Few-shot Image Generation via Adaptation-aware Kernel Modulation

Anonymous Author(s)
Affiliation
Address
email

Abstract

Few-shot image generation (FSIG) aims to learn to generate new and diverse samples given an extremely limited number of samples from a domain, *e.g.*, 10 training samples. Recent work has addressed the problem using transfer learning approach, leveraging a GAN pretrained on a large-scale source domain dataset and adapting that model to the target domain based on very limited target domain samples. Central to recent FSIG methods are *knowledge preserving criteria*, which aim to select a subset of source model's knowledge to be preserved into the adapted model. However, a *major limitation* of existing methods is that their knowledge preserving criteria consider *only source domain/source task*, and they fail to consider *target domain/adaptation task* in selecting source model's knowledge, casting doubt on their suitability for setups of different *proximity* between source and target domain. **Our work** makes two contributions. As our first contribution, we re-visit recent FSIG works and their experiments. Our important finding is that, under setups which assumption of close proximity between source and target domains is relaxed, existing state-of-the-art (SOTA) methods which consider only source domain in knowledge preserving perform *no better* than a baseline fine-tuning method. To address the limitation of existing methods, as our second contribution, we propose *adaptation-aware kernel modulation* to address general FSIG of different source-target domain proximity. Extensive experimental results show that the proposed method consistently achieves SOTA performance across source/target domains of different proximity, including challenging setups when source and target domains are more apart. Code / reproducibility details are included.

1 Introduction

Generative Adversarial Networks (GANs) [6, 7, 8] have been applied to a range of important applications including image generation [9, 8, 10], image-to-image translation [11, 12], image editing [13, 14], and data augmentation [15, 16, 17]. However, a critical issue is that these GANs often require large-scale datasets and computationally expensive resources to achieve good performance. For example, StyleGAN [9] is trained on Flickr-Faces-HQ (FFHQ) [9] that contains 70,000 images. However, in many practical applications only a few samples are available (*e.g.*, photos of rare animal species). Training a generative model is problematic in this low-data regime, where the generator often suffers from mode collapse or blurred generated images [18, 4, 19]. To address this, *few-shot image generation* (FSIG) studies the possibility of generating sufficiently diverse and high quality images, given very limited training data (*e.g.*, 10 samples). FSIG also attracts an increasing interest for some downstream tasks, *e.g.*, few-shot classification [17].

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

Table 1: Transfer learning for few-shot image generation: Various criteria are proposed to *augment* baseline transfer learning to preserve subset of source model's knowledge into the adapted model.

Method	**Knowledge preserving criteria**	Source domain/task aware	Target domain/adaptation aware
TGAN [1]	Not available	–	–
FreezeD [2]	Preservation of lower layers of the discriminator pre-trained on the *source* domain.	✓	✗
EWC [3]	Preservation of weights important to the *source* generative model pre-trained on the *source* domain.	✓	✗
CDC [4]	Preservation of pairwise distances of generated images by the *source* generative model pre-trained on the *source* domain.	✓	✗
DCL [5]	Preservation of multilevel semantic diversity of the generated images by the *source* generative model pre-trained on the *source* domain.	✓	✗
Our work	Preservation of kernels important in ***adaptation*** of source model to *target*.	✓	✓

FSIG with Transfer Learning. Recent works in FSIG are based on transfer learning approach [20] *i.e.*, leveraging the prior knowledge of a GAN pretrained on a large-scale, diverse source dataset (*e.g.*, FFHQ [9] or ImageNet [21]) and adapting it to a target domain with very limited samples (*e.g.*, face paintings [22]). As only very limited samples are provided to define the underlying distribution, standard fine-tuning of a pre-trained GAN suffers from mode collapse: the adapted model can only generate samples closely resembling the given few shot target samples [1, 4]. Therefore, recent works [3, 4, 5] have proposed to *augment* standard fine-tuning with different criteria to carefully preserve subset of source model's knowledge into the adapted model. Various criteria has been proposed (Table 1), and these *knowledge preserving criteria* have been central in recent FSIG research. In general, these criteria aim to preserve subset of source model's knowledge which is deemed to be useful for target-domain sample generation, *e.g.*, improving the diversity of target sample generation.

Research Gaps. One major limitation of existing methods is that they consider *only* source domain in preserving subset of source model's knowledge into the adapted model. In particular, these methods *fail to consider* target domain/adaptation task in selection of source model's knowledge (Table 1). For example, EWC [3] applies Fisher Information [23] to select important weights entirely based on the pretrained *source* model, and it aims to preserve these selected weights regardless of the target domain in adaptation. Similar to EWC, CDC [4] proposes an additional constraint to preserve pairwise distances of generated images by the *source* model, and there is no consideration of target domain/adaptation. These *target/adaptation-agnostic* knowledge preserving criteria in recent works raise question regarding their suitability in different source/target domain setups. It should be noted that existing work of FSIG (under very limited target samples) focuses largely on setups where source and target domains are in *close proximity* (semantically) *e.g.*, Human faces (FFHQ)→Baby faces [3, 5], or Cars→Abandoned Cars [4, 5]. It is unclear about their performance when source/target domains are more apart (*e.g.*, Human faces (FFHQ) → Animal faces).

