Conditioned Diffusion for Manufacturing Data: Improving Generation Controllability

*Jaewoong Kim, Dongso Kim, Kyongtae Park**
AI TF of Mobile Business Samsung Display Co., Ltd., South Korea
E-mail: jaewoong29.kim@samsung.com

Abstract

In the manufacturing industry, enhancing production quality is a crucial aspect of business success. Rapid detection and elimination of various defects during the production process are essential for maximizing profits. The significance of data in this regard has been increasingly recognized, with AI technology emerging as an effective solution to numerous industrial challenges. However, data-intensive companies with extensive know-how face a considerable challenge in obtaining balanced datasets for training AI models due to the imbalance between defective and normal data.

To address this issue, Generative Adversarial Networks (GANs) had been proposed as potential solutions. However, instability during GAN learning and mode collapse leading to insufficient diversity in generated results have limited their applicability in industrial settings. The emergence of diffusion models as a promising next-generation image generation technology has drawn significant attention due to their stability and ability to generate high-quality images that closely resemble real data.

In this paper, we propose an innovative approach to generating balanced datasets for manufacturing industries using the advantages of diffusion models. By transforming existing model to enable take additional embedding effectively, the diffusion model could learn numerical information that comes from the microscopic images of the defective circuits. Our methodology involves training the model from zero to fully adapt the structured condition of the data, enabling its application in manufacturing contexts where free-form prompts are not feasible.

Specifically, we extracted coordinate information from defective image data using image processing, and used as additional features to learn. We demonstrate that this approach effectively generates datasets to address data-imbalance problem with successful reflection of input information and the resulting images. By employing diffusion models to generate balanced manufacturing data, we aim to improve AI model performance in detecting and eliminating defects, ultimately leading to increased production efficiency and improved product quality.

Author Keywords

Image Generation; Diffusion; Variational AutoEncoder; DDPM; Data Imbalance;

1. Introduction

In the manufacturing industry, enhancing production quality is vital for business success. Rapid detection and elimination of various defects during production are essential for maximizing profits. Data's significance has grown, with AI technology addressing numerous industrial challenges. However, data-intensive companies face a challenge in obtaining balanced datasets due to the imbalance between defective and normal data.

Generative Adversarial Networks (GANs) were proposed as solutions but faced limitations due to instability during learning and insufficient diversity in generated results. Diffusion models emerged as a promising alternative, offering stability and high-quality images that closely resemble real data.

This paper introduces an innovative approach to generating balanced datasets for manufacturing industries using diffusion models. By adapting the model to effectively utilize additional embeddings, it learns numerical information from microscopic images of defective circuits. We extract coordinate information from defective image data and use it as additional features. Our methodology trains the model from scratch, enabling its application in manufacturing contexts where free-form prompts are not feasible. Our approach effectively generates balanced datasets with successful reflection of input information and high-quality images. By employing diffusion models to generate balanced manufacturing data, we aim to improve AI model performance in detecting and eliminating defects, leading to increased production efficiency and improved product quality.

2. Methods

Data Preparation

The image data utilized for training consisted of approximately 8000 pictures obtained through microscopic inspection of defective display pixel circuits. The types of defects included scratches and various forms of imperfections, totaling 10 distinct categories with around 200 to 1000 images per category.

To extract coordinate information from the images, we manually overlapped the images to identify recurring patterns. Given the consistent arrangement of pixels in a circuit's cycle, the overlapped image was cropped within 1.5x cycle and was thresholded to highlight only pixel regions. The resulting pixel highlight images were employed for template matching to determine the maximum similarity position's coordinates (Fig. 1). The extracted x, y coordinates from each image were combined with the corresponding defect label to formulate input prompts (Fig. 2). The input prompts were recorded and saved in a .txt file as captions.

