Aggregated Data Augmentation for Defects Using Enhanced Diffusion Models and Poisson Blending

1st Chen Duo
School of Astronautics
Harbin Institute of Technology
Harbin, China
chenduo@stu.hit.edu.cn

2nd Jin Jing
School of Astronautics
Harbin Institute of Technology
Harbin, China
jinjinghit@hit.edu.cn

3rd He Xujie
School of Astronautics
Harbin Institute of Technology
Harbin, China
hexujie@stu.hit.edu.cn

4thLiu Yi
School of Astronautics
Harbin Institute of Technology
Harbin, China
21B904020@stu.hit.edu.cn

5th Guo Yanan
School of Astronautics
Harbin Institute of Technology
Harbin, China
yananguohit@stu.hit.edu.cn

6th Ni Yanshu
School of Astronautics
Harbin Institute of Technology
Harbin, China
niyanshu_hit@163.com

2025 37th Chinese Control and Decision Conference (CCDC) | 979-8-3315-1056-5/25/$31.00 ©2025 IEEE | DOI: 10.1109/CCDC65474.2025.11090413

***Abstract*—Detecting defects of workpiece surface is crucial for maintaining operational reliability and safety in the industry. However, obtaining defect samples of workpiece such as cracks, pinholes, and folds, is labor-intensive and time-consuming. Moreover, the samples collected often exhibit class imbalance, leading models to easily fall into local optima. Currently methods, such as under-sampling and over-sampling, all rely on a limited number of real samples and suffer from a lack of data diversity. To address these chanlledges, this paper proposes the aggregated data augmentation for defects using enhanced diffusion models and poisson blending. For defects such as cracks and pinholes, we propose a Segmentation Driven Generation method combining Segmenting Everything In Context(Seggpt) segmentation with Open-Set Grounded Text-to-Image Generation(GLIGEN) prompt generation to create defect samples for corresponding categories. For more complex defects such as folds, we introduce a module named PoissonSeg that integrates Seggpt segmentation with the Poisson equation to generate samples of complex defects. Taking the steering wheel defect data set as an example, experiments results, including the comparisons and single-category ablation studies demonstrate using our proposed method significantly enhances the detection performance of six different defect detection algorithms on steering wheel defects, with increases of 0.144 in terms of mAP0.5, proving the effectiveness of the proposed method in data augmentation for defect detection.**

Keywords—industrial artificial intelligence, intelligent manufacturing, defect detection, data augmentation

I. Introduction

In recent years, the rapid development of the intelligent manufacturing industry has gradually pressured enterprises to become more competitive in response to global economic challenges. This increased demand for high-quality product production has heightened the need for intelligent inspection systems, making quality control a critical issue in the intelligent manufacturing sector. Automotive defect detection is a significant area within this field, involving comprehensive inspections to identify potential defects or malfunctions in vehicles. The steering wheel, being a crucial component for controlling the vehicle, is particularly important, as its surface defects can affect the driver's operational feel, reduce the stability and reliability of the steering wheel surface, and even compromise driving safety.

However, in real industrial scenarios, one of the major challenges in implementing defect detection using deep learning technology [1], [2] is the acquisition of defect samples. The data collected in practice often consists of a significantly higher number of normal samples compared to defect samples, and there is a substantial imbalance in the quantity of samples among different defect categories. This imbalance can cause the model to become biased towards categories with more samples during training, reducing its detection capability for certain categories. Such bias may also lead to overfitting, decreasing the model's convergence speed and resulting in poor performance on unseen samples. Therefore, addressing the issue of sample category imbalance in defect detection is crucial for enhancing the practical applicability and generalization of defect detection models.

The fundamental cause of sample category imbalance lies in the disparity in the number of real images for each category. Thus, the most direct and effective way to address this issue is to alter the sample quantity for each category to achieve balance.

One approach is through data processing methods such as under-sampling and over-sampling. Under-sampling reduces the number of samples in the majority class to balance the quantities between categories. T. M. Khoshgoftaar successfully improved fault detection performance under imbalanced data conditions using random under-sampling [3]. To avoid the loss of many valid samples with random under-sampling, which can reduce the final model accuracy, researchers have proposed under-sampling methods with selection strategies. One approach involves clustering algorithms: H. Chen introduced Euclidean distance [4], L. Li used the nearest neighbor decision rule [5], and A. Abdo applied fuzzy c-means clustering to determine the position of samples within a category [6]. Samples closer to the category center are deemed more important, while those farther from the center are randomly discarded. This method preserves representative samples in the corresponding class, reducing information loss. Another approach is based on sample density strategies, which utilize distances to give the distribution density of samples. By reducing under-sampling in dense areas and increasing it in sparse areas, the sample set can be balanced.

