Microcrack segmentation of 3D CT images based on multi-task learning

Junjie Peng
Capital Normal University

Wenbin Li
Capital Normal University

Suyu Liao
Capital Normal University

Tao Xu
Capital Normal University

Yining Zhu (ynzhu@cnu.edu.cn)
Capital Normal University

Research Article

Keywords: Microstructural cracks segmentation, Multi-task learning, Attention model, Atrous spatial pyramid pooling, 3D CT images

Posted Date: August 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1899621/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Microcrack segmentation of 3D CT images based on multi-task learning

Junjie Peng\textsuperscript{1}, Wenbin Li\textsuperscript{1}, Suyu Liao\textsuperscript{1}, Tao Xu\textsuperscript{1} and Yining Zhu\textsuperscript{1,2*}

\textsuperscript{1}\textit{School of Mathematical Sciences, Capital Normal University, Beijing, 100048, China.}
\textsuperscript{2}\textit{Shenzhen National Applied Mathematics Center, Southern University of Science and Technology, Shenzhen, 518055, China.}

*Corresponding author(s). E-mail(s): ynzhu@cnu.edu.cn; Contributing authors: 2200502173@cnu.edu.cn; Suyu.Liao@126.com; xutao_math@163.com;

Abstract

As a non-destructive testing technique, industrial CT (Computed Tomography) plays an important role in research and applications of the automatic defect detection for 3D CT images. In recent years, convolutional neural networks (CNNs) have shown great potential in image recognition and segmentation tasks. For industrial 3D CT image detection, how to segment microstructural cracks is the main task at present. The challenge of this task is to identify cracks with low grayscale contrast and small size under unbalanced sample conditions. To address these problems, we propose a crack segmentation method based on the multi-task deep convolutional neural networks. The network architecture consists of a segmentation task module for extracting cracks and a classification task module for improving accuracy, which share part feature maps. We introduce the ECA block based on attention mechanism from Chen et al. to improve the segmentation accuracy. Numerical experimental results show that the proposed method has the better performance compared with other methods. At the same time, the network proposed in this paper has good generalization ability. In the case of only using simulated images training, it still yields good prediction results for real CT images.

Keywords: Microstructural cracks segmentation, Multi-task learning, Attention model, Atrous spatial pyramid pooling, 3D CT images
1 Introduction

CT can be used for medical clinical diagnosis as well as industrial non-destructive testing. One of the applications of industrial CT is defect detection for workpieces, so automatic identification or segmentation of defects (such as cracks) in industrial 3D CT images is an important task. In this field, some research results have been achieved by traditional methods [1–3]. Chen et al. [4] used the threshold optimization method to effectively segment the asphalt X-ray CT images by reducing image inhomogeneity. A CT image crack segmentation method based on iterative phase consistency is proposed in the literature [5], and it enables high precision with little computational time and fast speed.

In recent years, with the rapid development of deep learning techniques, neural networks have performed well in the related applications of semantic segmentation. A series of typical network models for image segmentation have been proposed [9–13]. Long et al. [6] replaced the fully connected layers with the convolution kernels in the neural networks to achieve end-to-end image segmentation. Ronneberger et al. [7] proposed an encoder-decoder architecture with skip connections, which fully integrates high-resolution features and low-resolution features to enable precise with very small number of training images. The 3D U-Net architecture proposed by iek et al. [8] extended the convolutional layer and pooling layer in U-Net to three-dimension to process volumetric data and achieve 3D image segmentation. Segmentation methods based on CNNs are also applied to various fields including feature segmentation detection [14–16] and medical image segmentation [17–20].

Deep learning for industrial CT image crack segmentation has also made some progress. The object detection approach is used to detect cracks and defects in CT images [21], but it can only locate cracks and cannot extract cracks pixel by pixel. Qi et al. [22] proposed using the object detection model to find the crack region, and then using the threshold segmentation to extract the cracks in the crack region. These two segmentation methods do not allow for end-to-end, and even takes more time. Zhong et al. [23] proposed a network end-to-end for fine filamentous object segmentation in CT images.

