Supplementary Materials for
LFS-GAN: Lifelong Few-Shot Image Generation

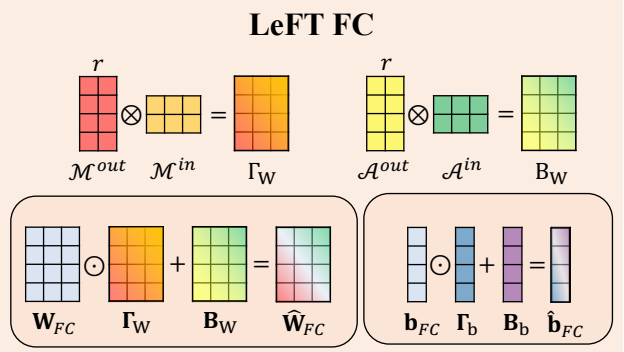


Figure 1: The reconstruction process of LeFT on fully-connected layers. The above example assumes that the rank r=2.

A. LeFT on Fully-Connected Layers

Fully-connected layers of StyleGAN2 [4] architecture consist of weight tensor $\mathbf{W}_{FC} \in \mathbb{R}^{d_{out} \times d_{in}}$ and bias $\mathbf{b}_{FC} \in \mathbb{R}^{d_{out}}$. We also apply LeFT to fully-connected layers to modulate these parameters like:

$$\hat{\mathbf{W}}_{FC} = \mathbf{W}_{FC} \odot \mathbf{\Gamma}_{\mathbf{W}} + \mathbf{B}_{\mathbf{W}}, \quad (1)$$

$$\hat{\mathbf{b}}_{FC} = \mathbf{b}_{FC} \odot \mathbf{\Gamma}_{\mathbf{b}} + \mathbf{B}_{\mathbf{b}}, \quad (2)$$

where $\mathbf{\Gamma}_{\{\mathbf{W},\mathbf{b}\}}$ and $\mathbf{B}_{\{\mathbf{W},\mathbf{b}\}}$ denote the modulation parameters of LeFT which are responsible for multiplication and addition to modulate weight and bias, respectively. Dimensions of these modulation parameters are equal to the dimensions of the original parameters - $\mathbf{W}_{FC}$ or $\mathbf{b}_{FC}$. Since $\mathbf{W}_{FC}$ is two-dimensional, we also apply a rank-constrained decomposition similar to the way in the convolution layers described on the main paper. The overall process is shown on Figure 1.

Similar to convolutional layers, we decompose $\{\mathbf{\Gamma}, \mathbf{B}\}_{\mathbf{W}}$ into $\{\mathcal{M}, \mathcal{A}\}^{out} \in \mathbb{R}^{d_{out} \times r}$ and $\{\mathcal{M}, \mathcal{A}\}^{in} \in \mathbb{R}^{r \times d_{in}}$. We can reconstruct the original modulation parameters via matrix-multiplication. For bias parameter, because of its single dimensionality, we do not apply weight decomposition.

Domains	Sketches	Female	Sunglasses	Male	Babies
FFHQ	**0.735**	0.253	0.571	0.309	0.531
Sketches		**0.697**	0.665	0.688	0.683
Female			**0.523**	0.266	0.480
Sunglasses				**0.498**	0.497
Male					**0.471**

Table 1: Pairwise LPIPS distance between domains. The large value represents that the two domains are perceptually distant.

B. Determination of Task Sequence

We constructed a sequence of tasks using perceptual distances between domains. To measure distance, we adopted Learned Perceptual Image Patch Similarity [10] (LPIPS). We computed pairwise LPIPS distance between two domain pair. The results are shown on Table 1. For the most challenging setting, we organized the task order by assigning the least similar task as the next task compared to the current task. As a result, we were able to decide the sequence of tasks as Sketches ($\mathcal{T}_1$) , Female ($\mathcal{T}_2$), Sunglasses ($\mathcal{T}_3$), Male ($\mathcal{T}_4$), and Babies ($\mathcal{T}_5$).

C. Implementation Detail

We used StyleGAN2 [4][1] as a backbone of our framework. Our training configurations came from [6, 12]. We adopted Patch Discriminator proposed in [6]. We trained our model using Adam optimizer [5] with the learning rate of 0.002. The batch size was set to 4. We experimented on a single GeForce RTX 3090 GPU. When we train our framework on each task, we froze the pre-trained weights and adopted LeFT modulators on them. We only saved the lightweight LeFT modulators after learning on each task. In the inference step, we loaded LeFT modulator to our backbone to generate the images of the previous tasks.

[1] https://github.com/rosinality/stylegan2-pytorch

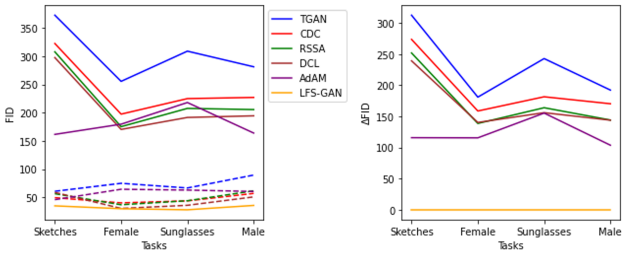


Figure 2: (left): We compared FID [1] scores after training on each tasks and last task in lifelong few-shot image generation task. The solid lines represent the results after the last task, while the dashed lines represent those of each tasks. (right): We visualized the difference of FID scores.

