



基于多尺度和注意力机制的混合监督金属表面缺陷检测

孙博言, 王洪元, 刘乾, 冯尊登, 唐郢

引用本文:

孙博言,王洪元,刘乾,冯尊登,唐郢. 基于多尺度和注意力机制的混合监督金属表面缺陷检测[J]. *智能系统学报*, 2023, 18(4): 886–893.

SUN Boyan,WANG Hongyuan,LIU Qian,FENG Zundeng,TANG Ying. Hybrid supervised metal surface defect detection based on multi-scale and attention[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(4): 886–893.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202205042>

您可能感兴趣的其他文章

基于注意力机制的显著性目标检测方法

Salient object detection method based on the attention mechanism

智能系统学报. 2020, 15(5): 956–963 <https://dx.doi.org/10.11992/tis.201903001>

基于小样本学习的LCD产品缺陷自动检测方法

An automatic small sample learning-based detection method for LCD product defects

智能系统学报. 2020, 15(3): 560–567 <https://dx.doi.org/10.11992/tis.201904020>

基于注意力融合的图像描述生成方法

An image caption generation method based on attention fusion

智能系统学报. 2020, 15(4): 740–749 <https://dx.doi.org/10.11992/tis.201910039>

基于跳跃连接金字塔模型的小目标检测

Skip feature pyramid network with a global receptive field for small object detection

智能系统学报. 2019, 14(6): 1144–1151 <https://dx.doi.org/10.11992/tis.201905041>

面向自闭症辅助诊断的无监督模糊特征学习新方法

A novel unsupervised fuzzy feature learning method for computer-aided diagnosis of autism

智能系统学报. 2019, 14(5): 882–888 <https://dx.doi.org/10.11992/tis.201808005>

多标记学习自编码网络无监督维数约简

Unsupervised dimensionality reduction of multi-label learning via autoencoder networks

智能系统学报. 2018, 13(5): 808–817 <https://dx.doi.org/10.11992/tis.201804051>

DOI: 10.11992/tis.202205042

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.Tp.20230321.1603.008.html>

基于多尺度和注意力机制的混合监督 金属表面缺陷检测

孙博言, 王洪元, 刘乾, 冯尊登, 唐郢

(常州大学 计算机与人工智能学院、阿里云大数据学院、软件学院, 江苏常州 213000)

摘要: 针对缺陷检测中被检测样品中因缺陷目标形状各异引起的无法提取有效特征的问题, 本文提出基于深度学习的缺陷检测模型。该模型使用改进后的多尺度特征融合模块, 在控制计算量的基础上解决识别不同大小缺陷的问题。通过引入非局部注意力机制模块, 模型对缺陷特征的提取能力得到加强; 在训练中使用混合监督训练, 探索模型所需要的标注量和检测准确度之间的关系。本文方法在 KSDD、KSDD2 和 STEEL 3 个数据集上都获得了比先进方法更好的精确度, 对于不同类型的缺陷都能提取到有判别力的特征。与先进的完全监督方法和无监督方法相比, 在数据集上精确度平均提高 0.8% 和 11%。

关键词: 缺陷; 检测; 特征提取; 学习算法; 学习系统; 图像处理; 金属; 产品品质; 深度学习

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2023)04-0886-08

中文引用格式: 孙博言, 王洪元, 刘乾, 等. 基于多尺度和注意力机制的混合监督金属表面缺陷检测 [J]. 智能系统学报, 2023, 18(4): 886-893.

英文引用格式: SUN Boyan, WANG Hongyuan, LIU Qian, et al. Hybrid supervised metal surface defect detection based on multi-scale and attention[J]. CAAI transactions on intelligent systems, 2023, 18(4): 886-893.

Hybrid supervised metal surface defect detection based on multi-scale and attention

SUN Boyan, WANG Hongyuan, LIU Qian, FENG Zundeng, TANG Ying

(School of Compute Science and Artificial Intelligence / Aliyun School of Big Data / School of Software, Changzhou University, Changzhou 213000, China)

Abstract: Aiming at the problem in defect detection that effective features cannot be extracted due to different shapes of defect targets in the detected samples, this paper presents a defect detection model based on deep learning, which uses an improved multi-scale feature fusion module to solve the problem of identifying defects of different sizes on the basis of controlling the amount of calculation. By introducing a non-local attention mechanism module, the model's ability of extracting defect features is enhanced. Furthermore, mixed-supervised training is used in training to explore the relationship between the amount of annotations required by the model and the detection accuracy. This method achieves better accuracy than the state-of-the-art methods on KSDD, KSDD2, and STEEL datasets, and can extract discriminative features for different types of defects. Compared with the state-of-the-art fully supervised and unsupervised methods, the average accuracy improvement on the dataset is 0.8% and 11%.

