LFS-GAN: Lifelong Few-Shot Image Generation

Juwon Seo[1*] Ji-Su Kang[2*] Gyeong-Moon Park[1†]

[1]Kyung Hee University, Yongin, Republic of Korea
[2]KLleon Tech., Seoul, Republic of Korea

{jwseo001, gmpark}@khu.ac.kr jisu.kang@klleon.io

Abstract

We address a challenging lifelong few-shot image generation task for the first time. In this situation, a generative model learns a sequence of tasks using only a few samples per task. Consequently, the learned model encounters both catastrophic forgetting and overfitting problems at a time. Existing studies on lifelong GANs have proposed modulation-based methods to prevent catastrophic forgetting. However, they require considerable additional parameters and cannot generate high-fidelity and diverse images from limited data. On the other hand, the existing few-shot GANs suffer from severe catastrophic forgetting when learning multiple tasks. To alleviate these issues, we propose a framework called Lifelong Few-Shot GAN (LFS-GAN) that can generate high-quality and diverse images in lifelong few-shot image generation task. Our proposed framework learns each task using an efficient task-specific modulator - Learnable Factorized Tensor (LeFT). LeFT is rank-constrained and has a rich representation ability due to its unique reconstruction technique. Furthermore, we propose a novel mode seeking loss to improve the diversity of our model in low-data circumstances. Extensive experiments demonstrate that the proposed LFS-GAN can generate high-fidelity and diverse images without any forgetting and mode collapse in various domains, achieving state-of-the-art in lifelong few-shot image generation task. Surprisingly, we find that our LFS-GAN even outperforms the existing few-shot GANs in the few-shot image generation task. The code is available at Github.

1. Introduction

Deep learning has achieved remarkable success in recent years, particularly in a single task learning on a large dataset such as ImageNet [9] or FFHQ [24]. However, obtaining a large amount of refined data for real-world applications is prohibitively expensive, and there are many domains where only limited data can be collected, such as the artistic domain. Additionally, when faced with long sequences of tasks over time, it is inevitable to train a new model for each target task.

In this context, many studies have recently highlighted the importance of lifelong few-shot learning. Lifelong few-shot learning combines two challenging settings: lifelong learning and few-shot learning. It seeks to overcome catastrophic forgetting [32, 12, 26] when learning a sequence of tasks over time, while learning from limited data without any overfitting problem. By integrating the concepts of lifelong and few-shot learning, lifelong few-shot learning holds great potential for real-world applications where data is limited or costly to obtain, and where learning a new model for each task is not practical.

Previous studies on lifelong few-shot learning have primarily focused on discriminative tasks [1, 31, 21, 38, 37, 39]. However, lifelong few-shot learning on generative tasks is unexplored before. Lifelong few-shot image generation task involves training a model to generate realistic and diverse images from handful training images, while continually learning new tasks and preserving the ability to generate images from the previous domains (see Figure 1). There are two key challenges in this setting. First, since the model learns tasks sequentially, it easily forgets the ability to generate samples of the previous tasks. Second, since the model learns from a biased and sparse distribution, it suffers from the mode collapse problem [2], i.e., it is prone to re-generate the same training samples or produce similar images regardless of the noises provided. Addressing these challenges is critical to enable the successful application of lifelong few-shot image generation in the real-world, where data is limited or costly to obtain.

To alleviate catastrophic forgetting, recent studies in the field of lifelong image generation have proposed a weight modulator inspired by the affine transformation [20], which enables generative models to learn task-specific informa-

*Equal contribution
†Corresponding author

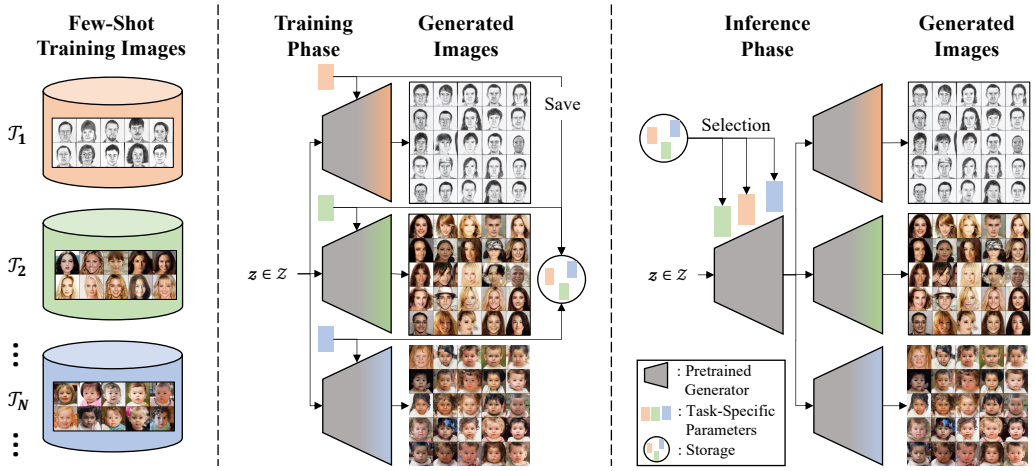


Figure 1: Illustration of our proposed lifelong few-shot image generation task. We construct a sequence of few-shot tasks. In the training phase, the model learns each task from the pretrained model. By updating only task-specific parameters applied to the model, the model can learn the current task without forgetting. In the inference phase, we can generate high-fidelity and diverse images of not only the current task but also the previous tasks by adopting task-specific parameters to the model.

tion in lifelong setting [47, 8, 42, 40]. However, conventional generative models [22, 24, 25, 23] consist of convolutional layers with high-dimensional weights, resulting in a significant increase of the number of parameters required to modulate them. As the number of tasks increases, the memory required to store these parameters becomes a significant burden. Furthermore, since lifelong generative models are designed to synthesize decent images from sufficient data, they suffer from mode collapse with handful training images. In the realm of few-shot image generation task, many works propose regularization-based methods to maintain the rich diversity of source models [28, 35, 45, 52]. However, they are prone to forget how to generate the previous tasks when learning a new task, since they fine-tune the models to learn the new task. A recent study [51] has introduced a modulation-based approach. This method divides the weights into be modulated and fine-tuned parts, based on their significance. Although this work adopts a modulation technique, it suffers from catastrophic forgetting due to its fine-tuning part.

