Accelerating Manufacturing Prototyping: A Continual Learning Approach for Imbalanced Sequential Image Generation

Jisu Shin[1], Sol Lee[1], Youngbin You[2], Jeongheon Park[2], Jungsoo Oh[2], Mun Yi[1]

[1]Korea Advanced Institute of Science and Technology (KAIST)
[2]Hankook Tire & Technology Co.,Ltd, Republic of Korea
{jisu3389, leesol4553, munyi}@kaist.ac.kr,
{ybyou11, jeongheon, jungsoo}@hankooktech.com

Abstract

Prototyping is fundamental in manufacturing to evaluate a new product before mass production; however, physical prototyping is time consuming, costly, and environmentally harmful. Diffusion models, which have achieved state-of-the-art performance in image generation, became a promising alternative. In the manufacturing industry, where new products are continuously developed, continual learning is essential. However, diffusion models suffer from catastrophic forgetting in continual learning scenarios, particularly when dealing with the imbalanced data which is common in real-world manufacturing environments. To address this challenge, we propose Catastrophic Forgetting Mitigation Regularization of Diffusion (CFMRD), a novel continual learning method that combines data-level adjustments with model prior distribution regularization. Using real tire manufacturing data, our approach reduces the forgetting rate measured via Mean Absolute Percentage Error (MAPE) to 0.1142%, outperforming baseline methods. It also achieves superior performance across all metrics of Average Final Quality (AFQ), closely approaching the performance of a model trained on all tasks simultaneously. Our study findings show that diffusion-based modeling is a practically viable approach for prototyping, contributing to the reduction of the product development cycle and environmental waste.

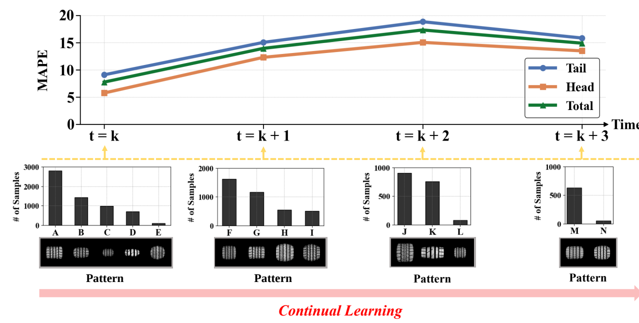

Figure 1: **Performance Degradation After Learning New Patterns.** This figure illustrates a common scenario in manufacturing, where model train with sequential tasks from imbalanced datasets. The yellow dashed lines represent the addition of new data during training. As the model learns new data, it suffers from *catastrophic forgetting*, leading to quality degradation for both head and tail labels. After learning Task 4 (final task), the model's Mean Absolute Percentage Error (MAPE) for tail labels increases to 16% which shows the inherent challenges of continual learning in such settings.

1 Introduction

Prototyping is one of the most important phases in product development within the manufacturing industry. It is an essential step for evaluating design specifications, functionality, and performance before mass production(Camburn et al. 2017). For example, in the tire manufacturing industry, companies must produce physical tire prototypes to measure tire footprints, which is essential for evaluating performance and safety. However, this traditional method is time-consuming, incurs high costs, and impacts the environment due to material consumption and waste during manufacturing and disposal processes. With the advancement of image generation models, especially diffusion models(Ho, Jain, and Abbeel 2020) demonstrating superior performance(Dhariwal and Nichol 2021), it has become possible to replace physical prototyping with virtual simulations.

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Nevertheless, these technologies are not widely used for prototyping, primarily due to the unique characteristics of the manufacturing data. It often involves the continual introduction of new categories, irregular dataset sizes, and significant label imbalance(Figure 1). For example, in the tire industry, some patterns developed in 2024 may have as many as 3,000 data samples, while specialized patterns, such as those for motorcycle tires, may have fewer than 10. Furthermore, the number of patterns developed each year is not constant, but tends to decrease over time.

These industrial characteristics worsen *catastrophic forgetting*, where models lose previously learned knowledge when updated with new data(Chrysakis and Moens 2020). In other words, updating the model for new tire patterns can degrade the performance of previously learned patterns. However, retraining the model with all previous data is unrealistic due to time and computational costs. For prototyping modeling, continuously developing and updating the model for new categories is more critical than for other tasks, making it even more challenging to apply the diffusion model to advanced tasks like prototyping.

Recent research actively addresses the catastrophic forgetting problem in image generation models, including diffusions. However, directly applying traditional continual learning methods from classification does not yield significant performance improvements(Zajac et al. 2023; Zhang et al. 2024). Although replay-based methods are effective in classification domains(Lopez-Paz and Ranzato 2017a), they cause data imbalance issues in diffusion models which induce mode collapse. Attempts to balance labels during memory sampling in classification tasks(Kim, Jeong, and Kim 2020), which is direct data-level adjustments, do not fundamentally solve issues such as forgetting tail patterns in the diffusion models because diffusion models learn more complex data distributions(Qin et al. 2023).

To overcome these challenges, we propose a novel continual learning method for diffusion models called *Catastrophic Forgetting Mitigation Regularization of Diffusion (CFMRD)*. This method combines data-level adjustment techniques like experience replay with soft distribution adjustment at the loss function level. Inspired by Class-balancing diffusion(Qin et al. 2023) which adjust imbalance prior distribution to the balance distribution during the sampling process, we designed an additional loss function that acts as a regularizer. This regularizer prevents the model from deviating from the distribution of previous task. It expands the learning distribution towards the tail classes of previous task rather than being biased towards the head classes of currrent task. As a result, we address the problem of high forgetting rates in low-frequency labels, ultimately achieving superior performance across all labels.

