CrackSegConnect: a Crack Inpainting Network based on Segmentation Model

Lizhou Chen, and Luoyu Zhou

Abstract—**Crack is one of the most common structure distresses which often appear in the engineering construction systems, and thus crack detection and evaluation is particularly important in structural health monitoring. However, defects in the captured crack image occur frequently due to the complex circumstance conditions, bringing difficulty to identification work and observation of crack property. Therefore, it is necessary to inpaint the crack before detection and evaluation. In this paper, we propose a crack inpainting network based on segmentation model. The network introduces segmentation results as prior knowledge, which can improve the inpainting results. The experimental results have demonstrated that the proposed network can inpaint the crack structure with high accuracy.**

Index Terms—**Crack Inpainting; Segmentation Model; Generative Adversarial Network; Crack Segmentation Perceptual Loss**

I. Introduction

CRACKS are the form of damages commonly appear in the engineering construction systems. Large quantity of manpower and material resources consumption have been plunged into engineering construction system maintenance, including inspection, repair and maintenance. Cracks are one of the most significant performance indexes of safety problems in road, bridge and other structures. Crack detection is one of the effective ways to solve the maintenance structure problems like roads and bridges. However, visual defects often occur due to the terrible environment, e.g. light insufficiency, structural ruins, occlusion and so on. These defects demolish the crack and increase the difficulty of crack detection and evaluation. It is necessary to develop an approach to repair this information that is missing for environmental reasons. Image inpainting techniques are suitable for these tasks. Although crack detection methods have become a hot topic over the past few years, less attention has been paid on crack inpainting [1-3].

Traditional image inpainting methods cannot deal with ruined crack images. Early works attempt to utilize background patches close to the defect of crack regions to solve this problem. Visual Memex [4] uses context information to nonparametrically model object relations and predict mask objects in scenes, while [5] uses implemented Bezier-exemplar hybrid based image inpainting method to eliminate out the wrinkle of the image. However, as these approaches assume the lost information can be found in the background area, they cannot hallucinate complex and non-repetitive structure contents, especially high-level semantic features.

Recent years, deep learning methods make rapid progresses in image processing. Convolutional neural networks (CNN) [6] and generative adversarial networks (GAN) [7] formulate the image inpainting as a background conditioned generation problem. The high-level semantic features produced by neural network are very likely to replace the local and background features utilized in traditional methods.

Contextual Autoencoder [8] introduces an encoder -decoder architecture network and a code-to-code layer to generate plausible inpainting results. GLGC [9] introduces local and global discriminator to determine the quality and global coherence of the local patches. Contextual Attention [10] introduces contextual attention to consider the patch-wise correlation, which enhances the image coherence and quality. Partial Convolution [11] redesigns the convolutional kernel and introduces an iterative mask generation process to improve the inpainting performance. Gated Convolution [12] adapts soft gating process to adaptively learn the mask generation process, reducing checkerboard effect and artifacts. Probabilistic Diverse GAN [13] adapts a feature representation method to give conditional information to a generator network to produce pluralistic results. Edge Connect [14] introduces a two-stage method which firstly generates edge map of the defect regions and then utilizes this edge map to provide prior information for the patch generation process. Furthermore, these inpainting methods are designed for normal objects like natural scene, faces, and so on. When it comes to crack, it is not suitable and cannot get higher performance as they don't take crack characteristics into account.

To this end, we present CrackSegConnect, a crack inpainting network based on segmentation model. The network include two stages. The first stage is an encoder -decoder architecture network, noted as Crack Structure Reasoning Network, which is used to reason the edge map of the masked crack region. The second stage is a generative inpainting network with gated convolutions and self-attention branches, noted as Crack Image Completion Network, which is used to generate high quality crack. Moreover, the network introduces segmentation results as prior knowledge, noted as Crack Segmentation Perceptual Loss, which can further improve the inpainting results.

Manuscript received August 24, 2022; revised December 28, 2022.

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61901059 and 51978079.

Lizhou Chen is a postgraduate student of Electronic and Information School of Yangtze University, Jingzhou 434023, China (email: 625075581@qq.com)

Luoyu Zhou is an associate professor of Electronic and Information School of Yangtze University, Jingzhou 434023, China (Corresponding author, email: luoyuzh@yangtzeu.edu.cn, ORCID: orcid.org/0000-0003-4417-1250)

