

DOI:10.13232/j.cnki.jnju.2023.03.005

## 基于 $U^2$ -Net 的金属表面缺陷检测算法

王昱翔<sup>1,2</sup>, 葛洪伟<sup>1,2\*</sup>

(1. 江南大学人工智能与计算机学院, 无锡, 214122; 2. 江苏省模式识别与计算智能工程实验室(江南大学), 无锡, 214122)

**摘要:** 金属表面缺陷检测旨在通过合理的算法判断工业生产中金属材料的表面图像中是否存在缺陷, 是计算机视觉领域在工业应用中的重要研究内容, 但现有的基于分割的金属表面缺陷检测算法存在抗干扰能力弱、容易背景误判和检测粒度不够细致等问题。针对具有各种干扰因子的金属表面图像, 提出一种基于  $U^2$ -Net 的金属表面缺陷检测算法 Attention- $U^2$ Net。首先, 为了解决实际生产中金属表面图像噪点过多导致背景误判和采样层获取信息不够细致的问题, 设计 U 型注意力编码模块, 可以在编码时增大缺陷区域权重的同时, 抑制背景噪声; 然后, 为了解决图像中存在的难检测样本和边界复杂问题, 设计具有权重的损失函数并结合多层次监督方法, 使检测算法更关注难检测样本和边界像素点, 提高预测准确度; 最后, 使用根据图像结果自动计算灰度阈值进行结果优化的算法, 最终生成高质量缺陷预测图。与六种常用的缺陷检测领域的像素分割方法在金属表面缺陷公开数据集上进行比较实验, 实验结果表明, 提出的算法的像素准确率、查准率、查全率、F-score、平均绝对误差和均方误差均取得了优良的结果, 证明该算法拥有较强的抗干扰因子能力, 最终缺陷预测图像像素准确率高, 边界明显。

**关键词:** 缺陷检测, 深度学习, 注意力编码, 图像优化, 图像分割

**中图分类号:** TP183

**文献标志码:** A

## Metal surface defect detection algorithm based on $U^2$ -Net

Wang Yuxiang<sup>1,2</sup>, Ge Hongwei<sup>1,2\*</sup>

(1. School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, 214122, China;

2. Jiangsu Provincial Engineering Laboratory of Pattern Recognition and Computational Intelligence, Jiangnan University, Wuxi, 214122, China)

**Abstract:** Metal surface defect detection aims to judge whether there are defects in the surface image of metal materials in industrial production through a reasonable algorithm, and is an important research content in the field of computer vision applied in industry. However, the existing segmentation-based metal surface defect detection algorithms still have problems such as weak anti-interference ability, easy background misjudgment and insufficient detection granularity. For metal surface images with various interference factors, this paper proposes a  $U^2$ -Net-based metal surface defect detection algorithm Attention- $U^2$ Net. First of all, in order to solve the problem of misjudgment of background and insufficient information obtained by sampling layer due to excessive noise of metal surface image in actual production, a U-shaped attention coding module is designed for suppressing background noise. Then, in order to solve the complex problems of difficult-to-detect samples and boundaries in images, a weighted loss function is designed and combined with a multi-level supervision method, so that the detection algorithm pays more attention to difficult-to-detect samples and boundary pixels and improves the prediction accuracy. Finally, an algorithm that automatically calculates the grayscale threshold according to the image results is used to optimize the results, and a high-quality defect prediction map is generated finally. By comparing with six pixel

基金项目: 国家自然科学基金(61806006), 江苏高校优势学科建设工程

收稿日期: 2023-03-17

\* 通讯联系人, E-mail: ghw8601@163.com

segmentation methods commonly used in the field of defect detection, the experimental results on the public data set of metal surface defects show that the algorithm has good performance. The experimental results show that the algorithm has strong anti-interference factor ability, and the final defect prediction image has high pixel accuracy and obvious boundaries.

**Key words:** defect detection, deep learning, attention coding, image optimization, image segmentation

制造业迅速发展,各种各样的工业产品不断涌现,其中金属材料制品是主要的加工产品之一。然而,在金属表面加工过程中,由于生产环境的局限性,会不可避免地产生表面缺陷,如果不及及时发现并处理将直接影响产品的表观质量和性能,严重时会导致产品的性能退化。因此,对金属表面缺陷检测算法的研究具有重要意义。

针对金属表面缺陷检测的问题,国内外学者进行了大量研究。现有的金属表面缺陷检测方法大体分两类:基于显示特征提取的传统方法<sup>[1-6]</sup>和基于自动特征提取的深度学习方法。然而,传统的金属表面缺陷算法无论是采用分类的方法还是采用分割的方法,均有对噪声干扰敏感、计算效率低和背景易与缺陷特征混杂等问题,无法有效地识别缺陷。基于深度学习的金属表面缺陷检测方法利用标注数据自动提取特征,避免了手工设计特征的困难,同时,此类方法抽象能力强,对同类目标的识别能力强。

基于深度学习的金属表面缺陷检测方法可分为三类:(1)基于整图分类网络的方法。2013年 Akram et al<sup>[7]</sup>首次将 DCNN (Dynamic Convolution Neural Network)应用在钢铁表面图像的分类任务中,其设计的深度卷积网络仅由单卷积层、单最大池化层和单全连接层组成,利用池化操作对池化前不同尺度的特征图进行金字塔池化操作,获得了固定大小的多尺度特征向量。Shang et al<sup>[8]</sup>提出一种铁轨缺陷的识别方法,首先利用 canny 算子提取图像边缘,定位铁轨区域,然后将处理后的图像利用预训练的 Inception V3 网络进行分类。Tao et al<sup>[9]</sup>提出三阶段级联深层神经网络,将三级目标检测网络的方法应用于金属表面的目标检测。(2)基于目标检测网络的方法。Zhang et al<sup>[10]</sup>提出一种基于旋转四边形的检测网络,对一种电连接器进行表面缺陷检测。Xue and Li<sup>[11]</sup>改进分类网络 GoogleNet,并将其作为 RFCN (Re-

gion-Based Fully Convolutional Networks)的骨干网络来检测钢轨表面裂纹。Chen et al<sup>[12]</sup>提出一种铁轨连接件检测的级联深层神经网络,分成三个阶段分别用于由粗到精的定位,再判断定位的目标是否存在缺陷。(3)基于像素分割网络的方法。Aslam et al<sup>[13]</sup>采用 U-Net 结构,使用基于像素分割的方法来检测金属表面的裂纹缺陷。Huang et al<sup>[14]</sup>采用 VGG16 作为骨干网络建立一种 FCN (Fully Convolutional Networks)网络,检测铁轨表面的裂纹和渗透两类主要缺陷。Han et al<sup>[15]</sup>首先利用 Fater R-CNN 的候选区与提取操作获得大量可能包含缺陷的图像块,然后利用分割网络 U-Net 进行分割,实现对多晶硅片上的缺陷的检测。这三类方法中,基于像素分割的方法可以获得缺陷详细的形态、尺寸和位置信息,在对金属产品质量进行分析与判断时有明显优势。

