



基于自适应全局定位算法的带钢表面缺陷检测

王延舒 余建波

Strip Surface Defect Detection Based on Adaptive Global Localization Algorithm

WANG Yan-Shu, YU Jian-Bo

在线阅读 View online: <https://doi.org/10.16383/j.aas.c210467>

您可能感兴趣的其他文章

基于深度学习的表面缺陷检测方法综述

A Survey of Surface Defect Detection Methods Based on Deep Learning

自动化学报. 2021, 47(5): 1017–1034 <https://doi.org/10.16383/j.aas.c190811>

基于贝叶斯CNN和注意力网络的钢轨表面缺陷检测系统

DeepRail: Automatic Visual Detection System for Railway Surface Defect Using Bayesian CNN and Attention Network

自动化学报. 2019, 45(12): 2312–2327 <https://doi.org/10.16383/j.aas.c190143>

缺陷检测技术的发展与应用研究综述

Review of Development and Application of Defect Detection

自动化学报. 2020, 46(11): 2319–2336 <https://doi.org/10.16383/j.aas.c180538>

一种面向散乱点云语义分割的深度残差特征金字塔网络框架

A Deep Residual Feature Pyramid Network Framework for Scattered Point Cloud Semantic Segmentation

自动化学报. 2021, 47(12): 2831–2840 <https://doi.org/10.16383/j.aas.c190063>

基于深度学习的桥梁裂缝检测算法研究

Research on Detection Algorithm for Bridge Cracks Based on Deep Learning

自动化学报. 2019, 45(9): 1727–1742 <https://doi.org/10.16383/j.aas.2018.c170052>

基于激光测量的航发叶片表面几何缺陷识别技术

Identification of Aero-engine Blade Surface Geometric Defects With Laser Measurement

自动化学报. 2020, 46(3): 594–599 <https://doi.org/10.16383/j.aas.c180080>

基于自适应全局定位算法的带钢表面缺陷检测

王延舒¹ 余建波¹

摘要 针对热轧带钢表面缺陷检测存在的智能化水平低、检测精度低和检测速度慢等问题,提出了一种基于自适应全局定位网络 (Adaptive global localization network, AGLNet) 的深度学习缺陷检测算法. 首先,引入一种残差网络 (Residual network, ResNet) 与特征金字塔网络 (Feature pyramid network, FPN) 集成的特征提取结构,减少缺陷语义信息在层级传递间的消失;其次,提出基于 TPE (Tree-structure Parzen estimation) 的自适应树型候选框提取网络 (Adaptive tree-structure region proposal extraction network, AT-RPN),无需先验知识的积累,避免了人为调参的训练模式;最后,引入全局定位回归算法,以全局定位的模式在复杂的缺陷检测中实现缺陷更精确定位. 本文实现一种快速、准确、更智能化、更适用于实际应用的热轧带钢表面缺陷的算法. 实验结果表明,AGLNet 在 NEU-DET 热轧带钢表面缺陷数据集上的检测速度保持在 11.8 帧/s,平均精度达到 79.90%,优于目前其他深度学习带钢表面缺陷检测算法. 另外,该算法还具备较强的泛化能力.

关键词 表面缺陷检测,深度学习,特征金字塔网络,自适应树型候选框提取,全局定位

引用格式 王延舒,余建波. 基于自适应全局定位算法的带钢表面缺陷检测. 自动化学报, 2024, 50(8): 1550–1564

DOI 10.16383/j.aas.c210467

Strip Surface Defect Detection Based on Adaptive Global Localization Algorithm

WANG Yan-Shu¹ YU Jian-Bo¹

Abstract A deep learning defect detection model based on adaptive global localization network (AGLNet) is presented to solve the problems of low intelligence, low detection accuracy and slow detection speed in hot-rolled strip surface defect detection. First, the feature extraction structure is combined with residual network (ResNet) and feature pyramid network (FPN) to reduce the disappearance of defect semantic information between layers transfers. Secondly, an adaptive tree-structure region proposal extraction network (AT-RPN) based on tree-structure Parzen estimation (TPE) algorithm is proposed, which does not need the accumulation of prior knowledge, and avoids the training model by manual parameter adjustment. Finally, a global localization regression algorithm is proposed to locate defects more accurately in complex defect detection using global positioning mode. In this paper, a fast, accurate, more intelligent and more applicable algorithm for surface defects detection of hot-rolled strips is realized. The experimental results show that the detection speed of AGLNet remains 11.8 frame/s and the average accuracy is 79.90%, which is better than other deep learning algorithms for strip surface defect detection on NEU-DET dataset. In addition, the algorithm has a strong generalization ability.

Key words Surface defects detection, deep learning, feature pyramid network (FPN), adaptive tree-structure region proposal extraction, global localization

Citation Wang Yan-Shu, Yu Jian-Bo. Strip surface defect detection based on adaptive global localization algorithm. *Acta Automatica Sinica*, 2024, 50(8): 1550–1564

热轧带钢在制造业中得到广泛运用,是制造业上游重要的原料之一. 但因其制作工艺复杂、制作

环节较多,在热轧带钢生产的过程中表面易于产生缺陷^[1]. 在竞争激烈的钢铁制造行业,高效的带钢缺陷检测可以帮助企业提高生产的效率,进而提高企业的竞争力^[2]. 传统的钢带表面检测方法主要是人工检测法和频闪光灯检测法^[3],均为非自动化检测方法,检测速度较慢,花费大量人力成本. 而计算机视觉检测方法是一种自动化的非接触的检测技术,早期多基于激光扫描检测法,后期发展出了基于摄像器件 CCD (Charge coupled device) 的检测方法^[4]. 相对于传统方法,计算机视觉检测具有分类能力强、检测精度高和检测成本低等优点. 而随着计算机技术、深度学习的发展,基于卷积神经网络 (Convolu-

收稿日期 2021-05-28 录用日期 2021-11-26

Manuscript received May 28, 2021; accepted November 26, 2021

国家重点研发计划 (2022YFF0605700), 国家自然科学基金 (92167107), 中央高校基本业务经费项目 (22120220575) 资助

Supported by National Key Research and Development Program of China (2022YFF0605700), National Natural Science Foundation of China (92167107), and Fundamental Research Funds for the Central Universities (22120220575)

本文责任编辑 徐德

Recommended by Associate Editor XU De

1. 同济大学机械与能源工程学院 上海 201804

1. School of Mechanical Engineering, Tongji University, Shanghai 201804

tional neural network, CNN) 的深度学习模型广泛应用于各种缺陷检测场景之中, 展现出更加卓越的检测性能^[5]. 其通过对底层特征的重组, 形成高维度的抽象特征和类别属性, 从而进行从局部特征到整体特征的图像精确识别^[6].

基于深度学习的缺陷检测是目标检测算法在实际生产中的应用, 任务是找出图像中的缺陷目标, 并确定其位置、大小和类别信息. 目前, 深度学习的检测算法主要分为两类, 一类是以 Faster R-CNN^[7]、Mask R-CNN^[8] 为代表的双阶段 (Two-stage) 检测算法; 另一类以 YOLO^[9]、SSD (Single shot multi-box detector)^[10] 为代表的单阶段 (One-stage) 检测算法. 双阶段检测算法将目标检测任务分成两个阶段来完成, 首先使用区域候选网络 (Region proposal network, RPN)^[11] 将原图分解成多个可能产生目标的候选区域, 再通过回归损失函数 (Regression loss function) 确定目标的位置信息, 通过分类损失函数 (Classification loss function) 确定目标的类别信息, 这种方法准确率比较高, 但是检测速度较慢. 凭借其在检测精度上的优越性, 双阶段检测方法在缺陷检测领域已经进行了诸多研究. Tao 等^[12] 将 Faster R-CNN 用于无人机电力巡检中绝缘子缺陷定位, 首先在自然环境下确定绝缘子区域, 再从绝缘子区域中实现缺陷检测. He 等^[13] 通过 Faster R-CNN 对带钢表面缺陷进行检测, 该网络的创新在于将 Backbone 中多级特征图组合为一个多尺度特征图, 在 NEU-DET 数据集上有较好的检测效果. 单阶段检测算法使用一个全卷积神经网络, 直接从整张图来预测包围框 (Bounding box) 的坐标, 以及包围框中包含物体的置信度和物体类别的概率, 是一个端到端的过程. 该方法速度较快但是精确度较低. Cheng 等^[14] 提出具有差异通道注意力机制和自适应性空间特征融合的 DEA_RetinaNet 深度学习网络. 该网络采用基于差分进化搜索的锚优化方法, 提高了网络的检测精度. 同时, 利用空间特征融合模块对卷积核提取的浅层和深层特征进行有效融合. 在检测带钢缺陷时达到 78.25% 的平均精度. Chen 等^[15] 使用 SSD 网络对接触网支撑装置上的紧固件缺陷区域进行定位, 使用不同层的特征图进行目标检测, 得到了较好的效果. Zhang 等^[16] 将 YOLO-v3 版本应用于桥梁表面缺陷定位, 其主要改进在于引入预训练权重、批再规范化 (Batch renormalization) 和 Focal loss, 进一步提高了缺陷检测率.

上述提到的双阶段深度学习方法均基于锚标定 (Anchor-based) 机制进行检测. 使用 Anchor 机制可以产生密集的锚框 (Anchor box), 使网络可以直接在此基础上进行目标分类和回归, 这有效地提

高了网络目标的召回能力, 对小目标检测来说提升非常明显. 同时, 在训练过程中加入了大量先验的参数, 让 Anchor-based 网络更容易训练且更加稳定, 使检测效果较为理想. 然而, Anchor-based 在实际训练和测试中也有诸多不足: 1) 在 Anchor-based 网络的训练过程中重复调节参数无疑浪费了时间; 2) Anchor-based 网络容易生成大量完全不包含目标的背景框, 导致正负样本严重不平衡, 浪费计算资源, 且易造成过拟合.

