Multi-scale Recurrent Residual U-Net for Anomaly Segmentation in Industrial Images

Haoyu Chen[1] and Shivani Sanjay Kolekar[1] and Kyungbaek Kim[1]

[1] Dept. of Artificial Intelligence Convergence, Chonnam National University, Gwangju, South Korea
leochy554@gmail.com
shivanikolekar@gmail.com
kyungbaekkim@jnu.ac.kr

Abstract. In recent years, there has been a rapid growth in automatic segmentation technology utilizing industrial imaging data. However, most manufacturing industries still use manual visual inspection of potential defects in final products, which requires a great number of manpower and is immensely time consuming task. Using automatic industrial image anomaly segmentation technology can greatly alleviate this problem, as it can reduce cost and time consumption along with improved quality control. Deep learning networks are widely used in industrial image data processing and interpretation due to their powerful feature extraction capabilities and efficient feature expression capabilities. To this end, this paper proposes a multi-scale recurrent residual U-Net model named MR2U-Net. The model introduces a multi-scale recurrent residual block to enhance the model's multi-scale industrial image anomaly segmentation ability. It uses a residual path between the downsampling and upsampling paths instead of ordinary skip connections, and narrows the semantics between feature maps for stitching difference. Compared with other popular segmentation networks, the well-trained MR2U-Net model has better stability for different types of component test results. For the evaluation, we use mean IOU coefficient and Dice coefficient of images where values are 0.5350 and 0.6490 respectively. The performance both have been significantly improved, providing a reliable solution for the automatic industrial image anomaly segmentation in large-scale industrial manufacturing.

Keywords: Industrial AI, Deep learning, Image Segmentation.

1 Introduction

In large-scale industrial production, segmentation of abnormal regions of industrial images is crucial to guarantee quality standards. The purpose of industrial image anomaly segmentation is to detect anomalous regions in individual images using various artificial intelligence-based methods. However, abnormal regions of parts in industrial production are often diverse, not only extremely complex, but also easily confused with normal regions, and it is very difficult for the model to extract effective

segmentation features, especially when abnormal samples are rare. Therefore, achieving accurate anomaly segmentation is a challenging task.

In recent years, deep learning has achieved great success in many different fields, especially in the field of computer vision (CV), such as face recognition, scene text detection, target tracking and automatic driving, etc., many abnormalities based on deep learning Region detection methods are also widely used in various industrial scenarios. For example, Zou et al. [1] proposed a new self-supervised learning scheme SPot-the-difference (SPD), which can regulate and compare self-supervised pre-training to be more suitable for anomaly detection and segmentation tasks. Nakazawa et al. [2] proposed a method to detect and segment abnormal wafer map defect patterns using a deep convolutional encoder-decoder neural network architecture.

Although great progress has been made in the research of industrial image anomaly region segmentation, there are still many problems. First of all, most of the data sets used by most researches now have fewer training samples, which are prone to overfitting. The generalization ability of the research results is poor, and the obtained models cannot assist employees in judging abnormal regions of industrial images. Secondly, the abnormal regions of industrial images are very complex and easily confused with the normal regions of the image, and it is very difficult for the model to extract effective segmentation features.

To address the above issues, this paper proposes a deep learning-based Multi-scale Recurrent Residual U-Net model (MR2U-Net) for industrial image anomaly segmentation. Firstly, a multi-scale recurrent residual block is introduced to enhance the network model's layered and multi-scale industrial image anomaly segmentation capabilities, and secondly, a residual convolutional layer is incorporated on the skip connection between the downsampling and upsampling paths to compensate for the loss from the downsampling. The difference between low-level features at early stages of the path and higher-level features from the upsampling path. Finally, good results were obtained in the industrial anomaly image test set.