但像素分割方法应用在金属表面缺陷检测中依然存在一些问题,例如,复杂的实际场景中采集的金属表面图片难免会有大量复杂的噪点信息,导致算法在检测缺陷时会存在各种误判,造成错检和漏检。生产方式的多种多样使产生的表面缺陷和瑕疵没有固定的模式,导致一部分难检测样本和瑕疵的复杂边界问题,也使算法在应对此类瑕疵时性能大打折扣。另外,基于图像像素分割的方法普遍对正负样本的分布非常敏感,而缺陷检测领域中缺陷处的像素点数量远小于图片总像素点,所以实际应用中必须考虑对此的优化。

现阶段使用的像素分割方法中,U<sup>2</sup>-Net 是 Qin et al<sup>[16]</sup>受到 U-Net<sup>[17]</sup>启发设计的深度学习网络结构,在图像分割领域显著性目标检测任务中具有更好的性能和效果。该方法提出一种新的 RSU (Residual U-blocks)模块并使用嵌套的 U 型结构,和 U-Net 方法相比,大大提高了从任意分辨率的输入特征图中提取多尺度特征的能力。

鉴于像素分割方法在金属表面缺陷检测中的

优势和U<sup>2</sup>-Net在图像分割领域的良好表现,本文针对金属表面缺陷检测任务对U<sup>2</sup>-Net网络进行改进,提出基于U<sup>2</sup>-Net的金属表面缺陷检测算法Attention-U<sup>2</sup>-Net (Att-U<sup>2</sup>-Net). 对U<sup>2</sup>-Net的编码器进行优化,提出融合注意力机制的编码模块,在有效保留局部和全局上下文信息的同时,增加了缺陷区域的权重,使主干网络能更好地应对背景噪点问题,并获得多尺度的信息. 加入 $L_{wbce}$ ,赋予特征像素点不同的权重,使模型能更好地解决复杂边界、难检测样本和正负样本分布不均衡问

题. 最后,在生成预测图时,采用OTSU方法进行图像处理,按照图像每个像素点的灰度自适应调整决策的阈值来优化最终的预测结果.

## 1 相关工作

**1.1 U<sup>2</sup>-Net 基本理论** 2020年Qin et al<sup>[16]</sup>提出一种简单而强大的深度学习网络架构U<sup>2</sup>-Net用于显著目标检测(Salient Object Detection, SOD). U<sup>2</sup>-Net的网络结构如图1所示,是一个两层嵌套的U型结构.

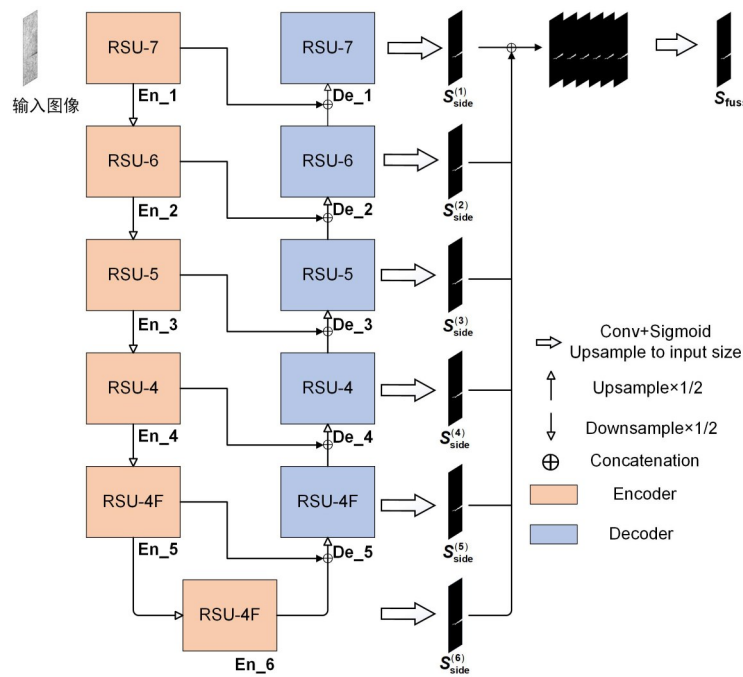


图1 U<sup>2</sup>-Net的网络结构

Fig. 1 The architecture of U<sup>2</sup>-Net

U<sup>2</sup>-Net与U-Net不一样的地方在于,它采用有交互的编码-解码器(Encoder-Decoder)结构,输入图像经过五次下采样并编码和五次上采样并编码处理,最后融合多级解码器输出缺陷概率预测图,得到缺陷概率预测结果. 底层的RSU采用RSU-F模块,能在不降低特征映射分辨率的情况下提取级内多尺度特征;顶层的RSU模块则具有类似U-Net的结构.

使用U<sup>2</sup>-Net分割图像有三个优势:能捕捉更多的上下文信息,提出的RSU模块可以融合不同尺度感受野的特征;这些RSU模块中使用了大量的池化操作,在增加整个架构深度的同时,没有显

著增加计算成本;该结构只建立在RSU模块上,没有使用任何特性分类来训练主干网络,比较灵活,可以适应不同的工作环境且性能损失很小.

**1.2 空间注意力和通道注意力** 2018年Hu et al<sup>[18]</sup>提出一种利用CNN中特征通道依赖性的创新网络SENet来进行图像识别. SENet显式地建模通道之间的相互依赖关系,自适应地重新校准通道的特征,使网络学习了通道之间的相关性,筛选针对通道的注意力,在稍微增加整个网络的计算量之后,识别效果有明显提升.

受到上述方法的启发,Badrinarayanan et al<sup>[19]</sup>研究网络架构中的注意力机制,提出CBAM

(Convolutional Block Attention Module) 模块. CBAM 模块包含通道注意力模块(Channel Attention Module, CAM)和空间注意力模块(Spatial Attention Module, SAM),其结构如图 2 所示.

CBAM 的通道注意力模块大体和 SENet 相同,主要区别是 CBAM 首先对输入特征图采取全局平均池化和全局最大池化,两种不同的池化意味着提取的高层次特征更丰富.接着,通过两个全连接层和相应的激活函数建模通道之间的相关性,合并两个输出得到各个特征通道的权重.最后,得到特征通道的权重之后,通过乘法逐通道加

权到原来的特征上,完成通道维度上的原始特征重标定,得到空间注意力调整的特征图.

CBAM 中的空间注意力模块和通道注意力模块相似,也是利用全局平均池化和全局最大池化,不同的是在通道这个维度上的操作,也就是把所有输入通道池化成两个实数,由  $(h \times w \times c)$  形状的输入得到两个  $(h \times w \times 1)$  的特征图.接着,用一个  $7 \times 7$  的卷积核,卷积后形成新的  $(h \times w \times 1)$  的特征图,将注意力模块特征图和输入的特征图相乘得到通道注意力调整的特征图.