为提升检测的速度和灵活性, Anchor-free 网络以其更灵活的机制, 摆脱了使用锚而带来的计算量过大、冗余框过多和人为介入过度等问题, 更加适用于实际的工业检测领域, 这也逐渐成为近些年来研究热点. Anchor-free 主要以两种不同的方式检测没有预设锚的目标^[17]. 一种方法是将锚框转化为关键点描述, 然后将问题转为关键点的检测与匹配问题, 例如 CornerNet^[18] 和 CenterNet^[19]. Jia 等^[20] 提出了一种改进的 CenterNet 来完成卫生陶瓷的缺陷检测. 另一种方法利用语义分割的思想, 将图像分解成数个小格, 对小格进行分类和回归进行密集预测, 提升了目标检测性能. Zhu 等^[21] 提出一种目标检测的特征选择无锚 (Feature selective anchor-free, FSAF) 模块. 该模块可以嵌入到具有特征金字塔结构的单点探测器中. Tian 等^[22] 提出全卷积单阶段 (Fully convolutional one-stage, FCOS) 检测器, 该检测器以每像素预测的方式解决了目标检测问题. 它避免了所有与锚点相关的超参数的设置, 取得了比以往基于锚点的一级检测器更好的检测效果. Kong 等^[23] 提出 FoveaBox 网络, 该网络增加了两个子网, 一个用于像素级分类, 另一个用于边界盒预测, 从而提高了检测性能. 但是, Anchor-free 方法较为灵活的机制, 会导致检测效果不太稳定.

在实际的工业生产中, 需要精确度和速度都维持较高水准的检测算法, 从而应对现实的检测问题. 所以, 为了保持 Anchor-based 网络的稳定性和精确性, 同时具有 Anchor-free 网络的灵活和高速的特点, 通过点或边定位思想来获取目标更精确定位的双阶段目标检测网络相继出现. Lu 等^[24] 提出 Grid R-CNN, 借助卷积层生成热力图的方式确定初始的网格点, 从而借助这些网格点确定定位框四条边的位置, 使网络对目标的位置信息更加敏感. Wang 等^[25] 提出边侧感知边界定位 (Side-aware boundary localization, SABL) 的检测方法, 通过提取出每条边的特征, 利用 Bucket 机制对四条边进行精确定位, 取代了传统方法对整个锚框的位置回归. 然而, 上述方法虽然在目标的检测精度上取得了进步, 但仍然存在如下问题: 1) 工业产品众多, 各

类产品缺陷类型多种多样,为保证多样化的缺陷检测,上述方法需要经过多次实验获取先验知识,获取最优化的网络参数设置,才能保持高精度的缺陷检测,花费了大量的人力资源和时间;2)各类缺陷形状差异较大,在单一产品上的缺陷模式众多,空间布局密集,上述方法对密集检测效果不佳。

为解决上述提到的问题,本文提出自适应全局定位网络(Adaptive global localization network, AGLNet),在提高缺陷检测精度的同时,提出自适应树型候选框提取算法(Adaptive tree-structure region proposal extraction network, AT-RPN),提高了检测速度,使检测更智能化,更加适用于各种复杂的工业缺陷检测环境。本文的主要贡献有如下几个方面:1)提出一种集成残差网络(Residual network, ResNet)和特征金字塔网络(Feature pyramid network, FPN)的特征提取网络结构,有效地利用 CNN 的各层级特征,减少了层级传递中的语义缺失,更适用于工业场景中小缺陷的检测,缓解表面缺陷空间布局密集导致的检测率低下问题;2)提出 AT-RPN 算法,在 RPN^[11]的基础上加入了基于 TPE(Tree-structured Parzen estimator)的自适应 Anchor 调节模块,以及 RoIAlign(Region of interest align)^[8]池化结构,实现了训练和测试过程锚高宽比(Anchor-ratio)参数的自主调节,减少了人为的介入,提高了检测精度,同时节约了运算成本;3)采用全局定位回归(Global localization regression)算法,该方法预测特征图中每个单元格与真实标定框(Ground-truth box)四个方向上的偏移量(Box offsets),经过二项分类预测(Binomial classification prediction)和稀疏选择器模块(Sparse selection module),获取筛选后的激活点参量加入回归运算,从而获取目标更精确的定位信息,同时保证检测的速度。实验中,使用 NEU-DET 数据集^[13]进行验证。结果表明,AGLNet 实现了端到端的热轧带钢表面缺陷检测,解决了 Anchor-based 网络需要大量先验知识的问题,同时也缓解了 Anchor-free 网络中不稳定的问题,在密集检测中效果更佳,平均精度均值达到 79.90%,具有较高的准确度和较好的实时性,适用于复杂的工业缺陷检测环境中。

1 自适应全局定位网络

1.1 AGLNet 结构

如图 1 所示,AGLNet 结构主要包括三部分:1)特征提取网络,采用残差网络和特征金字塔相结合(ResNet50_FPN)的特征信息提取结构。2)AT-RPN 候选框提取算法,结合提取到的特征信息和自

适应调节后的锚框参数,从而获取相应的候选区域。其中 RoIAlign 池化网络利用多尺度特征图和真实标定框信息筛选核实的候选区域特征图。3)全局定位回归算法,该模块对候选区域中每个子区域进行回归计算,获得目标的位置信息,然后利用分类损失函数获取目标的类别信息。

1.2 特征提取网络

特征提取网络采用 ResNet50_FPN 提取出原图像的多尺度的特征图,提取流程如图 1 中所示。ResNet50 对输入图像进行从低维度到高维度的特征提取,生成各自阶段的特征图{ C_2, C_3, C_4, C_5 };FPN 利用 ResNet50 生成的各阶段特征图,采取从高维度到低维度的上采样以及横向连接操作, C_5 经过 256 个 1×1 卷积得到 T_5 。同时, T_5 经过上采样得到的结果与特征图 C_4 经过 256 个 1×1 的卷积得到的结果进行张量相加得到 T_4 。 T_3 和 T_2 的获取流程同 T_4 。经过张量相加得到的{ T_5, T_4, T_3, T_2 }分别经过 256 个 3×3 的卷积,最终得到从高维度抽象特征到低维度底层特征。

1.3 AT-RPN 候选框提取算法

在目标检测中,使用多种形状大小的锚框作为初始预测,经过 RPN 和感兴趣区域(Region of interest, RoI)筛选后产生候选区域,以完成后续的目标回归和分类任务。自适应参数配置可以减少大量的电脑运算时间,以及获得更准确的结果,这使得锚框的配置成为重要的超参数。所以,本文采用了基于 TPE 的参数优化算法,形成了自适应的锚高宽比调节模块,应用于 AGLNet,通过迭代获得最优的参数设置,并减少了人为的介入。超参数 A_{scale} 和 A_{ratio} 可以借助锚的高、宽计算获得,如式 (1) 和式 (2) 所示,通过优化候选框高、宽的参数组合,以达到最优的检测效果。

$$A_{scale} = \sqrt{h \times w} \quad (1)$$

$$A_{ratio} = \frac{h}{w} \quad (2)$$

$$w \leq W, \quad h \leq H \quad (3)$$

其中, A_{scale} 是锚框尺度(Anchor scale),意味着锚框面积的开方; A_{ratio} 是锚框的高宽比; W, H 是候选区域的宽与高。

近些年来有很多关于超参数优化问题的研究,其中较为有效的方法是贝叶斯优化^[26]。而 TPE 优化算法是根据贝叶斯优化的思想,使用高斯混合模型(Gaussian mixture model, GMM)来学习超参数模型的方法,可以使用预期改进选择下一个超参

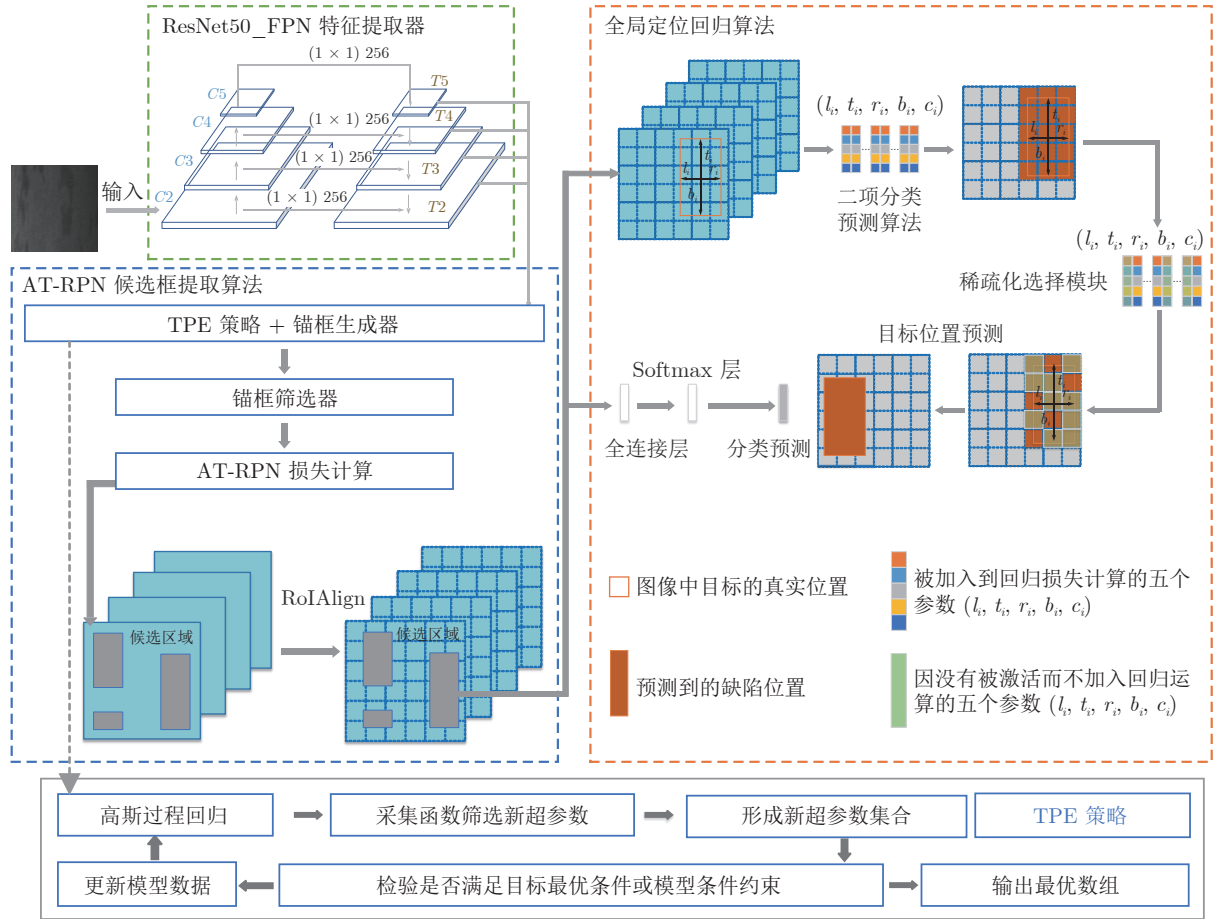


图1 AGLNet 结构

Fig.1 The structure of AGLNet

数. 相比于其他贝叶斯优化方法, TPE 在高维空间效果更好, 同时在速度上有显著的提升^[26].