Method	Sketches ($\mathcal{T}_1$)		Female ($\mathcal{T}_2$)		Sunglasses ($\mathcal{T}_3$)		Male ($\mathcal{T}_4$)		Babies ($\mathcal{T}_5$)	
	B	I	B	I	B	I	B	I	B	I
AdAM	0.250	0.395	0.342	0.428	0.352	0.490	0.229	0.443	0.407	0.498
LFS-GAN	**0.354**	**0.405**	**0.481**	**0.546**	**0.584**	**0.631**	**0.472**	**0.561**	**0.556**	**0.627**

Table 2: Comparison between LFS-GAN and state-of-the-art AdAM on generation diversity in lifelong few-shot image generation task. Here we denote B-LPIPS as 'B' and I-LPIPS as 'I'.

D. Additional Experiments on LFS Task

D.1. Comparison on Forgetting

In Figure 2, we visualized forgetting occurred in lifelong few-shot image generation task. We measured FID scores after the each task and after the last task to inspect the performance degradation. We compared our LFS-GAN with the existing few-shot GANs. We found that few-shot GANs suffered from catastrophic forgetting, while our LFS-GAN learned multiple tasks without forgetting. Surprisingly, AdAM [11] showed the reduced forgetting compared with other few-shot GANs. We explain this phenomenon that the weight modulators of AdAM alleviated catastrophic forgetting by recovering a part of previous domain's knowledge.

D.2. Additional Qualitative Results

We additionally sampled images from our LFS-GAN and state-of-the-art methods in lifelong few-shot image generation task, the results are shown on Figure 3. In this figure, which is also shown on Figure 5 of the main paper, lifelong GANs still generated images with a lot of distortions. On the other hand, few-shot GANs still failed to generate samples of the previous domain ($\mathcal{T}_1 \sim \mathcal{T}_4$).

In Figure 4, we also prepared other images sampled from our LFS-GAN on diverse target domains. To test our LFS-GAN on different source and target domain pairs, we trained our LFS-GAN from LSUN-Church [8] for a source domain to a sequence of target domains - Haunted houses, Van Gogh's house paintings, and Palace. The qualitative results are shown on Figure 5. We found that our LFS-GAN could also generate decent images within different source and target domain pairs.

D.3. Quantitative Comparison using I-LPIPS

In Table 2, we compared LFS-GAN with the existing state-of-the-art AdAM on generation diversity. LFS-GAN outperforms AdAM on both proposed and traditional metrics.

D.4. Experiments on Recent Generative Models

We tested recent unconditional image generation models - StyleSwin [9] and Latent Diffusion Models (LDM) [7] on both few-shot and lifelong few-shot image generation tasks. As shown on Table 3, both models also suffered from both catastrophic forgetting and overfitting. Therefore the challenge of lifelong few-shot image generation task is not limited StyleGAN2 [4]. Moreover, when applied LeFT to StyleSwin and LDM, both methods learned few data successfully without any forgetting. Thus, our proposed LeFT can be generalized on recent generative models. In this paper, we chose StyleGAN2 as our backbone because it still has been widely utilized and shown comparable performance to other recent generative models like StyleSwin and LDM in FFHQ 256x256 dataset. We compared generation performances of above architectures in Table 4.

E. Additional Experiments on FS Task

E.1. Qualitative Results

As shown on Figure 6-11, we trained state-of-the-art methods and our LFS-GAN framework on Sketches [6], Female [2], Sunglasses [6], Male [2], and Babies [6] from source domain of FFHQ [3], and on Abandoned cars from source domain of LSUN-Cars [8]. In these figures, lifelong GANs generated distorted images and showed a mode collapse problem in all tasks. Few-shot GANs generated images of better quality. However, we find that there happened a lot of distortions. Our LFS-GAN can generate images with reduced distortion and rich diversity compared to the existing state-of-the-art methods.

F. Additional Ablation Studies

F.1. Effect of Rank

In Figure 12, we compared FID scores in different rank. In the graph, the larger ranks (orange and green lines) tended to perform worse and diverge, while the smallest rank (blue line) consistently achieved convergence. To test our LFS-GAN in a semantically large domain gap between source and target, we set the source domain as LSUN-Church and the target domains as the existing facial do-