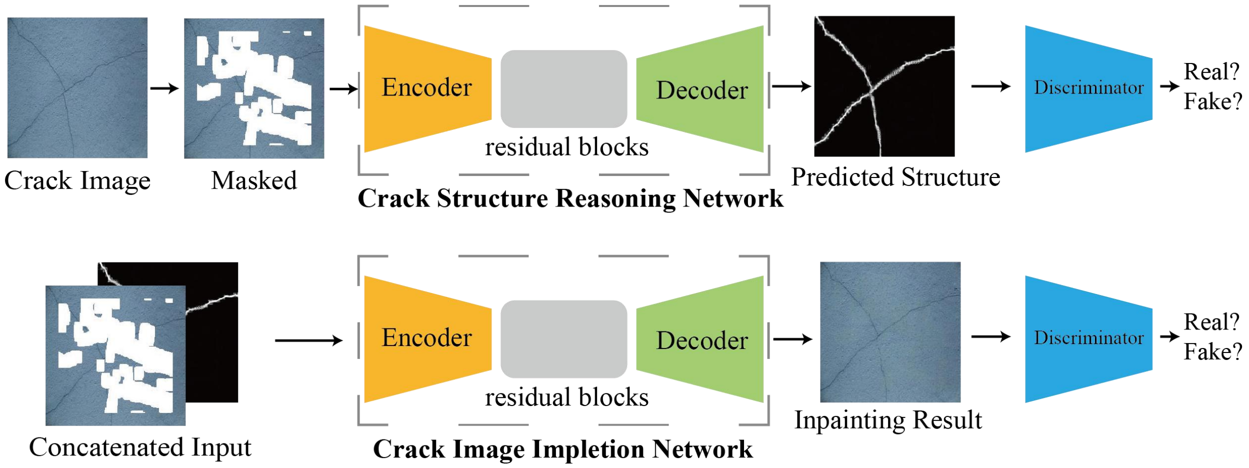

Fig. 1. The overall network structure. CrackSegConnect consists of two stages, Crack Structure Reasoning Network and Crack Image Impletion Network. The first one takes a crack image and mask as input and output the predicted structure. The second one concatenates the mask image and the predicted structure as input and output the inpainting result.

Experiments have demonstrated that the proposed network can generate high-quality crack inpainting results. Our contributions are summarized as follows:

1) We propose Crack Segmentation Perceptual (CSP) loss which can facilitate crack inpainting. With the help of CSP loss, our network can generate high-quality crack inpainting results, and thus the missing crack information can be effectively inpainted.

2) We introduce several techniques including gated convolutional block and self-attention mechanism to improve the inpainting method. Ablation studies show that these techniques can improve the crack inpainting performance.

3) Our unified feed-forward generative networks achieve high inpainting performance and produce high-quality results on cracks datasets.

The rest of this paper is organized as follows. Section II introduces the proposed network architecture. Section III presents experiment set and Section IV presents their results. Finally, Section V summarizes the conclusions and future work.

II. CrackSegConnect Architecture

Our overall architecture follows the backbone network similar to the method proposed by Kamyar Nazeri et al. [14]. We propose crack inpainting networks consist of two stages: 1) crack structure reasoning model that generates the crack edge map of the mask region, 2) crack image completion network that generates inpainting results of crack image via generated edge map and background information. Each stage consists of a generator and a discriminator. We introduce several techniques such as self-attention block and simultaneously propose a new loss function such as crack segmentation perceptual loss to improve the crack image inpainting performance. Fig. 1 shows the overall structure of our method.

A. Crack Structure Reasoning Network

The Crack Structure Reasoning Network follows the generator-discriminator architecture. The generator consists of an encoder block, 8 layers residual block and a decoder block. The discriminator is a patch-GAN discriminative network consists of 5 convolution layers. Let G_e be the edge generator, D_e be the edge discriminator, I_{gt} be the input ground truth image, I_{gray} be the grayscale image of I_{gt}, and E_{gt} be the edge maps. For the mask image, we set 1 for the mask region, 0 for the background. Let the mask be M, then the mask of ground truth grayscale $MI_{gray} = I_{gray} \odot (1-M)$. $\odot$ denotes Hadamard product. The mask of ground truth edge map is $ME_{gray} = E_{gray} \odot (1-M)$. Then, we use the edge generator to predict the edge map E_{pred}=G_e(MI_{gray}, ME_{gt}, M).

The E_{gt} and E_{pred} with conditional information I_{gray} are the inputs of the discriminator. The discriminator predicts whether the generated edge map is of high quality. The network is trained with an objective function, which is comprised of adversarial loss, feature-matching loss and smooth-L1 loss.

$$L_{G_{SRN}} = \lambda_{a\text{-}SRN} L_{a\text{-}SRN} + \lambda_{FM} L_{FM} + \lambda_{l1} L_{l1} \quad (1)$$

where $\lambda_{a\text{-}SRN}$=1, λ_{FM}=10 , λ_{l1}=0.05 are all regularization parameters. The adversarial loss follows the conditional GAN formulation and is defined as

$$\begin{aligned} L_{a\text{-}SRN} &= \mathrm{E}_{(E_{gt}, I_{gray})}\left[\log D_e(E_{gt}, I_{gray})\right] \\ &+ \mathrm{E}_{I_{gray}} \log\left[1 - D_e(E_{gt}, I_{gray})\right] \end{aligned} \quad (2)$$

The feature-matching loss L_{FM} compares the activation maps in the intermediate layers of the discriminator. This stabilizes the training process by forcing the generator to produce results with representations that are similar to real images. This is similar to perceptual loss [15], where activation maps are compared with those from the pre-trained VGG network. The feature matching loss L_{FM} is defined as

$$L_{FM} = \mathrm{E}\left[\sum_{i=1}^{L} \frac{1}{N_i} \left\| D_e^{(i)}(E_{gt}) - D_e^{(i)}(E_{pred}) \right\|_1\right] \quad (3)$$

where L is the final convolution layer of the discriminator, N_i is the number of elements in the i'th activation layer, and $D_e^{(i)}$ is the activation in the i'th layer of the discriminator.