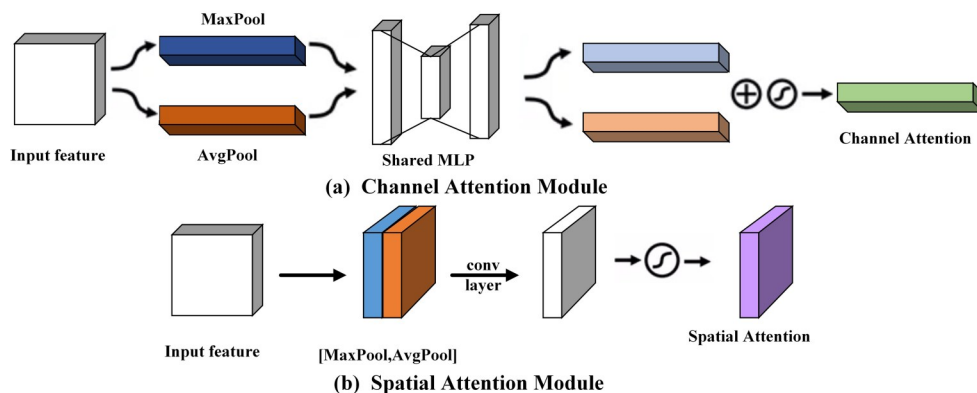


图 2 CBAM 的结构

Fig. 2 The architecture of CBAM

## 2 Attention-U<sup>2</sup>Net

**2.1 U 型空间注意力编码模块** 对于缺陷检测分割任务,局部和全局上下文信息都非常重要.现代的 CNN 设计,如 VGG<sup>[20]</sup>,ResNet<sup>[21]</sup>,DenseNet<sup>[22]</sup>等网络中,  $1 \times 1$  和  $3 \times 3$  的小型卷积滤波器由于其高效的计算性能是最常用的特征提取原件,但它们存在滤波器感受野太小而无法捕捉全局信息的缺点,所以底层输出特征图中只能包含局部特征.工业生产过程采集的金属表面图像的背景中易被误判为目标的背景噪声给缺陷检测任务带来严重的挑战,极大地影响模型的检测精度.视觉注意力机制的作用已在大量的计算机视觉任务中得到证实,能通过对显著区域赋予高的权重来帮助网络更加关注显著区域.

为了有效地保留局部和全局上下文信息并且增加缺陷区域的权重,受到 U<sup>2</sup>-Net 中 RSU 模块<sup>[16]</sup>

的启发,本文提出 USE (U-blocks Spatial-Attention Encoder) 编码模块.由于大多数的操作都应用于下采样的特征映射,所以这种结构的计算开销很小.

模型中使用的编码器模块细节如图 3 所示,对于编码器 En<sub>1</sub> 使用的 USE<sub>1</sub> 模块,输入的是原始输入图像,En<sub>i</sub> ( $i=2, 3, 4, 5$ ) 的 USE 模块输入的是上一层编码模块 C<sub>out1</sub> 的输出经过 1/2 采样处理之后的特征图.该编码器的内部结构类似 U<sup>2</sup>-Net 网络,首先通过输入卷积层将输入特征图转换成一个具有 C<sub>out</sub> 通道数的中间图  $F(x)$ ,这是提取局部特征的普通卷积层.然后,以中间图作为输入,学习提取和编码多尺度上下文信息  $U(F(x))$ .从梯度降采样特征图中提取多尺度特征,并通过逐步上采样、合并和卷积等方法将其编码到高分辨率特征图中,这一过程减少了大尺度直接上采样造成的细节损失.最终,通过逐像素



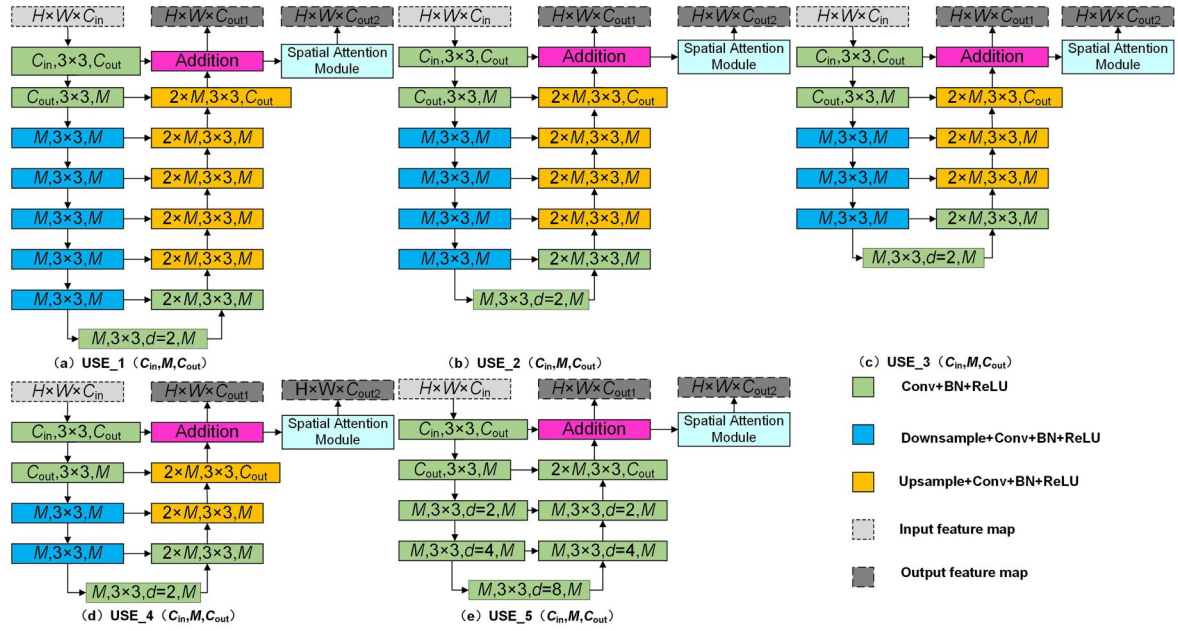


图3 U型空间注意力编码模块的结构

Fig. 3 The architecture of the U-blocks Spatial-Attention Encoder Module

求和  $F(x)$  和  $U(F(x))$  融合局部特征和多尺度特征得到特征图  $C_{out1}$ , 然后将  $C_{out1}$  通过空间注意力模块得到特征图  $C_{out2}$ .

将  $C_{out1}$  作为下层编码器的输入,  $C_{out2}$  作为同层对称解码器的输入.

由于编码器  $En_5$  处于网络结构的下层, 特征图经过上层编码器的多次下采样, 分辨率相对较低, 进一步对此特征图进行下采样会导致有用的上下文信息丢失. 因此, 在  $En_5$  的  $USE_5$  模块使用扩展卷积的方式代替合并和上采样的操作. 由图 3e 可以看到,  $USE_5$  中所有中间特征图和输入的特征图具有相同的分辨率.

