计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(10):122−133 [doi: 10.15888/j.cnki.csa.008739]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 STM R-CNN 的热轧带钢表面缺陷检测①

于　波[1], 张新凯[1,2], 王　卫[3]

[1](中国科学院 沈阳计算技术研究所, 沈阳 110168)
[2](中国科学院大学, 北京 100049)
[3](东北大学 计算机科学与工程学院, 沈阳 110169)
通信作者: 张新凯, E-mail: zhangxinkai19@mails.ucas.ac.cn

摘　要: 为了提高工业热轧带钢表面缺陷检测的检测精度, 将深度学习研究领域的前沿技术应用于带钢表面缺陷检测. 提出了一种以 Swin Transformer 作为骨干特征提取网络, 级联多阈值结构作为输出层的热轧带钢表面缺陷检测算法. 将 Transformer 结构应用于带钢表面缺陷检测领域, 与单纯基于卷积网络的深度学习目标检测算法相比, 能够达到更加精确的检测效果. 首先, 使用 Swin Transformer 作为骨干特征提取网络代替常规的残差网络结构, 增强特征网络对隐含在图像中的深层语义信息的摄取能力. 其次设计多级联检测结构, 设置逐级的 IoU 阈值, 实现检测精度与阈值提升的权衡. 最后使用柔性非极大值抑制 (Soft-NMS)、FP16 混合精度训练和 SGD 优化器等训练策略加速模型收敛和提升模型性能. 实验结果表明: 本文算法在工业热轧带钢数据集 (NEU-DET) 上相较于 YOLOv3、YOLOF、DeformDetr、SSD512 和 SSDLit 等深度学习算法都有更好的检测效果, 在裂纹 (crazing, Cr)、夹杂 (inclusion, In)、斑块 (patches, Pa)、麻点 (pitted surface, PS)、压入氧化铁皮 (rolled-inscale, RS)、以及划痕 (scratches, Sc) 等表面缺陷检测中训练速度和检测精度都有显著的提升, 漏检率显著降低.

关键词: 深度学习; 缺陷检测; Swin Transformer; NEU-DET; 热轧带钢; 机器视觉

引用格式: 于波,张新凯,王卫.基于 STM R-CNN 的热轧带钢表面缺陷检测.计算机系统应用,2022,31(10):122–133. http://www.c-s-a.org.cn/1003-3254/8739.html

Surface Defect Detection of Hot-rolled Strip Steel Based on STM R-CNN

YU Bo[1], ZHANG Xin-Kai[1,2], WANG Wei[3]

[1](Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)
[2](University of Chinese Academy of Sciences, Beijing 100049, China)
[3](School of Computer Science and Engineering, Northeastern University, Shenyang 110169, China)

Abstract: The cutting-edge technology in deep learning is applied to surface defect detection of strip steel for the accuracy improvement in surface defect detection of industrial hot-rolled strip steel. Therefore, a surface defect detection algorithm for hot-rolled strip steel is proposed, which takes Swin Transformer as the backbone feature extraction network and cascaded multi-threshold structure as the output layer. Compared with the deep learning target detection algorithm based solely on convolutional networks, the detection algorithm using the Transformer structure can achieve more accurate detection results. Specifically, first, Swin Transformer is used as the backbone feature extraction network to replace the conventional residual network structure and thus enhance the ability of the feature network to capture the deep semantic information implicit in an image. Secondly, a multi-cascade detection structure is designed, and step-by-step IoU thresholds are set to achieve the balance between detection accuracy and threshold improvement. Finally, training strategies such as soft non-maximum suppression (Soft-NMS), FP16 mixed precision training, and SGD optimizers are employed to accelerate model convergence and improve model performance. The experimental results reveal that the

① 基金项目: 辽宁省“兴辽英才计划”项目 (XLYC1907001)
收稿时间: 2022-01-17; 修改时间: 2022-02-15, 2022-02-24; 采用时间: 2022-03-01; csa 在线出版时间: 2022-07-07

proposed algorithm has better detection performance on the industrial hot-rolled strip steel data set (NEU-DET) than the deep learning algorithms such as YOLOv3, YOLOF, DeformDetr, SSD512, and SSDLit. Additionally, the training speed and detection accuracy are significantly improved in the surface defect detection of crazing (Cr), inclusion (In), patches (Pa), pitted surface (PS), polled-in scales (RS), scratches (Sc), and other surface defects, and the missed detection rate is greatly reduced.

Key words: deep learning; defect detection; Swin Transformer; NEU-DET; hot rolled strip steel; machine vision

1 引言

工业热轧带钢[1]是指厚度在 0.1–2 cm、宽度一般为 60–200 cm 的成卷带钢, 作为原材料被广泛应用于汽车、造船、电工设备、工程器械等设备的生产制造场景中, 但是由于生产工艺的限制, 在热轧生产过程中带钢表面会产生裂纹 (crazing, Cr)、夹杂 (inclusion, In)、斑块 (patches, Pa)、麻点 (pitted surface, PS)、压入氧化铁皮 (rolled-inscale, RS)、以及划痕 (scratches, Sc) 等表面缺陷[2,3]. 这些缺陷对带钢产品的外观质量、疲劳强度和抗腐蚀性等产生较大的影响. 带钢表面缺陷检测是热轧带钢生产质量检测中的重要环节. 传统检测方法[4]有人工检测、磁粉检测法、渗透检测法、涡流检测、X 射线检测以及超声波检测技术、机器视觉检测法等. 热轧带钢表面缺陷大小多在 0.1–5 cm 可视范围, 密集度较大, 分布不均匀, 类型为不规则点状、凸包或车条状, 颜色为灰白色、深灰色等, 缺陷类型差异在于色彩和形状差异. 不适用于超声波、射线等检测方法, 机器视觉检测方法具有安装简易、成本低廉、检测精度高等优点, 通常作为首选方案.

近些年随着机器学习技术的快速发展. 基于深度学习的图像识别检测技术成为表面检测任务的主要方法. 基于深度学习的目标检测算法从结构上可以分为一阶段检测算法和两阶段检测算法, 分别的代表是 SSD[5]、YOLO[6–8] 和 Faster R-CNN[9]、Mask R-CNN[10] 等. 一阶段检测算法不需要 RPN 阶段, 直接得到检测结果检测速度较快, 但检测精度较低[11]. 两阶段检测算法将检测任务分为两个阶段, 首先使用区域候选网络 (RPN) 产生候选区域, 然后使用检测网络检测候选区域的类别、位置, 这种方法的准确度较高但检测速度稍慢[12].

