Journal of Circuits, Systems, and Computers
Vol. 30, No. 15 (2021) 2150272 (16 pages)

DOI: 10.1142/S0218126621502728

World Scientific
www.worldscientific.com

Attention U-Net with Feature Fusion Module for Robust Defect Detection*

Yu-Jie Xiong[†,‡,§], Yong-Bin Gao[†], Hong Wu[†] and Yao Yao[†]

[†]*School of Electronic and Electrical Engineering,*
Shanghai University of Engineering Science,
Shanghai 201620, P. R. China

[‡]*Shanghai Key Laboratory of*
Multidimensional Information Processing,
East China Normal University,
Shanghai 200241, P. R. China
[§]*xiong@sues.edu.cn*

Received 18 November 2020
Accepted 17 April 2021
Published 25 May 2021

U-Net shows a remarkable performance and makes significant progress for segmentation task in medical images. Despite the outstanding achievements, the common case of defect detection in industrial scenes is still a challenging task, due to the noisy background, unpredictable environment, varying shapes and sizes of the defects. Traditional U-Net may not be suitable for low-quality images with low illumination and corruption, which are often presented in the practical collections in real-world scenes. In this paper, we propose an attention U-Net with feature fusion module for combining multi-scale features to detect the defects in noisy images automatically. Feature fusion module contains convolution kernels of different scales to capture shallow layer features and combine them with the high-dimensional features. Meanwhile, attention gates are used to enhance the robustness of skip connection between the feature maps. The proposed method is evaluated on two datasets. The best precision rate and MIoU of defect detection are 95.6% and 92.5%. The best F-score of concrete crack detection is 95.0%. Experimental results show that the proposed approach achieves promising results in both datasets. It demonstrates that our approach consistently outperforms other U-Net-based approaches for defect detection in low-quality images. Experimental results have shown the possibility of developing a mixture system that can be deployed in many applications, such as remote sensing image analysis, earthquake disaster situation assessment, and so on.

Keywords: U-Net; feature fusion module; attention gate; robust defect detection.

1. Introduction

In recent years, the metal industry rapidly rises around the world, and metal products are widely employed into all aspects of daily life and industrial manufacture.

*The paper was recommended by Regional Editor Tongquan Wei.
§Corresponding author.

The requirements of personal user and enterprise consumer on the appearance of metal products are increasingly stringent. If faulty products are not removed before delivery, such stuffs may cause immeasurable damage to lives and properties of the people. A conventional means of industrial inspection among assembly lines of varying products is manual testing. Manual testing often requires the tester to verify the products match with a expected example based on their individual's domain expertise, which is clearly a time-consuming activity. As the variety of the goods and the rising of the production, manual testing becomes an impossible task, every change made, big or small, could affect the entire appearance of the products. Encouraged by the strong demand of reducing the cost of social resources and improving the efficiency of industrial inspection, automatic defect detection becomes a new topic in both industrial and academic research. A large number of alternative approaches have been developed over the last decades to detect different kinds of surface defects.

This work is inspired by recent researches in the field of object detection. In order to combine more shallow layer features and focus on target structures without additional supervision, we proposed an attention U-Net with feature fusion module (FFM) for robust defect detection. The proposed approach adds FFM to the skip path for extracting shallow features, and suppress feature responses in irrelevant background regions using attention gates. In order to validate the effectiveness of our approach, exhaustive experiments on detecting the surface defects and concrete cracks are performed successfully in two datasets. The rest of this paper is organized as follows. Section 2 provides a brief overview of the related studies of defect detection. Section 3 presents a detailed description of the proposed method. The experimental results conducted on two datasets are reported in Sec. 4. The paper ends with a conclusion in Sec. 5.

2. Related Work

In the past, defect detection relied on manpower inspection, not only the results were unreliable, but also time consuming. Dangerous accidents may occur at any time, and it is urgent to build a system to diagnose hidden troubles effectively. To reduce unnecessary labor cost, researches on automatic defect detection are of great interests. Early researchers have focused on structural health monitoring. Chatzi *et al.*[1] developed a robust computational tool for the detection of cracks. A finite element-based method was proposed by Teidj *et al.*,[2] and it achieved a satisfactory performance for big changes of crack depth ratio and low noise. Yeum *et al.*[3] combined image processing with sliding windows to detect cracks of bolts. In order to remove the noise of images, fast gradient-teased algorithms[4] was utilized to enhance the detectability of the edges.

Deep learning techniques can learn required features automatically and are applied to solve challenging tasks. Chen *et al.*[5] proposed a framework for crack

detection based on convolutional neural networks (CNN). Li *et al.*[6] proposed an approach for simultaneous concrete defect detection and geolocalization in different places. Cha *et al.*[7,8] used CNN to extract features and detected defects of cracks and loosened bolts. In order to achieve pixel-level detection and improve the accuracy, Shelhamer *et al.*[9] proposed fully convolutional networks (FCN) for image semantic segmentation tasks. Yang *et al.*[10] implemented FCN to address the problem that conventional approaches were unable to identify and measure diverse cracks concurrently. U-Net[11] is a particular network architecture based on FCN. It has become the dominant method in the present research field of medical image processing, by achieving promising results with less training images. As a result, it motivated researchers to study the improvement of U-Net in different applications. Jin *et al.*[12] integrated the deformation convolution into the U-Net to capture various shape and scale information. Yuan *et al.*[13] added a partially dense-connection to aggregate informative feature maps which effectively promote the utilization of low-level features together with high-level features. Zhang *et al.*[14] fused the neighboring different scale feature maps by densely connection. Dolz *et al.*[15] extended the U-Net to utilize the multi-modal information and strengthen feature propagation by connecting each layer to following layers. Kolarik *et al.*[16] extended the U-Net to 3D Dense-U-Net for brain and spine segmentation. Qin *et al.*[17] proposed a two-level nested U-structure to capture textural information from both shallow and deep layers. A novel recurrent U-Net architecture[18] which encompasses several encoding and decoding layers was introduced in resource-constrained segmentation tasks.