**2.2 U型通道注意力编码模块**  $En_6$  作为网络中最底层的编码器使用 UCE (U-blocks Channel-Attention Encoder) 模块, 其与  $En_5$  具有相似的结构, 均使用扩展卷积的方式代替合并和上采样的操作来防止低分辨率下进行下采样导致细节信息的丢失, 具体结构如图 4 所示. 与  $USE_5$  不同的是, 低分辨率图像下特征图的空间信息更少, 空间注意力机制不能发挥较好的效果, 所以 UCE 中没有使用空间注意力模块, 而是使用了通道注意力模块, 增加低分辨率下通道特征的权重. 并且, UCE 模块只有一个输出图像  $C_{out1}$ , 和  $USE_5$  的输

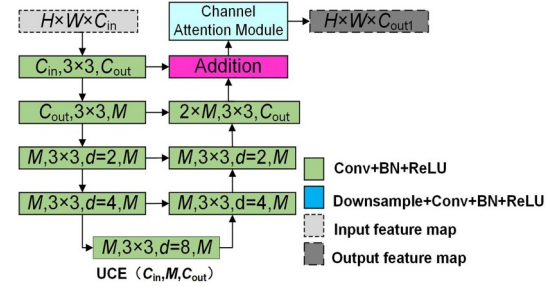


图4 U型通道注意力编码模块的结构

Fig. 4 The architecture of the U-blocks of Channel-Attention Encoder Module

出  $C_{out2}$  级联作为最下层解码器  $De_5$  的输入.

**2.3 OTSU 模块** OTSU 模块将模型输出的预测图  $S_{fuse}$  通过 OTSU 算法根据每个像素点的灰度计算合适的阈值, 并根据得到的阈值对预测图  $S_{fuse}$  进行二值化处理, 得到最终的预测图.

OTSU 算法又称为最大类间方差阈值分割算法, 由日本学者 Otsu 提出<sup>[23]</sup>, 是一种自适应的阈值确定的方法. 对于图像  $P(x, y)$ , 最佳二值化阈值记作  $T$ , 属于缺陷的像素点数占整幅图像的比例记作  $\omega_1$ , 其平均灰度为  $\mu_1$ ; 属于背景的像素点数占整幅图像的比例记作  $\omega_2$ , 其平均灰度为  $\mu_2$ ; 图像的总平均灰度记为  $\mu$ , 类间方差记为  $g$ , 如式(1)所示:

$$g = \omega_1 \times \omega_2 \times (\mu_1 - \mu_2)^2 \quad (1)$$

采用遍历 0~255 所有灰度的方法得到使类间方差最大的阈值  $T$ 。

**2.4 损失函数** 交叉熵损失函数广泛应用于目标检测,其计算如式(2)所示:

$$L_{bce}(P) = - \sum_{i=1}^H \sum_{j=1}^W \sum_{l=0}^1 t(G_{ij}=l) \log(Pr(P_{ij}=l|\varphi)) \quad (2)$$

其中,  $G_{ij}$  和  $P_{ij}$  为真值图  $G$  和预测图  $P$  在位置  $(i, j)$  的值.  $Pr(P_{ij}=l|\varphi)$  指给定所有参数  $\varphi$  的条件下  $P_{ij}=l$  的预测概率. 函数  $t(G_{ij}=l)$  如式(3)所示:

$$t(G_{ij}=l) = \begin{cases} 1, & G_{ij}=l \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

然而,在计算单个像素点的损失时,交叉熵给每个像素点赋予同样的权重,没有区分噪点和其他点对于检测目标的重要程度差异. 为了帮助网络感知边界和噪点并且更好地应对正负样本分布不均衡的情况,提升预测能力,本文使用  $\gamma$  权重来提高边界像素点的损失,将前四层解码器输出的

缺陷预测概率图  $S_{side}^{(1)}, S_{side}^{(2)}, S_{side}^{(3)}$  和  $S_{side}^{(4)}$  改进为  $L_{wbce}$ , 具体如式(4)所示:

$$L_{wbce}(P) = - \sum_{i=1}^H \sum_{j=1}^W \gamma \sum_{l=0}^1 t(G_{ij}=l) \log(Pr(P_{ij}=l|\varphi)) \quad (4)$$

最终的损失函数如式(5)所示:

$$L = \sum_{m=1}^M \omega_{side}^{(m)} L_{wbce side}^{(m)} + \omega_{side}^{(n)} L_{bce side}^{(n)} + \omega_{fuse} L_{bce(fuse)} \quad (5)$$

其中,  $L_{wbce side}^{(m)}$  ( $M=4$ ) 代表输出的缺陷预测概率图  $S_{side}^{(1)}, S_{side}^{(2)}, S_{side}^{(3)}$  和  $S_{side}^{(4)}$  的损失函数,  $L_{bce side}^{(n)}$  ( $N=6$ ) 代表输出的缺陷预测概率图  $S_{side}^{(5)}$  和  $S_{side}^{(6)}$  的损失函数,  $L_{bce(fuse)}$  是最后的融合输出缺陷预测概率图的损失函数.  $\omega_{side}^{(m)}, \omega_{side}^{(n)}$  和  $\omega_{fuse}$  是每个 loss 项的权重,本文中设置为 1.

**2.5 Attention-U<sup>2</sup>Net 网络结构** 本文提出的基于 U<sup>2</sup>-Net 的金属表面缺陷检测算法 Attention-U<sup>2</sup>Net 的结构如图 5 所示. 由图可见,网络结构由四部分构成:六级编码器、五级解码器、与解码器和最后一级编码器相连的预测图融合模型、OTSU 调整和归一化预测图模块. 模型对 U<sup>2</sup>-Net

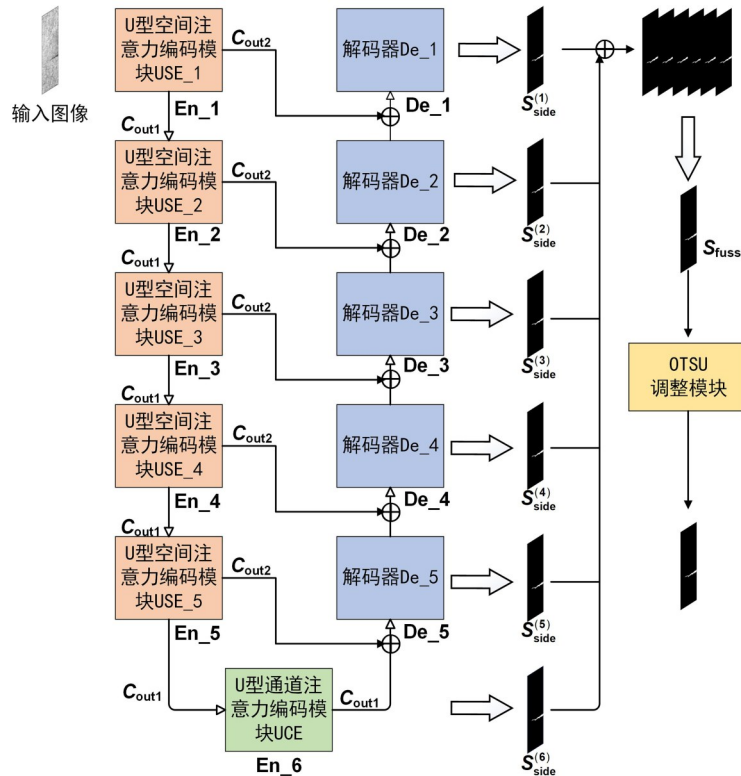


图 5 金属表面缺陷检测算法 Attention-U<sup>2</sup>Net 的网络结构

Fig. 5 The architecture of the Attention-U<sup>2</sup>Net Network

进行改进,使用U型空间注意力编码模块和U型通道注意力编码模块作为主干网络提取多层特征,最终生成概率映射图 $S_{fuse}$ 之后经过OTSU模块,通过预测图的每个像素点的灰度值计算合适的阈值,再对 $S_{fuse}$ 归一化处理得到最终的预测结果。

