Integrated APC-GAN and AttuNet framework for automated pavement crack pixel-level segmentation: A new solution to small training datasets

Tianjie Zhang
Department of Computer Science
Boise State University, Boise, Idaho, 83725
Email: tjzhang@u.boisestate.edu

Donglei Wang
Department of Civil Engineering
Boise State University, Boise, Idaho, 83725
Email: dongleiwang@u.boisestate.edu

Yang Lu
Department of Civil Engineering
Boise State University, Boise, Idaho, 83725
Email: yangfranklu@boisestate.edu

Word Count: 5,297 words + 1 table (250 words per table) = 5,547 words

Submitted [07/30/2022]

ABSTRACT

Pavement crack segmentation using deep learning methods can improve crack segmentation accuracy, but in many cases the training dataset is lacking or uneven, making it insufficient to train an accurate segmentation model. In this work, an integrated APC-GAN and AttuNet framework was proposed as an automated pavement surface crack pixel-level segmentation strategy for small training datasets. First, an automated pavement crack generative adversarial network (APC-GAN) was designed for the pavement cracks data as an image augmentation method, which was modified and improved from a traditional deep convolutional generative adversarial network (DCGAN). Then, a novel pixel level semantic segmentation structure, Attunet, was proposed by introducing the attention module into the convolutional network work structure. Another AttuNet-min was proposed by replacing the max pooling layer and activation function in AttuNet. In order to assess the performance of PC-GAN and AttuNet framework, an open-source dataset DeepCrack was used, which only contains 300 training images. The results show that APC-GAN could produce more clear and distinct pavement images than DCGAN and more diversity than traditional augmentation method. The AttuNet model with APC-GAN can reach higher accuracy in evaluation metrics than other augmentation methods. As for the segmentation model comparison, APC-GAN and AttuNet framework gain the highest value in recall, F1 score, mean Intersection over Union (mIoU) and mean pixel accuracy (mPA) among all models including U-Net, DeepLabv3, FCN and LRASPP, while the AttuNet-min gain the highest mean precision.

Keywords: Crack segmentation, GAN, attention module, deep learning

INTRODUCTION

Crack has become one of the primary defects in pavement, which seriously affects the service life of pavement (*1*). The traditional crack detection methods like counting cracks manually are labor-intensive and time-consuming (*2*). The automation of crack detection and segmentation has become the focus of research in civil engineering and highway agencies(*3*). Researchers have proposed a series of automated detection methods in pavement cracks detection. Image processing and machine learning are common and traditional methods in crack detection and extraction (*4; 5*). The main idea of popular machine learning-based crack inspection methods is to learn and summarize the characteristics of crack and then build a model to make predictions. Support Vector Machine (SVM) (*6*) and K-means algorithm (*7*) have already been used in the classification and segmentation of crack images. A crack detection approach based on Local Binary Patterns (LBP) with SVM was proposed by Cheng (*6*), which can extract the LBP feature of cracks and make a segmentation from each frame of the video taken from the road. Ai, D. et al. (*8*) used multi-neighborhood information to segment cracks and resulted with the F1 score of 0.8. The accuracy of the algorithm proposed by Kaddah, W. et al. (*9*) is about 75%, which is based on the improved minimum path method, an image processing method, to segment pavement cracks.

However, image processing and traditional machine learning are heavily dependent on an engineer's knowledge, which may limit the overall performance. Deep learning is becoming one of the most advanced pixel-level target detection methods in road condition inspection as it learns from large-scale data and requires little human involvement during training. Convolution Neural Network (CNN), a deep learning method, has been gradually utilized in road crack detection and segmentation (*10; 11*). For example, Bang, S et al. (*12*) used ResNet-152 to classify cracks and got the results as Precision of 77.68% and Recall of 71.98%. Cao Vu dung et al. (*13*) used Full Convolution Neural Network (FCN) to detect and segment cracks and got an average accuracy of 90%. Yahui Liu(*14*) proposed a DeepCrack dataset for crack segmentation, and six DeepCrack Structures which was mainly consisted of the extened FCN and the Deeply-Supervised Nets (DSN), and it showed a comparable result when compared with typical segmenting methods like AutoCrack and SegNet. Liu J W et al. (*15*) proposed a two-step pavement crack detection and segmentation method by firstly using a YOLO v3 to locate the crack area and then applying a modified U-Net model to segment the crack from this area. Chengjia Han(*16*) proposed a U-Net based CNN model, CrackW-Net, by adding a skip-level round-trip sampling block to segment the pavement images from the Crack500 dataset and a self-built dataset, and shows a good result.

Although deep learning is the most advanced pixel-level segmentation method, it requires a large amount and a wide diversity of annotated data to train the network(*17*). A small training dataset may cause the neural network overfitting and bad performance in robust. However, the cost of obtaining a large number of training samples is very high (*18*). To address this issue, some researchers developed deep learning architectures which can work with very few training images but still yields precise segmentations. The main idea of U-Net(*19*) is to replace pooling operators by upsampling operators to increase the output resolution. Also, it combines the high-resolution features with the upsampled output to learn more precise information based on small dataset. The attention module is popular in nature language process (NLP) and now it started to applied in the computer vision area(*20*). It can improve the sensitivity and efficiency of network to get rid of large amount data(*21*). For example, Wenjun Wang(*3*) proposed a pyramid attention network which uses pre-trained DenseNet121 and a feature pyramid attention module. It was tested on the Crack500 and MCD dataset and achieves a IoU of 0.6235. Xuezhi Xiang(*22*) proposed a pavement crack segmentation network based on BAM attention module and it got a mPA of 0.831, much higher compared to other networks like SegNet and CrackForest.