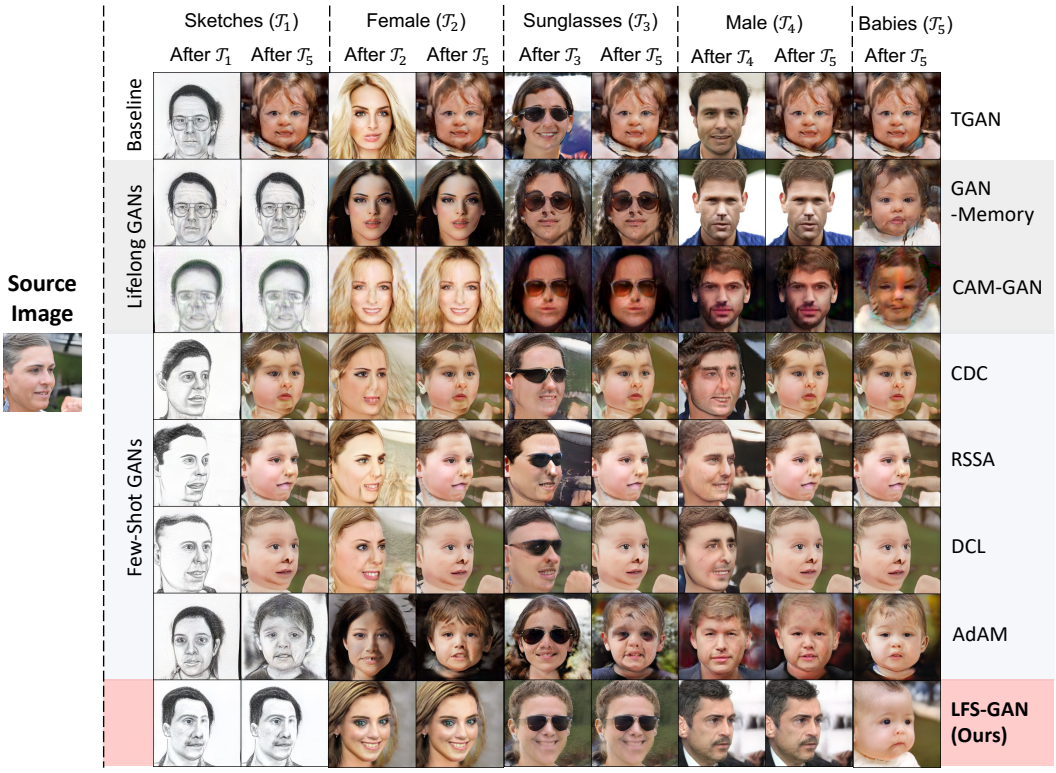


Figure 3: Qualitative results on lifelong few-shot image generation task.

Method	Setting	Sketches ($\mathcal{T}_1$)		Female ($\mathcal{T}_2$)		Sunglasses ($\mathcal{T}_3$)		Male ($\mathcal{T}_4$)		Babies ($\mathcal{T}_5$)	
		FID	B	FID	B	FID	B	FID	B	FID	B
StyleSwin	FS	81.79	0.25	59.29	0.31	65.23	0.34	67.95	0.34	111.46	0.34
StyleSwin	LFS	343.98	0.11	243.77	0.16	275.28	0.16	279.63	0.17	157.31	0.10
StyleSwin+LeFT	LFS	65.48	0.31	27.31	0.50	21.89	0.54	33.94	0.41	48.81	0.51
LDM	FS	112.86	0.20	105.73	0.25	85.32	0.34	106.09	0.23	144.79	0.32
LDM	LFS	213.09	0.03	174.48	0.06	248.94	0.12	200.17	0.13	295.15	0.05
LDM+LeFT	LFS	75.72	0.19	73.46	0.32	71.68	0.36	94.24	0.28	132.30	0.40

Table 3: Quantitative results of other methods on few-shot and lifelong few-shot image generation task. We denote B-LPIPS as 'B'.

Methods	FID (↓)
StyleGAN2	3.62
StyleSwin	2.81
LDM	4.98

Table 4: Comparison of generative models on FFHQ 256x256 dataset. To evaluate FID scores, each method samples 50,000 images.

mains. The results are shown on Figure 13. Note that the rank of 8 was the most effective in this large domain gap setting.

λ	Tasks					Average
	Sketches	Female	Sunglasses	Male	Babies	
0.25	0.211	0.430	0.550	**0.492**	0.498	0.436
0.5	0.309	<u>0.470</u>	<u>0.558</u>	0.434	<u>0.519</u>	<u>0.458</u>
1	**0.354**	**0.481**	**0.584**	0.472	**0.556**	**0.489**
2	0.220	0.431	0.512	<u>0.489</u>	0.493	0.429
4	<u>0.352</u>	0.448	0.510	0.453	0.504	0.453

Table 5: Ablation on λ of the cluster-wise mode seeking loss. We measured B-LPIPS on different λ.