Intuitively, if the output and ground truth images are similar, their high-level semantic features are closer. Therefore, we apply L1 distance to measure the similarity of the feature maps.

B. Training-Testing Framework

The architecture of Crack Image Completion Network is similar to the Crack Structure Reasoning Network, that is the generator-discriminator architecture. The generator consists of an encoder block, 8 layers residual block and a decoder block. The discriminator is a patch-GAN discriminative network consists of 5 convolution layers.

Given the input ground truth image I_{gt}, the input of network is the mask image $MI_{gt} = I_{gt} \odot (1-M)$ with the condition edge map $E_{merge} = E_{pred} \odot M + E_{gt} \odot (1-M)$. E_{pred} is the output of the crack structure reasoning network. Let G_{im} be the inpainting generator of the crack image completion network, and D_{im} be the discriminator, we get $I_{pred} = G_e(MI_{gt}, E_{merge})$.

Then our overall loss function of Crack Image Completion Network is composed of adversarial loss, perceptual loss, L1 loss, style loss, and the proposed crack segmentation perceptual loss.

$$L_{G_{ICN}} = \lambda_{a\text{-}ICN} L_{a\text{-}ICN} + \lambda_{perc} L_{perc} + \lambda_{l1} L_{l1} + \lambda_{st} L_{sty} + \lambda_{csp} L_{csp} \quad (4)$$

where $\lambda_{a\text{-}ICN}$=1, λ_{perc}=0.1, λ_{l1} =1, λ_{st}=250, λ_{csp}=0.05 are regularization parameters.

The adversarial loss and L1 loss are similar to the Crack Structure Reasoning Network (Section II.A). The perceptual loss and style loss [15, 16] are two common loss functions in the image generation field. Perceptual loss uses VGG-19 network to extract feature maps to compare semantic similarity. It is defined as

$$L_{perc} = \mathrm{E}\left[\sum_{i=1} \frac{1}{N_i} \left\| \phi_i(I_{gt}) - \phi_i(I_{pred}) \right\|_1 \right] \quad (5)$$

where ϕ_i is the activation map of the pre-trained network layer i. We use activation maps from different layers of the VGG-19 network pre-trained on the ImageNet dataset [17] as ϕ_i. These activation maps are also used to compute style loss which measures the differences between covariances of the activation maps. Given feature maps of sizes $C_j \times H_j \times W_j$, style loss is computed by

$$L_{style} = \mathrm{E}_j \left[\left\| G_j^{\phi}(MI_{pred}) - G_j^{\phi}(MI_{gt}) \right\|_1 \right] \quad (6)$$

where G_j^{ϕ} is a $C_j \times C_j$ Gram matrix constructed from activation maps ϕ_i. Style loss is shown by Sajjadi et al. [18] and it is an effective tool to combat “checkerboard” artifacts caused by transpose convolution layers [19].

The proposed crack segmentation perceptual loss L_{csp} compares the activation maps output of the intermediate layers of U-net crack segmentation model. It plays as a metric to the accuracy of the generated crack. The details of crack segmentation perceptual loss will be discussed in Section II.C.

C. Crack Segmentation Perceptual Loss

1) Crack Segmentation Network as a Feature Extractor

High-level crack semantic information is complex and hard to evaluate. The most common methods are pixel-wise metrics like MAE and SSIM [20]. However, these metrics are designed for comparison of clear crack images, lacking guidance on the process of crack image generation. Pixel-wise metrics might be good as indices at evaluation stage. However, during training stage, low quality inpainting results have extremely low scores on these metrics. These indices also fail into translation invariance and rotation invariance. Recently, high level semantic information is widely applied for image generation. This high-level semantic information is feature maps extracted from a well-pretrained network, like VGG [21] and inception [22] network. Therefore, the high-level information of crack image can also be represented by a pre-trained feature extractor. Fig. 2 shows that crack segmentation network works as a feature extractor through weighted sum of the feature maps.

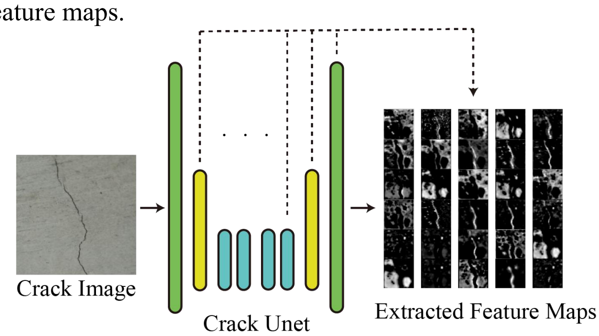

Fig. 2. Feature extractor of crack images.

VGG and inception network are two classical classification models. When it comes to crack images, classification model may not be available. For convenience, self-supervised methods like semantic segmentation should be taken into account. Semantic segmentation aims to consider the position and distribution of objects in images, and thus it can learn feature maps for representing the shape of objects. Therefore, trace and structure of the crack can be learned by a segmentation model. The feature maps extracted by a well pre-trained crack segmentation model can be utilized as a fine high-level semantic information of crack images.