Another alternative method is data augmentation. The most common strategy for data augmentation is the traditional augmentation methods like image random crop, image flip and adding noise(*23*). In 2014, Goodfellow(*24*) proposed the concept of generative adversarial networks (GANs), which can produce real-like images through a battle between a generator and discriminator. Alec Radford(*25*) proposed deep convolutional generative adversarial networks (DCGANs) based on the conception of GAN, and it showed good representations of images. Because of using the convolution structure, DCGAN is popular in computer vision and has been applied in many areas including pavement crack data augmentation. For example, Lili Pei(*26*) used variational autoencoder (VAE) to encode crack images and the results from VAE was input to DCGAN model to generate the fake images. Boqiang Xu(*17*) recaches a quite high identification accuracy in pavement crack classification tasks based on DCGAN and VGG16. However, there are still some problems in DCGAN. For example, the original DCGAN structure is more suitable for small size images like an image with resolution of 32 * 32 pixels. In pavement crack segmentation tasks, the resolution of images is normally larger than 256 * 256 pixels. Another problem is that the discriminator in DCGAN studies too fast which would lead the loss of discriminator to 0 very rapidly while the generator had not studied very well.

This paper proposed a framework for pavement crack segmentation which contains an automated pavement crack generative adversarial network (APC-GAN) and a new pixel-level pavement crack segmentation network, AttuNet. The APC-GAN was modified from DCGAN and designed for improving the generated road image quality. The kernel size of APC-GAN was enlarged in generator for capturing more information. Moreover, the convolutional layers were increased in both generator and discriminator to produce sharper images. Gaussian noise was added at the top of the discriminator to slow down its convergence speed. The AttuNet was modified from U-Net. An attention module was introduced to this structure which can extract cracks' features by fusing different channel information from different layers. Batch normalization was used both in APC-GAN and AttuNet to accelerate training. The proposed AttuNet combines the advantages of U-Net and attention module. Another AttuNet-min was proposed through replacing the max pooling layer by min pooling layer to make the network focus more on the crack in the road image. In this paper, an open-source dataset is used to verify the segmentation accuracy of the proposed method. The experimental results show that the APC-GAN provides more distinct and diversity images than DCGAN and traditional augmentation method. The proposed framework were compared with four classic CNN models including U-Net, DeepLab-resnet50(*27*), FCN-resnet50(*28*), and LRASPP_mobilenet_v3-large(*29*), and it showed a higher accuracy in crack segmentation. The organization of this study is outlined as follows. In METHODOLOGY, the basic theory and structure of APC-GAN and AttuNet is presented. In RESULTS, the proposed framework is applied to crack dataset and compares the performance with various CNN models and it shows the pixel-level segmentation results of the cracks.

METHODOLOGY

APC-GAN

APC-GAN was designed for the pavement crack segmentation tasks according to the shortages of DCGAN. The structure of APC-GAN was shown in Figure 1. The input is a random vector with length 4096. In Generator part, a G-Block (4×4) is comprised of a convolution layer with a kernel size 4×4, and then followed a batch normalization with momentum 0.7, an activation function Rectified Linear Unit (ReLU) and an up-sampling layer. In Discriminator part, a D-Block (3×3) is comprised of a convolutional layer whose kernel size is 3×3, and then followed a batch normalization with momentum 0.7, an activation function leaky Rectified Linear Unit (leakyReLU), a Dropout layer with parameter 0.25 and an average pooling layer. The convolution layer can process the image to produce a set of feature maps while the activation functions used in this APC-GAN model include ReLU, leakyReLU and Sigmoid, which can make the network learn a non-linear task. The average pooling in Discriminator was utilized to translate invariance and reduce the parameter size of the networks.

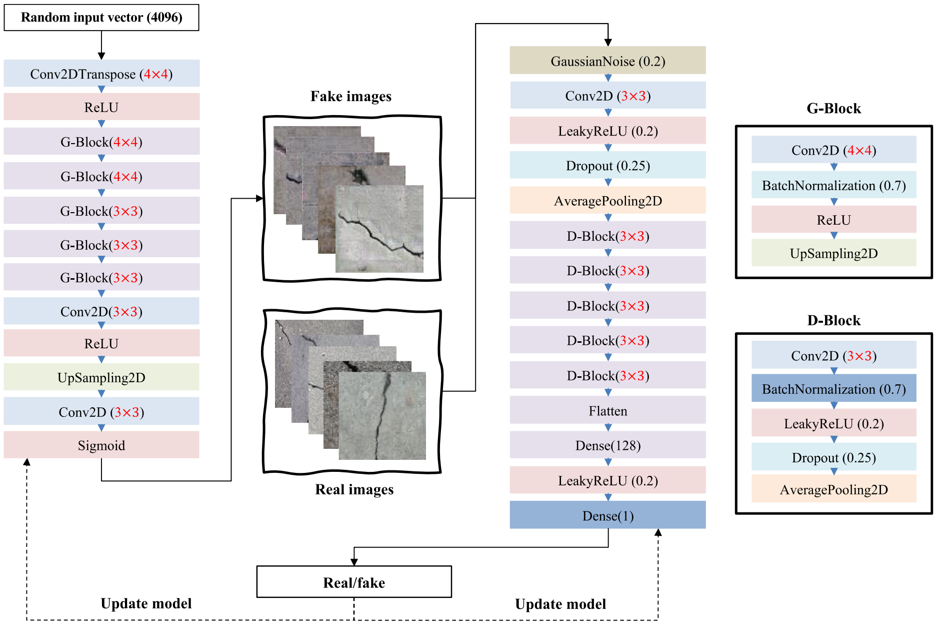

Figure 1 A structure diagram of the proposed APC-GAN model.

Normally, the DCGAN only works well at images with a low resolution like 32×32 pixels or 64×64 pixels. However, in a pavement crack segmentation task, a generated image with a resolution 256×256 pixels is required. Another problem in DCGAN is that the discriminator studies to fast leading the loss of discriminator to 0 very rapidly. It would lead to the situation that the loss cannot be used to update the generator although the generator did not learn well. In order to make the APC-GAN architecture for better results, some modifications were made. The contributions of this work can be summarized as follows:

1. Large kernel size was used. The kernel size was increased to 4×4 in generator and to 3×3 in discriminator. For generator, a large kernel at the top convolutional layers could cover more area and thus, could capture more information, which could maintain the smoothness of the image. For discriminator, a small kernel may cause the discriminator loss rapidly approach 0 while a larger kernel size can ease this situation.