ISSN 0097-996X/25/5603-1779-$1.00 © 2025

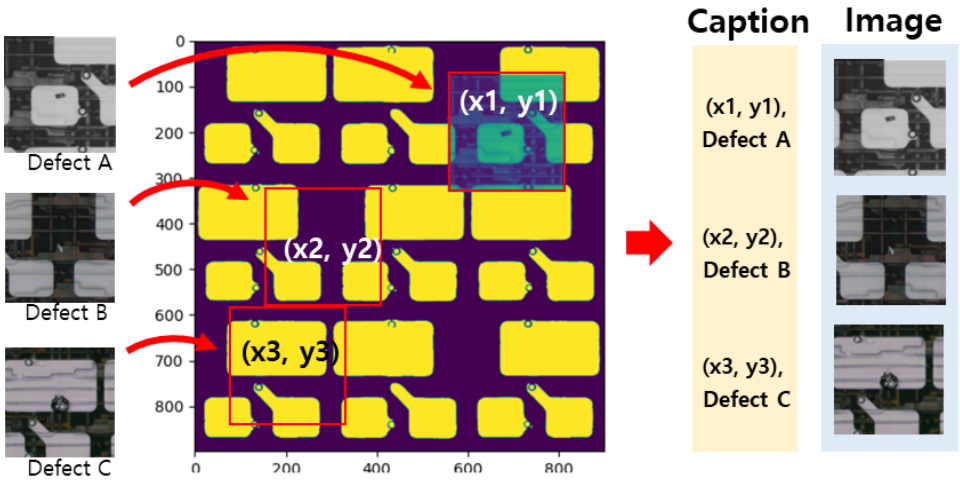

Figure 1. Work flow of data preparation: Each image is template-matched to the repetitive pattern of circuit and coordinate information is extracted to be merged with defect name into caption data.

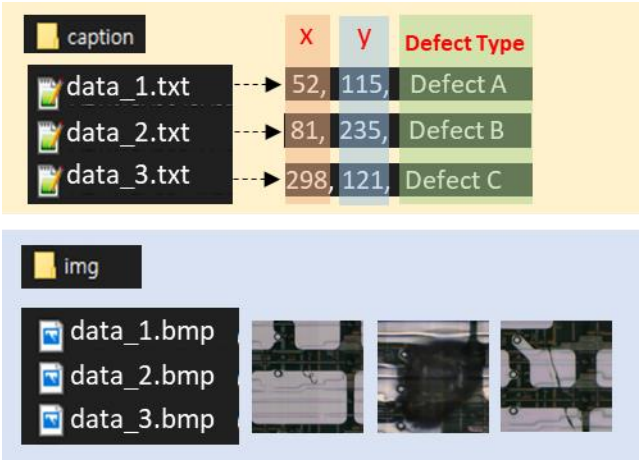

Figure 2. Exemplary data of caption and images. Caption data contains x, y coordinates and defect name, and image data is defective images of various defects included.

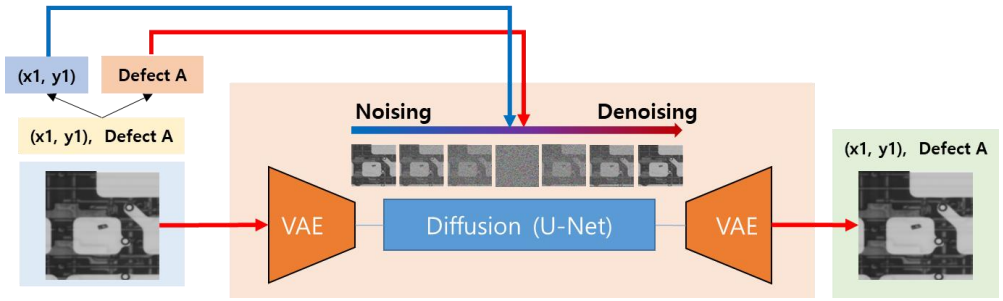

Figure 3. Description of embedding of coordinate and defect names into Diffusion model. A VAE model was used to enable latent space manipulation through encoding and decoding, while Unet-shaped diffusion model focuses on denoising data corrupted by noise.

Diffusion Model Architecture

The original diffusion model was inspired by the implementation available on an open source GitHub repository of "ExplainingAI-Code" (https://www.github.com/explainingai-code/StableDiffusion-PyTorch). For the core diffusion model, Denoising Diffusion Probabilistic Models (DDPM) was employed, while Vector-Quantized Variational Autoencoder (VQVAE) was used for the VAE model to ensure computational efficiency and training stability [2][3].

Upon initiating learning, the caption is first parsed to the x,y coordinates and defect name. Subsequently, the text of the defect name is transformed into a vector using OpenAI's Contrastive Language-Image Pretraining (CLIP) tokenizer[4]. These vectors are then merged with the x, y coordinate information and combined in the mid layer for feature fusion.