There are many semantic segmentation methods like U-net, Ladder-net and Deep Lab-v3 etc. In this paper, we choose U-net owing to its high efficiency. We adapt U-net model pre-trained on crack image dataset as the feature extractor, noted as CrackUnet. The CrackUnet architecture consists of a symmetric multi-scale encoder and decoder. The multi-scale encoder uses convolution layers to extract the feature and down-samples the input image, then outputs the multi-scale feature codes. The decoder up-samples and transforms the multi-scale code information from source domain to target domain and uses transpose convolution layers to calculate the output images. The U-net is trained by L1 loss, BCE loss, and dice loss [23]. We adopt crack image and its segmentation label as mask pair of train images.

2) Crack Segmentation Perceptual Loss

Image segmentation models aim to capture the position information, so crack image segmentation models can provide relatively accurate information of crack and background. Feature maps extracted from U-net models can be an evaluation metric for inpainting results and real images.

Given predicted image I_{gt} and its ground truth image I_{pred}, both feed them into the model. Let φ_i be the i_{th} layer activation feature map inferred by CrackUnet, we define the crack segmentation perceptual (CSP) loss by weighted sum of distance φ_i as

$$L_{csp} = \mathrm{E}\left[\sum_{i} \frac{1}{N_i} \left\| \varphi_i(I_{gt}) - \varphi_i(I_{pred}) \right\|_1 \right] \quad (8)$$

Fig. 4 shows the illustration of CSP loss through CrackUnet. If the inpainting images and ground truth images are closer at high-semantic feature space, the intermediate embedding of the feature extractor should be the same. The intermediate layer feature maps are similar, proving that high-level information represented by the CrackUnet is the same. It could work as a superior metric for crack inpainting quality. Experimental results show that the crack inpainting results can be improved with the CSP loss, as discussed in Section IV.

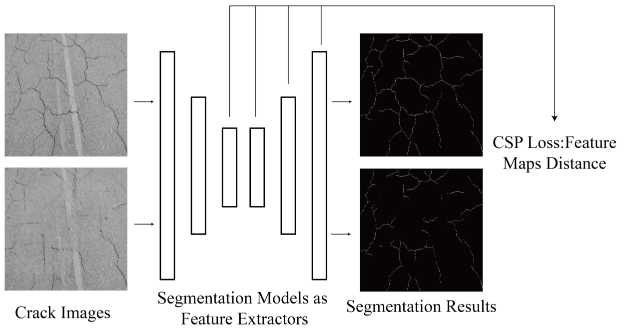

Fig. 3. Illustration of CSP loss through crack U-net.

Moreover, inspired by the superiority of gated convolution and self-attention, we adapt gated convolution and self-attention for crack inpainting, which also discussed in Section IV.

III. Experiments Set

A. Dataset

The dataset used in this paper is Concrete Crack Images for Classification Dataset and Crack Segmentation Dataset.

Concrete Crack Images for Classification Dataset (http://dx.doi.org/10.17632/5y9wdsg2zt.1) is proposed by the researchers from Middle East Technical University. It contains 20000 images of road cracks and is separated into training set of 16000 images and validation set of 4000 images. Crack Segmentation Dataset (https://www.kaggle.com/lakshaymiddha/crack-segmentation-dataset) contains around 11300 images. We divide it into training set of 9600 images and validation set of 1700 images. Fig.4 shows some typical examples of the datasets.

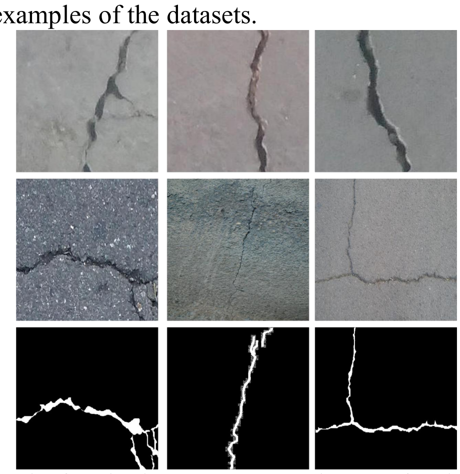

Fig.4. Sample of crack. Line 1: Concrete Crack Images for Classification Dataset. Line2,3: images and their masks in Crack Segmentation Dataset.

Concrete Crack Images for Classification Dataset (dataset 1) doesn't provide segmentation labels while Crack Segmentation Dataset (dataset 2) provides segmentation labels. For dataset 1, we apply canny operator to get the edge map of crack image. For dataset 2, we use the segmentation labels as edge maps.

B. Training Strategy

We use PyTorch to implement the network and run the experiments. The input size to the model is set to 256×256. The model is optimized using Adam optimizer. Generators G1, G2 are trained separately using edge map learning rate 10^{-4} until the losses plateau. We decrease the learning rate to 10^{-5} and continue to train G1 and G2 until convergence. Finally, we fine-tune the networks by removing D1, then train G1 and G2 end-to-end with learning rate 10^{-6} until convergence. Discriminators are trained with a learning rate one tenth of the generator.