Although the existing deep learning methods have improved the accuracy and efficiency of CT image crack segmentation to some extent, there are some challenges. First, most CT cracks are relatively thin, usually only 2 to 4 pixels in width, thus causing in an extreme imbalance between the cracks and the background voxels, resulting in low segmentation accuracy. Second, CT images are usually subject to noise, so that the contrast between cracks and background is not high, which limits the feature extraction ability of the model. In addition, the existing methods are often used for two-dimensional CT images, which are difficult to effectively apply to three-dimensional CT images.

To solve the above problems, a multi-branch segmentation network (MBSNet) for 3D CT crack images is proposed in this paper. This model introduces prior information by adding auxiliary branches, which makes the model pay more attention to the crack feature information, improves the
extremely unbalanced problem between cracks and background, and enhances the generalization ability of the model. The feature extraction ability of the network is further enhanced by adding the attention mechanism module. The MBSNet is designed for 3D CT images, which can make full use of the inter-slice information and improve the segmentation results. The numerical results show that in the aspect of fine crack image segmentation, the proposed method using 3D simulation image for training is superior to U-Net and 3D U-Net in both the simulation data and real 3D images, and has good robustness and generalization ability.

2 Methods

The proposed network architecture in this paper contains two task modules: a segmentation task module and a classification task module. The segmentation task module is used as the backbone network based on the 3D U-Net network to extract cracks in 3D CT images. The classification task module is introduced after the last encoding layer of the backbone network, and its main role is to distinguish between positive and negative samples in the input data, which plays an auxiliary role in the network. The classification task shares part of the higher-level feature maps with the segmentation task, which helps the network learn common general features improve the robustness of the model. The auxiliary task helps the network to obtain the global feature channel information of voxels and improve the imbalance between crack voxels and non-crack voxels.

In addition, the 3D features from the encoding part to the decoding part of the backbone network are combined through the concatenation with ECA modules [24]. The ECA module uses 1D convolution to capture local cross-channel information and learns the correct channel information for the purpose of suppressing irrelevant feature information, helping the network to extract crack feature information and improving the segmentation accuracy.

2.1 Classification architecture

The number of crack voxels in CT images is very small, accounting for less than 1% of the whole image voxel number, and in many local cubic patches, even there is no crack voxel. Therefore, the crack voxels are much less than the background voxels, which easily leads to the prediction results seriously biased towards non-crack voxels. To improve this problem, we introduce a classification task to the model. The classification task aims to distinguish between positive and negative training patches. Among them, the positive samples are those with more than a specific number of crack voxels, while the negative samples are those with a small proportion of crack voxels. The classification task and the segmentation task share part of the high-level feature maps, so the robustness of the model can be improved by adding auxiliary branches to reduce the noise [25] [26]. Since the high-resolution 3D features
can be learned from the encoder layer, the classification task is added at the end of the shrinkage path of the backbone network.

Fig 1 shows the architecture of the classification task. In order to learn global low-resolution 3D features, we introduce the ASPP (Atrous Spatial Pyramid Pooling) module [27] to the classification task. The SPP (Spatial Pyramid Pooling) module [28] and the dilated convolution [29] are combined in the ASPP module, as is shown in Fig 2. The ASPP module extracts information at various scales and the feature maps are integrated. The information extracted by the classification task contains both local and global feature maps, so it can improve the classification accuracy. In this paper, due to the tiny size of the experimental data, a large dilated rate does not work well, hence the ASPP module’s dilated rates are 2, 4 and 6. In addition, we use the global average pool to obtain the global feature maps and use 1×1 convolutions to generate the final classification probabilities.

