Crack Segmentation on FRP CT Images Using Residual Multi-scale Dilated Convolution and Reversed Cross Entropy Loss

Zhang Yuyao[1], Xu Yitao[2], Han Chenyu[2]
1 School of Mathematics, EPFL, Switzerland
2 School of Computer and Communication Science, EPFL, Switzerland

***Abstract*—This paper aims at solving the crack segmentation task on fiber-based plastic (FRP) CT images. After manually collecting the images and doing the labeling, we adopt a SOTA method on benchmark crack segmentation datasets as our baseline methods. Then we propose various modifications to improve the performance, including a loss function for robust classification, network modules aiming at dealing with cracks with different shapes and retaining high-frequency components in the images. We also propose a new architecture, Multi-scale Dilated Convolution (MDC), that can capture multi-shaped cracks with less parameters than existing methods. Our results show that the modifications boost the performance of the baseline method.**

I. Introduction

Glass fiber-reinforced plastics (GFRPs) are a type of composite material that is commonly used in various industries. Angle-ply GFRPs, which contain multiple layers oriented at specific angles, are one type of GFRP. When subjected to oscillating or random loads, FRPs can experience fatigue damage, resulting in the formation and propagation of cracks in the material. These cracks can lead to material failure. Thus, the study of fatigue is an important topic in the field of material engineering. Micro-CT scanners are used for the non-destructive imaging of materials, allowing for the detection of cracks. However, manual inspection of all cracks in CT images can be difficult and time-consuming.

Machine-learning-based methods[1] have been proposed for crack segmentation in buildings and pavements, but these methods tend to work best with long, uniform surface cracks. In GFRPs, cracks can have various shapes and can be small, causing low performance in the established method. Moreover, due to the composite nature of the material, different regions of the material can appear in different colors in CT images, while cracks hide those regions.

To address these challenges, we first collect 164 CT images of an FRP after being damaged and manually label the cracks on the images. We adopt a U-Net-based baseline method[1] that is SOTA on the current benchmark crack-detection dataset[2], [3], [4], [5], [6], [7]. Due to specific problems in our task, we propose to use Symmetric Cross-Entropy (SCE) loss to mitigate the noise in the manual labels as well as different network architecture modifications aiming at dealing with cracks with different shapes and retaining high-frequency components in the images. Modifications include Dense Atrous Convolution (DAC)[8], Residual Multi-scale Pooling(RMP)[9], High Frequency Extractor (HFE)[10], and a self-proposed module Multi-scale Dilated Convolution (MDC). Our optimization efforts resulted in satisfactory crack segmentation results for the material.

II. Relate Work

A. Crack Segmentation

Traditional methods based on image processing like thresholding[11]and morphological processing[12] usually leveraged manual feature generation and extraction to determine whether cracks were evident. Constraints like varying background and light conditions are also problems for these techniques.

In previous years, methods based on machine learning, like k-nearest neighbors[13], support vector machine[14], and artificial neural network[15][16] gained some achievements since they include less human intervention. However, this kind of approach barely uses multiple levels of feature extraction and may struggle to address the intricacies of real-world image data completely.

Deep learning-based methods turned white hot in recent years and rapidly became dominant in crack segmentation. Many useful techniques occur, from using FCN[17] to the trending encoder-decoder UNet[18] architectures, originally developed in the field of medical image segmentation. Zhou et al.[19], König et al.[20] used the FCN form network. Other works[21][22] also used Mask-RCNN[23], and Generative Adversarial Network (GAN)[24] to tackle the problem. Later works by Liu et al.[25], Lau et al. [26] etc. leveraged the UNet to obtain good results.

Yuki et al.[1] proposed a framework combining super-resolution and semantic segmentation networks to perform accurate crack segmentation on low-resolution images. The semantic segmentation part is a typical U-Net-structured network. They propose a new loss function based on several loss functions that focus on local properties around crack boundaries and solving the class imbalance problem in crack segmentation, including Boundary loss[27] and Combo loss[28]. The Combo loss that Yuki et al. use is a weighted sum of weight binary cross-entropy loss and Dice loss[29] or Generalized Dice loss[30], aiming at solving

the class imbalance problem. Our baseline is based on the method proposed in [1].

