

DOI:10.19651/j.cnki.emt.2314513

## 多级解码神经网络用于滚珠丝杠点蚀检测\*

赵慧锋<sup>1,2</sup> 李铁军<sup>1</sup>

(1. 沈阳化工大学装备可靠性研究所 沈阳 110142; 2. 沈阳化工大学机械与动力工程学院 沈阳 110142)

**摘要:** 由于滚珠丝杠点蚀区域小,环境干扰严重,缺陷难以及时检测。所以提出了一种多级解码神经网络,实现滚珠丝杠点蚀缺陷的分割。该网络由编码器、多级解码器和多尺度注意力模块组成。编码器由 Resnet34 组成,并引入 Ghost 模块构建了轻量化的多级解码器。为了融合多尺度特征并过滤冗余信息,设计了多尺度注意力模块。采用二值交叉熵函数,IOU 和 SSIM 函数组成的混合损失函数训练网络。在滚珠丝杠缺陷数据集上做了实验,多级解码神经网络在  $\max F_{\beta}$  指标上达到了 0.770 3,与其他方法相比,该网络取得了更好的分割结果,并且单张图片处理时间为 26 ms。为滚珠丝杠点蚀缺陷实时分割提供了一种新的方法。

**关键词:** 滚珠丝杠;缺陷检测;神经网络;图像分割

**中图分类号:** TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.2040

## Multi-level decoding neural network for pitting detection of ball screw

Zhao Huifeng<sup>1,2</sup> Li Tiejun<sup>1</sup>

(1. Equipment Reliability Institute, Shenyang University of Chemical Technology, Shenyang 110142, China;

2. School of Mechanical and Power Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China)

**Abstract:** Due to the small pitting area of the ball screw and the serious environmental interference, defects are difficult to detect in time. Therefore, a Multi-level decoding neural network is proposed to realize the segmentation of pitting defects in ball screws. The network consists of an encoder, a multi-level decoder and a Multi-scale Attention module. The encoder is composed of Resnet34, and the Ghost module is introduced to build a lightweight multi-level decoder. In order to fuse multi-scale features and filter redundant information, the Multi-scale Attention module is designed. A hybrid loss function composed of BCE function, IOU and SSIM function is used to train the network. Experiments on the ball screw defect dataset show that Multi-level decoding neural network achieves 0.770 3 in the  $\max F_{\beta}$  metrics, compared with other methods, which achieves better segmentation results, and the processing time of a single image is 26 ms. It provides a new method for real-time segmentation of ball screw pitting defects.

**Keywords:** ball screw; defect detection; neural network; image segmentation

## 0 引言

滚珠丝杠是机械传动系统的重要组成部分。在滚珠丝杠长时间运行过程中,点蚀缺陷会在某一时刻作为表面的单一小故障开始出现。然后滚珠会通过点蚀缺陷产生周期性的冲击,会加剧滚珠丝杠副的振动,最终导致传动系统失效。因此,及时检测滚珠丝杠上的点蚀缺陷对于保证机械传动系统的正常运行非常重要。当前针对滚珠丝杠缺陷检测大都基于震动信号做分析来判断是否出现故障<sup>[1]</sup>。但是震动信号在实际生产中容易受到各种干扰,且无法判断缺陷的具体位置和大小。采用视觉检测的方法可以定位具体的缺陷位置。但是传统视觉方法<sup>[2]</sup>依赖于低级特征或启发