![Fig. 1 Structure of the classification architecture.](image1)

![Fig. 2 Structure of the ASPP module.](image2)
The classification task uses the crack feature maps to determine whether the data are positive samples, the segmentation task and classification task share the backbone of the network, so more crack feature maps are obtained by the segmentation task, the influence of data imbalance on the segmentation accuracy can be reduced. The classification task plays the same job as the regularization term in the optimization problem, and the addition of positive and negative samples is the same as adding the crack’s a priori knowledge [25] [30]. The classification task can be used to limit the network’s solution space, prevent the network from focusing too much on the crack’s background information, and achieve the goal of minimizing the imbalance between the foreground and background voxels.

2.2 Network architecture

As shown in Fig 3, the MBSNet consists of a segmentation task and a classification task. In the segmentation task, there is an analysis and a synthesis path. In the analysis path, each layer consists of two 3×3×3 convolutions, each followed by a Parametric Rectified Linear Unit (PReLU) and a max pooling layer with a step size of 2 and a size of 2×2×2. Each layer of the synthesis path includes the upsampling of the feature maps, and then two 3×3×3 convolutions, each convolution has a PReLU behind. After each upsampling, the number of feature channels are in half. Fig 4 shows the architecture of ECA module [24], the ECA module makes the attention coefficients more focused on the crack features and suppresses irrelevant features, helping the network to learn the crack information and improve the segmentation accuracy.

The classification task is introduced as an additional branch after the last encoder layer of the segmentation task. The feature maps obtained from the segmentation task are passed through the ASPP module to obtain feature information at different scales. Then, the feature maps are turned into global feature vectors by a global average pooling operation, and finally the feature vectors are weighted for classification by 1×1 convolutions. The classification and segmentation tasks share the same parameters during training and cancel the classification network when predicting the results.

2.3 Loss function

The number of crack voxels account for less than 1% of the total voxels, resulting in an extreme imbalance between foreground and background regions, as well as an unequal distribution of positive and negative samples. Due to these problems, the recognition accuracy of positive samples is low, the gradient values created by positive samples contribute less to optimization update, and the network is unable to learn more useful feature maps. These problems can be improved by using an appropriate loss function. The loss function in this paper consists of two components, classification loss and segmentation loss.

In the classification task, the imbalance between positive and negative samples needs to be considered. Due to the limited number of positive samples,
the loss layer may be biased to the side with more samples, leading to failure in identifying cracks. Experiments have proved \cite{31} \cite{32} that the $L_{BCE}$ with weights added to the binary cross-entropy loss function can improve this problem. In the Eq (1), $g_i$ is the label value of the sample, a $g_i$ of 0 indicates that the sample is a negative sample, and a $g_i$ of 1 indicates that the sample is a positive sample. $p_i$ is the probability that sample $i$ is a positive sample, and $\alpha$ is the weight of the negative sample. By controlling the positive and negative
sample weights can effectively reduce the impact of sample imbalance. In our experiments, $\alpha$ is taken as 0.2.

$$L_{Cla} = -\frac{1}{N} \sum_{i=1}^{n} (g_i \log p_i + \alpha (1 - g_i) \log (1 - p_i))$$  \hspace{1cm} (1)$$

In the segmentation task, the number of foreground and background voxels is unbalanced, and it is difficult to distinguish crack voxels from non-crack voxels due to low contrast. Dice loss is a region-dependent loss function [10], in which the loss of a voxel is related to other voxels as well as the current voxel. Dice loss is able to establish the proper balance between foreground and background voxels, so performs well in strong imbalance. As a result, the Dice loss Eq (2) is used in our segmentation network. In the volume where the sum of voxels is $N$, $p_i$ is the probability that voxel $i$ is the voxel of crack, $g_i$ is 1 that the voxel is the crack, and $g_i$ is 0 that the voxel is the background.

$$L_{Seg} = \frac{2 \sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2}$$  \hspace{1cm} (2)$$

As shown in Eq (3), We obtain the loss function of the model by combining the loss functions of the segmentation task and the classification task, where the parameter $\mu$ is used to specify the weights of the auxiliary tasks. In our experiments, we take 0.25 for $\mu$.

$$Loss = L_{Seg} + \mu L_{Cla}$$  \hspace{1cm} (3)$$

3 Dataset and Evaluation metrics

3.1 Data preparation

In order to evaluate the performance of the network in this paper, we established a crack image dataset. As shown in Fig 5, the dataset consists of five sets of three-dimensional CT crack images which are named Crack I ~ Crack V. Crack I ~ Crack III are generated by computer simulation, Crack IV and Crack V are obtained by CT scanning and three-dimensional reconstruction of industrial samples containing cracks. The resolution of each slice and label is 512×512 pixels.