To reduce the model size and improve the speed, Constantin *et al.*[19] proposed the U-Net based on residual cell unit to reduce the parameters. Lian *et al.*[20] proposed a residual enhanced U-Net, which used the residual convolutions to accelerate network convergence. Ding *et al.*[21] proposed a light weight U-Net with few parameters and designed a multi-scale network to learn the scale-relevant density maps. Song *et al.*[22] proposed the DU-Net, which simplifies the structure of the U-Net to avoid overfitting. In order to further emphasize the semantic information and boost the feature representation, Oktay *et al.*[23] utilized attention gates to focus on target structures of varying shapes and sizes. Ni *et al.*[24] designed the attention module to learn global context for semantic segmentation of cataract surgical instruments. Abraham *et al.*[25] proposed the attention U-Net with feature pyramid and deep supervised layers. Fu *et al.*[26] proposed a two-stage attention aware method for train bearing shed oil inspection using spatial pyramid pooling.

3. The Proposed Method

3.1. *The standard U-Net*

The standard U-Net consists of a contracting path and an expanding path. Two 3×3 convolution operations and one 2×2 max-pooling layer with step of 2 pixels for

down-sampling form the fundamental convolution unit for contracting path which uses rectified linear unit (ReLU) for activation. Expanding path consists of reduplicative fundamental deconvolution units. Each fundamental deconvolution unit contains two 3×3 convolution operations followed by a ReLU function and a 2×2 convolution (up-convolution) for up-sampling. To ensure that the output image has the same resolution as the input image, the size of the feature map is doubled and the number of feature channels is halved in each up-sampling. Moreover, the concatenation operations are performed between the contracting path and expanding path. The size of up-convolution operation output o_{up} can be expressed as

$$o_{\text{up}} = s \times (i-1) + k - 2 \times p. \tag{1}$$

Accordingly, the size of pooling operation output o_{po} can be expressed as

$$o_{\text{po}} = \text{ceil}\left[\left(\frac{i + 2 \times p - k}{s}\right)\right] + 1, \tag{2}$$

where ceil function returns the smallest integer greater than or equal to the given value, i is the size of the input image, p is the parameter of padding setting, k is the size of the convolution kernel, and s is the step size.

The distance between the truth value and the predicted value is calculated by the binary cross-entropy loss function. The loss function can be calculated as

$$\text{Loss} = -\frac{1}{n} \overset{n}{\underset{i}{X}} (y_i \log(\hat{y}_i) - (1 - y_i) \log(1 - \hat{y}_i)), \tag{3}$$

where y_i is the expected value and $\hat{y}_i$ is the predicted value. Adam optimizer is used to accelerate the convergence speed. The update parameter θ_t can be defined as

$$\theta_t = \theta_{t-1} - \alpha \times \hat{m}_t / (\hat{v}_t + \varepsilon), \tag{4}$$

where α is the initial learning rate, $\hat{m}_t$ is the average value of the gradient, $\hat{v}_t$ is the standard deviation of the gradient, and ε is an infinitesimal positive number to ensure that the denominator is non-zero.

3.2. *The attention U-Net with feature fusion module*

Our proposed attention U-Net with feature fusion module (AU-FFM) is shown in Fig. 1. In this architecture, FFM is added in the skip path to extract more shallow layer features. On the other hand, in order to suppress the irrelevant areas of background and concentrate on the foreground areas, we stack soft attention gates in skip path at each level.

3.2.1. *Feature fusion module*

There are unexploited shape and texture information in the shallow layers. In order to extract more shallow layer features and combine them into the deep layer features, AU-FFM adds feature fusion module into the skip path. FFM is a residual structure

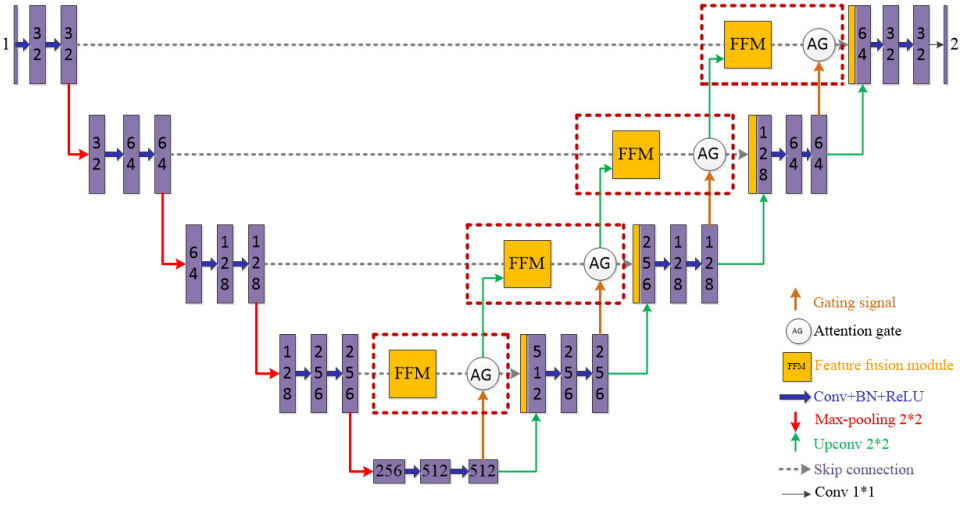

Fig. 1. The attention U-Net architecture with feature fusion module.