Contrary to under-sampling, over-sampling addresses data imbalance by increasing the number of samples in minority defect classes. Common simple over-sampling methods include random over-sampling, overlapping sampling, geometric transformation, and Synthetic Minority Over-sampling Technique (SMOTE). Yang proposed a variable-scale overlapping sampling strategy, conducting small-scale overlapping sampling for majority classes and large-scale overlapping sampling for minority classes [7]. However, the capability of this method to solve data imbalance is very

This study was partially funded by the Natural Science Foundation of China (12373107) and supported by the research and development project of a complex workpiece visual defect detection imaging system and software tools of China [No.2023ZXJ01A01].

limited, as overlapping sampling requires a long continuous sampling period to achieve good results. Without such an extended collection period, the increase in sample quantity from overlapping sampling is limited. Additionally, if the collection period is too long, it requires substantial storage space to store the data, posing a challenge for industrial monitoring systems. Researchers also proposed a method to define sample importance based on the number of majority samples around minority samples. If a minority sample is surrounded by many majority samples, its importance is higher. If a minority sample is entirely surrounded by majority samples, it might be identified as noise and removed from the data. ADASYN (Adaptive Synthetic Sampling) determines the participation ratio of each minority sample in generating new samples by normalizing the number of majority samples around it, emphasizing the creation of more minority samples near the classification boundary to balance the dataset [8]. Many clustering-based sampling strategies have also been proposed, where minority samples are first clustered, dividing the minority class into different subspaces and then over-sampling each subspace separately. Examples include k-means clustering, two-step clustering, hierarchical clustering, variable density clustering algorithms, and fuzzy c-means. Overall, samples obtained using the SMOTE method are easily influenced by the original sample distribution and lack consideration for sample dynamics, resulting in poor diversity of new samples.

Unlike the two previous methods, another approach utilizes deep learning techniques to learn the distribution characteristics of real samples to generate new ones. Currently, two widely used models in the generative domain are GAN and VAE and their variants. Generative Adversarial Networks (GANs) consist of a generator and a discriminator. The generator's task is to produce data that can deceive the discriminator, while the discriminator tries to distinguish between real and generated data. These two parts compete during training, continuously improving their performance, ultimately enabling the generator to produce high-quality samples. Variational Autoencoders (VAEs) are a type of autoencoder that transforms data into a latent space representation through an encoder and then reconstructs the data through a decoder, aiming to minimize the difference between input and reconstructed data. Unlike traditional autoencoders, VAEs generate a probability distribution during encoding, from which diverse outputs can be sampled. Liu used adversarial training classifiers to improve the robustness of rolling bearing fault diagnosis [9]. Zareapoor proposed an improved GAN model to generate erroneous samples and identify classifications [10]. DCGAN was used to generate dielectric line defects to address the insufficiency of defect samples [11]. Lu proposed a semantic label-enhanced VAE method for contact network component defect detection, named DefVAE [12]. Ferdousi R proposed a VAE-based rail defect synthetic image generation technique, combining weight decay regularization and image reconstruction loss to prevent overfitting [13].

However, in real industrial scenarios, defect samples are often difficult to obtain and very limited in number. For example, in the context of steering wheel defect detection, there are numerous defect categories and complex defect types, with imbalanced sample numbers across different categories, and the characteristics of defect samples are complex and different in size, while background interference exists, as depicted in Fig. 1. The limited number of samples makes it challenging for the model to learn effective features during training, and the imbalance in the number of samples across defect categories may lead to overfitting towards the majority class, affecting the model's generalization ability and resulting in poor performance in practical applications. Existing over-sampling and under-sampling methods are based on limited real samples, which may also lead to insufficient model generalization. Generative models such as GAN and VAE can generate high-quality samples, provided they have a sufficiently large dataset, making them unsuitable for steering wheel defect detection.

