DefectDiffusion: A Generative Diffusion Model for Robust Data Augmentation in Industrial Defect Detection

Adnan Md Tayeb, Hope Leticia Nakayiza , Heejae Shin, Seungmin Lee, Jae-Min Lee, Dong-Seong Kim
Networked Systems Laboratory, Department of IT Convergence Engineering,
Kumoh National Institute of Technology, Gumi, South Korea.
(mdtayebadnan, hopeleticia, shinheejae, tmdals6164, ljmpaul, dskim)@kumoh.ac.kr

***Abstract*—The accurate detection of industrial defects is critical for ensuring product quality and minimizing operational inefficiencies. However, deep learning models for defect detection often require large, balanced datasets, which are challenging to obtain in industrial settings due to the rarity and variability of defects. In this study, we propose DefectDiffusion, a novel generative diffusion model designed for robust data augmentation in industrial defect detection tasks. By leveraging the progressive noise reduction process inherent to diffusion models, DefectDiffusion synthesizes high-quality, diverse defect images that closely mimic real-world conditions. Unlike traditional augmentation techniques, our approach selectively augments defective regions while preserving the structural integrity of defect-free areas, ensuring realistic and meaningful data augmentation. Experimental results demonstrate that integrating DefectDiffusion-generated images significantly enhances the performance of state-of-the-art defect detection models, improving both precision and recall.**

***Index Terms*—Generative AI, Stable Diffusion, Few-Shot Learning, Defect generation, Defect detection**

I. Introduction

The rapid advancement of artificial intelligence (AI) in recent years has driven transformative changes across various sectors, with industrial applications benefiting significantly. AI has not only improved the efficiency of existing workflows but has also introduced novel opportunities in automation, quality control, and predictive maintenance. Among the most widely adopted AI approaches in industrial contexts is supervised learning—a versatile framework that supports the development of robust vision models to address complex industrial challenges. Supervised learning has demonstrated exceptional effectiveness in applications such as object detection [5], [11], [24], where accurate identification of objects in images or video streams is essential for tasks like quality inspection, sorting, and assembly line optimization. Additionally, supervised learning techniques have greatly advanced segmentation tasks [15], [23], enabling the division of images into meaningful regions, which is critical for defect detection, part recognition, and spatial mapping.

The widespread adoption of these methodologies in industrial settings is fueled by the growing availability of large-scale annotated datasets, advancements in computational power, and the development of sophisticated deep learning architectures. Consequently, supervised learning models have been deployed to enhance the precision, scalability, and efficiency of industrial systems, representing a significant leap in automation capabilities. However, these models are not without limitations. They require extensive labeled datasets, are susceptible to overfitting, and present challenges in achieving real-time inference within dynamic industrial environments.

To address the challenge of limited data in machine learning and computer vision tasks, researchers have explored a range of innovative strategies. For instance, DeVries and Taylor (2017) proposed augmenting datasets by artificially introducing defects or artifacts into pristine images. This method aimed to replicate imperfections that are often rare or difficult to capture in real-world scenarios. Similarly, Li et al. (2021) introduced advanced techniques for generating defective images, including cutting and pasting patches from defect-free images to simulate synthetic defects or transferring defect regions between images. However, while these methods provided valuable insights, the generated images often lacked realism and diversity, limiting their utility in training robust models.

The advent of generative AI, particularly deep learning-based generative models, has introduced transformative advancements in this domain. These models are capable of producing highly realistic, diverse, and contextually accurate images, thereby addressing the limitations of earlier methods. Recent progress in text-to-image generative models has been especially noteworthy, showcasing the ability to generate photorealistic images from textual descriptions. These models excel in capturing intricate details and producing images with varied features, perspectives, and attributes, thanks to large-scale training on extensive datasets covering diverse visual scenarios.

To further enhance image generation capabilities, researchers have investigated techniques that incorporate key features from reference images into diffusion models. This approach aims to improve the precision and contextual relevance of generated content, especially when working with limited training data. Notable contributions in this area include studies by Gal et al. (2022), Ruiz et al. (2023), Han et al. (2023), Chen et al. (2023a), and Chen et al. (2023b). For example, Chen et al. (2023a) introduced methods that enable subject-specific control, ensuring that generated images retain essential features of the reference while adapting to diverse scenarios. These advancements highlight the potential of generative models to overcome data limitations and produce realistic, contextually accurate outputs for a wide range of

applications.