在传统检测中, 需要首先设定锚框的高宽比和框的大小, 例如 $[0.5, 1, 2]$ 和 $[4, 8, 16]$. 但是, 这种办法具有较明显的不足: 只是对一张图片的多个像素点进行机械选取, 无法根据实际情况选取出最优化的框, 此过程生成约 20 000 张候选区域, 再根据交并比 (Intersection over union, IoU) 的标准进行筛选^[11], 这无疑浪费了大量的资源. 并且, 对于一些形状特殊的目标缺陷, 根据传统设置的参数无法选取到最优的结果. 由此启发了本文对锚框尺度 A_{scale} 和锚框高度比 A_{ratio} 两个超参数进行优化的想法, 进而打造最优质的候选框. AT-RPN 将锚框高 (h) 和宽 (w) 作为输入的超参数, 并通过式 (1) 和式 (2) 计算出锚框的高宽比和锚框尺度.

AT-RPN 超参数优化过程如下.

步骤 1. 设置超参数筛选域, 对超参数高 (h) 和宽 (w) 的取值范围进行定域.

步骤 2. 设置目标函数, AT-RPN 的损失函数

由分类损失函数和回归函数构成, 定义为

$$L(\{p_i\}, \{r_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \mu \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(r_i, r_i^*) \quad (4)$$

在 AT-RPN 算法中, 目标函数是损失函数, 目标值为损失函数的值, 高和宽作为超参数. 通过输入不同的高和宽, 反复迭代实现损失值的减少, 以达到最优的目的. 式中, p_i 表示第 i 个锚框的预测概率, p_i^* 表示 IoU 标定的概率. 当标定框包含目标时, $p_i^* = 1$; 当标定框不包含目标时, $p_i^* = 0$. N_{cls} , N_{reg} , μ 为固定值.

$L_{cls}(\cdot)$ 表示是否为目标的对数损失函数, 即

$$L_{cls}(p_i, p_i^*) = \lg[p_i \cdot p_i^* + (1 - p_i)(1 - p_i^*)] \quad (5)$$

$L_{reg}(\cdot)$ 表示修正参数后的回归损失函数, 定义为

$$L_{reg}(r_i, r_i^*) = \sum_{i \in \{x, y, w, h\}} smooth_L(r_i - r_i^*) \quad (6)$$

式中, $r_i = (r_x, r_y, r_w, r_h)$ 表示候选区域预测的距离参数, $r_i^* = (r_x^*, r_y^*, r_w^*, r_h^*)$ 表示与目标真实标定框的距离. $smooth_L(x)$ 函数定义为

$$smooth_L(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{否则} \end{cases} \quad (7)$$

r_i 和 r_i^* 的计算式分别为

$$\begin{aligned} r_x &= \frac{x - x_a}{w_a}, & r_y &= \frac{y - y_a}{h_a} \\ r_w &= \lg \frac{w}{w_a}, & r_h &= \lg \frac{h}{h_a} \\ r_x^* &= \frac{x^* - x_a}{w_a}, & r_y^* &= \frac{y^* - y_a}{h_a} \\ r_w^* &= \lg \frac{w^*}{w_a}, & r_h^* &= \lg \frac{h^*}{h_a} \end{aligned} \quad (8)$$

$$\begin{aligned} r_x^* &= \frac{x^* - x_a}{w_a}, & r_y^* &= \frac{y^* - y_a}{h_a} \\ r_w^* &= \lg \frac{w^*}{w_a}, & r_h^* &= \lg \frac{h^*}{h_a} \end{aligned} \quad (9)$$

式中, x^*, y^*, w^*, h^* 分别为目标真实标定框中心点横坐标和纵坐标、标定框宽度和高度; x_a, y_a, w_a, h_a 分别为预测出的锚框中心点横坐标和纵坐标、锚框宽度和高度.

步骤 3. 在模型被赋予到初始化的超参数数值之后, 建立高斯回归模型, 目的是使目标函数中的超参数的取值符合联合正态分布, 便于后续模型利用迭代过的历史信息进行优化取值.

通过以上 3 个步骤, AT-RPN 获取到一个待优化的超参数模型:

$$L: \chi \rightarrow \mathcal{R}, \quad x \in \chi \quad (10)$$

其中, χ 是超参数的范围, x 为模型中的超参数.

步骤 4. 转换待优化模型为代理函数, 即

$$x_{i+1} = \arg \max_x L(x; \Psi_i) \quad (11)$$

步骤 5. 使用采集函数 (Acquisition function), 在每次迭代中, 从现有的众多超参数中决定输入的超参数.

步骤 6. 选取一个输入的超参数 ($x_i \in \chi$), 输入到原模型 $L(x)$ 中, 得到的结果为

$$y_t = L(x; \Psi_i) + \epsilon \quad (12)$$

其中, ϵ 为平均值的高斯分布 $\epsilon \sim N(0, \sigma^2)$ 的噪音, σ 是噪音的方差.

步骤 7. 观察到该组值 x_i, y_i , 并加入到现有的观察到的数据中:

$$\Psi_{i+1} = \{\Psi_i, (x_{i+1}, y_{i+1})\} \quad (13)$$

式中, Ψ_i 为现有观察到的数据, 待新数据加入后进行下一轮迭代, 重复步骤 4 ~ 7.

在式 (11) 中, 采用期望函数 (Expected improvement, EI) 作为采集函数, 通过构造增益期望

函数来获取新测量值相较于历史最优值的增益的期望值, 使其可以在探索和利用之间选取一个较为平衡的点作为下一组超参数的标准. 期望函数中最常见的为

$$EI_{y^*}(x) = \int_{-\infty}^{y^*} (y^* - y)p(y|x) dy \quad (14)$$

式中, y^* 是目标函数的阈值, x 是建议的超参数, y 是使用超参数 x 的目标函数所取到的实际值, $p(y|x)$ 是表示在给出 x 后 y 的概率. TPE 优化模型在遵循贝叶斯优化的基础上, 设置:

$$p(x|y) = \begin{cases} l(x), & \text{若 } y \leq y^* \\ g(x), & \text{若 } y > y^* \end{cases} \quad (15)$$

在这里, 本文设置上一次运算的 y_t 值为本次运算的阈值 y^* . $l(x)$ 代表小于或等于上一次迭代中目标函数 y^* 的 x 的概率分布; $g(x)$ 代表大于上一次迭代中目标函数值 y^* 的 x 的概率分布. 为了优化目标函数的值, 不断进行迭代以获取 $l(x)/g(x)$ 的最大值.

在本模型中, 采用 EI 作为采集函数, 构建 TPE 算法的 EI 函数为

$$\begin{aligned} EI_{y^*}(x) &= \int_{-\infty}^{y^*} (y^* - y)p(y|x) dy = \\ &= \int_{-\infty}^{y^*} (y^* - y) \frac{p(y|x)p(y)}{p(x)} dy \end{aligned} \quad (16)$$

设置 $\gamma = p(y < y^*)$, γ 为目标函数的值低于阈值的概率, 且

$$\begin{aligned} p(x) &= \int p(x|y)p(y) dy = \\ &= \gamma l(x) + (1 - \gamma)g(x) \end{aligned} \quad (17)$$

根据前述, 可知:

$$\begin{aligned} EI_{y^*}(x) &= \int_{-\infty}^{y^*} (y^* - y)p(x|y)p(y) dy = \\ &= l(x) \int_{-\infty}^{y^*} (y^* - y)p(y) dy = \\ &= \gamma y^* l(x) - l(x) \int_{-\infty}^{y^*} p(y) dy \end{aligned} \quad (18)$$

则原 EI 函数可以转换为^[25]

$$EI_{y^*}(x) = \frac{\gamma y^* l(x) - l(x) \int_{-\infty}^{y^*} p(y) dy}{\gamma l(x) + (1 - \gamma)g(x)} \quad (19)$$

在每次迭代中, TPE 算法可以决定得到最优 EI 的 x^* 作为候选超参数点加入 Ψ_i 中, 进行下一步迭代, 直至达到最大迭代次数或终止时间. 本文提出的基于 TPE 的优化算法见算法 1 所示.

算法 1. TPE 算法

输入. 初始锚框宽和高 (w, h) 数组

输出. 锚框最优组合 (w, h)

- 1: 模型参数初始化;
- 2: For $i = 1, 2, 3, \dots$, 迭代开始;
- 3: 使用高斯过程回归处理数组;
- 4: 设置采集函数 $\arg \max_x L$ 并选择 x_{i+1} 输入到模型:

$$x_{i+1} = \arg \max_x L(x; \Psi_i)$$

- 5: 获取目标函数值 y_{i+1} ;
- 6: 将新超参数 (x_{i+1}, y_{i+1}) 加入到原超参数数组 $\Psi_{i+1} = \{\Psi_i; (x_{i+1}, y_{i+1})\}$, 原超参数数组更新;
- 7: 根据新的超参数数组更新统计模型;
- 8: 重复步骤 3 ~ 7, 直至迭代到最大次数或取得最优解;
- 9: 迭代结束.

图 2 为本文提出的 TPE 自适应 Anchor-ratio 调节模块流程图.

1.4 AT-RPN

输入图像在经过 ResNet50_FPN 卷积后, 得到不同尺度的特征图, 依次输入到 AT-RPN, 得到一组矩形候选区域. 每个候选区域都具有描述该区域包含缺陷目标的置信度. 在传统的 RPN^[11] 中, 通过如下步骤完成候选区域提取的工作: 1) 采用 256 个 3×3 的窗口依次在各层的特征图上进行滑动. 2) 在滑过的每一个位置上映射出三个具有固定尺寸和不同高宽比的锚框, 各层特征图的锚框尺度不同, 锚框配置为 $A_{\text{scale}} = [8, 16, 32, 64, 128]$, $A_{\text{ratio}} = [0.5, 1, 2]$. 3) 上述操作得到的 256 维特征图并列输入到回归层和分类层. 回归层对锚框的中心点坐标和宽高进行回归预测, 从而得到候选区域的位置信息; 分类层计算区域前景后景分数, 判断是否包含缺陷目标. 4) 采用非极大值抑制法 (Non-maximum suppression, NMS) 对获得的候选区域进行筛选, 减少冗余且不影响精度的区域. 本文提出的 AT-RPN 嵌入了基于 TPE 的锚优化模块, 在训练和测试时, 无需输入锚框的配置, 在初始化值赋予之后自动进行超参数优化, 获取最优的参数配置, 提高了网络的精度. 图 3 为 AT-RPN 的整体结构图.