式先验,由于光照、复杂纹理、油渍等因素的变化,在一些复杂环境中无法准确识别缺陷。

为了实现滚珠丝杠点蚀缺陷的快速分割,采用了显著性目标检测(salient object detection, SOD)技术<sup>[3]</sup>。SOD是计算机视觉中的一项关键技术,它可以分离出最显著的区域。SOD可应用于各种领域,如图像分割<sup>[4-5]</sup>、缺陷检测等。Song等<sup>[6]</sup>提出了一种端到端的钢铁缺陷分割网络。王盛等<sup>[7]</sup>改进 U-2Net 实现了太阳能电池片的缺陷分割。孙耀泽等<sup>[8]</sup>提出了轮对踏面损伤分割方法。但是当前还没有针对滚珠丝杠的点蚀缺陷分割的方法。而现有的分割方法如 U-Net<sup>[9]</sup>,BASNet<sup>[10]</sup>等,虽然在常见的数据集上取得

收稿日期:2023-09-02

\* 基金项目:国家自然科学基金(52275156)、辽宁省重点(一般)项目教育部(LJKZ0435)项目资助

了不错的结果,但面对工业缺陷,如滚珠丝杠小的点蚀缺陷时无法取得令人满意的分割效果,也无法满足实时检测的需求。虽然也有一些小而快的模型<sup>[11]</sup>,它只是通过更多地关注低分辨率的特征图,来确保速度。而且由于多次卷积计算,高层特征图丢失了很多细节,并不能很好的检测小尺度的目标。Chen 等<sup>[12]</sup>采用逐级融合策略来应对不同大小的目标,然而这些方法在不同尺度上重构信息的能力有限,而且简单的融合方法会引入冗余信息。Wang 等<sup>[13]</sup>通过添加注意力机制来过滤冗余信息,但每层的注意力层只能基于固定的尺度考虑当前信息的重要性,无法结合全局信息来度量不同尺度目标的重要程度,而且多次使用复杂的注意力机制增加了模型的计算量。

为了提升 SOD 方法在滚珠丝杠点蚀缺陷检测上的准确性和实时性,提出了多级解码神经网络(multi-level decoding neural network, MLDNet)。该网络与常规的 SOD 模型相比,在滚珠丝杠缺陷检测上具有更高的检测精

度和更快的检测速度。

## 1 MLDNet 结构

### 1.1 多级解码网络结构

MLDNet 由三部分组成:编码器、多级解码器和多尺度注意力(multi-scale attention, MSA)模块。整体网络结构如图 1。多级解码器由初级解码器和精炼解码器两部分组成。初级解码器初步解码缺陷区域,精炼解码器融合初级解码器和编码器的信息对缺陷区域进一步细化提升缺陷区域检测精度。引入了 Ghost 模块<sup>[14]</sup>来构建轻量化的多级解码器,减少模型参数,满足实时检测的需求。针对多层信息难以融合的问题,提出了 MSA 模块,它可以融合多个尺度的特征并过滤融合后的冗余信息,减少对小缺陷区域信息的干扰。在编码器部分为了保证运行的效率选择 Resnet34<sup>[15]</sup>作为编码器来保证模型适度的大小和合理的特征提取能力。

