

基于多尺度原型残差特征融合的连接器缺陷检测

程克林

上海赫立智能机器有限公司, 上海

收稿日期: 2023年6月26日; 录用日期: 2023年7月17日; 发布日期: 2023年7月28日

摘要

缺陷检测是工业生产中的重要一环,高效的缺陷检测方法能极大地提高工业流水线的生产效率。近年来,基于机器视觉的检测技术在降本增效方面取得了巨大进展,且能够进行全自动式检测,适用于各类检测场景。在工业连接器缺陷检测任务中,鉴于实际缺陷的外观变化可能很大,即缺陷区域可以有各种大小、形状和数量,针对这些问题本文提出了一种可用于工业连接器缺陷检测的基于多尺度原型残差特征融合的深度神经网络。具体来说,本文基于经典分割网络UNet进行改进,将不同尺度的特征和它们对应的原型残差特征融合,并使用自注意力机制来强化特征。同时,为了缓解实际检测环境中存在的正常样本和缺陷样本的数量不平衡问题,本文通过在正常样本上放置缺陷来扩充缺陷样本。实验表明,本文方法相比于基线方法表现出了更好的缺陷分割检测性能。

关键词

残差特征, 缺陷检测, 自注意力, 分割网络Unet, 深度神经网络

Connector Defect Detection Based on Multi-Scale Prototype Residual Feature Fusion

Clinton Cheng

Shanghai iHorry Machines Co., Ltd., Shanghai

Received: Jun. 26th, 2023; accepted: Jul. 17th, 2023; published: Jul. 28th, 2023

Abstract

Defect detection is an important part of industrial production, and efficient defect detection me-

thods can greatly improve the production efficiency of industrial assembly lines. In recent years, machine vision-based detection technology has made significant progress in cost reduction and efficiency improvement. This detection technology can perform fully automatic detection and is suitable for various detection scenarios. In the task of industrial connector defect detection, considering that the appearance of actual defects may vary greatly, that is, the defect region can have various sizes, shapes, and quantities, this paper proposes a deep neural network based on multi-scale prototype residual feature fusion for industrial connector defect detection. Specifically, this article improves on the classical segmentation network UNet by fusing features of different scales with their corresponding prototype residual features and using self-attention mechanism to enhance features. At the same time, to alleviate the imbalance in the number of normal and defective samples in the actual environment, we expand the defect samples by placing defects on normal samples. Experiments have shown that our proposed method exhibits better defect segmentation performance compared to the baseline method.

Keywords

Residual Feature, Defect Detection, Self-Attention, Segmentation Network Unet, Deep Neural Network

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

现代工业系统的飞速发展使得用户对产品性能的要求越来越高，而对于生产而言如何在提高产品性能的同时严格控制产品质量是需要重点关注的问题。然而，产品表面的划痕、斑点或孔洞等缺陷不仅会损害产品的美观性和使用舒适性，还可能对其性能产生不良影响。缺陷检测是减少产品缺陷不利影响的有效方法，传统的人工目检方法依赖于人工的专业性，可能存在因各种意外情况例如视觉疲劳等所带来的误检，同时传统检测手段的采样率较低、实时性也较差，难以适应现代工业的大规模生产要求[1]。

近年来，基于视觉传感器的机器视觉检测技术逐渐替代人工目检在实际生产中大量使用，这是一种非接触、无损的检测方法，通过光学设备和传感器自动接收和处理真实物体图像。在采集到物体图像后，使用一些传统的图像处理算法或神经网络模型对图像是否存在缺陷进行分类及对缺陷位置进行定位。机器视觉检测技术几乎是完全自动化的，可以大幅提高检测效率，同时保证了高度的实时性和准确性，减少人力成本，特别是对于一些大规模重复工业生产过程[2]。

