

基于通道混洗和注意力机制的轻量缺陷检测算法

黄轶凡¹, 黄之文², 金涛¹

¹上海理工大学光电信息与计算机工程学院, 上海

²上海理工大学机械工程学院, 上海

收稿日期: 2023年12月13日; 录用日期: 2023年12月31日; 发布日期: 2024年1月31日

摘要

缺陷检测在制造业中扮演着确保产品质量和效率的关键角色。由于缺陷通常没有固定形状, 并且受光照影响较大, 因此基于图像的高精度缺陷检测成为一项极具挑战性的任务。本文针对缺陷图像的特点, 提出了一种结合了通道混洗和注意力机制的U型结构的卷积神经网络。首先在跳跃连接中加入注意力门机制来提高网络的泛化能力; 其次, 针对上采样中容易产生边缘轮廓失真的问题, 采用了混合损失计算; 最后, 因使用了效率更高的编码解码器, 在模型参数和浮点运算数较低的情况下, 所提出的模型MIoU和F1指标达到了91.89, 94.67和64.82, 72.93。与FCN、U-Net、U-Net++进行了比较, 结果表明所提出的方法在表面缺陷检测领域优于相关方法。

关键词

表面缺陷检测, 图像分割, U型卷积神经网络, 注意力机制, 混合损失计算

A Lightweight Defect Detection Algorithm Based on Channel Shuffling and Attention Mechanism

Yifan Huang¹, Zhiwen Huang², Tao Jin¹

¹School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

²School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Dec. 13th, 2023; accepted: Dec. 31st, 2023; published: Jan. 31st, 2024

文章引用: 黄轶凡, 黄之文, 金涛. 基于通道混洗和注意力机制的轻量缺陷检测算法[J]. 建模与仿真, 2024, 13(1): 976-985. DOI: 10.12677/mos.2024.131094

Abstract

Defect detection plays a crucial role in ensuring product quality and efficiency in manufacturing. Due to the irregular shapes of defects and their susceptibility to lighting variations, image-based high-precision defect detection becomes a highly challenging task. In response to the characteristics of defect images, this paper proposes a U-shaped convolutional neural network with a hybrid loss calculation and attention mechanism. Firstly, an attention gate mechanism is introduced in the skip connections to enhance the network's generalization ability. Secondly, to address the issue of edge contour distortion during up-sampling, a hybrid loss calculation is employed. Lastly, leveraging more efficient encoding and decoding layers, the proposed model achieves MIoU and F1 scores of 91.89, 94.67 and 64.82, 72.93, respectively, with lower model parameters and floating-point operations. Comparative analysis against FCN, U-Net, and U-Net++ demonstrates the superior performance of the proposed method in the field of surface defect detection.

Keywords

Surface Defect Detection, Image Segmentation, U-Shaped Convolutional Neural Network, Attention Mechanism, Hybrid Loss Calculation

Copyright © 2024 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

表面缺陷检测在产品质量评估中扮演关键角色，是工业自动化中的重要组成部分。随着计算机视觉的快速发展，自动化计算机视觉检测成为提高工业自动化水平的主要手段。然而，大多数基于知识的缺陷检测算法[1] [2]，涉及特征提取和分类，存在准确性不足和灵活性差的问题。

近年来，深度学习模型，特别是以 CNN [3]为代表的模型，已成功应用于多个计算机视觉领域，包括表面缺陷检测。这些深度学习方法通过分类、对象检测和语义分割满足了工业需求。

基于分类算法如 VGG [4]能够区分是否含有缺陷，但无法定位缺陷位置；对象检测算法如 RCNN [5]能够定位缺陷但难以获取缺陷的几何位置；而基于深度学习的语义分割算法如 FCNs [6]、U-Net [7]、U-Net++ [8]提供了精细的分割信息，提高了检测准确性。近年涌现的深度学习算法进一步提高了语义分割的效果，使得缺陷检测可以更精确地在特定语义区域进行。

如今，图像分割技术被深度学习模型的广泛使用所掩盖，深度学习模型提供了更好的分割结果。Badrinarayanan 等人提出了一种基于全卷积网络(FCN)的名为 SegNet [9]的架构，该架构使用一种新颖的反池化技术来执行非线性上采样。