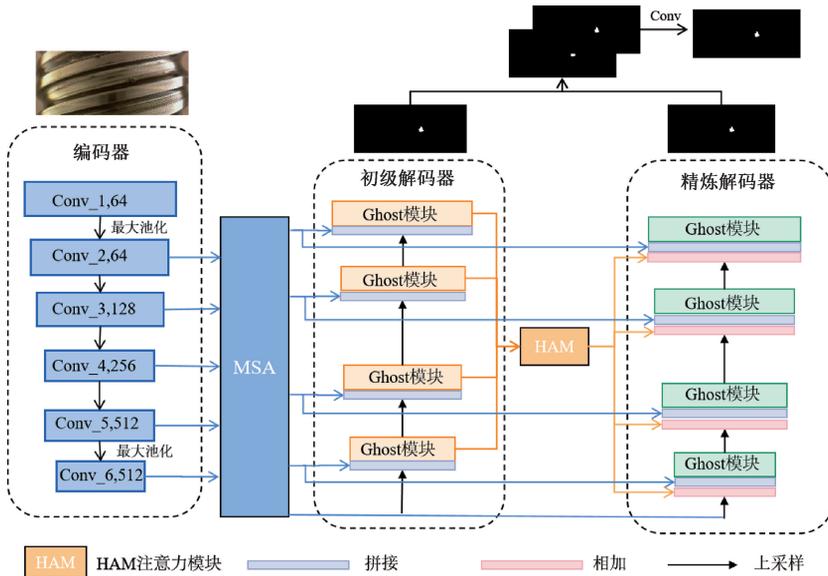


图 1 多级解码神经网络结构

### 1.2 多尺度注意力

传统的编码解码结构没有考虑不同尺度信息的融合。在特征提取的过程,浅层特征图包含了大量的细节信息,但是缺乏足够的定位信息。为满足滚珠丝杠大量小的点蚀缺陷检测的需求,并能够根据不同大小的缺陷特征过滤冗余信息,于是设计了 MSA 模块,网络结构如图 2 所示。常规的融合方式如 FPN<sup>[16]</sup>,将高层特征图经过多次的上采样与低层特征图融合,这样会造成高层信息流失。并且经过多次上采样的高层特征图会对低层特征图的信息造成干扰。而该方法首先将  $X_5$  与  $X_1$ 、 $X_2$  分别融合,然后将  $X_5$  与  $X_4$ 、 $X_3$  分别融合。采用跨尺度融合的方法将高层特征图直接与低层特征图融合,减少信息流失。在特征图融合后添加了一个混合注意力模块<sup>[17]</sup>(hybrid attention

module, HAM) 过滤融合后特征图的冗余信息。

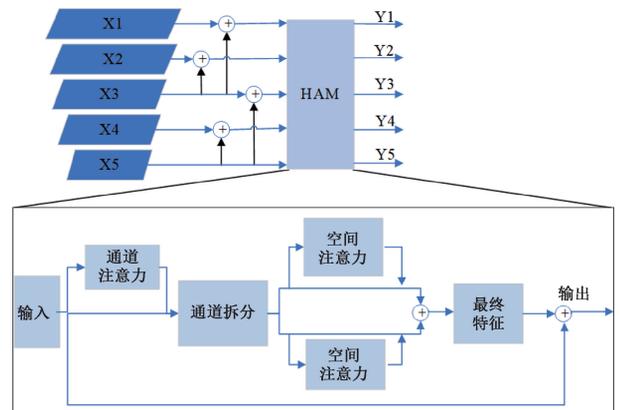


图 2 MSA 模块结构

HAM 兼顾通道与空间的注意力,并减缓注意力机制的复杂性,实现了性能和模型复杂度之间的良好权衡。融合后的五层特征图共享同一个 HAM 注意力模块,不仅可以学习不同尺度的信息而且可以减少参数量。

### 1.3 多级解码器

常规的解码方法如 U-Net 采用单次解码的方法,特征利用不充分。为了进一步利用不同阶段的信息细化特征,添加了精炼解码器。精炼解码器将编码器和初级解码器提取的特征进行全面的融合,融合后的特征对于不同阶段的信息做了进一步学习和细化。最后融合不同阶段的解码特征得到最终的结果。精炼解码器会先将初级解码器获得的多层特征经过一个 HAM 注意力机制过滤冗余信息,然后将初级解码特征和本层的特征相加融合,最后与编码器提取的特征进行拼接,再次解码特征。通过多级解码的方式可以进一步融合不同特征,解码得到更精细的目标区域。由于普通的卷积存在大量计算冗余,为了减少参数量多级解码器由 Ghost 轻量化模块组成。Ghost 模块是一种轻量化结构可以在保证精度的同时减少模型的参数量,结构如图 3。

