Deep Data Augmentation for Defect Detection Enhancement: A Diffusion Model Based Approach

Yi Peng[a], Yaping Zhang[b,*]

School of Information Science and Technology, Yunnan Normal University, Kunming, Yunnan, 650500, China
[a]728008978@qq.com, [b]zhangyp@ynnu.edu.cn
*Corresponding author

Abstract: Weld defect detection is a crucial step in industrial production processes. To effectively identify these defects, the X-ray inspection method based on non-destructive testing is commonly employed. Addressing the challenges of limited sample size and class imbalance in X-ray images, this study proposes an enhanced diffusion model algorithm to augment samples, thereby improving the defect detection capability for rare categories. Experimental results prove the enhanced dataset's detection performance surpassing that of the original dataset. The improvement is notable, with a 5.1% enhancement on the WDD dataset. This paper presents a viable data augmentation solution for small-sample weld seam defect detection.

1. Introduction

Welding plays a crucial role in industrial production, with the manufacturing and processing of most metal products relying on welding technology. However, practical production often encounters various defects in welding structures, such as porosity, posing a significant threat to product quality. Therefore, quality inspection of welding components becomes urgent and necessary. Among the various methods, X-ray inspection [1] stands out as one of the most widely adopted techniques in the industry. Nevertheless, traditional X-ray welding seam defect detection methods heavily rely on manual film inspection, which is inefficient and has a high leakage rate, making it unable to meet the requirements of modern production.

Currently, artificial intelligence technologies, represented by deep learning, have found widespread applications in various fields such as natural language processing, image recognition, speech recognition, autonomous driving, and intelligent manufacturing. In the field of defect detection, defect detection based on deep learning has garnered attention due to its strong adaptability, proficiency in extracting complex features, and high accuracy. Although deep learning can improve the accuracy of weld seam defect detection, it requires a substantial amount of training and test data.

However, in the field of weld seam defect detection, obtaining a sufficient quantity of industrial sample data is quite challenging, making the exploration of effective data augmentation methods crucial. Up to now, the most common and proven effective data augmentation methods primarily

involve traditional geometric and optical transformations, such as translation, rotation, flipping, scaling, distortion, brightness, and color adjustments. Despite the adoption of these traditional data augmentation methods, overfitting issues persist, as these minor modifications to images hardly add additional content information, thus failing to address the aforementioned problems effectively.

Another data augmentation method involves synthesizing images using generative models. Typical generative models include Generative Adversarial Networks (GAN) [2], Variational Autoencoders (VAE) [3], and Diffusion Models [4].

The Diffusion Model stands out as the latest prominent representative in the field of generative models. Diffusion models and their variants have achieved significant success in image generation, with widespread applications in medical image synthesis. However, their full potential has yet to be realized in the field of vision-based non-destructive detection.

This study aims to explore the applicability of the Diffusion Model in augmenting weld seam data by enhancing the characteristics of X-ray welding seam images and combining other methods. The goal is to augment the dataset of welding seam defect images, thereby improving defect detection accuracy.

2. Related work

In order to construct applicable deep learning models, avoiding overfitting is a focal point of research. Data augmentation is a powerful method for achieving this goal, artificially increasing the size of the training dataset by distorting or oversampling the data. Augmented data represents a more comprehensive set of potential data points, thereby minimizing the distance between the training, validation, and future test sets. Data distortion augmentation involves transforming existing images while preserving their labels. This includes geometric and color transformations, random erasure [5], adversarial training [6], and neural style transfer [7], among other enhancement methods. Oversampling augmentation involves creating synthetic instances and adding them to the training set, incorporating methods such as image blending [8], feature space augmentation [9], and generative models.

Geometric and color transformations based on image data were among the earliest methods used for data augmentation. Although these methods are simple and common, excessive usage can lead to the generation of overly homogeneous data samples, including many samples with little practical application value. Noise injection-based data augmentation involves randomly adding noise, such as Gaussian noise, to the original images. Moreno-Barea et al. [10] tested noise injection on nine datasets from the UCI repository, and the results indicated that adding noise to images can enhance the robustness of Convolutional Neural Networks (CNN).

Generative modeling is an effective data augmentation method that utilizes generative models to synthesize new artificial samples while preserving features similar to those in the original dataset. Bowles et al. [11] describe generative modeling as a way to "unlock" additional information from the dataset. Frid-Adar et al. [12] tested the effectiveness of DCGANs in generating medical images of liver lesions. They observed that, in addition to classical data augmentation, the inclusion of samples generated by DCGAN could further enhance detection accuracy. Jiangsha et al. [13], with a limited set of extracted weld seam defect images and manually drawn labels, utilized GAN to generate weld defect images, which achieved a 10% performance improvement on the GDXray [14] dataset higher than traditional data augmentation methods alone.