To address these challenges, we propose a novel framework Lifelong Few-Shot Generative Adversarial Network (LFS-GAN). Our LFS-GAN learns a new task via a powerful weight modulation technique called the Learnable Factorized Tensor (LeFT) that captures task-specific knowledge with low memory costs while freezing learned weights from the source task. Our proposed LeFT reduces the memory burden by decomposing the weight tensor and restoring it during the forward operation to modulate the weight. This method enables the efficient and effective generation of high-quality images. Furthermore, we propose a cluster-wise mode seeking loss to improve the diversity of generated images. The mode seeking loss [30] has shown its effect to diversify the generated images of GANs. However, in a low-data circumstance, a simple application of mode seeking loss shows less effect because the generated images tend to be similar to the training images. Thus, we alter the mode seeking loss to be effective in our task and achieve greater diversity. Lastly, we find that the intra-cluster LPIPS cannot capture the imbalanced generation with respect to training images. To resolve this issue, we propose a novel diversity measure called Balanced Inter- and Intra-cluster LPIPS (B-LPIPS) to accurately evaluate generation diversity in our task.

Our main contributions can be summarized as follows:

- To the best of our knowledge, we formulate and tackle a challenging lifelong few-shot image generation task for the first time.
- We introduce a novel weight modulation technique, called Learnable Factorized Tensor (LeFT), which enables the generative model to learn new tasks without forgetting and significant parameter growth.
- To enhance diversity in the generated images, we propose a cluster-wise mode seeking loss that maximizes the relative distances of intermediate latent codes, feature maps, and images.
- Extensive experiments, including our novel metric B-LPIPS, demonstrate that LFS-GAN outperforms the current state-of-the-art methods on generating high-

quality and diverse images not only in lifelong few-shot image generation task but also in few-shot image generation task.

2. Related Work

Lifelong Image Generation. Recently, several methods have been proposed to alleviate catastrophic forgetting in generative models that are trained on a sequence of tasks continuously. One such approach is Lifelong GAN [48], which uses a knowledge distillation [16] based technique to prevent catastrophic forgetting. As a result, even when the generative model learns a new task, it minimally loses the previously acquired knowledge. Another method, MeRGAN [44], prevents catastrophic forgetting by utilizing memory replay techniques, such as joint retraining and aligning replays. During joint retraining, replayed samples are utilized, while the aligning replay forces the current generator to generate the same samples as the auxiliary generator. GAN-Memory [8] proposes a non-forgetting lifelong image generation algorithm by using additional memory to learn the newly arrived task. It introduces variants of modulation algorithms, such as FiLM [36] and AdaFM [50], to learn the current task. Despite the effectiveness of these methods, they have significant limitations in the low-data regime, such as severe mode collapse. Additionally, these methods require a large number of parameters for each new task, resulting in accelerating overfitting of the network. In contrast, our proposed approach takes into consideration the scarcity of training data in few-shot learning scenarios and employs a significantly reduced number of trainable parameters, effectively mitigating the risk of over-fitting. LoRA [18] also studies the efficient fine-tuning technique of Transformers [5] in natural language processing area. However, its decomposition and reconstruction schemes are too simple to apply to convolution layers of generative models.

Few-Shot Image Generation. Recently, there has been significant progress in few-shot classification tasks [11, 4, 3, 19]. This has led to a great interest in few-shot image generation, both in conditional and unconditional settings. The goal of few-shot image generation task is to generate realistic and diverse images from a limited number of training samples. FUNIT [28] explores image-to-image translation between source and target domains in few-shot context. Fusion-based methods, such as F2GAN [17], LoFGAN [14], and AGE [10], have also studied few-shot image generation in conditional settings, where there are a fixed number of training images per class or category. Our work focuses on a more challenging setting, where we have only a limited number of images per dataset or domain. Other studies on the unconditional setting are FastGAN [29], which proposes a fast and stabilized GAN architecture and a self-supervised learning method, and MoCA [27], which employs a prototype memory with an attention mechanism [41]. However, these studies tackle scenarios where the number of training samples is more than a hundred, while our approach addresses the extreme few-shot setting, where only ten training samples are available.

Many recent works have shown that leveraging pre-trained networks trained on large datasets can be effective in the extreme low-data regime. For instance, TGAN [43] argues that fine-tuning a network from a large source network can lead to effective results in few-shot setting. FreezeD [33] fine-tunes a pre-trained GAN by freezing the earlier layers of the discriminator. BSA [34] fine-tunes the pre-trained network by adapting batch statistics, while EWC [28] uses the Fisher information matrix to prevent changes in important weights. CDC [35] preserves pairwise distances among generated samples, and RSSA [45] utilizes a self-correlation matrix for structural consistency. DCL [52] maximizes mutual information using contrastive loss [6].

Recent work, AdAM [51], adopts a modulation-based approach in few-shot image generation, splitting the weights into modulated and fine-tuned components to generate appropriate images in large domain gaps. Motivated by AdAM, our proposed approach also employs a modulation-based approach. Unlike AdAM, our method shows no forgetting in lifelong few-shot image generation and exhibits superior generation quality and diversity.

3. Method

In this section, we first formulate lifelong few-shot image generation task (Section 3.1). To learn a generative model for this task, we propose a novel framework - Lifelong Few-Shot GAN (LFS-GAN). In Section 3.2, we introduce a lightweight modulation technique - Learnable Factorized Tensor. To enhance the generation diversity of our LFS-GAN, we adopt a variant of mode seeking loss, described in Section 3.3. Furthermore, we point out shortcomings of the existing metric for detecting imbalanced generation and propose a novel diversity measure i.e., Balanced Inter- and Intra-LPIPS (B-LPIPS) in Section 3.4.