To further enhance image generation, researchers have explored various techniques to integrate key elements from reference images into diffusion models, thereby improving content precision when working with a limited set of images [1], [2], [4], [6], [21], [25], [26]. Chen et al. [1] pioneered a complete parameter adaptation approach, which involves modifying the entire diffusion model to better align with the reference images. Han et al. [6] proposed a method that uses SVG decomposition with a small set of trainable parameters to prevent catastrophic forgetting in scenarios with few reference images. This approach helps the model align more accurately with reference images while reducing the risk of overfitting. Additionally, Chen and colleagues [2] developed an image-conditioned adapter that retains essential characteristics from the reference images without requiring optimization of the network parameters.

However, these approaches still require a substantial number of images to effectively train the model for generating new images [13], [19], [27], [29]. In industrial settings, for instance, developing a large dataset of defect images is both challenging and costly. To address this issue, some studies [3], [10], [12], [16] have explored zero-shot and few-shot techniques for defect generation, but these methods still fall short of addressing the data insufficiency problem in real-world scenarios. Wang et al. (2020), Zhao, Cong, and Carin (2020), and Robb et al. (2020) have investigated few-shot image generation by leveraging pretrained models that can adapt from large domains to smaller ones. However, these approaches primarily focus on transferring entire images rather than emphasizing critical defect regions. Addressing the unique distribution of defects and defect-free areas individually could enhance defect generation methods.

In line with this, our study introduces DefectDiffusion, a novel few-shot defect image generation method that produces new defect images using only a small number of existing defect samples. DefectDiffusion operates in two primary stages: training on defect-free images and generating defect images.

In the first stage, a Stable Diffusion model is trained to generate a diverse collection of high-quality, defect-free images. This model serves as the backbone for generating intermediary images that retain high visual fidelity and intricate detail. In the second stage, defects are extracted from a limited set of defect images using binary masks to isolate the defective regions. These defects are then transformed—adjusting their size, shape, and position—to introduce variability. Finally, these transformed defects are seamlessly blended into the defect-free images using an advanced blending technique, resulting in realistic and visually coherent defect images.

This approach enhances defect analysis by expanding the availability of diverse and realistic defect images with minimal data. We evaluated DefectDiffusion on our Steel Surface defect dataset, and the results highlighted its effectiveness. The augmented dataset generated by DefectDiffusion achieved an 11.9% improvement in Mean Average Precision (mAP) for steel surface defect detection, demonstrating a significant enhancement in detection performance enabled by our methodology.

To summarize, our contributions are as follows:

- We propose a novel model, DefectDiffusion, capable of generating annotated defect images, which can be utilized for robust data augmentation.
- We demonstrate the advantages of the proposed method in industrial applications by conducting experiments on a steel surface dataset. The results highlight the improved performance in defect detection.

II. Preliminary

A. Diffusion Models

Generative diffusion models (DDMs) [8], [9], [17], [18], [22], [28] have emerged as powerful probabilistic frameworks for synthesizing data by progressively reversing the effects of Gaussian noise applied to an initial sample. These models operate through two primary phases: a forward diffusion process, where noise is incrementally added to an original data sample x_0, and a reverse diffusion process, which reconstructs the data by iteratively denoising. The forward process follows a Markov Chain of length T, transitioning the clean sample into a near-isotropic Gaussian distribution as T increases. The model is trained to invert this process and recover the original data distribution.

Given a sample x_0 from the data distribution $q(x)$, the forward diffusion process adds noise step-by-step, controlled by a predefined variance schedule $\{\beta_t \in (0,1)\}_{t=1}^{T}$. The transition probability at each step is represented as:

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}),$$

and the complete process is given by:

$$q(\mathbf{x}_{1:T} \mid \mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t \mid \mathbf{x}_{t-1}).$$

After sufficiently many steps, the distribution of x_T converges to a standard Gaussian.

In the reverse process, the model attempts to recover the original data by sequentially removing the noise, following the probability distribution:

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T) \prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t),$$

where the transition probabilities are parameterized as:

$$p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)).$$

For simplicity, the covariance $\boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)$ is often modeled as a diagonal matrix $\sigma_t^2 \mathbf{I}$.