In this study, using actual tire industry data, we prove that our method, CFMRD, solves the catastrophic forgetting problem even in realistic scenarios with increasingly unbalanced label distribution. Our research is the first to identify the continual learning challenges of diffusion models on unbalanced datasets and to propose a new method that simultaneously performs data-level adjustments and model prior distribution adjustments.

By enabling diffusion models to be effectively used for prototyping, our method significantly shortens new product development time and reduces environmental burdens. Furthermore, our method contributes to reducing the time and cost required not only for prototyping but also for retraining models. We expect this research play a significant role in solving catastrophic forgetting problems of diffusions across various industrial fields in the future. Our contributions are as follows:

- We are the first to identify and address the continual learning challenges in diffusion models caused by the continuous emergence of new categories and unbalanced data in manufacturing.
- We propose CFMRD, a novel methodology that combines data-level adjustments with model prior distribution adjustments to minimize catastrophic forgetting, even when learning from unbalanced data.
- By validating our model's effectiveness with real manufacturing data, we demonstrate that diffusion models can be effectively applied to core tasks like prototyping, contributing to cost savings and environmental benefits.

2 Related Works

2.1 Conditional Image Generation

Generative Adversarial Networks (GANs)(Goodfellow et al. 2014) and diffusion models (Ho, Jain, and Abbeel 2020) are foundational methods in image generation. Recently, diffusion models are regarded as powerful tool for generating high-resolution and detailed images(Dhariwal and Nichol 2021). Diffusion models achieve this by progressively denoising images through a two-phase process: forward diffusion, where noise is incrementally added, and reverse diffusion, where noise is removed to restore the original image(Ho, Jain, and Abbeel 2020). Conditional image generation using diffusion models has been researched in many ways(Batzolis et al. 2021; Hung et al. 2023). These models generate images reflecting specific attributes based on input conditions, controlled through two main approaches: predicting noise that matches given attributes(Rombach et al. 2022), and using classifier-free guidance(CFG), which trains both conditional and unconditional models simultaneously to adjust the strength of conditions during image generation(Ho and Salimans 2022).

Conditional image generation can be also categorized based on the type of condition used: label-based and text-based. Label-based methods generate images for specific categories based on class labels, whereas text-based conditional generation visually represents natural language description. Text-based approach has been successfully demonstrated by large transformer-based models, such as OpenAI's Dall-E(Ramesh et al. 2021) and Google's Imagen(Saharia et al. 2022).

These improvements have broadened applications across fields such as healthcare(Khader et al. 2022), autonomous driving(Tang et al. 2019), and art creation(Wang, Chen, and Wang 2024). In contrast, the application of conditional image generation in manufacturing remains limited, due to challenges such as catastrophic forgetting and the complexities of data. Therefore, further exploration of conditional image generation in manufacturing—especially approaches that address product development processes like prototyping—holds substantial potential to advance industries such as smart manufacturing.

2.2 Continual Learning in Image Generation

Continual learning is the technique that enables a model to sequentially learn a series of tasks $(T^{(1)}, T^{(2)}, \ldots, T^{(T)})$ integrating new information without forgetting the knowledge of previous task(Delange et al. 2021). Traditionally, continual learning mainly used in classification tasks, which has limited output spaces such as one-hot class. These methods can be categorized into three approaches: *Regularization-based*(Aljundi et al. 2018), *Replay-based*(Wu et al. 2018; Chaudhry et al. 2019) and *Parameter-isolation-based*(Masana, Tuytelaars, and Van de Weijer 2021).

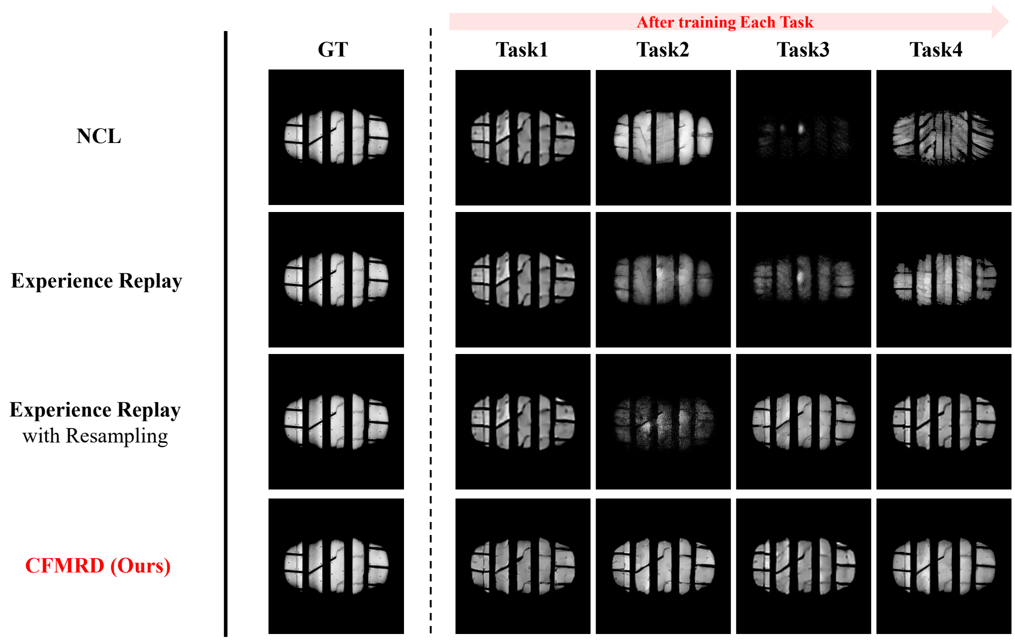

Figure 2: **Qualitative Comparison of Generated Tire Footprint Images** after training each task on continual learning setups. Columns represent the results of generating the same pattern (originally learned in Task 1) after training on subsequent tasks (Task 2, Task 3, and Task 4). Rows compare the proposed **CFMRD (Ours)** approach with baseline methods: **NCL**, **Experience Replay** and **Experience Replay with Resampling**. The proposed method demonstrates the superior performance for mitigating catastrophic forgetting across tasks.