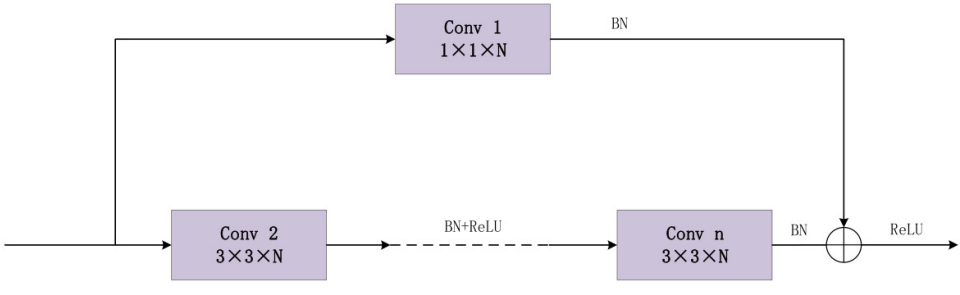

Fig. 2. An example of the feature fusion module.

that composed of a series of 3×3 convolution operations and a 1×1 convolution operation, as shown in Fig. 2. The process of the FFM is described as follows: the input of FFM is the combination of the basic feature maps of the fundamental convolution unit and the up-sampled output from the subsize feature maps with attention gate. Then, a series of convolution operations are conducted to interior feature fusion. While, the depth-wise separable (1×1) convolutions are used for cross-channel feature fusion. Besides, batch normalization is used to reduce parameters of networks. Therefore, these improvements of FFM can obtain better generalization ability and reduce the calculation cost.

The convolution operations and short links in the FFM extract more shallow layer features, and guarantee that extracted features contain not only semantic information, but also the effective textural and location information. Semantic information can be used to distinguish the defects from the background, and textural and location

information can accurately segment the defects from the background. Therefore, the defects can be located to the corresponding position of the original image accurately.

3.2.2. *Attention gate*

We add attention gates (AG) in skip path at each scale to concatenate the down-sampling features and up-sampling features. The structure of attention gate is illustrated in Fig. 3. Input feature map x^l is scaled with the attention coefficient α. The input feature map x^l and gating vector $g^{l'}$ are utilized together for computing α. The up-sampling operation of feature maps is achieved by bilinear interpolation. The value range of attention coefficient for each pixel i is $\alpha_i \in [0, 1]$. Attention coefficients are used to prune feature responses and identify salient image regions. Therefore, only the activations associated with the particular target are retained. In the default setting, a single scalar attention value is computed for each pixel vector x^l of layer l. In order to determine focus regions, gating vector g adds an additional branch. The features $g^{l'}$ are obtained from the decoder layer l' and they are coarse features. These coarse features are used for gating operation, to disambiguate the noisy information and uncorrelated effects. x^l and $g^{l'}$ provide local and contextual information, and are fused for computing attention coefficients.

First, the 1×1 convolution operation is performed on the gating vector $g^{l'}$, while the 2×2 convolution operation is performed on the input feature map x^l of layer l, parameterized by W_{x}^T and W_{g}^T, respectively. $g^{l'}$ has the same size as x^l. To obtain the intermediate feature maps, the $g^{l'}$ and the processed x^l are added in element-wise. the ReLU function is performed to transform the intermediate feature maps nonlinearly. Then, the 1×1 convolution is used to perform linear transformation, and the sigmoid activation function is utilized to normalize the attention coefficients. AGs calculate attention coefficients $\alpha_i^l \in [0, 1]$ for each pixel i of layer l, it is formulated as follows:

$$\alpha_i^l = \sigma_2(\Psi^T(\sigma_1(W_x^T x_i^l + W_g^T g_i^{l'} + b_{g^{l'}})) + b_\Psi), \tag{5}$$

where b denotes the bias and Ψ denotes the linear transformations. The attention coefficient α_l' is used to scale the input x^l of layer l, it can be expressed as

$$\hat{x}_i^l = \alpha_i^l \cdot x_i^l. \tag{6}$$

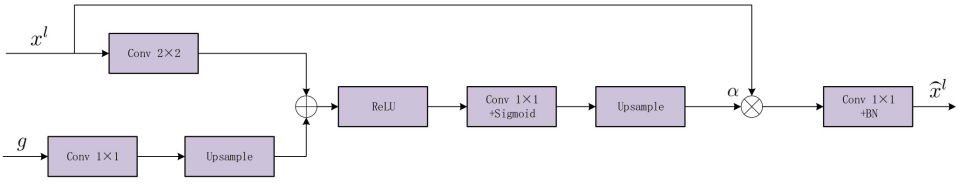

Fig. 3. The structure of attention gate in our proposed architecture.

Lastly, the output feature maps $\hat{x}_i^l$ are concatenated with corresponding up-sampling feature maps in the decoder. AGs parameters can be trained with the standard backward propagation. The update rule for AGs in layers can be expressed as

$$\frac{\partial(\hat{x}_i^l)}{\partial(\Phi^{l-1})} = \alpha_i \frac{\partial(f(x_i^{l-1}; \Phi^{l-1}))}{\partial(\Phi^{l-1})} + \frac{\partial(\alpha_i)}{\partial(\Phi^{l-1})} x_i^l, \tag{7}$$

where Φ denotes the convolutional parameters need to be updated, and the initialized gradient item is scaled with α_i.

The AGs are incorporated into the proposed architecture to highlight notable features that are passed through the skip path. The AGs extract coarse scale information and use them to disambiguate noisy and irrelevant responses. Stacking AGs in skip path at multiple scales aggregate more context information and make better segmentation performance.