B. Robustness in Classification and Stronger Data Augmentation

There are several works that focus on training more robust classification models and data augmentation. Wang et al. proposed a simple but efficient method called Symmetric Cross Entropy (SCE) to make the neural network more robust to noise in labels during training. The idea is to change the order of target and predicted labels in the formula of normal cross-entropy loss (CE loss) and combine the SCE with normal CE loss. A data augmentation method proposed to enable multi-background training in object detection tasks is introduced in YOLOv4[31], called Mosaic. The idea is to concatenate several random images in the train set at a random cutting point and combine the ground truth bounding boxes to let the model see different backgrounds in one image. We propose introducing theses methods to our task.

III. Method

Our method first contains pre-processing of the CT images since there is no label. Then, we adopt the segmentation pipeline introduced in [1] and discard the super-resolution part since that is meaningless to our project. We modify the loss function and network architecture in [1] to make it more adapted to our goal.

A. Build the dataset

There are 3 directions of the CT scan in our dataset, namely X, Y, Z. The Z direction has few cracks on the images and the cracks are of different patterns than X and Y directions. Hence, we discard the Z direction. The images in the same direction come from a single scan sequence where consecutive images are not very different. We pick one image every 15 images and form a dataset consisting of 82 images from X direction and 82 images from Y. The labeling is done manually by us. The aspect ratio of images in X direction is larger than 3 while the one in Y direction is smaller than 0.3. We cut 5 square images from a single image in X or Y direction with overlapping, and rotate all images at 90,180,270 degrees to form a dataset of 3280 images. We split train and test sets with a ratio of 0.9, resulting in 2952 train images and 328 test images.

B. Symmetric cross-entropy loss and the whole loss function

The labeling is done manually, so noise in the label is inevitable. To mitigate the influence of noisy labels, we propose to use an established noise-resistant classification loss function introduced in [32], called symmetric cross-entropy loss (SCE loss). Normal cross-entropy (CE) loss measures the distance between two discrete distributions over the same set of states: $l_{ce} = -\sum_{k=1}^{K} q(k|x) \log p(k|x)$, where x is the input image, K is the total number of classes, k is the current

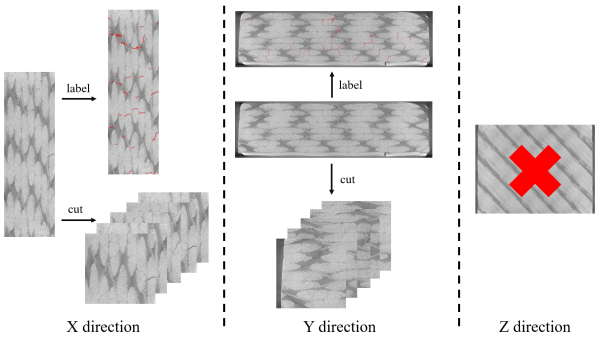

Figure 1. Example images in our dataset. We discard Z direction and cut images in X and Y directions. The labeling is done manually by ourselves.

image class. $q(k|x)$ is the ground truth class distribution, which is a one-hot vector whereas $p(k|x)$ is the predicted class distribution. A reversed version of CE where the order of p and q is exchanged: $l_{rce} = -\sum_{k=1}^{K} p(k|x) \log q(k|x)$, where the $\log 0$ problem is solved by adding a small constant to 0 entries in $q(k|x)$. The SCE loss is a weighted sum of CE and RCE, given in equation 1:

$$l_{sce} = \alpha l_{ce} + \beta l_{rce}, \tag{1}$$

Want et al.[32] proved that such a simple modification of the original cross-entropy loss enables robust training against noise in labels. The final form of our loss function is based on the loss proposed in [1] and l_{sce}, as shown in equation 2.

$$\mathcal{L}_{total} = \kappa \mathcal{L}_{Boundary} + (1-\kappa)[(1-\gamma)\mathcal{L}_{Dice} + \gamma l_{sce}], \tag{2}$$

where $\mathcal{L}_{Boundary}$ is Boundary loss[27] and $\mathcal{L}_{Dice}$ is Dice loss [29]. This loss function is similar to $\mathcal{L}_{BC}$ introduced in [1] except for the cross-entropy part. We adopt this loss since it performs the best among all other loss functions in [1] in the high-resolution segmentation task.