2. The number of convolutional layers was increased in APC-GAN compared to the original DCGAN. A small number of convolution operators, especially in generator, would make the produced images very blurry while more layers can help capture additional information which can eventually add sharpness to the final produced images.

3. A batch normalization layer was followed by the convolutional layer. Batch normalization acts as a regularize which can reduce the accelerating training and improve the generated image quality. The function can be described in **Equation 1.**

$$y = \frac{x - E[x]}{\sqrt{Var[x] + \epsilon}} * \gamma + \beta \qquad (1)$$

Where γ and β are learnable parameter vectors ($\gamma = 1, \beta = 0$), ϵ is a value added to the denominator for numerical stability ($\epsilon = 1e - 5$). E[x] and Var[x] are the mean and variance of input x.

4. A Gaussian noise layer was added as the first layer of the discriminator. It can prevent the discriminator from studying too quick.

When applied the APC-GAN model in the data augmentation, the initial learning rate was set to 0.0001, and adjusted during training, where the learning rate would decrease by 15% for each 10000 steps. Binary Cross Entropy Loss (BCELoss) was used as the loss function and Adam was utilized as the optimizer to update the network. The batch size of the dataset was set to 16.

AttuNet

An AttuNet was proposed as a novel crack segmentation approach in this work. Figure 2 shows the main architecture of the proposed structure, in which the details of each operation are presented. This structure was modified from U-Net. The output is a N_c channel map of the probabilities where N_c is the number of classes (N_c=1 in this work). Four attention modules were introduced to connect different layers. Each attention module has two inputs, one from the current layer and the other from a previous layer. Then, the output from attention module would be concatenated with the current layer. The attention module filters the features from different layers rather than connect the features directly. Each convolutional operation followed a batch normalization to standardize the inputs and stabilize the learning procedure.

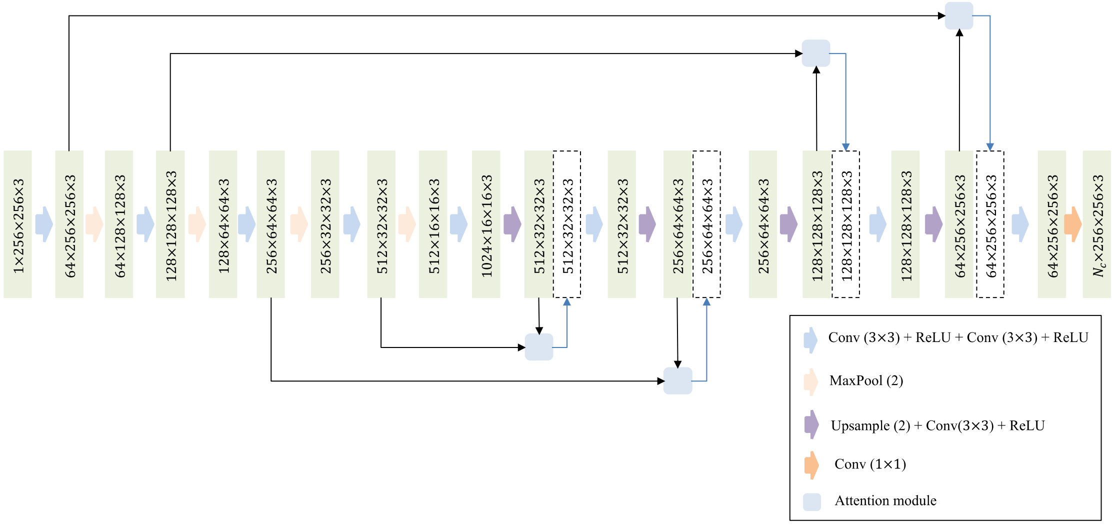

Figure 2 The structure of the proposed AttuNet.

There are some differences made in the AttuNet compared to U-Net model:

1. When applying image feature extraction, it will cause the lack of spatial local information and loss of pixel positioning, which will lead to the precision loss in the final crack segmentation. Therefore, an attention module was introduced to increase the accuracy of the model. The details of attention module were shown in Figure 3. The attention module got two inputs g and x from the former layers. Each input was processed through a convolution layer with a kernel size of 1×1 and a batch normalization. They are added together where the module can fuse the features under the two different scales. Then, the fused features were processed with ReLU activation, a convolution operation with a kernel size 1×1, a batch normalization and a sigmoid function. At last, the result from the attention module would concatenate with the input x. By doing this, the attention module can fuse the different features from different scale layers to improve the consistency of the feature map and thus improve the model performance, as well as decreasing the data usage. An attention module can refine the pavement crack to make it effectively guide the AttuNet training. Moreover, the convolution layer in the attention module can extracts cracks' features by fusing different channel information from different layers.

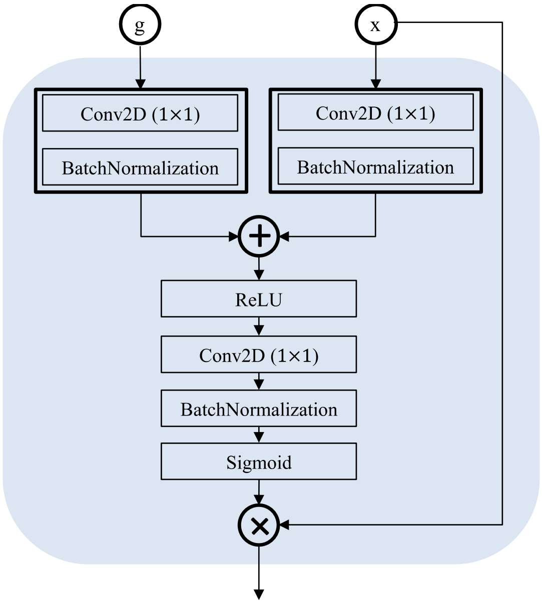

Figure 3 The structure of the attention module.