4. Experimental Results

This section describes the experiments, including datasets, hardware configuration, optimization methods, parameters, etc. We employ two datasets to evaluate the proposed method. The proposed methods in this paper are all performed on a workstation with i7-8700k CPU @ 3.7 GHz, 32GB RAM and an NVIDIA GTX1080 GPU with 8 GB memory. The code is based on the PyTorch framework.

4.1. *Datasets*

We used concrete crack dataset[27] and our SUES-Washer dataset to evaluate the performance of our methods. Concrete crack dataset includes 57 images as train set and 24 images as test set. Totally, 84 images are taken at different places of the Huazhong University of Science and Technology Campus. Each image is manually labeled with a crack location, and the resolution of each image is $512 * 512$. SUES-Washer dataset contains 400 images for training the models and 100 images for testing the models. In order to improve the generalization capacity of the model, we adopted the mix-up method[28] to augment the dataset, and finally augmented the training dataset to 800 and 1,200 images. The examples of data augmentation are shown in Fig. 4.

4.2. *Optimization and parameters*

In order to speed up the training and adjust the learning rate, Adam algorithm is used to optimize the model. The learning rate of the Adam is set to 0.0001, the exponential decay rate of the first moment estimate is set to 0.9, and the exponential

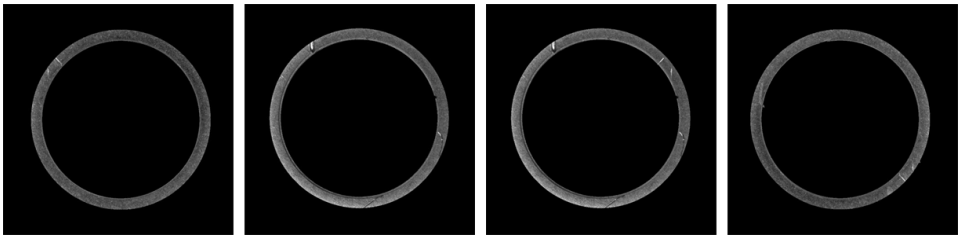

Fig. 4. Examples of data augmentation on SUES-Washer dataset. (a) and (b) are original images, while (c) and (d) are generated images.

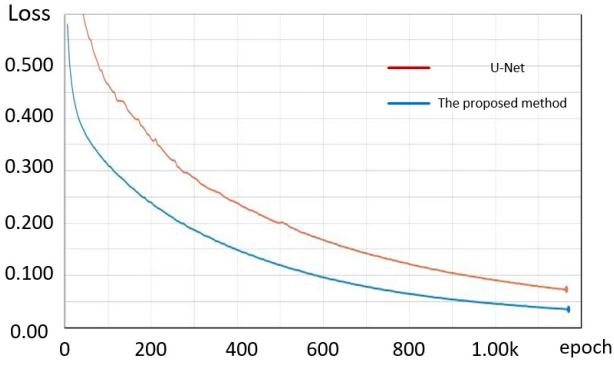

Fig. 5. The comparison of the loss changes between standard U-Net and the proposed method.

decay rate of the second moment is 0.999. The main parameter settings of the networks are shown in Table 1. The comparison of the loss change between standard U-Net and the proposed method is shown in Fig. 5. Compared with the standard U-Net, it is noted that our method has a faster convergence speed and lower loss in training phase.

4.3. *Evaluation metrics*

In this paper, the performance of the different methods is objectively evaluated by the mean intersection over union (MIoU), precision (P), recall (R) and F-score (F).

Table 1. The details of parameters settings of training.

Parameter	Value
Loss	Cross Entropy Loss
Epoch	1,500
Optimizer	Adam
Momentum	0.9
Learning Rate	0.0001

These evaluation metrics are widely used in the field of image segmentation. The MIoU, precision (P), recall (R) and F-score (F) are defined as follows:

$$\text{MIoU} = \frac{1}{k+1} \sum_{i=0}^{k} \frac{\text{TP}}{\text{FN} + \text{FP} + \text{TP}}, \tag{8}$$

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{9}$$

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{10}$$

$$F = \frac{2 \times P \times R}{P + R}, \tag{11}$$

where k is the number of classes, TP is the number of true positive samples, FP is the number of false negative samples and FN is the number of false negative samples.

4.4. *Experiments and analysis*

First, we compare the performance of our proposed method with other exsiting approaches. In order to validate the robustness of our proposed method, standard Gaussian noise and salt & pepper noise are added into test samples of SUES-Washer dataset, respectively. We test the performance of U-Net, DU-Net (DU),[22] U-Net + FFM (U-FFM), Attention U-Net (AU-Net) and attention U-Net with feature fusion module (AU-FFM) on SUES-Washer dataset with 400/800/1,200 training images.

In the experiments using 400 training images (as shown in Table 2), our best precision and mean intersection over union (MIoU) reaches 95.61% and 92.51%, respectively, which outperforms others. Compared with original U-Net, it can be seen that FFM is contributed to better performance (up to 3.42% and 2.51%).

Table 2. The performance comparison of different methods on SUES-Washer dataset with 400 training images.

Dataset	Method	Precision	MIoU
Original	U-Net	0.8958	0.8562
	DU-Net	0.9001	0.8672
	U-FFM	0.9300	0.8813
	AU-Net	0.9407	0.9111
	AU-FFM (ours)	0.9561	0.9251
With Gaussian noise	U-Net	0.8436	0.8096
	DU-Net	0.8556	0.8203
	U-FFM	0.9110	0.8897
	AU-Net	0.9289	0.9012
	AU-FFM (ours)	0.9314	0.9193
With salt & pepper noise	U-Net	0.7939	0.7787
	DU-Net	0.8052	0.7801
	U-FFM	0.9486	0.8982
	AU-Net	0.9292	0.8673
	AU-FFM (ours)	0.9518	0.9186

Table 3. The performance comparison of different methods on SUES-Washer dataset with 800 training images.