相较于基于图像处理和统计学习的传统缺陷检测方法，基于深度神经网络的检测算法能够处理更为复杂和多样的工业图像，例如对于那些背景复杂，缺陷微小的样本，传统方法几乎无法成功检测和定位缺陷，而基于神经网络的方法则展现出卓越的性能优势，能够成功识别此类缺陷[3] [4] [5]。Augustauskas 等人基于 U-Net 网络使用空间金字塔池化和“瀑布”型连接来提取更好的缺陷特征[6]。李原等人修改了原始 U-Net 网络的编码模块，在残差块之间构建稠密连接从而增强浅层特征的深度延展[7]。Roth 等人利用核心集思想构建由正常样本特征块组成的特征库，通过比较输入样本特征块和特征库中特征块的相似度来分类和定位缺陷[4]。郑明明等人针对于不规则的缺陷区域，利用可变形卷积的优秀形变建模能力提取特征，并利用双分支注意力来优化特征[8]。尽管基于深度学习的缺陷检测方法已取得令人瞩目的进展，但它们对于那些实际检测场景中所出现的多变缺陷的检测能力仍然不足，这些缺陷可能是

细微的、难以辨别的，并且具有各种各样的外观。

本文提出了一种基于多尺度原型残差特征融合的缺陷检测模型并用于工业连接器的检测。该模型首先使用在大规模自然图像数据集 ImageNet [9]上预训练的特征提取器上提取多尺度的特征，每个尺度的特征都将被构建为多个能够代表该尺度下一些正常样本图像的特征原型。将每个尺度下的特征与在对应尺度下离其最近的原型做差值。然后将差值特征连接到原始特征上，这样便获得了包含多尺度残差特征的多尺度特征。融合了残差特征的多尺度特征再使用自注意力机制[10]进行特征强化以提升最终分割检测效果。同时我们还设计了一种用于扩充连接器缺陷样本的数据增广技术，通过将已知连接器缺陷样本的缺陷区域粘贴到正常的连接器样本上来获得伪缺陷样本并加入到训练数据中。在监督学习过程中，我们结合使用了平滑 L1 损失[11]和 Focal 损失[12]，其中 Focal 损失是一种用于缓解类不平衡问题的有效方法。

2. 提出方法

2.1. 缺陷数据扩充

在实际工业缺陷检测场景中，由于现代生产线对产品质量的严格控制，产品的良品率通常能够保证在非常高的水平，导致在收集数据时难以收集到缺陷样本，以致最后构建的数据集正常样本和缺陷样本的数量不平衡。这种不平衡的现象将导致在深度神经网络的训练过程中，正常样本的反向传播梯度将淹没缺陷样本的反向传播梯度以致最终检测模型的性能不佳。为此，本文针对于工业连接器图像设计了一种缺陷数据扩增技术，通过在正常的连接器图像上拼接从含有缺陷的连接器获得的缺陷部分来模拟缺陷样本，既增加了缺陷样本的数量，又在一定程度上缓解了类不平衡问题。

具体来说，首先对训练数据集中所有已经采集到的带缺陷连接器图像进行初次数据增强操作，可选的操作有对比度变换、清晰度变换、过度曝光和直方图均衡等，随机选取其中的两种操作对原始带缺陷图像进行组合变换，这样做的目的是为了模拟真实图像采集环境下可能出现的光照，采集设备镜头污染或故障等问题导致的图像失真。然后从这些变换后的带缺陷图像中提取出缺陷部分，这些缺陷部分将被施加旋转变换、剪切变换和平移变换中的一种以模拟缺陷部分位置和形状的差异。这些变换后的缺陷部分将被拼接回正常样本图像上以获取模拟异常图像，具体的拼接方法如下所示：

$$A = (1 - \alpha)(M \odot N) + M_{\text{invert}} \odot N + \alpha R, \quad (1)$$

其中 A 表示生成的模拟缺陷图像， R 表示缺陷区域， M 和 M_{invert} 分别表示缺陷区域的二值化掩码和其反转掩码， N 表示正常样本图像， α 表示不透明度参数， \odot 表示逐元素乘法运算。