C. Mosaic Data Augmentation

We refer readers to Appendix Section I for details of Mosaic Data Augmentation. We do not adopt this strategy in the following experiments.

D. Network Architecture

The overall network structure is based on the UNet model (Fig.2), with 4 four additional modules, Dense Atrous Convlolution (DAC), Residual Multi-scale Pooling(RMP), Multi-scale Dilated Convolution with by splitting channels(MDC) and High Frequency Extractor(HFE).

1) Dense Atrous Convolution (DAC): In our dataset, the cracks are in different shapes and sizes. Therefore, to leverage the multiscale features of the crack images, followed by Xiang et al. [8], we added the DAC module to capture deeper and wider information. We refer users to Appendix Section IV for details of DAC module.

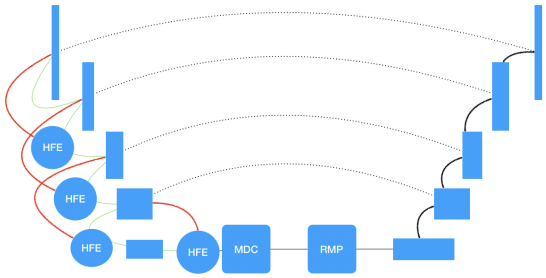

Figure 2. The overall architecture of the modified UNet model. Red(green) lines represent the tensor from the shallower(deeper) layer in the down-sampling parts, and the black lines stand for the up-sampling parts. The dotted lines are the skip connections in UNet.

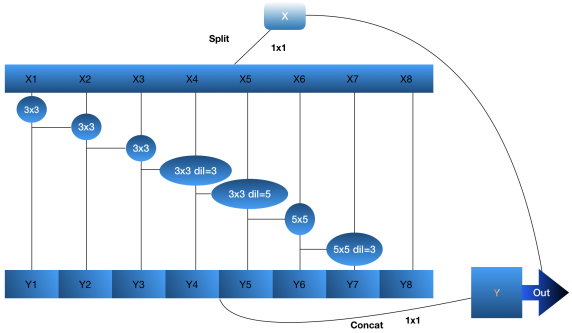

Figure 3. The graphical illustration of Multi-scale Dilated Convolution by Splitting Channels. The input feature map is fed into a 1x1 convolution, and the channels are split into 8 subsets. Cascade convolutions with kernels of multiple sizes and dilation rates are used for each subset. Then all the subsets are concatenated and fed into a 1x1 convolution, and finally stacked with the input map.

2) Residual Multi-scale Pooling(RMP): To extract more semantic information, we implanted the Residual Multi-kernel Pooling, introduced by Gu et al.[9], in our network architecture. We refer users to Appendix Section IV for details of RMP module.

3) Multi-scale Dilated Convolution (MDC) by Splitting Channels: Noticing that the previous DAC module would result in too many parameters, we tried to build a module that has comparable performance but requires less computational burden. Inspired by Hu et at.[10], we introduced the Multi-scale Dilated Convolution by splitting the channels of the input feature map. We first do a 1x1 convolution to the input X and then split the channels into eight feature map subsets $X_i,\ i \in \{0, \ldots, 8\}$. This channel splitting reduces the computational complexity significantly. For each subset, we conduct a cascade convolution pipeline. X_1 is fed into a 3x3 convolution to get Y_1 and X_2 combined with Y_1 are then fed together into another 3x3 convolution, by such process we get $Y_i,\ i \in \{1, 2, \ldots, 7\}$. There are two kernel sizes, 3 and 5(see figure 3). However, dilation is added for the 4, 5, and 7th convolutions. Dilated convolution[33] enlarges the kernel by inserting "holes" into the kernel entries. This will help enlarge the receptive field to capture more information. Finally, we add a residual connection for X_8 to get Y_8. We then concatenate all the Y_is to get the resulting feature map Y and output the stack of Y and X after skip connection. We will compare the performance and number of parameters in the experiment section.