在训练过程中, 给每个锚框分配一个二进制标签, 采用 IoU 的方式, 判断该区域属于目标还是背景. IoU 表示锚框和目标真实标框之间的交集面积和并集面积之比, 定义为

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (20)$$

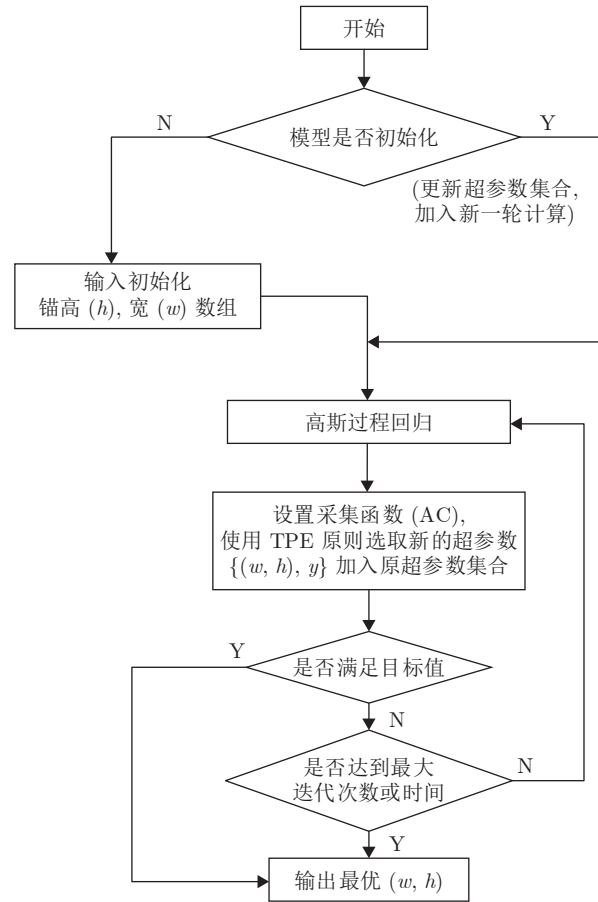


图 2 TPE 自适应 Anchor-ratio 调节模块流程图
Fig. 2 Flow chart of TPE adaptive anchor-ratio adjustment module

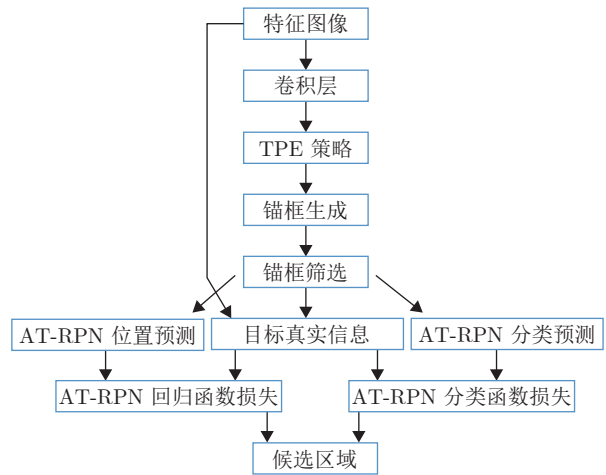


图 3 AT-RPN 整体结构图
Fig. 3 Whole structure of AT-RPN

$$I = \begin{cases} 1, & 0.7 < IoU \leq 1 \\ 0, & 0.3 < IoU \leq 0.7 \\ -1, & 0 \leq IoU \leq 0.3 \end{cases} \quad (21)$$

式中, I 为分配的二进制标签, 当 IoU 大于等于 0.7 时, 标记该区域为正样本 ($I = 1$), 加入训练; 当 IoU 小于等于 0.3 时, 标记该区域为负样本 ($I = -1$), 加入训练; 而剩余的样本会影响模型精度, 故不参与训练 ($I = 0$).

1.5 RoIAlign

AGLNet 采用 RoIAlign 方法将候选区域池化为固定尺寸的特征图, 以便后续的目标分类和位置回归的操作. 在传统的双阶段检测网络中, 一般使用 RoIPooling 的方法经过两次量化过程固定特征图: 1) 将候选区域的边界量化为整数坐标值; 2) 将量化后的边界区域平均分割成 $k \times k$ 个单元, 对每一个单元的边界进行量化. RoIPooling 反向传播具体计算为

$$\frac{\partial L}{\partial x_i} = \sum_r \sum_j [i = i^*(r, j)] \frac{\partial L}{\partial y_{rj}} \quad (22)$$

式中, r 表示候选区域的序列, i 代表像素点的序列, x_i 表示池化前特征图上的像素点, y_{rj} 代表池化后的第 r 个候选区域的第 j 个点, $i^*(r, j)$ 代表最大池化过程中选出的最大像素点所在的坐标. 当池化过程中某一个像素值采用了当前点 x_i 的像素值, 即 $i = i^*(r, j)$ 时, 才会传回梯度. 但是, 在整个过程中存在一些不足, 最终结果是通过两次四舍五入的量化过程得到的, 由此产生了一定的误差, 影响到目标检测的精确性. 为解决该问题, 本文采用 RoIAlign, 取消量化过程, 使用双线性内插的方法得到坐标浮点数像素点上的图像值, 从而将上述问题转化为连续操作, 具体过程如下: 1) 遍历每一个候选区域, 保持该区域的浮点数边界不变; 2) 将候选区域平均分割为 $k \times k$ 个单元, 保持每个单元边界不变; 3) 在每个单元中计算固定的 4 个坐标位置, 用双线性内插方法计算该 4 个位置的值, 再进行最大池化的操作. 双线性内插法是做了两次一维的线性插值, 用 4 个最邻近估计给定灰度, 即

$$(u', v'), (u', v' + 1), (u' + 1, v'), (u' + 1, v' + 1) \quad (23)$$

新图像的输入像素点 (u_0, v_0) 必定落在原始图像的 4 个像素点中间, 具体计算为

$$\begin{aligned} g(u_0, v_0) = & (1 - \alpha)(1 - \beta)g(u', v') + \\ & \alpha(1 - \beta)g(u' + 1, v') + \\ & \beta(1 - \alpha)g(u', v' + 1) + \\ & \alpha\beta g(u' + 1, v' + 1) \end{aligned} \quad (24)$$

式中, α 设为在平面直角坐标系中, x 轴上 u' 到 u_0 的长度与 u' 到 $u' + 1$ 的长度的比值; β 设为在平面直角坐标系中, y 轴上 v' 到 v_0 的长度与 v' 到 $v' + 1$ 的长度的比值.

RoIAlign 的反向传播公式为

$$\begin{aligned} \frac{\partial L}{\partial x_i} = & \sum_i \sum_j [d(i, i^*(r, j)) < 1] \times \\ & (1 - \Delta h)(1 - \Delta w) \frac{\partial L}{\partial y_{rj}} \end{aligned} \quad (25)$$

式 (25) 表示每一个与 $i^*(r, j)$ 横、纵坐标差值小于 1 的点都接受与此对应的点 y_{rj} 回传的梯度, $d(\cdot, \cdot)$ 表示两点之间的距离, Δh 和 Δw 表示 x_i 与 $i^*(r, j)$ 横纵坐标的差值.

1.6 全局定位回归算法

在传统的双阶段网络中, 通过包围框回归 (Bounding-box regression) 确定目标在图像中的位置信息^[7-8]. 通过第一阶段网络获得目标的候选框 $P(x_p, y_p, w_p, h_p)$, x_p, y_p 是候选框中心点的横、纵坐标信息, w_p, h_p 是候选框的宽和高. 目标的真实位置信息通过 $G(x_g, y_g, w_g, h_g)$ 来表示, x_g, y_g 是目标真实标定框中心点的横、纵坐标信息, w_g, h_g 是真实标定框的宽和高信息. 在传统的回归方法 (例如 Faster R-CNN) 中, $\Delta x, \Delta y, \Delta w, \Delta h$ 是预测单个框的偏移量, 具体计算为

$$\begin{aligned} \Delta x = \frac{x_g - x_p}{w_p}, \quad \Delta y = \frac{y_g - y_p}{h_p} \\ \Delta w = \lg \frac{w_g}{w_p}, \quad \Delta h = \lg \frac{h_g}{h_p} \end{aligned} \quad (26)$$

传统的方法 Faster R-CNN 是通过预测上述 4 个偏移量来确定目标的位置, 在候选框和真实标定框之间存在很大位移时, 精确定位会变得很困难, 影响最终的检测精度. 在实际的工业缺陷检测中, 缺陷的形状各异, 需要更加精确的位置信息, 因此本文引入了全局定位回归方法, 如图 1 所示. 在经过 RoIAlign 之后, 网络可以获得被分为 $k \times k$ 单元的特征图, 将每一个单元的特征信息输入到全卷积网络中, 计算每个单元特征 u_i 在各自位置 (x_i, y_i) 到真实标定框 G 左上角和右下角的距离, 借此进行位置偏移量的预测, 以确定目标的精确位置^[27]. 本文方法摆脱了锚框的限制, 使用对全局各点的偏移量的回归计算, 使网络对目标的位置更加敏感, 获得精确的位置信息. 特征图各单元格位置偏移量具体计算为

$$\begin{aligned} l_i = \frac{x_i - x_l}{w_p}, \quad t_i = \frac{y_i - y_t}{h_p} \\ r_i = \frac{x_r - x_i}{w_p}, \quad b_i = \frac{y_b - y_i}{h_p} \end{aligned} \quad (27)$$

其中, l_i, t_i, r_i, b_i 分别表示单元格到真实标定框左边界、上边界、右边界和下边界的距离; x_i, y_i 表示单元格位置的横、纵坐标, x_l, y_t 和 x_r, y_b 分别表示

真实标定框左上角和右下角的横、纵坐标。

然而, 将特征图的单元格全部输入到全卷积中, 无疑会花费大量计算, 并且单元格具有前景和背景的信息, 具有背景信息的单元格加入回归中会降低最终检测的效果. 所以, 进行了两次稀疏化的操作: 首先, 标签化每个单元格, 使用二项分类预测第一次的稀疏化操作, 通过该预测算法, 将每一个单元格标签化, 判断每个单元是属于前景还是背景特征. 若该单元格为前景信息, 则判断 c_i 为 1; 若不是前景信息, 则判断 c_i 为 0. 并只允许具有前景信息的单元格加入运算, 定义如下:

$$c_i = \begin{cases} 1, & u_i \in G \\ 0, & \text{否则} \end{cases} \quad (28)$$

c_i 作为单元格分类判断的输出量加入到位置偏移量运算中, 形成 5 个条件的回归 l_i, t_i, r_i, b_i, c_i . 在训练时, 二项分类预测输出值 c_i 传入 sigmoid 激活函数计算二值交叉熵损失. 在上述操作完成后, 进入第二个稀疏化选择模块进行筛选, 在全连接层中加入 Dropout 稀疏化操作, 使神经网络向前迭代的过程中随机让网络的 50% 的节点停止工作, 并且不更新权重, 有效地避免了输入太多点后产生过拟合, 同时显著减少了运算量.