Keywords: defect; detector; feature extraction; learnin galgorithm; learning system; image processing; metal; quality of product; deep learning

产品质量检测作为工业生产中的重要一环^[1-3]。起初人们主要是利用传统机器学习算法对这一问题进行研究^[4-5]。随着卷积神经网络(convolutional neural networks, CNN)^[6]等深度学习模型在计算机视觉领域的大规模应用, 利用深度学习的方法进行缺陷检测也逐渐成为一个热门的研究方向。根据标注程度的不同, 基于深度学习的缺陷检测可以分为 3 种: 1) 对全部数据进行图像级以及像素级标注(全监督学习), 典型代表为以 CNN 为主干网络的方法^[7-10]; 2) 对全部数据仅进行图像

级标注(弱监督学习)^[11-13]; 3) 对数据不进行任何标注(无监督学习), 主要以 VAEs(variational autoencoders)^[14-16]和 GANs(generative adversarial nets)^[17-18]为主。此外 Bozic 等^[19]利用混合监督训练对缺陷检测进行了研究。

2014 年, 人们开始了对注意力机制的研究并开始逐步扩展到不同的领域^[20-21]。2017 年 Wang 等^[22]从机器翻译出发, 将自注意力看作是一种非局部的平均值, 从而将其转化为一般的非局部滤波操作, 提出 Non-Local Block, 从而适用于计算机图像问题当中; 2019 年 Dong 等^[23]将注意力机制引入到缺陷检测当中, 提出一个基于金字塔特征融合和全局上下文注意网络的表面缺陷检测方

收稿日期: 2022-05-25. 网络出版日期: 2023-03-23.

基金项目: 国家自然科学基金项目(61976028); 2022 年江苏省研究生科研创新计划(KYCX22_3066).

通信作者: 王洪元. E-mail: hywang@cczu.edu.cn.

法——PGA(pyramid feature fusion and global context attention network for automated surface defect detection) 网络。但是这种方法更关注局部上下文的特征, 面对较大的缺陷则效果很差。

上述几种方法面对复杂的缺陷类型都略有不足, 同时在精确度与样本标注量之间没有做到很好的平衡。基于此, 本文提出一个融合了注意力机制的缺陷检测模型, 并在其中加入了一个多尺度融合单元; 将全监督学习和弱监督学习结合起来, 对模型进行混合监督训练^[18], 探索样本标注量对精确度的影响。

本文的主要贡献如下:

1) 使用改进的多尺度特征融合模块, 加强模型对不同形状的缺陷特征的提取能力并探索尺度对精确度的影响。

2) 融合 Non-Local Block 注意力机制单元, 增强模型对于缺陷特征的识别能力, 提高缺陷检测的精确度。

3) 使用混合监督训练探索像素级标注数量对精确度的影响。

1 多尺度注意力检测模型

缺陷检测具有缺陷目标大小不一、特征表达能力弱、特征信息比较少等问题, 针对上述问题, 本文以保持低计算量并提高模型在缺陷特征的表达能力的目标进行一系列探索, 主要进行了如下改进:

1) 改进了多尺度特征融合网络, 在控制计算量的基础上, 加强模型对于形状各异特征的提取能力。

2) 在网络中融入 Non-Local Block 注意力机制模块, 优化网络中对特征的表达能力, 丰富特征图中缺陷的特征信息。

3) 设计不同数量的像素级标注, 利用混合监督训练探索标注级别对精确率的影响。

1.1 缺陷检测模型

如图 1 所示, 模型主要分为 2 个模块, 分别为 L_1 和 L_2 模块, 其中 L_1 模块使用像素级数据进行训练, 提高模型像素级的识别能力; L_2 模块则使用图像级数据进行训练, 提高模型图像级的识别能力。网络架构细节如表 1 所示。对于一张输入图片, 首先利用 L_1 模块 4 个卷积层提取图像特征, 在每一层卷积层后设置 2×2 的最大池化层, 以减小图像的大小并保留更多的特征。相比于利用具有步长的卷积层降低图像大小, 这种操作更有利于保留更多的特征。选取 L_1 模块第 1 层和最后一层输出的特征图送入多尺度特征融合模块进行进一步处理。此外最后一层卷积分别输出单通道和 1024 通道的特征图, 单通道的特征图将利用 Non-Local Block 对其进行特征加强后与 1024 通道特征图进行映射连接形成一个 1025 通道的特征图。经过注意力机制单元处理后的单通道特征图能够避免网络使用大量的特征图, 减少对大量参数的过度拟合, 同时加强模型对于缺陷特征的关注。利用 L_2 模块对输出的 1025 通道的特征图进行操作并分别输出 2 个 32 通道的特征图。最后将这 2 个特征图与多尺度特征融合单元输出的 3 通道特征图进行映射连接送入全连接层, 生成最终的分类。