Dataset	Method	Precision	MIoU
Original	U-Net	0.9087	0.8614
	DU-Net	0.9127	0.8802
	U-FFM	0.9481	0.9154
	AU-Net	0.9457	0.9110
	AU-FFM (ours)	0.9623	0.9274
With Gaussian noise	U-Net	0.8834	0.8141
	DU-Net	0.8956	0.8232
	U-FFM	0.9409	0.9175
	AU-Net	0.9306	0.8933
	AU-FFM (ours)	0.9591	0.9194
With salt & pepper noise	U-Net	0.8129	0.7942
	DU-Net	0.8265	0.8052
	U-FFM	0.9563	0.8986
	AU-Net	0.9401	0.8990
	AU-FFM (ours)	0.9463	0.9122

Meanwhile, with the help of attention gate, the precision and MIoU are up to 4.49% and 5.47%. The results validate the discriminability of attention gate and feature fusion module.

In the experiments using 800 training images (as shown in Table 3), our best precision and MIoU reaches 96.23% and 92.74%, respectively. Compared with standard U-Net, AU-FFM improves the precision and MIoU up to 5.36% and 6.60%. Similar results are also obtained in the experiments using 1,200 training images (as shown in Table 4), our best precision and MIoU reaches 96.95% and

Table 4. The performance comparison of different methods on SUES-Washer dataset with 1,200 training images.

Dataset	Method	Precision	MIoU
Original	U-Net	0.9112	0.8670
	DU-Net	0.9223	0.8765
	U-FFM	0.9518	0.9148
	AU-Net	0.9609	0.9198
	AU-FFM (ours)	0.9695	0.9330
With Gaussian noise	U-Net	0.8979	0.8655
	DU-Net	0.9021	0.8802
	U-FFM	0.9518	0.9199
	AU-Net	0.9535	0.9149
	AU-FFM (ours)	0.9620	0.9246
With salt & pepper noise	U-Net	0.8598	0.8035
	DU-Net	0.8621	0.8202
	U-FFM	0.9575	0.8977
	AU-Net	0.9305	0.8871
	AU-FFM (ours)	0.9732	0.9019

93.30%, respectively. The results reveal the data augmentation by a mixture of standard samples and typical noises is contributed to stability of the all methods with hierarchical networks. Nevertheless, with the increasing amount of augment samples, the improvements of different networks are decreasing gradually.

As shown in Tables 2–4, no matter how many images are used for training, the results always represented that the U-Net is greatly affected by noise and has poor robustness. The loss of shallow layer features leads to the decrease of U-Net accuracy. DU-Net avoids over-fitting and improves the precision and MIoU, but it also loses the texture features and its also greatly affected by noise. FFM extracts more shallow layer features and combines them with the up-sampling feature map. So that FFM helps U-Net to locate the defect edge more accurately. AG is highly beneficial to defect localization and identification. Experimental results demonstrate that the proposed method (AU-FFM) has higher accuracy and MIoU, and significantly improves the detection performance in low-quality images.

The detection results of the U-Net, DU-Net and AU-FFM (ours) on the images with normal illumination from SUES-Washer dataset are shown in Fig. 6. With normal illumination, all three methods detect the defects successfully. But U-Net and DU-Net without enough local features cannot locate the defects accurately. FFM

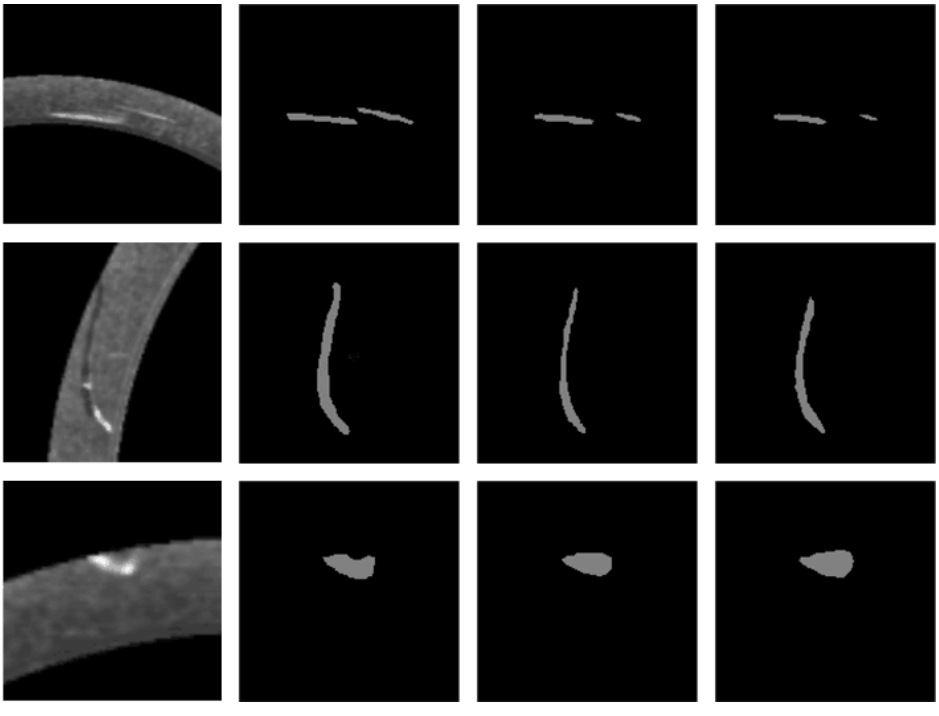

Fig. 6. Detection results on the images with normal illumination from SUES-Washer dataset: (a) Original images, (b) AU-FFM (ours), (c) DU-Net and (d) U-Net.

