

Automated Methods for Panel Defect Image Generation and Assisting Defect Detection

Rui Zheng, Xiaojun Tang, Jingru Wang, Yuyu Liu, Xingqun Jiang

BOE Technology Group Co., Ltd, Beijing, China

Abstract

In panel production, defective products are inevitable. By utilizing deep learning models for automated defect detection and classification, defective items can be swiftly filtered out, aiding in the diagnosis of defect causes and enabling faster resolution of production issues. However, training deep learning models to achieve high accuracy typically relies on a substantial volume of samples. In real production line applications, many types of defects lack sufficient samples, which limits model performance. This paper introduces a method that leverages prompts and random noise to generate defect images, allowing the detection model to select training samples from the generated images. These selected samples are then integrated with real samples to optimize the model, thereby enhancing detection performance for rare defect types. The defect images are created using the inpainting method of the Stable Diffusion (SD) model. Multiple mask generation techniques are applied during the generation process to meet the location requirements of different defect types. The generated images exhibit randomness and diversity, and a pre-trained detection model filters out redundant or irrelevant images. This refined, unique, and varied dataset helps the detection model achieve higher accuracy. For instance, by adding 500 generated samples to the original 6 real samples, the F1 score increases by approximately 5.4%.

Author Keywords

Panel Production; Defect Detection; Defect Generation; Automatic Sample Selection; Stable Diffusion

1. Introduction

The rapid development of Stable Diffusion (SD) [2] technology in recent years has led to its widespread application across fields like image generation, image editing, and even text-to-video [8]. Among them, image generation has had a great impact on the traditional image field. For example, DiffBIR [9] leverages diffusion models to transform low-quality images into high-resolution images, achieving notable denoising and super-resolution results. Another algorithm, DiffuMask [10], generates annotated sample images containing target objects by utilizing text parsing and diffusion models, supporting applications in segmentation, detection, and recognition tasks. SD's inpainting capabilities allow controlled image generation based on masks, key points and so on, creating realistic and flexible visuals ideal for creative work or data generation, especially in the production of datasets which is of small amount.

In industrial production, defect samples emerge occasionally, and using datasets of small amount to train models to solve defect detection tasks often poses significant challenges. Data generation is a common and effective solution. Various methods for defect generation have been developed. Early approaches involved minor adjustments to defect images in terms of color and appearance, while later approaches applied GANs (Generative Adversarial Networks) to generate targeted samples. For instance, [1] introduced a GAN method with controllable defect area and intensity for industrial image defect generation. This approach

enables the intentional creation of hypothetical defect samples that might reasonably occur. Considering that GAN models can suffer from mode collapse and sometimes may produce inconsistent image quality, SD (Stable Diffusion) algorithm has emerged as a promising alternative for defect image generation. SD models produce stable, diverse, and high-quality images, allowing effective model fine-tuning with a small dataset size to yield satisfactory outcomes.

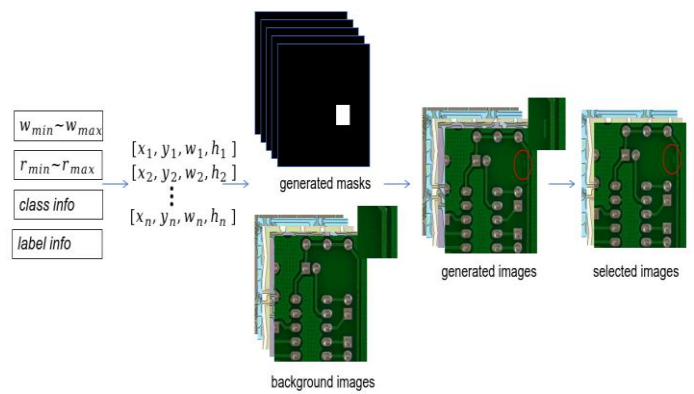


Figure 1. Diagram of automatic defect image generation and selection

In this study, we employ the inpainting method in SD (Stable Diffusion) to generate defect samples. Specifically, we create a 9-channel input image by combining the original training sample, a mask of the objects on the training sample, and the original training sample with the targets erased. Then the inpainting method generates a 3-channel output sample where the masks determine the locations and approximate shapes of the generated defects. As shown in Figure 1, the mask is defined by the bounding box's left-top point, width, and height. The mask can be automatically generated from them. To accommodate the varied needs of industrial defect generation, we devised a strategy combining different mask generation methods, creating tailored solutions for different defect types. Additionally, we propose a method of using existing samples to train high performance and low performance models in combination to screen out erroneous samples and difficult samples to assist the original samples training model. This approach helps train the model on more complex image data while reducing manual intervention, as models can largely handle the data selection process. Filtering and incorporating challenging samples improve the detection model's accuracy, addressing the limitations of direct generated samples, which may require further human inspection to remove blank or erroneous images.