基于深度学习的带钢表面缺陷检测普遍使用卷积神经网络作为特征提取并分类[13]. Fu 等[14]提出的一种端到端的卷积神经网络, 实现了带钢表面缺陷的高精度分类; He 等[15]、He 等[16]通过 GAN 网络生成未标注数据进行数据增强, 再进行缺陷分类, 解决了缺陷样本数量较少难以训练的问题. 以上检测方法解决了带钢表面缺陷分类问题, 但没有解决缺陷定位问题. 李维刚等[17]改进 YOLOv3 算法的带钢表面缺陷检测, 是在传统 YOLOv3 的基础上进行的微调; 程婧怡等[18]提出的改进 YOLOv3 进行金属表面缺陷检测, 通过融合浅层与深层特征图, 实现检测性能的提升; 刘亚姣等[19]提出的改进 YOLOv3 算法实现型钢表面缺陷检测, 通过使用可变形卷积替代特征网络部分层和 K-means 聚类方法优化先验框, 兼顾检测的速度和精度. 以上改进 YOLOv3 系列算法实现缺陷的分类与定位, 但是检测精度受限制于速度, 不是很理想. He 等[20]使用 Defect Detection Network 实现了端到端带钢表面缺陷检测, 检测准确率范围在 70%–80%; 常海涛等[21]提出的 Faster R-CNN 在工业 CT 图像缺陷检测中的应用, 将两阶段的 Faster R-CNN 应用于经过处理后的缺陷图像检测领域, 检测效果较为理想. 上述方法在目标检测方面已经取得不错的成绩, 检测效果也基本能够满足应用需求. 随着技术的更新迭代, 更先进的算法不断地被设计出, 相比传统的全卷积深度学习方法, 新的方法在性能和精度上比传统的技术更优秀.

本文目的是将深度学习领域前沿算法应用于传统的工业热轧带钢表面缺陷检测中. 热轧带钢表面常见瑕疵缺陷种类繁多, 缺陷形状差异较大且特征相似, 面积大小差异较大. 在数据集数量较少的情况下, 难以实现算法的快速收敛, 不易于精确检测. 针对存在的问题, 本文主要工作如下:

(1) 使用 Swin Transformer[22]作为骨干特征提取网络代替常规的残差网络结构, 增强特征网络对隐含在图像中的深层语义信息的摄取能力;

(2) 设计 Multi-Stage R-CNN 级联结构, 实现多阈值逐级上升策略, 实现检测精度的提升;

(3) 使用柔性非极大值抑制 (Soft-NMS)、FP16 混合精度训练和 SGD 优化器等训练策略加速模型收敛、减少训练时间和提升模型性能;

2 相关技术

2.1 Swin Transformer 架构图

2021 年微软亚洲研究中心首次提出基于 Transformer[23] 的骨干架构 (Swin Transformer), 同年 10 月, 获 CVPR2021 最佳论文奖. 与以往所了解的基于卷积运算的骨干架构 (VggNet、ResNet、DenseNet 和 Darknet) 不同, 它是一种新的 Vision Transformer 结构[24], 其最基本的计算模块不再是传统的卷积模块, 而是采用 Transformer 结构结合图像处理的设计思想实现, 代替传统的骨干网络用于视觉任务, 从输入图像或者视频中提取特征.

如图 1, 首先通分片模块将输入的彩色图像分片成不重叠的 patch. 每个 patch 被视为一个“token”, 其特征被设置为原始像素 RGB 值的串联. 使用了 4×4 的窗口大小, 因此每个窗口的特征维数为 4×4×3=48. 在这个原始特征上应用线性嵌入层, 将其投影到 C (C=96) 维. 将线性的图片块送入 Swin Transformer Block, 如图 2, 将输出的图像进行图片块融合, 经过 Patch Merging 模块后长宽减半, 深度增加一倍, 实现了视野的扩大, 依次循环执行 Swin Transformer Block 后进行融合, 获取更大的视野. 阶段 1–4 分别输出特征图, 作为下一阶段的特征图提取的输入. 如图 2 所示, Swin Transformer Block 结构主要是 Layer Norm, MLP[25], Window multi-head self-attention (W-MSA) 和 shifted Window multi-head self-attention (SW-MSA) 组成, 与经典的 Transformer Block 结构的主要区别在于 W-MSA 和 SW-MSA 模块.

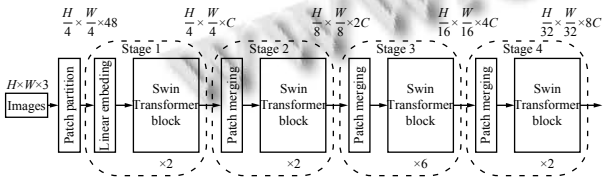


图 1 Swin Transformer 骨干网络架构图

2.2 区域预选网络 (RPN)

区域预选网络, 主要用于从输入的特征图生成可能存在目标的候选框区域. 以全卷积网络为基础, 其与后面目标检测部分的卷积神经网络参数共享.

Anchors generator 生成锚框为 0.5、1.0、2.0 三种, cls_logits 和 box_pred_0 结合锚框得到预测框, 将预测框与目标进行损失计算, 在训练时通过反向传播调整卷积参数.

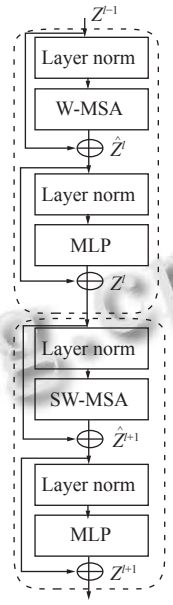


图 2 Swin Transformer Block 结构图

将 box_pred_0 传送到下一级 R-CNN 网络用于进一步的参数训练.

如图 3, 来自特征提取网络的特征图进入 RPN 网络, 先经过一个 3×3 卷积, 再分别经过一个 1×1 卷积, 分别产生类别划分和边界框预测, 分类预测主要用来二分类预测前景和背景, 边界框作为 multi-stage R-CNN 级联网络的输入. 同时预测的结果会结合 anchors generator 生成区域预选框和 label 进行损失计算.

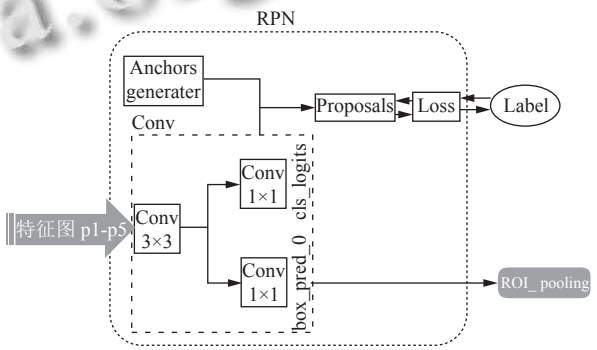


图 3 区域建议框网络

RPN 和 R-CNN 网络损失函数为:

$$L(\{p_i\},\{t_i\})=\frac{1}{N_{\text{cls}}}\sum_i L_{\text{cls}}(p_i,p_i^*)+\lambda\frac{1}{N_{\text{reg}}}\sum_i p_i^* L_{\text{reg}}\left(t_i,t_i^*\right) \tag{1}$$

其中, i表示锚框 index, p_i表示第i个锚框为真实标签的

预测概率, p_i^*代表对应正样本为 1, 负样本为 0, 保证了当锚框为负样本时, 没有边界框回归损失. t_i代表预测的第i个锚框的边界框回归值, t_i^*代表的是对应的第i个真实框值, 计算锚框与真实框的偏移量. N_{cls}是最小批次规模, N_{reg}是 anchor location 的数量. L_{cls}为交叉熵损失, L_{reg}为$Smooth_{L_1}$Loss.