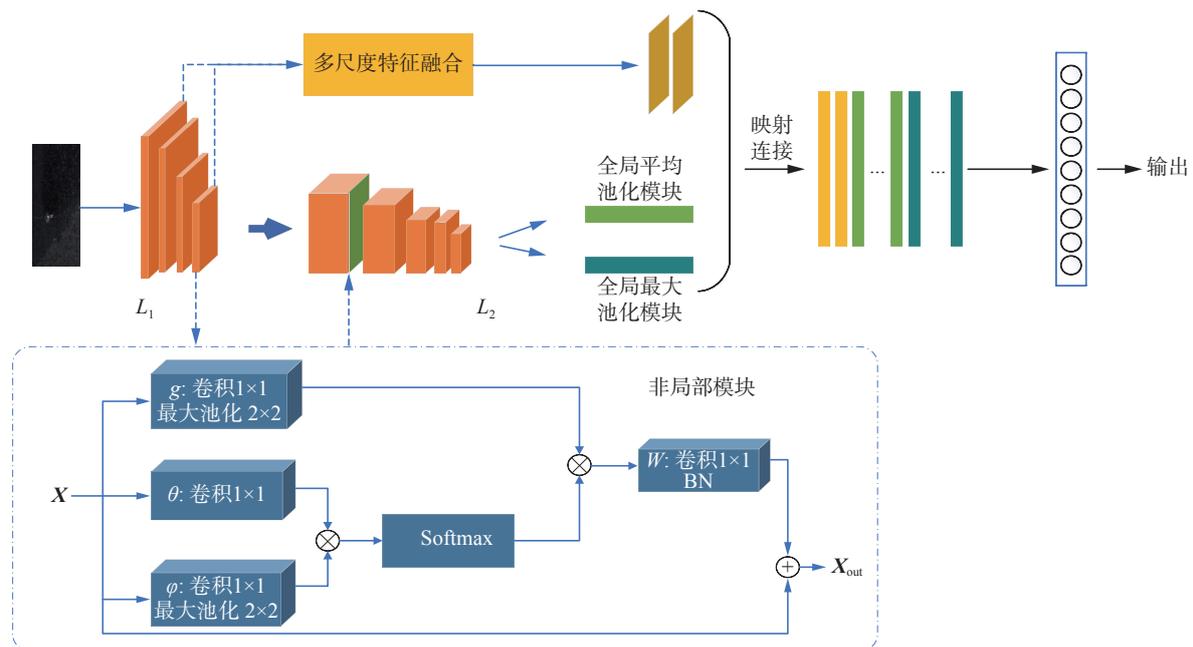


图 1 网络图
Fig. 1 Network diagram

表 1 网络架构细节
Table 1 Network architecture details

| 模块 | 卷积层 | 卷积核 | 特征数量 |
|----------|---------------------------------|--------------|-------|
| L_1 模块 | 输入: 图片 | — | 3/1 |
| | $2 \times \text{Conv2D}(I_1)$ | 5×5 | 32 |
| | Max-pool | 2×2 | 32 |
| | $3 \times \text{Conv2D}$ | 5×5 | 64 |
| | Max-pool | 2×2 | 64 |
| | $4 \times \text{Conv2D}$ | 5×5 | 64 |
| | Max-pool | 2×2 | 64 |
| | $\text{Conv2D}(I_2)$ | 5×5 | 1024 |
| | $\text{Conv2D}(I_3)$ | 1×1 | 1 |
| | Non-Local Block ^[22] | 输入 I_3 | — |
| 输出 I_4 | | — | 1 |
| 多尺度模块 | 输入 I_1/I_3 | — | 32/1 |
| | 输出 $I_5/I_6/I_7$ | — | 1/1/1 |
| L_2 模块 | 输入: I_2, I_4 | — | 1025 |
| | Max-pool | 2×2 | 1025 |
| | Conv2D | 5×5 | 8 |
| | Max-pool | 2×2 | 8 |
| | Conv2D | 5×5 | 16 |
| | Max-pool | 2×2 | 16 |
| | Conv2D | 5×5 | 32 |
| | Max-pool(I_8) | 2×2 | 32 |
| | Avg-pool(I_9) | 2×2 | 32 |
| | $I_5 + I_6 + I_7 + I_8 + I_9$ | — | 67 |
| 全连接层 | — | 1 | |

本文将 2 个模块的损失组合成一个统一的联合损失:

$$L_{\text{total}} = \lambda \cdot \gamma \cdot L_1 + (1 - \lambda) \cdot L_2 \quad (1)$$

式中: λ 作为平衡因子, 平衡 2 个模块在整个模型中

的作用; γ 作为像素级标注是否存在的指示器。

1.2 多尺度特征融合模块

多尺度特征融合模块主要是为了解决由于不同深度的卷积层感受野大小不一导致的目标检测能力的差距问题, 通过将不同尺度的特征融合在一起, 可以有效提高模型对于各种类型缺陷的检测能力。

本文在前人的基础上, 改进了多尺度特征融合模块。模块的改进思路主要有以下 2 点: 1) 突出多尺度的作用。本文仅通过池化操作快速改变模型的通道数便于后期的特征融合; 同时避免深层卷积对多尺度的影响, 突出多尺度中不同尺度对于模型的影响。2) 降低多尺度的层数。过多的层数会导致参数量剧增, 缺陷检测需要用较少的时间实现更好的效果。同时本文模型中相邻的不同尺度特征图之间差异并不明显, 从而对模型性能提升较小。而在弱监督训练下, 由于缺陷没有明显的标注, 较多的尺度图的组合会提供更多的信息, 但是相邻尺度间的不良影响仍然存在。

模块的构成如图 2 所示, 本文仅使用第 1 层和最后一层的特征, 忽略中间层的特征。第 1 层具有较大的感受野, 可以提高对较小目标的识别准确率。同时, 对于第 1 层的特征只需要一个池化操作, 随后本文提取了特征的最大值以获得更有判别力的缺陷特征。对于深层卷积输出的特征图, 为了突出特征图中的明显特征, 利用 1×1 的卷积加强了特征的提取, 此外, 还进行了取平均值和取最大值操作。由于缺陷检测样本中存在大量噪声, 本文利用最深层特征的平均值有效抑制噪声对模型的影响。最后利用映射连接融合不同尺度。

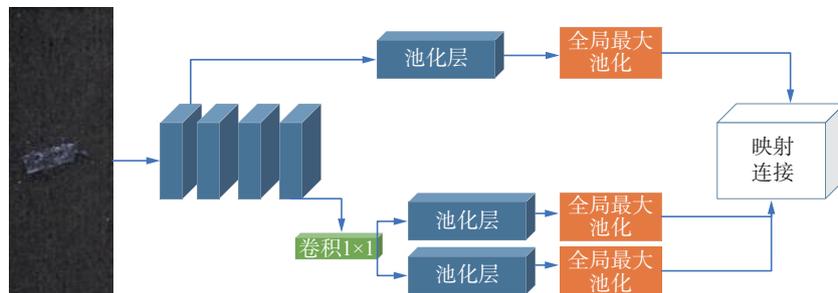


图 2 多尺度特征融合模块

Fig. 2 Multi-scale feature fusion module

1.3 注意力模块

工业产品的缺陷检测具有环境复杂, 目标小等特点, 导致缺陷特征信息比较少, 为了提高模型的特征表达能力, 本文在模型中融入了注意力机制模块 Non-local Block, 丰富特征图中的缺陷信息。相比较于其他注意力机制, Non-local Block