In image generation task, continual learning is learning sequential tasks, where each task is defined by a specific set of conditions. However, due to the complexity of output spaces distribution which is image, traditional methods are not satisfactory enough to address catastrophic forgetting in image generation(Zhang et al. 2024). Specifically, replay-based method, which store samples from previous tasks in a memory buffer or generate them to replay alongside new data, are effective for classification tasks. However, it faces significant challenges in image generation. Replay often induces data imbalance, which leads to the mode collapse; the model over-optimizes for certain classes while neglecting others, resulting in degradation of overall performance(Lopez-Paz and Ranzato 2017b; Srivastava et al. 2017).

Beyond traditional methods, new approaches for conditional diffusion—such as C-LoRA(Smith et al. 2023) and Generative Distillation(Masip et al. 2023)—have been proposed. However, these methods assume datasets with uniform distributions and simple conditions, like text descriptions or categorical labels. Additionally, they rely on parameters from previous models, leading to increased storage and computational costs as models become more complex. These limitations restrict their applicability to real-world scenarios like manufacturing, where conditions are inherently complex and datasets are imbalanced. To address these challenges, this study proposes a novel approach that prevents catastrophic forgetting even in the context of complex and imbalanced dataset distributions.

3 Methods

3.1 Problem Definition

In this study, we address the challenge of continual learning in image generation using tire data as a representative case from the manufacturing industry. The dataset is based on real-world data collected from a global tire manufacturing company. It consists of tabular data with 82 columns of specification (*spec*) information including *pattern name, tire mold specifications* and 3 columns of test conditions (*condition*) such as *test load*.

Task	Total Size	# of Patterns	Max	Min
Task 1	11,500	10	3,219	14
Task 2	5,750	8	1,870	5
Task 3	4,600	6	1,408	16
Task 4	3,450	5	968	10

Table 1: Data Distribution per Task

Based on domain knowledge—*the pattern name column plays a pivotal role in the spec data, as other attributes such as mold are closely correlated with the pattern name*—we define each task $T^{(t)}$ based on different patterns. For example, **Task 1** consists of 15,000 samples distributed across 10 patterns,with the largest pattern containing 3,219 samples and the smallest only 14. Subsequent tasks consist of fewer patterns and samples, as detailed in Table 1. This reflects the real-world data distribution in the manufacturing industry.

Our objective is to develop a model that can effectively learn from this imbalanced dataset without forgetting previously learned patterns. Formally, each task $T^{(t)}$ is defined as:

- **Input space** $C^{(t)}$: Tire specifications and test conditions
- **Output space** $I^{(t)}$: Tire footprint images
- **Training set** $D^{(t)}$: A set of samples $\{(c_j^{(t)}, i_j^{(t)})\}_{j=1}^{|D^{(t)}|}$

The goal is to learn a mapping function f that generalizes across all tasks:

$$f : \bigcup_{t=1}^{T} C^{(t)} \rightarrow \bigcup_{t=1}^{T} I^{(t)} \tag{1}$$

This function f aims to generate tire footprint images (I) from tire specifications and test condition data (C).

3.2 Preliminaries

Stable Diffusion Stable Diffusion(Rombach et al. 2022) is a generative model that operates in a compressed latent space, where an input image x is encoded into a latent representation z_0 using a Variational Autoencoder (VAE)(Kingma 2013). In this latent space, the diffusion process iteratively adds Gaussian noise to z_0 over multiple time steps, resulting in a noisy latent representation z_t at each time step t. This process is defined as:

$$z_t = \sqrt{\bar{\alpha}_t} z_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, 1) \tag{2}$$

where $\bar{\alpha}_t$ controls the noise schedule.

To generate high-quality images, the model employs a denoising network ϵ_θ to predict the noise ϵ in z_t. The denoising process is conditioned on both the time step t and external information y, which guides the generation process. To achieve this, the external condition y is transformed into a feature representation by a condition encoder $\tau_\theta(y)$. The objective function for training is:

$$\mathcal{L}_{LDM} = \mathbb{E}_{\mathcal{E}(x), y, \epsilon \sim \mathcal{N}(0,1), t} \left[\| \epsilon - \epsilon_\theta(z_t, t, \tau_\theta(y)) \|_2^2 \right] \tag{3}$$

This objective reconstructs the latent representation z_0 to generate images aligned with the given condition.

In this study, we extend this model to retain the performance for previous conditions y across tasks. The conditions y are provided as tabular data consisting of tire specifications (*spec*) and test conditions (*condition*).