4) High Frequency Extractor: During down-sampling, we will lose some high-frequency information like the edges of the cracks. However, this high-frequency information plays a crucial role for neural networks to discover the cracks. In Hu et al.'s[10] work for remote sensing, they introduced the High-Frequency Extractor (HFE) to detect the edge of the clouds. Also inspired by the attention strategy in work by Lin et al[34], we apply this method while down-sampling. As shown in figure 2 the HFE module takes two input feature maps from the previous layer as the deeper one F_D and the map F_S from the layer before the shallower one. F_S is fed into a 1x1 convolution then down-sampled, and is denoted as $down(\phi(F_S))$. Then we add the F_D with the difference between F_D and F_S and then feed the output $F_S + \gamma(F_D - down(\phi(F_S))$ to the next encoder(We conducted the ablation study for this γ, and put the result in the Appendix Section V).

IV. Experiment

We first experiment on our modification of loss function and data augmentation. After selecting the best scheme, we conduct experiments on all combinations of our modifications on the network architecture. All our experiments run for 40000 epochs except for the final best model, which we run for 100000 epochs. We set batch size = 6, the same as in [1]. We use Adam optimizer with a learning rate of 0.00001. For testing, we set the prediction threshold range from 0.01 to 0.99 with a 0.01 interval. We use averaged IoU (AIU)[35] and the best IoU among all thresholds to measure the performance of our model. Our baseline model achieves a 0.3569 AIU score and 0.4071 best IOU (threshold:0.39).

A. Symmetric Cross-entropy Loss

We first conduct thorough experiments on SCE loss. To explore the trade-off between reversed CE and normal CE in the SCE loss, we set $\alpha = 1-\beta$ in equation 1. We experiment with $\beta = [0.1, 0.2, 0.3, 0.5, 0.7, 0.8, 0.9]$, since more extreme trade-off cases need more fine-grained parameter tuning. Results are shown in table I.

From table I, we can see that a larger weight on reversed CE loss can contribute to a better AIU score while preserving the best IoU. The result indicates that symmetric CE loss makes the model generalize better than the baseline model. Hence, we use SCE loss in the following experiments and set $\beta = 0.9$.

Metric \ β	0.1	0.2	0.3	0.5	0.7	0.8	0.9
AIU	0.3564	0.3614	0.3625	0.3667	0.3676	0.3664	**0.3742**
Best IoU (threshold)	0.4046(0.37)	**0.4085(0.4)**	0.4049(0.4)	0.4049(0.43)	0.4041(0.37)	0.4031(0.36)	0.4063(0.42)

Table I
RESULTS OF SCE LOSS. THE LARGER WEIGHT ON REVERSED CE LOSS CONTRIBUTES TO BETTER AVERAGE PERFORMANCE WHILE KEEPING THE BEST IOU(40000 ITERATIONS).

Metric \ Method	Basline	DAC	MDC	DAC+RMP	MDC+RMP	DAC+RMP+HFE	MDC+RMP+HFE
AIU	0.3564	0.3717	0.3762	0.3827	0.3768	0.3774 (Gamma=0.3)	**0.3832(Gamma=0.1)**
Best IoU (threshold)	0.4046(0.37)	0.4131(0.3)	0.4130(0.36)	**0.4182(0.37)**	0.4139(0.35)	0.4175(0.3)	0.4169(0.41)
Parameters	29.5M	46.8M	**30.4M**	47.1M	**30.7M**	48.7M	**32.3M**

Table II
RESULTS OF NETWORK MODULE EXPERIMENTS(40000 ITERATIONS).

Images Predictions Ground Truth

Figure 4. Example of our model prediction. Note that our model does not make errors on the wrong label.

Metric \ Method	Basline	MDC+RMP+HFE
AIU	0.4247	0.4406
Best IoU (threshold)	0.4510 (0.35)	0.4630(0.32)

Table III
RESULTS OF OUR BEST MODEL TRAINED FOR 100000 ITERATIONS.