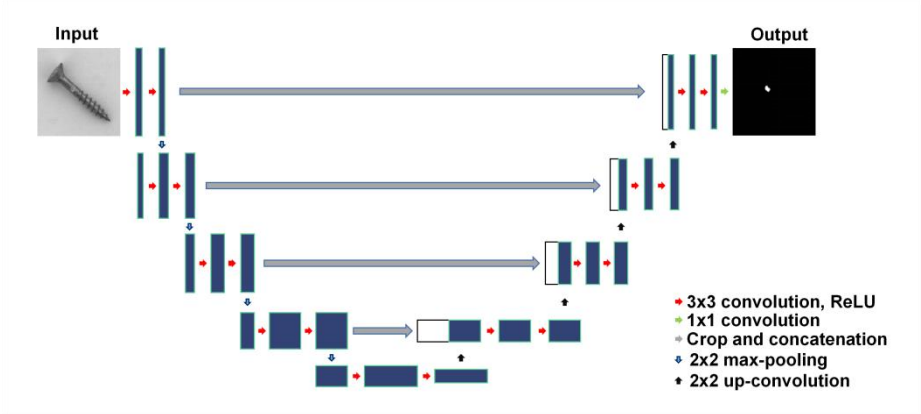

Fig. 1. U-Net model framework.

2 Related Work

U-Net Model. In 2014, Long et al. [3] used fully convolutional neural network (FCN) for end-to-end segmentation of natural images for the first time, achieving a breakthrough from traditional machine learning-based methods to deep learning methods. In 2015, Ronneberger et al. [4] proposed the U-Net network, and its structure is shown in Fig. 1. U-Net is an FCN-based network. Both encoders and decoders and skip connections are used to perform more accurate segmentation on a small number of training images. The difference between U-Net and FCN is that the U-Net network is left-right symmetric, the left side is the encoding path for capturing context information, and the right side is the decoding path for accurate positioning to restore the feature map size. After the output feature maps of each layer of the encoder are copied and cut, they are fused with the deconvoluted feature maps of the corresponding decoder, and then used as the input of the next layer to continue upsampling. The U-Net network has a large number of feature channels in the upsampling process, which enables it to pass context information to layers with higher resolution.

During the training process, since the images of abnormal areas of industrial parts are very complex, U-Net uses traditional convolution and pooling operations in the encoder to extract features. This feature extraction method can easily cause the model to fail to extract all useful feature information, and some features are lost. In addition, U-Net uses traditional convolution and deconvolution in the decoder to restore the feature map, which will cause a certain loss of feature information, making the network unable to fully restore the complex feature information of the image. Therefore, the U-Net network has become the research object of many researchers in the field of image segmentation. He et al. [5] made the network better preserve the feature maps in deeper neural networks by adding ResNet units in the U-Net network, and provided improved performance for deeper networks. Oktay et al. [6] suppress the feature responses irrelevant to the background region by adding an attention mechanism in the skip connections of U-Net, reducing the number of parameters and computational burden brought by the increase of network depth.

Recurrent Convolutional Neural Network. Recurrent Neural Network (RNN) is a special class of neural networks capable of processing data. The traditional feed-forward neural network only points to the output layer through the value of the activation function in the hidden layer, while RNN sends the result value of the activation function at the hidden layer node to the output layer and returns it to the next hidden layer node Computational inputs form a feedback loop that is the opposite of traditional feedforward networks. In the continuous development of deep learning, RNN has gradually been introduced into the convolutional neural network (CNN) to form a recurrent convolutional neural network (RCNN). RCNN uses a circular convolutional layer (RCL) instead of a convolutional layer. RCL does not output to the pooling layer after extracting the features of the input layer, but uses a changed cyclic neural network for processing, and uses the method of adding empty data to the feature layer data.

3 Methodology

In order to solve the general problem of the traditional segmentation model's ability to segment industrial abnormal images, this paper refers to the U-shaped structure in the U-Net model, improves the convolutional layer and skip connections in the U-Net model, and proposes a multi-scale recurrent residual U-Net model (MR2U-Net), whose structure is shown in Fig. 2. The MR2U-Net model refers to the concept of recurrent convolution layer[7], and uses a multi-scale recurrent residual convolution layer to replace the convolution layer in U-Net in the encoding and decoding process, which not only increases the depth of the model and effectively retains the features in the image through recurrent convolution, but also It can solve the problem of gradient disappearance caused by the deepening of network layers, and at the same time alleviate the over-fitting problem caused by the small amount of data, and efficiently extract important features of industrial abnormal images. Next, in order to alleviate the difference between the encoder-decoder features, this paper uses the residual path instead of the skip connection in U-Net, by taking the convolution operation before the corresponding feature connection in the encoder and decoder. Through the nonlinear operation of the 3×3 convolutional layer and the 1×1 residual structure, the decoding part can better restore the original image, thereby improving the segmentation accuracy. Finally, the softmax activation function is used to perform binary classification on the decoding results to realize the segmentation of abnormal regions and backgrounds.