Class-Balancing Diffusion Models Class-Balancing Diffusion Models (CBDM) (Qin et al. 2023) address the degradation in diffusion model performance in imbalanced dataset. Diffusion models assume that the prior distribution of data is balanced. However, in scenarios where the actual data distribution is imbalanced, this assumption leads to overfitting on head classes and poor performance on tail classes. CBDM softly adjust the prior distribution during the sampling step, modifying the reverse diffusion process $p_\theta(x_{t-1}|x_t, y)$, where x_t represents the noise at time step t and y serves as the conditional label guiding the generation. The adjusted distribution is formulated as:

$$p_\theta^\star(x_{t-1}|x_t, y) = p_\theta(x_{t-1}|x_t, y) \frac{p_\theta(x_{t-1})}{p_\theta^\star(x_{t-1})} \frac{q^\star(x_t)}{q(x_t)} \tag{4}$$

where $q(x_t)$ represents the original data distribution, and $q^\star(x_t)$ is the target balanced distribution. Technically this adjustment is reformulated as a mean squared error (MSE) loss term, making it computationally feasible to integrate into the training process. Our study extends this loss term to induce the diffusion model generates images that are not biased toward the current task. Through MSE loss, we adjust the model to sample images in prior distributions across tasks, enabling it to generate high-quality images not only for the current task but also for the previous tasks.

Experience Replay Experience Replay(Chaudhry et al. 2019) is an approach to address the problem of *catastrophic forgetting* in continual learning. It involves sampling data from previous tasks and storing it in a memory buffer, which is used with new data during training.

3.3 Imbalance Sequential Image Generation

In this study, we propose CFMRD (Catastrophic Forgetting Mitigation Regularization of Diffusion), a novel training method that integrates data and sampling-level approach.It addresses the catastrophic forgetting problem in continual setting with imbalance data, a common challenge in the manufacturing industry.

In continual learning setting, diffusions easily forget the distribution of previous tasks $q_{t-1}(y)$, and becomes biased toward the distribution of the current task $q_t(y)$. To address this, we adjust the prior distribution at the loss function level to an idealized distribution $q^*(y)$, that uniformly incorporates data across tasks $t = 1, 2, \ldots, T$. It is formulated as:

$$q^*(y) = \frac{1}{\sum_{t=1}^{T} |Y^{(t)}|}, \; \forall y \in \bigcup_{t=1}^{T} Y^{(t)} \tag{5}$$

Here, T denotes the total number of tasks, and $Y^{(t)}$ represents the set of labels in task t. By aligning the model's prior distribution to this distribution, CFMRD minimizes forgetting while achieving superior performance across head and tail classes. Inspired by Classifier-Free Guidance (CFG)(Ho and Salimans 2022), we train the model on both real and adjusted distributions, softly guiding the model toward the ideal distribution. This approach stabilizes training process preventing mode collapse.

Training Algorithms This section provides an overview of the training algorithm for the proposed method, CFMRD. At the beginning of each new task k, a fixed-size memory buffer $\mathbf{B}$ is initialized with balanced samples from all tasks $\{1, \ldots, k\}$. The buffer $\mathbf{B}$ is then combined with the current task's data to construct the training dataset. When sampling data into the buffer, oversampling or downsampling is applied depending on the availability of samples for each pattern. During training, the model follows the standard noise prediction framework utilized in the stable diffusion(Rombach et al. 2022), with regularization losses with

Algorithm 1: Training Algorithm for CFMRD

Inputs: Training data $\{(y^{(i)}, x_0^{(i)})\}_{i=1}^{M}$, memory buffer B, current task K, batch size b, number of categories N, regularization coefficients τ, γ, and probability p

1: Initialize memory buffer B with balanced data from tasks $\{1, \ldots, k\}$
2: Combine the memory buffer B with the current task's data to form the training set
3: **for** each batch of size b from the data loader **do**
4:   **for** each sample $(y^{(i)}, x_0^{(i)})$ in the batch **do**
5:     Encode input to latent space using VAE: $z_0^{(i)} = \text{VAE\_encoder}(x_0^{(i)})$
6:     Sample $t \sim \mathcal{U}(\{1, \ldots, T\})$ and $\epsilon^{(i)} \sim \mathcal{N}(0, I)$
7:     Compute the noise of latent representation: $z_t^{(i)} = \sqrt{\bar{\alpha}_t} z_0^{(i)} + \sqrt{1 - \bar{\alpha}_t} \epsilon^{(i)}$
8:     Compute condition embedding $\tau_\theta(y^{(i)})$
9:     Predict noise $\hat{\epsilon}^{(i)} = \epsilon_\theta(z_t^{(i)}, t, \tau_\theta(y^{(i)}))$
10:   **end for**
11:   Compute latent diffusion loss: $\mathcal{L}_{LDM} = \frac{1}{b} \sum_{i=1}^{b} \|\epsilon^{(i)} - \hat{\epsilon}^{(i)}\|^2$
12:   **if** random probability $< p$ **then**
13:     Sample N data from buffer B
14:     **for** each data $(y^{(j)}, z_0^{(j)})$ sampled from buffer **do**
15:       Compute noisy latent and predict noise as in lines 6–8
16:     **end for**
17:     Compute Past-to-Current Regularization losses: $\mathcal{L}_{pc} = \tau \cdot \frac{1}{N} \sum_{i,j} \|\hat{\epsilon}^{(i)} - \text{sg}(\hat{\epsilon}^{(j)})\|^2$
18:     Compute Current-to-Past Regularization losses: $\mathcal{L}_{cp} = \gamma \cdot \tau \cdot \frac{1}{N} \sum_{i,j} \|\text{sg}(\hat{\epsilon}^{(i)}) - \hat{\epsilon}^{(j)}\|^2$
19:     Total loss: $\mathcal{L} = \mathcal{L}_{LDM} + \mathcal{L}_{pc} + \mathcal{L}_{cp}$
20:   **else**
21:     Total loss: $\mathcal{L} = \mathcal{L}_{LDM}$
22:   **end if**
23:   Update with total loss $\mathcal{L}$
24: **end for**

a probability p (e.g., $p = 20\%$). Specifically, an additional N samples is drawn from the memory buffer $\mathbf{B}$, and corresponding regularization terms are computed. For these samples, the model predicts noise $\hat{\epsilon}^{(j)}$, and two types of regularization losses are applied:

- **Past-to-Current Regularization** (L_{pc}): This term ensures that the representations of current task samples (i) do not deviate significantly from the representations of samples from ideal distribution containing previous tasks (j). It induce consistent noise predictions across tasks:

$$L_{pc} = \tau \cdot \frac{1}{N} \sum_{i,j} \left\| \hat{\epsilon}^{(i)} - \text{sg}(\hat{\epsilon}^{(j)}) \right\|^2 \tag{6}$$

  Here, $\text{sg}(\cdot)$ denotes the stop-gradient operation and τ is a scaling factor for the regularization term.