2. Related Work

Data augmentation is a widely used technique in deep learning tasks, designed to expand the training dataset by generating new samples, which will improve model's accuracy and enhance

model's generalization ability. Methods for data augmentation fall into two main categories: traditional techniques and deep learning-based approaches.

Traditional data augmentation methods includes geometric transformations, color adjustments, region occlusion and so on. Geometric transformations commonly involve scaling, flipping, translation, rotation, and cropping. Color transformations adjust brightness, contrast, and saturation of the image, while occlusion methods use techniques like random pixel erasure and region masking within an image. These methods can be used in combination or individually.

There are some more advanced image processing methods based on deep learning for data augmentation, such as Mixup, which mixes two images in a certain proportion, CutMix, which pastes a part of one image onto another, Mosaic [6], which uses online random occlusion, and so on. There are also some data augmentations, getting by model generated, which can use GANs to produce samples similar to real data, though GANs are often limited by stability issues and variable image quality. Consequently, Stable Diffusion (SD) models are increasingly preferred for generating high-diversity image data. For example, the Lora [3] model can be fine-tuned with just 5-10 images and related prompts to generate novel target samples. And DiffuMask can use prompts to generate annotated segmentation samples, the samples are created only by the description and the description can be random and imaginative, it's a pretty good job. MaskDiffusion [7] can generate point-wise labels for an image either, and [4] leverages text-to-image models to generate defect samples in potentially defect-prone regions, helping to address the scarcity of defect data. This paper's approach satisfies the need to create defect samples in appropriate areas while also filtering samples to enhance the detection model's performance. By generating masks that provide target's bounding box labels and finally automatically filtering samples, the method improves the model's accuracy, using both synthetic and original samples when training the model.

3. Defect Image Generation

The SD-based inpainting method enables image completion within masked regions. With a well-trained SD model supported by ample samples, it is possible to achieve highly accurate image generation as long as the mask provided during inference is known and precise. In real-world applications, however, defect samples are often scarce, and preparing inference masks manually is both labor-intensive and prone to subjective limitations that reduce mask diversity. To address these issues, this paper proposes an automatic mask generation and sample filtering system, as illustrated in Figure 2.

The generated mask is generated by the selected box info, and the inpainting SD model was finetuned on existing industrial samples. The generated samples are the target defect images, which can be generated by the inpainting SD model with the input, concatenated noise, generated mask, defect-free background image, during the inference process. Once the target defect images are generated, pre-trained high performance detection model delete failed samples, and pre-trained low performance detection model choose only the difficult samples that will be useful for model training. These selected samples, along with the original dataset, are then used to train the target model M3. This M3 model ultimately serves as the final model deployed on the production line.

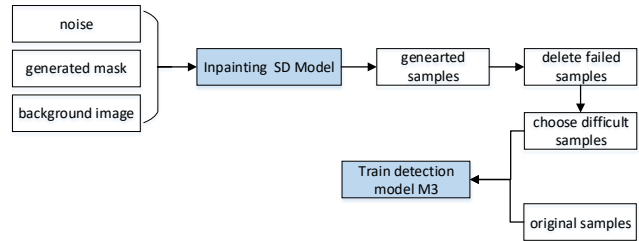


Figure 2. The Method for Training Detection Model Using Generated samples

Defect Mask Generation Strategy:

The defect images generated during production come in various forms, with variations in both shape and position, as illustrated in Figure 3. Broadly, these defects can be categorized into three types: defects with variable locations and shapes, defects with fixed locations and shapes, and defects with non-fixed locations but tendencies toward specific width and height dimensions. Therefore, the strategy for generating defect masks must be adapted to the specific defect type, varying as necessary to align with the characteristics of each defect category.