2.2. 分割网络

本文所使用的分割网络整体基于 U-Net 结构[13]进行设计，使用在 ImageNet 数据集上预训练的 ResNet18 网络作为编码器主干。输入图像由编码器编码为多尺度特征，使用 ResNet18 中前三个尺度的特征来构建多尺度特征原型库。注意，构建特征原型库这一步是离线操作的，网络参数冻结，不对网络做任何优化。在得到多尺度特征原型库后，在训练或推理过程中，输入图像的多尺度特征将从对应尺度的特征原型库中找到最相似的原型并与其做差值然后将差值连接到原始特征上输入到自注意力模块中进行特征强化。具体网络结构如图 1 所示。

2.2.1. 原型生成与原型残差特征计算

对于训练数据集中的所有正常样本 D_N ，使用预训练的 ResNet18 网络提取每个正常样本的多尺度特征，ResNet18 网络包含四个残差连接块，每个残差连接块的最后一个卷积层的输出作为这个尺度下的特征，本文使用前三个残差连接块的输出特征来构建多尺度正常样本原型库。具体来说，对于第 i 个残差

连接块，其输出特征表示为 $F_i \in \mathbb{R}^{c_i \times h_i \times w_i}$ ($i \in 1, 2, 3$)，其中 c 表示特征图的通道数， h 表示特征图的高度， w 表示特征图的宽度。对于所有正常样本在第 i 个尺度下特征集合 \mathcal{F}_i ，使用 K-Means 聚类算法将这些特征聚类为 K 个簇，每个簇的聚类中心被视为一个原型，每个原型被认为可以代表一类相似正常样本的特征。最终，我们可以获得前三个尺度的正常样本原型库，每个尺度的原型库 \mathcal{P}_i 中均含有 K 个原型。

在训练过程中，对于输入图像 x ，其在第 i 个尺度下的特征表示为 F_i^x ，首先使用 L2 距离从第 i 个尺度的正常样本原型库中找到离 F_i^x 最近的原型 P_i^x ，即

$$P_i^x = \arg \min_{P_i^k \in \mathcal{P}_i} \|F_i^x - P_i^k\|_2. \quad (2)$$

然后，由特征 F_i^x 与其最邻近的原型 P_i^x 得到残差特征 G_i^x ，即

$$G_i^x = F_i^x \ominus P_i^x, G_i^x \in \mathbb{R}^{c_i \times h_i \times w_i}, \quad (3)$$