比如目前比较流行 Liang-Chieh Chen 等人提出了 DeepLab v3+ [10]，在编码器下采样 4 倍后，会有一个分支连接到解码器部分，而解码器的输出前会有一个直接 4 倍双线性插值操作，该网络的一些设计思想，如多尺度信息的捕获和融合，对裂缝检测的网络模型设计具有一定的参考价值。

为了使神经网络具备强大的特征表示能力，目前已经提出了几种多尺度和多层次特征提取的裂缝检测新方法[11]，这些方法均使用特征金字塔模块作为多尺度特征提取器，在多种分辨率条件下捕获丰富的上下文信息。此外，Fei 等人[12]提出了一种用于 3D 沥青路面裂缝检测的 CrackNet-V 高效深度神经网络，

使用多个小卷积核(3×3)的卷积层来增加网络结构的深度, 以此在不增加额外参数的基础上提高准确性和计算效率。

鉴于工业数据获取困难, 使得传统语义分割难以应用于表面缺陷检测, 我们采用 U-Net 结构, 通过少量样本训练即可获得高性能。为了兼顾效率和准确性, 我们选择 ShuffleNet Unit [13]块作为语义分割模型的编码器和解码器的基础结构, 以提高效率。同时, 在 U-Net 级联通道上加入注意力门机制[14], 以重点关注相关区域的激活。这项工作的主要贡献包括提高语义分割模型的效率和精度, 增强对不同尺度物体的分割鲁棒性, 并引入注意力机制以加强相关区域的学习。

2. 模型方法

本小节首先介绍结合了 ShuffleNet Unit、通道混洗和注意力门机制的改进 U-Net, 然后描述了 ShuffleNet Unit 和注意力门机制的具体结构。

2.1. 总体分割网络结构

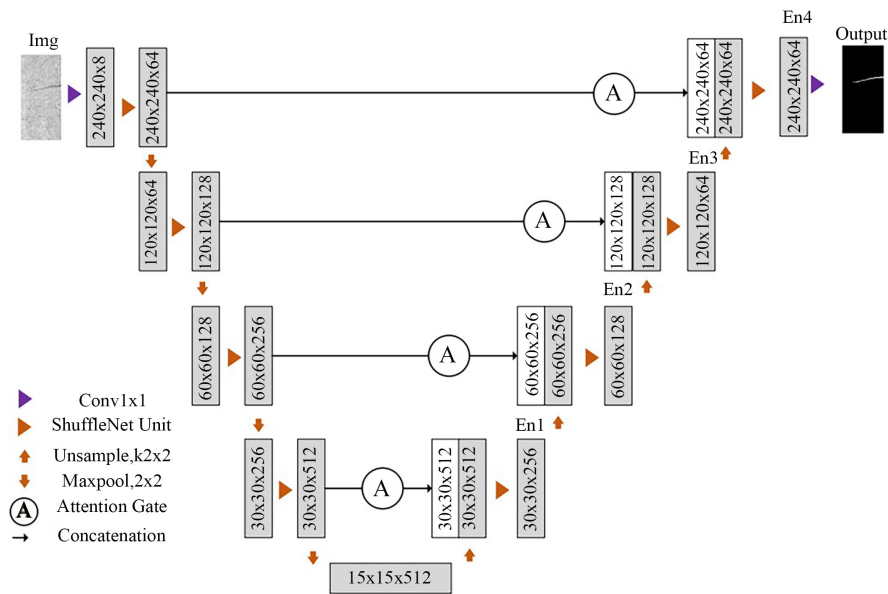


Figure 1. Enhanced U-Net architecture (illustrated using a 240×240 resolution image input as an example). Each square or cube represents an input image, feature map, or output prediction mask. The size of the input image, feature map, or output prediction mask is indicated within the respective square or cube

图 1. 改进 U-Net 架构(以 240×240 分辨率图像输入为例)。每个正方形或立方体代表一个输入图像、特征图或输出预测掩模。输入图像、特征图或输出预测掩模的大小标记在正方形或立方体的内部

图 1 体现了本文的分割结构, 结合 ShuffleNet Unit 和注意力门机制, 旨在提高分割精度的同时保持效率。ShuffleNet Unit 中的 GConv (Group convolutions)有效减少了传统 CNN 所需的密集计算量。然而, 由于组之间不共享特征图, 每个卷积核只能看到一部分输入特征, 降低了输出特征集合的表达能力。为了对抗 GConv 导致的组间特征不通的问题, 我们在 GConv 计算后对输出进行通道混洗处理, 使得接下来的组卷积可以在每个组输出的部分通道集合上进行计算。