- **Current-to-Past Regularization** (L_{cp}): This term ensures that the representations of past task data (j) align with the representation of data from current task (i). It prevents the model from being overly influenced by past tasks, ensuring a balanced learning process for the current task:

$$L_{cp} = \gamma \cdot \tau \cdot \frac{1}{N} \sum_{i,j} \left\| \text{sg}(\hat{\epsilon}^{(i)}) - \hat{\epsilon}^{(j)} \right\|^2 \tag{7}$$

  Here, γ scales the contribution of this term. L_{cp} provides a mechanism for effectively learning new tasks while preserving past knowledge.

The total loss is computed as:

$$L = \begin{cases} L_{\text{LDM}} + L_{pc} + L_{cp} & \text{if regularized } (p = 20\%) \\ L_{\text{LDM}} & \text{otherwise} \end{cases}$$

where L_{LDM} is the standard noise prediction loss from the latent diffusion framework.

L_{pc} prevents the current task representations from shifting too far from past task representations, while L_{cp} aligns past task representations with the current task. By incorporating these regularization terms together, the model ensures a balance between retaining knowledge from past tasks and adapting to the current task. Therefore this training method effectively mitigating catastrophic forgetting.

4 Experiment setup

4.1 Implementation Details

This study utilizes the Stable Diffusion model from scratch as the backbone where the pretrained-VAE *'sd-vae-ft-mse'* from Stability AI was used to construct the latent space. The model trained and evaluated on an NVIDIA A6000 GPU with a batch size of 64 and a learning rate of 0.0001. Due to the absence of prior research on continual learning for image generation under imbalanced data, we followed a class-incremental learning (Liu et al. 2023): pre-training on Task 1 for 1,300 epochs and fine-tuning on subsequent tasks for 800 epochs. The training process is based on 1,000 diffusion steps and 16 worker threads. To mitigate *catastrophic forgetting*, the strength of the regularization terms were set as $p = 0.2$, $\tau = 0.0005$ and $\gamma = 0.25$. The buffer size B for experience replay was set as 200, which accounts for less than 1% of the total dataset.

4.2 Baselines

Non-Continual Learning (NonCL) train a single model on the data from all tasks together. While this approach is not under a continual learning (CL) setup, it serves as an **upper bound** for CL baselines since there is no forgetting. Naive Continual Learning (NCL) is the simplest baseline in CL, where the same model is continually trained without applying any CL techniques to mitigate forgetting. This approach directly suffers from the catastrophic forgetting problem and it serves as an **lower bound**. Experience Replay (ER) combines samples from previous tasks with the current task data to train the model. We adopt reservoir sampling for managing the memory buffer, following the approach described

Method	Evaluation	MAPE (↓)	LPIPS (↓)	PSNR (↑)
NonCL (upper target)	AFQ	9.3903	0.1345	17.2204
	FR	–	–	–
NCL (lower target)	AFQ	14.2250	0.1836	15.1938
	FR	4.5071	0.0388	-0.3412
Experience Replay	AFQ	13.0452	0.1525	15.4149
	FR	3.8702	0.0222	0.010
Experience Replay with Resampling	AFQ	12.6200	0.1430	15.8298
	FR	2.8414	0.0131	**0.7139**
CFMRD (Loss Only)	AFQ	11.9554	0.1573	14.8846
	FR	1.4803	0.0287	-1.0655
CFMRD (Ours)	AFQ	**9.8499**	**0.1349**	**16.1166**
	FR	**0.1142**	**0.0028**	-0.1187

Table 2: **Quantitative Comparison of Continual Learning Performance** across various methods, based on Average Final Quality (AFQ) and Forgetting Rate (FR). CFMRD outperforms all baselines, achieving the best AFQ and FR values in MAPE_mean and LPIPS, as highlighted in **bold**. These results demonstrate that CFMRD effectively improves the generation quality at the final step while minimizing knowledge forgetting across previous tasks.

in (Riemer et al. 2018) and (Zhang et al. 2024). Experience Replay with Resampling extends the basic ER approach by adding resampling techniques to balance the data distribution in replayed samples.

4.3 Evaluation Metrics

To evaluate the performance of the proposed method, we categories metrics into two types: Image Generation Performance Metrics and Continual Learning Evaluation Metrics.

Image Generation Performance Metrics To evaluate the performance of the proposed model, we utilize mean MAPE of three physical attributes of tire footprint which are essential for assessing tire performance in real-world scenarios. It captures the Mean Absolute Percentage Error (MAPE) between the ground truth and generated tire footprint images across three key physical properties: contact length, width, and area. The MAPE for each property is computed as the mean of the absolute percentage error, defined as MAPE $= \frac{\text{GT}-\text{Generated}}{\text{GT}} \times 100$, where GT represents the ground truth. By quantifying the deviation between the generated and actual footprint attributes, MAPE offers a assessment of the model's capability to accurately reproduce the characteristics of tire footprint.