其中 \ominus 表示逐元素相减运算。

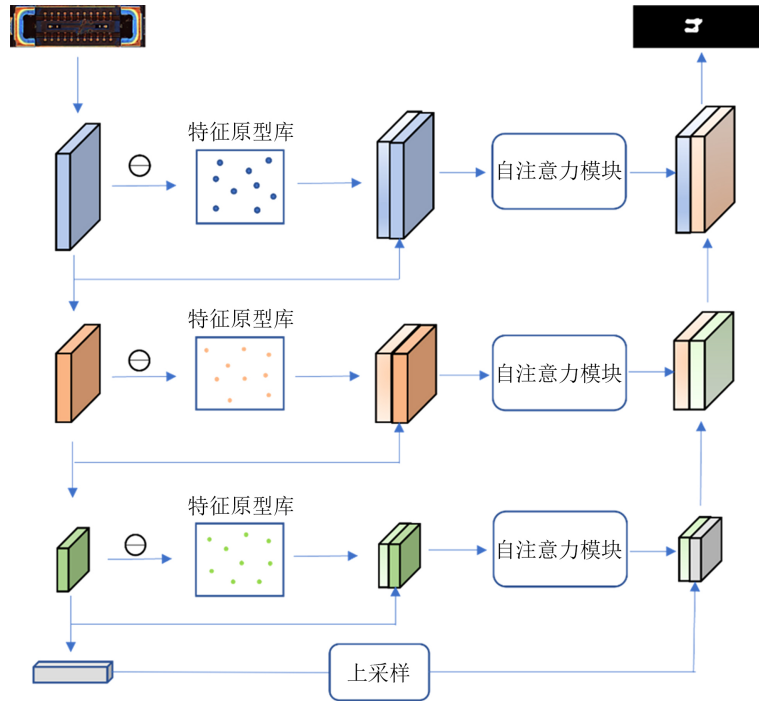


Figure 1. Overall network structure diagram
图 1. 网络整体结构图

2.2.2. 自注意力机制

使用自注意力机制能够增强有效特征，同时抑制不必要的特征。本文使用的自注意力模块由空间自注意力模块和通道自注意力模块组合而成。在深度学习中，每一个特征图都可以被视为一个目标检测器，通道注意力的目的就是将注意力集中在那些有意义的目标上，而空间注意力与通道注意力不同，空间注意力更加关注一个特征图内部的信息是哪一种信息，与通道注意力相辅相成。自注意力模块的输入特征首先经过通道自注意力，再经过空间自注意力，具体流程如式(4)所示：

$$\begin{aligned} F' &= M_c(F) \otimes F \\ F'' &= M_s(F') \otimes F', \end{aligned} \quad (4)$$

其中, M_c 表示通道自注意力, M_s 表示空间自注意力, F 表示输入特征, \otimes 表示张量积。在本文自注意模块的输入 F 为原始多尺度特征 F_i^x 与对应残差特征 G_i^x 的拼接特征。通道自注意力和空间自注意力分别由下式表示:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))), \quad (5)$$

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])), \quad (6)$$

其中, MLP 表示多层感知器, AvgPool 和 MaxPool 分别表示特征图的平均池化操作和最大池化操作。需要注意的是对于通道自注意力, 平均池化和最大池化是在通道层面操作的, 压缩了空间维度, 即 $M_c(F) \in \mathbb{R}^{c \times b \times l}$; 而对于空间自注意力, 平均池化和最大池化则是在空间层面操作的, 压缩了通道维度, 即 $M_s(F) \in \mathbb{R}^{l \times h \times w}$ 。自注意力机制如图 2 所示。

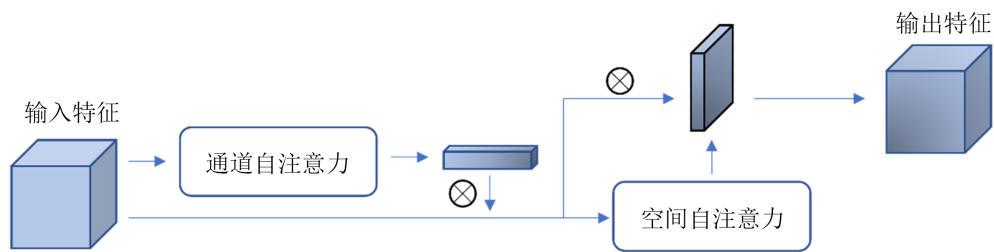


Figure 2. Self attention mechanism
图 2. 自注意力机制

2.2.3. 损失函数

在模型训练过程中, 我们使用平滑 L1 损失来拟合预测的缺陷分割掩码和真实的缺陷分割掩码, 具体如下所示:

$$\text{Smooth}_{L1} = \begin{cases} 0.5(y_i - \hat{y}_i)^2, & \text{若 } |y_i - \hat{y}_i| < 1 \\ |y_i - \hat{y}_i| - 0.5, & \text{其它,} \end{cases} \quad (7)$$