编码器En\_1,En\_2,En\_3,En\_4和En\_5分别使用USE\_1,USE\_2,USE\_3,USE\_4和USE\_5结构,En\_6使用UCE模块。由于在En\_5和En\_6中特征图的分辨率相对较低,进一步对这些特征图进行下采样会丢失有用的上下文信息,所以编码模块的结构较特殊,其具体结构见图3e和图4。USE模块的两个输出 $C_{out1}$ 和 $C_{out2}$ 分别作为下级编码模块的输入和同层对称解码器的输入。

解码阶段,五个解码器模块使用与U<sup>2</sup>-Net类似的RSU解码模块。每个解码器将来自前一级的上采样特征映射和来自其同层对称解码器的特征映射的级联作为输入,其内部具体实现如图6所示。

网络首先通过 $3 \times 3$ 卷积和Sigmoid函数中的En\_6,De\_5,De\_4,De\_3,De\_2和De\_5生成六个输出缺陷预测概率图 $S_{side}^{(6)}, S_{side}^{(5)}, S_{side}^{(4)}, S_{side}^{(3)}, S_{side}^{(2)}$ 和 $S_{side}^{(1)}$ 。然后,将输出的缺陷预测概率图向上采样至与输入图像大小一致,并通过级联操作相融合,然后通过 $1 \times 1$ 卷积层和一个Sigmoid函数,生成缺陷预测概率映射图 $S_{fuse}$ 。

OTSU调整模块将缺陷预测概率映射图 $S_{fuse}$ 作为输入,计算 $S_{fuse}$ 的自适应阈值,然后依靠此阈值对 $S_{fuse}$ 做灰度图像二值化处理得到最终的预测图。

### 3 实验结果与分析

**3.1 数据集和实验环境** 为了验证本文算法的有效性,使用Tabernik et al<sup>[24]</sup>提供和注释的缺陷生产项目公开数据集KolektorSDD。数据集包括399张图片,其中52张为有明显缺陷的图片,347张为无瑕疵图片,这些图片来源于真实的工业环境中捕捉的钢铁表面,均有像素级别标注。每张

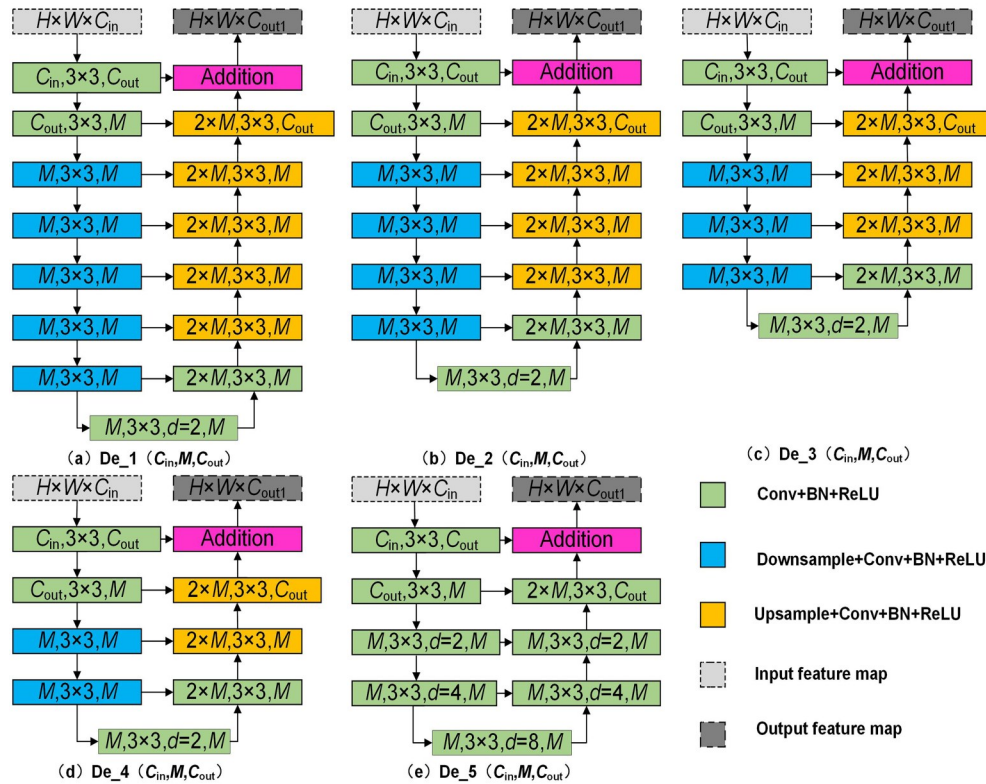


图6 解码器模块的结构

Fig. 6 The architecture of the Decoder Module



图片包含复杂的背景噪点信息,有很大的挑战性,在钢铁表面缺陷检测领域有很大的研究空间.实验环境为 python3.6, pytorch 1.10.0, GeForce RTX 2080.

**3.2 评价指标** 使用七项评价指标来度量模型的性能,包括平均像素的准确率(Accuracy,  $Acc$ )、查准率(Precision,  $P$ )、查全率(Recall,  $R$ )、 $F_\beta$ -score( $\beta=1, 2$ )、平均绝对误差(Mean Absolute Error,  $MAE$ )和均方误差(Mean Squared Error,  $MSE$ ). 由于本文及其他像素分割方法的评价指标均是根据像素级别的图像实验结果进行计算,所以下文中将平均像素的准确率、查准率和查全率简称为准确率、查准率和查全率.

平均像素的准确率、查准率和查全率可以评估预测图的像素级预测结果和真值图的差别,其计算方法如式(6)所示:

$$\begin{aligned} Acc &= \frac{TP + TN}{ALL} \\ P &= \frac{TP}{TP + FP} \\ R &= \frac{TP}{TP + FN} \end{aligned} \quad (6)$$

其中,  $TP, TN, FP, FN, ALL$  分别表示预测图中结果为预测为缺陷且正确、预测为背景且正确、预测为缺陷且错误、预测为背景且错误以及所有结果的像素点数量.