B. Network Modules

We run experiments on different combinations of modules. The Baseline, only DAC, only MDC, DAC + RMP, MDC + RMP, DAC + RMP + HFE, and MDC + RMP + HFE. The results are in Table II. Under the setting of 40000 iterations, we find that DAC+RMP gets the highest best IoU score of 0.4182(0.4046), and MDC+RMP+HFE gets the highest AIU of 0.3832(0.3564). Only by comparing DAC and MDC we can find that the best IoU is roughly the same, but the AIU increases by 0.0045, it has more robustness. Importantly, the number of parameters is reduced by 1.6M(16391744), it reduces about 1/3 of the computation complexity, which is significant. For RMP, after adding it, both AIU and IoU increased for MDC and DAC. And by adding HFE, we increased both AIU and Best IoU for MDC+RMP, and for DAC + RMP although it led to a Degradation of performance after adding HFE, it is still better than the baseline a lot, and we think this degradation is due to the slow convergence, which further shows the advantage of our MDC module.

C. Final Results

Finally, we select three models that show promising results during the 40000-iteration trial run, namely DAC+RMP, MDC+RMP, and MDC+RMP+HFE (Gamma=0.1), to run for 100000 iterations and compare them to the baseline which is also trained for 100000 epochs. The results are shown in table III. We refer readers to Appendix Section III for more results on 100000-iteration experiments.

D. Visual Results

Here we show an example of our model prediction in figure 4. Our model predicts almost all cracks correctly. It is worth noticing that our model does not make errors on the wrong label, as indicated by the red arrow. We refer readers to Appendix Section II for more results of our best model.

V. CONCLUSION

In this project, we aim to solve the crack segmentation problem on FRP CT images. After analysis of specific problems that emerged in our task, we propose to use Symmetric Cross-entropy loss to mitigate the manual label noise in our dataset. Moreover, various network architecture modification schemes are proposed to capture different-shaped cracks better and retain high-frequency components in the images, including Dense Atrous Convolution (DAC), Residual Multi-scale Pooling(RMP), Multi-scale Dilated Convolution with by splitting channels(MDC) and High-Frequency Extractor(HFE). Thorough experiments show that the best combination of the network architecture modification is our self-proposed MDC+RMP+HFE with $\gamma = 0.1$. our method largely improves the performance of the baseline method and achieves robust segmentation against noise in labels.

REFERENCES

[1] Y. Kondo and N. Ukita, "Crack segmentation for low-resolution images using joint learning with super-resolution," in *2021 17th International Conference on Machine Vision and Applications (MVA)*. IEEE, 2021, pp. 1–6.

[2] R. Amhaz, S. Chambon, J. Idier, and V. Baltazart, "Automatic crack detection on two-dimensional pavement images: An algorithm based on minimal path selection."

[3] L. Zhang, F. Yang, Y. D. Zhang, and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 3708–3712.

[4] Q. Zou, Y. Cao, Q. Li, Q. Mao, and S. Wang, "Cracktree: Automatic crack detection from pavement images," *Pattern Recognition Letters*, vol. 33, no. 3, pp. 227–238, 2012.

[5] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei, and H. Ling, "Feature pyramid and hierarchical boosting network for pavement crack detection," *arXiv preprint arXiv:1901.06340*, 2019.

[6] Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, "Automatic road crack detection using random structured forests," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3434–3445, 2016.

[7] M. Eisenbach, R. Stricker, D. Seichter, K. Amende, K. Debes, M. Sesselmann, D. Ebersbach, U. Stoeckert, and H.-M. Gross, "How to get pavement distress detection ready for deep learning? a systematic approach." in *International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 2039–2047.

[8] C. Xiang, W. Wang, L. Deng, P. Shi, and X. Kong, "Crack detection algorithm for concrete structures based on super-resolution reconstruction and segmentation network," *Automation in Construction*, vol. 140, p. 104346, 2022.

[9] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE transactions on medical imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[10] K. Hu, D. Zhang, and M. Xia, "Cdunet: Cloud detection unet for remote sensing imagery," *Remote Sensing*, vol. 13, no. 22, p. 4533, 2021.