需要注意的是, ShuffleNet Unit 在编码器-解码器架构中, 编码器和解码器中相应位置的 ShuffleNet Unit 块类型应尽可能一致, 以确保更好的编码器和解码器兼容性。此外, 我们采用 U-Net 结构的跳跃连

接方式，并在最大池化和反卷积之间引入了注意力门机制。注意力门有助于抑制图像中不相关区域的特征激活，从而降低误报。这种方法既保留了来自编码部分的上下文信息，又引入了多个路径进行反向传播梯度，有助于减轻梯度消失问题。

2.2. ShuffleNet Unit 和通道混洗模块结构

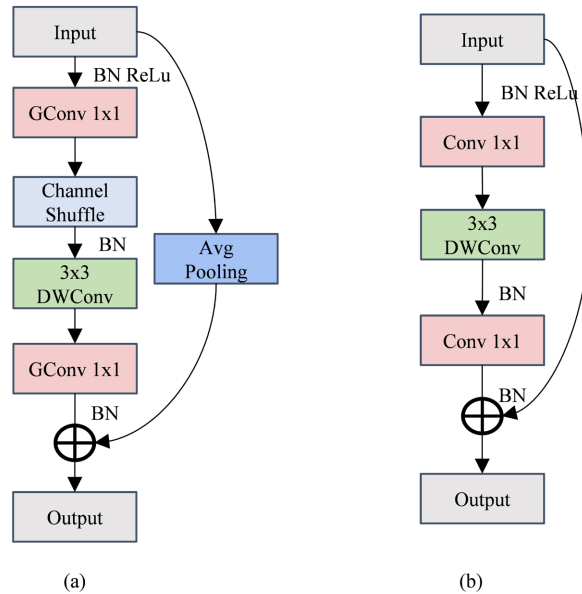


Figure 2. The ShuffleNet Unit structure (a) as the ShuffleNet Unit module, (b) As the ResNet convolutional block

图 2. ShuffleNet Unit 结构(a)为 ShuffleNet Unit 模块, (b)为 ResNet 卷积块

ShuffleNet Unit 如图 2 所示，是一个由 5 层组成的残差单元：首先是 1×1 分组卷积，接着添加了一个通道混洗操作，然后经过 3×3 深度可分离卷积，采用 1×1 卷积将输入映射成与输出相同的尺寸，最后包含 3×3 平均池化的短路连接，将输入直接加到输出上。需要注意的是，对原始输入采用 3×3 平均池化，以获得与输出相同大小的特征图，然后将该特征图与输出进行连接，而不是直接相加。这样的设计旨在降低计算量和参数大小。

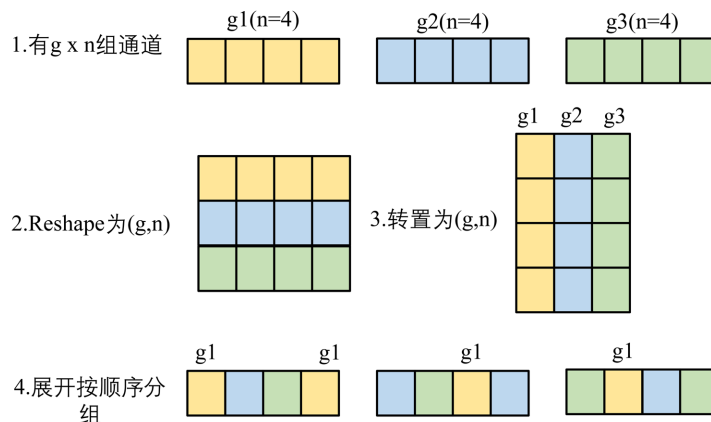


Figure 3. Channel shuffle module

图 3. 通道混洗模块

通道混洗结构, 如图 3 所示, 在小型网络中, 逐点卷积的昂贵性导致了有限通道的严格约束, 这会显著损失精度。为了解决这个问题, 一个直接的方法是引入通道稀疏连接。通过确保每个卷积操作仅在相应的输入通道组上进行, 组卷积可以显著降低计算损失。然而, 如果多个组卷积堆叠在一起, 会产生一个副作用: 某个通道的输出仅来自一小部分输入通道, 这降低了通道组之间的信息流通, 从而减弱了信息表示能力。为了解决这个问题, 我们允许组卷积能够获取不同组的输入数据, 使得输入和输出通道之间形成全连接。具体而言, 对于上一层输出的通道, 我们可以进行混洗(Shuffle)操作, 将其分成几个组合并到下一层, 从而有效解决上述副作用。