Multi-scale Recurrent Residual Block. The Multi-scale Recurrent Residual block used in this paper is shown in Fig. 3, which replaces all the convolutional layers in the U-Net structure in order to learn multi-scale image features. Firstly, the 5×5 and 7×7 convolutional blocks are decomposed by a series of 3×3 convolutional blocks with smaller size, and then the output is obtained from each 3×3 convolutional block and concatenated to extract spatial features at different scales. The outputs of the second and third 3×3 convolutional blocks are effectively close to the outputs of the 5×5 and 7×7 convolutional blocks, respectively, thereby reducing the number of network parameters and speeding up the training speed of the network. At the same time, the recurrent convolution operation is performed on the series convolutional blocks, and when t = 3 time steps, a feedforward subnetwork with the maximum depth of 4 and the minimum depth of 1 is formed, including a convolution layer and a subsequence of three recurrent convolutional layers to deepen the number of layers of the network and enhance the feature extraction ability. In addition, a 1×1 convolution residual connection is added to the module, so that the network can obtain more spatial information.

Suppose that the input sample x_n in the multi-scale recurrent layer is located at the n'th layer, and for the pixel unit (i, j) located on the n'th feature map in the layer, its output M at the step size t is given by the following formula:

$$M_{ijk}^{n}(t) = (w_k^f)^T x_n^{f(i,j)}(t) + (w_k^r)^T x_n^{r(i,j)}(t-1) + b_k \quad (1)$$

Where $x_n^{f(i,j)}$ is the feedforward input, $x_n^{r(i,j)}$ is the recurrent input of the n'th layer, w_k^f is the feedforward weight, w_k^r is the recurrent weight, and b_k is the bias of the n'th feature map. The output M is activated by the RuLU function with the following formula:

$$f(M_{ijk}^n(t)) = Max(0, M_{ijk}^n(t)) \tag{2}$$

In MR2U-Net, the final output of the multi-scale recurrent layer is passed through the residual connection, assuming that the output of the multi-scale recurrent residual block is x_{n+1}, then it can be calculated as follows:

$$x_{n+1} = x_n + f(M_{ijk}^n(t)) \tag{3}$$

Where x_n and x_{n+1} correspond to the input versus output of the multi-scale recurrent residual block.

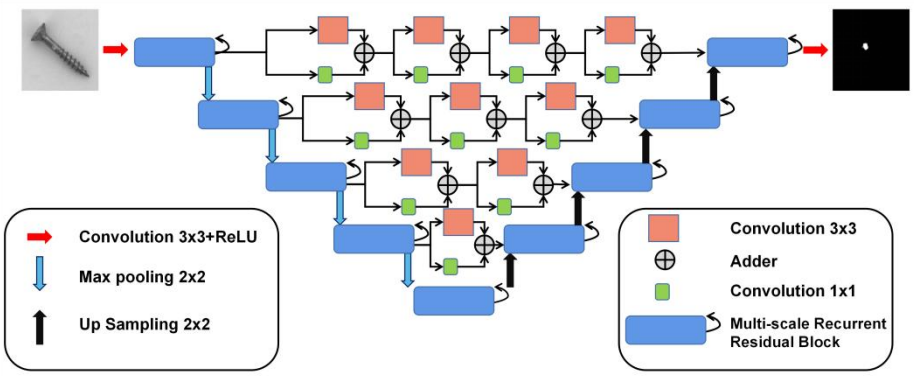

Fig. 2. MR2U-Netmodel framework.