3.1. Lifelong Few-Shot Image Generation Task

In this section, we first define lifelong few-shot image generation task. As illustrated in Figure 1, given a sequence of tasks $T = \{\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_N\}$, each task consists of a dataset which contains the images, denoted as $\mathcal{D}^t = \{x_i^t\}_{i=1}^k$. Here, k means the number of training samples, so we can call our task as k-shot image generation task. With few training samples, the model easily converges to the biased distribution, which causes the model to be overfitted. Moreover, as the given sequence of tasks becomes longer, the trained model significantly forgets the previous tasks. Our goal is to train a model to have the following ability: after training on the t^{th} task, the trained model can generate

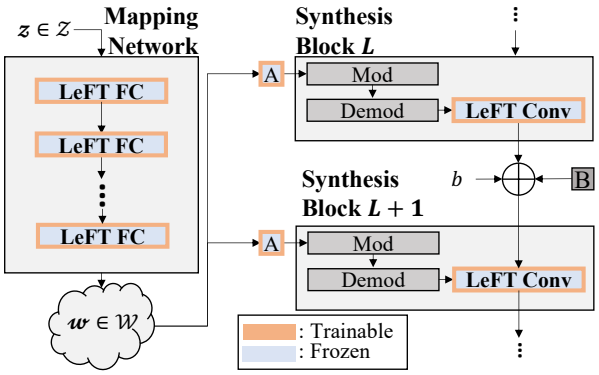


Figure 2: StyleGAN2 generator architecture with our proposed Learnable Factorized Tensor (LeFT). "A" and "B" denote the affine transformation and noise injection used in the original StyleGAN2 implementation, respectively. We apply LeFT on FC layers and conv. layers to modulate the original weights. The original weights are kept frozen during learning a new task while only LeFT is trainable.

the realistic and diverse samples of both the current task $\mathcal{T}_t$ and the previous tasks $\{\mathcal{T}_1, ..., \mathcal{T}_{t-1}\}$.

3.2. Learnable Factorized Tensor

To train a generative model on a sequence of tasks without catastrophic forgetting on the previous tasks, we adopt a modulation-based approach. Inspired by the style transfer literature [20], we modulate the pretrained weights by affine transformation. The affine transformation consists of two operations - multiplication and addition. We use StyleGAN2 [25] as our backbone, and each convolution layer has a weight tensor $\mathbf{W} \in \mathbb{R}^{c_{out} \times c_{in} \times k \times k}$. Here, c_{out} and c_{in} denote the size of the output and input channels, respectively, and k is the kernel size. We modulate the original weight tensor $\mathbf{W}$ to obtain the modulated weight tensor $\hat{\mathbf{W}} \in \mathbb{R}^{c_{out} \times c_{in} \times k \times k}$ like:

$$\hat{\mathbf{W}} = \mathbf{W} \odot \mathbf{\Gamma} + \mathbf{B}, \tag{1}$$

where $\mathbf{\Gamma} \in \mathbb{R}^{c_{out} \times c_{in} \times k \times k}$ and $\mathbf{B} \in \mathbb{R}^{c_{out} \times c_{in} \times k \times k}$ are the task-specific modulation parameters which are responsible for multiplication and addition, respectively, and $\odot$ is a Hadamard-product. In training on each task, we set the original weight $\mathbf{W}$ frozen and only the modulation parameters $\mathbf{\Gamma}, \mathbf{B}$ trainable. By training only the modulation parameters, we can learn new tasks while not updating the pretrained weights. Therefore, the model can learn multiple tasks without forgetting. We also apply the same approach to the fully-connected layers which consist of the weight tensor $\mathbf{W}_{FC} \in \mathbb{R}^{d_{out} \times d_{in}}$, where d_{out} and d_{in} are the dimensions of the output and input, respectively. In Figure 2, we demonstrate which layers of the StyleGAN2 generator are modulated. However, as the weight tensor of convolution layers is high-dimensional, storing task-specific modulation parameters for each task results in a serious memory burden. Furthermore, it is well-known that adapting a number of parameters to learn each low-shot task accelerates the over-fitting problem. To address these issues, we propose a novel weight decomposition technique called Learnable Factorized Tensor (LeFT). A key operation of LeFT is a rank-constrained decomposition. For example, we first reshape one of the modulation parameter $\mathbf{\Gamma}$ to be three-dimensional $\mathbf{\Gamma} \in \mathbb{R}^{c_{out} \times c_{in} \times K}$, where K is a square of kernel size. Therefore, the original tensor can be expressed using three two-dimensional matrices: $\mathcal{M}_1^{out} \in \mathbb{R}^{c_{out} \times r}$, $\mathcal{M}_1^{inst} \in \mathbb{R}^{r \times r \cdot K}$, and $\mathcal{M}_2^{in} \in \mathbb{R}^{c_{in} \times r}$, where r is a rank. We can reconstruct the original tensor by:

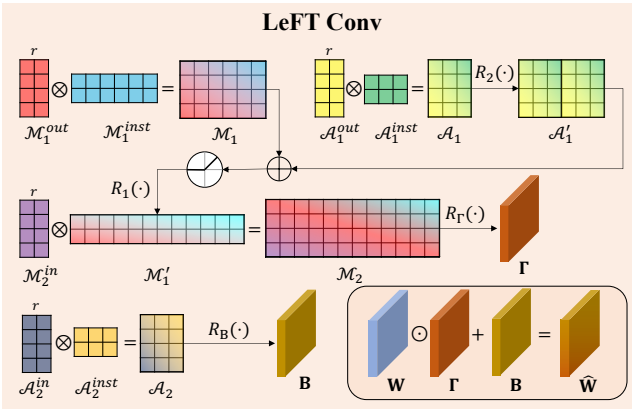


Figure 3: The reconstruction process of LeFT. The above example assumes that the rank r=2.