For evaluating similarity with ground truth (GT), we employ LPIPS (Learned Perceptual Image Patch Similarity)(Zhang et al. 2018)and PSNR (Peak Signal-to-Noise Ratio)(Xu et al. 2022) as evaluation metrics to assess the similarity between generated images and GT. LPIPS is a perceptual similarity metric that measures the distance between latent features extracted from pre-trained neural networks offering a more human-aligned evaluation of visual quality. In our experiment we employ AlexNet(Krizhevsky, Sutskever, and Hinton 2012) as pre-trained neural network. On the other hand, PSNR is a traditional pixel-wise metric that compares pixel intensities, with higher values indicating closer fidelity to the ground truth.

Continual Learning Evaluation Metrics Average Final Quality (AFQ)(Lopez-Paz and Ranzato 2017a) evaluates the model's overall performance on all tasks at the end of the training process. For a model trained on T tasks, AFQ is defined as:

$$AFQ = \frac{1}{N}\sum_{n=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T} m(f^{(T)}, \mathcal{T}_n^{(t)})\right] \tag{8}$$

where $m(f^{(T)}, \mathcal{T}^{(t)})$ represents the model's performance on task t after it has been trained on all T tasks and N represents the total number of patterns across all tasks.

Forgetting Rate (FR)(Chaudhry et al. 2018) quantifies the extent of catastrophic forgetting by calculating the difference in performance on the previous tasks as the model learns new tasks. For a model trained on T tasks, FR for task t is defined as:

$$FR^{(t)} = \frac{1}{N}\sum_{n=1}^{N}\Bigg[\frac{1}{T-t}\sum_{i=t+1}^{T}\Big(m(f^{(i)}, \mathcal{T}_n^{(t)}) - m(f^{(t)}, \mathcal{T}_n^{(t)})\Big)\Bigg] \tag{9}$$

where $m(f^{(i)}, \mathcal{T}^{(t)})$ is the model's performance on task t for pattern n after learning task $i > t$ and $m(f^{(t)}, \mathcal{T}^{(t)})$ is the performance on task t for pattern n directly after training on task t.

5 Experiment Results

5.1 Quantitative Evaluation

Table 2 provides a comparison of the performance of different methods under the continual learning setting.

Catastrophic Forgetting in Diffusion Models Naive Continual Learning (NCL) shows quality degradation when training sequentially on multiple tasks, resulting in poor image generation quality compared to Non-Continual Learning(NonCL). Specifically, NCL achieves an AFQ_{MAPE} of 14.2250 and AFQ_{LPIPS} of 0.1836, whereas NonCL shows superior performance with an AFQ_{MAPE} of 9.3903, AFQ_{LPIPS} of 0.1345, and AFQ_{PSNR} of 15.1938. Furthermore, the forgetting rates for NCL are relatively high, with FR_{MAPE} of 4.5071 and FR_{LPIPS} of 0.0388. These findings confirm that diffusion models suffer from catastrophic forgetting in continual learning setting in real-world manufacturing data.

Performance of CFMRD ER shows a slight improvement over NCL, achieving an AFQ_{MAPE} of 13.0452 and AFQ_{LPIPS} of 0.1525, but it still suffers from high forgetting rates, with an FR_{MAPE} of 3.8702. ER with Resampling performs better than ER, achieving an AFQ_{MAPE} of 12.6200, AFQ_{LPIPS} of 0.1430 and AFQ_{PSNR} of 15.8289. It shows relatively high performance in PSNR; however, the high FR_{MAPE} indicates that it struggles to generate accurate footprint images, despite having a similar number of pixels. Furthermore, it still exhibits quality degradation as indicated by an FR_{MAPE} of 2.8414 and FR_{LPIPS} of 0.0131. When only the regularization term of CFMRD is applied $CFMRD_{\text{loss only}}$, the forgetting rate is further reduced, with an FR_{MAPE} of 1.4803 and achieves an AFQ_{MAPE} of 11.9554. It outperforms both ER and ER with Resampling in MAPE. By incorporating both the regularization and data-level adjustments (CFMRD), our method achieves the best performance across all metrics, with an AFQ_{MAPE} of 9.8499, AFQ_{LPIPS} of 0.1349 and AFQ_{PSNR} of 16.1166. and the lowest forgetting rates as FR_{MAPE} of 0.1142 and FR_{LPIPS} of 0.0028. It proves superior performance, comparable to the NonCL.

Comparison of Forgetting Rates Between Head and Tail Classes It is essential to generate high-quality images for both head and tail patterns in prototyping. For comparison, we define head and tail patterns based on the concept of data imbalance as described in (Zhu, Guo, and Xue 2020). Specifically, patterns with a cumulative data contribution exceeding 50% of the total dataset are classified as head patterns, while the remaining patterns are classified as tail patterns. According to this definition, the top 5 patterns are categorized as head patterns, and the remaining 24 patterns as tail patterns. Figure 3 shows the comparison of forgetting rates, measured as FR_{MAPE}, between head and tail patterns across different methods. Traditional methods such as NCL and ER struggle with catatrophic forgetting, especially for tail patterns. For instance, in the NCL, FR_{MAPE} of head pattern is 3.2623% and tail pattern is 4.8349%. ER shows more imbalance, with FR_{MAPE} of head patterns at 1.4193% and tail patterns at 4.5175%. In contrast, our proposed method CFMRD, consistently reduces FR_{MAPE} across both categories, achieving the lowest forgetting rates. Specifically, FR_{MAPE} for head patterns is reduced to 0.9934%, and for tail patterns, it is further improved to -0.0011%.