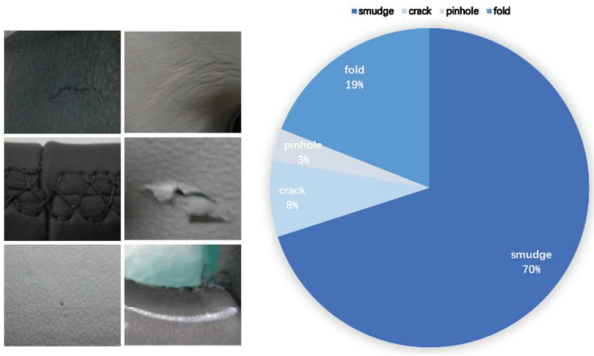

Fig. 1. Example images of steering wheel defects and the pie chart of label counts in steering wheel dataset.

In this paper, to address the aforementioned issues, we propose the aggregated data augmentation for defects using enhanced diffusion models and poisson blending. For surface defects such as cracks and pinholes, we propose a Segmentation Driven Generation method that combines our Seggpt [14] segmentation with GLIGEN [15] prompt generation. After the Seggpt segmentation delineates the target area, the GLIGEN generation model creates defect samples of the corresponding category at the selected target locations. For pinhole-type micro-defects, we have specially designed a scaling generation step to achieve better results. In the case of complex wrinkle defects, we propose PoissonSeg module that integrate Seggpt segmentation with the Poisson equation. By applying the Poisson equation to the target area selected by Seggpt, we achieve the generation of samples for complex defects, addressing the imbalance in the number of samples for different defect categories in the original dataset. In this study, we took the steering wheel defect dataset as an example to design and conduct comparative experiments as well as single-category ablation experiments. The experimental results indicate that utilizing our proposed method to balance the number of samples for each category significantly enhances the detection performance of six different defect detection algorithms for steering wheel defects. Incorporating generated samples of a specific defect category into the training set also notably improves the detection performance for that particular category and most other defect categories.

Our main contributions are as follows:

1) In view of the difficulty in collecting defect samples and the unbalanced number of collected defect samples of various categories in the field of defect detection, we propose the aggregated data augmentation for steering wheel defects using enhanced diffusion models and poisson blending.

2) For simple defects, we propose a Segmentation driven generation method, which combines the functions of

segmentation model and diffusion model to achieve simple defect generation. For complex defects, we propose PoissonSeg module, which uses Poisson equation to perform image fusion of defects in the segmented target region.

3) Our results show a positive effect on the defect detection algorithms, whether it is the defect detection algorithm of CNN structure or transformer structure.

II. METHODOLOGY

A. Overview

In this section, we propose a novel data augmentation framework that synergizes diffusion models and image fusion techniques. This approach is specifically designed to mitigate the issue of sample number imbalances across various defect categories. Our methodology incorporates the Seggpt segmentation model, the GLIGEN generative model, and Poisson equation-based image fusion to achieve a comprehensive augmentation strategy.

The primary objective of our method is to alleviate the detrimental effects of sample number disparities among defect categories. The overall architecture of our proposed framework is depicted in Fig. 2. To accommodate the varying complexities of different defect types, we employ two distinct generation pathways, each tailored to optimize the augmentation process for specific defect characteristics. Taking the defect detection problem of the steering wheel surface as an example, the overall process is as follows: with the defect-free steering wheel image as input, Seggpt is used to segment the image to different degrees and select different target areas by providing example images. Defects are divided into two Generation paths according to the complexity of defect features. Simple defects are generated in the target area by Segmentation Driven Generation method. For small defects, corresponding scaling modules are set to improve the quality of defects generated. PoissonSeg module is used for image fusion of complex defects to obtain new defect samples.

B. Segmentation Driven Generation method

Surface defects such as cracks and pinholes, exhibit an imbalance in sample quantities within the dataset. Given their relatively simple characteristics and the broad range of potential target regions, we propose a Segmentation Driven Generation method. This module first employs the Seggpt foreground extractor to identify the target region where the steering wheel is located. Subsequently, it uses real defect images and textual descriptions as prompts for the GLIGEN generator to produce the corresponding defect images. Specifically, for pinhole defects, which are particularly small, we incorporate a scaling module before and after the GLIGEN generator to enhance the generation quality. This approach ensures that the generated samples are both accurate and effective, addressing the sample imbalance issue for these defect types.

- Foreground Extractor：

Seggpt is a versatile segmentation model capable of handling multiple segmentation tasks by framing them as a contextual coloring problem. The core idea is to assign random colors to each data sample, compelling the model to rely on contextual information rather than colors to determine the task. During training, Seggpt dynamically creates paired images with similar contexts for each training image, sharing the same random color mapping. This approach ensures the model continually encounters contextual information during training, enhancing its generalization ability. Additionally, Seggpt introduces spatial and feature integration strategies for context integration. After training, Seggpt can segment input images based on contextual features derived from example images. We leverage this capability to perform prompt-based segmentation on input images to isolate the foreground regions (i.e., the target area of the steering wheel) while filtering out background regions, thereby providing a basis for defect generation in the next step, as depicted in Fig. 3.