Rather than directly predicting the clean data, the model typically learns to estimate the noise ϵ_t added during the forward process. The original sample x_0 is reconstructed by subtracting the predicted noise from the noisy input at each step. The training objective is defined as the minimization of

the residual error between the true noise and the predicted noise:

$$L_t = \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}_t}\left[\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t, t)\|^2\right],$$

where $\bar{\alpha}_t$ represents a cumulative product of variance terms. This objective ensures the model learns to denoise effectively at each step, enabling robust sample generation.

B. Stable Diffusion

Latent Diffusion Models (LDMs) [20] are a popular variant of diffusion models (DDMs) that perform operations in a latent space rather than the pixel space, significantly reducing training time and inference costs. In an LDM, an encoder $E(\cdot)$ is used to compress an input image into a lower-dimensional latent representation z, where the diffusion and denoising processes take place. A decoder $D(\cdot)$ subsequently reconstructs the image from this latent representation, yielding $\tilde{x} = D(z)$.

One prominent example of an LDM is Stable Diffusion, which utilizes cross-attention layers to incorporate various conditioning inputs, such as text. The training process for Stable Diffusion leverages two main regularization strategies:

- **KL-regularization (KL-reg):** Aligns the learned latent representation with a standard normal distribution.
- **VQ-regularization (VQ-reg):** Integrates a vector quantization layer within the decoder, similar to VQGAN but with the quantization layer embedded in the decoder.

Stable Diffusion, trained on a large-scale dataset of natural images, has demonstrated remarkable performance across diverse tasks. Its pre-trained weights are publicly available, enabling its adoption as a foundational model for a wide array of downstream applications.

III. Method

Given input defect free image $x_i \in \mathbb{R}^{W\times H\times C}$, a few number of defect images $x_d \in \mathbb{R}^{W\times H\times C}$ along with a binary mask $x_m \in \{0,1\}^{W\times H}$ where a value of 1 identifies the defective regions. The main goal is to generate new defect images $\hat{y} \in \mathbb{R}^{W\times H\times C}$. In this context, we refer to the input x_i as the defect free images, x_d as the defective image, x_g is the intermediate generated image by Stable diffusion and $\hat{y}$ as the generated image. DefectDiffusion consists of two key steps: (i) training Stable Diffusion on defect-free images, and (ii) generating defect images. The overall process of DefectDiffusion is illustrated in Figure 1 and can be defined mathematically as below:

$$\hat{y} = (x_m \cdot x_d) + (x_g \cdot (1 - x_m)) \tag{1}$$

A. Training on Defect-free Images

In the initial phase of our methodology, we focus on training a Stable Diffusion model as the foundational backbone for DefectDiffusion. The primary goal in this stage is to harness the exceptional generative capabilities of Stable Diffusion to create a diverse array of high-quality defect-free images. These images will later serve as critical inputs for the subsequent tasks of defect generation and inpainting. Stable Diffusion has established itself as a state-of-the-art framework for producing photorealistic images with remarkable detail and consistency across a wide range of applications. By embedding this powerful model into DefectDiffusion, we aim to ensure that the generated outputs not only achieve superior fidelity but also replicate the nuanced textures and structural coherence necessary for real-world scenarios.

Given a clean, defect-free input image $x_i \in \mathbb{R}^{W\times H\times C}$, the Stable Diffusion model processes this input to synthesize a corresponding output image $x_g \in \mathbb{R}^{W\times H\times C}$. This output x_g represents a pivotal intermediate result within the DefectDiffusion pipeline, laying the groundwork for precise defect modeling and robust image augmentation in subsequent phases.

B. Generation of Defects

In the defect generation training phase, we utilize a diffusion model to synthesize realistic defect patterns while preserving the quality of the background. This process involves several key steps designed to iteratively train the model using annotated data, ensuring precise defect generation and maintaining a realistic background. The methodology can be broken down as follows:

1) Forward Diffusion Step: Initially, noise is systematically added to the entire image, following the principles of the forward diffusion process in a diffusion model. This step is critical for introducing stochastic variations in the image, which forms the basis for learning to reconstruct and generate specific patterns in subsequent steps.

2) Annotation-Based Mask Creation: From the corresponding annotations, binary masks are generated. These masks distinguish the foreground (defect region) from the background. To introduce variation and increase robustness, several augmented masks are created from the original binary mask:

- **Foreground Preservation:** The mask ensures that the defect region remains intact.
- **Size Variability:** The masks are resized to create slightly larger or smaller versions of the defect region, simulating natural variations in defect size.