The dataset is divided into two parts: the training set and the test set. The training set consists of Crack I and Crack II, where Crack I consists of 256 slices and their labels, and Crack II consists of 512 slices and their labels. To make the training set similar to the real 3D CT images, we created tiny, low-contrast crack voxels and introduced random noise to them. The test set contains Crack III, Crack IV and Crack V. Crack III consists of 300 slices and labels. Crack IV and Crack V consist of 350 and 300 slices. To create labels, we manually labeled each 2D CT slice pixel by pixel along the crack shape and
Springer Nature 2021 LATEX template

Microcrack segmentation of 3D CT images based on multi-task learning

Fig. 5 The dataset and its 3D visualization results. Each column shows a set of three-dimensional CT crack images, from left to right are Crack I ~ Crack V. The first and third rows are an example of crack image with its label, and the second and fourth rows are the 3D visualization of the data and its label.

Binarized the labeled images, with 0 indicating the crack voxel and 1 indicating the background voxel.

During the training process, the 3D training images are cropped into image patches of size $64 \times 64 \times 64$ and normalized as the input to the network. The data enhancement is achieved by flipping and rotating the volumetric data. We count the total number of crack voxels $N$ in each patch label of the above training data, and consider samples with $N$ greater than a given threshold $\phi$ as positive samples, otherwise as negative samples, as labels for the classification task. In this paper, $\phi$ is 500.

3.2 Evaluation metrics

In order to evaluate the accuracy of crack segmentation, Recall, Mean Intersection over Union (MIoU) and F1 Score (F1) were adopted to evaluate the approach performance, where TP, FP, TN and FN are the true positive, false positive, true negative and false negative rates.
Recall = $\frac{TP}{TP + FN}$ \hfill (4)  

$\text{MIoU} = \frac{TP}{TP + FN + TP}$ \hfill (5)  

Precision = $\frac{TP}{TP + FP}$ \hfill (6)  

$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$ \hfill (7)

The results of crack segmentation are used to measure and analyze the properties of cracks. By comparing with the crack property labels, the results of the quantified crack properties are evaluated. Crack volume, surface area, average width and crack complexity are calculated in 3D crack images. The estimated relative error is calculated as an important indicator to measure the performance of crack quantification method. We used morphological dilation, erosion, and denoising for each slice, and then calculated the total number of voxels. The sum of voxel volume is regarded as the volume of three-dimensional crack. We extract the boundary in each slice, and the sum of the surface area of the boundary is regarded as the surface area of the three-dimensional crack. The ratio of surface area to volume represents the complexity of cracks.

4 Experiments and results

4.1 Training configuration

Our model is built with python 3.7 under the framework of pytorch 1.8.0 on a machine equipped with an NVIDIA Tesla V100S GPU with 32GB of memory. The initial learning rate of the network was 0.01, we used the stochastic gradient descent solver for network training, batch size is 32, and we ran 200 training iterations. The learning rate was periodically updated 5 times by cosine in the whole process.

4.2 results

3D U-Net [8], U-Net [7] and SegNet [9] are widely used as representative image segmentation networks. Therefore, in this paper, the above three methods are selected for comparison. Fig 6 shows some typical results on these methods and the proposed network. The 3D visual results in Fig 7 show that the segmentation cracks of the method in this paper have greater similarity with the labels.