Bias	r	# of Trainable Params.	Sketches		Female		Sunglasses		Male		Babies		Average	
			FID	B-LPIPS	FID	B-LPIPS	FID	B-LPIPS	FID	B-LPIPS	FID	B-LPIPS	FID (↓)	B-LPIPS (↑)
	1	108K	**34.66**	**0.354**	**29.59**	0.481	**27.69**	**0.584**	**35.44**	**0.472**	**41.48**	**0.556**	**33.77**	**0.489**
	2	192K	35.19	0.237	35.96	0.447	34.85	0.537	43.58	0.417	51.53	0.479	40.22	0.423
w/	4	358K	38.08	0.248	39.86	0.383	42.92	0.463	44.67	0.422	49.49	0.493	43.00	0.402
	8	695K	35.85	0.163	40.68	0.405	45.74	0.468	54.15	0.331	64.90	0.486	48.27	0.370
	16	1,380K	40.74	0.232	48.46	0.304	53.00	0.407	56.14	0.349	77.65	0.367	55.20	0.332
	1	**54K**	37.52	0.223	34.11	**0.492**	33.14	0.448	40.19	0.423	52.15	0.494	39.42	0.416
	2	96K	35.67	0.274	34.59	0.404	34.07	0.418	41.46	0.417	52.04	0.536	39.56	0.410
w/o	4	180K	41.66	0.256	36.86	0.402	40.87	0.423	42.62	0.370	59.98	0.556	44.40	0.402
	8	350K	41.22	0.201	42.25	0.364	43.94	0.468	53.35	0.314	62.93	0.524	48.74	0.374
	16	704K	44.67	0.213	46.88	0.313	46.04	0.401	57.69	0.307	66.82	0.395	52.42	0.326

Table 6: Ablation on the bias and the rank of LeFT (detailed).

Activation	Sketches		Female		Sunglasses		Male		Babies		Average	
	FID	B-LPIPS	FID	B-LPIPS	FID	B-LPIPS	FID	B-LPIPS	FID	B-LPIPS	FID (↓)	B-LPIPS (↑)
Identity	45.25	0.285	30.77	0.475	32.66	0.525	39.64	0.462	51.00	0.509	39.87	0.451
Sigmoid	34.16	0.250	**28.92**	0.452	31.10	0.496	41.77	0.383	52.85	0.503	37.76	0.417
Tanh	40.86	0.292	30.92	0.458	31.24	0.523	37.97	0.442	50.25	0.486	38.25	0.440
LeakyReLU	**33.43**	0.277	49.65	0.451	30.77	0.505	**30.96**	0.449	**32.39**	0.504	35.42	0.437
GELU	41.94	0.322	29.67	0.472	28.55	0.541	36.84	0.418	51.88	0.488	37.78	0.448
SiLU	41.06	0.251	32.78	0.418	32.82	0.519	36.57	0.390	58.17	0.510	40.28	0.417
ReLU	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 7: Ablation on the activation functions of LeFT (detailed).

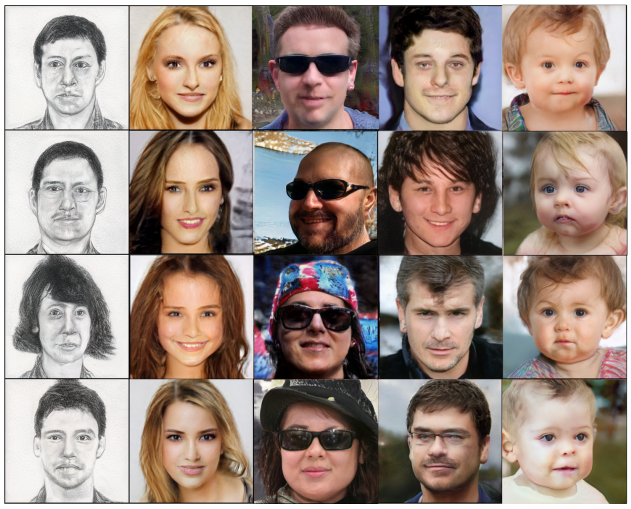


Figure 4: Diverse samples generated from LFS-GAN. Each column represents each target domain (i.e., Sketches, Female, Sunglasses, Male, and Babies respectively.)