3) Foreground Isolation with Noise: A mask is randomly selected from the pool of generated masks. This mask is then multiplied with the noisy image, isolating the foreground region (defect area) by setting the background pixels to zero. This ensures that only the defect region, with its corresponding noise, is visible to the model.

4) Model Training with Foreground Focus: The resulting image, containing only the noisy foreground region, is fed into the diffusion model. The model learns to reconstruct and generate the specific defect patterns within the masked foreground area. This step is crucial for ensuring the model's ability to focus on defect regions during the training process.

5) Background Preservation Using Inverse Mask: To maintain the integrity of the background, the generated image is multiplied with the inverse of the mask. This operation ensures

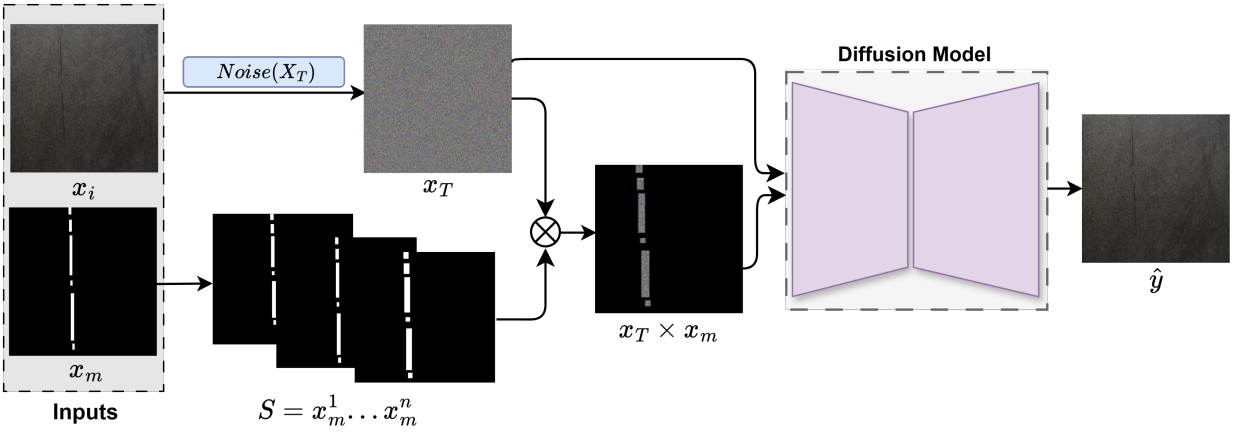

Fig. 1. Overview of the proposed DefectDiffusion

that the background remains unaffected and precisely matches the original image, preventing any loss of detail or introduction of artifacts.

6) Integration of Foreground and Background: Finally, the defect region generated by the model is integrated with the preserved background to form the complete image. This composite image is then used for further training and evaluation, ensuring that both the defect and the background are accurately represented.

This approach ensures that the diffusion model learns to generate defects with high precision while maintaining the original quality of the background. The use of augmented masks and the integration of inverse masks enhance the robustness of the model, allowing it to generalize well to diverse scenarios. By carefully balancing defect generation and background preservation, this methodology provides a reliable framework for defect simulation and analysis, which is particularly valuable in scenarios where annotated datasets are limited.

IV. Experiment

To verify the effectiveness of DefectDiffusion, we conduct experiment on our Steel Surface dataset to extend the dataset and tested it for defect detection task.

A. Dataset

We conducted our experiments using a custom Steel Surface dataset containing 400 samples of pristine, defect-free steel images. The dataset also includes two types of defects: wrinkles and nozzles, with 20 samples representing wrinkles and 25 samples showcasing nozzle defects. All images were captured using a high-resolution imaging setup to ensure clarity and detail.

To prepare the dataset, defects were manually annotated, and corresponding binary masks were generated to facilitate precise defect localization. Both the images and masks were resized to dimensions of 600x600 pixels to standardize input dimensions. Initially, the Stable Diffusion model was trained exclusively on the defect-free samples to establish a foundational model capable of replicating the unique attributes of defect-free steel surfaces. Following this baseline training, we synthesized 3,000 defect-containing images, comprising 1,500 samples for each defect type. These synthetic images were partitioned into training (80%), testing (10%), and validation (10%) subsets. A summary of the dataset's composition, including its usage across the experimental pipeline, is presented in Table I.