图 4 为 AGLNet 方法 (图 4(a)) 与传统回归的 Faster R-CNN (图 4(b)) 和关键点定位算法 Grid R-CNN (图 4(c)) 之间的网络结构比较. 由图 4 可知, Faster R-CNN 将候选框和真实标定框中心点之间的差值与两框高和宽之间的差值, 输出为单一向量进行回归预测, 对目标的位置缺少有效的监督, 无法得到较好的检测效果. Grid R-CNN 采用了多关键点定位的方法, 首先通过热力图 (Heatmap) 的方式找到候选区域中目标的 9 个关键点位置信息, 通过预测点的加权求和得到预测框的边界信息, 该方法采用多点监督的方式, 具有更高的敏感性, 可以获得更好的检测效果. 本文方法对候选区域的单元格进行两次筛选, 获取对定位起到关键作用的单元格位置信息加入到回归运算中, 对目标位置具有更高的敏感性.

2 实验结果分析

2.1 实验细节

实验采用的硬件配置为 Intel (R) Core (TM) i7-8700K 处理器, GTX 1080 Ti 显卡, 软件环境为 CUDA10.0 和 cuDNN7.6, 开发环境为 Ubuntu 18.04. AGLNet 模型通过 Pycharm 和开源的深度学习框架 Pytorch1.1.0 来完成. 在训练中使用随机

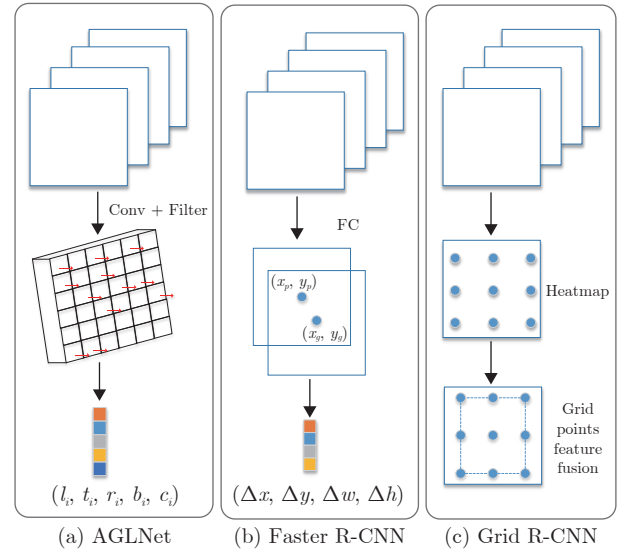


图 4 AGLNet 与 Faster R-CNN 和 Grid R-CNN 的比较
Fig.4 Comparison of AGLNet with Faster R-CNN and Grid R-CNN

梯度下降 (Stochastic gradient descent, SGD) 优化器优化网络参数, 学习率为 0.0015, Momentum 为 0.9, 一共迭代 24 000 次. 在目标检测中, 评价网络性能时需要同时兼顾精确率 (Precision, P) 和召回率 (Recall, R), 一般使用平均精度 (Average precision, AP) 和平均精度均值 (Mean average precision, mAP) 来评价网络的性能. AP 表示某一类缺陷的检测精度, 计算式如式 (29) 所示. mAP 是所有类别检测精度的均值, 计算式如式 (30) 所示:

$$AP(i) = \int_0^1 P(R) dR \quad (29)$$

$$mAP = \frac{\sum_{i=1}^n AP(i)}{n} \quad (30)$$

其中, i 表示某一类别的缺陷, $AP(i)$ 为某一类别的检测精度, n 为总类别数. P 表示精度, R 表示召回率, P 和 R 的计算式分别为

$$P = \frac{TP}{TP + FP} \quad (31)$$

$$R = \frac{TP}{TP + FN} \quad (32)$$

式中, TP 为成功预测的正例, FN 为被误判为负例的正例, FP 为被误判为正例的负例. $TP + FP$ 即表示为检测出的正样本总数, $TP + FN$ 表示所有正样本数. 通过计算 IoU 并设定阈值确定分类结果是否正确, 从而确定 TP 、 FP 和 FN 的数量.

FPS (Frame per second) 为网络模型每秒可以检测的图片数量, 用于评价网络模型的检测速度,

具体计算为

$$FPS = \frac{N_{\text{figure}}}{A_{\text{time}}} \quad (33)$$

式中, N_{figure} 为检测图片的总数, A_{time} 为检测总时间.

2.2 带钢表面缺陷检测结果分析

本次实验采用 NEU-DET 数据集^[13] 验证 AGLNet 模型在工业缺陷检测场景下的精度和速度. 该数据集是关于热轧带钢的 6 类表面缺陷的图片和标签, 每类缺陷图片为 300 张, 共 1800 张, 图片大小为 200×200 像素. 缺陷类型如图 5 所示, 分别为裂纹 (Cracking)、夹杂 (Inclusion)、斑块 (Patches)、麻点 (Pitted-surface)、压入氧化 (Rolled-in_scale) 和划痕 (Scratches).

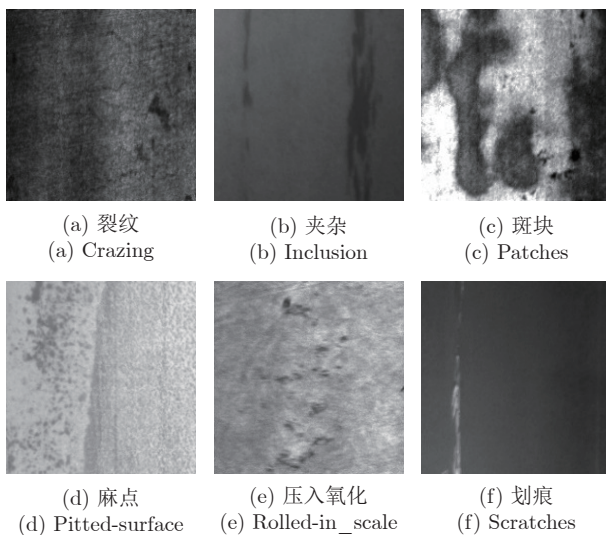


图 5 NEU-DET 数据集热轧带钢表面缺陷
Fig.5 Surface defects of hot rolled strip in NEU-DET dataset

随机划分训练集和测试集比例为 7 : 3, 即测试集为 1260 张, 测试集 540 张. 图片储存为 .eps 格式的灰度图, 缺陷的位置和类别信息以 VOC 数据集的格式储存为 .xml 文件. 由于热轧带钢表面的复杂性, 同类缺陷的外观具有较大差异, 而不同种类缺陷色差和形状具有相似部分, 且与背景颜色差异度低, 目标缺陷的高宽比具有很大的差异性, 因而增加了精确位置定位的难度.

2.2.1 带钢表面缺陷检测结果分析

为了检验 AGLNet 的检测精度, 选取了 4 种经典的目标检测模型 (Faster R-CNN, RetinaNet, FCOS, Grid R-CNN) 和 YOLO 系列算法 (YOLO-v1, YOLO-v2, YOLO-v3, YOLO-v4, YOLO-v5, YOLOF) 与 AGLNet 模型进行对比实验. Faster R-

CNN, RetinaNet, FCOS, Grid R-CNN 均采用 ResNet50 作为特征提取网络; 为了保证 YOLO 系列算法的完整性和高效性, YOLO 系列算法没有使用 ResNet50 替代原本特征提取网络. 如表 1 对比实验结果所示, 相比于其他方法, AGLNet 实验方法取得了最高的平均精度均值. 相较于精度较高但速度较慢的经典双阶段模型 Faster R-CNN, AGLNet, 在 NEU-DET 数据集上的 mAP 依然提高了 0.70% 的精度. 相较于兼顾检测速度和精度的 RetinaNet 而言, AGLNet 的 mAP 提高了 4.54%. 对比于同样基于 Anchor-free 思想的 FCOS 来说, AGLNet 提高了 4.72% 的 mAP. 相较于同样采用点偏移量回归定位的双阶段网络 Grid R-CNN, AGLNet 的 mAP 提高了 6.76%. 就 YOLO 系列检测算法而言, AGLNet 比 YOLO-v1, YOLO-v2, YOLO-v3, YOLO-v4, YOLO-v5, YOLOF 分别提高了 17.00%, 13.37%, 10.50%, 1.91%, 3.08% 和 2.58%. 从单一缺陷角度分析, AGLNet 在划痕缺陷的检测上有显著的提升, 达到了最高的检测精度 96.64%. 另外, AGLNet 检测麻点缺陷时保持最优的检测精度, 达到了 91.67%; 在夹杂和斑点两种缺陷的检测上也具有十分突出的检测结果; 在检测裂纹和压入氧化缺陷特征时, 仍维持较高的水平.

2.2.2 检测结果分析

为进一步评估网络的性能, 表 2 展示了 AGLNet, Grid R-CNN 和 Faster R-CNN 在 NEU-DET 数据集的对比测试结果. Faster R-CNN, Grid R-CNN 和 AGLNet 均属于双阶段网络, 通过 AGLNet 与其他两种算法的对比可以得出, 凭借 AGLNet 对于缺陷特征的高敏感性的特点, 可以在实际的缺陷检测中检测到更细小的缺陷信息, 最终获取更准确的缺陷定位, 生成更多比较细小的缺陷检测框. Grid R-CNN 和 AGLNet 都是基于 Anchor-free 思想的检测模型, 通过表 2 中对比可以得出, 此类思想的检测算法均会生成数量较多、较为细小的缺陷框, AGLNet 识别出来的小缺陷位置更加准确. 综上结果, AGLNet 凭借其图像的高敏感性, 降低了多个缺陷空间位置聚集导致的干扰, 对于夹杂、斑点、麻点以及划痕这四类空间分布密集的缺陷具有突出的检测效果. 但是, 由于 AGLNet 对图像缺陷的敏感程度过高, 其在检测裂痕缺陷时, 提取出过多的缺陷特征, 将实际的缺陷区域划分成多个缺陷区域, 导致检测出的区域和实际标定的区域出现了较大偏差, 致使该缺陷类型的检测精度相对较低.