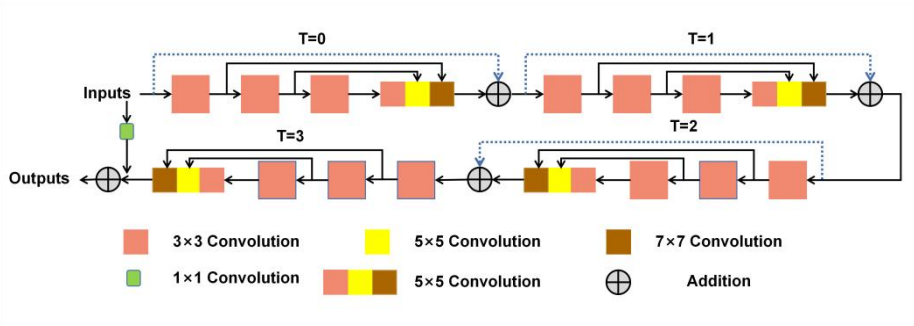

Fig. 3. Expanded multi-scale recurrent residual convolution structure for t = 3.

4 Experimental Results and Evaluation

4.1 Dataset

This experiment uses the MVTec AD dataset [8], which contains 5354 industrial images of different object and texture categories. For each category, it has normal (no defect) and abnormal images. We selected 1258 abnormal images of 15 categories for experiments, which contain more than 70 different types of defects. We randomly select 80% of the images as the training set and 20% of the images as the testing set. All training and test images (including ground truth) are resized to 128 × 128 to speed up model training.

4.2 Implementation

In this experiment, a computer with Intel Core i7-10700 2.9 GHz CPU (with 16GB memory) and NVIDIA GeForce RTX 2060 SUPER GPU (with 8GB video memory) is used for training and testing. The network model uses the Adam optimization algorithm, where the initial learning rate Set to 10-4, and the number of training epoch is set to 200.

4.3 Performance Indicators

This experiment uses Intersection-Over-Union (IOU) and Dice Coefficient (DC) to evaluate defect segmentation performance. The formulas of IOU and DC are as follows:

$$IOU = \frac{TP}{TP+FN+FP} \tag{4}$$

$$DC = \frac{2TP}{2TP+FN+FP} \tag{5}$$

where TP, FP and FN denote the number of pixels of true positive, false positive and false negative.

4.4 Experimental results and analysis

For the purpose of evaluation of MR2U-Net in the segmentation of industrial image abnormal regions, the MVTec AD dataset is compared with U-Net[4], Residual U-Net[5], Attention U-Net[6] Three models based on deep learning are used to segment and process abnormal regions of industrial images.All the methods above use the training set for training and the test set for evaluation. The results are shown in Table 1. The mean IOU and DC of the MR2U-Net model are as high as 0.5350 and 0.6490, respectively, and the results obtained are better than other schemes. Specifically, with regard to mean IOU and DC, our method achieves the best performance in 9 out of a total of 15 categories. For categories where our model did not achieve first place in segmentation performance, it still achieves comparable performance to the best

competing methods. Fig. 4 shows the segmentation results of MR2U-Net on the MVTec AD dataset. It can be seen that the segmentation results obtained by the MR2U-Net model are very close to the ground truth, which reflects that the MR2U-Net model has strong industrial image anomalies Segmentation ability.

Table 1. Performance comparison of MR2U-Net with other networks.

Category	U-Net[4]		Residual UNet[5]		Attention UNet[6]		MR2U-Net	
	IOU	DC	IOU	DC	IOU	DC	IOU	DC
Bottle	0.5976	0.7298	0.7052	0.8117	0.6476	0.7532	0.6547	0.7608
Cable	0.4676	0.5724	0.5283	0.6243	0.5400	0.6534	**0.5459**	**0.6624**
Capsule	0.1873	0.2853	0.1934	0.2711	0.2780	0.3865	**0.3579**	**0.4615**
Carpet	0.5289	0.6513	0.4928	0.6100	0.4731	0.5952	0.5063	0.6250
Grid	0.3399	0.4564	0.3944	0.5338	0.1390	0.2193	0.3283	0.4368
Hazel nut	0.7237	0.8274	0.7169	0.8273	0.7242	0.8210	0.7047	0.8132
leather	0.5824	0.7204	0.6440	0.7722	0.6410	0.7689	**0.6791**	**0.7974**
Metal nut	0.4614	0.5937	0.3582	0.4676	0.3408	0.4691	**0.5517**	**0.6814**
Pill	0.2498	0.3614	0.5699	0.6937	0.5147	0.6303	**0.6330**	**0.7420**
Screw	0.2176	0.2944	0.2178	0.2718	0.2336	0.3037	**0.2662**	**0.3341**
Tile	0.8128	0.8904	0.8080	0.8840	0.8121	0.8873	0.7787	0.8662
Toothbrush	0.3223	0.4266	0.3219	0.4489	0.3326	0.4480	**0.3340**	**0.4682**
Transistor	0.0768	0.1340	0.3800	0.5207	0.3576	0.4966	**0.4816**	**0.6221**
Wood	0.2219	0.3192	0.5221	0.6367	0.5658	0.6872	**0.5820**	**0.7062**
Zipper	0.5892	0.7309	0.6476	0.7781	0.5959	0.6510	0.6209	0.7584
Mean	0.4253	0.5330	0.5000	0.6101	0.4797	0.5847	**0.5350**	**0.6490**