[11] S. Wang and W. Tang, "Pavement crack segmentation algorithm based on local optimal threshold of cracks density distribution," in *International Conference on Intelligent Computing*. Springer, 2011, pp. 298–302.

[12] M. R. Jahanshahi, F. Jazizadeh, S. F. Masri, and B. Becerik-Gerber, "Unsupervised approach for autonomous pavement-defect detection and quantification using an inexpensive depth sensor," *Journal of Computing in Civil Engineering*, vol. 27, no. 6, pp. 743–754, 2013.

[13] M. R. Jahanshahi, S. F. Masri, C. W. Padgett, and G. S. Sukhatme, "An innovative methodology for detection and quantification of cracks through incorporation of depth perception," *Machine vision and applications*, vol. 24, no. 2, pp. 227–241, 2013.

[14] G. Moussa and K. Hussain, "A new technique for automatic detection and parameters estimation of pavement crack," in *4th International Multi-Conference on Engineering Technology Innovation, IMETI*, vol. 2011, 2011.

[15] B. J. Lee and H. D. Lee, "Position-invariant neural network for digital pavement crack analysis," *Computer-Aided Civil and Infrastructure Engineering*, vol. 19, no. 2, pp. 105–118, 2004.

[16] T. Saar and O. Talvik, "Automatic asphalt pavement crack detection and classification using neural networks," in *2010 12th Biennial Baltic Electronics Conference*. IEEE, 2010, pp. 345–348.

[17] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[19] S. Zhou and W. Song, "Concrete roadway crack segmentation using encoder-decoder networks with range images," *Automation in Construction*, vol. 120, p. 103403, 2020.

[20] F. Çelik and M. König, "A sigmoid-optimized encoder–decoder network for crack segmentation with copy-edit-paste transfer learning," *Computer-Aided Civil and Infrastructure Engineering*, 2022.

[21] L. Attard, C. J. Debono, G. Valentino, M. Di Castro, A. Masi, and L. Scibile, "Automatic crack detection using mask r-cnn," in *2019 11th international symposium on image and signal processing and analysis (ISPA)*. IEEE, 2019, pp. 152–157.

[22] K. Zhang, Y. Zhang, and H.-D. Cheng, "Crackgan: Pavement crack detection using partially accurate ground truths based on generative adversarial learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 2, pp. 1306–1319, 2020.

[23] W. He, C. Li, X. Nie, X. Wei, Y. Li, Y. Li, and S. Luo, "Recognition and detection of aero-engine blade damage based on improved cascade mask r-cnn," *Applied Optics*, vol. 60, no. 17, pp. 5124–5133, 2021.

[24] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[25] Z. Liu, Y. Cao, Y. Wang, and W. Wang, "Computer vision-based concrete crack detection using u-net fully convolutional networks," *Automation in Construction*, vol. 104, pp. 129–139, 2019.

[26] S. L. Lau, E. K. Chong, X. Yang, and X. Wang, "Automated pavement crack segmentation using u-net-based convolutional neural network," *IEEE Access*, vol. 8, pp. 114 892–114 899, 2020.

[27] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. B. Ayed, "Boundary loss for highly unbalanced segmentation," in *International conference on medical imaging with deep learning*. PMLR, 2019, pp. 285–296.

[28] S. A. Taghanaki, Y. Zheng, S. K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, and G. Hamarneh, "Combo loss: Handling input and output imbalance in multi-organ segmentation," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 24–33, 2019.

[29] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. IEEE, 2016, pp. 565–571.

[30] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2017, pp. 240–248.

[31] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[32] Y. Wang, X. Ma, Z. Chen, Y. Luo, J. Yi, and J. Bailey, "Symmetric cross entropy for robust learning with noisy labels," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 322–330.

[33] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[34] F. Lin, J. Yang, J. Shu, and R. J. Scherer, "Crack semantic segmentation using the u-net with full attention strategy," *arXiv preprint arXiv:2104.14586*, 2021.

[35] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei, and H. Ling, "Feature pyramid and hierarchical boosting network for pavement crack detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1525–1535, 2019.