拥有更高的计算效率并可以快速捕捉较长范围内的 2 个像素点的相互依赖。

本文利用 Non-Local Block 单元对单通道特征图进行了进一步处理。如图 1 所示, 对于输入特征图 X , 令其分别通过 g 、 θ 和 ϕ 操作生成 3 个特征图。 g 、 θ 和 ϕ 是一个 1×1 的卷积操作。令经过 θ 和

ϕ 作用后的矩阵进行矩阵乘法并使用 SoftMax 激活函数对结果进行处理, 使其结果映射到 (0,1) 之间, 获得注意力系数矩阵 f , 此矩阵表示 2 个像素点之间的关联性。方法 g 代表某一点的线性映射, 使 f 和 g 相乘, 最终获得结果 y_i 。

整个注意力机制模块如下所示:

$$X_{out} = W \cdot y_i + x_i \quad (2)$$

式中: W 为一个 1×1 卷积操作, 目的是将输出的通道图转变为与输入相同, 并保证整个注意力机制模块的“即插即用”; x_i 为一个残差连接, 避免 Non-Local Block 模块对整个模型产生扰乱; y_i 为 Non-Local 单元。主要的结果如下所示:

$$y_i = \text{SoftMax}(\theta \cdot \varphi(x_i))g(x_i) = \frac{e^{\theta(x_i^T) \cdot \varphi(x_i)}}{\sum_{vj} \theta(x_i^T) \cdot \varphi(x_i)} \cdot w_g x_i \quad (3)$$

与正常样品相比, 缺陷的特征会存在很大的不同, 利用 Non-Local 注意力机制单元可以使当前像素点与其他像素点建立有效的联系, 以权重数值的模式加强模型对于这些区域的关注, 以提取更明显的缺陷特征。

1.4 混合监督训练

本文通过使用混合监督训练的方法, 尝试探索不同数量的像素级标注对于精确率的影响。将图像的标注分为图像级标注和像素级标注, 弱监督训练、混合监督训练和完全监督训练如图 3 所示, 其中对错号(即钩叉号)分别代表图像级标注的正常图像和缺陷图像, 像素级标注则为在缺陷图像上对缺陷进行按像素标记。

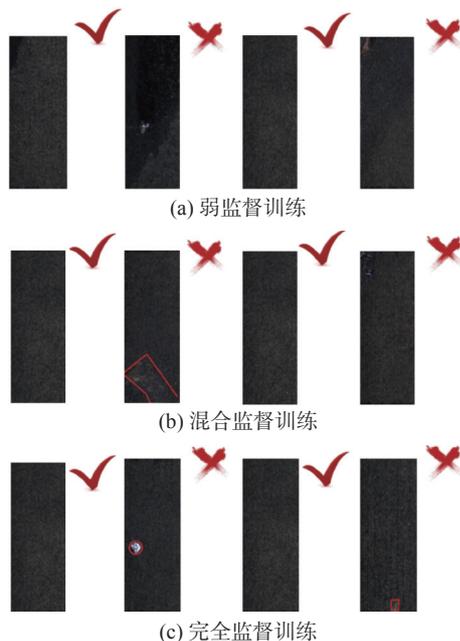


图 3 不同训练方式对照
Fig. 3 Comparison of different training methods

具体设定如下:

- 1) 弱监督训练: 只有图像级标注, 无像素级标注, ($N=0$);
- 2) 混合监督训练: 所有图像具有图像级标注, 部分缺陷图像具有像素级标注 ($0 < N < N_{all}$);
- 3) 完全监督训练: 所有图像既具有图像级标注也具有像素级标注, ($N = N_{all}$)。

其中混合监督训练的像素级标注数量设为 N , 并且像素级标注一定是有缺陷样本。

图像的标注程度由联合损失函数中的 γ 决定, 当 γ 为 1 时, 图像为正常样品和带像素级标注的缺陷样本; 当 γ 为 0 时, 图像为无任何标注的缺陷样本。对于 L_1 模块, 即便在弱监督情况下, 模型的像素级识别能力也会得到训练, 只有当训练数据为无任何标注的缺陷样本时, L_1 模块才会被禁用。任何情况下, L_2 模型都将被训练用于从图像级角度识别图像是否有缺陷。

2 仿真验证与结果

2.1 数据集和评价指标

本文实验使用的数据集是当前缺陷检测领域 3 个主流的数据集: KolektorSDD、KolektorSDD2 和 Severstal Steel defect dataset。KolektorSDD 数据集^[2] 总共包含 399 张图片, 其中 52 张为有缺陷图像, 347 张为正常图片。

KolektorSDD2 数据集^[19] 总共包含 3 335 张图片, 其中 356 张有缺陷图像, 2 979 张正常图像。该数据集在缺陷图片中融入了多种不同类型的缺陷(划痕、小斑点等)。该数据集于 2021 年公布, 是缺陷检测领域内最新的数据集。上述 2 个数据集都来源于 Kolektor Group d.o.o 在真实生产环境中捕获的金属表面的图片。