- Defect Generation with GLIGEN

GLIGEN is an image generation model based on latent diffusion models (LDMs). Diffusion models are effective for text-to-image generation, involving a forward diffusion process that adds noise and a reverse denoising process to generate new samples. The forward process learns a latent representation of images, and the diffusion model is trained in the latent space to reduce computational costs. The training objective is to gradually generate clearer samples from noise. The process of reverse denoising is the process of generating new data from noise samples. This process can be viewed as learning a de-noising neural network f_θ, used to estimate the distribution of the sample. The training goal of the diffusion model is to optimize the parameters of the neural network by minimizing the difference from the real noise. A common loss function is the mean square error (MSE) loss:

$$L = \mathbb{E}_{t,x_0,\varepsilon}[\| \varepsilon - \varepsilon_\theta(x_t, t) \|^2] \tag{1}$$

$\epsilon_\theta(x_t, t)$ is the output of the model, x_t is the sample at the time step t. x_{t-1} is the sample from the previous step, ε_t is the noise extracted from the standard normal distribution. GLIGEN aims to enhance control over generated images, allowing sample generation through text input as well as input conditions like prompt images, key points, and bounding boxes. By adding new gated self-attention layers and a continual learning strategy, GLIGEN successfully incorporates new positional information (e.g., prompt images, key points, bounding boxes) into large-scale text-to-image generation models without altering the original model weights, significantly improving control and quality of generated images. We denote $v = [v_1, \cdots, v_M]$ as the visual feature tokens of an image. Instructions based on the text-to-image model are defined as a combination of header and entity based [15]:

$$\text{Instruction: } y = (c, e) \tag{2}$$

$$\text{Caption: } c = [c_1, \cdots, c_L] \tag{3}$$

$$e = [(e_1, l_1), \cdots, (e_N, l_N)] \tag{4}$$

where L is the caption length, N is the number of entities to ground, e is the semantic information of the grounding entity. In certain cases, both semantic and spatial information can be represented with l alone. GLIGEN freezes the original self-attention layer and cross-attention layer, and adds a gated attention layer to it, thereby introducing new location information:

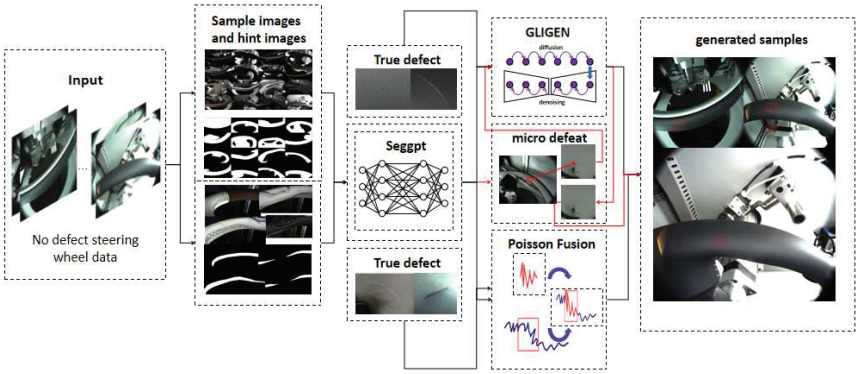

Fig. 2. The overall architecture of the data augmentation framework.

$$v = v + SelfAttn(v) \tag{5}$$

$$v = v + CrossAttn(v, h^c) \tag{6}$$

$$v = v + \beta \cdot \tanh(\gamma) \cdot TS(SelfAttn([v, h^e])) \tag{7}$$

Where h^e is as follows:

$$h^e = MLP(f_{text}(e), Fourier(l)) \tag{8}$$

where Fourier is the Fourier embedding, and MLP is a multi-layer perceptron that first concatenates the two inputs across the feature dimension.

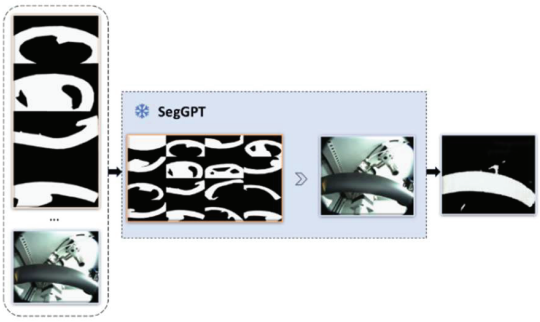

Fig. 3. Seggpt splits the foreground area.