Training Diffusion Model

For encoding and decoding of Images, we trained VQVAE for 50 epochs and subsequently DDPM for an additional 200 epochs. The number of timesteps for noising and denoising processes was set to 1000. Since the dataset lacked normal data and all captured images had distinct coordinates, it was impossible to form image pairs based on direct defect presence. Therefore, simple denoising was performed using classifier-free-guidance (CFG) with a value of 1.

Image Generation

To generate an image, the trained model takes a prompt which contains x, y coordinate and defect name in a string as an input. This input prompt undergoes parsing to provide conditions for controlling the denoising procedure through the defect name.

GUI implementation

The entire process was packaged into a user-friendly GUI program to enable seamless usage for users without prior knowledge of Diffusion. To enhance the user experience, only the number of epochs and timesteps were provided as inputs for hyperparameters that significantly impact image generation quality. Additionally, options were provided for introducing jittering in input coordinates or selecting completely random coordinates during image generation to ensure a diverse range of coordinate conditions.

Evaluation of Image Generation Quality

To assess the enhancement in image generation quality during DDPM training, we saved a model file every 10 epochs. Upon completion of DDPM learning, we loaded each model file and generated 10 images from the locations where randomly-selected real image's coordinates for comparison. The comparison was conducted by calculating Structural Similarity Index Measure (SSIM) and Learned Perceptual Image Patch Similarity (LPIPS)[5] between the generated and actual images, excluding the regions of varying shapes where defects were present in both sets of images. These metrics were then plotted against epochs.

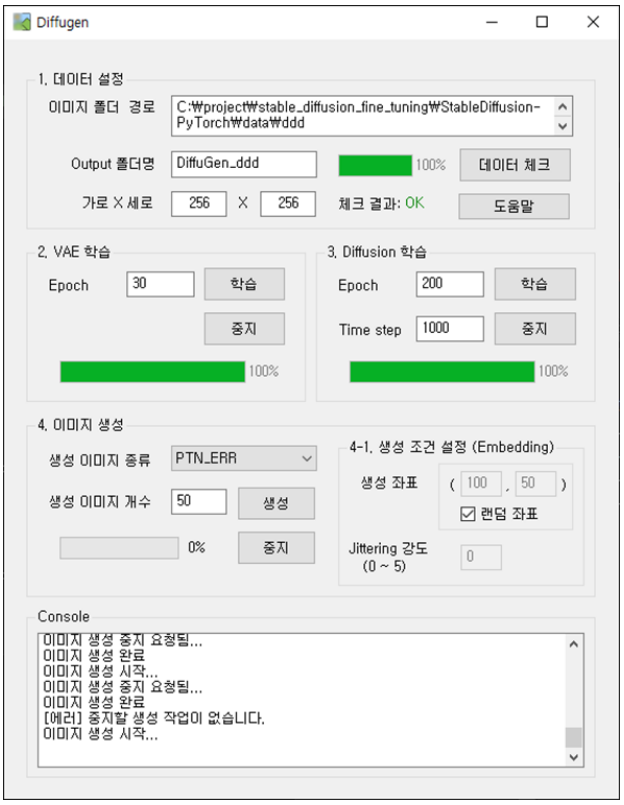

Figure 4. Screen shot of the UI implementation. Most of the hyper-parameters and configure settings were hidden for seamless usability of various users.

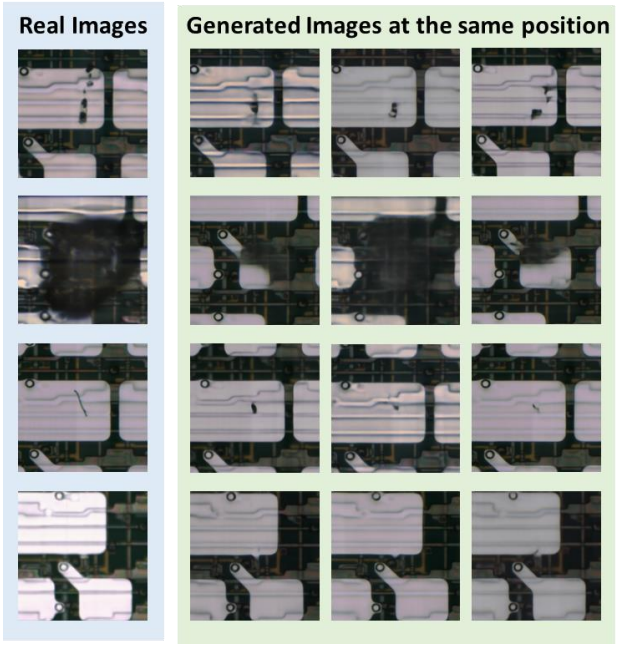

Figure 5. Results of the defective image generations. All images were generated with controlled coordinates to align with those of the images in the left column.