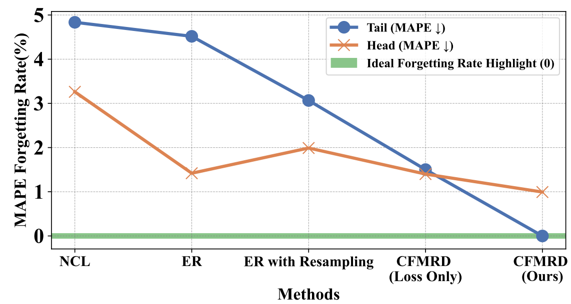

Figure 3: The line graph illustrates the Forgetting Rate via MAPE(%) for head and tail patterns across different methods. The proposed CFMRD method achieves the lowest forgetting rates for both head and tail patterns.

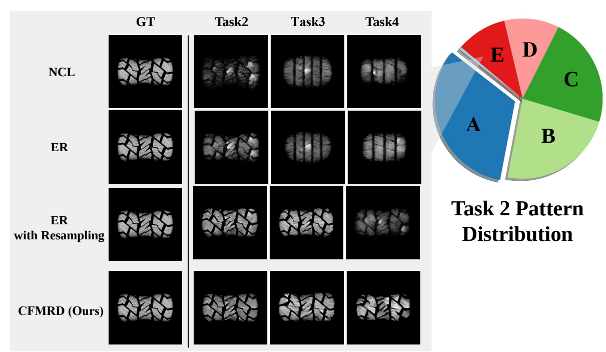

Figure 4: Generated images of the tail pattern (E) in continual learning settings.

5.2 Qualitative Evaluation

Figure 2 illustrates the qualitative performance of each method in continual learning settings. In NCL, the pattern in tire footprint image become progressively blurred. After learning the task 4, it generates entirely different tire footprints. ER demonstrates some improvement compared to NCL but still generates blurry and distorted images as they train learn the task. Similarly, ER with Resampling shows instability, with tire footprint images distorted after certain tasks. In contrast, CFMRD maintains high fidelity with less blurring or distortion. These results confirm that our method effectively mitigates catastrophic forgetting.

5.3 Ablation Study

Complementary of Each Loss Term In our proposed method, we incorporate both Current-to-Past Regularization (L_{cp}) and Past-to-Current Regularization (L_{pc}) to balance learning between past and current tasks. To validate the necessity of these two terms, we conducted an ablation study by removing L_{cp}, while keeping L_{pc}. When L_{cp} was removed, the metrics were as follows: FR_{MAPE} of 1.091%, FR_{LPIPS} of 0.0106 and FR_{PSNR} of 0.0093. These results indicate that the model exhibits relatively low forgetting rates, suggesting that L_{pc} effectively prevents the model from forgetting the distribution of past tasks. However, AFQ were as follows: AFQ_{MAPE} of 13.9193%, AFQ_{LPIPS} of 0.1649 and AFQ_{PSNR} of 15.1694. The relatively high AFQ_{MAPE} and AFQ_{LPIPS} values suggest that, while L_{pc} minimizes forgetting, the model struggles to effectively learn new tasks without L_{cp}.Through this ablation study, we demonstrate the complementary roles of L_{cp} and L_{pc}. While L_{pc} focuses on minimizing forgetting by adjusting the model to the past task distribution, L_{cp} ensures the model to adapt to current tasks.

6 Conclusion & Limitation

This study addresses the challenge of continuous image generation under imbalanced data distributions, a common issue in real-world manufacturing industry where new products are continuously developed. Since diffusion-based models learned complex data distribution, it face significant difficulties in such scenarios with existing method. To address this, we proposed a novel methodology that adjust the distribution of model both at the input level and at the sampling level. Our experiments show that the proposed approach effectively mitigates catastrophic forgetting, particularly for tail classes. As a result, the stable diffusion model generates high-quality images across all patterns in sequential setting. This demonstrates the potential of diffusion-based models for practical applications, such as manufacturing prototyping, which require continual model updates. However, our study is limited to four tasks due to resource constraints. Future work will focus on validating the proposed method on more diverse tasks to enhance the applicability of diffusion models in manufacturing industry.

7 Acknowledgments

This work was supported by Hankook Tire & Technology Co., Ltd, Republic of Korea, and some experiments were conducted using the instruments of Hankook Tire & Technology Co., Ltd. The authors extend their sincere thanks to Hankook Tire & Technology Co., Ltd for their help and support.

References

Aljundi, R.; Babiloni, F.; Elhoseiny, M.; Rohrbach, M.; and Tuytelaars, T. 2018. Memory Aware Synapses: Learning What (Not) to Forget. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 139–154.

Batzolis, G.; Stanczuk, J.; Schönlieb, C.-B.; and Etmann, C. 2021. Conditional image generation with score-based diffusion models. *arXiv preprint arXiv:2111.13606*.

Camburn, B.; Viswanathan, V.; Linsey, J.; Anderson, D.; Jensen, D.; Crawford, R.; and Wood, K. 2017. Design Prototyping Methods: State of the Art in Strategies, Techniques, and Guidelines. *Design Science*, 3: e13.

Chaudhry, A.; Dokania, P. K.; Ajanthan, T.; and Torr, P. H. 2018. Riemannian Walk for Incremental Learning: Understanding Forgetting and Intransigence. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

Chaudhry, A.; Rohrbach, M.; Elhoseiny, M.; Ajanthan, T.; Dokania, P. K.; Torr, P. H.; and Ranzato, M. 2019. On Tiny Episodic Memories in Continual Learning. *arXiv preprint*.