For surface defects such as smudges, cracks, and pinholes, we utilize GLIGEN's ability to generate images based on multiple input conditions. We provide real defect sample images as prompt images and textual descriptions as text inputs, generating corresponding defects under these dual prompts. Prior to this, it is necessary to determine the exact location of the defect in the original image. The overall process involves using Seggpt to extract the foreground (steering wheel) from the image to identify candidate regions for defect generation, selecting a suitable region for the defect, and then using GLIGEN to generate the defect at this location. For pinhole defects, given their small size, using the original sample image as the prompt may result in information loss and invisible defects in the generated samples. To address this, we first enlarge the pinhole region in the original sample image, use the enlarged image for defect generation, and then resize the result back to its original dimensions, thereby obtaining visible small defect samples, as shown in Fig. 4.

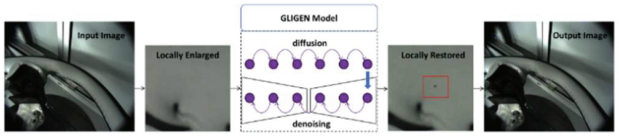

Fig. 4. Small defect scaling generation process.

C. PoissonSeg module

In the context of imbalanced sample categories for defects, certain defect types exhibit complex characteristics and specific defect locations. For instance, wrinkle defects possess intricate and variable texture information, and their occurrence is typically localized to fixed regions, rendering diffusion-based sample generators unsuitable. Therefore, we propose a sample generator based on Poisson distribution. This module also employs the Seggpt foreground extractor to identify target regions in defect-free steering wheels. However, the prompt image in this case focuses on the fixed regions where wrinkles commonly occur, rather than the entire steering wheel. Subsequently, using the defect areas from real defect samples as a foundation, the Poisson equation-based image fusion technique is utilized to generate the defect samples. This approach ensures the accurate and effective synthesis of complex defect samples, thereby addressing the imbalance in sample numbers for these specific defect types.

The core idea of Poisson fusion is not to directly overlay the two images to be fused but to generate a new image in the fusion region of the target image based on the guidance field (essentially the gradient field) of the source image. In other words, by providing the gradient field of the source image, the target image generates the fused part according to its characteristics, following the gradient field of the source image, achieving a more natural fusion effect. The fundamental principle of Poisson fusion is solving the Poisson equation. If the source image is g, the target image is f_*, the selected fusion region is Ω, and the boundary is $\partial\Omega$, then the corresponding Poisson equation is:

$$\nabla^2 f_* = \nabla g \text{ in } \Omega \tag{9}$$

∇^2 is a Laplace operator, ∇g is the gradient of the source image, Satisfy the condition:

$$f|_{\partial\Omega} = f^*|_{\partial\Omega} \tag{10}$$

For wrinkle defects on the steering wheel, which typically occur in the outer and inner connection areas due to skin pulling, using GLIGEN for defect generation may not ensure reasonable placement, such as generating wrinkles on the upper or lower surfaces of the steering wheel. Given the complexity of wrinkle textures, the quality of generated samples may be poor. Therefore, we employ Poisson equation-based image fusion, using real defect samples and defect-free steering wheel images to create new defect samples. Specifically, for each defect type, Seggpt is used to segment the target region of the defect-free steering wheel image (e.g., the stitching area for thread defects, the inner and outer connection area for wrinkles). The final fusion location is determined within the target region, and Poisson fusion is performed using the defect region from real defect samples and the defect-free steering wheel image at the selected location to create new defect samples.

III. Experiments And Results

A. Comparative Experiments

Experimental Design: To investigate the effectiveness of the proposed data augmentation method, which integrates diffusion models and image fusion, in addressing the imbalance of defect sample categories, we designed the following comparative experiments using the steering wheel surface defect dataset. We selected representative object detection algorithms in the field of defect detection, including YOLOV5, YOLOV7, FasterRCNN, DINO, RetinaNet, and SSD. The detection performance of these algorithms was compared under three scenarios: without any augmentation, using default traditional augmentation methods (e.g. flipping, scaling), and using the proposed data augmentation method that integrates diffusion models and image fusion.