Appendix for Crack Segmentation on FRP CT Images Using Residual Multi-scale Dilated Convolution and Reversed Cross Entropy Loss

Zhang Yuyao[1], Xu Yitao[2], Han Chenyu[2]
1 School of Mathematics, EPFL, Switzerland
2 School of Computer and Communication Science, EPFL, Switzerland

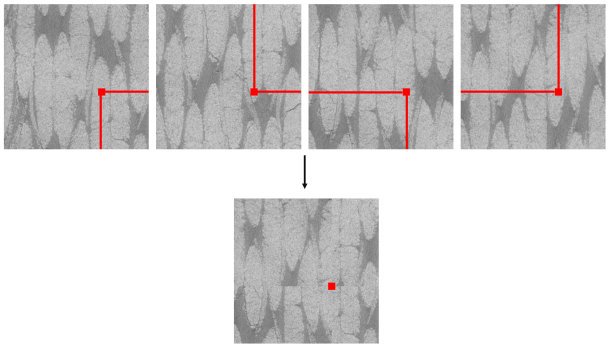

Figure 1. Example of Mosaic image augmentation. Selecting a random point in the four images, Mosaic algorithm cuts the images horizontally and vertically with an intersection on that point and picks one part in each image to form a new image.

I. Mosaic Data Augmentation

A. Details

Proposed in [1], Mosaic data augmentation has been proved to make the model generalize better via introducing different backgrounds in training images. Although this is a technique for object detection models, we assume image segmentation could also benefit from it. Given four images, Mosaic algorithm first selects the same random position in each image and cut the images horizontally and vertically with an intersection on that point. Then the algorithm randomly picks one part from each image and form the final image, called Mosaic image, as shown in figure 1. We perform Mosaic augmentation on our training images and keep the test set unchanged.

B. Experiment

Because of the two randomnesses (random images and random cutting point) in Mosaic augmentation, the number of extra train images that we can obtain is almost independent of the actual number of train images. We tried two schemes for Mosaic data augmentation, one is online and the other is offline. Online augmentation does the augmentation during the training and picks random images from the current batch, resulting in a larger batch size during training. The offline version has the access to the whole train set and randomly picks images from it to do the augmentation. For the online version, we enlarge the batch size from 6 to 12 by doing Mosaic six times and for the offline version, we generate 4000 new train images. Results are given in table I.

	Online	Offline
AIU	0.3706	0.3643
best IoU(threshold)	0.4009(0.46)	0.3981(0.45)

Table I
Results of training with Mosaic data augmentation. Both online and offline versions do not provide performance gain(40000 iterations).

Unfortunately, we can see that both online and offline versions do not provide performance gain. We assume that the reason is that Mosaic augmentation breaks the continuity of the original cracks and might lead to wrong fitting in some weird cracks that do not exist in the test set or even in the real world, causing a performance decrease.

II. Prediction Results

Here we show some predictions of our best model in figure 2. We can see that most of the cracks are detected. Moreover, notice that in the third column, the top part has a wrong label (a crack that is not there, the left one of the two vertical cracks). Our model successfully predicts the true crack position and is robust to noise in the label.

III. Results of 100000-epoch Experiments

We show the concrete results of all 100000-iteration experiments in table.

IV. Dense Atrous Convolution(DAC) and Residual Multi-scale Pooling(RMP)

Dense Atrous Convolution(DAC) module was first introduced by Gu et al.[2] for medical image segmentation. By performing four atrous convolutions on the same feature map, DAC can increase the receptive field. This leads to less sacrifice of resolution. Here we use a combination of kernels with sizes 1, 3, and 5 from top to bottom, which contributes to better performance. (Fig.3)

RMP consists of four convolution parts with kernel sizes 6, 5, 3, and 2. The input feature map is fed into these convolution layers and up-sampled by a 1x1 convolution.

Method / Metric	Basline	DAC+RMP	MDC+RMP	MDC+RMP+HFE
AIU	0.4247	0.4368	0.4380	0.4406(γ=0.1)
Best IoU (threshold)	0.4510 (0.35)	0.4601 (0.29)	0.4581(0.34)	0.4630(0.32)
Parameters	29.5M	47.1M	30.7M	32.3M

Table II
RESULTS OF 100000-ITERATION EXPERIMENTS.

