检测效果最好，使用 Swin-Transformer 作骨干网络，改进 GIoU Loss 策略和最佳锚框比，优化模型特征提取和训练走向，有效的解决了一阶段系列算法边界框预测错误和回归不精准的问题，精度更甚改进的 Faster RCNN；相比于 YOLOX，虽然检测速度稍慢，但实际检测效果相对提升较大，不存在漏检，概率分数和定位精度高，检测鲁棒性更强，适合于实际应用。

本文 MS-RetinaNet 算法检测的完整工艺流程如图 13 所示，正确的工艺检测流程按照实线箭头方向执行，首先检测到工艺执行 L-i-s 润滑 S-m1，其次检测到 2 个 S-m1 同时存在并装配完成，接着检测到 C-s 调试 S-m1 连接处滑杆的吻合度，调试完后装配 S-m2；先检测到 L-i-s 润滑 S-m2，T-l-a 螺纹锁固 S-m2，执行完成后检测到 2 个 S-m2 都存在且装配完成；继续检测到 4 个 S-m3 逐个装配完成；4 个 N-s 装配完成；最后检测到使用 D-i 测试山板中山的高度差；每一步检测出待检测目标，再执行下一步，循环往复。图中虚线箭头指向的是当前错误工艺步骤，当算法检测到错误时，即当前正确为 C-s 调试 S-m1，而检测到 L-i-s 润滑 S-m2，算法模块配合后端进行

错误提示;继续执行检测当前工艺步骤任务,直到后续输入检测到 C-s 调试 S-m1 的步骤,再执行流程的下一步检测;其中任何一步出现误检和漏检,都会导致后端系统误动作,造成工艺流程检测的出错,无法起到时序检测的效果,故算法精度必须严格把控。

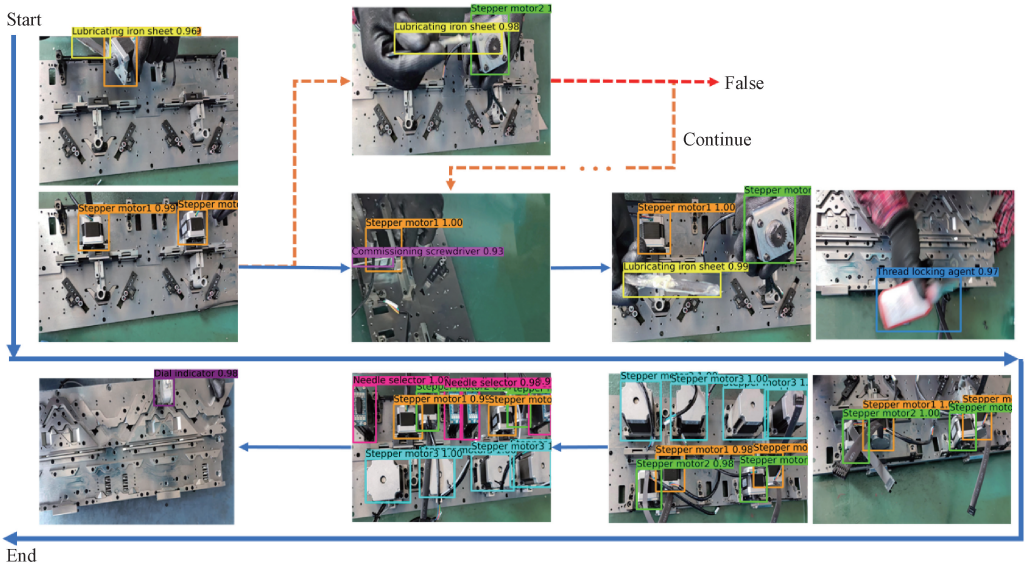


图 13　完整的工艺流程检测

Fig. 13　Detection of complete process

3　结　论

本文基于 RetinaNet 算法模型,将多头滑动窗口自注意力结构的 Swin-Transformer 引入,增加特征提取的全局精确度,减少模型量化误差;选择最小结构 Tiny,最大程度减少计算量,有效地缓解了 Transformer 结构的厚重感;采用改进的 GIoU Loss 策略,损失退化时,通过判定因子式引导模型训练方向,来提升边界框回归效果;根据实际样本,改进固定锚框比率,提高多尺度样本的召回率;以优异的检测精度搭配设计的时序检测头,准确检测出工艺流程的正误;基本达到实时精准检测的要求,从图像端实现工艺流程的时序检测。后续研究对模型检测速度进一步优化,并使其能够定量分析检测结果,应用于更多工业场景。

参考文献

[1] 邓中民, 胡灏东, 于东洋, 等. 结合图像频域和空间域的纬编针织物密度检测方法[J]. 纺织学报, 2022, 43(8): 67-73.
DENG ZH M, HU H D, YU D Y, et al. Density detection method of weft knitted fabrics making use of combined image frequency domain and spatial domain[J]. Journal of Textile Research, 2022, 43(8): 67-73.