3. Results and Discussion

The image generation results (Fig. 5) demonstrate that the model effectively creates plausible defect images in various shapes while accurately reflecting the input coordinate information. This indicates that the additional embedding of coordinate information was properly learned during training. The generated images exhibit varying textures and contrasts, which stem from the lack of quantification of these features in the training caption data. In reality, this diversity is a result of different conditions and equipment environments encountered during image capture, as evidenced by the original image dataset itself. By encoding more attributes into the captions and including them as embeddings for learning, greater controllability over image generation can be achieved based on these results.

To utilize generated images further for AI training purposes, it is essential that the model accurately replicates not only the pixels but also the distribution of circuits in the background. The similarity metrics evaluated outside of defect regions, such as SSIM and LPIPS scores, both increased as epochs progressed, reaching approximately 0.9 in similarity around epoch 200 (Fig. 6). Despite the inherent blurring issue present in VQVAE, these high similarity scores suggest that the generated images possess sufficient potential for use as training data.

Despite the high generation performance, there are some drawbacks that need to be acknowledged. One of the first issues is the requirement for a large amount of data. In practice, successful diffusion learning calls for at least 7000 pieces of data or more. This implies that this diffusion technology is suitable for manufacturing applications where sufficient image data is available. Otherwise the performance may not fully materialize in limited data.

Another limitation is the relatively low proportion of successful generation of defect shapes, which indicates that obtaining sufficient quantities of specific defect images requires extensive repeated generation. Diffusion models learn denoising processes through probabilistic representations instead of rule-based predictions in their generation process. As a result, stable diffusion requires writing numerous conditions for generating desired images similar to how it is necessary to write many conditions for successful image generation in the case of prompt writing. To generate wanted defect images with higher probability, additional embeddings representing various image characteristics need to be incorporated.

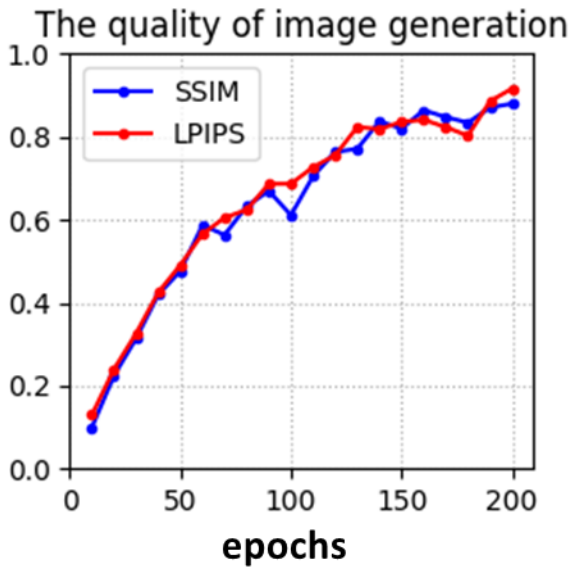

Figure 6. The evaluation results of the generation quality. The values of both SSIM and LPIPS increases along with DDPM train epoch.

4. References

[1] Goodfellow, I. et al., Generative Adversarial Networks. 2014. https://arxiv.org/abs/1406.

[2] Ho, J. et al., Denoising Diffusion Probabilistic Models. 2020

https://arxiv.org/abs/2006.11239

[3] Oord A. et al., Neural Discrete Representation Learning. 2018. https://arxiv.org/abs/1711.00937

[4] Radford, A. et al., Learning Transferable Visual Models From Natural Language Supervision. 2021. https://arxiv.org/abs/2103.00020

[5] Zhang, R. et al., The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. 2016. https://arxiv.org/abs/1801.03924