$F_\beta$ -score 是统计学中的概念,可以用来衡量二分率模型的精确度,兼顾了分类模型的查准率和查全率,和单一的评价指标相比,更能体现模型的整体性能.在像素分割任务中,常使用  $F_\beta$ -score 作为查全率和查准率的加权调和平均来统合评估模型的性能,其计算如式(7)所示:

$$F_\beta\text{-score} = \frac{(1 + \beta^2) \times P \times R}{\beta^2 \times P + R} \quad (7)$$

超参数  $\beta$  决定在合并查准率和查全率这两个评价指标时查全率所占的权重.  $\beta=0.5$  时,查准率比查全率更重要,  $F_\beta$ -score 变为  $F_{0.5}$ -score, 也称  $F_{0.5}$  分数;  $\beta=1$  时,查全率和查准率同等重要,即具有同等权重,  $F_\beta$ -score 变为  $F_1$ -score, 又称平衡  $F$  分数;  $\beta=2$  时,查全率比查准率更重要,  $F_\beta$ -score 变为  $F_2$ -score.

本文中,为了整体评价模型的性能和实际场景中的需求,将  $\beta$  分别设置为 1 和 2 进行实验对比,即使用  $F_1$ -score 和  $F_2$ -score 作为整体评价模型在实际任务中性能的评价指标.

$MAE$  和  $MSE$  评估预测图和真值图之间的像素级平均误差和均方误差,它们的值越小,说明误差越小,如式(8)和式(9)所示:

$$MAE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W |P(i, j) - G(i, j)| \quad (8)$$

$$MSE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (P(i, j) - G(i, j))^2 \quad (9)$$

其中,  $P$  和  $G$  分别为预测图和真值图,输入图的维度为  $H \times W$ ,  $P(i, j)$  为预测图中  $(i, j)$  处的像素灰度,  $G(i, j)$  为真值图中  $(i, j)$  处的像素灰度.

**3.3 参数设置** 使用 KolektorSDD 中的 299 张图片为训练集,剩下的 100 张图片为测试集来评估模型的性能.在数据扩增阶段,采用随机旋转  $90^\circ$ 、镜像翻转和添加噪点的方式将训练集中的图片扩增到 1000 张. Batchsize 设置为 6, 训练轮数为 100. 为了获得良好的实验效果,损失函数的超参数设置为  $\gamma=2$ .

**3.4 实验结果对比分析** 将本文基于  $U^2$ -Net 的金属表面缺陷检测算法与其他常用的图像像素分割方法进行对比,包括 U-Net<sup>[17]</sup>, Att-Unet<sup>[25]</sup> (Attention-Unet),  $U^2$ -Net<sup>[16]</sup>, NestedUNet<sup>[26]</sup>, Res-Unet++<sup>[27]</sup> (ResUnetPlusPlus) 和 R2U-Net<sup>[28]</sup>. 为了保证公平性,所有方法都使用相同的评估代码.

**3.4.1 定量评估** 表 1 给出了本文方法与其他六种方法在 KolektorSDD 数据集上不同指标的对比结果,表中黑体字表示结果最优.由表可见,本文方法的各项指标均超过其他方法,其平均像素的准确率、查准率、查全率以及  $MAE$  和  $MSE$  比其他方法都有显著提升.其他方法的  $MAE$  均高于 0.85,而本文方法直接将  $MAE$  降低到 0.277; 本文方法的平均像素准确率为 99.8%,其他算法均小于 99.3%; 本文方法的平均像素查准率为 79.4%,其他算法均小于 77.4%; 本文方法的平均像素查全率为 84.3%,其他方法中查全率最高的 Nested-UNet 仅有 82.3%,且其查准率为 73.7%,明显低于本文方法的 79.4%. 本文方法



表1 本文方法同其他六种方法在KolektorSDD数据集上的性能比较

Table 1 Performance of our method and six other methods on KolektorSDD dataset

|                     | <i>Acc</i>   | <i>P</i>     | <i>R</i>     | $F_1^-$<br>score | $F_2^-$<br>score | <i>MAE</i>   | <i>MSE</i>   |
|---------------------|--------------|--------------|--------------|------------------|------------------|--------------|--------------|
| U-Net               | 0.977        | 0.682        | 0.789        | 0.732            | 0.850            | 10.81        | 50.63        |
| Att-Unet            | 0.987        | 0.762        | 0.812        | 0.759            | 0.790            | 10.36        | 30.88        |
| U <sup>2</sup> -Net | 0.982        | 0.763        | 0.809        | 0.785            | 0.799            | 0.858        | 0.794        |
| Nested-Unet         | 0.989        | 0.737        | 0.823        | 0.778            | 0.804            | 10.10        | 10.23        |
| Res-Unet++          | 0.993        | 0.774        | 0.805        | 0.789            | 0.798            | 0.889        | 0.908        |
| R2U_Net             | 0.992        | 0.761        | 0.813        | 0.786            | 0.802            | 10.01        | 10.28        |
| 本文方法                | <b>0.998</b> | <b>0.794</b> | <b>0.843</b> | <b>0.812</b>     | <b>0.833</b>     | <b>0.277</b> | <b>0.001</b> |

和未改进的U<sup>2</sup>-Net方法相比,平均像素准确率从98.2%提升到99.8%,提升了1.6%,而MAE和MSE分别降低了0.581和0.793,性能的提升效果较明显。

**3.4.2 定性评估** 为了进一步验证本文方法的优良性能,图7展示了本文方法与其他方法的可视化对比。

图7a的绿框内为其他方法对多噪点无瑕疵表面预测的误判结果,红框中的内容为本文方法的误判结果。可以看出,U-Net和R2U\_Net在应

对此类图片时出现了大面积的误判现象,Att-Unet,U<sup>2</sup>-Net和NestedUNet存在明显的亮斑误判,ResUnet++在边界点出现亮斑误判,而本文方法仅在红框所示区域内有不明显的误判斑点。

图7b的绿框内为其他方法对多噪点不明显瑕疵表面预测的细节效果,红框中的内容为本文方法的细节效果。可以看出,U-Net未能有效识别此处的不明显裂纹缺陷;Att-Unet,U<sup>2</sup>-Net,NestedUNet和ResUnet++均检测出不明显裂纹缺陷,但对比mask图像,在绿框范围内均有像素不连续和细节边界不明显的缺点,也没有检测出裂纹下方的细纹分支;ResUnet++和R2U\_Net对裂纹末端的检测存在缺失;本文方法生成的图像有明显边界和良好的检测效果,而且,在红框标识范围内,对其他方法中没有检出的细纹分支和裂纹末端均有明显的检测效果。

**3.4.3 结果分析** 从定量评估和定性评估两个方面可以看出,本文方法和其他常用的像素分割方法相比,在检测金属表面缺陷时存在优势。定量评估中直观体现算法对数据集中瑕疵像素点检测性能的评价指标,如准确率、查准率和查全率,本文算法均有较大的优势。这是因为网络的前五层编码器使用了USE的结构,这种结构在引入空