如式 (1) 所示, RPN 网络的损失函数为分类交叉熵损失和边界框回归损失, R-CNN 网络中损失函数为多分类交叉熵损失和边界框回归损失.

(1) 二值交叉熵损失公式 (cross entropy loss): RPN 网络中的分类器将候选框分为前景和背景, 是一个二分类问题. 预测结果只有两个, p和$1-p$, 如式 (2):

$$L_{\text{cls}} = -\left[p_i^* \cdot \log(p_i) + \left(1-p_i^*\right) \cdot \log(1-p_i)\right] \tag{2}$$

其中, p_i表示第i个锚框预测为真实标签的概率, p_i^*当为正样本时为 1, 为负样本时为 0;

(2) 在 R-CNN 网络中将会用到多分类交叉熵损失函数:

$$L_{\text{cls}} = \frac{1}{N}\sum_j L_j = -\frac{1}{N}\sum_j \sum_{c=1}^{M} y_{jc} \log\left(p_{jc}\right) (3) \tag{3}$$

其中, M 为类别的数量, y_{jc}是符号函数 (0 或 1), 如果样本j的真实类别等于 c 取 1, 否则取 0; p_{jc}是观测样本j属于类别 c 的预测概率.

(3) 边界框回归损失:

$$L_{\text{reg}}\left(t_i, t_i^*\right) = \sum_i Smooth_{L_1}\left(t_i - t_i^*\right) \tag{4}$$

$$Smooth_{L_1} = \begin{cases} \dfrac{1}{2}x^2, \text{ if } |x| < 1 \\ |x| - \dfrac{1}{2}, \text{ otherwise} \end{cases} \tag{5}$$

$$t_i = \left[t_x, t_y, t_w, t_h\right],\ t_i^* = \left[t_x^*, t_y^*, t_w^*, t_h^*\right] \tag{6}$$

$$t_x = \frac{x - x_a}{w_a},\ t_y = \frac{y - y_a}{h_a},\ t_w = \log\left(\frac{w}{w_a}\right),\ t_h = \log\left(\frac{h}{h_a}\right) \tag{7}$$

$$t_x^* = \frac{x^* - x_a}{w_a},\ t_y^* = \frac{y^* - y_a}{h_a},\ t_w^* = \log\left(\frac{w^*}{w_a}\right),\ t_h^* = \log\left(\frac{h^*}{h_a}\right) \tag{8}$$

其中, x, y, w, h分别表示真实 (标签) 框的坐标位置, x_a, y_a, w_a, h_a分别表示锚框预测框的坐标. 采用式 (5) $Smooth_{L_1}$回归损失函数, 计算预测框与真实框的损失函数.

2.3 检测结构网络 (R-CNN)

如图 4 所示, 经典的 Faster R-CNN 的检测结构网络. 因为是单一 IoU 阈值, 所以无论阈值如何设置, 检测结果都具有高度的对抗性. 如果阈值设置的很高, 那么预测的 bounding box 与真实的 real bounding box 就会包含很多背景, 使得网络很难取得正样本数据. 如果阈值较低, 网络可以获得更多的正样本, 但是其中会包含较多的非真实样本. 所以, 很难通过一个单一的网络模型去实现阈值的设置.

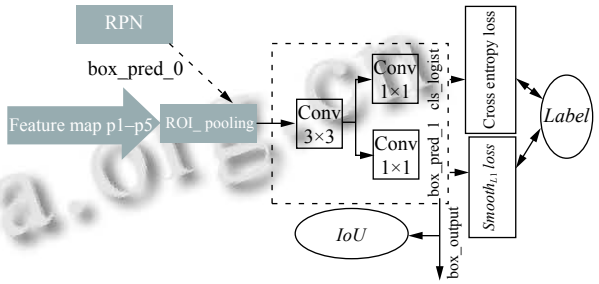


图 4 Faster R-CNN RPN 检测网络

3 本文算法

3.1 算法流程图

为了提升热轧带钢表面缺陷检测的精度, 如图 5、本文在原 Faster R-CNN 网络框架的基础上通过替换残差特征提取网络为 Swin Transformer 特征提取网络, 设计多级阈值检测器 (multi-stage R-CNN), 通过逐级提升的阈值, 提高检测精度. 采用随机垂直、水平翻转、多尺度训练和数值填充等数据增强方式提升训练数据的质量. 采用柔性非极大值抑制 (Soft-NMS)、FP16 混合精度训练和 SGD 优化器等训练策略实现模型加速收敛、减少训练时间.

3.2 Swin Transformer 网络原理

Swin Transformer 是在 Google 提出的 ViT 模型的思路启发下实现的一次巨大提升. Swin Transformer 的设计思路: 直接把图像分成固定大小的图片块, 通过线性变换得到 patch embedding, 类似于自然语言处理中的词嵌入, 将图片序列化输入 Transformer 后进行特征提取分类等操作. 如图 6、图 7 所示, 首先将 200×200×3 pix 的图片切分成 4×4 的图片块, 将图片块展平成线性维度, 再转化为 tokens embedding, 在 tokens embedding 的基础上添加位置 embedding. 将其输入到自定义数量的 Transformer Encoder 模块.

微软亚洲研究中心同时提出不同规模的预训练模型 Swin-T、Swin-S、Swin-B、Swin-L, 本文采用 Swin-T: $C = 96$预训练模型, 是以 ImageNet-1K 数据预训练模型

作为本文算法的骨干网络.

本文实验采用的数据集格式: H=200 pix, W=200 pix.