2.2.3 检测速度分析

为了评估 AGLNet 的检测速度, 表 3 将模型每

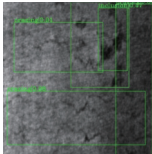
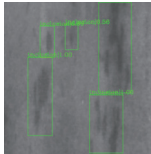
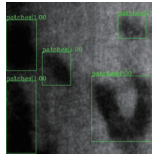
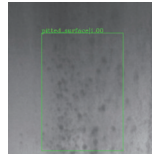
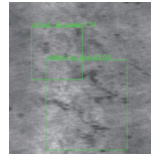
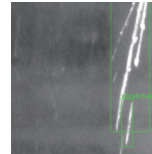
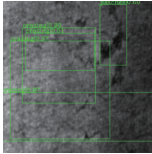
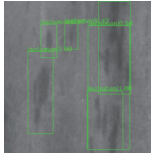
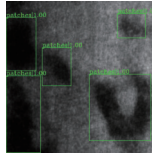
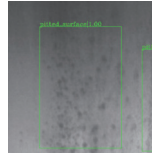
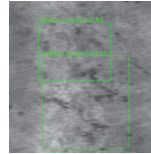
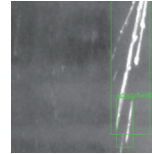
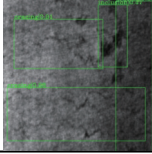
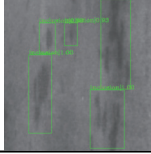
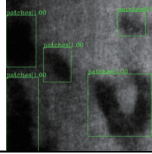
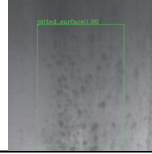
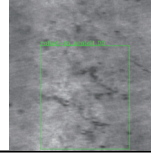
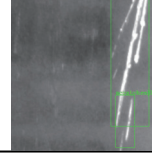
表 1 各个模型在 NEU-DET 数据集的缺陷检测平均精度结果 (%)

Table 1 The average precision results of defect detection for each model in the NEU-DET dataset (%)

方法	平均精度均值	裂纹	夹杂	斑块	麻点	压入氧化	划痕
Faster R-CNN	79.20	71.31	84.63	82.92	80.17	80.31	75.87
RetinaNet	75.36	53.02	78.74	93.33	91.37	62.21	73.49
FCOS	75.18	52.41	75.03	91.48	84.85	62.86	84.43
Grid R-CNN	73.14	41.52	78.68	86.23	86.47	59.74	86.21
YOLO-v1	62.90	42.35	63.42	68.23	66.49	69.37	67.53
YOLO-v2	66.53	47.35	70.47	72.23	65.82	65.49	77.84
YOLO-v3	69.40	68.39	61.88	71.44	68.33	72.66	73.71
YOLO-v4	77.99	64.87	70.84	93.24	83.83	69.52	85.63
YOLO-v5	76.82	62.42	75.76	84.23	81.27	64.59	92.63
YOLOF	77.32	63.48	71.82	90.56	85.21	64.24	88.63
AGLNet	79.90	54.72	83.31	88.63	91.67	64.42	96.64

表 2 AGLNet、Grid R-CNN and Faster R-CNN 基于 NEU-DET 数据集的对比测试结果

Table 2 Comparison results of AGLNet, Grid R-CNN and Faster R-CNN based on NEU-DET dataset

	裂纹	夹杂	斑块	麻点	压入氧化	划痕
AGLNet						
Grid R-CNN						
Faster R-CNN						

秒浮点运算次数 (Floating point operations per second, FLOPs)、模型参数量 (Parameters, Params) 和模型每秒处理图片量 (FPS) 与其他检测器进行比较. FLOPs 表示所需的计算能力, Params 参数大小描述所需的内存. 如表 3 所示, AGLNet 的 FLOPs 低于其他两款双阶段的模型 Faster R-CNN 和 Grid R-CNN, 在检测的精度上高于这两款模型. 这表明 AGLNet 在提高计算精度的同时, 降低了模型的计算量. 相较基于 Anchor-free 思想的 RetinaNet, FCOS 和 Grid R-CNN, AGLNet 的 FLOPs 高于 RetinaNet, 低于 FCOS 和 Grid R-CNN. 相较追求速度的 YOLO 系列的算法, AGLNet 在 FLOPs 方面是不足的, 但是精度高于两者的精度. 综上, 在追求检测精度和检测速度的前提下, AGLNet 保持着最高的检测精度, 同时 AGLNet 也有较高的 FPS, 表明 AGLNet 模型在检测速度上也较为优

秀, 能够满足实际生产中实时缺陷检测的速度要求.

2.2.4 模型评估

为了全面评估模型的性能, 分析模型尚存在的不足, 本文结合上述实验结果进行更充分的分析. 根据表 1, 以裂纹类型的缺陷为例, RetinaNet, FCOS, Grid R-CNN 以及 AGLNet 对其的检测精度分别是 53.02%, 52.41%, 41.52% 以及 54.72%, 检测效果都不太理想. 根据 AP 的计算式 (29) 可得出, 高水平的 AP 需要精度和召回率同时保持较高水平才能实现, 因此本文进行如下实验, 分别将 IoU 的阈值设置为 0.5 和 0.75, 探究 IoU 阈值设置对模型的检测结果、召回率、平均检测精度的影响, 实验结果如表 4 所示. 其中, gts 为测试集中真实存在的全部缺陷数量, Dets 表示检测出来的缺陷总数, Recall 代表检测出来的缺陷比例, mAP 表示检测出来的缺陷的平均精度.

表 3 各模型 FLOPs, Params 和 FPS 对比结果
Table 3 Comparison of FLOPs, Params and FPS of each model

方法	FLOPs (GMAC)	Params (M)	FPS (帧/s)
Faster R-CNN	408.36	98.25	~8.2
RetinaNet	239.32	37.74	~12.3
FCOS	438.68	89.79	~9.3
Grid R-CNN	329.51	64.32	~10.2
YOLO-v3	89.45	27.84	~15.4
YOLOF	151.47	63.24	~13.4
AGLNet	273.95	79.80	~1.8

表 4 各类缺陷在不同 IoU 阈值下的测试结果
Table 4 Detection results of various defects under different IoU thresholds

IoU 阈值	缺陷类型	gts	Dets	Recall	mAP
IoU0.5	裂纹	139	1 886	0.935	54.72
IoU0.75	裂纹	139	1 823	0.923	47.48
IoU0.5	夹杂	181	1 188	0.945	83.31
IoU0.75	夹杂	181	1 163	0.932	82.17
IoU0.5	斑块	151	627	0.960	88.63
IoU0.75	斑块	151	591	0.942	89.45
IoU0.5	麻点	88	689	0.955	91.67
IoU0.75	麻点	88	636	0.938	89.24
IoU0.5	压入氧化	126	1 034	0.893	64.42
IoU0.75	压入氧化	126	1 051	0.882	59.66
IoU0.5	划痕	117	317	0.991	96.64
IoU0.75	划痕	117	322	0.986	92.79
IoU0.5	全部缺陷	802	5 741	0.947	79.90
IoU0.75	全部缺陷	802	5 586	0.934	76.79

实验结果表明, 在使用 AGLNet 测试 NEU-DET 数据集时, 提高了 IoU 的阈值设置, 会显著降低平均检测精度的水平, 共降低了 3.11% 的平均检测精度; 但是对召回率的影响不大, 仅降低了 1.3%. 在 IoU0.5 的情况下, 平均召回率达到了 0.947, 裂纹和压入氧化两个缺陷的召回率也分别达到了 0.935 和 0.893; 在 IoU0.75 的情况下, 平均召回率达到了 0.934, 裂纹和压入氧化两个缺陷的召回率也分别达到了 0.923 和 0.882. 上述情况表明, AGLNet 对目标检测的能力是极佳的, 可以识别出原图中绝大多数的缺陷目标. 综上, 提高 IoU 的阈值对召回率影响不大, 但是对精确度影响显著, 说明已识别出来的高质量正样本是足够的. 提高阈值, 去除低质量正样本后, TP 仍然可以保持较高的水平.

同时, 在 IoU0.5 的情况下, 6 种缺陷的召回率都相对较高, 但裂纹和压入氧化的 mAP 却很低. 通过统计, 裂纹数据集中, 真正的缺陷总数为 139 个,

但是 AGLNet 检测出来 1 886 个缺陷目标, 负样本数远远超过正样本的数量, 导致计算的精确度较低, 这是最终没有获得理想结果的主要原因.

经过具体分析检测的结果图片, 裂纹和压入氧化等缺陷有如下几个特点, 从而加大了图片的检测难度: 1) 缺陷图片的分辨率差; 2) 缺陷的空间布局密集; 3) 真实标定框内有多处非缺陷的正常区域; 4) 缺陷呈线型不规则的结构.

图 6 中, 以两个缺陷的图片为例, 左列为真实标定框的位置图片, 右列为 AGLNet 检测出来的区域. 经过对比, 在原始标注的图中, 由于缺陷过于不规则, 人为标注的真实标定框的区域其实包含了一部分缺陷区域和一部分正常区域. 在检测结果图中, AGLNet 算法仅将缺陷区域识别出来, 而将正常区域去除, 导致输出了多个且较小的检测目标框. 所以 AGLNet 凭借其对图片的高敏感度, 仅将图中缺陷的位置检测出并输出, 并没有输出真实标定框中正常区域. 虽然 AGLNet 能够更精确地检测出缺陷的位置, 但由于与真实标定框的位置不相符, 从而导致在最终计算检测结果与真实标定框的重叠率时, 结果变差, 最终的 mAP 结果不太理想.