$$\mathcal{M}_1 = \mathcal{M}_1^{out} \otimes \mathcal{M}_1^{inst}, \tag{2}$$
$$\mathcal{M}_1' = R_1(\mathcal{M}_1), \tag{3}$$
$$\mathcal{M}_2 = \mathcal{M}_2^{in} \otimes \mathcal{M}_1', \tag{4}$$
$$\mathbf{\Gamma} = R_\Gamma(\mathcal{M}_2), \tag{5}$$

where $\otimes$ is a matrix-multiplication operation, and $R_1 : \mathbb{R}^{c_{out} \times r \cdot K} \rightarrow \mathbb{R}^{r \times c_{out} \cdot K}$ and $R_\Gamma : \mathbb{R}^{c_{in} \times c_{out} \cdot K} \rightarrow \mathbb{R}^{c_{out} \times c_{in} \times k \times k}$ are the reshaping functions. For the addition parameter $\mathbf{B}$, we reconstruct by using two-dimensional matrices: $\mathcal{A}_2^{in} \in \mathbb{R}^{c_{in} \times r}$ and $\mathcal{A}_2^{inst} \in \mathbb{R}^{r \times K}$:

$$\mathcal{A}_2 = \mathcal{A}_2^{in} \otimes \mathcal{A}_2^{inst}, \tag{6}$$
$$\mathbf{B} = R_B(\mathcal{A}_2), \tag{7}$$

where $R_B : \mathbb{R}^{c_{in} \times K} \rightarrow \mathbb{R}^{c_{out} \times c_{in} \times k \times k}$ is a function operating repetition and reshaping. By the above efficient decomposition and reconstruction scheme, we can reduce the number of modulation parameters significantly about less than 1%.

From Equation 2, recovering a three-dimensional tensor through multiplication is similar to the operation of multi-layered perceptron (MLP). In MLP, using bias and activation functions can generally improve performance. To apply bias and activation to the LeFT, we additionally introduce two matrices: $\mathcal{A}_1^{out} \in \mathbb{R}^{c_{out} \times r}$ and $\mathcal{A}_1^{inst} \in \mathbb{R}^{r \times K}$. These two matrices are also rank-constrained. We can reconstruct the bias $\mathcal{A}_1 \in \mathbb{R}^{r \times c_{out} \cdot K}$ as:

$$\mathcal{A}_1 = \mathcal{A}_1^{out} \otimes \mathcal{A}_1^{inst}, \tag{8}$$

$$\mathcal{A}_1' = R_2(\mathcal{A}_1), \tag{9}$$

where $R_2 : \mathbb{R}^{c_{out} \times K} \rightarrow \mathbb{R}^{r \times c_{out} \cdot K}$ is a repetition and reshaping function. Instead of Equation 3, we add the bias $\mathcal{A}_1$ to the intermediate matrix $\mathcal{M}_1$ and apply activation function $\mathrm{act}(\cdot)$ as follows:

$$\mathcal{M}_1' = R_1(\mathrm{act}(\mathcal{M}_1 + \mathcal{A}_1')). \tag{10}$$

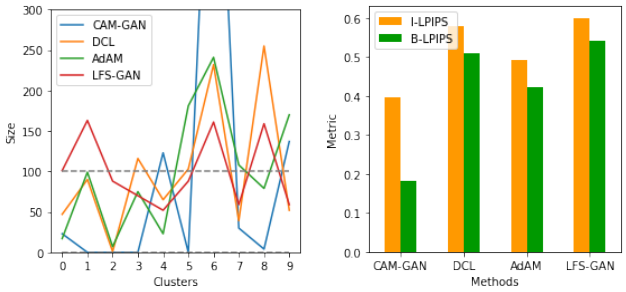
We experimentally find that using ReLU function as the activation function is most effective. The overall process is shown in Figure 3. We also apply LeFT on fully-connected layers. The decomposition and reconstruction processes are described in the Supplementary.

3.3. Cluster-Wise Mode Seeking Loss

Diverse image generation is a major interest when the number of training images is extremely small. Due to the lack of concerns on diversity, the existing modulation-based methods [8, 40, 51] show inferior diversity performance compared to regularization-based [28, 35, 45, 52] methods. Furthermore, rigid regularizations to preserve the diversity of source domain often result in unnatural distortions on generated images. To enhance diversity within the target domain, the mode seeking loss [30] and its variants [7] have shown diverse image generation. The mode seeking loss introduced in [30] is as follows:

$$\mathcal{L}_{ms} = \min(\frac{\Delta z}{\Delta I}), \tag{11}$$

where I and z are the generated images and the input noise vectors, respectively, and Δ is a mean absolute error as a distance measure. The mode seeking loss maximizes the distance of images with respect to the distance of the input noise vectors. However, in few-shot setting, the generated images are prone to be close to the given training images. In this situation, applying the original mode seeking loss is not effective. Therefore, we propose a cluster-wise mode seeking loss - a variant of mode seeking loss which is effective in few-shot scenario. Initially, we set clusters as many as B, a batch size of the training images. We generate $4B$ images and assign each of them to the perceptually closest cluster. Different from [30], we utilize the intermediate latent vector $w \in \mathcal{W}$ and the feature map F_l of layer l. We



(a) The size of each cluster. (b) I-LPIPS vs. B-LPIPS.

Figure 4: We measure the number of each cluster and its corresponding I-LPIPS and B-LPIPS on learning Babies task. The number of fake image is 1000 and the gray line represents the ideal number 100. We observe (a): the distribution of each cluster's size is highly imbalanced, (b): the I-LPIPS cannot reflect this biased distribution.

maximize w, F_l, and I with respect to z, w, and w within each cluster c_i. Thus, our cluster-wise mode seeking loss is computed as:

$$d_w = \frac{\Delta w}{\Delta z},\ d_F = \frac{1}{L}\sum_{l=1}^{L}\frac{\Delta F_l}{\Delta w},\ d_I = \frac{\Delta I}{\Delta w}, \tag{12}$$

$$\mathcal{L}_{cms} = \min(\frac{1}{\frac{1}{B}\sum_{i=1}^{B}\mathbb{E}_{z,w,F,I\sim c_i}\left[d_w + d_F + d_I\right]}). \tag{13}$$

We apply this cluster-wise mode collapse loss to update the generator. The total loss functions for the generator and the discriminator the are:

$$\mathcal{L}^G = \mathcal{L}_{adv}^G + \lambda\mathcal{L}_{cms}, \tag{14}$$

$$\mathcal{L}^D = \mathcal{L}_{adv}^D, \tag{15}$$

where $\mathcal{L}_{adv}^G$ and $\mathcal{L}_{adv}^D$ are the non-saturating adversarial losses proposed in [13], and λ is a hyper-parameter that controls the effect of the proposed cluster-wise mode seeking loss. We experimentally find that using $\lambda = 1$ is the most effective.