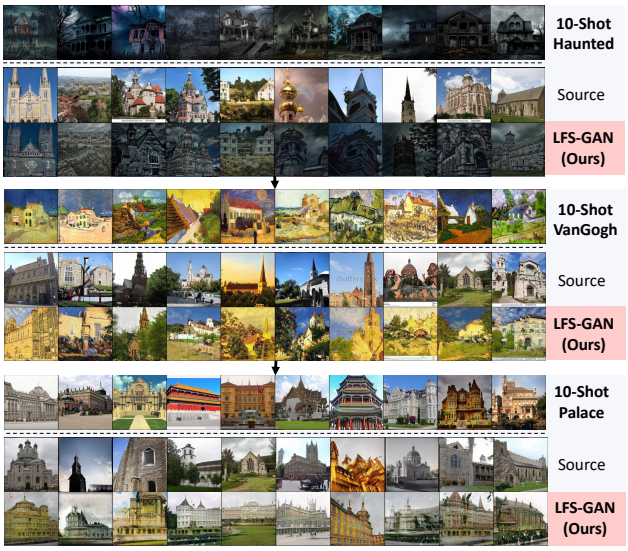


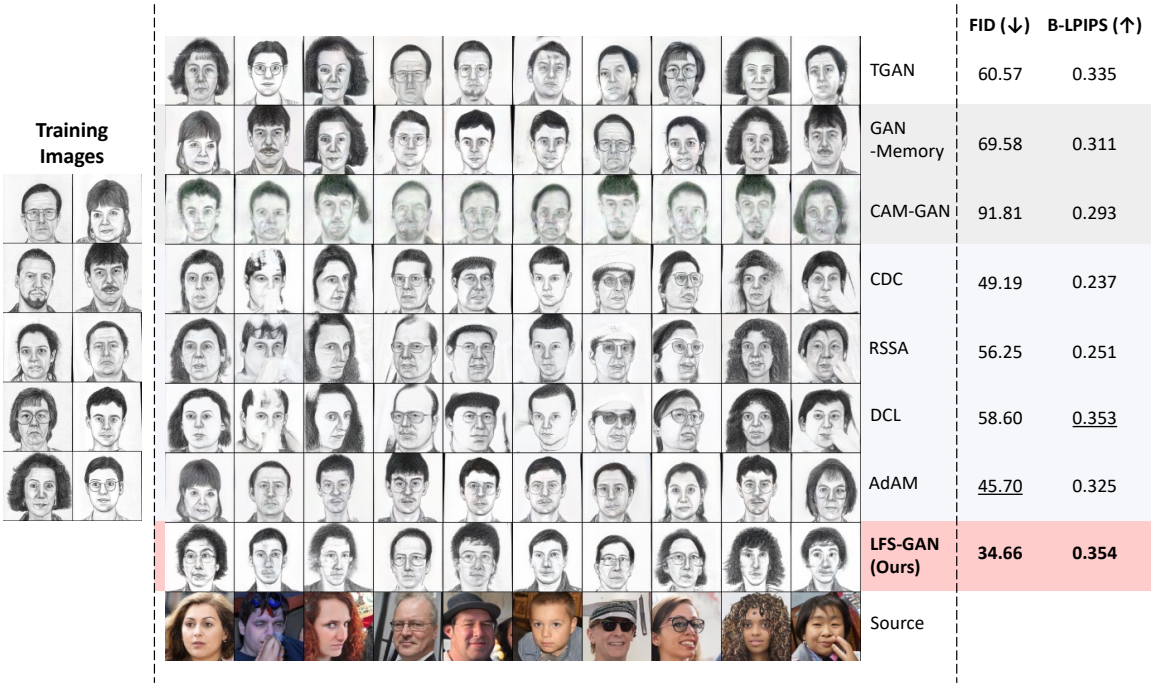
Figure 5: Qualitative results of our LFS-GAN on lifelong few-shot image generation task. The source domain is LSUN-Church [8], and the target domains are Haunted houses, Van Gogh's house paintings, and Palace.

F.2. The Number of Training Images

In Figure 14, we presented the performance difference according to the number of training images. We observed that as the number of training images increased, the generation quality also increased.

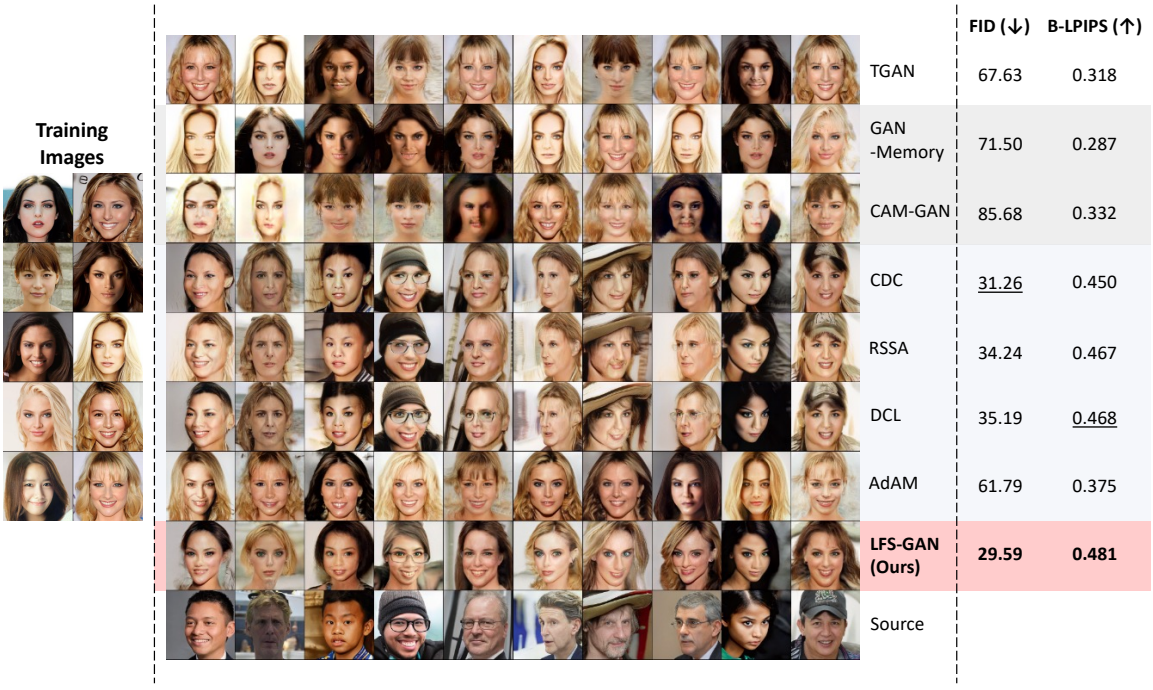
F.3. λ of Cluster-Wise Mode Seeking Loss