2. Each convolution layer was followed by a batch normalization layer, which can standardize the parameters and speed up the training procedure.

3. Root Mean Squared Propagation (RMSProp) with a momentum 0.9 was utilized as the optimizer to update the network. BCEWithLogitsLoss was used as the loss function, which combines a sigmoid layer and the Binary Cross Entropy in one single class. This loss is more numerically stable than using a plain Sigmoid followed by a BCELoss.

The proposed AttuNet structure combines the attention modules within the CNN network which can make the network works well with a small dataset. This is because the attention module can fuse the features from different layers which can make the model study faster.

For the crack segmentation task, another version of AttuNet, called AttuNet-min, were designed in this work. In this version, the max pooling layer was replaced by the min pooling layer. This is because that the crack pixels always have relatively small value in an image, using a min pooling layer can keep the crack information accurately when down size the images. At the same time, the ReLU was replaced by LogSigmoid function as shown in **Equation 2**.

$$LogSigmoid(x) = \log\left(\frac{1}{1+\exp\left(-x\right)}\right) \qquad (2)$$

When the input x gets smaller, the absolute value of output from LogSigmoid would be larger. Thus, the LogSigmoid function would give more weight on the small pixels while the ReLU function pays more attention on the brighter part.

Overall procedure

An overall procedure to evaluate and validate the usage of the proposed framework was shown in Figure 4. A DeepCrack dataset was utilized in this work to test the performance of the CNN networks. The DeepCrack dataset is an open-source dataset published in GitHub (https://github.com/yhlleo/DeepCrack). This dataset consists of 537 RGB crack images with manually annotated segmentations. The image has a resolution of 544 * 384 pixels. The images were divided into two subsets: 300 images for training and 237 images for testing. As we can see, the number of images is relatively low for a deep learning training.

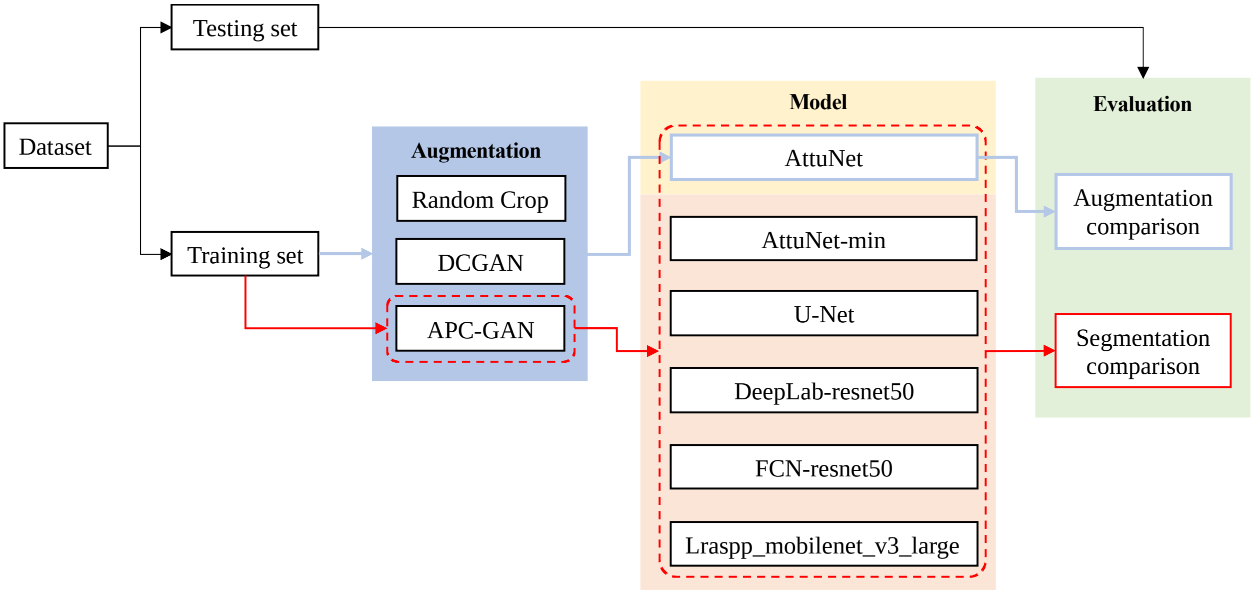

Figure 4 The overall procedure of experiments.

The blue line in Figure 4 shows the procedure for comparison and evaluation of the augmentation methods including Random Crop, DCGAN and APC-GAN. Each augmentation method in this work generated 300 images in total. The generated images would be annotated manually and then combined with the DeepCrack training dataset. Then, the proposed AttuNet would be trained on the augmented dataset. The data augmentation methods were compared and assessed by the results from the segmentation.

The red line in Figure 4 shows the process for comparison and evaluation of different segmentation models including AttuNet, AttuNet-min, U-Net, DeepLab-resnet50, FCN-resnet50, and LRASPP_mobilenet_v3-large. In order to evaluate the performance of the proposed AttuNet and AttuNet-min in the pavement cracks segmentation work. Other four segmentation methods including U-Net, DeepLab_resnet50, FCN-resnet50 and lraspp_mobilenet_v3_large were introduced and compared. We fine-tuned these four methods. The number of classes was set to 1. In the training procedure, the initial learning rate was set to 0.00001 and the batch size of the dataset was set to 16. The training epochs was set to 300. BCEWithLogitsLoss was used as the loss function and the RMSProp was utilized as the optimizer to update the network parameters. Before importing to the deep learning structure, all the images including the crack image and its label, would be reshaped to 256 * 256 pixels.

The data augmentation methods and CNN models were all implemented in Python and computed under the following machine speculations: Windows 10, Intel(R) Core (TM) i9-10900X CPU, NVIDIA RTX A4000 with 16 GB memory, 64GB RAM.