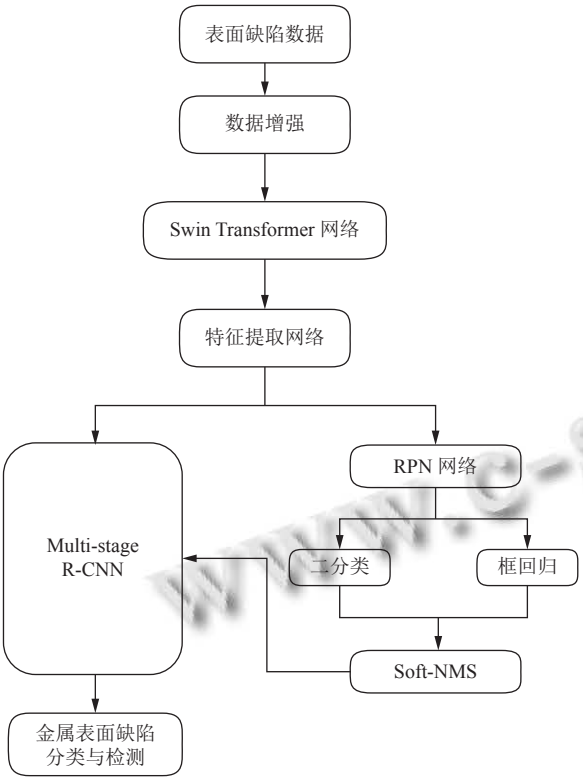


图 5 STM R-CNN 热轧带钢表面缺陷检算法测流程图

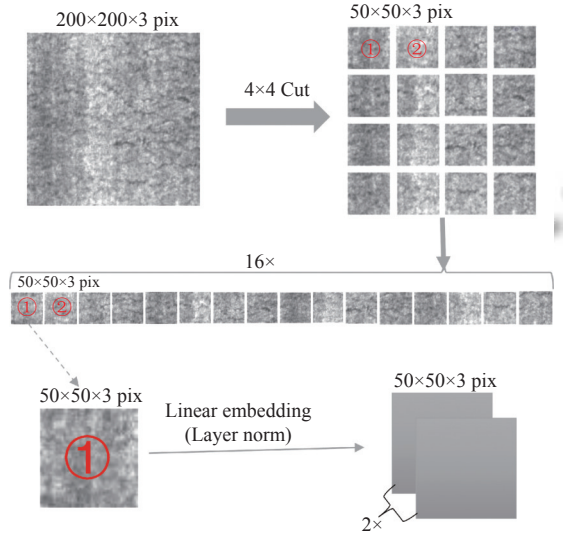


图 6 Swin Transformer 数据处理过程

骨干特征提取网络总共 4 个阶段, 每个阶段输出一张特征图, 输出的特征图尺寸分别为:

$$\frac{H}{4}\times\frac{W}{4}\times C;\ \frac{H}{8}\times\frac{W}{8}\times 2C$$

$$\frac{H}{16}\times\frac{W}{16}\times 4C;\ \frac{H}{32}\times\frac{W}{32}\times 8C$$

Transformer 的基本运算单元是 self-attention 结构.

如式 (9)、图 8 所示, self-attention 运算单元: $a^1, a^2, \cdots, a^n$是序列化数据输入, $b^1, b^2, \cdots, b^n$是不改变维度输出, Q (query), K (key), V (value)是输入序列中编码 W^q、W^k、W^v 是训练参数矩阵, d_k 是编码的维度 (dimension).

$$Attention(Q,K,V)=Softmax\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V \tag{9}$$

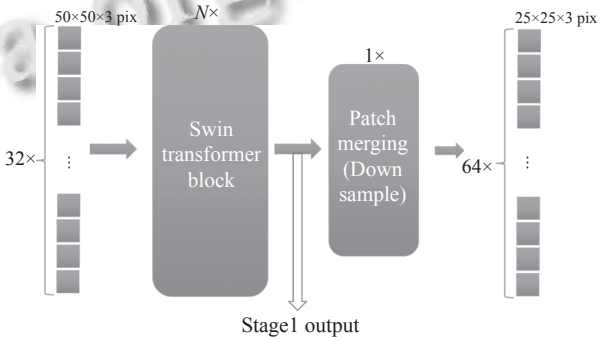


图 7 Swin Transformer 数据处理过程

3.3 多级联检测网络 (multi-stage R-CNN)

为了提升模型的检测能力, 通过构建一个级联架构, 检测器模块的阈值不断提高, 分别是 (0.55, 0.65, 0.75). 通过使用前一阶段的回归输出进行重采样, 一些极端值通过增加 IoU 阈值被移除, 优化深层检测器, 提升整体性能. 如式 (10) 所示, 交并比值 (IoU), 预测框 (bounding box predict) 和真实框 (real bounding box) 的交集和并集的比值计算:

$$IoU=\frac{(A\cap B)}{A\cup B} \tag{10}$$

由于 bounding box 通常包括一个目标和一些背景, 很难确定检测是正样本还是负样本. 所以, 这通常由 IoU 的阈值来决定. 如果 IoU 超过阈值, 则被认为是正样本. 因此, 采用 Soft-NMS 策略:

$$y=\begin{cases} g_y, & IoU(x,y)\geqslant u \\ g_y(1-IoU(x,y)), & IoU(x,y)<u \end{cases} \tag{11}$$

其中, g_y是检测框分数, $IoU(x,y)$为检测框与真实框的交并比, u为设定的阈值.

如图 9, 骨干网络的特征图同时输入到 RPN 和 multi-stage R-CNN 网络部分, 接收来自 RPN 网络的

box_pred_0, 通过级联多个 R-CNN 模块, 每个模块设置不同的 *IoU* 阈值实现 box_pred 的调整. 最终结果, 分类预测为 *n* 个 R-CNN 模块的 cls_logist 的平均值, 边界框预测为最后一个 R-CNN 的输出 box_pred_3.