B. Implementation Details

Our network implementation and experiments were conducted using Python 3.8.10 and PyTorch 1.13.1 on a high-performance computing setup comprising four NVIDIA GeForce RTX 3090 GPUs, each with 24GB of dedicated memory. For optimization, we employed the AdamW optimizer [14] with a learning rate of 1×10^{-5}, maintaining the default beta values of 0.9 and 0.999. The input image resolution was standardized to 255×255, and the training batch size was set to 16 by default to ensure effective utilization of system resources and stable model convergence.

TABLE I
Overview of the Steel Surface dataset used in this experiment

Defect category	Train	Test	Validation
Wrinkles	1,200	150	150
Nozzles	1,200	150	150

C. Result Analysis

The primary goal of this study is to produce high-quality synthetic images tailored for industrial applications. Considering the high cost and effort associated with manual data annotation, it is essential that the generated images include precise annotations of defect locations. As demonstrated in Figures **??** and 2, our approach effectively generates realistic and visually convincing images. The visual characteristics of the generated images align closely with the geometries

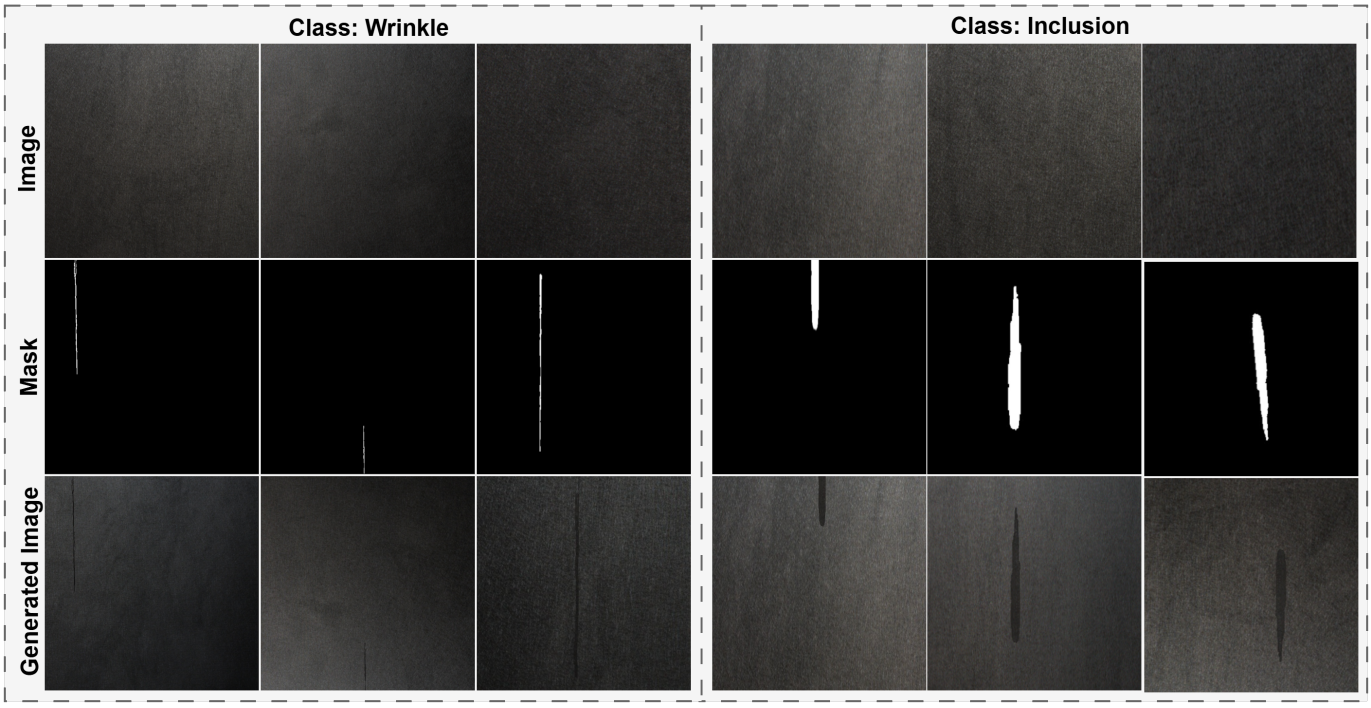

Fig. 2. Examples of real defect images and synthetic images generated by DefectDiffusion for defects in a steel surface dataset.

provided in the input guides. Moreover, the defects visible in these synthetic images accurately correspond to the defect patterns indicated in the accompanying masks.