However, GANs suffer from training instability and demand for extensive data, making them potentially suboptimal for constrained datasets. The Diffusion Model is a novel generative model that surpasses GANs regarding result quality and training stability, providing a new approach to enhancing image datasets to drive machine vision systems. The diffusion model consists of two

components: a forward diffusion process and a reverse inverse diffusion process. During the diffusion process, Gaussian noise is gradually added to the image, transforming it into random noise. In the reverse diffusion process, this procedure is reconstructed to obtain a new image.

Introducing the diffusion model into weld defect image data augmentation represents a novel direction. This paper, proposes a diffusion model that improves the noise estimation network to expand the weld defect image dataset. IDDPM [15] is a relatively classic diffusion model. Compared to the original diffusion model, it mainly improves in optimizing the Noise Schedule, using importance sampling, and improving sampling speed. In this paper, we further improve the IDDPM model according to the characteristics of weld seam images.

3. Materials & Methods

3.1 Improved Noise Estimation Network

Currently, in the image generation task based on the diffusion model, the noise estimation network (NE-Net) draws inspiration from the core structure of Convolutional Neural Networks (CNN) and adopts a U-Net architecture [16]. The U-Net network is an encoder-decoder structure where the encoder captures low and high-level features, and the decoder combines these features to generate the final image result. Skip connections play a crucial role in this architecture, aiding in the transmission of spatial information lost during pooling operations, thereby restoring complete spatial resolution in the encoding-decoding process. However, for U-Net, using a simple skip connection scheme to model global multiscale contexts without considering semantic differences, may not always achieve the desired results. Therefore, the Channel-wise Cross Fusion Transformer (CCT) module [17] is proposed to improve U-Net, which guided the fused multiscale channel-wise information to effectively connect to the decoder features, replacing simple skip connections.
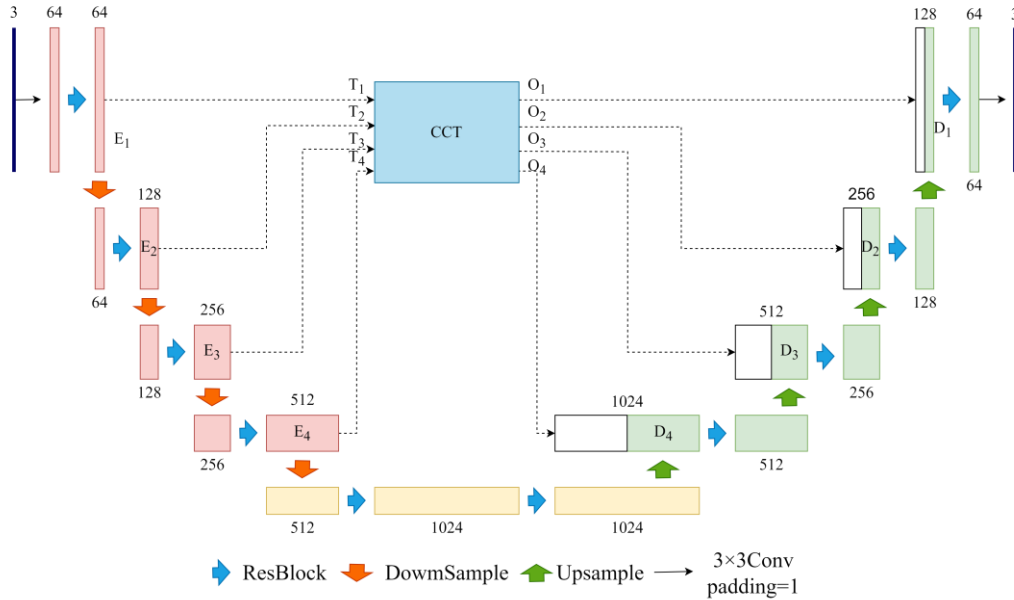

Figure 1: Improved noise estimation network

This paper presents an improved version of the noise estimation network, called Channel-wise Cross Fusion Transformer for Noise Estimation (NE-CCT), to enhance the model's performance in image generation tasks, as illustrated in Figure 1. By embedding the CCT module in skip connections, the model's ability to express image features is enhanced. To further increase the receptive field without increasing parameters, this paper adopts dilated convolutions with varying

dilation rates in the encoder instead of traditional convolution operations [18]. Therefore, this paper introduces a Coordinate Attention (CA) module [19] in the bottom two layers of the NE-CCT network. These improvement measures aim to optimize the performance of the noise estimation network, making it better suited for image generation tasks, enhancing its ability to handle multi-scale information, reducing computational burden, and improving the capture of image details.

3.2 Weld Defect Detection Dataset WDD

The Weld Defect Dataset (WDD) is constructed from 994 X-ray images of defective weld seams collected from a collaborative company. These images have a resolution of 400×600 and encompass four categories of defects: porosity, incomplete penetration, cracks, and undercut. We partitioned the dataset, with 723 images utilized for the training set, 121 images for the validation set, and 150 images for the test set.