其中 y_i 表示真实分割掩码, \hat{y}_i 表示预测分割掩码。

同时, 为了进一步地缓解正常样本和缺陷样本的类不平衡问题以提高分割结果的鲁棒性, 本文还引入了 Focal 损失函数, 即

$$\text{Focal} = -\alpha(1 - \hat{y}_i)^\gamma \log(\hat{y}_i) \quad (8)$$

其中 α 用于调节正负样本损失之间的比例, $(1 - \hat{y}_i)^\gamma$ 用来减低易分样本的损失贡献。最终, 本文用于优化网络参数的损失函数为

$$\text{Loss} = \text{Smooth}_{L1} + \lambda \cdot \text{Focal}. \quad (9)$$

3. 实验部分

3.1. 数据集

实验所用的数据集均使用工业相机采集自真实生产流水线中的工业连接器样本, 通过人工缺陷检测方法确定缺陷样本和正常样本以及标注缺陷样本的缺陷部分。在该数据集中一共包含 2000 个样本, 其中含 1500 个正常样本和 500 个带缺陷样本, 其中训练集包含 1200 个正常样本和 300 个带缺陷样本, 验证集包含 300 个正常样本和 200 个带缺陷样本用于测试模型性能。

3.2. 实验部署

本文的所有实验均基于 Pytorch 深度学习框架，使用 NVIDIA TITAN Xp 图形加速卡进模型的训练和推理，所使用的 CPU 为 i7 6700K，内存容量为 32GB。所有采集图像及其对应分割掩码在被输入网络前裁剪为 512*512 分辨率。我们使用 SGD 优化器进行模型参数优化，初始学习率设置为 2×10^{-4} ，权重衰减为 10^{-5} ，共训练 7400 个轮次，每个轮次的批量大小设置为 64，每个批中包含 32 个正常样本和 32 个带缺陷样本，其中带缺陷样本是由真实标注缺陷样本和模拟缺陷样本混合而成。对于每个尺度的特征原型库中的原型数量，我们均设置为 20。Focal 损失中的 α 和 γ 分别设置为 0.5 和 4，最终损失中的 λ 设置为 3。网络参数由 ImageNet 上预训练的 ResNet18 网络参数初始化，注意，特征原型库在网络开始训练前便已离线构建完成。

3.3. 实验结果

本文实验结果如图 3 和图 4 所示。图 3 展示了本文方法对部分带缺陷连接器图像分割可视化结果，可以发现本文方法能够非常准确地分割出正确的缺陷区域，图中红色区域表示缺陷预测结果。



Figure 3. Prediction effect of defects in partial connector images with defects

图 3. 部分带缺陷连接器图像缺陷预测效果

图 4 显示了本文的方法在模型训练过程中的损失函数曲线和验证集像素分类准确率，可以发现本文的方法的损失函数能快速收敛到一个较低的水平而验证集像素分类准确率则快速上升。