TABLE II
AVERAGE RESULTS FOR DEFECT DETECTION ON STEEL SURFACES, SPECIFICALLY FOR WRINKLE AND NOZZLE DEFECTS.

Used data	mAP	Precision	Recall
Real only	68.54	78.12	82.71
Synthetic only	92.32	90.25	95.7
Real + Synthetic	85.45	83.89	82.35

To assess the authenticity of the generated data, we employ the Frechet Inception Distance (FID) [7], which compares the distribution of synthetic images to the Steel surface dataset. This evaluation yields an FID score of 99.57, indicating the realism of the generated images. Further, we explore the practical applicability of these annotated synthetic images in industrial settings by leveraging a well-established object detection model, YoLo-NAS. The model was trained under four distinct scenarios: i) training exclusively on real data; ii) training on real data with basic augmentation techniques; iii) training solely on synthetic data; and iv) pre-training on synthetic data followed by fine-tuning on real data.

The evaluation on a separate test dataset reveals that incorporating synthetic data into the training pipeline consistently enhances defect detection across all metrics. The quantitative results, presented in Table II, demonstrate that combining real and synthetic data leads to a significant improvement in detection performance compared to using either real or synthetic data alone. Notably, the mean average precision (mAP) increased by approximately 10% when both real and synthetic data were utilized for training, underscoring the value of synthetic images in boosting model performance.

The qualitative results are illustrated in Figures 2. Figure 2 presents a comparison between real and synthetic images for the wrinkle defect category. In this figure, the top row displays real images, while the bottom row shows synthetic images generated for wrinkles. Upon comparison, it is evident that the synthetic images closely resemble the real ones. Although there are some discrepancies, such as deeper black regions in the background of the synthetic images, the wrinkles themselves are quite similar in appearance to those in the real images. The generated wrinkles vary in size and shape, reflecting a diverse range of defect characteristics that are present in the real dataset.

V. CONCLUSION

In this paper, we have presented DefectDiffusion, an innovative technique aimed at augmenting steel surface defect datasets to enhance the performance of defect analysis. Our method utilizes an image-to-image diffusion model, specifically Stable Diffusion, to generate highly realistic defect-free images. These images form the basis for a blending technique that seamlessly incorporates defects, resulting in an enriched dataset of defect images. The application of DefectDiffusion to our steel surface dataset has yielded a substantial improvement in performance, with defect analysis accuracy increasing by approximately 20%.

Future work will focus on expanding the scope of Defect-Diffusion by applying it to diverse datasets and evaluating its effectiveness across multiple domains, thereby broadening its applicability and impact.

ACKNOWLEDGMENT

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the IITP grant funded by the Korea government(MSIT) (IITP-2025-RS-2020-II201612, 33%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 33%) and by the MSIT, Korea, under the ITRC support program(IITP-2025-RS-2024-00438430, 34%)

REFERENCES

[1] H. Chen, Y. Zhang, X. Wang, X. Duan, Y. Zhou, and W. Zhu. Disenbooth: Disentangled parameter-efficient tuning for subject-driven text-to-image generation. *arXiv preprint arXiv:2305.03374*, 2023.

[2] W. Chen, H. Hu, Y. Li, N. Rui, X. Jia, M.-W. Chang, and W. W. Cohen. Subject-driven text-to-image generation via apprenticeship learning. *arXiv preprint arXiv:2304.00186*, 2023.

[3] Y. Duan, Y. Hong, L. Niu, and L. Zhang. Few-shot defect image generation via defect-aware feature manipulation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 571–578, 2023.

[4] R. Gal, Y. Alaluf, Y. Atzmon, O. Patashnik, A. H. Bermano, G. Chechik, and D. Cohen-Or. An image is worth one word: Personalizing text-to-image generation using textual inversion. *arXiv preprint arXiv:2208.01618*, 2022.