As shown on Table 5, we inspected the effect of λ of our proposed cluster-wise mode seeking loss. While we found



	FID (↓)	B-LPIPS (↑)
TGAN	60.57	0.335
GAN -Memory	69.58	0.311
CAM-GAN	91.81	0.293
CDC	49.19	0.237
RSSA	56.25	0.251
DCL	58.60	<u>0.353</u>
AdAM	<u>45.70</u>	0.325
LFS-GAN (Ours)	**34.66**	**0.354**

Figure 6: Qualitative comparison with state-of-the-art lifelong GANs and few-shot GANs on Sketches in few-shot image generation task.

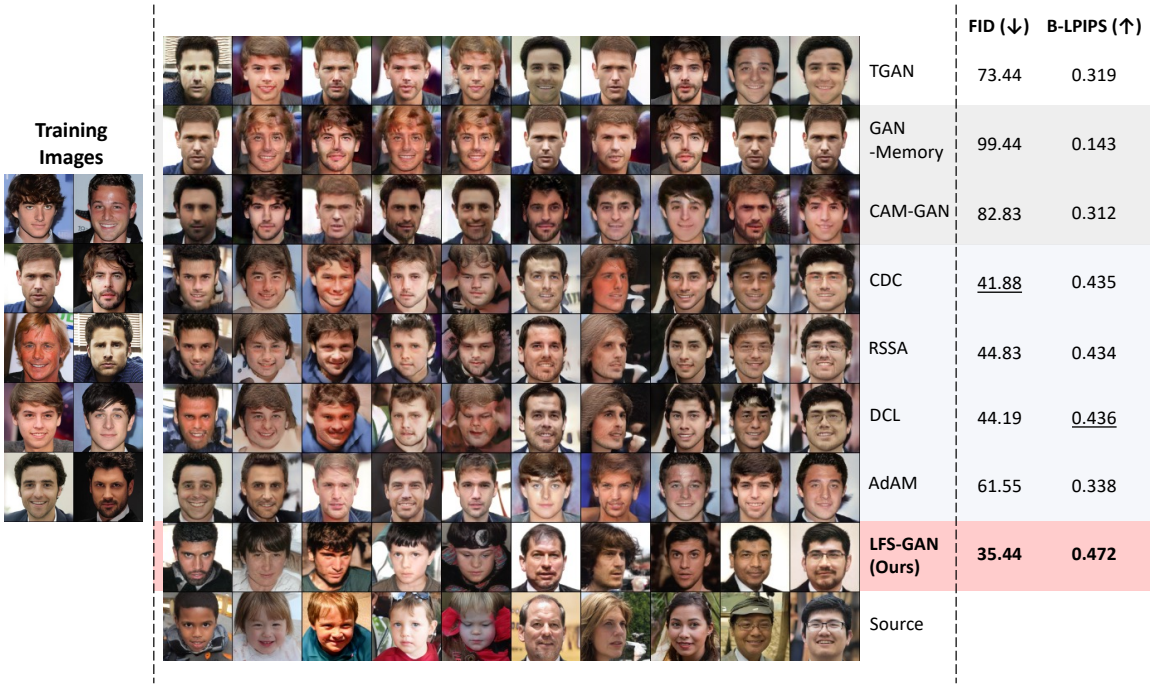


	FID (↓)	B-LPIPS (↑)
TGAN	67.63	0.318
GAN -Memory	71.50	0.287
CAM-GAN	85.68	0.332
CDC	<u>31.26</u>	0.450
RSSA	34.24	0.467
DCL	35.19	<u>0.468</u>
AdAM	61.79	0.375
LFS-GAN (Ours)	**29.59**	**0.481**

Figure 7: Qualitative comparison with state-of-the-art lifelong GANs and few-shot GANs on Female in few-shot image generation task.

	FID (↓)	B-LPIPS (↑)
TGAN	72.81	0.391
GAN -Memory	87.02	0.169
CAM-GAN	86.81	0.333
CDC	36.03	0.505
RSSA	44.01	0.510
DCL	33.05	0.517
AdAM	45.55	0.392
LFS-GAN (Ours)	**27.69**	**0.584**
Source		

Figure 8: Qualitative comparison with state-of-the-art lifelong GANs and few-shot GANs on Sunglasses in few-shot image generation task.



	FID (↓)	B-LPIPS (↑)
TGAN	73.44	0.319
GAN -Memory	99.44	0.143
CAM-GAN	82.83	0.312
CDC	41.88	0.435
RSSA	44.83	0.434
DCL	44.19	0.436
AdAM	61.55	0.338
LFS-GAN (Ours)	**35.44**	**0.472**
Source		

Figure 9: Qualitative comparison with state-of-the-art lifelong GANs and few-shot GANs on Male in few-shot image generation task.