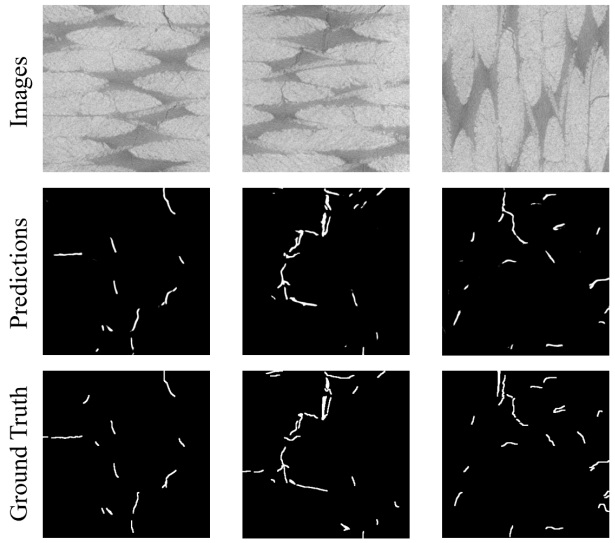

Figure 2. Examples of the prediction results.

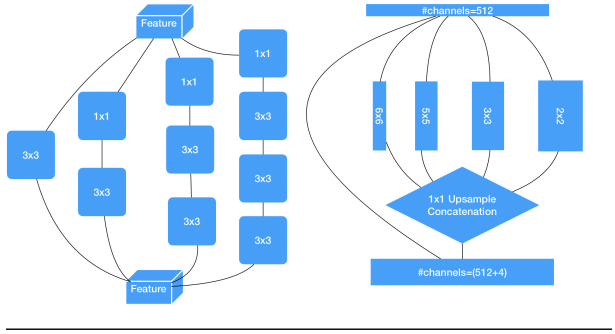

Figure 3. The graphic representation of DAC(Left) and RMP(Right). The DAC uses four cascade pipelines of convolutions with different sizes, which can enlarge the receptive field to extract more information. The RMP feds the input feature map into four convolutions of sizes 2, 3, 5, and 6, then the resulting feature map is upsampled to the height and weight size as the input to combine the extracted information with the input map

This 1x1 convolution can also reduce the computational cost. The outputs are finally concatenated with the original feature map to gather information.

V. ABLATION STUDY FOR γ IN HFE MODULE

We've also done an ablation study for the parameter γ in the HFE module(40000 iterations). For both DAC+RMP+HFE and MDC+RMP+HFE we set $\gamma = 0.1, 0.3, 0.5, 0.7, 0.9$ respectively, and it turned out for DAC+RMP+HFE, $\gamma = 0.3$ is the optimal choice, and for MDC+RMP+HFE $\gamma = 0.1$ is the optimal choice. The results are in Table III. and Table IV. We think the degradation in performance is because too many high-frequency features are captured, which causes damage to the original information.

Metric \ γ	0	0.1	0.3	0.5	0.7	0.9
AIU	0.3827	0.3753	**0.3774**	0.3689	0.3518	0.3214
Best IoU (threshold)	0.4182(0.37)	0.4142(0.39)	**0.4175(0.3)**	0.4076(0.32)	0.3866(0.35)	0.3515(0.35)

Table III
RESULTS OF 40000-ITERATION EXPERIMENTS FOR DAC+RMP +HFE.

Metric \ γ	0	0.1	0.3	0.5	0.7	0.9
AIU	0.3768	**0.3832**	0.3699	0.3683	0.3335	0.3149
Best IoU (threshold)	0.4139(0.35)	**0.4169(0.41)**	0.4056(0.35)	0.4013(0.39)	0.3636(0.37)	0.3414(0.42)

Table IV
RESULTS OF 40000-ITERATION EXPERIMENTS FOR MDC+RMP +HFE.

REFERENCES

[1] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[2] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE transactions on medical imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.