Evaluation metrics

Precision (P), Recall (R), F1 score (F1), Intersection over Union (IoU) and pixel accuracy (PA) were utilized to evaluate the semantic segmentation results.

(1) P can measure how accurate your predictions are. The precision can be calculated by Equation 3 where TP is true positive and FP is false positive.

$$P = \frac{TP}{TP+FP} \qquad (3)$$

(2) R suggests the level of sensitivity for prediction results. Recall can be calculated by Equation 4 where FN is false negative.

$$R = \frac{TP}{TP+FN} \qquad (4)$$

(3) F1 is defined based on the harmonic average of Precision and Recall. It can be calculated using Equation 5.

$$F_1 = \frac{2PR}{P+R} \qquad (5)$$

(4) IoU measures the overlap between 2 areas. It was used to measure how much the predicted areas overlaps with the ground truth. IoU was calculated according to Equations 6.

$$IoU = \sum_{i,j}^{k} \frac{p_{ii}}{p_{ij}+p_{ji}-p_{ii}} \qquad (6)$$

Where p_{ij} represents the number of pixels belonging to class i but predicted as class j.

The mean Intersection over Union (mIoU) was calculated according to Equation 7.

$$mIoU = \frac{1}{2}\sum_{i,j}^{k=2} \frac{p_{ii}}{p_{ij}+p_{ji}-p_{ii}} \qquad (7)$$

(5) PA is a semantic segmentation metric that denotes the percentage of pixels that are accurately classified in the image. It can be calculated by Equation 8.

$$PA = \sum_{i}^{k} \frac{p_{ii}}{t_i} \qquad (8)$$

Where t_i is the total number of pixels that is labeled as class i.

Since there are two classes present in this work: crack and background, the mean pixel accuracy (mPA) was calculated to represents the class average accuracy, shown in Equation 9.

$$mPA = \frac{1}{2} \sum_{i}^{k=2} \frac{p_{ii}}{t_i} \qquad (9)$$

RESULTS

A traditional image augment method, random crop, and a DCGAN were used in this work to compare with the proposed APC-GAN method. Some samples of the original image and the generated images from random crop, DCGAN, APC-GAN were shown in Figure 5.

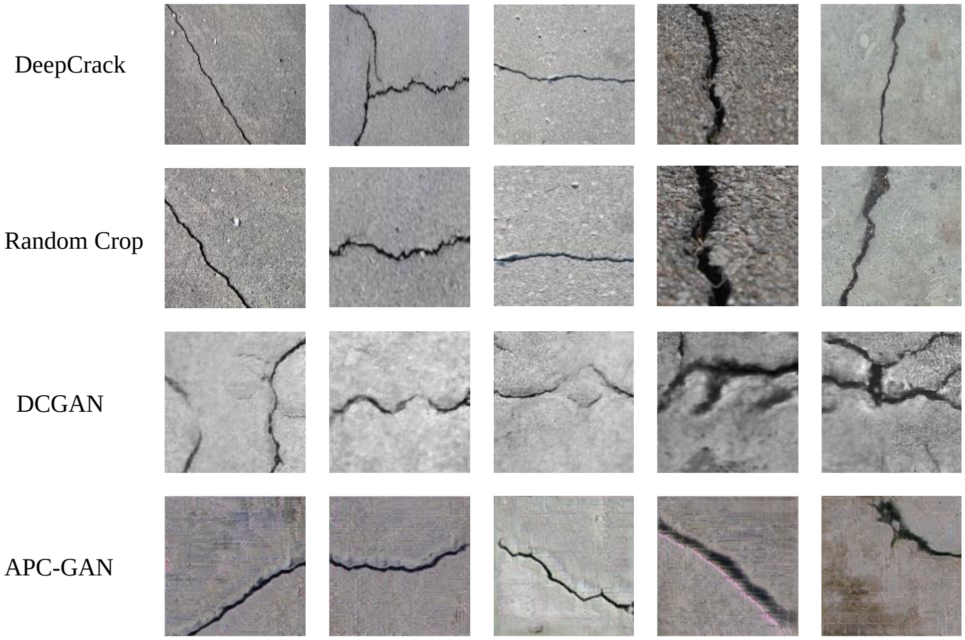

Figure 5 The raw images from DeepCrack and the generated images from random crop, DCGAN and APC-GAN.

As we can see from Figure 5., compared to the DCGAN, the images generated from APC-GAN are more distinct and sharper. The images produced from random crop are clear and distinct than images generated from GANs as the image was cropped from the original image directly. However, these images are not as much diversity as the images produced from DCGAN and APC-GAN.

The generated images were added to the original training data respectively. And the augmented training dataset was used to train the AttuNet model. The precision, recall, F1 score, mIoU, mPA were calculated and shown in Table 1.

TABLE 1 Comparison of different image augmentation methods using the same segmentation model.

Model	Data	Augmentation	P	R	F1	mIoU	mPA
AttuNet	DeepCrack	None	**0.950**	0.839	0.892	0.812	0.839
		APC-GAN	0.947	**0.868**	**0.906**	**0.836**	**0.868**
		DCGAN	0.949	0.851	0.897	0.822	0.851
		Random Crop	**0.950**	0.856	0.900	0.827	0.856

As we can see from TABLE 1, the dataset DeepCrack with a APC-GAN augmentation can make the segmentation neural network gain the highest recall, F1 score, mIoU and mPA value among all the augmentation methods used in this work. It means that by using the proposed APC-GAN, the accuracy of semantic segmentation method can be improved in a small dataset pavement crack detection task.

In order to evaluate that the proposed AttuNet and AttuNet-min have good performance in the small data pavement cracks segmentation work. Other four segmentation methods including U-Net, DeepLab_resnet50, FCN-resnet50 and lraspp_mobilenet_v3_large were introduced and compared. These methods were trained on the DeepCrack train dataset with a PC-GAN augmentation. The training procedure was shown in Figure 6.