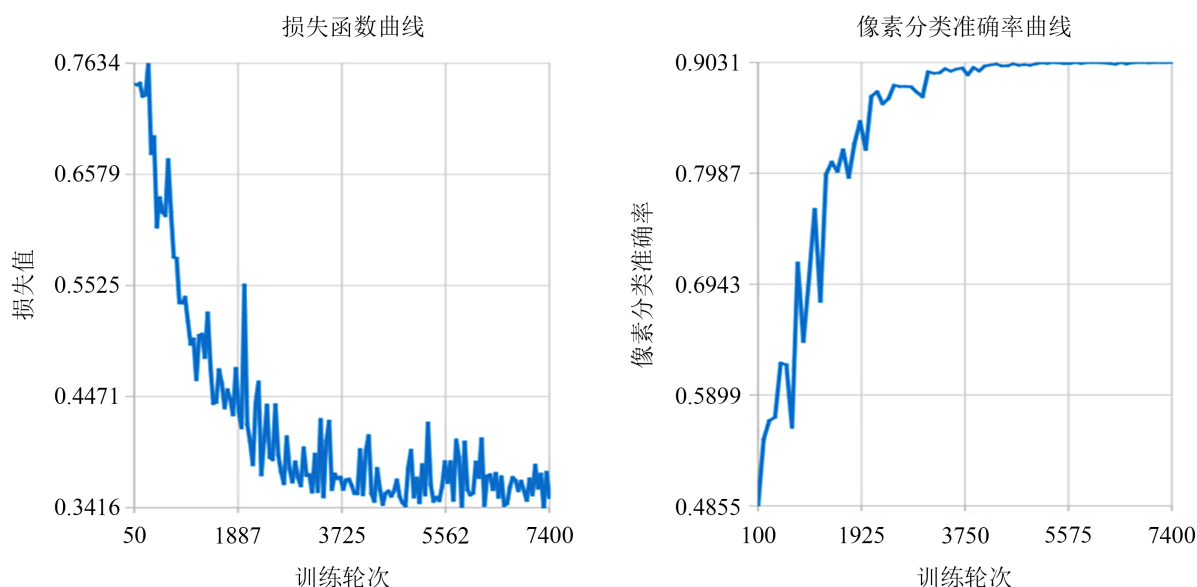


Figure 4. Loss function curve and pixel classification accuracy curve of verification set

图 4. 损失函数曲线及验证集像素分类准确率曲线

表 1 展示了原始 U-Net 网络和本文提出方法的像素分类准确率差异, 以及原型残差特征和自注意力机制的有效性。可以发现, 在 U-Net 网络上单独应用原型残差特征机制能够提升 2.3% 的像素分类准确率, 原型残差特征与原始特征融合能够使得模型在优化过程中正常样本的特征更加向具有代表性的正常样本特征(即原型)靠近, 从而获得更为紧凑的特征表示, 隐式地优化了正常样本与缺陷样本之间的分类边界。而单独应用自注意力机制则能够提升 1.2% 的像素分类准确率, 自注意力机制使得模型更加关注那些有辨别力的特征而忽略那些噪声特征, 而在缺陷检测中, 噪声信息是普遍存在的, 因此我们所采用的自注意力机制进一步地优化了样本的特征表示。本文所提出的方法同时使用了原型残差特征和自注意力机制, 充分结合了两种方法的优点, 相较于原始 U-Net 提升了 3.9% 的分类准确率达到 90.3% 的最终像素分类准确率。说明本文所提出的方法能够有效地对工业连接器进行缺陷分割检测。

Table 1. Effect analysis of different components

表 1. 不同组件的效果分析

方法	像素分类准确率(%)
UNet	86.4
UNet + 原型残差特征	88.7
UNet + 自注意力机制	87.6
UNet + 原型残差特征 + 自注意力机制(本文)	90.3

图 5 展示了原型数量对模型性能的影响, 需注意原型数量实际上就表示了每个尺度下所有正常样本特征的聚类数量。可以观察到当原型数量较少时, 分类准确率会出现下降, 原因是较少的原型数量容易导致原型之间的差异变小, 不利于模型学习有判别力的特征表示。相反地, 过多的原型将导致每个原型趋近于单个的正常样本特征, 这将使得在生成残差特征的过程中, 当前样本与原型的残差特征近似于当前样本与其最近邻的单个正常样本的残差特征, 而从最近的聚类原型获得的残差特征通常要比前者更具代表性。

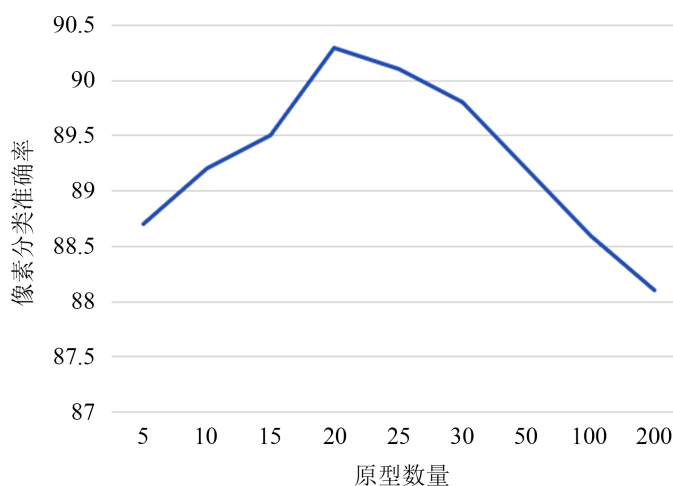


Figure 5. Pixel classification accuracy curve corresponding to different number of prototypes