Dataset: The data was sourced from a self-developed online inspection system for steering wheel surface defects, comprising 1420 real images, with 1101 images in the training set and 319 images in the test set. The number of annotated instances for each defect category is shown in Table 1.

TABLE I. Table of Labeled Samples in the Steering Wheel Defect Dataset

Category	Training Set Labels	Test Set Labels
smudge	1341	428
cracks	147	55
pinholes	63	18
folds	364	110

The original dataset exhibits significant disparities in the number of labels for each category, not being in the same magnitude. In this experiment, we generated 3480 defect images using the proposed data augmentation method, including 946 images with cracks, 1333 with pinholes, and 1201 with folds. After incorporating the generated samples into the training set, the number of training labels for each category was balanced. To validate the effect, the test set remained unchanged.

Evaluation Metrics: Precision is defined as the proportion of true positive instances among all instances predicted as positive, while Recall is the proportion of true positive instances among all actual positive instances. In the practical application scenario of steering wheel defect detection, both missed and false detections are critical concerns. Therefore, we adopted the mean average precision (mAP) at an IoU threshold of 0.5 (mAP0.5) as the evaluation metric to comprehensively consider both aspects. The detection accuracy of the model is evaluated by calculating the average value of all types of AP (mAP), which is defined in Equation:

$$mAP = \frac{1}{C}\sum_{c=1}^{C} AP(c) \tag{11}$$

where C is the number of target categories.

Training Settings: All experiments were conducted on a system with an AMD EPYC 7402 CPU (24 cores, 96 threads), 503 GB RAM, and NVIDIA 3090 GPUs with 192 GB VRAM. All algorithms were trained using the SGD optimizer for 120 epochs with a batch size of 4. The default traditional augmentation methods refer to the default data augmentation methods provided in the code of each algorithm. The quantitative experimental results are shown in Table 2.

It can be observed from the table that mAP0.5 values of the six defect detection algorithms have increased progressively after using the data expansion method of integrated diffusion model and image fusion proposed in this paper. This indicates that augmenting the training samples using the traditional enhancement and the proposed method to balance the number of training labels for each defect category effectively improves the detection performance of various defect detection algorithms on steering wheel defects.

TABLE II. Quantitative Experimental Results

	YOLO5	**Fasterrcnn**
base	0.311	0.523
traditional enhancement	0.491	0.539
data expansion	0.557	0.552
	DINO	**YOLO7**
base	0.612	0.083
traditional enhancement	0.629	0.220
data expansion	0.670	0.303
	Retinanet	**SSD**
base	0.345	0.218
traditional enhancement	0.470	0.327
data expansion	0.494	0.470

B. Single-Category Ablation Experiments

Experimental Design: To investigate the impact of the generated samples obtained through our proposed method on

different defect categories across various models, we designed single-category ablation experiments for different algorithms. Given that YOLOV7, RetinaNet, and SSD had relatively lower mAP0.5 values in the comparative experiments, we selected the more robust models YOLOV5, FasterRCNN, and DINO for these ablation studies. For each of the three algorithms, we incorporated generated samples of a single defect category into the training set while keeping the test set unchanged, and observed the performance of the models on each defect category after the addition of single-category defect samples. The dataset and experimental settings remained consistent with the comparative experiments. The quantitative experimental results are shown in Table 3, Table 4 and Table 5, and the corresponding bar charts are shown in Fig. 5, 6, and 7.

TABLE III. YOLOV5 QUANTITATIVE RESULTS

	smudges	**cracks**	**pinholes**	**folds**	**All**
base	0.409	0.297	0.051	0.485	0.311
traditional enhancement	0.506	0.557	0.210	0.691	0.491
add pinholes	0.534	0.574	0.249	0.714	0.518
add cracks	0.492	0.632	0.339	0.663	0.532
add folds	0.540	0.580	0.145	0.720	0.496

TABLE IV. FASTERRCNN QUANTITATIVE RESULTS

	smudges	**cracks**	**pinholes**	**folds**	**All**
base	0.576	0.569	0.156	0.794	0.523
traditional enhancement	0.586	0.640	0.143	0.790	0.539
add pinholes	0.603	0.665	0.293	0.749	0.578
add cracks	0.589	0.641	0.205	0.754	0.547
add folds	0.615	0.669	0.083	0.803	0.543