IV. Experiments and Analysis

In order to demonstrate the inpainting performance of our proposed network, we compare the quantitative and qualitative results with several baseline models, including Edge Connect [14], Gated Convolution [12], Contextual Attention [10] and Patch Match [24]. For ablation study, we explore the property of our techniques, including CSP loss, gated convolution and self-attention.

For the crack inpainting results, they are evaluated by the following metrics: 1) peak signal-to-noise ratio (PSNR), 2) structural similarity index (SSIM), 3) mean average error (MAE).

A. Qualitative Comparison and Quantitative evaluation

Results compared with several image inpainting models are shown in Fig. 5. The edge map of the mask crack is generated clearly and sharply by the crack structure reasoning network. By introducing self-attention, the crack structure can be reasoned well, providing better precondition to the crack image completion network. The completion network takes generated crack structure as precondition and can achieve high-quality crack. One-stage methods [12, 10, 24] only concentrate on the background information and loss the information of crack traces. These one-stage methods obtain unsatisfactory inpainting results with many residual masks, as shown in Fig.5 (e~g). In contrast, two-stage methods, including EdgeConnect [14] and our method, can obtain satisfactory results by implementing crack structure reasoning network and crack image completion network, as shown in Fig.5 (c~d). Moreover, by introducing CSP loss, our method can produce both precise and high-quality inpainting results. Especially when masks are relatively large and only very limited crack information is visible, our method can generate more accurate crack information. As CSP loss forces the network to consider the accuracy of the crack structure, lost crack information of the mask area can be reconstructed at a great extent. Obviously, the thin cracks can be inpainted in Fig.(c), but they are defective in Fig.(d). More comparative samples with EdgeConnect [14] are shown in Fig.6.

The evaluation values for crack inpainting results from dataset 1 and dataset 2 are respectively shown in Table 1 and Table 2. Obviously, our method can achieve relatively higher scores than other methods, which demonstrates that our proposed method is very successful in crack inpainting.

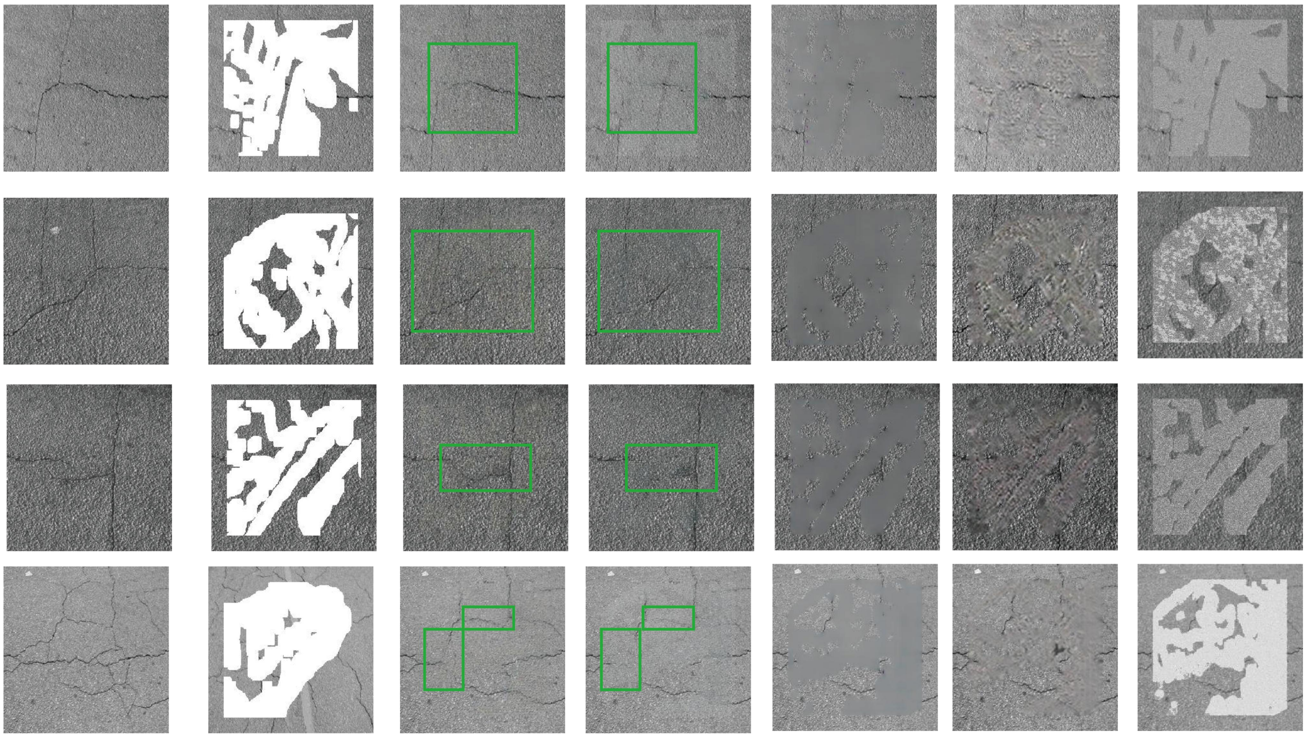

Fig. 5. Result samples by our method and comparative methods. (a) original images, (b) mask images, (c) our method, (d) EdgeConnect [14] , (e) Gated Convolution [12], (f) Contextual Attention [10], (g) Patch Match [24]