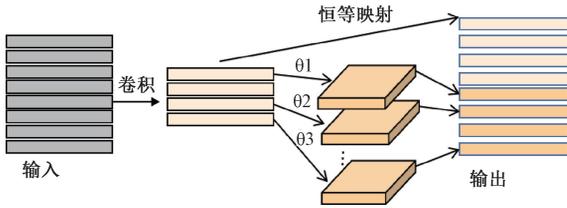


图 3 Ghost 模块结构

Ghost 模块首先采用少量的  $1 \times 1$  大小的普通卷积生成通道数小于原始特征的内在特征。然后使用廉价操作如深度卷积,根据内在特征生成更多特征,最后融合两部分的特征生产新的特征图。

### 1.4 损失函数

为了获得高质量的分割区域和清晰的边界,损失函数由三部分,如式(1)所示。

$$L = L_{BCE} + L_{IOU} + L_{SSIM} \quad (1)$$

式中: $L_{BCE}$ 、 $L_{SSIM}$ 、 $L_{IOU}$  分别表示二值交叉熵(binary cross entropy, BCE)损失、SSIM 损失和 IOU 损失。

BCE 是二值分类和分割中应用最广泛的损失。定义如式(2)所示。

$$L_{BCE} = -(y \log(x) + (1 - y) \log(1 - x)) \quad (2)$$

SSIM 捕获图像中的结构信息,可以用于图像质量评估计算方式如式(3)所示。

$$L_{SSIM} = 1 - \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3)$$

其中, $\mu_x$ 、 $\mu_y$  和  $\sigma_x$ 、 $\sigma_y$  分别为图像的均值和标准差。 $C_1$  和  $C_2$  是小的正常数,以避免除零。

IOU 是度量两个集合的相似性而提出的,作为和分割的标准评价度量方式如式(4)所示。

$$L_{IOU} = 1 - \frac{\sum_{i=1}^H \sum_{j=1}^W S(i,j)G(i,j)}{\sum_{i=1}^H \sum_{j=1}^W (S(i,j) + G(i,j) - S(i,j)G(i,j))} \quad (4)$$

其中, $G(i,j) \in \{0,1\}$  表示像素的真值标记, $S(i,j)$  表示作为显著目标的预测概率。

## 2 实验结果

### 2.1 实验设置

为了验证提出的模型的有效性,选择在 BSDData<sup>[18]</sup> 滚珠丝杠缺陷数据集上进行了实验。利用 Pytorch 框架实现的网络模型。使用 Adam 优化器来训练网络,权重衰减为  $1 \times 10^{-4}$  初始学习率为 0.001。在训练过程中,每个图像首先调整为  $256 \times 256$ ,然后随机裁剪为  $224 \times 224$  大小。

### 2.2 评估指标

为了根据标签评价预测图的质量,采用了 SOD 中常用的两个评价指标:F-measure ( $F_\beta$ )<sup>[2]</sup> 和平均绝对误差(mean absolute error, MAE)<sup>[19]</sup>。MAE 计算预测像素和真实值的平均绝对误差,如式(5)所示。

$$MAE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W |S(i,j) - G(i,j)| \quad (5)$$

$F_\beta$  用于综合衡量精确度和召回率,计算方式如式(6),通常取最大值( $\max F_\beta$ ), $\beta^2$  在实验中选择默认值 0.3。

$$F_\beta = \frac{(1 + \beta^2) \times Precision \times Recall}{\beta^2 \times Precision + Recall} \quad (6)$$

### 2.3 实验结果对比分析

将提出的模型和其他先进的 SOD 模型(U-Net<sup>[9]</sup>、BASNet<sup>[10]</sup>、R3Net<sup>[20]</sup>、PoolNet<sup>[21]</sup>、U-2Net<sup>[22]</sup>) 在滚珠丝杠缺陷数据集上做了检测对比,用论文中提供的默认参数做了实验。

为了评估分割显著性对象的质量,图 4 中展示了模型检测结果的可视化对比。可以看到由于缺陷区域非常小,其他常规的分割方法虽然能找到缺陷的大致位置,但只能获得一个模糊的缺陷区域,分割的边缘比较模糊。本文提出的方法 MLDNet,分割出的缺陷区域更加准确,轮廓也更加清晰。