Severstal Steel defect 数据集远大于其他 2 个数据集, 包含 4 类 12 568 张灰度图像, 其中包含各种缺陷。图像来源于 Severstal 公司在平板钢生产时的图形。在实验中, 本文仅使用数据集的一个子集进行训练。

评估模型性能时, 由于工业环境下对缺陷检测要求只有有无缺陷 2 种情况, 所以本文采用工业环境下常用的精确度 (average precision, AP) 作为评价指标, 其计算方式为

$$A = \int_0^1 p(r) dr \quad (4)$$

$$p = \frac{T}{T + F} \quad (5)$$

式中: A 为精确度, p 为模型的准确率, T 为正确划分为正例的样本数, F 为错误划分为正例的样本数。

2.2 实验设计

对于像素级标注图像的数量 N , 本文在进行不同数量的实验后, 分别选定了最具有代表性的数目。KSDD 数据集设计了 $N=\{0,5,15,20,\text{all}\}$; KSDD2 数据集设定 $N=\{0,16,53,126,246\}$ 。上述 2 个数据集训练时, Epoch 为 50, 学习率为 0.01, BatchSize 为 1。STEEL 数据集: 选择了数据集中部分缺陷图像 N_{all} 作为样本, $N_{\text{all}}=\{300,750\}$, $N=\{0,10,50,150,300,\text{all}\}$, 此外无论 N_{all} 为多少, 都将选取全部的正样本图像, Epoch 为 90, 学习率为 0.1, BatchSize 为 1。

整个实验基于 Ubuntu16.04、Cuda10 和 Cudnn 7.6 的环境和 Python 3.8、Pytorch 1.9.0 和 torchvision 0.10.0 深度学习框架实验, 实验的硬件配置包括 1 块 GPU: GTX 2080Ti (显存: 11 GB)。

2.3 实验结果与对比

为了验证所提出方法的有效性, 本文在 3 个大规模数据集上进行了实验并和先进的方法进行比较。与本文比较的方法有 2020 年 Dong 等人提出的方法^[23]F-anoGan^[18]、2019 年 Bergmann 等^[16]提出的方法、2021 年 Bozic 等^[19]提出的方法。在 KSDD 数据集上如图 4(a), 本文方法在弱监督的情况下 AP 值达到 88.11%, 远超过目前文献 [16, 18] 提出的无监督算法; 仅需要 5 张带有像素级标注, 准确度便可达到 99.9%, 此时效果与文献 [19, 23] 提出的完全监督算法基本相同; 完全监督情况下, AP 为 100%。在 KSDD2 数据集上如图 4(b), 弱监督训练下本文的 AP 为 83.5%, 远超同类的所有方法; 在完全监督情况下, AP 为 94.9%, 高于同类方法。在 STEEL 数据集上如图 4(c), 文献 [19] 300 为当 $N=300$ 时文献 [19] 提出的方法结果。表 2 为 KSDD 数据集结果对比, 表 3 为 KSDD2 数据集结果对比, 表 4 为 STEEL 数据集结果对比。

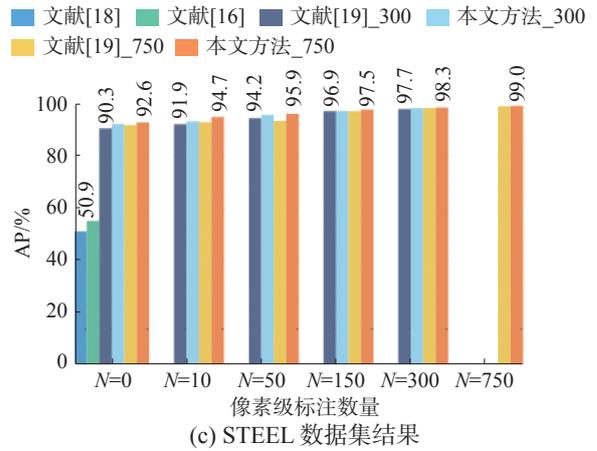
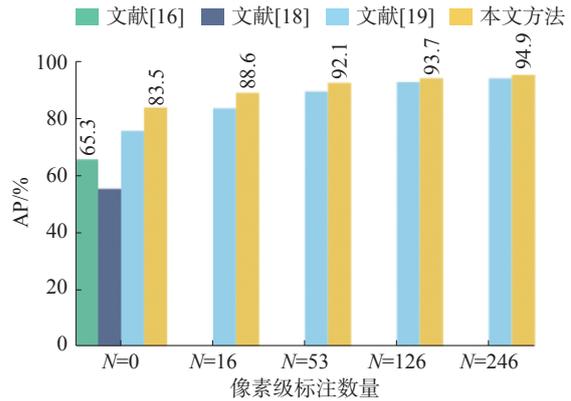
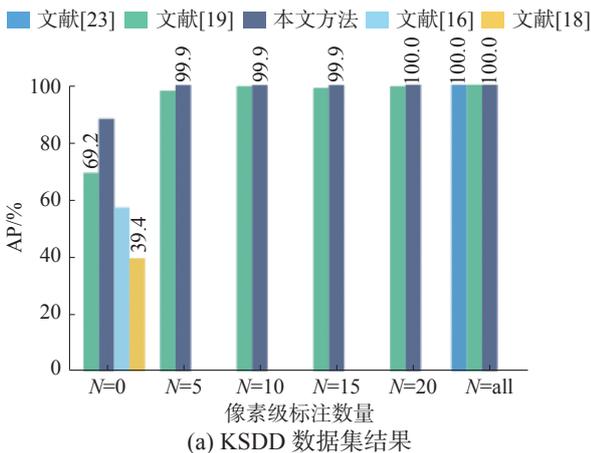