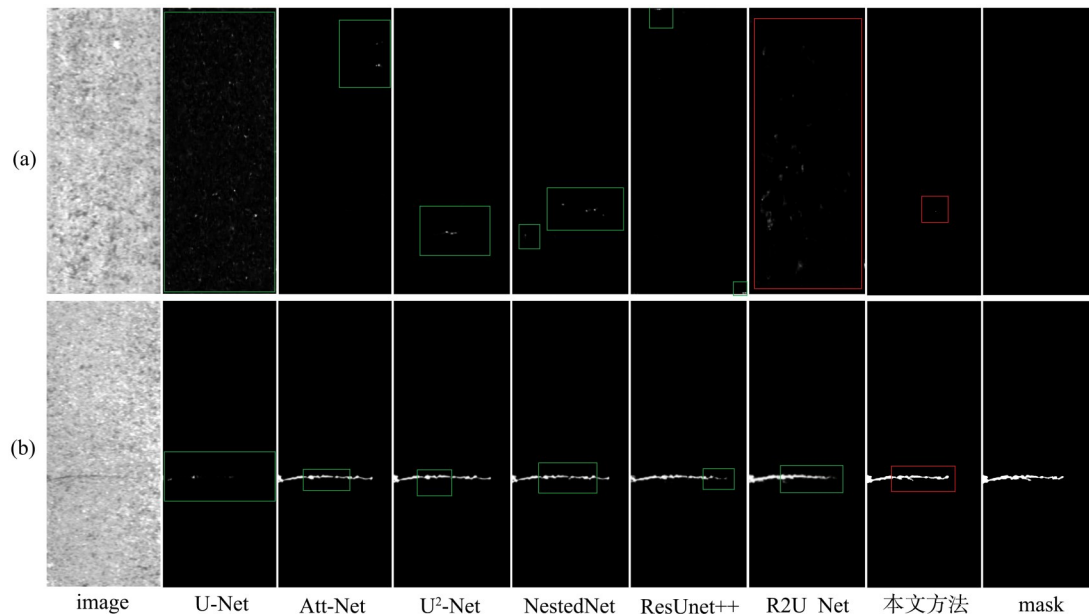


图7 本文方法与其他方法的可视化对比

Fig. 7 Visual comparison of our model with other methods

间注意力之后可以在图片分辨率较高的情况下更关注图像的空间信息,并利用这些空间信息细化特征;最后一层,空间信息经过前几层的下采样已经损失过多,编码器输入图像的分辨率较低,这里使用 UCE 的结构来引入通道注意力机制,有助于在提取特征时更关注输入图像的通道信息特征,特别是在分辨率较低时,图像的通道信息比损失过多的空间信息更重要. 本文方法在深层和浅层中引入不同的注意力机制,设计了两种结构不同的编码器,有利于算法在浅层和深层中获取更多的图像信息,减少信息损失,使算法在面对复杂噪点和边界较复杂的瑕疵时拥有更优秀的性能. 这一优点在定性评估中也可以看到,对瑕疵的检测,本文方法的结果更细致,误判情况也更少.

**3.5 消融实验** 为了验证各个模块的有效性,在  $U^2$ -Net 的基础上,先用改进的 U 型空间注意力编码模块替换原本的编码模块,再用 U 型通道注意力编码模块替换原本的编码模块. 实验结果如表 2 所示,表中黑体字表示结果最优. 由表可见,用 U 型空间注意力编码模块替换之后,算法的各项指标均有提升;用 U 型通道注意力编码模块替换之后,查准率和 MAE 有明显提升;最终加入 OTSU 改进之后,各项指标取得了最好的结果.

表 2 不同模块的消融实验

Table 2 Ablation study for different modules

| USE | UCE | OTSU | Acc          | P            | R            | $F_1^-$<br>score | $F_2^-$<br>score | MAE          | MSE          |
|-----|-----|------|--------------|--------------|--------------|------------------|------------------|--------------|--------------|
|     |     |      | 0.982        | 0.763        | 0.809        | 0.785            | 0.799            | 0.858        | 0.794        |
| ✓   |     |      | 0.987        | 0.765        | 0.823        | 0.792            | 0.811            | 0.370        | 0.346        |
| ✓   | ✓   |      | 0.992        | 0.790        | 0.820        | 0.805            | 0.814            | 0.347        | 0.135        |
| ✓   | ✓   | ✓    | <b>0.998</b> | <b>0.794</b> | <b>0.843</b> | <b>0.812</b>     | <b>0.833</b>     | <b>0.277</b> | <b>0.001</b> |

还可以看出,在  $U^2$ -Net 的基础上,使用本文设计的 USE 模块替换原本网络中前五层的编码器使算法的各项评价指标均有提升,其中查全率、MAE 和 MSE 的提升较明显. 这是因为用 USE 替换之后,加大了高分辨率图像,即网络浅层输入图像空间信息的权重. 和原网络相比,USE 模块使改进后的网络利用了额外的空间权重信息,并且因为不同层次的图像分辨率不同,使空间信息特征的尺度变多,有效融合之后提升了网络的性能,

也就提升了查全率、MAE 和 MSE.

在此基础上,使用本文设计的 UCE 模块替换原本网络中的第六层编码器,使查准率和 MSE 的提升较为明显,但查全率有略微的下降. 这是因为在低分辨率条件下空间信息会被较低的图像质量稀释,而通道信息的权重占主导因素,在深层网络中引入通道注意力机制的编码模块可以使网络在低分辨率条件下有效利用低分辨率图像的通道信息特征,使用 UCE 模块可以增加网络对于这类图像通道信息的权重,在之后的多尺度特征融合中增加了可以利用的信息特征尺度,使网络有效利用了通道特征,也提升了瑕疵点预测的查准率.

最终使用 OTSU 模块之后,算法性能达到最优. 在复杂的工业环境中,算法检测金属表面缺陷的性能一部分依赖采集的图像质量,而使用 OTSU 这种针对不同输入图像的自适应阈值确定的方法可以有效减轻输入图像整体质量对检测性能的影响. 由于实验使用的数据集中存在大量这类图像,所以本文方法引入 OTSU 模块后,算法性能被提升至最优,从表 2 的实验结果也可以直观地看出这一点.

**3.6 损失函数中超参数的调优实验** 为了使网络具有更好的性能,对损失函数中的超参数  $\gamma$  进行调优,超参数  $\gamma$  的作用是调节边界像素点损失权重,提升网络对边界与噪点的感知和更好地应对正负样本分布不均衡情况的能力. 由于超参数  $\gamma$  设置为 1 时,式(4)中的损失函数和式(2)中的交叉熵损失函数具有同等效果,所以当  $\gamma = 1$  时,损失函数改进后的实验结果等效于式(2)中的损失函数. 从表 3 可见,当  $\gamma = 2$  时,本文使用的损失函数效果最优(如表中黑体字所示),也优于交叉熵损失函数. 理论上,超参数  $\gamma$  的设置和本文使用的数据集相关,在边界更复杂的情况下,可以适当增加  $\gamma$  的大小,这会使算法的性能更加优秀.