Chrysakis, A.; and Moens, M. F. 2020. Online Continual Learning from Imbalanced Data. In *International Conference on Machine Learning*, 1952–1961. PMLR.

Delange, M.; Aljundi, R.; Masana, M.; Parisot, S.; Jia, X.; Leonardis, A.; Slabaugh, G.; and Tuytelaars, T. 2021. A Continual Learning Survey: Defying Forgetting in Classification Tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Dhariwal, P.; and Nichol, A. 2021. Diffusion models beat GANs on image synthesis. *Advances in Neural Information Processing Systems*, 34: 8780–8794.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. In *Advances in neural information processing systems*, volume 27.

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

Ho, J.; and Salimans, T. 2022. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*.

Hung, A. L.; Zhao, K.; Zheng, H.; Yan, R.; Raman, S. S.; Terzopoulos, D.; and Sung, K. 2023. Med-cDiff: Conditional Medical Image Generation with Diffusion Models. *Bioengineering*, 10(11): 1258.

Khader, F.; Mueller-Franzes, G.; Arasteh, S. T.; Han, T.; Haarburger, C.; Schulze-Hagen, M.; Schad, P.; et al. 2022. Medical Diffusion: Denoising Diffusion Probabilistic Models for 3D Medical Image Generation. *arXiv preprint arXiv:2211.03364*.

Kim, C. D.; Jeong, J.; and Kim, G. 2020. Imbalanced Continual Learning with Partitioning Reservoir Sampling. In *Computer Vision–ECCV 2020*, 411–428. Springer International Publishing.

Kingma, D. P. 2013. Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114*.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25.

Liu, X.; Cao, X.; Lu, H.; Xiao, J.; Bagdanov, A. D.; and Cheng, M.-M. 2023. Class Incremental Learning with Pre-Trained Vision-Language Models. *arXiv preprint arXiv:2310.20348*.

Lopez-Paz, D.; and Ranzato, M. 2017a. Gradient Episodic Memory for Continual Learning. In *Advances in Neural Information Processing Systems*, volume 30.

Lopez-Paz, D.; and Ranzato, M. 2017b. Gradient Episodic Memory for Continual Learning. In *Advances in Neural Information Processing Systems*.

Masana, M.; Tuytelaars, T.; and Van de Weijer, J. 2021. Ternary Feature Masks: Zero-Forgetting for Task-Incremental Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3570–3579.

Masip, S.; Rodriguez, P.; Tuytelaars, T.; and van de Ven, G. M. 2023. Continual Learning of Diffusion Models with Generative Distillation. *arXiv preprint*.

Qin, Y.; Zheng, H.; Yao, J.; Zhou, M.; and Zhang, Y. 2023. Class-balancing diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18434–18443.

Ramesh, A.; Pavlov, M.; Goh, G.; Gray, S.; Voss, C.; Radford, A.; and Sutskever, I. 2021. Zero-shot text-to-image generation. In *International Conference on Machine Learning*, 8821–8831. PMLR.

Riemer, M.; Cases, I.; Ajemian, R.; Liu, M.; Rish, I.; Tu, Y.; and Tesauro, G. 2018. Learning to Learn Without Forgetting by Maximizing Transfer and Minimizing Interference. *arXiv preprint arXiv:1810.11910*.

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

Saharia, C.; Chan, W.; Saxena, S.; Li, L.; Whang, J.; Denton, E.; and Salimans, T. 2022. Photorealistic text-to-image diffusion models with deep language understanding. *arXiv preprint arXiv:2205.11487*.

Smith, J. S.; Hsu, Y.-C.; Zhang, L.; Hua, T.; Kira, Z.; Shen, Y.; and Jin, H. 2023. Continual Diffusion: Continual Customization of Text-to-Image Diffusion with C-LoRA. *arXiv preprint*.

Srivastava, A.; Valkov, L.; Russell, C.; Gutmann, M. U.; and Sutton, C. 2017. VEEGAN: Reducing Mode Collapse in GANs Using Implicit Variational Learning. In *Neural Information Processing Systems*.

Tang, S.; Yan, X.; Xu, Y.; and Liu, X. 2019. Multimodal conditional image synthesis for self-driving simulations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 202–201.

Wang, B.; Chen, Q.; and Wang, Z. 2024. Diffusion-Based Visual Art Creation: A Survey and New Perspectives. *arXiv preprint arXiv:2408.12128*.

Wu, C.; Herranz, L.; Liu, X.; Wang, Y.; van de Weijer, J.; and Raducanu, B. 2018. Memory Replay GANs: Learning to Generate Images from New Categories without Forgetting. In *Neural Information Processing Systems*.

Xu, Q.; Xu, Z.; Philip, J.; Bi, S.; Shu, Z.; Sunkavalli, K.; and Neumann, U. 2022. Point-NeRF: Point-Based Neural Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5438–5448.

Zajac, M.; Deja, K.; Kuzina, A.; Tomczak, J. M.; Trzcinski, T.; Shkurti, F.; and Milos, P. 2023. Exploring Continual Learning of Diffusion Models. *arXiv preprint arXiv:2303.15342*.

Zhang, H.; Zhou, J.; Lin, H.; Ye, H.; Zhu, J.; Wang, Z.; and Liang, Y. 2024. CLoG: Benchmarking Continual Learning of Image Generation Models. *arXiv preprint*.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 586–595.

Zhu, R.; Guo, Y.; and Xue, J. 2020. Adjusting the Imbalance Ratio by the Dimensionality of Imbalanced Data. *Pattern Recognition Letters*, 133: 217–223.