4. Results

This section will evaluate the proposed improved diffusion model in terms of both the quality of generated images and its effectiveness in data augmentation for defect detection tasks.

During the data augmentation experiment, only the training set was augmented, while the test set remained unchanged. To validate the effectiveness of the synthetic defect images generated by the proposed data augmentation method, it was necessary to select a mature and stable detection network. In this study, we opted for YOLOv5s as the defect detection network, a widely used and proven-effective model.

For the specific application of weld seam defect detection and recognition, we employed several traditional augmentation methods, including random rotation (-10° to 10°), random Horizontal flipping (with a probability of 0.5), and Gaussian blur. Furthermore, we compared the combination of our method with traditional methods. The experimental results are presented in Table 1.

The results indicate that applying our method alone achieves the best recall. Combining our method and random rotation is the most effective augmentation method overall. Compared to the baseline dataset, it resulted in a 7.9% accuracy improvement, a 5.5% recall improvement, and a 5.1% mAP improvement. It is worth noting that the combination of our proposed method with traditional methods shows a significant improvement in the comprehensive metric mAP compared to using a single traditional method.

Table 1: Performance comparison of defect detection of YOLOv5s on the WDD dataset augmented by different methods

	Precision	Recall	mAP
Baseline dataset	0.616	0.554	0.581
IDDPM	0.452	0.667	0.604
ours	0.556	**0.678**	0.613
Horizontal flip	0.575	0.64	0.568
Random Rotation	**0.720**	0.525	0.597
Gaussian noise	0.553	0.541	0.579
Ours + Horizontal flip	0.593	0.631	0.603
Ours + Random Rotation	0.695	0.609	**0.632**
Ours + Gaussian noise	0.641	0.542	0.582

5. Conclusions

We found that by improving the skip-layer connection part of the noise prediction network in the diffusion model, we can effectively improve the quality of samples generated by the diffusion model, which is proven in the experimental results. In addition, we introduced the diffusion model into the small-sample image data augmentation task, and using the images generated by the diffusion model for data augmentation effectively alleviated the overfitting problem in the detector. In addition, the proposed method is compatible with traditional data augmentation techniques, and their combined use significantly improves the overall detection performance of the model. Our method still has shortcomings, such as a large number of parameters and long computation time. Improving the computational efficiency of the model is an important area for future improvement.

References

[1] Nacereddine N, Zelmat M, Belaifa S S, et al. Weld defect detection in industrial radiography based digital image processing [J]. Transactions on Engineering Computing and Technology, 2005, 2: 145-148.

[2] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[3] Kingma D P, Welling M. Auto-encoding variational bayes [J]. arXiv preprint arXiv:1312.6114, 2013.

[4] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models [J]. Advances in neural information processing systems, 2020, 33: 6840-6851.

[5] Zhong Z, Zheng L, Kang G, et al. Random erasing data augmentation[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 13001-13008.

[6] Moosavi-Dezfooli S M, Fawzi A, Frossard P. Deepfool: a simple and accurate method to fool deep neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2574-2582.

[7] Gatys L A, Ecker A S, Bethge M. A neural algorithm of artistic style [J]. arXiv preprint arXiv:1508.06576, 2015.

[8] Inoue H. Data augmentation by pairing samples for images classification [J]. arXiv preprint arXiv:1801.02929, 2018.

[9] Konno T, Iwazume M. Icing on the cake: An easy and quick post-learnig method you can try after deep learning[J]. arXiv preprint arXiv:1807.06540, 2018.

[10] Moreno-Barea F J, Strazzera F, Jerez J M, et al. Forward noise adjustment scheme for data augmentation[C]// 2018 IEEE symposium series on computational intelligence (SSCI). IEEE, 2018: 728-734.

[11] Bowles C, Chen L, Guerrero R, et al. Gan augmentation: Augmenting training data using generative adversarial networks [J]. arXiv preprint arXiv:1810.10863, 2018.

[12] Frid-Adar M, Diamant I, Klang E, et al. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification[J]. Neurocomputing, 2018, 321: 321-331.

[13] Jiangsha A, Tian L, Bai L, et al. Data augmentation by a CycleGAN-based extra-supervised model for nondestructive testing[J]. Measurement Science and Technology, 2022, 33(4): 045017.

[14] Mery D, Riffo V, Zscherpel U, et al. GDXray: The database of X-ray images for nondestructive testing [J]. Journal of Nondestructive Evaluation, 2015, 34(4): 42.

[15] Nichol A Q, Dhariwal P. Improved denoising diffusion probabilistic models[C]//International Conference on Machine Learning. PMLR, 2021: 8162-8171.

[16] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]// Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.

[17] Wang H, Cao P, Wang J, et al. Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer[C]//Proceedings of the AAAI conference on artificial intelligence. 2022, 36(3): 2441-2449.

[18] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017.

[19] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.