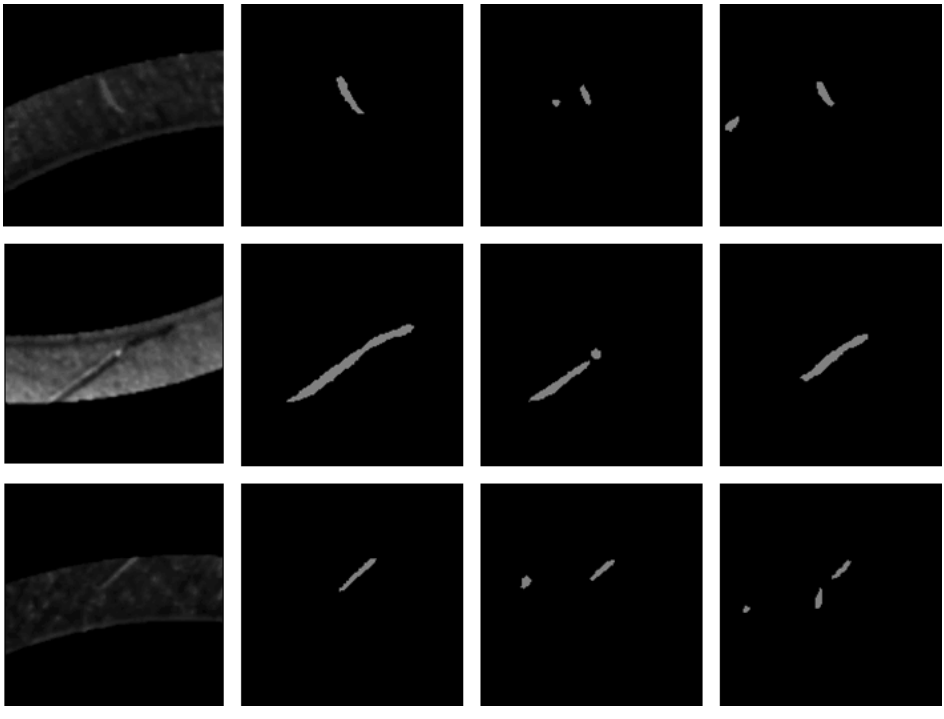

Fig. 7. Detection results on the images with low illumination from SUES-Washer dataset: (a) Original images, (b) AU-FFM (ours), (c) DU-Net and (d) U-Net.

extracts more shallow layer features and combines them with the up-sampling feature maps. Thus, our proposed method achieves a better results. Figure 7 shows the detection results of the U-Net, DU-Net and AU-FFM on the images with low illumination from SUES-Washer dataset. The results show that AU-FFM has the best performance with insufficient lighting conditions. It is also demonstrated that the proposed method (AU-FFM) has great robustness in low-quality conditions.

To further test the applicability of our proposed AU-FFM, we evaluate it on concrete crack dataset.[27] In particular, concrete crack detection is a difficult task due to shape-variability and background contrast. AU-FFM is compared with Liu's method[27] in terms of precision, recall and F-score. The results are shown in Table 5. The precision, recall and F-score of our methods are improved by 3%, 5% and 5%,

Table 5. Comparisons of U-Net and AU-FFM (ours) on concrete crack dataset.

Methods	Precision	Recall	F-score
Liu *et al.*[27]	90%	91%	90%
AU-FFM (ours)	93%	96%	95%

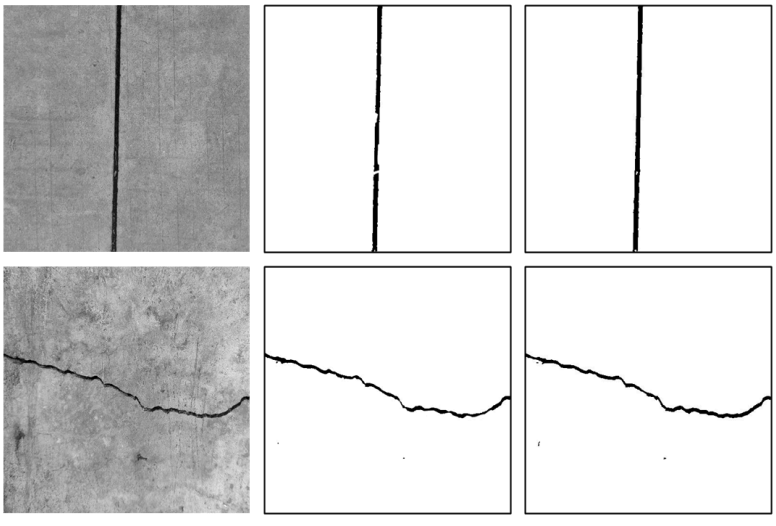

Fig. 8. Detection results of concrete cracks under simple background, (a) Original images (b) U-Net (c) AU-FFM.

respectively. The results of the two methods with different background are shown in Figs. 8 and 9. It can be found that U-Net is greatly affected by noises. But on the other hand, AU-FFM still shows good performance with noise and rough background for detecting concrete cracks.