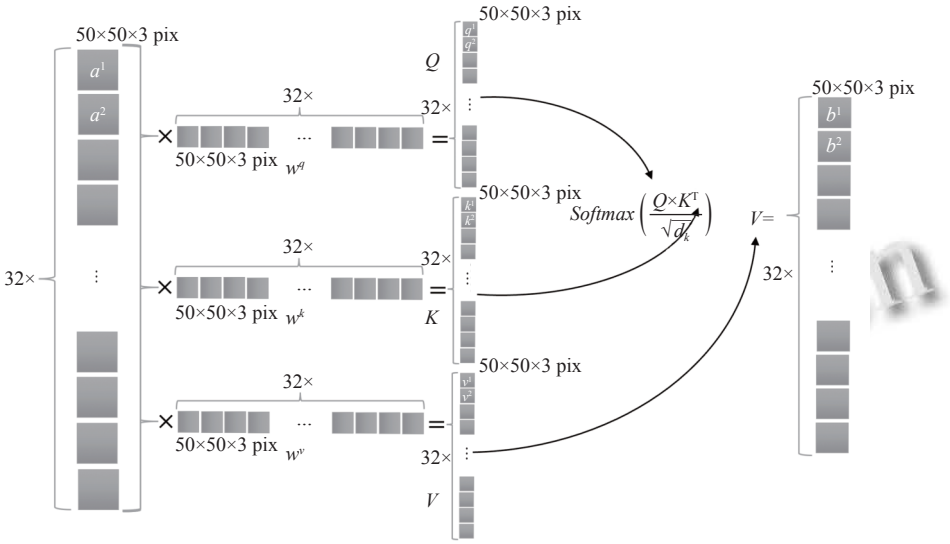


图 8 Self-Attention 运算单元

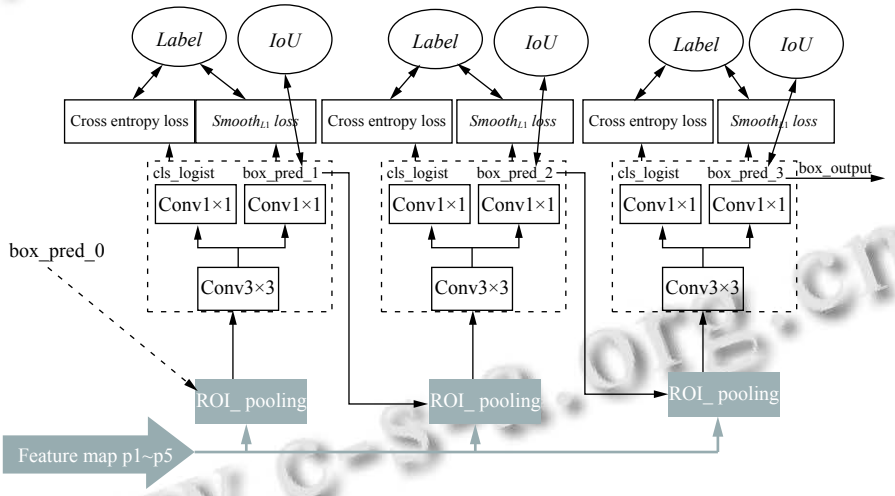


图 9 级联检测网络

4 实验及结果分析

4.1 实验

4.1.1 公开数据集

本文采用公开的热轧带钢 NEU-DET 数据集[20,26,27]作为缺陷检测的训练测试数据, 缺陷分别为裂纹、夹杂、斑块、麻点、压入氧化铁皮、以及划痕等 6 种表面缺陷图像, 每类 300 张, 尺寸为 200×200 pix, 共计 1 800 张缺陷图像. 如图 10 所示.

如图 10 所示, 其中图 10(a) 为 crazing 龟裂纹图像, 裂纹密集数量较多, 颜色较浅难发现; 图 10(b) inclusion 内含物杂质图像、特征明显、易于区别; 图 10(c) patches 斑块, 色斑特征明显, 数量较少, 规模较大; 图 10(d) pitted surface 点蚀图像, 麻点密集, 数量较多; 图 10(e) rolled-inscale 轧制氧化皮, 特征不明显, 密集度较大; 图 10(f) scratches 刮痕图像, 特征明显, 数量较少.

由于该数据集数量较少, 采用垂直随机 0.5 概率随即翻转、多尺度训练数据增强方法提升数据数量. 在模型训练时将数据集按照 1:1:8 的比例划分为测试集、验证集和训练集, 测试集 180 张、验证集 180 张、训练集 1 440 张.

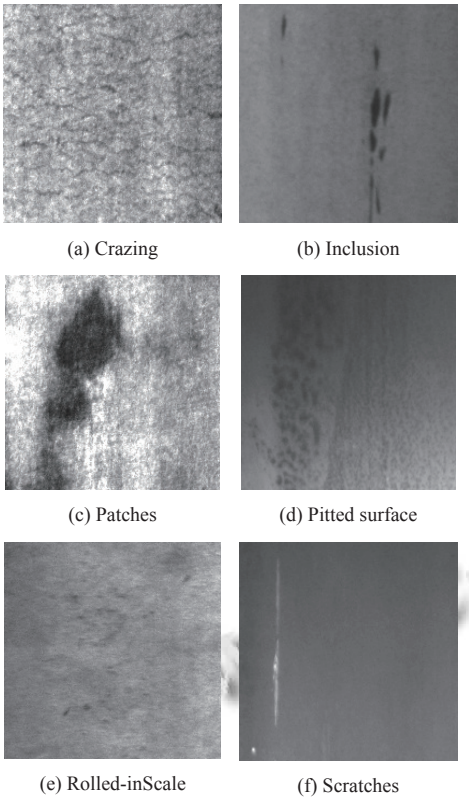


图 10 东北大学公开热轧带钢表面缺陷检测数据集

4.1.2 训练策略

(1) FP16 混合精度训练

混合精度训练是在减少精度损失的情况下利用半精度浮点数加速训练. 使用 FP16 即半精度浮点数存储权重和梯度. 在减少占用内存的同时起到了加速训练的效果. 图 11 为 IEEE 标准中的 FP16 格式.

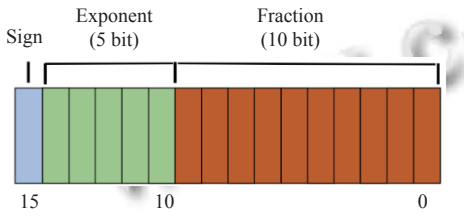


图 11 IEEE 标准中的 FP16 格式

FP16 取值范围是$5.96\times10^{-8}\sim65504$, 而 FP32 则是$1.4\times10^{-45}\sim3.4\times1038$, 从 FP16 的取值范围可以看出, 用 FP16 代替原 FP32 神经网络计算虽然会产生精度损失, 但是 FP16 相比 FP32 内存占用会大减少, 英伟达 GPU 支持加速运算, 节约训练时间.

(2) 柔性非极大值抑制 (Soft-NMS)

非极大值抑制 (NMS) 算法存在将相邻检测框的分数强制归零的情况, 将会存在重叠真实物体被强制归零, 导致检测失败, 平均检测精度 (average precision, AP) 下降.

传统 NMS 重置函数:

$$s_i=\begin{cases} s_i, & IoU(M,b_i)<N_t \\ 0, & IoU(M,b_i)\geqslant N_t \end{cases} \tag{12}$$

其中, s_i为检测框分数, $IoU(M,b_i)$为真实框与检测框的交并比函数, N_t为设置的重叠阈值.

采用柔性非极大值抑制 (Soft-NMS) 算法, 通过降低重叠检测框的分数而不是强制归零操作, 保留重叠检测框.

Soft-NMS 重置函数:

$$s_i=\begin{cases} s_i, & IoU(M,b_i)<N_t \\ s_i(1-IoU(M,b_i)), & IoU(M,b_i)\geqslant N_t \end{cases} \tag{13}$$

检测框超过重叠阈值时, 重置检测框分数线性衰减, 与M距离较近的检测框衰减程度加大, 而距离较远的检测框受影响程度较小.

(3) 实验优化策略

式 (14) 采用随机梯度下降 (SGD) 优化器, 目标函数$J(\theta)$, 模型参数集合θ, 累计梯度v_t, 学习率η为 0.002 5, 动量γ为 0.9, 无梯度限制. 学习策略采用预热策略, 预热迭代次数 500 次, 预热起始学习比率 0.001, 学习率衰减起止 16、19 step.

$$v_t=\gamma v_{t-1}+\eta\cdot\nabla_\theta J(\theta)\theta=\theta-v_t \tag{14}$$

实验环境如表 1 所示.