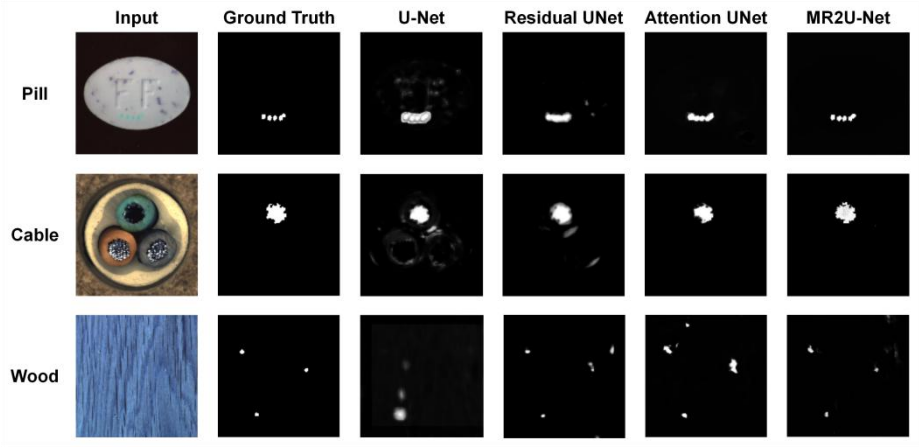

Fig. 4. Segmentation results on some test images from the MVTec AD dataset.

5 Conclusion

In this paper, Our develop and analyze a method for segmenting abnormal regions from industrial images. Based on the U-Net network structure, a multi-scale recurrent residual mechanism is introduced to enhance the feature extraction ability of the model for target regions. As the number of network layers increase, the overall robustness of the network is improved. Furthermore, our model outperforms current popular segmentation models on anomaly region segmentation in industrial images. we plan to improve our model to achieve optimized performance, in the future.

Acknowledgments

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Innovative Human Resource Development for Local Intellectualization support program(IITP-2022-RS-2022-00156287) supervised by the IITP(Institute for Information & communications Technology Planning Evaluation). This results was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE) (2022RIS-002)

References

1. Zou Y, Jeong J, Pemula L, et al. SPot-the-Difference Self-supervised Pre-training for Anomaly Detection and Segmentation[C]//European Conference on Computer Vision. Springer, Cham, 2022: 392-408..
2. Nakazawa T, Kulkarni D V. Anomaly detection and segmentation for wafer defect patterns using deep convolutional encoder–decoder neural network architectures in semiconductor manufacturing[J]. IEEE Transactions on Semiconductor Manufacturing, 2019, 32(2): 250-256.
3. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
4. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.
5. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
6. Oktay O, Schlemper J, Folgoc L L, et al. Attention u-net: Learning where to look for the pancreas[J]. arXiv preprint arXiv:1804.03999, 2018.
7. Zhou J, Hong X, Su F, et al. Recurrent convolutional neural network regression for continuous pain intensity estimation in video[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2016: 84-92.
8. Bergmann P, Batzner K, Fauser M, et al. The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection[J]. International Journal of Computer Vision, 2021, 129(4): 1038-1059.