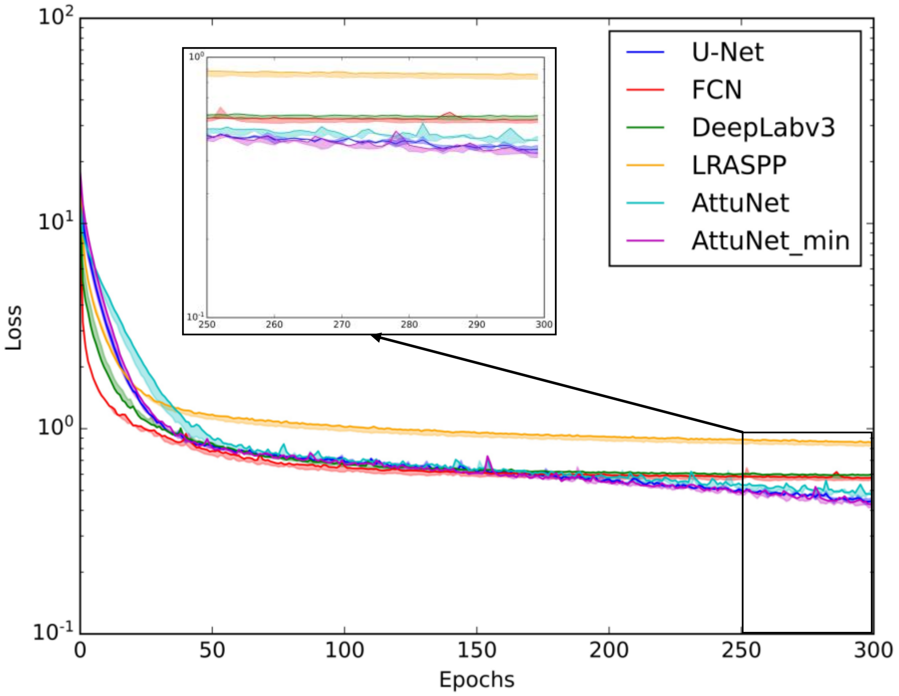

Figure 6 The Loss changes during the training procedure.

Figure 6 shows the training procedure where the loss of the model was decrease with the epochs. And as we can see, the LRASPP did not decrease rapidly around 50 epochs as other models all decreased to around 0.8 at 50 epochs. Compared to the FCN and DeepLabv3, the loss of U-Net, AttuNet and AttuNet-min continuously decreased after 200 epochs and finally reached at about 0.4. Because these models are all using the same loss function BCEWithLogitsLoss, the loss can actually reflect the performance of the models. It indicts that the U-Net, AttuNet and AttuNet-min learned more during the training than other three methods.

In order to evaluate and compare the performance between different deep learning segmentation models statistically, each model was trained and tested three times. The mean value and standard deviation were calculated. The evaluation metrics of each model were shown in Figure 7.

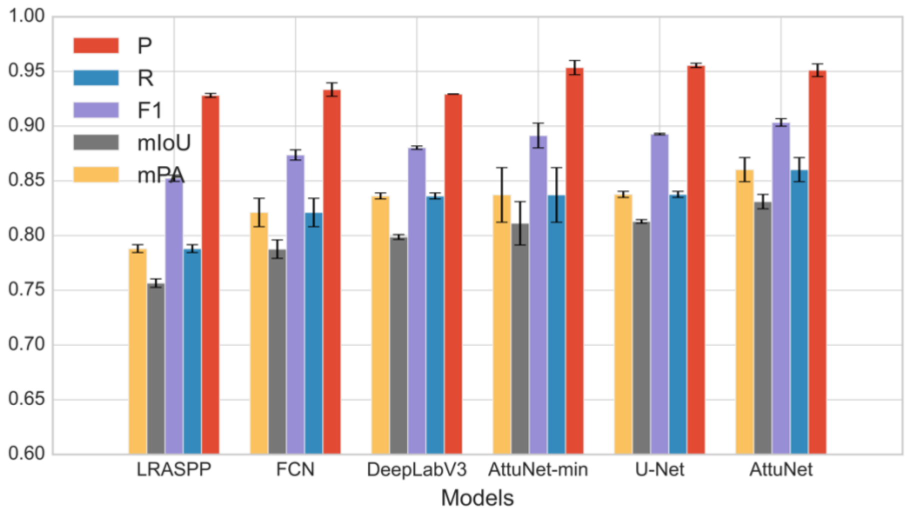

Figure 7 Comparison of different models on the test data.

The models in Figure 7 were ranked by the mIoU from lowest to highest. It shows that the AttuNet got the highest mIoU (0.831) among all the models. It also got the highest value in mPA, followed by AttuNet-min and U-Net, which means that the AttuNet has the highest percentage to predicts the pixels accurately as labeled in image. The mIoU and mPA are two most important indexes showing the segment ability of the method as both of them counted and compared each pixel. A higher value in mIoU and mPA means more pixels were classified accurately. The AttuNet-min got the highest precision (0.96) among all CNN models. It means that the 96% of the predicted cracks or background are originally labeled as cracks or background.

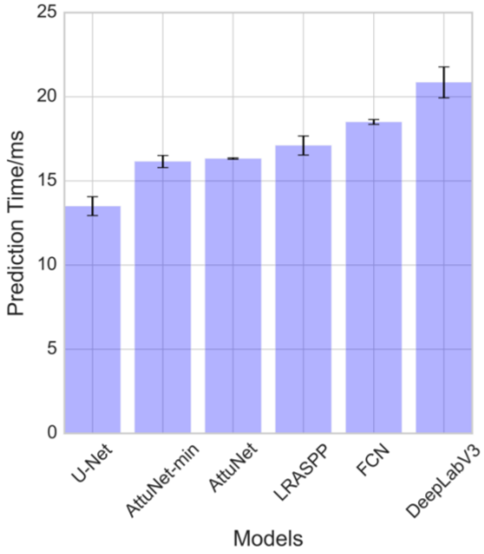

Figure 8 The comparison of prediction time of each model.