假设输入特征图的形状是, 深度可分离卷积模块的通道数是 m , g 代表组的个数, 那么 ResNet 卷积块的计算量如下所示:

$$1*1*c*m*h*w + 3*3*m*m*h*w + 1*1*m*c*h*w = s = hw(2cm + 9m^2) \tag{1}$$

ShuffleNet Unit 卷积块计算量如式(2)

$$(1*1*c*m*h*w)/g + 3*3*m*h*w + (1*1*m*c*h*w)/g = hw(2cm/g + 9m) \tag{2}$$

可见, ShuffleNet unit 需要的计算量最小。

2.3. 注意力门机制

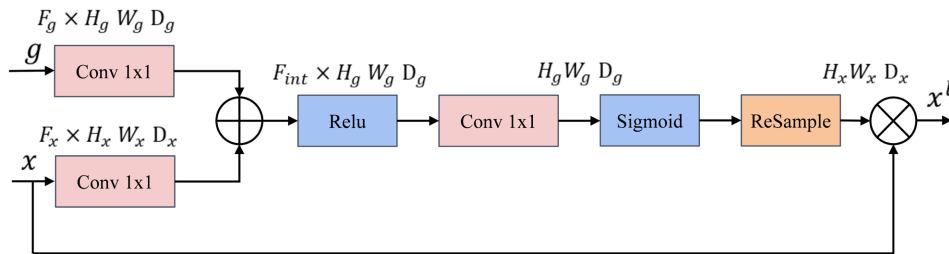


Figure 4. Attention gate structure
图 4. 注意力门的结构

我们所提出的模型中, 图 4 展示了注意力门的具体表示。在图像处理中, 注意力机制致力于聚焦于特定区域, 将其他区域置于忽略之中, 类似于人类视觉注意力的运作方式。就像人的视觉可以集中在某一点或区域, 同时抑制周围区域。通过引入注意力门, 我们能够抑制图像中与任务无关的区域特征激活, 从而有效减少误报的可能性。在当前研究中, 这一机制有助于推动模型参数在与缺陷分割相关的空间区域进行更新。具体实现上, 跳跃连接通过注意力门将编码器与相应的解码器连接起来, 如图 1 所示。注意力门接收两个输入, 一个来自相应的编码器, 包含该层中的所有上下文和空间信息, 另一个输入是来自下面解码器层的门控信号。注意力门的输出随后传递给解码器进行连接。我们所提出的模型中, 图 4 展示了注意力门的具体表示。

2.4. 改进损失函数计算

交叉熵 - Dice 混合损失函数是一种将交叉熵损失函数和 Dice 损失函数结合在起使用的损失函数。交叉熵损失函数用于评估实际输出与期望输出之间的接近程度, 而 Dice 损失函数用于评估两个样本的相似性。通过将这两个损失函数进行混合, 可以综合考虑模型分类准确性和相似性。定义训练数据集为 $D = \{X^n, Y^n\}$, 这里的 $X_n = \{x_i^{(n)}, i = 1, \dots, M\}$ 是输入的图像, $Y_n = \{y_i^{(n)}, i = 1, \dots, M\}$ 而是缺陷标注图, M

表示每个图像中的像素数量，其损失函数定义如式 2：

$$L_n = -\frac{1}{N} \sum_{i=1}^M \left(y_i^{(n)} \cdot \log(F(x_n)) + (1 - y_i^{(n)}) \cdot \log(1 - F(x_n)) + \text{DiceLoss} \right) \quad (3)$$

其中 L_n 是每层解码器的输出， $F(x_n)$ 为模型预测样本为缺陷的概率。

Dice 损失函数是一种用于评估两个样本相似度的度量函数，其定义为：

$$\text{DiceLoss} = 1 - \left(\frac{2 * |Y \cap P|}{|Y| + |P|} \right) \quad (4)$$

其中， Y 和 P 分别表示模型的输出和期望的输出， $Y \cap P$ 表示 Y 和 P 之间的交集。通过将交叉熵损失函数和 Dice 损失函数进行混合，可以综合考虑模型分类准确性和相似性，从而更好地训练模型。