TABLE V. DINO QUANTITATIVE RESULTS

	smudges	**cracks**	**pinholes**	**folds**	**All**
base	0.748	0.646	0.216	0.839	0.612
traditional enhancement	0.716	0.660	0.306	0.832	0.629
add pinholes	0.733	0.798	0.309	0.792	0.658
add cracks	0.740	0.780	0.322	0.812	0.664
add folds	0.742	0.733	0.302	0.837	0.654

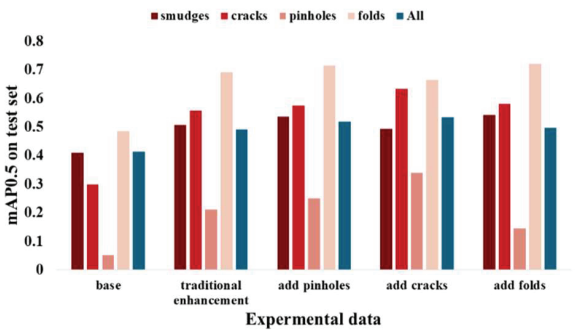

Fig. 5. Histogram of quantitative experiment results of YOLOV5

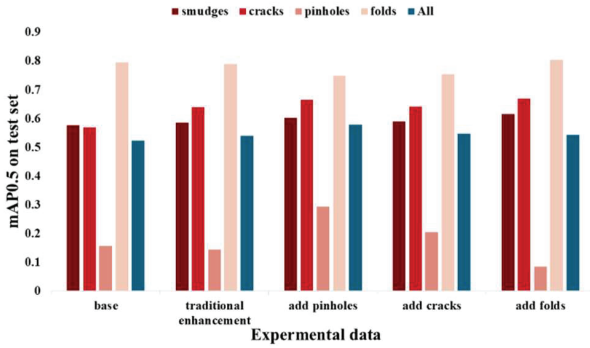

Fig. 6. Histogram of quantitative experiment results of FasterRCNN

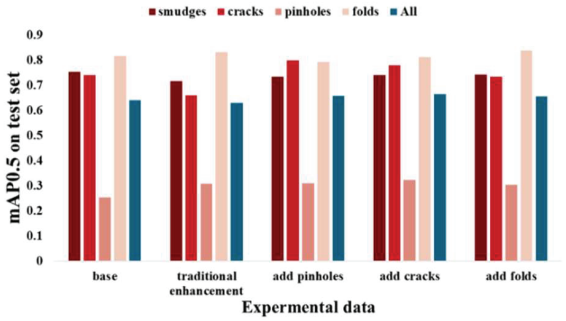

Fig. 7. Histogram of quantitative experiment results of DINO

From the result tables, the following observations can be made:

i) For all three defect detection algorithms, adding single-category defect samples to the training set led to an increase in the mAP0.5 value for that specific category. For example, YOLOV5's mAP0.5 for the crack category increased from 0.297 in the original dataset and 0.557 with traditional augmentation to 0.632 after adding generated crack samples. Similarly, the mAP0.5 for each generated defect category improved across the three models. This indicates that incorporating single-category defect samples generated using the proposed data augmentation method enhanced the models' ability to learn features of that specific defect type, thereby improving their generalization performance on that category.

ii) After adding defect samples of a particular category, the detection performance of the three models improved for all defect categories. For instance, the total mAP0.5 of YOLOV5 on the test set increased from 0.311 to 0.532 after adding crack defect samples. This suggests that expanding the dataset with defect samples of a single category enhances the models' ability to learn some fundamental features, thereby improving overall detection performance.

iii) YOLOV5 and FasterRCNN exhibited significant improvements in detecting pinhole defects after adding generated pinhole samples, with YOLOV5's performance increasing from 0.051 to 0.249 and FasterRCNN's from 0.156 to 0.293. Conversely, DINO showed a smaller improvement, from 0.216 to 0.309. For minor defects, CNN-based local feature extraction methods are more sensitive to the increased sample size compared to transformer-based global feature extraction methods, hence the more substantial improvements for CNN-based defect detection algorithms.

iv) Adding defect samples of a particular category also improved the detection performance for most other defect categories. For instance, FasterRCNN's performance on pinholes increased from 0.156 to 0.293 after adding pinhole samples, while performance on cracks and smudges also improved from 0.569 to 0.665 and 0.576 to 0.603, respectively. This indicates that the introduction of defect samples of one category enhances the model's learning of some fundamental features, which is beneficial for detecting most other defect categories as well.