图 5. 不同原型数量对应的模型像素分类准确率曲线

4. 结语

为了解决工业连接器缺陷分割中可能存在的缺陷区域多变问题, 本文提出了一种基于多尺度原型残差特征的缺陷分割模型。通过将原型残差特征与原始特征融合, 使得到的特征更加鲁棒。同时, 结合使用空间自注意力和通道自注意力机制来进一步强化特征。实验结果表明, 本文所提出的方法能够实现比基线模型更好的分割性能, 准确地分割出工业连接器的缺陷部分。

参考文献

- [1] 罗东亮, 蔡雨萱, 杨子豪, 章哲彦, 周瑜, 白翔. 工业缺陷检测深度学习方法综述[J]. 中国科学: 信息科学, 2022, 52(6): 1002-1039.
- [2] Ren, Z., Fang, F., Yan, N., *et al.* (2022) State of the Art in Defect Detection Based on Machine Vision. *International Journal of Precision Engineering and Manufacturing-Green Technology*, **9**, 661-691. <https://doi.org/10.1007/s40684-021-00343-6>
- [3] Bergmann, P., Fauser, M., Sattlegger, D., *et al.* (2020) Uninformed Students: Student-Teacher Anomaly Detection with Discriminative Latent Embeddings. *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 4183-4192. <https://doi.org/10.1109/CVPR42600.2020.00424>
- [4] Roth, K., Pemula, L., Zepeda, J., *et al.* (2022) Towards Total Recall in Industrial Anomaly Detection. *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 18-24 June 2022, 14318-14328. <https://doi.org/10.1109/CVPR52688.2022.01392>
- [5] Wu, K., Zhu, L., Shi, W., *et al.* (2022) Self-Attention Memory-Augmented Wavelet-CNN for Anomaly Detection. *IEEE Transactions on Circuits and Systems for Video Technology*, **33**, 1374-1385. <https://doi.org/10.1109/TCSVT.2022.3211839>
- [6] Augustauskas, R. and Lipnickas, A. (2020) Improved Pixel-Level Pavement-Defect Segmentation Using a Deep Autoencoder. *Sensors*, **20**, 2557. <https://doi.org/10.3390/s20092557>
- [7] 李原, 李燕君, 刘进超, 范衡, 王庆林. 基于改进 Res-UNet 网络的钢铁表面缺陷图像分割研究[J]. 电子与信息学报, 2022, 44(5): 1513-1520.
- [8] 郑明明, 刘胜全, 马前. 基于可变形卷积融合双注意力机制的缺陷检测方法[J]. 东北师大学报(自然科学版), 2023, 55(2): 52-61. <https://doi.org/10.16163/j.cnki.dslkxb202201180002>
- [9] Deng, J., Dong, W., Socher, R., *et al.* (2009) ImageNet: A Large-Scale Hierarchical Image Database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, 20-25 June 2009, 248-255. <https://doi.org/10.1109/CVPR.2009.5206848>
- [10] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. <https://doi.org/10.1109/CVPR.2018.00745>
- [11] Girshick, R. (2015) Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- [12] Lin, T.Y., Goyal, P., Girshick, R., *et al.* (2017) Focal Loss for Dense Object Detection. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 2980-2988. <https://doi.org/10.1109/ICCV.2017.324>
- [13] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. Springer International Publishing, Cham, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28