3. 实验

3.1. 数据集

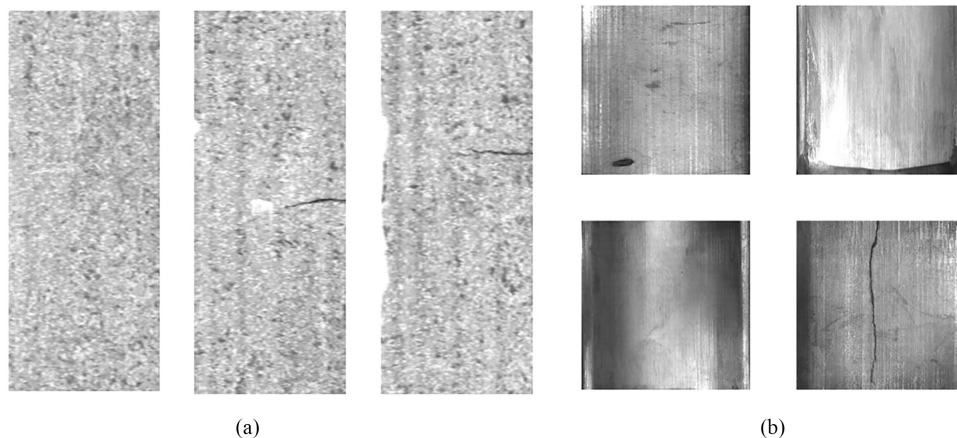


Figure 5. Defect examples (a) Kolektor surface defect dataset, (b) Industrial magnetic tile defect dataset
图 5. 缺陷样例(a)Kolektor 表面缺陷数据集(b)工业磁瓦缺陷数据集

图 5 为本文缺陷数据集示意图，本文使用了如下两个数据集：

1) 一个名为 Kolektor 表面缺陷数据集(Kolektor SDD)的数据集[15]。该数据集由 Kolektor Group d 提供并注释的有缺陷的电气换向器图像构建。具体而言，在嵌入电气换向器的塑料表面上观察到微观碎片或裂纹。每个换向器的表面积被捕获在八个不重叠的图像中。

2) 在工业磁瓦缺陷数据集上评估了我们提出的方法[16]。本文仅仅关注于数据集中的气孔、裂纹、磨损、断裂、和自由(无缺陷)五种缺陷类型。

我们的数据集分为三个子集。数据集的 80% 是训练集，10% 是验证集，10% 是测试集。我们遵循蒙特卡罗交叉验证方法并对数据集进行了三个任意划分。如果给定的示例包含异常，我们将其称为正例；如果示例不存在异常，我们将其称为负例。

3.2. 实验环境和评价函数

在训练设置方面，我们使用 PyTorch 框架构建网络，采用 SGD 优化器，动量项为 0.9，权重衰减为 $1e-4$ 。初始学习率设为 0.01，并采用“Poly”衰减策略。所有实验在 NVIDIA GeForce RTX 3090 Ti 24 GB GPU 上执行，批量大小为 8，最大 epoch 为 100。

评估模型性能时，基于混淆矩阵，包括真阳性(TP)、假阳性(FP)、真阴性(TN)和假阴性(FN)。各类别的 $F1$ 分数计算如式(5)：

$$F1 = 2 \cdot \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (5)$$

其中，精度 precision 如式(6)，召回率 recall 如式(8)：

$$\text{precision} = \frac{TP}{TP + FP} \quad (6)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (7)$$

对每个类别， IoU 定义为预测值与真实值的交并比：

$$IoU = \frac{TP}{TP + FP + FN} \quad (8)$$

此外， $mIoU$ 表示所有类别 IoU 的平均值， $Ave.F1$ 得分是 $F1$ 所有类别的平均值。

在模型运行效率方面，使用 $Flops$ (浮点运算数)反映模型计算效率和时间复杂度。实验中，通过输入一张 240×240 的图像比较不同模型的 $Flops$ 。 $Flops$ 越小说明模型计算效率越好，而 $params$ 表示算法参数个数，反映算法占用内存的大小。较小的参数值使得算法更容易部署到移动设备或嵌入式设备中。