IV. Conclusion

In this paper, to address the small number of surface defect samples and the uneven number of samples among different categories, we propose the aggregated data augmentation for defects using enhanced diffusion models and poisson blending. By utilizing the GLIGEN generative model and Poisson equation-based image fusion, our approach not only increases the number of samples for each defect category, enriching the diversity of defect samples and enhancing the model's detection performance, but also balances the sample distribution across categories. Experimental results on the steering wheel defect dataset show that the six defect detection algorithms all exhibited improved performance after data augmentation using our method. Incorporating single-category defect samples into the training set significantly enhanced the detection performance for that category as well as for most other defect categories, validating the potential application of our method in other practical defect detection scenarios.

In future research, we will further refine and optimize our approach, particularly by enhancing the quality and diversity of generated defect samples. We will also explore more comprehensive metrics and methodologies to assess sample diversity, examining in detail how generated samples differ in defect characteristics, morphology, and distribution. Such analyses will help substantiate this method's capability for producing high-quality, diversified samples. Moreover, we plan to investigate computational efficiency and time complexity in greater depth, including extensive evaluations of the method's runtime and computational overhead across datasets of various scales and under different hardware configurations. Through these evaluations, we aim to identify strategies that effectively balance algorithmic complexity and execution efficiency while maintaining the overall enhancement performance.

References

[1] A. Saberironaghi, J. Ren, and M. El-Gindy, "Defect detection methods for industrial products using deep learning techniques: a review," Algorithms, vol. 16, no. 2, pp. 95, 2023.

[2] Y. Kahraman and A. Durmuşoğlu, "Deep learning-based fabric defect detection: A review," Textile Research Journal, vol. 93, no. 5-6, pp. 1485-1503, 2023.

[3] T. M. Khoshgoftaar and K. Gao, "Feature selection with imbalanced data for software defect prediction," in Proc. Int. Conf. Mach. Learn. Appl., Dec. 2009, pp. 235–240.

[4] H. Chen, C. Li, W. Yang, J. Liu, X. An, and Y. Zhao, "Deep balanced cascade forest: A novel fault diagnosis method for data imbalance," ISA Trans., vol. 126, pp. 428–439, Jul. 2022.

[5] L. Li et al., "Transformer fault diagnosis based on hybrid sampling and support vector machines," Electr. Power, vol. 54, no. 12, pp. 150–155, 2021.

[6] A. Abdo, H. Liu, H. Zhang, J. Guo, and Q. Li, "A new model of faults classification in power transformers based on data optimization method," Electr. Power Syst. Res., vol. 200, Nov. 2021, Art. no. 107446.

[7] J. Yang, G. Xie, and Y. Yang, "An improved ensemble fusion autoencoder model for fault diagnosis from imbalanced and incomplete data," Control Eng. Pract., vol. 98, May 2020, Art. no. 104358.

[8] Z. Xing, Y. Liu, Q. Wang, and J. Li, "Research on intelligent diagnostic techniques for rolling bearings based on unbalanced data sets," in Proc. Int. Conf. Phys., Comput. Math., Xiamen, China, Dec. 2022, p. 3034.

[9] H. Liu, J. Zhou, Y. Xu, Y. Zheng, X. Peng, and W. Jiang, "Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks," Neurocomputing, vol. 315, pp. 412–424, Nov. 2018.

[10] M. Zareapoor, P. Shamsolmoali, and J. Yang, "Oversampling adversarial network for class-imbalanced fault diagnosis," Mech. Syst. Signal Process., vol. 149, Feb. 2021, Art. no. 107175.

[11] Y. Lyu, Z. Han, J. Zhong, C. Li, and Z. Liu, “A GAN-based anomaly detection method for isoelectric line in high-speed railway,” in Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC), May 2019, pp. 1–6.

[12] T. Lu, Z. Wang, Y. Shen, X. Shao, and Y. Tang, “DefVAE: A defect detection method for catenary devices based on variational autoencoder,” IEEE Trans. Instrum. Meas., vol. 72, pp. 1-12, 2023, Art. no. 3532912.

[13] R. Ferdousi, C. Yang, M. A. Hossain, et al., “Generative model-driven synthetic training image generation: An approach to cognition in rail defect detection,” arXiv preprint arXiv:2401.00393, 2023.

[14] X. Wang, X. Zhang, Y. Cao, W. Wang, C. Shen and T. Huang, "SegGPT: Towards Segmenting Everything In Context," 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2023, pp. 1130-1140.

[15] Y. Li et al., "GLIGEN: Open-Set Grounded Text-to-Image Generation," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 22511-22521.