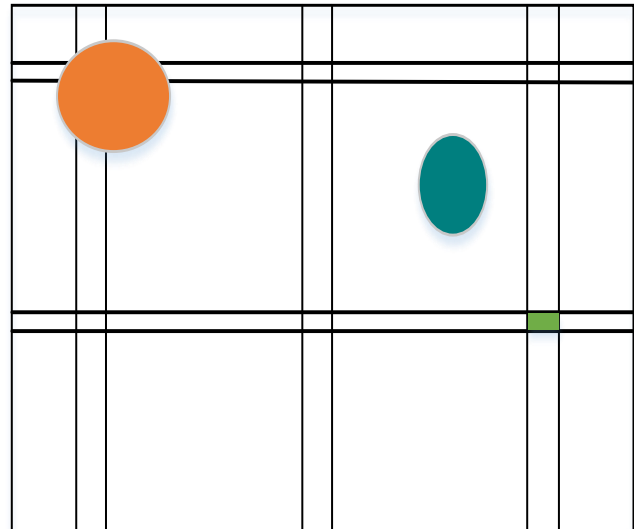


Figure 3. Distribution of defects of different categories

The mask generation for the types of defect with random locations and shapes, often referred to as a “random generation method”, must avoid generating mask with truly random values, as this would reduce the accuracy of the good sample generating rate. Our experience indicates that if the model's instructions significantly diverge from the patterns it had learned, it either fails to generate the target image or simply generates errors. In this paper, we analyze the width ranges and aspect ratio ranges of all targets within each category of original samples, using these ranges as the boundaries for randomly selecting width and aspect ratio values. With the chosen width, the target box height is calculated based on the aspect ratio, and the target position is randomly set within the image.

Another mask generation approach applies to defects that consistently occur in fixed locations and shapes, such as the green rectangular defect in Figure 3, which appears on a specific circuit component. In this case, the original image is used as the

background input, and defects are re-generated at the original location, but with slight variations in shape, as shown in Figure 4(b). This ensures that the generated images meet the requirements while also exhibiting diversity.

For defect categories without fixed locations but with standardized shapes, an alternative to random generation involves using a fixed shape ratio at variable positions. This process includes randomly selecting a target size from the real training samples, scaling it proportionally to the generated sample's dimensions, and then randomly positioning it within an intact background. This method preserves the original target's dimensions but introduces a varied background, thereby enriching the training dataset.

Inpainting Model Train and Inference:

In the process of training the inpainting model, we train the model one by one according to the defect category to achieve the best generation performance for each category. The model takes a 9-channel input, consisting of the training sample, the training sample's mask, and a training sample with the target removed. It generates a 3-channel image output. The time step (represented by t) used for adding noise during the training process are random values in the range of 0-1000. The noise addition process of the original samples x_0 is shown in formula 1, where x_t represents the noisy image at step t , β_s denotes the noise level introduced at step s , and N represents a Gaussian distribution. $q(x_t | x_0)$ represents the input sample, after adding t -step noise, and is fed into the model together with the mask m and the training sample (background image) b to predict the added noise. The principle is shown in formula 3, where $\epsilon_\theta(x_t, t)$ is the noise predicted by the model, c is the control condition. The model is optimized by calculating the mean squared error loss between the predicted noise and the actual added noise.

$$\begin{cases} q(x_t | x_0) = N(x_t; \sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)I) \\ \bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s) \end{cases} \quad (1)$$

$$x_0 = \frac{x_t - \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}} \quad (2)$$

$$\epsilon_\theta(x_t, t) = \text{Unet}(q(x_t | x_0), m, b, t, c) \quad (3)$$

In the inference process, a noise x_t is randomly initialized, a mask m is created according to the generation requirements, and a selected high-quality background sample b is used. These three components are concatenated and fed into the pre-trained model to obtain a noise prediction $\epsilon_\theta(x_t, t)$. Using Equation 2, the noise is removed from $\epsilon_\theta(x_t, t)$ to obtain x_0 . Next, following Equation 4, the background portion is superimposed onto x_0 . Then this process iterates: further noise is added to x_{t-1} , with m and b concatenated, the result is fed back into the model to predict and subtract noise, yielding x_0 , then following Equation 4. Repeating this process for a set number of time steps results in the final denoised, generated image.

$$x_0 = x_0 \cdot m + b(1 - m) \quad (4)$$

4. Defect Images Filtering

From a practical perspective, generated defect images require the removal of incorrect or redundant ones. To achieve this, we first trained a high-performance YOLOv8-L [5] model using the original data. Generated defect images were then processed through YOLOv8-L, where a low-confidence threshold was applied. If the model detected a target with confidence exceeding this threshold, the generated image was deemed valid. If the target was undetected or the confidence was below the threshold, the image was

discarded. Remaining samples then advanced to subsequent filtering steps.