3.3. 比较实验

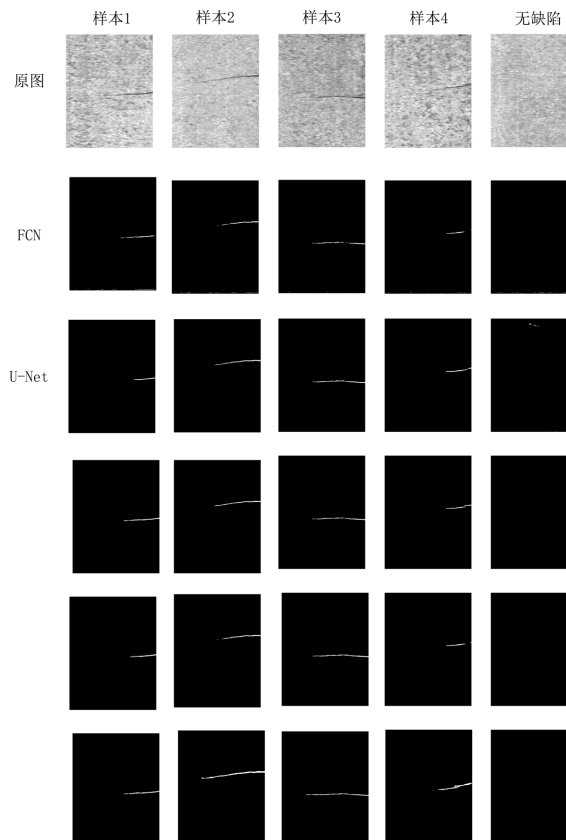


Figure 6. Comparison of detection results on the Kolektor dataset
图 6. Kolektor 数据集上的检测结果对比

我们对所提出的架构进行了与用于语义分割的最先进模型的比较。我们从测试集中随机选择了图像，对所提出的分割结构与 FCN、U-Net、U-Net++ 等模型进行了图像分割，并将结果展示在图 6 中。

通过观察图中的分割效果，明显可以看出与其他分割方法相比，在缺陷边缘处，完整度上与真实更加接近，而且输出噪声更低。所提出的模型在分割这些缺陷区域方面表现得更为接近。值得注意的是，样例 2 及其与样例 4 区域是最具挑战性的分割区域。在这方面，大多数分割模型都表现不佳，而我们提出的分割网络却令人满意地成功分割了这些复杂区域。

Table 1. Segmentation results comparison on the Kolektor dataset

表 1. Kolektor 数据集上的分割结果比较

Algorithm	FLOPs (GMac)	Params (M)	<i>mIoU</i> (%)	<i>F1</i> (%)
FCN	39.842	18.644	87.23	90.23
U-Net	62.791	17.263	89.23	92.35
U-Net++	312.615	47.189	91.73	94.56
Ours	42.844	23.252	91.89	94.67

从表 1 和表 2 可以看出，本文提出的分割网络比其他语义分割算法具有最高的 *mIoU*，同时具有较低的 FLOP 的 Params。U-Net++ 结合了类 DenseNet 结构，密集的跳跃连接提高了梯度流动性，这略微提高了其 *mIoU*。然而，它的 FLOPs 和 Params 变得非常大，不利于实时性和低内存要求。

同时给出了给出了各种方法在磁瓦数据集上的输出结果，本文的方法在不同的表面缺陷数据集下，*mIoU* 和 *F1* 上均优于其他对比方法。

Table 2. Comparison of segmentation results on the magnetic tile defect dataset

表 2. 磁瓦缺陷数据集上的分割结果比较

Algorithm	FLOPs (GMac)	Params (M)	<i>mIoU</i> (%)	<i>F1</i> (%)
FCN	39.842	18.644	62.41	71.34
U-Net	62.791	17.263	63.93	71.74
U-Net++	312.615	47.189	64.29	72.23
Ours	42.844	23.252	64.82	72.93

3.4. 消融实验

从表 3 中可以看出，加入多尺度融合和混合损失计算对 U-Net 的 *mIoU* 和 *F1* 指数提升较大，但会使 U-Net 本就较大的 FLOPs 更大。而将 ShuffleNet Unit 作为 backbone 能在减少 FLOPs 的同时，对 *mIoU* 和 *F1* 指数有一定的提升。而本文的架构将两者同时加入 U-Net 中，这样在保证 *mIoU* 和 *F1* 指数有较好的改进，同时 FLOPs 相比 U-Net 有更好的表现。