表 1 总结了不同模型的  $\max F_\beta$  和 MAE 指标在滚珠丝杠缺陷数据集上的实验结果。

从表 1 可知与其他模型对比 MLDNet 的  $\max F_\beta$  取得最大值达到了 0.7703,并且 MAE 最小。同时 MLDNet 具有最快的检测速度,单张图片处理时间为 26 ms。

为了验证各个模块的有效性,在滚珠丝杠数据集上做了消融实验。每次移除一个模块做了对比,以验证每个模块的作用,实验结果如表 2 所示。

从表 2 可知经过多级解码器进一步融合后的特征比采用单级解码器的效果要好,并且加入 MSA 后模型的性能也有提升,证明了 MSA 模块的效果。最后加上混合损失函数比单纯采用 BCE 可以取得更优的结果。

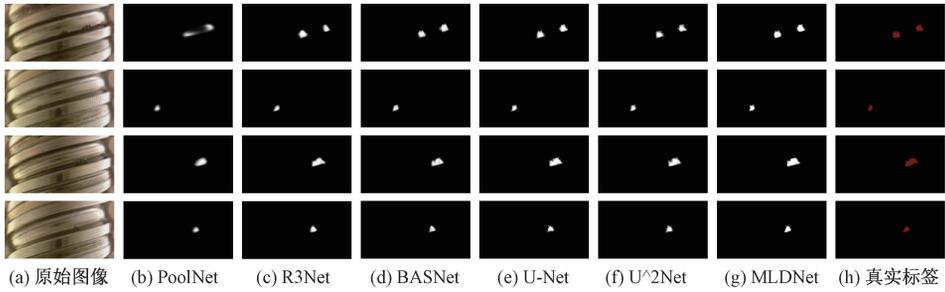


图 4 检测结果可视化对比

表 1 不同模型对比结果

| 模型      | MaxF <sub>β</sub> ↑ | MAE ↓           | 时间/ms     |
|---------|---------------------|-----------------|-----------|
| PoolNet | 0.322 7             | 0.008 28        | 37        |
| R3Net   | 0.408 0             | 0.002 83        | 32        |
| BASNet  | 0.746 1             | 0.000 65        | 74        |
| U-Net   | 0.763 7             | 0.000 62        | 34        |
| U^2Net  | 0.760 2             | 0.000 82        | 29        |
| MLDNet  | <b>0.770 3</b>      | <b>0.000 61</b> | <b>26</b> |

注:箭头向上表示越大越好箭头向下表示越小越好。

表 2 消融实验结果

| 结构            | MaxF <sub>β</sub> | MAE             |
|---------------|-------------------|-----------------|
| 单级解码+BCE      | 0.762 7           | 0.001 26        |
| 多级解码+BCE      | 0.767 2           | 0.001 01        |
| 多级解码+MSA+BCE  | 0.768 6           | 0.000 93        |
| 多级解码+MSA+混合损失 | <b>0.770 3</b>    | <b>0.000 61</b> |

### 3 结 论

针对实际生产中滚珠丝杠点蚀缺陷难以检测的问题,提出了 MLDNet。采用 Resnet34 作为编码器提取特征,引入 Ghost 模块构建解码器。设计了 MSA 模块融合多尺度特征。采用了 BCE,SSIM 和 IOU 组成的混合损失函数引导网络获得高质量的分割区域。在滚珠丝杠缺陷数据集进行了实验,MLDNet 与常规的分割方法相比 maxF<sub>β</sub> 和 MAE 取得最优的结果,同时单张图片检测时间为 26 ms。实验证明了本文提出的方法的有效性,为滚珠丝杠点蚀缺陷实时检测提供了一种新方法。

### 参考文献

[1] 聂从辉,周光,刘迪一,等. HHT-SVM 在滚珠丝杠副疲劳点蚀失效诊断中的应用[J]. 组合机床与自动化加工技术,2020(12):80-84,89.  
 [2] ZHU W, LIANG S, WEI Y, et al. Saliency optimization from robust background detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Columbus, OH, USA, 2014: 2814-2821.