The segmentation results of the two-dimensional SegNet and U-Net network model crack images are smoother, and the crack boundary contours is quite different from the label. In contrast, 3D CNN segmentation results are closer to the real cracks. The 3D convolutional kernel can fully extract inter-slice information from multiple frames, so the results of 3D U-Net and the
Microcrack segmentation of 3D CT images based on multi-task learning

Fig. 6 Examples of crack segmentation results. From left to right: crack image, label image, SegNet, U-Net, 3D U-Net, the proposed method.

MBSNet are more accurate than U-Net and SegNet. Moreover, the MBSNet has better accuracy for crack voxels compared with other networks. The method in this paper can also achieve better results when the contrast between cracks and background is low. The pixel-level error of the label can lead to a lower score due to the small size of the crack. However, the method proposed in this paper can still produce good results in the case of labeling errors. As shown in the Table 1, Table 2 and Table 3, the crack properties of the MBSNet are closer to reality than other methods in terms of crack measurement. Table 4, Table 5 and Table 6 shows that the proposed method outperforms the other methods, which demonstrates the effectiveness of the method.

Table 1 Measurement results of Crack characteristics in Crack III.

<table>
<thead>
<tr>
<th>CNN</th>
<th>surface area/error</th>
<th>volume/error</th>
<th>average width/error</th>
<th>complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>356769</td>
<td>536962</td>
<td>3.01</td>
<td>0.6649</td>
</tr>
<tr>
<td>SegNet</td>
<td>222614/37.60%</td>
<td>308875/42.68%</td>
<td>2.76/8.31%</td>
<td>0.7343</td>
</tr>
<tr>
<td>U-Net</td>
<td>261437/26.72%</td>
<td>358124/33.31%</td>
<td>2.72/9.63%</td>
<td>0.7416</td>
</tr>
<tr>
<td>3D U-Net</td>
<td>300735/15.71%</td>
<td>311777/41.94%</td>
<td>2.07/31.23%</td>
<td>0.9700</td>
</tr>
<tr>
<td>MBSNet</td>
<td>328893/7.81%</td>
<td>457340/14.83%</td>
<td>2.78/7.64%</td>
<td>0.7199</td>
</tr>
</tbody>
</table>

In order to verify the classification task is effective in real data, we did a series of ablation experiments. As shown in the Table 7 and Table 8, the 3D U-Net with the classification task (3D CU-Net) have better crack segmentation results on the Recall, MIoU and F1, indicating that adding the classification
Microcrack segmentation of 3D CT images based on multi-task learning

![Image](image.png)

**Fig. 7** The 3D visualization segmentation results. From the first row to the third row: Crack III, Crack IV and Crack V.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Measurement results of Crack characteristics in Crack IV.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>surface area/error</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Label</td>
<td>380840</td>
</tr>
<tr>
<td>SegNet</td>
<td>214618/43.65%</td>
</tr>
<tr>
<td>U-Net</td>
<td>184477/51.56%</td>
</tr>
<tr>
<td>3D U-Net</td>
<td>231678/39.17%</td>
</tr>
<tr>
<td>MBSNet</td>
<td>283834/25.47%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Measurement results of Crack characteristics in Crack V.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>surface area/error</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Label</td>
<td>80993</td>
</tr>
<tr>
<td>SegNet</td>
<td>65960/18.56%</td>
</tr>
<tr>
<td>U-Net</td>
<td>101615/25.46%</td>
</tr>
<tr>
<td>3D U-Net</td>
<td>78852/2.64%</td>
</tr>
<tr>
<td>MBSNet</td>
<td><strong>79721/1.57%</strong></td>
</tr>
</tbody>
</table>

The classification task can improve the segmentation accuracy of the method in this paper, which proves the effectiveness of the classification task.
Table 4  Segmentation accuracy evaluation of Crack III.