3.4. Balanced Inter- and Intra-Cluster LPIPS

Introduced in [35], intra-cluster LPIPS (I-LPIPS) has been widely used to measure diversity in few-shot image generation task [52, 51]. In I-LPIPS metric, the fake images are clustered to the real training images by the nearest LPIPS metric [49]. As a result, I-LPIPS is computed by averaging pairwise LPIPS distance within each cluster. However, we empirically find that in some cases, the size of the clusters can be highly imbalanced. Furthermore, there are cases where some clusters contain no generated images that are related to the corresponding training images. In such a

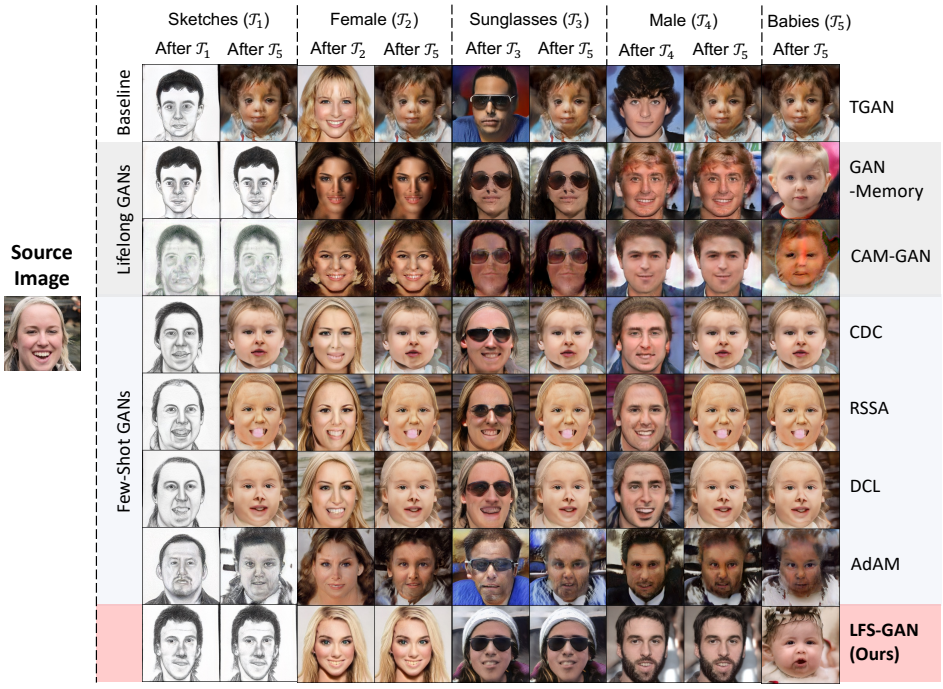


Figure 5: Qualitative comparison on lifelong few-shot image generation task. Given source image of left, we visualize the generated image of after each task and after the last task.

situation, intra-cluster LPIPS can not account for this imbalance. Figure 4 shows an example of the above situation.

To measure diversity in aware of cluster balance, we propose a novel metric called Balanced Inter- and Intra-Cluster LPIPS (B-LPIPS). We introduce a concept of entropy to reflect the balance of each cluster. For cluster c_i, we can compute the proportion $p(c_i)$ of the cluster:

$$p(c_i) = \frac{\|c_i\|}{N}, \tag{16}$$

where $\|c_i\|$ is the size of the cluster c_i and N is the total number of generated images. The B-LPIPS are computed as a weighted sum of pairwise LPIPS of each cluster:

$$w_i = -p(c_i) \cdot \log_{10} p(c_i), \tag{17}$$

$$\text{B-LPIPS} = \sum_{i=1}^{k} w_i \cdot \text{P-LPIPS}(c_i), \tag{18}$$

where k denotes the number of training images and $\text{P-LPIPS}(c_i)$ is a pairwise LPIPS within cluster c_i.

4. Experiments

4.1. Exeperimental Setup

Datasets. We used FFHQ [24], LSUN-Church, and LSUN-Cars [46] as the source domains. The target domains were (i) Sketches [35], (ii) Female [22], (iii) Sunglasses [35], (iv) Male [22], (v) Babies [35], (vi) Van Gogh's house (vii) Haunted house, (viii) Palace, and (ix) Abandoned cars. For source domains of LSUN-Church and LSUN-Cars, we presented experimental results in the Supplementary.

Baselines. Since our work tackled lifelong few-shot for the first time. We established our work on the baseline - (i) TGAN [43]. We evaluated two distinct approaches to our proposed lifelong few-shot image generation task - (1) lifelong GANs and (2) few-shot GANs. For lifelong GANs, we considered two methods: (ii) GAN-Memory [8] and (iii) CAM-GAN [40] are modulation-based lifelong GANs. For few-shot GANs, we evaluated four methods: (iv) CDC [35], (v) RSSA [45], and (vi) DCL [52] are regularization-based few-shot GANs and (vii) AdAM [51] is a modulation-based few-shot GAN.