## 4 结论

本文提出一种基于  $U^2$ -Net 的金属表面缺陷检测算法 Attention- $U^2$ Net. 为了解决实际生产中金属表面图像噪点过多导致算法背景误判和采样

表3 超参数 $\gamma$ 的消融实验Table 3 Ablation study for hyper-parameter  $\gamma$ 

| $\gamma$ | Acc          | P            | R            | $F_1^-$<br>score | $F_2^-$<br>score | MSE          | MAE          |
|----------|--------------|--------------|--------------|------------------|------------------|--------------|--------------|
| 1        | 0.993        | 0.754        | 0.799        | 0.776            | 0.789            | 0.672        | 0.003        |
| 2        | <b>0.998</b> | <b>0.794</b> | <b>0.843</b> | <b>0.812</b>     | <b>0.833</b>     | <b>0.277</b> | <b>0.001</b> |
| 3        | 0.995        | 0.805        | 0.749        | 0.775            | 0.760            | 0.378        | 0.005        |
| 4        | 0.993        | 0.534        | 0.893        | 0.668            | 0.787            | 10.38        | 0.013        |

层获取信息不够细致的问题,提出一种U型注意力编码模块,可以在编码采样时增大缺陷区域权重的同时,抑制背景噪声.为了解决图像中的难检测样本和边界复杂问题,设计了具有权重的损失函数,结合多层次监督方法,使算法更加关注难检测样本和边界像素点,提高预测准确度.最后,使用根据图像结果自动计算灰度阈值进行结果优化的方法,最终生成高质量缺陷预测图.大量的仿真和对比实验表明,提出的算法和U<sup>2</sup>-Net方法相比,平均像素准确率从98.2%提升到99.8%,提升了1.6%,MAE和MSE分别降低了0.581和0.793,提升效果明显.与其他算法相比,本文方法的各项指标均有显著提升,且应对复杂干扰因子、难检测样本和边界复杂等问题的效果较好.

虽然算法的各项指标均有提升,但它的网络结构嵌套仍然比较复杂,计算量和浅层深度学习网络相比还是较大.后续将改进深层的网络结构,优化编码器和解码器的数量,使其既能适应任意分辨率的图像也不需要复杂深层的网络结构.

#### 参考文献

- [1] Nand G K, Noopur, Neogi N. Defect detection of steel surface using entropy segmentation//2014 Annual IEEE India Conference. Pune, India: IEEE, 2014:1-6.
- [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] Ghorai S, Mukherjee A, Gangadaran M, et al. Automatic defect detection on hot-rolled flat steel products. IEEE Transactions on Instrumentation and Measurement, 2013, 62(3):612-621.
- [4] Liu Y, Xu K, Wang D D. Online surface defect identification of cold rolled strips based on local binary pattern and extreme learning machine. Metals, 2018, 8(3):197.
- [5] Niskanen M, Kauppinen H, Silven O. Real-time aspects of SOM-based visual surface inspection//Proceedings Volume 4664, Machine Vision Applications in Industrial Inspection X. San Jose, CA, USA: SPIE, 2002:123-134.
- [6] Yazdchi M, Yazdi M, Mahyari A G. Steel surface defect detection using texture segmentation based on multifractal dimension//2009 International Conference on Digital Image Processing. Bangkok, Thailand: IEEE, 2009:346-350.
- [7] Akram M W, Li G W, Jin Y, et al. CNN based automatic detection of photovoltaic cell defects in electroluminescence images. Energy, 2019(189): 116319.
- [8] Shang L D, Yang Q S, Wang J N, et al. Detection of rail surface defects based on CNN image recognition and classification//2018 20<sup>th</sup> International Conference on Advanced Communication Technology. Chuncheon, Korea (South): IEEE, 2018:45-51.
- [9] Tao X, Zhang D P, Ma W Z, et al. Automatic metallic surface defect detection and recognition with convolutional neural networks. Applied Sciences, 2018, 8(9):1575.
- [10] Zhang J B, Zhang Z T, Su H, et al. Quality inspection based on quadrangular object detection for deep aperture component. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(10):5938-5948.
- [11] Xue Y D, Li Y C. A fast detection method via region-based fully convolutional neural networks for shield tunnel lining defects. Computer-Aided Civil and Infrastructure Engineering, 2018, 33(8):638-654.
- [12] Chen J W, Liu Z G, Wang H R, et al. 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.
- [13] Aslam Y, Santhi N, Ramasamy N, et al. Localization and segmentation of metal cracks using deep learning. Journal of Ambient Intelligence and Humanized Computing, 2021, 12(3):4205-4213.

- [14] Huang H W, Li Q T, Zhang D M. Deep learning based image recognition for crack and leakage defects of metro shield tunnel. *Tunnelling and Underground Space Technology*, 2018(77):166—176.
- [15] Han H, Gao C Q, Zhao Y, et al. Polycrystalline silicon wafer defect segmentation based on deep convolutional neural networks. *Pattern Recognition Letters*, 2020(130):234—241.
- [16] Qin X B, Zhang Z C, Huang C Y, et al.  $U^2$ -Net: Going deeper with nested U-structure for salient object detection. *Pattern Recognition*, 2020(106):107404.
- [17] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation// *Proceedings of the 18<sup>th</sup> International Conference on Medical Image Computing and Computer Assisted Intervention*. Springer Berlin Heidelberg, 2015: 234—241.
- [18] Hu J, Shen L, Sun G. Squeeze-and-excitation networks// *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE, 2018: 7132—7141.
- [19] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481—2495.
- [20] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition// *The 3<sup>rd</sup> International Conference on Learning Representations*. San Diego, CA, USA: ICLR, 2014: 234—241.
- [21] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition// *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE, 2016: 770—778.
- [22] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks// *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE, 2017: 2261—2269.
- [23] Otsu N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems Man & Cybernetics*, 2007, 9(1): 62—66.
- [24] Tabernik D, Šela S, Skvarč J, et al. Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing*, 2020, 31(3): 759—776.
- [25] Oktay O, Schlemper J, Le Folgoc L, et al. Attention U-Net: Learning where to look for the pancreas. 2018, arXiv:1804.03999.
- [26] Zhou Z W, Siddiquee M M R, Tajbakhsh N, et al. UNet++: A nested U-Net architecture for medical image segmentation// *The 4<sup>th</sup> International Workshop on Deep Learning in Medical Image Analysis*. Springer Berlin Heidelberg, 2018: 3—11.
- [27] Jha D, Smedsrud P H, Riegler M A, et al. ResUNet++: An advanced architecture for medical image segmentation// *2019 IEEE International Symposium on Multimedia*. San Diego, CA, USA: IEEE, 2019: 225—255.
- [28] Alom M Z, Yakopcic C, Hasan M, et al. Recurrent residual U-Net for medical image segmentation. *Journal of Medical Imaging*, 2019, 6(1): 014006.

(责任编辑 杨可盛)