If we want to use the segmentation model into the real time crack detection work, the prediction time per image is an important factor. A faster prediction time of one model means this model is more suitable to the real time job. As shown in Figure 8, the models were ranked by the prediction time. As we can see, the U-Net model consumes least time while the DeepLabv3-resnet50 consumes the largest time.

The mean prediction time of AttuNet and AttuNet-min are 16.32 ms and 16.15 ms, respectively, which perform better than LRASPP, FCN and DeepLabV3.

Figure 9 shows some samples with cracks in various scenes and their segmentation results using different methods.

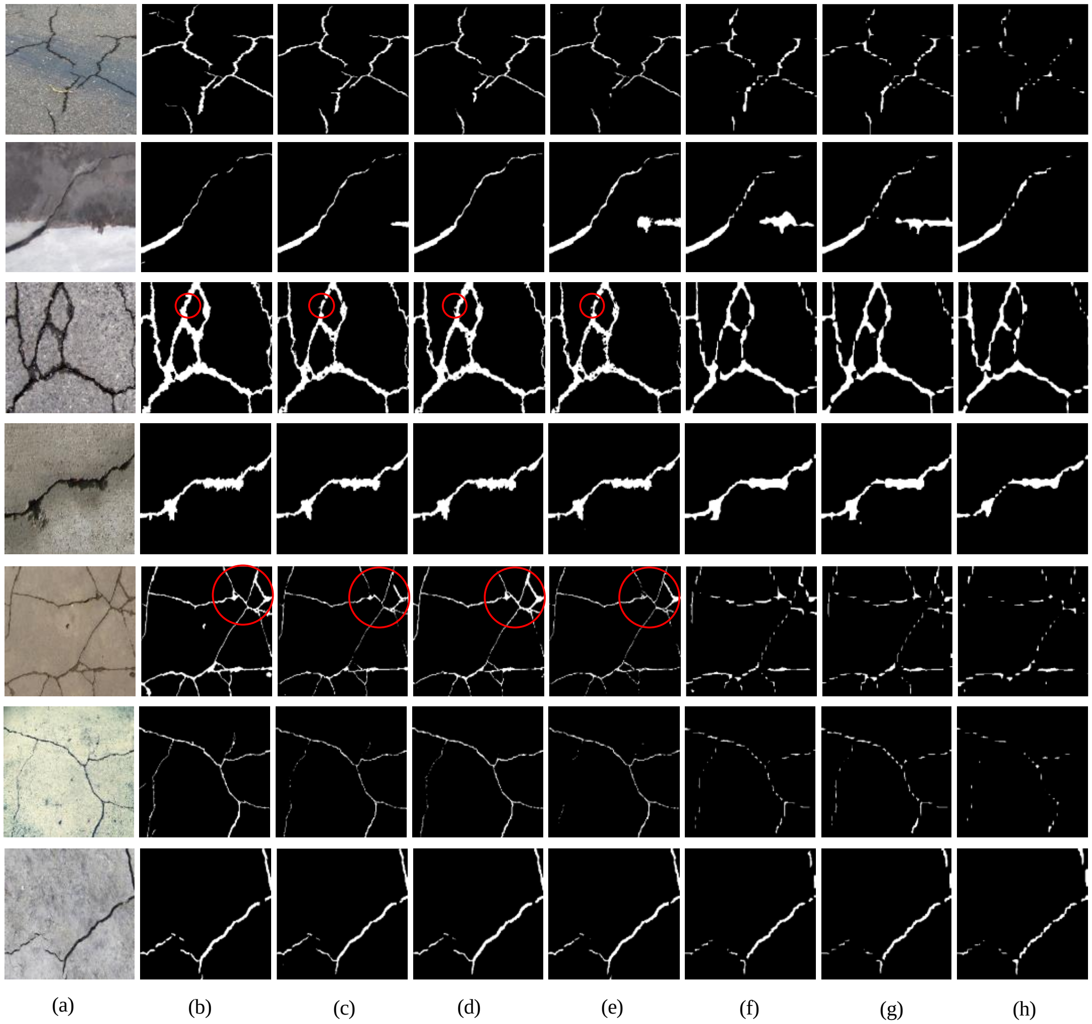

Figure 9 Several samples with cracks in various scenes and their segmentation results using different methods: (a) Original image (b) Ground Truth (c) AttuNet (d) AttuNet_min (e) U-Net (f) FCN_resnet50 (g) Deeplabv3_resnet50 (h) LRASPP_mobilenet_v3_large

As shown in Figure 9, it is obvious that the results from AttuNet, AttuNet_min and U-Net are more continuous and complete than the segmented results from FCN-resnet50, DeepLab-resnet50 and LRASPP_mobilenet_v3_large. The segmented part in red circle shows that the cracks were segmented more entirely by AttuNet_min than by AttuNet and U-Net. It shows that the AttuNet-min has a good performance in the continuous of the cracks as the segmentation image is much closer to the ground truth. This is mainly because in AttuNet_min, the max pooling layer was replaced by a min pooling layer and the activation function logsigmoid focused more on the small pixel value. Thus, the dark part in real image would be paid more attention in AttuNet-min which makes the segmentation results continuously.

CONCLUSIONS

This paper proposed a novel pixel-level crack segmentation stragety for a pavement crack inspection with small dataset. This situation is very common in road maintenance as the cost of obtaining a large number of pavement top-view images and labeling these manually is very high. However, a small training dataset may cause the neural network overfiting and bad model performance in robust. The proposed framework comsisted of APC-GAN and AttuNet, which can accurately segment images with small dataset.