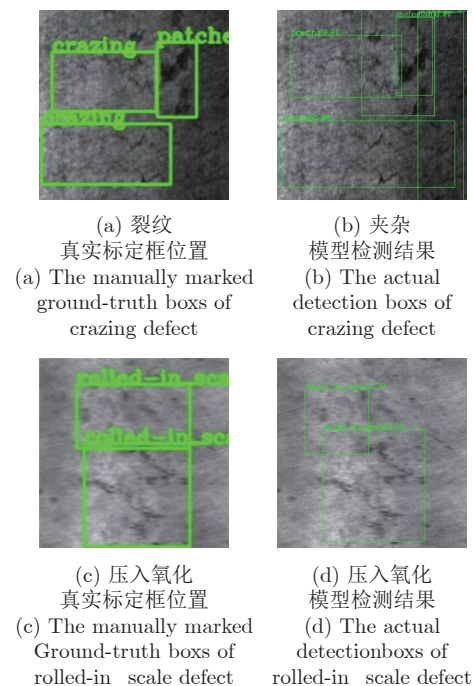


图 6 AGLNet 模型下裂纹和压入氧化缺陷检测结果与人工标注位置对比

Fig. 6 Comparison between inspection results of crazing and rolled-in_scale defects under AGLNet model and manually marked positions

2.2.5 消融实验

为了验证创新模块的有效性, 在仍然采取全局

表 5 消融实验结果
Table 5 Results of ablation experiments

序号	ResNet50_FPN	ResNet50	AT-RPN	RPN	mAP (%)	FPS	GPU 存贮占用量 (MiB)
1	✓		✓		79.90	11.8	5568
2	✓			✓	78.64	10.3	7039
3		✓	✓		77.97	12.2	5024
4		✓		✓	76.82	10.6	6436

定位回归算法的基础上, 实施消融实验, 验证其他网络模块对 AGLNet 检测效果的提升. 结果如表 5 所示, 使用 ResNet50_FPN 的模型和 AT-RPN 达到了最高的 mAP, 且速度相对较高. 同时, 对上述四次实验的 mAP 和 FPS 进行了对比评估, 对比结果如表 6 所示. 对比实验 1 和实验 2 以及对比实验 3 和实验 4 表明, AT-RPN 模块可以有效地提高 AGLNet 的检测精度和速度. 对比实验 1 和实验 3 以及对比实验 2 和实验 4 表明, FPN 加入后会提高网络检测的精度, 同时影响检测的速度. 经过测试, 由 FPN 模块和 AT-RPN 模块组成的模型比原始模型平均精度均值提高了 3.08%, FPS 提高了 1.2. 结果表明, 在提高检测精度的同时, 提高了检测速度.

表 6 消融实验对比结果

序号	对比实验	mAP 提升 (%)	FPS 提升	节约显存占用率 (%)
1	实验 1/实验 2	1.26	1.5	20.89
2	实验 3/实验 4	1.15	1.6	21.93
3	实验 1/实验 3	1.93	-0.4	-10.82
4	实验 2/实验 4	1.82	-0.3	-9.36
5	实验 1/实验 4	3.08	1.2	13.49

为了更全面地检测 AT-RPN 的效果, 本文对 AT-RPN 算法、RPN 算法和 AABO 算法分类以及回归损失函数的变化进行了对比. AABO 优化算法同样是一种基于贝叶斯思想的 Anchor 优化器, 通过 TPE 和 Bandit-based 算法的结合以及优化锚框参数选择, 获取到更好的检测结果, 提高了检测速度. 对比结果如图 7 和图 8 所示. 三次实验网络均采用 ResNet50_FPN 作为特征提取器, 网络第二阶段均采用全局定位回归作为位置回归方式. 红线表示 AT-RPN 算法中函数的损失值, 绿线表示 RPN 算法中函数的损失值, 蓝线表示 AABO 算法中函数的损失值. 在图 7 和图 8 中, AT-RPN 和 AABO 的分类损失函数和位置回归函数的损失值下降得更快, 并且均比传统 RPN 算法中的损失函数更早达到稳定值. AABO 与 AT-RPN 相比较, AT-RPN

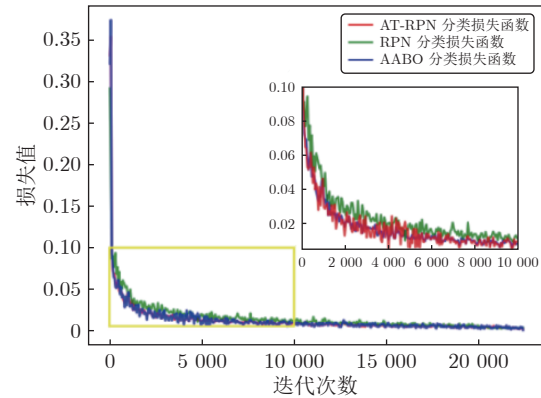


图 7 AT-RPN、RPN 和 AABO 的分类损失函数变化对比

Fig. 7 The change of classification loss function of AT-RPN, RPN and AABO

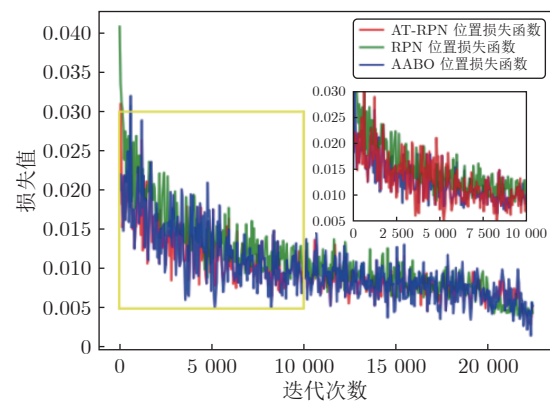


图 8 AT-RPN、RPN 和 AABO 的位置回归损失函数变化对比

Fig. 8 The change of location regression loss function of AT-RPN, RPN and AABO

的损失函数下降的过程更加稳定. 另外, 根据表 6 消融实验对比结果可知, 使用 AT-RPN 后, 可以节省 20% 以上对电脑显存的占用, 大大降低了缺陷检测算法对电脑 GPU 的占用率, 更适用于实际的工业环境中. 所以, 损失函数对比实验和消融实验对比结果表明, AT-RPN 可以大大提升网络的检测精度和网络检测速度, 并减少深度学习算法对计算机显存占用, 对实际工业生产应用有较大的帮助.

2.3 印刷电路板缺陷检测结果分析

为了更加全面地验证 AGLNet 模型的性能, 本文采用新的数据集进行测试评估. 印刷电路板 (Printed circuit board, PCB) 是各种电子元器件的载体, 在电子通讯等多个领域有着广泛的应用. 随着技术的不断发展, 电子产品向更轻、更薄转变, PCB 也朝着更薄更小的方向迈进. 传统人工检测模式逐渐不适用于目前的质检过程, 所以人们开始着力于研究使用人工智能检测的方法取代传统人工检测的模式, 以达到 PCB 缺陷检测速度更快、精度更高的目的.

扩展实验在 PCB-Master^[28] 数据集上进行 (<http://robotics.pkusz.edu.cn/resources/dataset/>). 该数据集每张图片的平均像素大小为 $2\ 777 \times 2\ 138$. PCB 缺陷类型一共有 6 种, 分为漏孔 (Missing_hole)、鼠咬 (Mouse_bite)、断路 (Open_circuit)、短路 (Short)、毛刺 (Spur) 和余铜 (Spurious_copper). 一张 PCB 的图片中包含多种且多个缺陷, 缺陷的统计结果如表 7 所示, 缺陷的高宽比如图 9 所示. 随机划分训练集和测试集比例为 7 : 3, 即训练集为 485 张图片, 测试集为 207 张图片. 由于印刷电路板缺陷非常细小、缺陷微弱, 所以一般用高分辨率的图像进行记录.

表 7 PCB-Master 数据集基本信息
Table 7 Basic information of PCB-Master dataset

缺陷类型	图像数量	缺陷数量
漏孔	115	497
鼠咬	115	492
断路	115	482
短路	115	491
毛刺	115	488
余铜	116	503
全部缺陷总计	693	2 953

表 8 各个模型在 PCB-Master 数据集上测试结果
Table 8 Test results of each model on PCB-Master dataset

	Faster R-CNN	RetinaNet	FCOS	Grid R-CNN	YOLO-v3	YOLOF	AGLNet
AP (漏孔) (%)	87.43	91.54	90.73	95.55	85.83	94.22	99.45
AP (鼠咬) (%)	84.90	90.50	85.24	93.37	79.25	93.35	95.17
AP (断路) (%)	86.15	89.65	84.74	91.45	74.73	88.63	92.93
AP (短路) (%)	89.45	92.16	92.83	99.70	83.23	99.70	99.70
AP (毛刺) (%)	86.91	95.26	91.50	95.36	82.62	98.86	99.65
AP (余铜) (%)	86.53	87.48	88.03	90.48	73.10	95.39	94.22
mAP (%)	86.90	91.10	88.85	94.32	79.79	95.03	96.85
FPS (帧/s)	~4.20	~6.67	~5.41	~5.88	~9.52	~7.69	~6.25

为了与之前的实验形成对比, 本文同样选取了 Faster R-CNN, RetinaNet, FCOS, Grid R-CNN, YOLO-v3, YOLOF 和 AGLNet 七种算法对 PCB-Master 数据集进行训练和测试, 测试结果如表 8 所示. 相较于基于 Anchor-free 思想的 RetinaNet, FCOS, Grid R-CNN 检测模型, AGLNet 在 PCB-Master 数据集上的 mAP 分别提高了 5.75%, 8.00%, 2.53% 的精度. 相较于经典的双阶段的 Faster R-CNN 和单阶段的 YOLO-v3 检测网络, AGLNet 在 PCB-Master 数据集上的 mAP 分别提高了 9.95% 和 17.06% 的精度. 相较于 2021 年最新的 YOLOF 检测模型, AGLNet 在该数据集上的检测平均精度高出 1.82%. 从单一缺陷角度分析, AGLNet 在漏孔、鼠咬、断路、短路、毛刺缺陷上均有较高的检测精度, 在短路和毛刺两个缺陷检测效果最为显著, 高达 99.70% 和 99.65%. 在余铜缺陷的检测中, 略逊色于 YOLOF, 但仍然保持着 94.22% 的检测精度. 评估 AGLNet 在 PCB-Master 数据集上的检测速度, 表 8 展示了各模型之间 FPS 的比

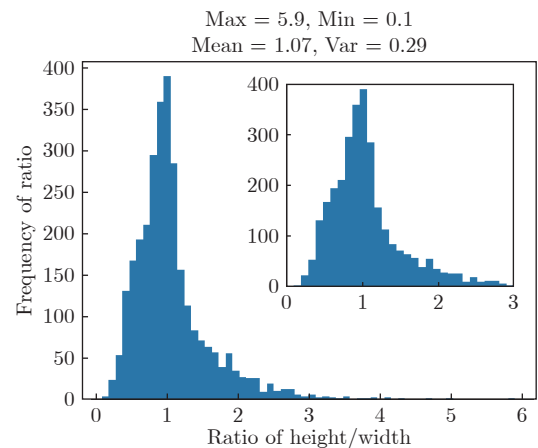


图 9 PCB-Master 数据集集中的高宽比统计结果
Fig.9 Statistical results of aspect ratio in PCB-Master dataset

较. 如表 8 所示, AGLNet 的检测速度约为 6.25 帧/s, 高于 Faster R-CNN, Grid R-CNN, FCOS 模型, 略低于 RetinaNet 和 YOLOF 模型, 低于 YOLO-v3 模型.