图 4 3 个数据集训练结果

Fig. 4 Three datasets training results

表 2 KSDD 数据集结果对比
Table 2 Comparison of KSDD dataset results

| 方法 | N=0 | N=5 | N=10 | N=15 | N=20 | N=all |
|--------|------|------|------|------|------|-------|
| 文献[16] | 57.1 | — | — | — | — | — |
| 文献[18] | 39.4 | — | — | — | — | — |
| 文献[23] | — | — | — | — | — | 100 |
| 文献[19] | 66.8 | 97.9 | 99.5 | 98.9 | 99.5 | 100 |
| 本文 | 88.1 | 99.9 | 99.9 | 99.9 | 100 | 100 |

表 3 KSDD2 数据集结果对比
Table 3 Comparison of KSDD2 dataset results

| 方法 | N=0 | N=16 | N=53 | N=126 | N=246 |
|--------|------|------|------|-------|-------|
| 文献[16] | 65.3 | — | — | — | — |
| 文献[18] | 55.0 | — | — | — | — |
| 文献[19] | 73.3 | 83.2 | 89.1 | 92.4 | 93.7 |
| 本文 | 83.5 | 88.6 | 92.1 | 93.7 | 94.9 |

表 4 STEEL 数据集结果对比
Table 4 Comparison of STEEL dataset results

| 数据集 | 方法 | N=0 | N=10 | N=50 | N=150 | N=300 | N=750 |
|--------------|--------|-------------|-------------|-------------|-------------|-------------|-------------|
| STEEL 300 | 文献[16] | 54.9 | — | — | — | — | — |
| | 文献[18] | 50.9 | — | — | — | — | — |
| | 文献[19] | 90.3 | 91.9 | 94.2 | 96.9 | 97.7 | — |
| | 本文 | 92.0 | 93.0 | 95.5 | 97.0 | 98.1 | — |
| STEEL 750 | 文献[19] | 91.6 | 92.7 | 93.2 | 96.9 | 98.1 | 98.8 |
| | 本文 | 92.6 | 94.7 | 95.9 | 97.5 | 98.3 | 99.0 |

弱监督情况下与无监督方法相比, AP 提高 1 倍左右, 其他情况下也有 2% 以上的提高。可以看出, 本文提出的方法和先进的方法相比有明显的提升。特别地, 由于多尺度模块和注意力模块的运用, 在 KSDD 数据集上取得了极佳的效果, 38% 的像素级标注便可以与完全监督的方法相媲美。综合以上分析, 可以看出本文的方法在混合监督的情况下取得的精确度均高于同类方法, 证明了本文方法的优越性。

2.3.1 融合模块的有效性

表 5 是在 KSDD 和 KSDD2 数据集上的消融研究, Mus 代指多尺度融合模块, Non-Local 为注意力机制模块, 对于 KSDD 数据集, 弱监督情况下, 相比较于 BaseLine, 各种方法都有很大提升; 其他标注情况下, AP 也达到了行业领先的结果。对于 KSDD2 数据集, 同时使用 2 个模块的时候效果最优, 弱监督情况下有 10% 的提高, 混合监督时有平均 3% 的提高, 完全监督情况也提升了 1% 左右。

表 5 在 KSDD 和 KSDD2 数据集上的消融实验
Table 5 Ablation experiments on KSDD and KSDD2 datasets

| 数据集 | 标注数量 | 对比方法 | Mus | Non-Local | Mus+Non-Local |
|-------|-------|------|------|-----------|---------------|
| KSDD | N=0 | 66.8 | 73.4 | 71.9 | 88.1 |
| | N=5 | 97.9 | 99.8 | 99.5 | 99.9 |
| | N=10 | 99.5 | 99.6 | 99.6 | 99.9 |
| | N=15 | 98.9 | 99.8 | 99.6 | 99.9 |
| | N=20 | 99.5 | 99.9 | 99.7 | 100 |
| | N=all | 100 | 100 | 99.9 | 100 |
| | KSDD2 | N=0 | 73.3 | 79.5 | 78.0 |
| N=16 | | 83.2 | 86.6 | 87.1 | 88.6 |
| N=53 | | 89.1 | 90.5 | 90.7 | 92.1 |
| N=126 | | 92.4 | 93.1 | 93.2 | 93.6 |
| N=246 | | 93.7 | 94.3 | 94.8 | 94.9 |

2.3.2 多尺度融合中层数选取的分析

在使用多尺度融合模块时, 一个很重要的问题是不同尺度的组合对于模型的结果有巨大的影响。如图 5 所示, 本文利用 KSDD2 数据集分别选取了不同层进行组合。本文将卷积层分为 4 个单元, 其中第 1 个单元以 1 代替, 图 5 横坐标 (1,4) 即为第 1 层和第 4 层组合。本文在不同尺度组合下模型效果如图 5 所示, 全监督训练情况下任何多尺度组合方式结果都比其他级别有更好的表现。同时, 全监督训练下不同尺度对结果影响很小。当对数据的像素级标注量为 16 (即 N=16) 或弱监督训练 (即 N=0) 时, 不同组合对结果产生的影响幅度最大, 其效果最差的尺度组合和效果最好的尺度组合会使 AP 相差近 11%。大部分情况下, 只需要融入 2 层组合层数便可以达到最好的效果, 随着融合层数的增加, 结果都有着不同幅度的下降, 可以明显看出尺度越多不仅不会提高精确度反而会造成一定程度的降低。