In this paper, only a total of 300 color images used for training. The performance of APC-GAN was evaluated and it shows a better ability in producing sharper contrast and more diversity images compared to DCGAN and Random Crop. It improves the recall, F1 score, mIoU and mPA of the segmentation model. The proposed AttuNet model combines the attention module and batch normalization layer with the CNN network. It gets the testing mean IoU (0.831), which is higher than the classic CNN models including U-Net, DeepLabv3, FCN and LRASPP. Compared with AttuNet, the AttuNet-min achieved a more continuous segmentation result by applying the min pooling layer and LogSigmoid activation function.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: Tianjie Zhang, Yang Lu; data collection: Donglei Wang; analysis and interpretation of results: Tianjie Zhang; draft manuscript preparation: Tianjie Zhang. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Wang, W., M. Wang, H. Li, H. Zhao, K. Wang, C. He, J. Wang, S. Zheng, and J. Chen. Pavement crack image acquisition methods and crack extraction algorithms: A review. *Journal of Traffic and Transportation Engineering (English Edition),* Vol. 6, No. 6, 2019, pp. 535-556.
2. Yu, Y., M. Rashidi, B. Samali, A. M. Yousefi, and W. Wang. Multi-image-feature-based hierarchical concrete crack identification framework using optimized SVM multi-classifiers and D–S fusion algorithm for bridge structures. *Remote Sensing,* Vol. 13, No. 2, 2021, p. 240.
3. Wang, W., and C. Su. Convolutional neural network-based pavement crack segmentation using pyramid attention network. *Ieee Access,* Vol. 8, 2020, pp. 206548-206558.
4. Lin, W., Y. Sun, Q. Yang, and Y. Lin. Real-time comprehensive image processing system for detecting concrete bridges crack. *Computers and Concrete, An International Journal,* Vol. 23, No. 6, 2019, pp. 445-457.
5. Cao, W., Q. Liu, and Z. He. Review of pavement defect detection methods. *Ieee Access,* Vol. 8, 2020, pp. 14531-14544.
6. Chen, C., H. Seo, C. H. Jun, and Y. Zhao. Pavement crack detection and classification based on fusion feature of LBP and PCA with SVM. *International Journal of Pavement Engineering*, 2021, pp. 1-10.
7. Huyan, J., W. Li, S. Tighe, R. Deng, and S. Yan. Illumination compensation model with k-means algorithm for detection of pavement surface cracks with shadow. *Journal of Computing in Civil Engineering,* Vol. 34, No. 1, 2020, p. 04019049.
8. Ai, D., G. Jiang, L. S. Kei, and C. Li. Automatic pixel-level pavement crack detection using information of multi-scale neighborhoods. *Ieee Access,* Vol. 6, 2018, pp. 24452-24463.
9. Kaddah, W., M. Elbouz, Y. Ouerhani, V. Baltazart, M. Desthieux, and A. Alfalou. Optimized minimal path selection (OMPS) method for automatic and unsupervised crack segmentation within two-dimensional pavement images. *The Visual Computer,* Vol. 35, No. 9, 2019, pp. 1293-1309.
10. Chen, C., H. Seo, and Y. Zhao. A novel pavement transverse cracks detection model using WT-CNN and STFT-CNN for smartphone data analysis. *International Journal of Pavement Engineering*, 2021, pp. 1-13.
11. Lau, S. L., E. K. Chong, X. Yang, and X. Wang. Automated pavement crack segmentation using u-net-based convolutional neural network. *Ieee Access*, Vol. 8, 2020, pp. 114892-114899.
12. Bang, S., S. Park, H. Kim, and H. Kim. Encoder–decoder network for pixel-level road crack detection in black-box images. *Computer-Aided Civil and Infrastructure Engineering,* Vol. 34, No. 8, 2019, pp. 713-727.
13. Dung, C. V. Autonomous concrete crack detection using deep fully convolutional neural network. *Automation in Construction*, Vol. 99, 2019, pp. 52-58.
14. Liu, Y., J. Yao, X. Lu, R. Xie, and L. Li. DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing,* Vol. 338, 2019, pp. 139-153.
15. Liu, J., X. Yang, S. Lau, X. Wang, S. Luo, V. C. S. Lee, and L. Ding. Automated pavement crack detection and segmentation based on two-step convolutional neural network. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 35, No. 11, 2020, pp. 1291-1305.
16. Han, C., T. Ma, J. Huyan, X. Huang, and Y. Zhang. CrackW-Net: A novel pavement crack image segmentation convolutional neural network. *Ieee Transactions on Intelligent Transportation Systems*, 2021.
17. Xu, B., and C. Liu. Pavement crack detection algorithm based on generative adversarial network and convolutional neural network under small samples. *Measurement*, Vol. 196, 2022, p. 111219.
18. Zhang, Y., and K. V. Yuen. Crack detection using fusion features-based broad learning system and image processing. *Computer-Aided Civil and Infrastructure Engineering,* Vol. 36, No. 12, 2021, pp. 1568-1584.
19. Ronneberger, O., 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.

20. Wan, H., L. Gao, M. Su, Q. Sun, and L. Huang. Attention-based convolutional neural network for pavement crack detection. *Advances in Materials Science and Engineering,* Vol. 2021, 2021.
21. Oktay, O., J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, and B. Kainz. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.
22. Xiang, X., Y. Zhang, and A. El Saddik. Pavement crack detection network based on pyramid structure and attention mechanism. *IET Image Processing,* Vol. 14, No. 8, 2020, pp. 1580-1586.
23. Mazzini, D., P. Napoletano, F. Piccoli, and R. Schettini. A novel approach to data augmentation for pavement distress segmentation. *Computers in Industry,* Vol. 121, 2020, p. 103225.
24. Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. *Advances in neural information processing systems,* Vol. 27, 2014.
25. Radford, A., L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434,* 2015.
26. Pei, L., Z. Sun, L. Xiao, W. Li, J. Sun, and H. Zhang. Virtual generation of pavement crack images based on improved deep convolutional generative adversarial network. *Engineering Applications of Artificial Intelligence,* Vol. 104, 2021, p. 104376.
27. Chen, L.-C., G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587,* 2017.
28. Long, J., 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.
29. Howard, A., M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, and V. Vasudevan. Searching for mobilenetv3.In *Proceedings of the IEEE/CVF international conference on computer vision*, 2019. pp. 1314-1324.