仅将 ShuffleNet Unit 作为 backbone 虽然能减少 FLOPs，但分割效果一般，而加入了多尺度融合和 ShuffleNet Unit 能同时兼顾效率与精度。

Table 3. Ablation experiments on the magnetic tile defect dataset

表 3. 在磁瓦缺陷数据集上的消融实验

Algorithm	FLOPs (GMac)	<i>mIoU</i> (%)	<i>F1</i> (%)
U-Net	62.791	63.81	71.34
U-Net+ShuffleNet Unit	38.542	64.14	72.19

续表

U-Net+AttentionGate	66.938	64.03	72.41
Ours	42.844	64.82	72.93

4. 结论

基于缺陷图像特点和最新分割模型设计思想,我们提出了一种采用U型全卷积神经网络的缺陷检测模型。通过在公开数据集 Kolektor 和工业磁瓦缺陷数据集上的测试,我们获得了比当前主流模型更出色的结果。在这两个数据集上,Params 和 Flops 为 42.844 和 23.252 时, $MIoU$ 和 $F1$ 指标分别达到 91.89, 94.67 和 64.82, 72.93, 获得了比目前主流模型更优的结果。该方法也可通过适当地修改应用于其他语义分割的场景,比如桥梁裂缝检测、铁道表面缺陷检测。

基金项目

这项工作得到了国家重点研发计划(2020YFB2007501)的支持。

参考文献

- [1] Koch, C., Georgieva, K., Kasireddy, V., et al. (2015) A Review on Computer Vision Based Defect Detection and Condition Assessment of Concrete and Asphalt Civil Infrastructure. *Advanced Engineering Informatics*, **29**, 196-210. <https://doi.org/10.1016/j.aei.2015.01.008>
- [2] Chen, Y., Ding, Y., Zhao, F., et al. (2021) Surface Defect Detection Methods for Industrial Products: A Review. *Applied Sciences*, **11**, Article No. 7657. <https://doi.org/10.3390/app11167657>
- [3] LeCun Y, Boser, B., Denker, J.S., et al. (1989) Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, **1**, 541-551. <https://doi.org/10.1162/neco.1989.1.4.541>
- [4] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. Arxiv:1409.1556
- [5] Ren, S., He, K., Girshick, R., et al. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149.
- [6] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 07-12 June 2015, 3431-3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- [7] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention MICCAI 2015: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Munich, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [8] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., et al. (2019) Unet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Transactions on Medical Imaging*, **39**, 1856-1867. <https://doi.org/10.1109/TMI.2019.2959609>
- [9] Badrinarayanan, V., Kendall, A. and SegNet, R.C. (2015) A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. Arxiv:1511.00561
- [10] Chen, L.C., Zhu, Y., Papandreou, G., et al. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 801-818. https://doi.org/10.1007/978-3-030-01234-2_49
- [11] Yang, F., Zhang, L., Yu, S., et al. (2019) Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection. *IEEE Transactions on Intelligent Transportation Systems*, **21**, 1525-1535. <https://doi.org/10.1109/TITS.2019.2910595>
- [12] Fei, Y., Wang, K.C.P., Zhang, A., et al. (2019) Pixel-Level Cracking Detection on 3D Asphalt Pavement Images through Deep-Learning-Based CrackNet-V. *IEEE Transactions on Intelligent Transportation Systems*, **21**, 273-284. <https://doi.org/10.1109/TITS.2019.2891167>
- [13] Zhang, X., Zhou, X., Lin, M., et al. (2018) Shufflenet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23

June 2018, 6848-6856. <https://doi.org/10.1109/CVPR.2018.00716>

- [14] Oktay, O., Schlemper, J., Folgoc, L.L., *et al.* (2018) Attention U-Net: Learning Where to Look for the Pancreas. Arxiv:1804.03999
- [15] Tabernik, D., Šela, S., Skvarč, J., *et al.* (2020) Segmentation-Based Deep-Learning Approach for Surface-Defect Detection. *Journal of Intelligent Manufacturing*, **31**, 759-776. <https://doi.org/10.1007/s10845-019-01476-x>
- [16] Huang, Y., Qiu, C. and Yuan, K. (2020) Surface Defect Saliency of Magnetic Tile. *The Visual Computer*, **36**, 85-96. <https://doi.org/10.1007/s00371-018-1588-5>