Using the original image data, we trained a lower-performance YOLOv8-S model. Defect samples that have passed through YOLOv8-L were subsequently processed by YOLOv8-S with a high-confidence threshold. If YOLOv8-S detected a target with confidence exceeding this threshold, we then considered that samples contained features which lightweight model had already well-learned and removed this set of data. The remaining samples, containing unique or complex features, were retained and combined with the original data to further enhance the training of the final target detection model.

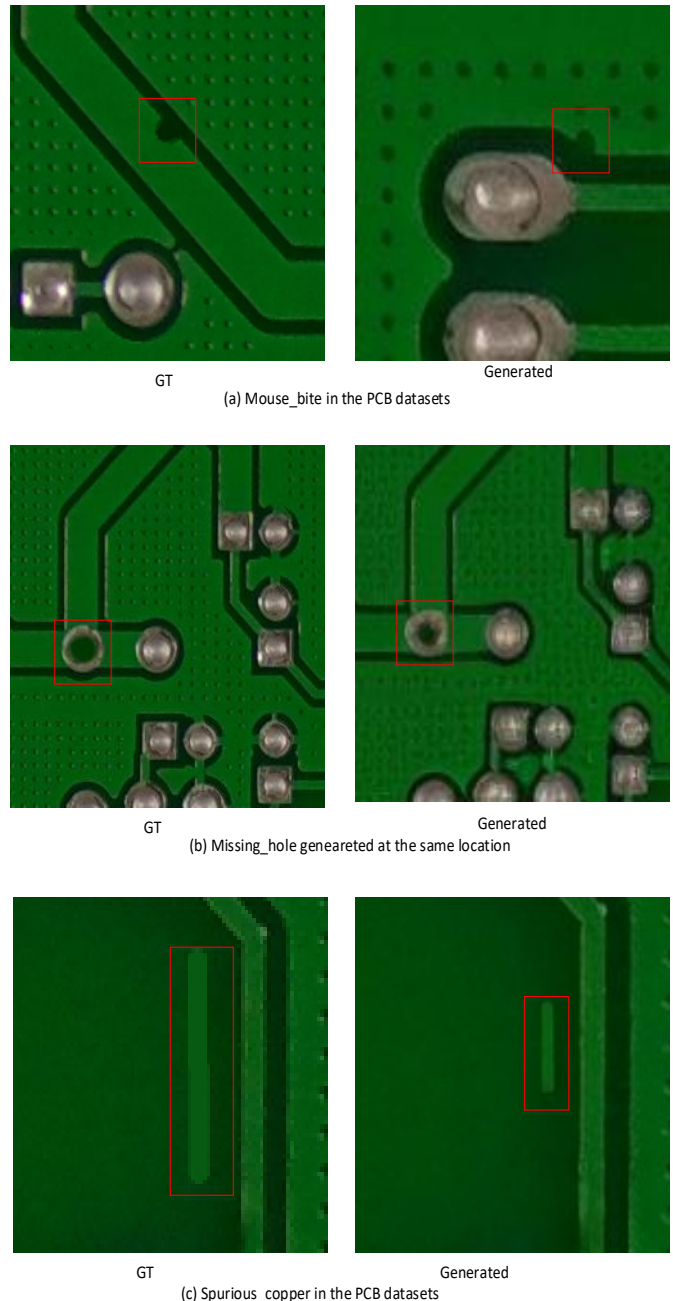


Figure 4. Generations under different mask generation strategies

5. Experimental Results

We conducted the experiments on a V100 machine. Taking into account the unique characteristics of industrial defect images, where defect location and shape are often tied to the defect category, we trained one generative model for one certain defect. And for this study, we developed specific strategies for generating various types of defect images. Figure 4 presents the generated image results, with defects highlighted in red boxes. Fig4 (a) shows both real samples (GT) and the output of randomly generated defect image. Mouse_bite is one type of defects in the PCB datasets, the defects are characterized by having the same color and shape resembling that of a mouse bite, which aligns well with real-world conditions while their shapes and locations of the generated defects are randomly assigned. Fig4(b) showcases the defect diversity achieved by regenerating defect at the original location in the same image, maintaining identical target dimensions while allowing for shape variability within the same location. Fig4(c) illustrates defects with scaled original dimensions but placed randomly over non-defective images. Because there are relatively few target samples of the certain shapes, we maintained a few target aspect ratios, scaling its dimensions based on the ratios of the original and generated new images. All the above experiments confirm that these mask generation approaches are effective, and help us to create more ideal images.