[5] M. Golam, A. Riztia, P. Daely, A. Khatun, J.-H. Kim, J.-H. Lee, Y.-R. Cho, D.-S. Kim, and J. M. Lee. Development of the system architecture for black ice detection to prevent transportation calamities. In *Proceedings of the Symposium of the Korean Institute of Communications and Information Sciences*, pages 692–693, 2022.

[6] L. Han, Y. Li, H. Zhang, P. Milanfar, D. Metaxas, and F. Yang. Svdiff: Compact parameter space for diffusion fine-tuning. *arXiv preprint arXiv:2303.11305*, 2023.

[7] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[8] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020.

[9] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020.

[10] G. Huang, I. Laradji, D. Vazquez, S. Lacoste-Julien, and P. Rodriguez. A survey of self-supervised and few-shot object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4071–4089, 2022.

[11] M. A. Khatun, D.-S. Kim, J. M. Lee, and J.-H. Kim. Comparison between vision transformer and cnn for ice image classification. In *Proceedings of the Symposium of the Korean Institute of Communications and Information Sciences*, pages 1822–1823, 2023.

[12] E. Levin and O. Fried. Differential diffusion: Giving each pixel its strength. *arXiv preprint arXiv:2306.00950*, 2023.

[13] Hanxi Li, Zhengxun Zhang, Hao Chen, Lin Wu, Bo Li, Deyin Liu, and Mingwen Wang. A novel approach to industrial defect generation through blended latent diffusion model with online adaptation. *arXiv preprint arXiv:2402.19330v2*, Mar 2024. These authors contributed equally to this work.

[14] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2017.

[15] Adnan Md Tayeb and Tae-Hyong Kim. Unestformer: Enhancing decoders and skip connections with nested transformers for medical image segmentation. *IEEE Access*, 12:190996–191009, 2024.

[16] C. Mou, X. Wang, J. Song, Y. Shan, and J. Zhang. Dragondiffusion: Enabling drag-style manipulation on diffusion models. *arXiv preprint arXiv:2307.02421*, 2023.

[17] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, I. Sutskever, and M. Chen. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*, 2021.

[18] A. Q. Nichol and P. Dhariwal. Improved denoising diffusion probabilistic models. In *Proceedings of the International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021.

[19] S. Niu, B. Li, X. Wang, and H. Lin. Defect image sample generation with gan for improving defect recognition. *IEEE Transactions on Automation Science and Engineering*, 17(3):1611–1622, 2020.

[20] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.

[21] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22500–22510, 2023.

[22] J. Song, C. Meng, and S. Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

[23] A. M. Tayeb, K.-H. Lee, J.-Y. Seo, G.-W. Kim, H.-B. Park, H.-J. Lee, and T.-H. Kim. Facial skin condition analysis based on pore and wrinkle segmentation. In *Proceedings of the Symposium of the Korean Institute of Communications and Information Sciences*, pages 1648–1649, 2023.

[24] Adnan Tayeb, Md Alam, Ayesha Khatun, Golam Mohtasin, Md Facklasur Rahaman, Md Javed Ahmed Shanto, and Dong-Seong Kim. Pureparking: A decentralized, secure framework for parking space sharing using blockchain. 06 2024.

[25] Adnan Md Tayeb, Mst Ayesha Khatun, Mohtasin Golam, Md Facklasur Rahaman, Ali Aouto, Oroceo Paul Angelo, Minseon Lee, Dong-Seong Kim, Jae-Min Lee, and Jung-Hyeon Kim. Smartrsd: An intelligent multimodal approach to real-time road surface detection for safe driving, 2024.

[26] Adnan Md Tayeb, Hope Leticia Nakayiza, Heejae Shin, Seungmin Lee, Chaesoo Lee, YeongHun Lee, Dong-Seong Kim, and Jae-Min Lee. DefectGen: Few-Shot Defect Image Generation Using Stable Diffusion for Steel Surface Analysis. In *2024 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 2092–2097. IEEE, 2024.

[27] J. Wei, F. Shen, C. Lv, Z. Zhang, F. Zhang, and H. Yang. Diversified and multi-class controllable industrial defect synthesis for data augmentation and transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 4444–4452, 2023.

[28] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M.-H. Yang. Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys*, 56(4):1–39, 2023.

[29] G. Zhang, K. Cui, T.-Y. Hung, and S. Lu. Defect-gan: High-fidelity defect synthesis for automated defect inspection. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2523–2533, 2021.