Metrics. In the evaluation of the generation performance, we adopted three metrics. Firstly, the Fréchet Inception Distance (FID) score [15] was used to measure the similarity between the generated images and real images. A lower FID score indicates a higher quality of generation. Secondly, the balanced inter- and intra-cluster LPIPS (B-LPIPS), which is proposed in our paper, was used to measure the diversity of the generated images. Finally, we used the intra-cluster LPIPS (I-LPIPS) [35], as a traditional metric for the auxiliary measure of generation diversity. A higher B-LPIPS or I-LPIPS demonstrates a greater diversity of the generated images, as they are more distinct from one another. We

Method		Task Order										Average	
		Sketches ($\mathcal{T}_1$)		Female ($\mathcal{T}_2$)		Sunglasses ($\mathcal{T}_3$)		Male ($\mathcal{T}_4$)		Babies ($\mathcal{T}_5$)			
		FID	B-LPIPS	FID	B-LPIPS	FID	B-LPIPS	FID	B-LPIPS	FID	B-LPIPS	FID (↓)	B-LPIPS (↑)
TGAN (Baseline) [43]		372.89	0.157	255.53	0.238	309.13	0.247	281.43	0.129	171.19	0.203	278.03	0.195
Lifelong GANs	GAN-Memory [8]	69.58	0.311	71.56	0.287	87.02	0.169	99.44	0.143	177.73	0.150	101.05	0.212
	CAM-GAN [40]	91.81	0.293	85.68	0.332	86.81	0.333	82.83	0.312	146.20	0.181	98.66	0.290
Few-Shot GANs	CDC [35]	322.72	0.205	197.40	0.427	244.94	0.463	227.00	0.381	69.98	0.454	208.41	0.386
	RSSA [45]	308.00	0.285	175.20	0.440	207.58	0.484	205.49	0.405	76.70	0.481	194.59	0.419
	DCL [52]	297.73	0.307	170.31	0.435	191.54	0.490	194.42	0.443	77.22	0.487	186.25	0.432
	AdAM [51]	161.48	0.250	179.69	0.342	217.99	0.352	163.87	0.299	110.08	0.407	166.82	0.330
LFS-GAN (Ours)		**34.66**	**0.354**	**29.59**	**0.481**	**27.69**	**0.584**	**35.44**	**0.472**	**41.48**	**0.556**	**33.77**	**0.489**

Table 1: Quantitative results on **lifelong few-shot image generation** task. We measured each metric after the last task - Babies ($\mathcal{T}_5$). The bold value represents the best result and the underlined value represents the second best result.

sampled 5,000 images to compute FID score and 1,000 images to compute both B-LPIPS and I-LPIPS. In the main paper, we chose B-LPIPS as a default metric for measuring diversity. The diversity comparsion using I-LPIPS is presented in the Supplementary.

4.2. Results on LFS Task

Qualitative Results. At first, we evaluated state-of-the-art methods on our proposed lifelong few-shot image generation task. Figure 5 shows a qualitative result of state-of-the-art methods and LFS-GAN. As seen on the figure, TGAN and other few-shot GANs suffered from catastrophic forgetting. Moreover, they showed degraded quality in the last task. It was because they learned a single model on a sequence of task while ruining the model due to the biased and scarce distribution of each task. On the other hand, the images from lifelong GANs had a lot of distortions and were similar to the training images. This result demonstrated that lifelong GANs suffered from mode collapse. Compared to other methods, ours could generate high-quality and diverse images without forgetting.

Quantitative Results. As seen on Table 1, TGAN could not generate neither high-quality nor diverse images in our task. Lifelong GANs could generate high-quality images compared to TGAN. However, they could not generate diverse images. It was because they had no concerns about the low-data circumstance, thus they suffered from mode collapse. On the other hand, few-shot GANs generated diverse images, but they failed to generate high-quality images. We argue that this phenomenon was because of severe catastrophic forgetting happening during learning on a sequence of tasks. Unlike other methods, AdAM showed the alleviated forgetting. It is due to its training scheme of separating the weights into be fine-tuned and modulated, thus it could recover some amount of the knowledge of the previous tasks. However, our LFS-GAN could generate both high-quality and diverse images. Furthermore, we evaluated state-of-the-art methods on training efficiency. In Table 2, we find that our LFS-GAN achieved the most efficient parameter consumption to learn a new task.

Method		# of Trainable Params.	% w.r.t. Backbone
Baseline (TGAN) [43]		30.0M	100%
Lifelong GANs	GAN-Memory [8]	5.3M	17.7%
	CAM-GAN [40]	2.3M	7.7%
Few-Shot GANs	CDC [35]	30.0M	100.0%
	RSSA [45]	30.0M	100.0%
	DCL [52]	30.0M	100.0%
	AdAM [51]	18.9M	63.0%
LFS-GAN (Ours)		**0.1M**	**0.3%**

Table 2: Comparison on training efficiency. In the first column, there is the number of parameters to learn a new task of each method. Since we use StyleGAN2 as our backbone, the second column represents the percentage with respect to the number of parameters of the StyleGAN2 generator.

4.3. Results on FS Task

Different from the proposed lifelong few-shot image generation task, few-shot image generation task aims to generate decent and diverse images on a single target domain consisting of few-shot data.

Qualitative Results. We present the generated samples of state-of-the-art methods and LFS-GAN on the diverse target domains in the Supplementary. As seen on figure, similar to the results of lifelong few-shot image generation task, lifelong GANs could not generate neither high-quality nor diverse images in few-shot image generation task. On the other hand, few-shot GANs achieved superior performance compared to lifelong GANs. However, there were several distortions on generated images. We insist that these distortions came from strong regularizations. They restrict the ability to learn target domain, resulting in distortion.

Quantitative Results. As seen on Table 3, the existing lifelong GANs failed to generate high-quality or diverse images. Surprisingly, we find that our LFS-GAN also outperformed the existing few-shot GANs on few-shot image generation task.

Method		Tasks										Average	
		Sketches		Female		Sunglasses		Male		Babies			
		FID	B-LPIPS	FID	B-LPIPS	FID	B-LPIPS	FID	B-LPIPS	FID	B-LPIPS	FID (↓)	B-LPIPS(↑)
Baseline (TGAN)		60.57	0.335	67.63	0.318	72.81	0.391	73.44	0.319	114.29	0.414	77.75	0.355
Lifelong GANs	GAN-Memory	69.58	0.311	71.50	0.287	87.02	0.169	99.44	0.143	177.73	0.150	101.05	0.212
	CAM-GAN	91.81	0.293	85.68	0.332	86.81	0.333	82.83	0.312	146.20	0.181	98.66	0.290
Few-Shot GANs	CDC	49.19	0.237	31.26	0.450	36.03	0.505	41.88	0.435	64.75	0.496	44.62	0.425
	RSSA	56.25	0.251	34.24	0.467	44.01	0.510	44.83	0.434	72.45	0.486	50.36	0.430
	DCL	58.60	0.353	35.19	0.468	33.05	0.517	44.19	0.436	66.10	0.508	47.43	0.456
	AdAM	45.70	0.325	61.79	0.375	45.55	0.392	61.55	0.338	91.13	0.421	61.14	0.370
LFS-GAN (Ours)		**34.66**	**0.354**	**29.59**	**0.481**	**27.69**	**0.584**	**35.44**	**0.472**	**41.48**	**0.556**	**33.77**	**0.489**

Table 3: Quantitative results on **few-shot image generation** task. We conducted each task independently. The bold value represents the best result and the underlined value represents the secondary best result.