为了评估 AGLNet 在 PCB-Master 数据集上各种缺陷中 AP 与召回率的情况, 表 9 展示了模型在测试集上各类缺陷的检测结果. gts 为缺陷的数量, Dets 为模型检测出的缺陷数量, Recall 为模型在测试集上的召回率. 如表 9 可知, AGLNet 的召回率在各个缺陷上都取得了很高的结果, 可以完全识别所有的断路和毛刺缺陷. PCB 缺陷检测结果如图 10 所示.

表 9 PCB-Master 测试集检测数据统计
Table 9 Data statistics of PCB-Master defect detection test set

缺陷类别	gts	Dets	Recall	AP
漏孔	169	696	0.998	0.995
鼠咬	142	665	0.990	0.952
断路	142	667	0.990	0.929
短路	132	590	1.000	0.997
毛刺	143	687	1.000	0.997
余铜	137	644	0.979	0.942
全部缺陷总计	865	3949		

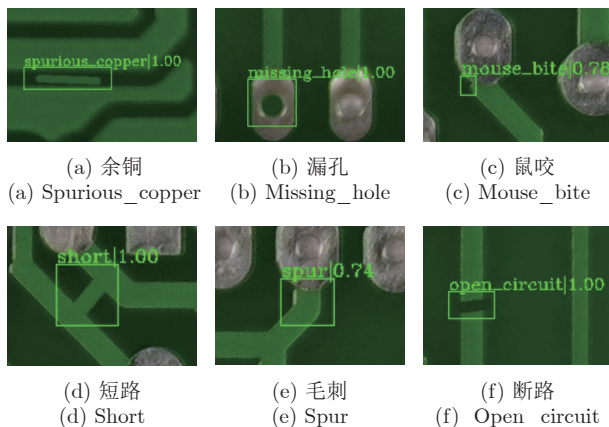


图 10 PCB-Master 检测结果
Fig. 10 PCB-Master test results

综合比较, YOLOF, RetinaNet 和 YOLO-v3 模型虽然检测速度高于本文的 AGLNet 模型, 但是 AGLNet 的检测精度却远高于这三个模型. 此外, AGLNet 模型在速度和精度上都高于 Faster R-CNN, Grid R-CNN, FCOS 模型. 综上所述, 综合比较各类模型在 PCB-Master 数据集上的检测结果, 本文可以得知 AGLNet 在检测精度上展现了最好的效果, 远高于其他模型; 同时, 在模型检测的速

度上也保持着不错的结果. 所以, AGLNet 凭借其对于图片的敏感性, 可以对细小的缺陷实现高精度的检测, 满足了印刷电路板质检的基本要求, 是一个相对有价值的模型.

3 结束语

针对热轧钢带表面缺陷检测精度低、速度慢、智能化低等问题, 本文提出了 AGLNet, 采用结合后的残差网络和特征金字塔网络作为特征提取结构, 提取目标缺陷的高维度到低维度特征, 并使用本文创新提出的 AT-RPN 自适应候选框提取算法模块, 无需先验知识的测试积累, 抛弃人为调参的模式, 最后使用本文引入的全局定位回归算法改变过去包围框回归的传统思路, 采用全局定位回归的新模式获取复杂缺陷更精确的定位, 以及分类损失函数获取目标的类别信息, 从而实现一种快速、准确、更智能化、更适用于实际工业应用的带钢表面缺陷检测的算法.

通过实验可以得到以下结论: 1) 通过结合后的残差网络和特征金字塔网络的特征提取结构, 可以减少层级之间传递导致的语义缺失, 有效缓解缺陷形状差异较大、空间位置密集的问题; 2) AT-RPN 自适应候选框提取模块, 在提高了检测精度的同时, 也大大提高了检测的速度; 3) 全局定位回归算法抛弃传统对包围框进行回归的模式, 采用多点监督定位的方法实现了目标缺陷的更精确定位; 4) 与其他经典算法相比, AGLNet 算法通过 NEU-DET 数据集进行验证, 平均精度均值达到对比实验中最优的 79.90%, 同时 FPS 达到优异的 11.8 帧/s; 5) AGLNet 具有较强的泛化能力, 在 PCB-Master 印刷电路板缺陷数据集的检测中, 平均精度为各算法中最优的 96.85%, FPS 约为 6.25 帧/s, 实现了快速、精确的智能化的缺陷检测.

References

- 1 Wang Dian-Hong, Gan Sheng-Feng, Zhang Wei-Min, Lei Wei-Xin. Strip surface defect image classification based on double-limited and supervised-connect Isomap algorithm. *Acta Automatica Sinica*, 2014, **40**(5): 883-891 (王典洪, 甘胜丰, 张伟民, 雷维新. 基于监督双限制连接 Isomap 算法的带钢表面缺陷图像分类方法. *自动化学报*, 2014, **40**(5): 883-891)
- 2 Song K C, Yan Y H. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Applied Surface Science*, 2013, **285**: 858-864
- 3 Neogi N, Mohanta D K, Dutta P K. Review of vision-based steel surface inspection systems. *EURASIP Journal on Image and Video Processing*, 2014, **2014**(1): Article No. 50
- 4 Xu Zhi-Xiang, Lu Hong, Shen Jian. Camera calibration and its error analysis. *Acta Automatica Sinica*, 1993, **19**(1): 115-117 (许志祥, 卢宏, 沈剑. 摄像机定标及其误差分析. *自动化学报*, 1993, **19**(1): 115-117)

- 5 Li Shao-Bo, Yang Jing, Wang Zheng, Zhu Shu-De, Yang Guan-Ci. Review of development and application of defect detection technology. *Acta Automatica Sinica*, 2020, **46**(11): 2319–2336 (李少波, 杨静, 王铮, 朱书德, 杨观赐. 缺陷检测技术的发展与应用研究综述. *自动化学报*, 2020, **46**(11): 2319–2336)
- 6 Liu Guo-Liang, Yu Jian-Bo. Application of neural-symbol model based on stacked denoising auto-encoders in wafer map defect recognition. *Acta Automatica Sinica*, 2022, **48**(11): 2688–2702 (刘国梁, 余建波. 基于堆叠降噪自编码器的神经-符号模型及在晶圆表面缺陷识别. *自动化学报*, 2022, **48**(11): 2688–2702)
- 7 Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(6): 1137–1149
- 8 He K M, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, **42**(2): 386–397
- 9 Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 779–788
- 10 Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y, et al. SSD: Single shot MultiBox detector. In: Proceedings of the 14th European Conference on Computer Vision (ECCV). Amsterdam, The Netherlands: Springer, 2016. 21–37
- 11 Lin T Y, Dollár P, Girshick R, He K M, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 936–944
- 12 Tao X, Zhang D P, Wang Z H, Liu X L, Zhang H Y, Xu D. Detection of power line insulator defects using aerial images analyzed with convolutional neural networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020, **50**(4): 1486–1498
- 13 He Y, Song K C, Meng Q G, Yan Y H. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. *IEEE Transactions on Instrumentation and Measurement*, 2020, **69**(4): 1493–1504
- 14 Cheng X, Yu J B. RetinaNet with difference channel attention and adaptively spatial feature fusion for steel surface defect detection. *IEEE Transactions on Instrumentation and Measurement*, 2021, **70**: Article No. 2503911
- 15 Chen J W, Liu Z G, Wang H R, Núñez A, Han Z W. Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network. *IEEE Transactions on Instrumentation and Measurement*, 2018, **67**(2): 257–269
- 16 Zhang C B, Chang C C, Jamshidi M. Concrete bridge surface damage detection using a single-stage detector. *Computer-Aided Civil and Infrastructure Engineering*, 2020, **35**(4): 389–409
- 17 Zhang S F, Chi C, Yao Y Q, Lei Z, Li S Z. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020. 9756–9765
- 18 Law H, Deng J. CornerNet: Detecting objects as paired keypoints. *International Journal of Computer Vision*, 2020, **128**(3): 642–656
- 19 Duan K W, Bai S, Xie L X, Qi H G, Huang Q M, Tian Q. CenterNet: Keypoint triplets for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea: IEEE, 2019. 6568–6577
- 20 Jia X G, Yang X Q, Yu X H, Gao H J. A modified CenterNet for crack detection of sanitary ceramics. In: Proceedings of the 46th Annual Conference of the IEEE Industrial Electronics Society (IECON). Singapore: IEEE, 2020. 5311–5316
- 21 Zhu C C, He Y H, Savvides M. Feature selective anchor-free module for single-shot object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 840–849
- 22 Tian Z, Shen C H, Chen H, He T. FCOS: Fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea: IEEE, 2019. 9626–9635
- 23 Kong T, Sun F C, Liu H P, Jiang Y N, Li L, Shi J B. FoveaBox: Beyond anchor-based object detection. *IEEE Transactions on Image Processing*, 2020, **29**: 7389–7398
- 24 Lu X, Li B Y, Yue Y X, Li Q Q, Yan J J. Grid R-CNN. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 7355–7364
- 25 Wang J Q, Zhang W W, Cao Y H, Chen K, Pang J M, Gong T, et al. Side-aware boundary localization for more precise object detection. In: Proceedings of the 16th European Conference on Computer Vision (ECCV). Glasgow, UK: Springer, 2020. 403–419
- 26 Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyperparameter optimization. In: Proceedings of the 24th International Conference on Neural Information Processing Systems (NIPS). Granada, Spain: Curran Associates Inc., 2011. 2546–2554
- 27 Cao J L, Cholakkal H, Anwer R M, Khan F S, Pang Y W, Shao L. D2Det: Towards high quality object detection and instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020. 11482–11491
- 28 Ding R W, Dai L H, Li G P, Liu H. TDD-net: A tiny defect detection network for printed circuit boards. *CAAI Transactions on Intelligence Technology*, 2019, **4**(2): 110–116



王延舒 同济大学机械与能源工程学院硕士研究生。2020年获四川大学学士学位。主要研究方向为机器学习, 深度学习, 视觉检测与识别。

E-mail: 2030211@tongji.edu.cn

(**WANG Yan-Shu** Master student at the School of Mechanical Engineering, Tongji University. He received his bachelor degree from Sichuan University in 2020. His research interest covers machine learning, deep learning, and visual inspection and identification.)



余建波 同济大学机械与能源工程学院教授。2009年获上海交通大学博士学位。主要研究方向为机器学习, 深度学习, 智能质量管控, 过程控制, 视觉检测与识别。本文通信作者。

E-mail: jbyu@tongji.edu.cn

(**YU Jian-Bo** Professor at the School of Mechanical Engineering, Tongji University. He received his Ph.D. degree from Shanghai Jiao Tong University in 2009. His research interest covers machine learning, deep learning, intelligent quality control, process control, and visual inspection and identification. Corresponding author of this paper.)