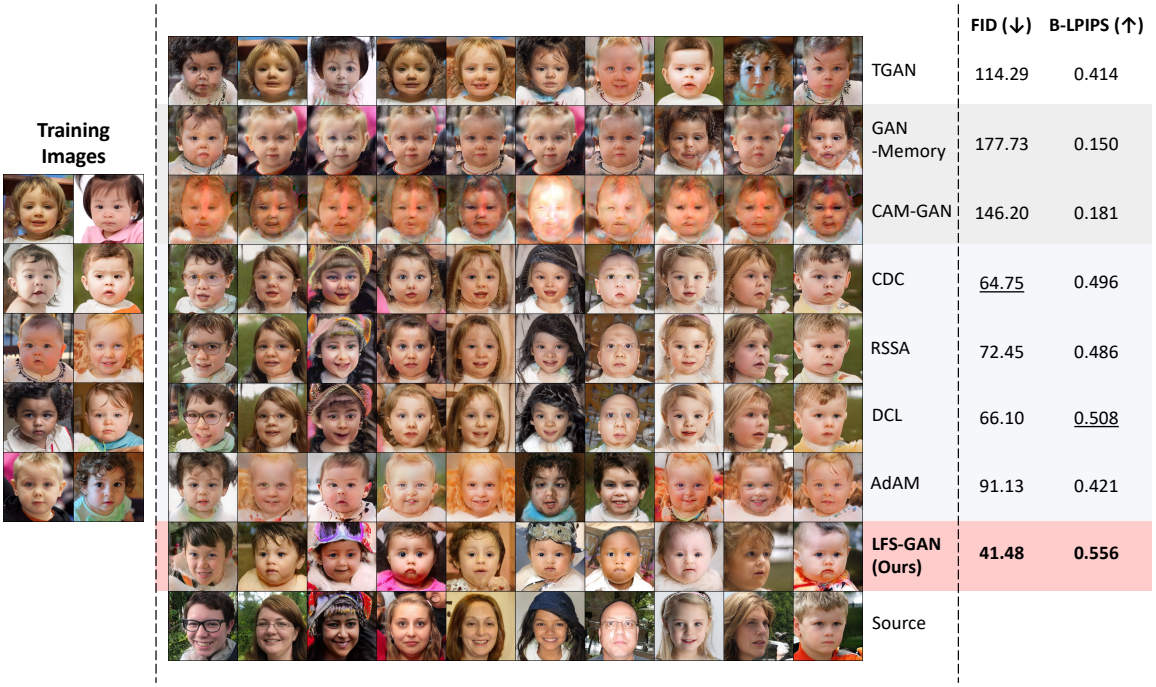


Figure 10: Qualitative comparison with state-of-the-art lifelong GANs and few-shot GANs on Babies in few-shot image generation task.

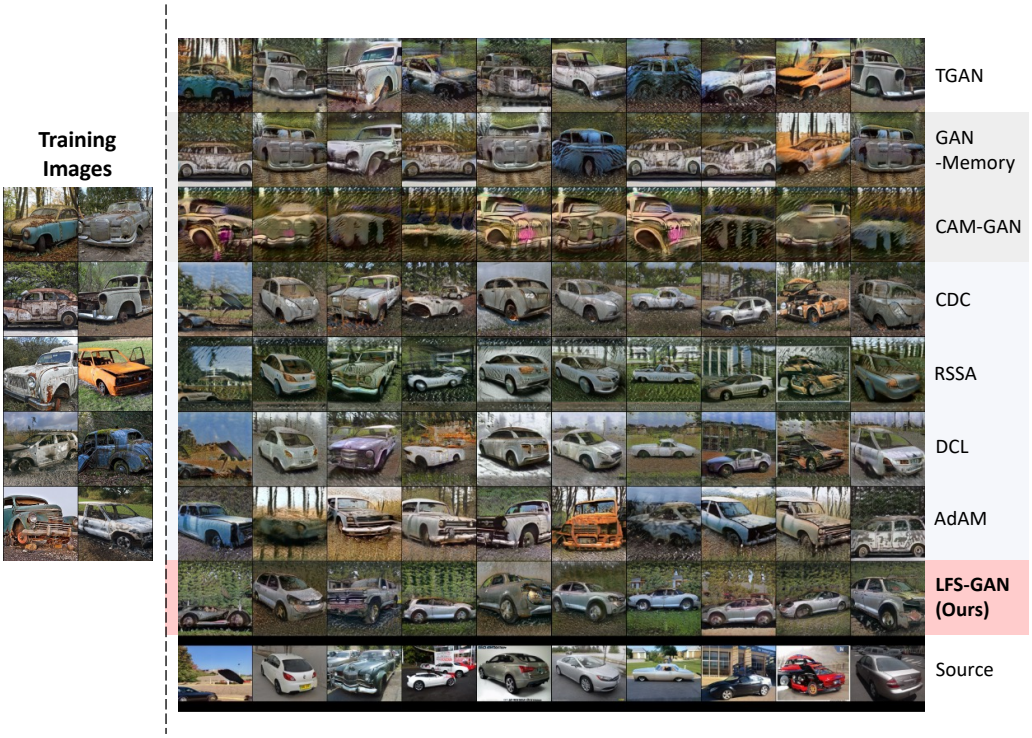


Figure 11: Qualitative comparison with state-of-the-art lifelong GANs and few-shot GANs on Abandoned cars in few-shot image generation task.