表 1 实验环境

实验环境	版本型号
操作系统	Ubuntu 18.04
CPU	Intel(R) Xeon(R) CPU
GPU	NVIDIA Tesla P100-PCIE
内存	16 GB
IDE	Jupyter Notebook
深度学习框架	PyTorch

4.2 结果分析

4.2.1 模型分析

模型评价指标, 计算精确率 P (precision)、召回率 R (recall)、平均精度 AP (average precision) 和平均精度均值 mAP (mean average precision) 作为网络模型性能评估标准. mAP 是所有类别平均精度均值, 通常用来评价检测模型的整体性能.

$$P=\frac{TP}{TP+FP} \tag{15}$$

$$R=\frac{TP}{TP+FN} \tag{16}$$

$$AP=\int_0^1 P(R)dR \tag{17}$$

$$mAP=\frac{1}{n}\sum_{i=1}^{n}(AP)_i \tag{18}$$

表 2、式 (15) 和式 (16) 中的 *TP* (true positive)、*TN* (true negative)、*FP* (false positive)、*FN* (false negative) 分别代表: 将正样本预测为正类的个数 (*TP*); 将负样本预测为负类的个数 (*TN*); 将负样本预测为正类的个数 (*FP*); 将正样本预测为负类的个数 (*FN*).

表 2 精确率和召回率表

	Positive	Negative
Positive	*TP*	*FP*
Negative	*FN*	*TN*

4.2.2 不同策略模型消融实验

4 种策略分别是 ResNet50+R-CNN (Faster R-CNN)、Swin Transformer+R-CNN、ResNet50+Multi-stage R-CNN、Swin Transformer+Multi-stage R-CNN (本文). 如表 3、表 4 所示, 通过消融实验平均精度均值 (*mAP*) 和平均精确度 (*AP*), 可以更加准确地了解到模型在修改过程前后的性能提升.

如表 3 所示, 基础策略模型为 ResNet50+R-CNN, 不同策略下 mAP 和 mAP_50 提升.

表 3 不同策略消融实验平均精度均值 (*mAP*) 对比 (%)

策略	mAP	mAP_50	mAP_75
ResNet50+R-CNN (Faster R-CNN)	23.2	65.3	10.1
Swin Transformer+R-CNN	24.6	70.9	7.7
ResNet50+Multi-stage R-CNN	27.0	68.2	14.4
Swin Transformer+Multi-stage R-CNN (本文)	29.2	72.8	14.0

(1) Swin Transformer+R-CNN 实现 1.4% 和 5.6% 的平均精度均值提升;

(2) ResNet50+Multi-stage R-CNN 实现 3.8%、2.9% 和 4.3% 的平均精度均值提升;

(3) Swin Transformer+Multi-stage R-CNN (本文) 实现 6%、7.5% 和 3.9% 的平均精度均值提升.

如表 4 所示, 本文设计的算法模型 (本文), 在裂纹、夹杂、斑块、麻点、压入氧化铁皮、以及划痕等缺陷上实现 5.2%、3.4%、7.7%、1.5%、8.9%、9.4% 的平均精度提升.

表 4 6 种缺陷在不同策略消融实验中平均精度 (*AP*) 对比 (%)

策略	Crazing	Inclusion	Patches	Pitted_Surface	Rolled-inScale	Scratches
ResNet50+R-CNN(Faster R-CNN)	9.9	22.9	34.7	32.3	18.3	21.2
Swin Transformer+R-CNN	12.7	21.0	40.7	29.8	21.7	21.8
ResNet50+Multi-stage R-CNN	11.8	23.0	41.3	35.3	22.3	27.9
Swin Transformer+Multi-stage R-CNN (本文)	15.1	26.3	42.4	33.8	27.2	30.6

4.2.3 可视化模型训练 Loss 曲线

通过观察训练过程中的损失参数变化, 有助于了解模型训练设置, 进而有助于提升模型的训练精度.

如图 12 所示, 更换骨干网络后的 Swin Transformer+R-CNN 模型相较于 ResNet50+R-CNN 模型在训练时波动范围较大, 波峰波谷大于 0.1, 难以实现 Loss 曲线的快速收敛.

ResNet50+Multi-stage R-CNN 模型快速实现收敛, 波峰波谷范围小于 0.1 而且随着训练轮数的增加, 实现收敛.

Swin Transformer+Multi-stage R-CNN (本文) 模型相较于前 3 个模型实现快速收敛, 波峰波谷波动范围小于 0.1, 和本文采取的训练策略相关, 是 4 个模型中训练效果最好的模型.

4.2.4 实验模型评价

根据表 2 和式 (17)、式 (18), 计算精确率 *P* 和召回率 *R*, 生成图 13、在阈值逐级提升下的 P-R 图. 如图 13 可以发现随着阈值 (*IoU*) 的提升模型的 *AP* 值 (面积) 不断下降.

通过将本文设计的模型与当前主流的物体检测算法, 在相同数据集和实验环境下进行检测对比实验, 表 5、表 6 为对比实验数据.

如表 5 所示, 本文设计的算法模型在 mAP、mAP_50 和 mAP_75 上, 相较 YOLOF 实现 3.4%、5.9% 和 0.4% 的精度提升; 相较于 YOLOv3 实现 15.4%、25.9% 和 8.5% 的精度提升; 相较 Deformable DETR 实现 20.7%、53.1% 和 7.8% 的精度提升; 相较于 SSDLit 实现 13.9%、24.5% 和 9.9% 的精度提升; 相较于 SSD512 实现 20%、38.9% 和 12.4% 的精度提升.