To further verify the effectiveness of generated images on defect detection tasks, we designed and trained generative models using various amounts of training data to yield models of differing performance levels. The original training data amount ranges from 6 to 200, and then the models each generated 500 defect images, which were trained again both with and without integration into the original training dataset. The test images were always the same 100 images for the model trained with only real training images and with extra generated images. As shown in the table, we used F1 score to evaluate the model's detection and classification capability, and all the results in the table indicate an improvement in F1 score. When the initial dataset size is limited, a more pronounced effect is observed.

Table 1. Comparison of different amounts of training data and performance improvement of detection models

Real Train Images	Real Test Images	Generated Train Images	Real	Real + Generated
			F1 Score	F1 Score
6	100	500	0.639	0.693 +5.4%
12	100	500	0.766	0.811 +5.5%
25	100	500	0.874	0.938 +5.4%
50	100	500	0.919	0.937 +1.8%
100	100	500	0.969	0.977 +0.8%
200	100	500	0.977	0.978 +0.1%

6. Conclusion

This study introduces a method that automatically generates masks based on defect categories to guide data generation in an inpainting SD model. After image generation, using high and low performance detection models trained on the original dataset can efficiently filters out samples that are true and underrepresented in the target model's training data. Then combining the generated samples with the original samples to train the target model is more effectively. Experiments confirmed that the mask generation approach aligns well with real-world industrial needs, and the model-based image data filtering method proves to be highly effective, reducing manual labor while delivering excellent filtering outcomes. Additionally, generating defect image data significantly enhances detection model performance. In practical application, it is obvious that precise object position labeling boosts mAP scores for the model which will be evaluated about the location performance. However, due to gaps between bounding boxes on the mask and the actual generated targets, further exploration is needed to create precise labels for the targets during sample generation process. This refinement aims to establish a stronger foundation for detection and segmentation tasks.

7. References

- [1] Niu S , Li B , Wang X ,et al.Region- and Strength-Controllable GAN for Defect Generation and Segmentation in Industrial Images[J].IEEE Transactions on Industrial Informatics, 18[2024-10-30]. DOI:10.1109/TII.2021.3127188.
- [2] Rombach R , Blattmann A , Lorenz D ,et al.High-Resolution Image Synthesis with Latent Diffusion Models[J]. 2021.DOI:10.48550/arXiv.2112.10752.
- [3] Hu E J , Shen Y , Wallis P ,et al.LoRA: Low-Rank Adaptation of Large Language Models[J]. 2021.DOI:10.48550/arXiv.2106.09685.
- [4] Pan Z , Shenoy R , Balakrishnan K , et al. 32-1: Improving QD Backplane Defect Image Generation Using Automatic Masking in Diffusion Models[C]//SID Symposium Digest of Technical Papers. 2024, 55(1): 409-412.
- [5] Varghese R , Sambath M .YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness[C]//2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS).0[2024-10-30].DOI:10.1109/ADICS58448.2024.10533619.
- [6] Bochkovskiy A , Wang C Y , Liao H Y M .YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. 2020.DOI:10.48550/arXiv.2004.10934.
- [7] Kawano Y , Aoki Y .MaskDiffusion: Exploiting Pre-Trained Diffusion Models for Semantic Segmentation[J].IEEE Access, 12[2024-10-30].DOI:10.1109/ACCESS.2024.3456442.
- [8] Sun R , Zhang Y , Shah T ,et al.From Sora What We Can See: A Survey of Text-to-Video Generation[J]. 2024.
- [9] Lin X , He J , Chen Z ,et al.DiffBIR: Towards Blind Image Restoration with Generative Diffusion Prior[J]. 2023.
- [10] Wu W , Zhao Y , Shou M Z ,et al.DiffuMask: Synthesizing Images with Pixel-level Annotations for Semantic Segmentation Using Diffusion Models[J].ArXiv, 2023, abs/2303.11681.DOI:10.48550/arXiv.2303.