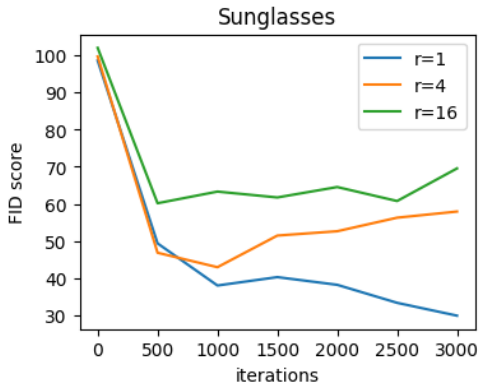


Figure 12: Comparison of FID scores in different ranks.

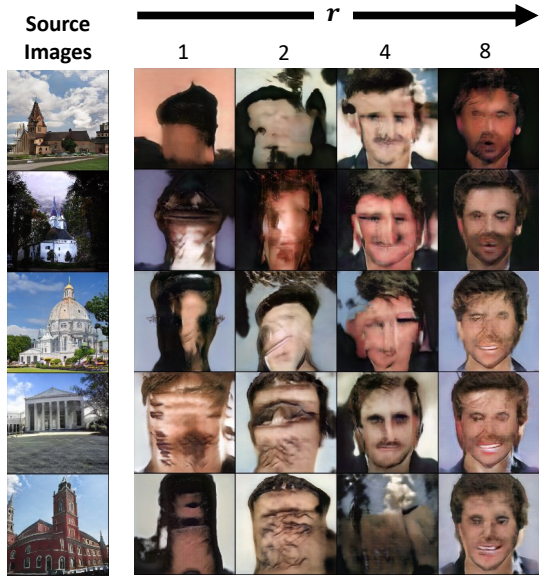


Figure 13: Qualitative results on the distant domain pair (LSUN-Church → Male).

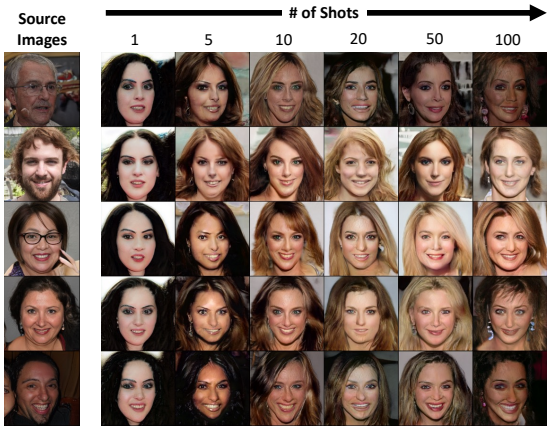


Figure 14: Qualitative results on different number of training images.

Maximize $\Delta w/\Delta z$	Maximize $\Delta F/\Delta w$	Maximize $\Delta I/\Delta w$	Sketches	Female	Sunglasses	Male	Babies	Average
			0.221	0.427	0.499	0.418	0.551	0.423
		✓	0.282	0.469	0.473	0.447	0.511	0.436
	✓		0.267	0.450	0.526	0.397	0.486	0.426
✓			0.278	0.446	0.537	0.415	0.500	0.435
✓	✓	✓	**0.354**	**0.481**	**0.584**	**0.472**	**0.556**	**0.489**

Table 8: Ablation on the maximization target of the cluster-wise mode seeking loss (detailed). We measuresd B-LPIPS on different maximization target in our proposed cluster-wise mode seeking loss.

that λ=1 was the most effective, our LFS-GAN showed comparable performance on different λ values.

F.4. Detailed Results of Ablation Studies

We described the detailed results of ablation studies which were previously stated in the main paper on Table 6, 7, and 8. We found that using bias, the rank of 1, and ReLU activation generally showed the superior performances compared to other options. Furthermore, we confirmed that maximizing the relative distance of intermediate latent vectors (w), feature maps (F), and generated images (I) was the most effective for enriching diversity.

References

[1] 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. 2

[2] 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

[3] 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. 2

[4] 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. 1, 2

[5] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 1

[6] 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. 1, 2

[7] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022. 2

[8] 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. 2, 4

[9] Bowen Zhang, Shuyang Gu, Bo Zhang, Jianmin Bao, Dong Chen, Fang Wen, Yong Wang, and Baining Guo. Styleswin: Transformer-based gan for high-resolution image generation. In *CVPR*, 2022. 2

[10] 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. 1

[11] 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

[12] 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. 1