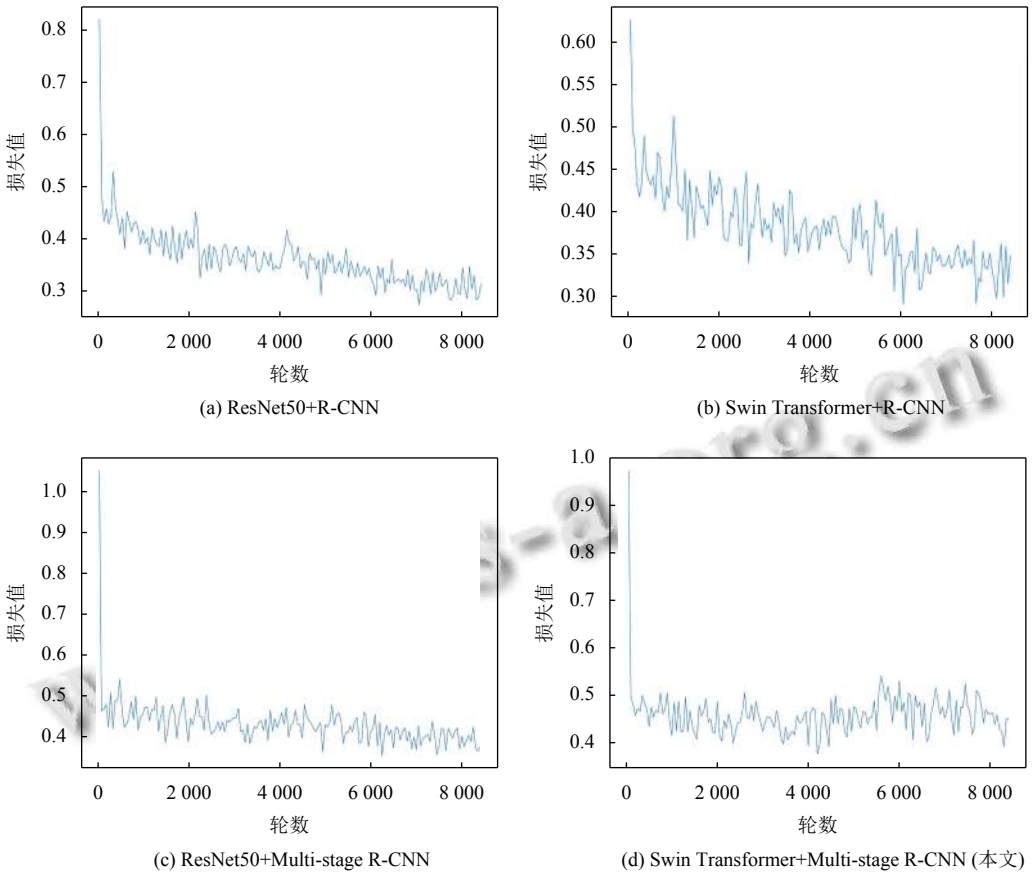


图 12 不同策略消融实验下的损失函数收敛曲线图

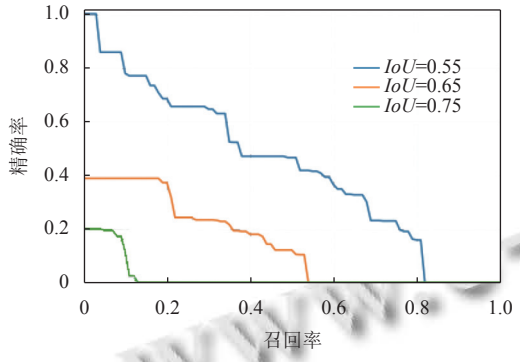


图 13 *IoU* 为 0.55、0.65、0.75 的 P-R 图

表 5 不同算法的平均精度均值 (*mAP*) 对比 (%)

算法	mAP	mAP_50	mAP_75
YOLOF[28]	25.8	66.9	13.6
YOLOv3[8]	13.8	46.9	5.5
Deformable DETR[29]	8.5	19.7	6.2
SSDLit[5]	15.3	48.3	4.1
SSD512[5]	9.2	33.9	1.6
本文	29.2	72.8	14.0

如表 6 所示, 本文设计的算法模型在 6 种缺陷上的检测精度 (*AP*), 相较 YOLOF 实现 1.9%、–1.4%、0.8%、4.6%、4.1%、10.6% 的检测精度提升; 相较 YOLOv3 实现 7.2%、18.2%、24.5%、2.5%、22.5%、17.5% 的检测精度提升; 相较 Deformable DETR 实现 11.8%、17.9%、30.5%、20%、20.3%、24.1% 的检测精度提升; 相较 SSDLit 分实现 7.2%、11.5%、9.3%、13.9%、18.5%、19.7% 的检测精度提升; 相较 SSD512 实现 12%、17.4%、22.2%、24.3%、22.5%、21.6% 的检测精度提升.

表 6 6 种缺陷在不同算法下平均精度 (*AP*) 对比 (%)

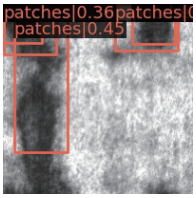
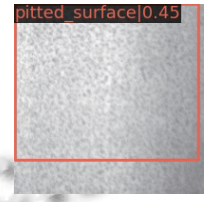
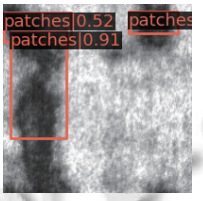
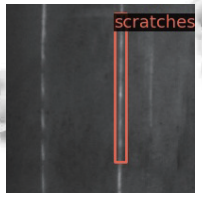
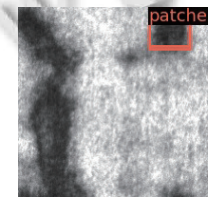
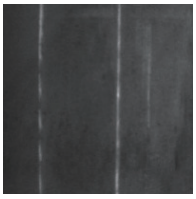
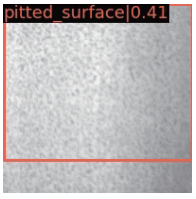
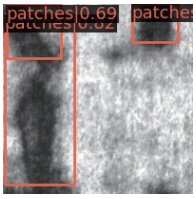
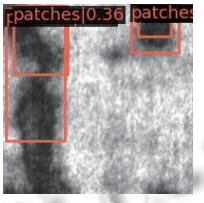
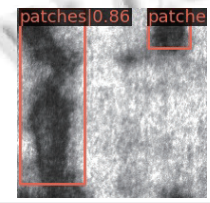
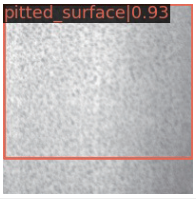
算法	Cr	In	Pa	PS	RS	Sc
YOLOF[28]	13.2	27.7	41.6	29.2	23.1	20.0
YOLOv3[8]	7.9	8.1	17.9	31.3	4.7	13.1
Deformable DETR[29]	3.3	8.4	11.9	13.8	6.9	6.5
SSDLit[5]	7.9	11.5	33.1	19.9	8.7	10.9
SSD512[5]	3.1	4.1	19.9	16.4	2.9	9.0
本文	15.1	26.3	42.4	33.8	27.2	30.6

如表 5、表 6, 本文设计的模型在检测精度上性能相优于 YOLOF、YOLOv3、Deformable DETR、SSDLit、SSD512 等主流模型.

4.2.5 实验效果

为了体现本文设计的算法检测精度的优越性, 将目前主流的几种检测算法与本文算法进行效果对比, YOLOF[28]、YOLOv3[8]、Deformable DETR[29]、SSDLit[5] 和 SSD512[5] 等算法, 对比实验采用的数据集相同, 输入图像尺寸一致, 如表 7 所示, 不同算法检测效果.

表 7 检测效果对比表

算法	Patches	Scratches	Pitted Surface
YOLOF[28]			
YOLOv3[8]			
Deformable DETR[29]			
SSDLit[5]			
SSD512[5]			
本文			

实验效果表明, 本文所设计的热轧带钢表面缺陷检测算法模型, 在公开数据集 NEU-DET 上的检测效果较好; 检测精度优于其他检测算法, 缺陷检测效果无论是在消融实验还是在与主流算法对比实验中均表现较好. 相比其他检测算法存在检测精度较低、存在漏检、误检等问题, 本文设计的算法, 在检测精度、检测正确率等评价指标方面都有较优异的性能表现.