<table>
<thead>
<tr>
<th>CNN</th>
<th>Recall</th>
<th>MIoU</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SegNet</td>
<td>0.7515</td>
<td>0.8031</td>
<td>0.7460</td>
</tr>
<tr>
<td>U-Net</td>
<td>0.8613</td>
<td>0.7990</td>
<td>0.7480</td>
</tr>
<tr>
<td>3D U-Net</td>
<td>0.7525</td>
<td>0.8300</td>
<td>0.7937</td>
</tr>
<tr>
<td>MBSNet</td>
<td>0.8136</td>
<td>0.8505</td>
<td>0.8227</td>
</tr>
</tbody>
</table>

Table 5  Segmentation accuracy evaluation of Crack IV.

<table>
<thead>
<tr>
<th>CNN</th>
<th>Recall</th>
<th>MIoU</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SegNet</td>
<td>0.2831</td>
<td>0.6120</td>
<td>0.3659</td>
</tr>
<tr>
<td>U-Net</td>
<td>0.2896</td>
<td>0.6166</td>
<td>0.3768</td>
</tr>
<tr>
<td>3D U-Net</td>
<td>0.3131</td>
<td>0.6220</td>
<td>0.3879</td>
</tr>
<tr>
<td>MBSNet</td>
<td>0.3582</td>
<td>0.6304</td>
<td>0.4132</td>
</tr>
</tbody>
</table>

Table 6  Segmentation accuracy evaluation of Crack V.

<table>
<thead>
<tr>
<th>CNN</th>
<th>Recall</th>
<th>MIoU</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SegNet</td>
<td>0.3526</td>
<td>0.6397</td>
<td>0.4378</td>
</tr>
<tr>
<td>U-Net</td>
<td>0.4669</td>
<td>0.6904</td>
<td>0.5509</td>
</tr>
<tr>
<td>3D U-Net</td>
<td>0.4075</td>
<td>0.6702</td>
<td>0.5114</td>
</tr>
<tr>
<td>MBSNet</td>
<td>0.6181</td>
<td>0.7298</td>
<td>0.6321</td>
</tr>
</tbody>
</table>

Table 7  Performance Comparison of classification net in Crack IV.

<table>
<thead>
<tr>
<th>CNN</th>
<th>Recall</th>
<th>MIoU</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D U-Net</td>
<td>0.3131</td>
<td>0.6220</td>
<td>0.3879</td>
</tr>
<tr>
<td>3D CU-Net</td>
<td>0.3360</td>
<td>0.6253</td>
<td>0.3987</td>
</tr>
<tr>
<td>MBSNet</td>
<td>0.3582</td>
<td>0.6304</td>
<td>0.4132</td>
</tr>
</tbody>
</table>

Table 8  Performance Comparison of classification net in Crack V.

<table>
<thead>
<tr>
<th>CNN</th>
<th>Recall</th>
<th>MIoU</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D U-Net</td>
<td>0.4075</td>
<td>0.6702</td>
<td>0.5114</td>
</tr>
<tr>
<td>3D CU-Net</td>
<td>0.4812</td>
<td>0.6945</td>
<td>0.5627</td>
</tr>
<tr>
<td>MBSNet</td>
<td>0.6181</td>
<td>0.7298</td>
<td>0.6321</td>
</tr>
</tbody>
</table>
5 Conclusion

In this paper, a multi-task approach based on volumetric convolutional neural networks is proposed to solve the small size and low contrast problem in unbalanced 3D crack images. The algorithm is composed of two tasks: a segmentation task and a classification task. The segmentation task is based on the 3D U-Net architecture to extract the cracks. The feature channels from the encoder path to the decoder path are propagated through skip connections with the ECA modules to help the model locate relevant regions and focus more on the crack features. At the end of the contracting path of the segmentation task, we introduce the classification task with the ASPP block. This classification task introduces more priori information and effectively improves the problem of extreme imbalance between background and foreground voxels. The network is trained with simulated datasets and performs well on both simulated and real datasets, indicating that it is robust.

References


Microcrack segmentation of 3D CT images based on multi-task learning


Microcrack segmentation of 3D CT images based on multi-task learning