Bias	r	# of Trainable Params.	Average	
			FID (↓)	B-LPIPS (↑)
w/	1	108K	**33.77**	**0.489**
	2	192K	40.22	0.423
	4	358K	43.00	0.402
	8	695K	48.27	0.370
	16	1,380K	55.20	0.332
w/o	1	**54K**	39.42	0.416
	2	96K	39.56	0.410
	4	180K	44.40	0.402
	8	350K	48.74	0.374
	16	704K	52.42	0.326

Table 4: Ablation on the bias and the rank of LeFT.

Activation	Average	
	FID (↓)	B-LPIPS (↑)
Identity	39.87	0.451
Sigmoid	37.76	0.417
Tanh	38.25	0.440
LeakyReLU	35.42	0.437
GELU	37.78	0.448
SiLU	40.28	0.417
ReLU	**33.77**	**0.489**

Table 5: Ablation on the activation functions of LeFT.

Maximize			Average
$\Delta w/\Delta z$	$\Delta F/\Delta w$	$\Delta I/\Delta w$	B-LPIPS (↑)
			0.423
		✓	0.436
	✓		0.426
✓			0.435
✓	✓	✓	**0.489**

Table 6: Ablation on the maximization target of the cluster-wise mode seeking loss.

4.4. Analysis

Ablation on the LeFT components. We first inspected the effect of the bias term and rank of LeFT in Table 4. In most cases, the bias was responsible for improving both quality and diversity. We found that while LeFT without bias and of rank of 1 (the sixth row) reduced the number of trainable parameters a lot, the bias term was more crucial for generating high-quality and diverse images. As seen on Table 5, we observed that applying activation functions on LeFT generally improved the quality of generated samples and we selected to use ReLU as an activation function of LeFT by its decent performance in general.

Ablation on the cluster-wise mode seeking loss. Finally, we experimented on which part to be maximized in cluster-wise mode seeking loss (see Table 6). In this table, not to apply cluster-wise mode seeking loss (the first row) showed no effect on enhancing diversity compared to the finalized method (the last row). In general, we decided to maximize all parts by cluster-wise mode seeking loss for improved diversity and a slight gain in quality.

5. Conclusion

In this paper, we formulate and tackle the challenging lifelong few-shot image generation for the first time. To generate high-quality and diverse images in our task, we propose a novel framework Lifelong Few-Shot GAN, LFS-GAN for short. In LFS-GAN, we learn each task by introducing a novel weight modulation technique Learnable Factorized Tensor (LeFT). When learning each task, we only train LeFT parameters while freezing the original weights, thus we can achieve lifelong few-shot image generation without forgetting. Moreover, we propose a variant of mode seeking loss - cluster-wise mode seeking loss to enhance the diversity of generated images with less affecting the quality. Extensive experiments demonstrate that our LFS-GAN achieves state-of-the-art in generating high-quality and diverse images in both lifelong few-shot image generation task and few-shot image generation task.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1G1A1094379), and in part by MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2023-RS-2023-00258649) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation), and in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea Government (MSIT) (Artificial Intelligence Innovation Hub) under Grant 2021-0-02068, and by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.RS-2022-00155911, Artificial Intelligence Convergence Innovation Human Resources Development (Kyung Hee University)).

References

[1] Afra Feyza Akyürek, Ekin Akyürek, Derry Wijaya, and Jacob Andreas. Subspace regularizers for few-shot class incremental learning. In *International Conference on Learning Representations*, 2022. 1

[2] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017. 1

[3] Sungyong Baik, Janghoon Choi, Heewon Kim, Dohee Cho, Jaesik Min, and Kyoung Mu Lee. Meta-learning with task-adaptive loss function for few-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9465–9474, 2021. 3

[4] Sungyong Baik, Myungsub Choi, Janghoon Choi, Heewon Kim, and Kyoung Mu Lee. Meta-learning with adaptive hyperparameters. *Advances in neural information processing systems*, 33:20755–20765, 2020. 3

[5] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020. 3

[6] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020. 3

[7] Yen-Chi Cheng, Chieh Hubert Lin, Hsin-Ying Lee, Jian Ren, Sergey Tulyakov, and Ming-Hsuan Yang. Inout: diverse image outpainting via gan inversion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11431–11440, 2022. 5

[8] Yulai Cong, Miaoyun Zhao, Jianqiao Li, Sijia Wang, and Lawrence Carin. Gan memory with no forgetting. *Advances in Neural Information Processing Systems*, 33:16481–16494, 2020. 2, 3, 5, 6, 7

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. 1

[10] Guanqi Ding, Xinzhe Han, Shuhui Wang, Shuzhe Wu, Xin Jin, Dandan Tu, and Qingming Huang. Attribute group editing for reliable few-shot image generation. *arXiv preprint arXiv:2203.08422*, 2022. 3

[11] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pages 1126–1135. PMLR, 2017. 3

[12] Robert M. French. Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, 3(4):128–135, 1999. 1

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 5

[14] Zheng Gu, Wenbin Li, Jing Huo, Lei Wang, and Yang Gao. Lofgan: Fusing local representations for few-shot image generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8463–8471, October 2021. 3

[15] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. 6

[16] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distiling the knowledge in a neural network. *In NIPS workshop on Deep Learning and Representation Learning*, 2015. 3