TABLE I
QUANTITATIVE RESULTS FROM CONCRETE CRACK IMAGES FOR CLASSIFICATION DATASET (DATASET 1)
(* HIGHER IS BETTER, † LOWER IS BETTER)

Model/metric	PSNR*	SSIM*	MAE †
CrackSegConnect (Ours)	**32.18**	**0.82**	**0.0176**
EdgeConnect [14]	29.93	0.79	0.0184
Gated Convolution [12]	27.65	0.74	0.0203
Contextual Attention [10]	26.45	0.76	0.0210
Patch Match [24]	24.30	0.67	0.0304

TABLE II
QUANTITATIVE RESULTS FROM CRACK SEGMENTATION DATASET (DATASET 2) (* HIGHER IS BETTER, † LOWER IS BETTER)

Model/metric	PSNR*	SSIM*	MAE †
CrackSegConnect (Ours)	**30.56**	**0.89**	0.0385
EdgeConnect [14]	26.80	0.84	0.0404
Gated Convolution [12]	25.82	0.62	0.0448
Contextual Attention [10]	24.37	0.68	0.0472
Patch Match [24]	21.65	0.54	**0.0243**

B. Ablation Study

For ablation study, we conduct controlled trials which study the effect of CSP loss, gated convolution, as well as self-attention for crack image completion network and crack structure reasoning network. Results show that both these techniques can improve the performance of the inpainting quantitatively and qualitatively.

1) The ablation study on CSP loss for crack image completion network

CSP loss is a precise evaluate metric totally designed for crack images, and thus it benefits the training a lot and can generate clear and sharp crack inpainting images. For different layers of crack segmentation model, guidance of the loss to the network training process is different. We study the inpainting performance by using different layers of model. For experiment convenience, we only use the crack image pair, including its edge map and ground truth, to train the crack image completion model.

To explore the effects of CSP loss in different layer, we respectively use each single layer and some layer combination to train our model (CrackSegConnect Network). More detailed results are shown in Fig.7. It's shown to us that the network obtain best PSNR using the 3rd and 7th layer of CrackSegConnect. We observed that 3rd layers and 7th layers are respectively the final layers of encoder and decoder. The reason of this phenomenon is that the final layers can represent high-level global information, which can guide the inpainting network to form better results.

2) The ablation study on gated convolution and self-attention for crack image completion network

Gated convolution can handle the defect regions better than vanilla convolution. Self-attention layers have powerful capacity to associate information of different positions. The experiment shows that it can achieve better inpainting results with self-attention layers. We compare the baseline model and the improved model with self-attention and gated convolution. Moreover, to measure the upper bound of the crack image completion network's capacity, we only train the network with ground truth crack edge information (ground truth label), rather than generated edge information by crack structure reasoning network. The quantitative results are shown in Table III.

TABLE III
QUANTITATIVE RESULTS OF CRACK IMAGE COMPLETION NETWORK WITH GATED CONVOLUTION AND SELF-ATTENTION (TRAINED WITH GROUND TRUTH LABEL). (* HIGHER IS BETTER, † LOWER IS BETTER.)

Model/metric	SSIM*	MAE †
Crack Image Completion Network	32.30	0.0138
Crack Image Completion Network + Self-attention + Gated Convolution	**32.93**	**0.0128**