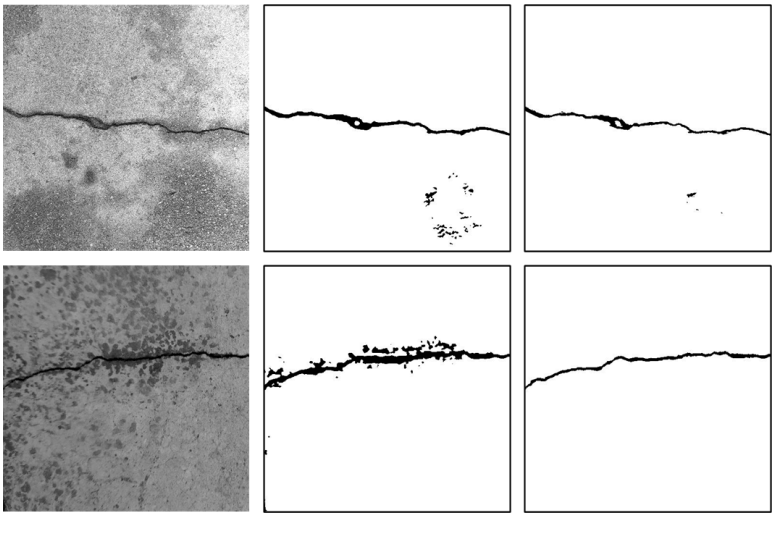

Fig. 9. Detection results of concrete cracks under complex background: (a) Original images, (b) U-Net and (c) AU-FFM.

5. Conclusion and Discussion

In this paper, we presented a novel attention U-Net with feature fusion module for robust defect detection. Our approach focuses on target structures without additional supervision and combines more shallow layer features. FFM is added in the skip path to extract more shallow layer features. In order to help our model to suppress the irrelevant areas of background and concentrate on the target areas, we stack soft attention gates in skip path at each scale. The proposed architecture is robust to the different variations in defect size, texture and shape, which represents its ability in detecting the noise images. Experimental results show that the proposed method is highly beneficial for defect task. For future work, we will attempt to achieve higher detection accuracy and reduce the detection time. The proposed methods will also be further evaluated in different applications such as sensing image analysis, earthquake disaster situation assessment, and so on.

Acknowledgments

This work is jointly sponsored by the National Key Research and Development Program of China (Grant No. 2019YFC1509202), National Natural Science Foundation of China (Grant No. 62006150), Shanghai Young Science and Technology Talents Sailing Program (Grant No. 19YF1418400), Shanghai Key Laboratory of Multidimensional Information Processing (Grant No. 2020MIP001), and Fundamental Research Funds for the Central Universities.

References

1. E. Chatzi, B. Hiriyur, H. Waisman and A. Smyth, Experimental application and enhancement of the XFEM-GA algorithm for the detection of flaws in structures, *Comput. & Struct.* **89** (2011) 556–570.
2. S. Teidj, A. Khamlichi and A. Driouach, Identification of beam cracks by solution of an inverse problem, *Procedia Technol.* **22** (2016) 86–93.
3. C. Yeum and S. Dyke, Vision-based automated crack detection for bridge inspection, *Comput.-Aided Civ. Inf. Eng.* **30** (2015) 759–770.
4. A. Beck and M. Teboulle, Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems, *IEEE Trans. Image Process.* **18** (2009) 2419–2434.
5. F. Chen and M. Jahanshahi, NB-CNN: Deep learning-based crack detection using convolutional neural network and Nave Bayes data fusion, *IEEE Trans. Ind. Electron.* **65** (2018) 4392–4400.
6. R. Li, Y. Yuan, W. Zhang and Y. Yuan, Unified vision-based methodology for simultaneous concrete defect detection and geolocalization, *Comput.-Aided Civ. Inf. Eng.* **33** (2018) 527–544.
7. Y. Cha, K. You and W. Choi, Vision-based detection of loosened bolts using the Hough transform and support vector machines, *Autom. Constr.* **71** (2016) 181–188.
8. Y. Cha, W. Choi and O. Buyukozturk, Deep learning-based crack damage detection using convolutional neural networks, *Comput.-Aided Civ. Inf. Eng.* **32** (2017) 361–378.