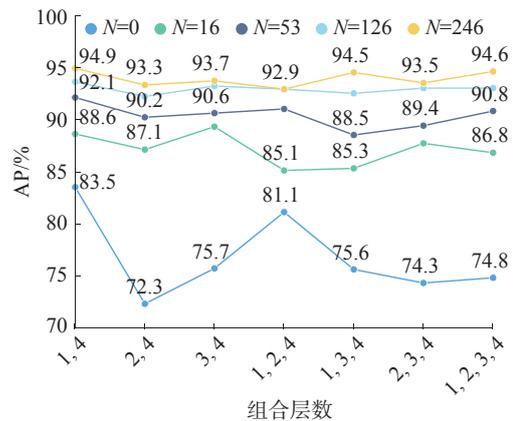


图 5 不同级别训练在不同尺度组合下的结果
Fig. 5 Results of different levels of training under different scale combinations

综上所述, 针对本文提出的多尺度特征融合模块, 发现: 1) 不同尺度融合效果的影响程度随着像素级标注数量的提高而降低; 2) 在相同训练条件下, 尺度数量与模型效果呈非线性。

2.3.3 像素级标注数量和准确值关系的分析

此外, 本文在 KSDD 和 KSDD2 数据集上研究像素级标注量对实验的影响如图 6, 对于 KSDD 数据集, 当 N=5 时, 即像素级标注总量仅为 9%, 但是此时的结果已经与完全监督效果相似。对于 KSDD2 数据, 当标注数量 N=16, 占总缺陷图像的 6% 时, AP 获得了幅度最大的提高。两个数据集都在最小标注的情况下即低于 10% 的像素级标注时获得幅度最大的提高。

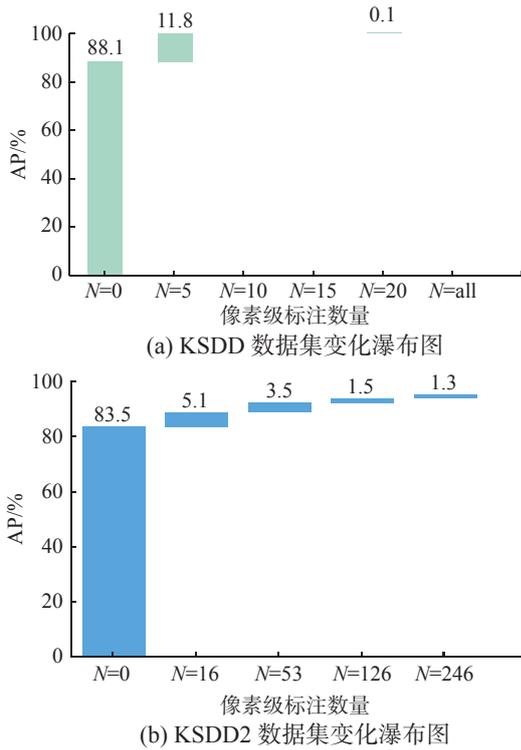


图 6 不同标注量对 AP 的提升

Fig. 6 Improvement of AP with different annotations

2.3.4 注意力机制模块有效性分析

Non-Local 注意力机制模块可以直接专注于范围内任意的 2 个点之间的联系, 不过分关注周围点的信息, 可以提高模型对于较长缺陷的识别能力, 如图 7 所示。本文针对 Non-Local 注意力机制模块的效果在 KSSD 和 KSSD2 数据集上进行了验证, 表 4 的结果显示该模块在弱监督情况下对模型的效果提升最明显, AP 比原始模型提高 5%。在完全监督的情况下, AP 也有 1% 的提升。

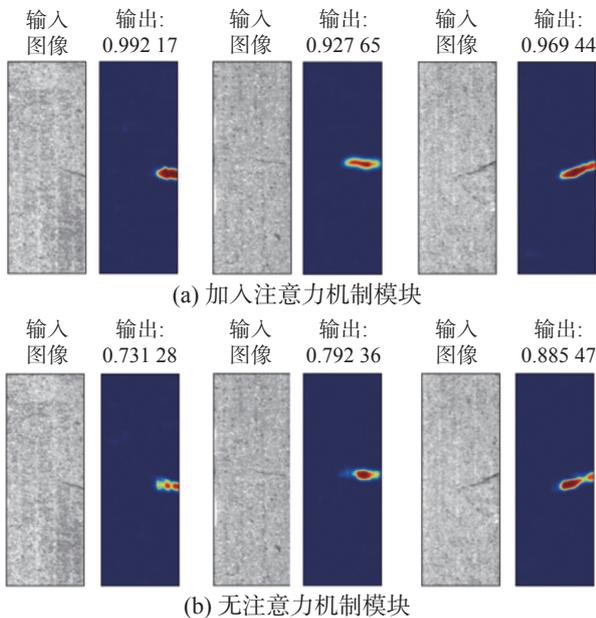


图 7 注意力机制的影响

Fig. 7 Influence of attention mechanism

3 结束语

本文关注基于多尺度和注意力机制的缺陷检测问题。针对当前缺陷目标形状各异和特征提取能力较差的问题, 设计了一个多尺度特征融合单元, 提高模型对于不同大小缺陷特征的提取能力, 在模型中融入了一个注意力机制单元, 加强模型对于缺陷特征的提取。为了进一步评估多尺度的作用, 本文设计了一个不同尺度组合实验, 评估当前模型下最合适多尺度组合。利用混合监督训练证明本模型对 10% 以内的数据进行像素级标注可以获得最大幅度的提高。在 3 个大规模数据集上的优异表现证明了本文方法的有效性。未来可以针对该模型的普适性做进一步探索。