[2] 陈瑶,张云伟,雷金辉,等. 基于视觉的四足动物骨架及行走步态特征提取方法[J]. 电子测量与仪器学报, 2022,36(2):68-77.
CHEN Y, ZHANG Y W, LEI J H, et al. Visual based feature extraction method for quadruped skeleton and walking gait[J]. Journal of Electronic Measurement and Instrumentation, 2022,36(2): 68-77.

[3] KHAN A I, AL-HABSI S. Machine learning in computer vision[J]. Procedia Computer Science, 2020, 167: 1444-1451.

[4] 史朋飞,韩松,倪建军,等. 结合数据增强和改进 YOLOv4 的水下目标检测算法[J]. 电子测量与仪器学报,2022,36(3):113-121.
SHI P F, HAN S, NI J J, et al. Underwater target detection algorithm based on data augmentation and improvement of YOLOv4[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(3): 113-121.

[5] 彭继慎,孙礼鑫,王凯,等. 基于模型压缩的 ED-YOLO 电力巡检无人机避障目标检测算法[J]. 仪器仪表学报,2021,42(10):161-170.
PENG J SH, SUN L X, WANG K, et al. Model compression based obstacle avoidance target detection algorithm for ED-YOLO power inspection unmanned aerial vehicles[J]. Chinese Journal of Scientific Instrument, 2021,42(10): 161-170.

[6] MANSOUR R F, ESCORCIA-GUTIERREZ J, GAMARRA M, et al. Intelligent video anomaly detection

and classification using faster RCNN with deep reinforcement learning model[J]. Image and Vision Computing, 2021, 112: 104229.

[7] 田翔, 张良. 改进的 R-C3D 时序行为检测网络[J]. 信号处理, 2021, 37(3): 447-455.
TIAN X, ZHANG L. Improved R-C3D temporal action detection network[J]. Journal of Signal Processing, 2021, 37(3): 447-455.

[8] 杨津. 基于深度学习的无锚框时序行为检测方法[D]. 杭州:杭州电子科技大学,2022.
YANG J. Anchor-free temporal action detection method based on deep learning[D]. Hangzhou: Hangzhou University of Electronic Science and Technology, 2022.

[9] CHEN Z, XIE L, NIU J, et al. Visformer: The vision-friendly transformer[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 589-598.

[10] LI B, LIU R, CHEN T, et al. Weakly supervised temporal action detection with temporal dependency learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(7): 4473-4485.

[11] LIU X, WANG Q, HU Y, et al. End-to-end temporal action detection with transformer[J]. arXiv preprint arXiv:2106.10271, 2021.

[12] ALE L, ZHANG N, LI L. Road damage detection using RetinaNet[C]. 2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018: 5197-5200.

[13] ZHANG J. A novel one-stage object detection network for multi-scene vehicle attribute recognition[J]. EAI Endorsed Transactions on Scalable Information Systems, 2022, 9(36): e7.

[14] AHMED B, GULLIVER T A, ALZAHIR S. Image splicing detection using mask-RCNN[J]. Signal, Image and Video Processing, 2020, 14(5): 1035-1042.

[15] YU Z, SHEN Y, SHEN C. A real-time detection approach for bridge cracks based on YOLOv4-FPM[J]. Automation in Construction, 2021, 122: 103514.

[16] GHOSH S, CHAKI A, SANTOSH K C. Improved U-Net architecture with VGG-16 for brain tumor segmentation[J]. Physical and Engineering Sciences in Medicine, 2021, 44(3): 703-712.

[17] 史雨馨,朱继杰,凌志刚. 基于特征增强 YOLOv4 的无人机检测算法研究[J]. 电子测量与仪器学报,2022, 36(7):16-23.
SHI Y X, ZHU J J, LING ZH G. Research on unmanned aerial vehicle detection algorithm based on feature enhancement YOLOv4[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(7): 16-23.

[18] WANG F, RAO Y, LUO Q, et al. Practical cucumber leaf disease recognition using improved Swin Transformer and small sample size[J]. Computers and Electronics in Agriculture, 2022, 199: 107163.

[19] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.

[20] GUO H, GUO C, XU B, et al. MLP neural network-based regional logistics demand prediction[J]. Neural Computing and Applications, 2021, 33(9): 3939-3952.

作者简介

李玮,2019 年于台州学院获得学士学位,现为青岛科技大学自动化与电子工程学院硕士研究生,主要研究方向为深度学习、计算机视觉等。
E-mail: 972217816@qq.com

Li Wei received his B. Sc. degree from Taizhou University in 2019. Now he is a M. Sc. candidate at the School of Automation and Electronic Engineering, Qingdao University of Science and Technology. His main research interests include deep learning, computer vision, etc.

高林(通信作者),2004 年于华东理工大学获得硕士学位,2012 年于华东理工大学获得博士学位,现为青岛科技大学副教授,硕士生导师,主要研究方向为数据挖掘、人工智能等。
E-mail: gaolin0619@126.com

Gao Lin (Corresponding author) received his M. Sc. from East China University of Science and Technology in 2004, and Ph. D. degree from East China University of Science and Technology in 2012. Now he is an associate professor and master's supervisor of Qingdao University of Science and Technology. His main research interests include data mining, artificial intelligence, etc.