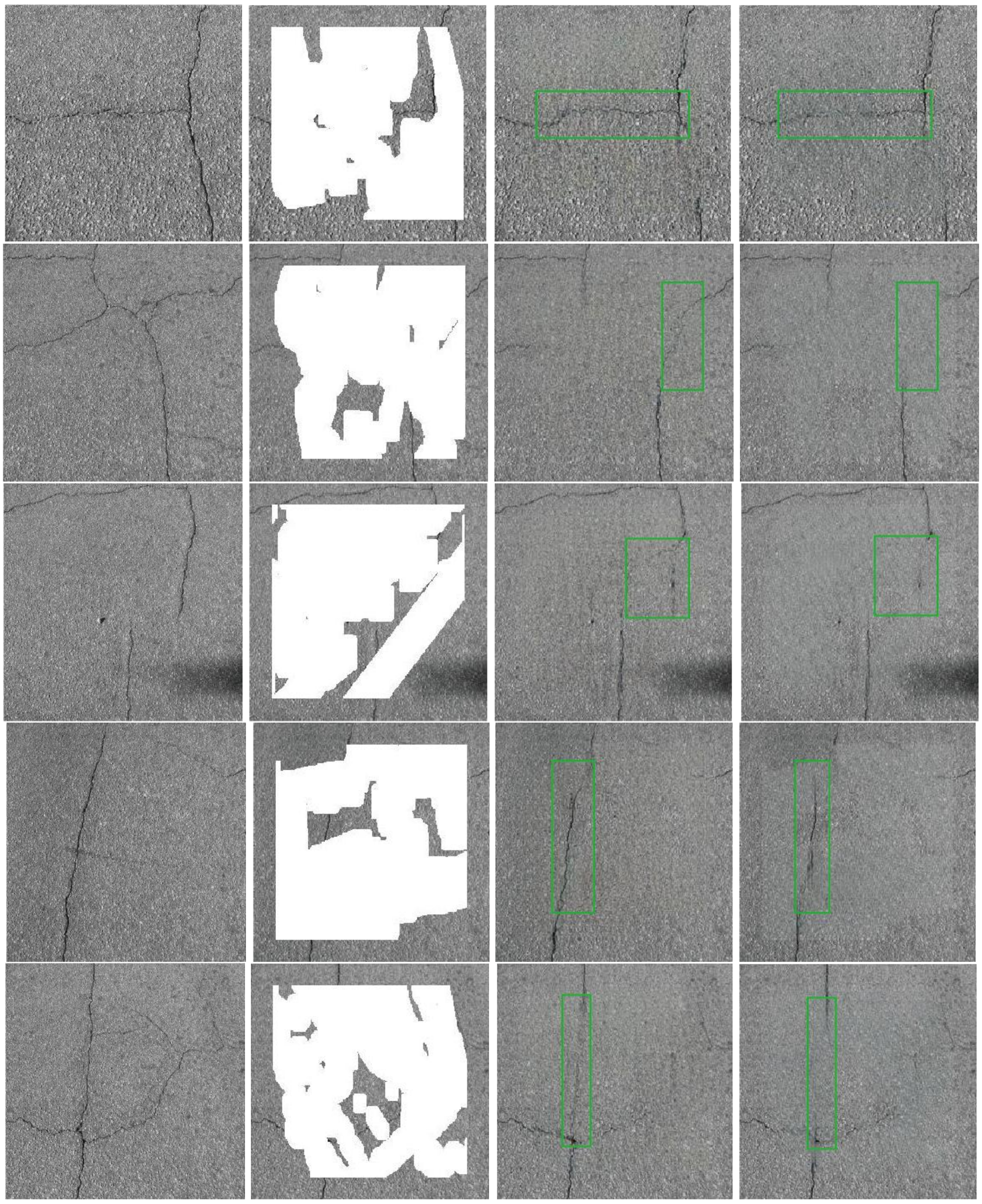

Fig. 6. Result samples from our models and EdgeConnect. (a) original images, (b) mask images, (c) CrackSegConnect (our method), (d) EdgeConnect [14].

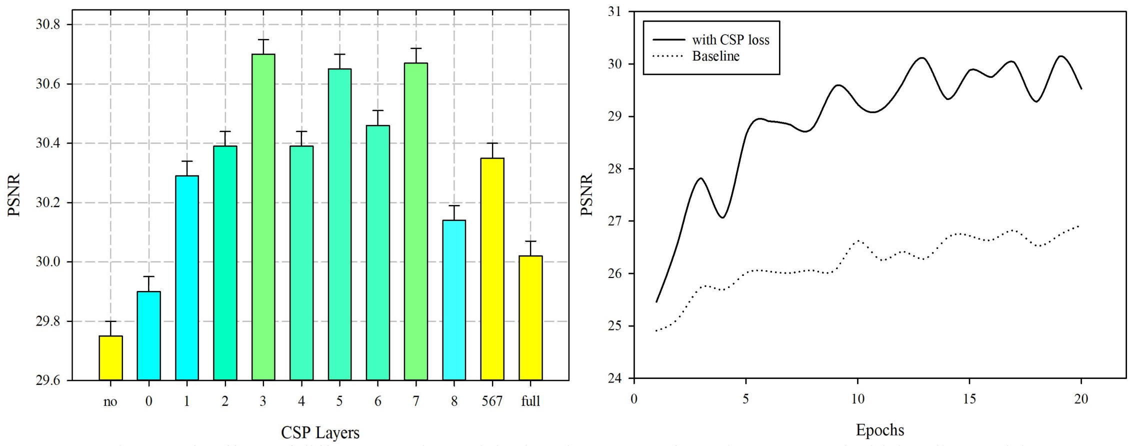

Fig. 7. Left: effects of different CSP loss. Right: learning curves of CSP loss compared with baseline model.

3) The ablation study on self-attention for crack structure reasoning network.

Self-attention layers have powerful capacity to associate information of different positions, and thus are very suitable for edge model to generate clear edge maps. For comparison, we test the effects of Crack Structure Reasoning Network with and without self-attention. For generated edge maps, we measure the quality of our results using three metrics, including precision, recall, and F1 score. The quantitative results are shown in Table IV, which presents our model can achieve better edge structure reasoning results.

TABLE IV
QUANTITATIVE RESULTS FROM CRACK SEGMENTATION DATASET
(* HIGHER IS BETTER, † LOWER IS BETTER)

Model/metric	Precision*	Recall*	Dice(F1-score)*
Crack Image Completion Network	0.1104	0.0836	0.0950
Crack Image Completion Network + Self-attention + Gated Convolution	**0.1325**	**0.1565**	**0.1435**

V. CONCLUSION

In this paper, we propose CrackSegConnect, a crack inpainting network based on segmentation model. The network firstly reasons the crack structure of the defect regions and then generates inpainting results. It can generate sharp crack structure and high-quality inpainting results. Qualitative results and quantitative comparisons demonstrate that our methods can achieve superior performance.