作为 Transformer 骨干结构的视觉图像检测算法, 能够在检测性能上远远优于同类的检测架构, 检测性能超越主流的卷积架构检测网络, 在技术发展的过程中有助于技术的更迭.

5 结语

本文针对热轧带钢表面缺陷检测问题, 设计的

STM R-CNN 检测算法, 实现了缺陷的精确分类与定位. 首先, 采用 region proposal network 作为 baseline, 采用 cross entropy loss 和 $Smooth_{L1}Loss$ 进行分类和边界框回归. 其次采用 Swin Transformer 替换残差网络作为骨干特征提取网络, 设计 multi-stage R-CNN 多级联结构实现多阈值检测.

通过实验得到了以下结论: 在消融实验中本文设计的算法在 mAP、mAP_50 和 mAP_75 三个指标上比基础算法分别实现 6%、7.5% 和 3.9% 的精度提升; 在裂纹、夹杂、斑块、麻点、压入氧化铁皮以及划痕等类别缺陷上分别实现 5.2%、3.4%、7.7%、1.5%、8.9%、9.4% 的精度提升.

本文设计的算法模型在 mAP、mAP_50 和 mAP_75 上, 相较 YOLOF 实现 3.4%、5.9% 和 0.4% 的精度提升; 相较于 YOLOv3 实现 15.4%、25.9% 和 8.5% 的精度提升; 相较 Deformable DETR 实现 20.7%、53.1% 和 7.8% 的精度提升; 相较于 SSDLit 实现 13.9%、24.5% 和 9.9% 的精度提升; 相较于 SSD512 实现 20%、38.9% 和 12.4% 的精度提升.

实验结果证明, 本文设计的算法模型在检测的精确度还是快速收敛方面都优于 YOLOF、YOLOv3、Deformable DETR、SSDLit、SSD512 等主流模型. 但是, 本文算法没有能够完全抛弃卷积网络结构, 没有完全实现理想算法模型. 下一步, 将结合 DETR 算法思想, 将 RPN 和 R-CNN 层网络完全由 Transformer 结构替代, 实现彻底摆脱卷积运算结构. 届时也将会对模型进行参数优化和训练策略调整, 实现 Transformer 结构对视觉处理领域的完全占领.

参考文献

1 那宝魁. 钢铁企业质量体系中的设计和过程控制. 钢铁, 1997, (8): 73–74. [doi: 10.13228/j.boyuan.issn0449-749x.1997.08.017]

2 本刊编辑. 热轧板常见十四种缺陷. 新疆钢铁, 2020, (3): 24, 37, 40, 43, 46.

3 任海鹏, 马展峰. 基于复杂网络特性的带钢表面缺陷识别. 自动化学报, 2011, 37(11): 1407–1412.

4 Li SB, Yang J, Wang Z, *et al*. Review of development and application of defect detection technology. Acta Automatica Sinica, 2020, 46(11): 2319–2336.

5 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot MultiBox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.

6 Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 779–788.

7 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6517–6525.

8 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv: 1804.02767v1, 2018.

9 Ren SQ, He KM, Girshick R, *et al*. 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. [doi: 10.1109/TPAMI.2016.2577031]

10 He KM, Gkioxari G, Dollár P, *et al*. Mask R-CNN. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2980–2988.

11 裴伟, 许晏铭, 朱永英, 等. 改进的 SSD 航拍目标检测方法. 软件学报, 2019, 30(3): 738–758. [doi: 10.13328/j.cnki.jos.005695]

12 石杰, 周亚丽, 张奇志. 基于改进 Mask RCNN 和 Kinect 的服务机器人物品识别系统. 仪器仪表学报, 2019, 40(4): 216–228.

13 He D, Xu K, Zhou P. Defect detection of hot rolled steels with a new object detection framework called classification priority network. Computers & Industrial Engineering, 2019, 128: 290–297.

14 Fu GZ, Sun PZ, Zhu WB, *et al*. A deep-learning-based approach for fast and robust steel surface defects classification. Optics and Lasers in Engineering, 2019, 121: 397–405. [doi: 10.1016/j.optlaseng.2019.05.005]

15 He Y, Song KC, Dong HW, *et al*. Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network. Optics and Lasers in Engineering, 2019, 122: 294–302. [doi: 10.1016/j.optlaseng.2019.06.020]

16 He D, Xu K, Zhou P, *et al*. Surface defect classification of steels with a new semi-supervised learning method. Optics and Lasers in Engineering, 2019, 117: 40–48. [doi: 10.1016/j.optlaseng.2019.01.011]

17 李维刚, 叶欣, 赵云涛, 等. 基于改进 YOLOv3 算法的带钢表面缺陷检测. 电子学报, 2020, 48(7): 1284–1292. [doi: 10.3969/j.issn.0372-2112.2020.07.006]

18 程婧怡, 段先华, 朱伟. 改进 YOLOv3 的金属表面缺陷检

测研究. 计算机工程与应用, 2021, 57(19): 252–258. [doi: 10.3778/j.issn.1002-8331.2104-0324]

19 刘亚姣, 于海涛, 王江, 等. 基于深度学习的型钢表面多形态微小缺陷检测算法. 计算机应用, 2021: 1–8. http://kns.cnki.net/kcms/detail/51.1307.TP.20211014.1312.010.html. (2021-11-22).

20 He Y, Song KC, Meng QG, *et al*. 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. [doi: 10.1109/TIM.2019.2915404]

21 常海涛, 苟军年, 李晓梅. Faster R-CNN 在工业 CT 图像缺陷检测中的应用. 中国图象图形学报, 2018, 23(7): 1061–1071. [doi: 10.11834/jig.170577]

22 Liu Z, Lin YT, Cao Y. Swin Transformer: Hierarchical vision transformer using shifted windows. Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 9992–10002.

23 Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.

24 Dosovitskiy A, Beyer L, Kolesnikov A, *et al*. An image is worth 16x16 words: Transformers for image recognition at scale. Proceedings of the 9th International Conference on Learning Representations. ICLR, 2021.

25 Tang CX, Zhao YC, Wang GT, *et al*. Sparse MLP for image recognition: Is self-attention really necessary? arXiv: 2109.05422, 2021.

26 Tao X, Zhang D, Ma W. Automatic metallic surface defect detection and recognition with convolutional neural networks. Applied Sciences, 2018, 8(9): 1575.

27 Bao Y, Song K, Liu J. Triplet-graph reasoning network for few-shot metal generic surface defect segmentation. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1–11.

28 Chen Q, Wang YM, Yang T, *et al*. You only look one-level feature. Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13034–13043.

29 Zhu XZ, Su WJ, Lu LW, *et al*. Deformable DETR: Deformable transformers for end-to-end object detection. Proceedings of the 9th International Conference on Learning Representations. ICLR, 2021.

(校对责编: 牛欣悦)