[3] ACHANTA R, HEMAMI S, ESTRADA F, et al. Frequency-tuned salient region detection [C]. 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009:1597-1604.  
 [4] 张平,佟昆宏,王学珍. 基于改进 U-net 网络的液压管路分割方法[J]. 电子测量与仪器学报,2023,37(1): 123-129.  
 [5] 邸拴虎,杨文瀚,廖苗,等. 基于 RA-Unet 的 CT 图像肝脏肿瘤分割[J]. 仪器仪表学报,2022,43(8):65-72.  
 [6] HE Y, SONG K, MENG Q, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features [J]. IEEE Trans. Instrum. Meas. 69(2020):1493-1504.  
 [7] 王盛,吴浩,彭宁,等. 改进 U^2-Net 的太阳能电池片缺陷分割方法[J]. 国外电子测量技术,2023,42(2):177-184.  
 [8] 孙耀泽,高军伟. 基于 DeepLabV3+ 的轮对踏面损伤分割算法[J]. 电子测量技术,2022,45(23):113-118.  
 [9] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [J]. Medical Image Computing and Computer-Assisted Intervention MICCAI 2015, 2015, Volume 9351.  
 [10] QIN X, ZHANG Z, HUANG C, et al. BASNet: Boundary-aware salient object detection [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, 2019:7471-7481.  
 [11] WU Z, SU L, HUANG Q. Cascaded partial decoder for fast and accurate salient object detection[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Long Beach, CA, USA:3902-3911.  
 [12] CHEN S, TAN X, WANG B, et al. Reverse attention for salient object detection [C]. Computer Vision-ECCV 2018. Springer International Publishing, Cham: 236-252.  
 [13] WANG T, ZHANG L, WANG S, et al. Detect globally, refine locally: A novel approach to saliency

- detection [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Salt Lake City, UT;3127-3135.
- [14] HAN K, WANG Y, TIAN Q, et al. GhostNet: More features from cheap operations[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020: 1577-1586.
- [15] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Las Vegas, NV, USA, 2016: 770-778.
- [16] LIN T, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, 2017: 936-944.
- [17] LI G, QI F, ZHA L, et al. HAM: Hybrid attention module in deep convolutional neural networks for image classification[J]. Pattern Recognition, Volume 129, 2022, 108785, (n. d.).
- [18] SCHLAGENHAUF T, LANDWEHR M. Industrial machine tool component surface defect dataset [C]. Date in Brief, 2021.
- [19] PERAZZI F, KRAHENBUHL P, PRITCH Y, et al. Saliency filters: Contrast based filtering for salient region detection [C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Providence, RI, 2012:733-740.
- [20] DENG Z, HU X, ZHU L, et al. R3Net: Recurrent residual refinement network for saliency detection, in: proceedings of the twenty-seventh international joint conference on artificial intelligence, international joint conferences on artificial intelligence organization [J]. Stockholm, Sweden, 2018;684-690.
- [21] LIU L, HOU Q, CHENG M, et al. A Simple pooling-based design for real-time salient object detection[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, 2019;3912-3921.
- [22] QIN X, ZHANG Z, HUANG C, et al. U<sup>2</sup>-Net: Going deeper with nested u-structure for salient object detection[C]. Pattern Recognition, 2020.

#### 作者简介

赵慧锋, 硕士, 主要研究方向为计算机视觉。

E-mail: 1506225973@qq.com

李铁军(通信作者), 博士, 主要研究方向为机械系统热流固耦合特性, 智能制造。

E-mail: tiejunli2018@qq.com