9. E. Shelhamer, J. Long and T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* **39** (2014) 640–651.
10. X. Yang, H. Li, Y. Yu, X. Luo, T. Huang and X. Yang, Automatic pixel-level crack detection and measurement using fully convolutional network, *Comput.-Aided Civ. Inf. Eng.* **33** (2018) 1090–1109.
11. O. Ronneberger, P. Fischer and T. Brox, U-net: Convolutional networks for biomedical image segmentation, *Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, 2015, pp. 234–241.
12. Q. Jin, Z. Meng and T. Pham, DUNet: A deformable network for retinal vessel segmentation, *Knowl.-Based Syst.* **178** (2019) 149–162.
13. M. Yuan, Z. Liu and F. Wang, Using the wide-range attention U-Net for road segmentation, *Remote Sens. Lett.* **10** (2019) 506–515.
14. J. Zhang, Y. Jin and J. Xu, MDU-Net: Multi-scale densely connected U-Net for biomedical image segmentation, preprint (2018), arXiv: 1812.00352v2.
15. J. Dolz, I. Ayed and C. Desrosiers, Dense multi-path U-Net for ischemic stroke lesion segmentation in multiple image modalities, *Proc. Int. MICCAI Brainlesion Workshop*, Granada, Spain, 2018, pp. 271–282.
16. M. Kolarik, R. Burget and V. Uher, Optimized high resolution 3D Dense-U-Net network for brain and spine segmentation, *Proc. Int. Conf. Telecommunications and Signal Processing*, Athens, Greece, 2019, pp. 237–240.
17. X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. Zaiane and M. Jagersand, U^2-Net: Going deeper with nested U-structure for salient object detection, *Pattern Recognit.* **106** (2020) 107–404.
18. W. Wang, K. Yu, J. Hugonot, P. Fua and M. Salzmann, Recurrent U-Net for resource-constrained segmentation, *Proc. IEEE Int. Conf. Computer Vision*, Seoul, Korea, 2019, pp. 2142–2151.
19. A. Constantin, J. Ding and Y. Lee, Accurate road detection from satellite images using modified U-net, *Proc. IEEE Asia Pacific Conf. Circuits and Systems*, Chengdu, China, 2018, pp. 423–426.
20. S. Lian, L. Li and G. Lian, A global and local enhanced residual U-Net for accurate retinal vessel segmentation, *IEEE Trans. Comput. Biol. Bioinform.* **14** (2015) 1–10.
21. L. Ding, K. Zhao and X. Zhang, A lightweight U-Net architecture multi-scale convolutional network for pediatric hand bone segmentation in X-Ray image, *IEEE Access* **7** (2019) 436–445.
22. L. Song, W. Lin and Y. Yang, Weak micro-scratch detection based on deep convolutional neural network, *IEEE Access* **7** (2019) 547–554.
23. O. Oktay, J. Schlemper and L. Folgoc, Attention U-Net: Learning where to look for the pancreas, *Proc. Conf. Medical Imaging with Deep Learning*, Amsterdam, Netherlands, 2018, pp. 1–10.
24. Z. Ni, G. Bian and X. Zhou, RAUNet: Residual attention U-Net for semantic segmentation of cataract surgical instruments, *Proc. Int. Conf. Neural Information Processing*, 2019, pp. 1–11.
25. N. Abraham and N. Khan, A noval focal tversky loss function with improved attention U-Net for lesion segmentation, *Proc. Int. Symp. Biomedical Imaging*, Venice, Italy, 2019, pp. 1–5.
26. X. Fu, K. Li and J. Liu, A two-stage attention aware method for train bearing shed oil inspection based on convolutional neural networks, *Neurocomputing* **380** (2020) 212–224.

27. Z. Liu, Y. Cao and Y. Wang, Computer vision-based on concrete crack detection using U-net fully convolutional networks, *Autom. Constr.* **104** (2019) 129–139.
28. H. Zhang, M. Cisse and Y. Dauphin, Mixup: Beyond empirical risk minimization, *Proc. Int. Conf. Learning Representations*, Vancouver, Canada, 2018, pp. 1–13.

1. Attention U-Net with feature fusion module for robust defect detection

Accession number: 20212310448168
Authors: Xiong, Yu-Jie (1, 2); Gao, Yong-Bin (1); Wu, Hong (1); Yao, Yao (1)
Author affiliation: (1) School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai; 201620, China; (2) Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai; 200241, China
Corresponding author: Xiong, Yu-Jie(xiong@sues.edu.cn)
Source title: Journal of Circuits, Systems and Computers
Abbreviated source title: J. Circuits Syst. Comput.
Issue date: 2021
Publication year: 2021
Article number: 2150272
Language: English
ISSN: 02181266
CODEN: JCSME7
Document type: Article in Press
Publisher: World Scientific
Abstract: U-Net shows a remarkable performance and makes significant progress for segmentation task in medical images. Despite the outstanding achievements, the common case of defect detection in industrial scenes is still a challenging task, due to the noisy background, unpredictable environment, varying shapes and sizes of the defects. Traditional U-Net may not be suitable for low-quality images with low illumination and corruption, which are often presented in the practical collections in real-world scenes. In this paper, we propose an attention U-Net with feature fusion module for combining multi-scale features to detect the defects in noisy images automatically. Feature fusion module contains convolution kernels of difierent scales to capture shallow layer features and combine them with the high-dimensional features. Meanwhile, attention gates are used to enhance the robustness of skip connection between the feature maps. The proposed method is evaluated on two datasets. The best precision rate and MIoU of defect detection are 95.6% and 92.5%. The best F-score of concrete crack detection is 95.0%. Experimental results show that the proposed approach achieves promising results in both datasets. It demonstrates that our approach consistently outperforms other U-Net-based approaches for defect detection in low-quality images. Experimental results have shown the possibility of developing a mixture system that can be deployed in many applications, such as remote sensing image analysis, earthquake disaster situation assessment, and so on. © World Scientific Publishing Company
Number of references: 28
Main heading: Feature extraction
Controlled terms: Crack detection - Image segmentation - Medical imaging - Remote sensing
Uncontrolled terms: Convolution kernel - Defect detection - Earthquake disaster - High dimensional feature - Low illuminations - Multi-scale features - Remote sensing images - Unpredictable environments
Classification code: 746 Imaging Techniques
Numerical data indexing: Percentage 9.25e+01%, Percentage 9.50e+01%, Percentage 9.56e+01%
DOI: 10.1142/S0218126621502728
Compendex references: YES
Database: Compendex

Data Provider: Engineering Village

Web of Science® ISI Web of Knowledge℠

经检索“**Web of Science**”，下述学术期刊属《**Science Citation Index Expanded (SCI-EXPANDED)**》收录来源刊。

Source title: JOURNAL OF CIRCUITS SYSTEMS AND COMPUTERS
ISSN / eISSN: 0218-1266 / 1793-6454
Publisher ： WORLD SCIENTIFIC PUBL CO PTE LTD , 5 TOH TUCK LINK, SINGAPORE, SINGAPORE, 596224

上海工程技术大学·图书馆·信息咨询部

检索人（签章）：林康治

检索日期：2021 年 9 月 6 日