In addition, we propose CSP loss, a powerful loss function to extract high level representation of crack images by a pre-trained semi-supervised crack segmentation model. With the help of CSP loss, we can generate high quality crack images while comparative methods tend to ignore some important thin crack information.

For future works, there are two main aspects. The main shortcoming of our method is that the model size is relatively large. Furthermore, it relies on large amount of training data. Therefore, one of the future works is the lightweight model that reduces the reliance on training data without degrading performance. The other aspect is crack segmentation perceptual loss. How to train more efficient segmentation models based on very limited data remains a challenging issue. Meanwhile, how to effectively utilize middle layers of segmentation models that can represent abundant layer information to assist inpainting process is also worth exploring.

REFERENCES

[1] F. Panella, A. Lipani, J. Boehm, “Semantic Segmentation of Cracks: Data Challenges and Architecture”, *Automation in Construction*, 135, 1-13, 2022.

[2] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei, H. Ling, “Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection”, *IEEE Transactions on Inteligent Transportation Systems*, 21(4), pp. 1525-1535, 2020.

[3] W. Wang, C. Su, “Semi-supervised Semantic Segmentation Network for Surface Crack Detection”, *Automation in Construction*, 128, 1-10, 2021.

[4] T. Malisiewicz, A. Efros, “Beyond Categories: The Visual Memex Model for Reasoning about Object Relationships”, *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, 22, 1222-1230, 2009.

[5] K. Sim, C. Toa, Z. Lim, L. Kathy, “Bezier-Exemplar Hybrid based Image Inpainting Method in Computerized Colorimetric Protein Estimation System”, *IAENG International Journal of Computer Science*, 49(2), pp463-468, 2022.

[6] Y. LeCun, Y. Bengio, G. Hinton, “Deep Learning”, *Nature*, 521, 436 -444, 2015.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, “Generative Adversarial Nets”, *Advances in Neural Information Processing Systems*, 27, 2672-2680, 2014.

[8] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, A. Efroset, “Context Encoders: Feature Learning by Inpainting”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.2536-2544, 2016.

[9] S. Iizuka, E. Simo-Serra, H. Ishikawa, “Globally and Locally Consistent Image Completion”, *ACM Transactions on Graphics*, 36, 1-14, 2017.

[10] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, T. Huang, “Generative Image Inpainting with Contextual Attention”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5505-5514, 2018.

[11] G. Liu, F. A. Reda, K. J. Shih, T. Wang, A. Tao, B. Catanzaro, “Image Inpainting for Irregular Holes using Partial Convolutions”, *European Conference on Computer Vision*, 85-100, 2018.

[12] J. Yu, Z. Lin, J. Yang, J. Yang, X. Shen, X. Lu, T. Huang, “Free-Form Image Inpainting With Gated Convolution”, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4471-4480, 2019.

[13] H. Liu, Z. Wan, W. Huang, Y. Song, X. Han, J. Liao, “Pd-GAN: Probabilistic Diverse Gan for Image Inpainting”, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9371-9381, 2021.

[14] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, M. Ebrahimi, “Edgeconnect: Structure Guided Image Inpainting using Edge Prediction”, *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019.

[15] J. Johnson, A. Alahi, L. Fei-Fei, “Perceptual Losses for Real-time Style Transfer and Super-Resolution”, *European Conference on Computer Vision*, 694-711, 2016.

[16] L. A. Gatys, A. S. Ecker, M. Bethge, “Image Style Transfer using Convolutional Neural Networks”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2414-2423, 2016.

[17] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge”, *International Journal of Computer Vision*, 115, 211-252, 2015.

[18] M. S. Sajjadi, B. Scholkopf, M. Hirsch, “Enhancenet: Single Image Super-Resolution through Automated Texture Synthesis”, *Proceedings of the IEEE International Conference on Computer Vision*, 4491-4500, 2017.

[19] A. Odena, V. Dumoulin, C. Olah, E. Simoncelli, “Deconvolution and Checkerboard Artifacts”, *Distill*, 1(10), 2016.

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, et al. “Image Quality Assessment: from Error Visibility to Structural Similarity”, *IEEE Transactions on Image Processing*, 13(4), 600-612, 2004.

[21] K. Simonyan, A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, *arXiv preprint* arXiv:1409.1556, 2014.

[22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, “Going Deeper with Convolutions”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1-9, 2015.

[23] J. Zhuang, “LadderNet: Multi-Path Networks based on U-Net for Medical Image Segmentation”, *arXiv preprint* arXiv:1810.07810, 2018.

[24] C. Barnes, E. Shechtman, A. Finkelstein, D. Goldman “PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing”, *ACM Trans. Graph*, 28, 1-12, 2009.

Lizhou Chen received the bachelor's degree in electronic information from SichuanNormal University, Chengdu,China, in 2020, where he is currently pursuing the M.S. degree in signal and information processing in Yangtze University, Jingzhou, China. His current research interests include structural health monitoring and image processing..

Luoyu Zhou received the Ph.D. degree in information engineering from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, (CIOMP, CAS), Changchun, China, in 2013. He is currently an associate professor with the School of Electronics and Information, Yangtze University, China. His current research interests include computer vision and artificial intelligence.