[17] Yan Hong, Li Niu, Jianfu Zhang, Weijie Zhao, Chen Fu, and Liqing Zhang. F2gan: Fusing-and-filling gan for few-shot image generation. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 2535–2543, 2020. 3

[18] Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. 3

[19] Shell Xu Hu, Da Li, Jan Stühmer, Minyoung Kim, and Timothy M. Hospedales. Pushing the limits of simple pipelines for few-shot learning: External data and fine-tuning make a difference. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9068–9077, June 2022. 3

[20] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, pages 1501–1510, 2017. 1, 4

[21] Xisen Jin, Bill Yuchen Lin, Mohammad Rostami, and Xiang Ren. Learn continually, generalize rapidly: Lifelong knowledge accumulation for few-shot learning. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 714–729, 2021. 1

[22] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability

and variation. In *International Conference on Learning Representations*, 2018. 2, 6

[23] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. In *Proc. NeurIPS*, 2021. 2

[24] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1, 2, 6

[25] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020. 2, 4

[26] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017. 1

[27] Tianqin Li, Zijie Li, Andrew Luo, Harold Rockwell, Amir Barati Farimani, and Tai Sing Lee. Prototype memory and attention mechanisms for few shot image generation. In *International Conference on Learning Representations*, 2022. 3

[28] Yijun Li, Richard Zhang, Jingwan Cynthia Lu, and Eli Shechtman. Few-shot image generation with elastic weight consolidation. *Advances in Neural Information Processing Systems*, 33:15885–15896, 2020. 2, 3, 5

[29] Bingchen Liu, Yizhe Zhu, Kunpeng Song, and Ahmed Elgammal. Towards faster and stabilized gan training for high-fidelity few-shot image synthesis. In *International Conference on Learning Representations*, 2021. 3

[30] Qi Mao, Hsin-Ying Lee, Hung-Yu Tseng, Siwei Ma, and Ming-Hsuan Yang. Mode seeking generative adversarial networks for diverse image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1429–1437, 2019. 2, 5

[31] Pratik Mazumder, Pravendra Singh, and Piyush Rai. Few-shot lifelong learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 2337–2345, 2021. 1

[32] Michael McCloskey and Neal J. Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. volume 24 of *Psychology of Learning and Motivation*, pages 109–165. Academic Press, 1989. 1

[33] Sangwoo Mo, Minsu Cho, and Jinwoo Shin. Freeze the discriminator: a simple baseline for fine-tuning gans. *arXiv preprint arXiv:2002.10964*, 2020. 3

[34] Atsuhiro Noguchi and Tatsuya Harada. Image generation from small datasets via batch statistics adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2750–2758, 2019. 3

[35] Utkarsh Ojha, Yijun Li, Jingwan Lu, Alexei A Efros, Yong Jae Lee, Eli Shechtman, and Richard Zhang. Few-shot image generation via cross-domain correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10743–10752, 2021. 2, 3, 5, 6, 7

[36] Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. 3

[37] Chengwei Qin and Shafiq Joty. LFPT5: A unified framework for lifelong few-shot language learning based on prompt tuning of t5. In *International Conference on Learning Representations*, 2022. 1

[38] Guangyuan Shi, Jiaxin Chen, Wenlong Zhang, Li-Ming Zhan, and Xiao-Ming Wu. Overcoming catastrophic forgetting in incremental few-shot learning by finding flat minima. *Advances in Neural Information Processing Systems*, 34, 2021. 1

[39] Xiaoyu Tao, Xiaopeng Hong, Xinyuan Chang, Songlin Dong, Xing Wei, and Yihong Gong. Few-shot class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12183–12192, 2020. 1

[40] Sakshi Varshney, Vinay Kumar Verma, PK Srijith, Lawrence Carin, and Piyush Rai. Cam-gan: Continual adaptation modules for generative adversarial networks. *Advances in Neural Information Processing Systems*, 34, 2021. 2, 5, 6, 7

[41] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 3

[42] Vinay Kumar Verma, Kevin J Liang, Nikhil Mehta, Piyush Rai, and Lawrence Carin. Efficient feature transformations for discriminative and generative continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13865–13875, 2021. 2

[43] Yaxing Wang, Chenshen Wu, Luis Herranz, Joost van de Weijer, Abel Gonzalez-Garcia, and Bogdan Raducanu. Transferring gans: generating images from limited data. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 218–234, 2018. 3, 6, 7

[44] Chenshen Wu, Luis Herranz, Xialei Liu, Joost van de Weijer, Bogdan Raducanu, et al. Memory replay gans: Learning to generate new categories without forgetting. *Advances in Neural Information Processing Systems*, 31, 2018. 3

[45] Jiayu Xiao, Liang Li, Chaofei Wang, Zheng-Jun Zha, and Qingming Huang. Few shot generative model adaption via relaxed spatial structural alignment. *arXiv preprint arXiv:2203.04121*, 2022. 2, 3, 5, 6, 7

[46] Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*, 2015. 6

[47] Mengyao Zhai, Lei Chen, and Greg Mori. Hyper-lifelonggan: Scalable lifelong learning for image conditioned generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2246–2255, 2021. 2

[48] Mengyao Zhai, Lei Chen, Frederick Tung, Jiawei He, Megha Nawhal, and Greg Mori. Lifelong gan: Continual learning for conditional image generation. In *Proceedings of the*

IEEE/CVF International Conference on Computer Vision, pages 2759–2768, 2019. 3

[49] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 5

[50] Miaoyun Zhao, Yulai Cong, and Lawrence Carin. On leveraging pretrained gans for generation with limited data. In *International Conference on Machine Learning*, pages 11340–11351. PMLR, 2020. 3

[51] Yunqing Zhao, Keshigeyan Chandrasegaran, Milad Abdollahzadeh, and Ngai man Cheung. Few-shot image generation via adaptation-aware kernel modulation. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. 2, 3, 5, 6, 7

[52] Yunqing Zhao, Henghui Ding, Houjing Huang, and Ngai-Man Cheung. A closer look at few-shot image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9140–9150, 2022. 2, 3, 5, 6, 7