参考文献:

- [1] 陶显, 侯伟, 徐德. 基于深度学习的表面缺陷检测方法综述 [J]. *自动化学报*, 2021, 47(5): 1017-1034.
TAO Xian, HOU Wei, XU De. A survey of surface defect detection methods based on deep learning[J]. *Acta automatica sinica*, 2021, 47(5): 1017-1034.
- [2] TABERNIK D, ŠELA S, SKVARČ J, et al. Segmentation-based deep-learning approach for surface-defect detection[J]. *Journal of intelligent manufacturing*, 2020, 31(3): 759-776.
- [3] KUMAR A. Computer-vision-based fabric defect detection: a survey[J]. *IEEE transactions on industrial electronics*, 2008, 55(1): 348-363.
- [4] ABDI H, WILLIAMS L J. Principal component analysis[J]. *Wiley interdisciplinary reviews: computational statistics*, 2010, 2(4): 433-459.
- [5] 虞祖耀, 王洪元, 张继. 基于机器视觉的织布瑕疵在线检测 [J]. *计算机工程与设计*, 2016, 37(10): 2851-2856.
YU Zuyao, WANG Hongyuan, ZHANG Ji. On-line detection of weaving defects based on machine vision[J]. *Computer engineering and design*, 2016, 37(10): 2851-2856.
- [6] FUKUSHIMA K, MIYAKE S. Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition[M]. Berlin: Springer Berlin Heidelberg, 1982: 267-285.
- [7] PARK J K, KWON B K, PARK J H, et al. Machine learning-based imaging system for surface defect inspection[J]. *International journal of precision engineering and manufacturing-green technology*, 2016, 3(3): 303-310.
- [8] KYEONG K, KIM H. Classification of mixed-type defect patterns in wafer Bin maps using convolutional neural networks[J]. *IEEE transactions on semiconductor manu-*

- facturing, 2018, 31(3): 395–402.
- [9] HE Yu, SONG Kechen, MENG Qinggang, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features[J]. *IEEE transactions on instrumentation and measurement*, 2020, 69(4): 1493–1504.
- [10] LI Feng, XI Qinggang. DefectNet: toward fast and effective defect detection[J]. *IEEE transactions on instrumentation and measurement*, 2021, 70: 1–9.
- [11] GE Ce, WANG Jing, WANG Jingyu, et al. Towards automatic visual inspection: a weakly supervised learning method for industrial applicable object detection[J]. *Computers in industry*, 2020, 121: 103232.
- [12] LI Qizhu, ARNAB A, TORR P H S. Weakly- and semi-supervised panoptic segmentation[M]. *Computer Vision - ECCV 2018*. Cham: Springer International Publishing, 2018: 106–124.
- [13] SALEH F, ALIAKBARIAN M S, SALZMANN M, et al. Built-in foreground/background prior for weakly-supervised semantic segmentation[M]. *Computer Vision-ECVCV 2016*. Cham: Springer International Publishing, 2016: 413–432.
- [14] MINHAS M S, ZELEK J. Semi-supervised anomaly detection using autoencoders[J]. *Journal of computational vision and imaging systems*, 2019: 03674.
- [15] BERGMANN P, LÖWE S, FAUSER M, et al. Improving unsupervised defect segmentation by applying structural similarity to autoencoders[EB/OL]. (2018-07-05)[2022-05-25]. <https://arxiv.org/abs/1807.02011>.
- [16] BERGMANN P, FAUSER M, SATTLEGGER D, et al. Uninformed students: student-teacher anomaly detection with discriminative latent embeddings[C]// *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2020: 4182–4191.
- [17] XIN Zihao, WANG Hongyuan, QI Pengyu, et al. Printed surface defect detection model based on positive samples[J]. *Computers, materials & continua*, 2022, 72(3): 5925–5938.
- [18] SCHLEGL T, SEEBOCK P, WALDSTEIN S M, et al. F-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks[J]. *Medical image analysis*, 2019, 54: 30–44.
- [19] BOZIC J, TABERNIK D, SKOCAI D. Mixed supervision for surface-defect detection: from weakly to fully supervised learning[J]. *Computers in industry*, 2021, 129: 103459.
- [20] MNIH V, HEES N, GRAVES A, et al. Recurrent models of visual attention[EB/OL]. (2014-06-24)[2022-05-25]. <https://arxiv.org/abs/1406.6247>.
- [21] XU K, BA J L, KIROUS R, et al. Show, attend and tell: neural image caption generation with visual attention[C]// *Proceedings of the 32nd International Conference on International Conference on Machine Learning-Volume 37*. New York: ACM, 2015: 2048–2057.
- [22] WANG Xiaolong, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]// *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2018: 7794–7803.
- [23] DONG Hongwen, SONG Kechen, HE Yu, et al. PGANet: pyramid feature fusion and global context attention network for automated surface defect detection[J]. *IEEE transactions on industrial informatics*, 2020, 16(12): 7448–7458.

作者简介:



孙博言, 硕士研究生, 主要研究方向为计算机视觉、缺陷检测。



王洪元, 教授, 主要研究方向为计算机视觉、模式识别和智能系统, 江苏省人工智能专委会常务理事, 常州市计算机学会副理事长, CCF常州副主席。主持国家自然科学基金项目3项, 获吴文俊人工智能科技进步三等奖。发表学术论文150余篇, 出版专著1部、教材3部。



刘乾, 硕士研究生, 主要研究方向为计算机视觉。