Contributions. In this paper we take an important step to address these research gaps for FSIG. Specifically, our work makes two contributions. **As our first contribution**, we re-visit existing state-of-the-art (SOTA) algorithms and their experiments. Importantly, we observe that when the close proximity assumption is relaxed in experiment setups and source/target domains are more apart, existing SOTA methods perform *no better* than a baseline fine-tuning method. Our observation suggests that recent methods considering only source domain/source task in knowledge preserving may not be suitable for *general* FSIG when source and target domains are more apart. To validate our claims, we introduce additional experiments with different source/target domains, analyze their proximity qualitatively and quantitatively, and examine existing methods under a unified framework.

Informed by our analysis, **as our second contribution**, we propose an *adaptation-aware kernel modulation* approach to address general FSIG of different source/target domain proximity. In marked contrast to existing works which preserve knowledge important to *source* task, our method aims to preserve subset of source model's knowledge that are important to the *target* domain and the *adaptation* task. More specifically, we propose an *importance probing* algorithm to identify kernels which encode important knowledge for adaptation to the target domain. Then, we preserve the knowledge of these kernels using a parameter-efficient *rank-constrained kernel modulation*. We conduct extensive experiments to show that our proposed method consistently achieves SOTA

performance across source/target domains of different proximity, including challenging setups when source/target domains are more apart. Our main contributions are summarized as follows:

- We revisit existing FSIG methods and experiment setups. Our study uncovers issues with existing methods when applied to source/target domains of different proximity.
- We propose an adaptation-aware kernel modulation method for FSIG. Our method consistently achieves SOTA performance both visually and quantitatively across source/target domains with different proximity. We include our code in Supplementary.

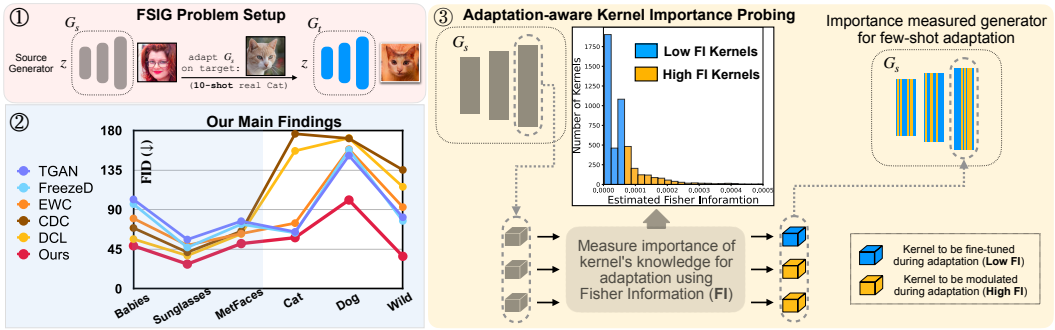


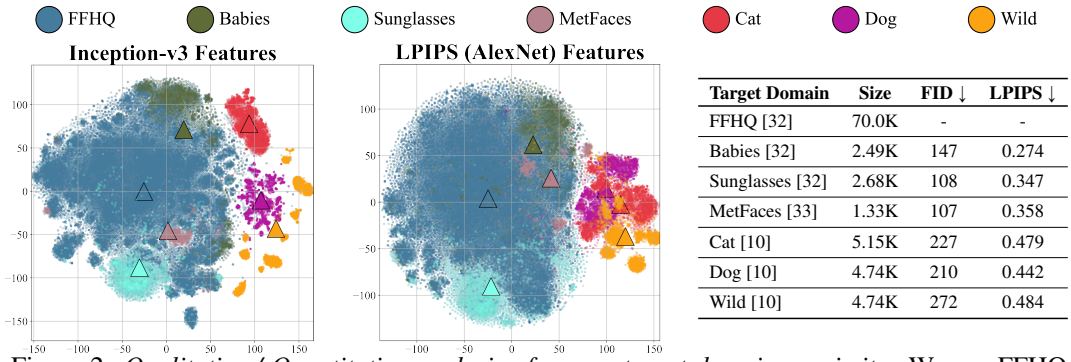
Figure 1: *Overview and our contributions.* ①: We consider the problem of FSIG with Transfer Learning using very limited target samples (*i.e.*10-shot). ②: Our work makes two contributions, • We discover that when the close proximity assumption between source-target domain is relaxed, SOTA FSIG methods (EWC [3], CDC [4], DCL [5]) which consider only source domain/source task in knowledge preserving perform no better than a baseline fine-tuning method (TGAN [1]) (Sec 3). • We propose a novel adaptation-aware kernel modulation for FSIG that achieves SOTA performance across source / target domains with different proximity (Sec 4). ③ Schematic diagram of our proposed Importance Probing Mechanism: We measure the importance of each kernel for the target domain after probing and preserve source domain knowledge that is important for target domain adaptation (Sec 4). The same operations are applied to discriminator. Best viewed in color.

2 Related Work

Few-shot image generation. Conventional few-shot learning [24, 25, 26] aims at learning a discriminative classifier for classification [26, 27], segmentation [28, 29] or detection [30, 31] tasks. Differently, few-shot image generation (FSIG) [4, 3, 5] aims at learning a generator for new and diverse samples given extremely limited samples (*e.g.*, 10 shots). Transfer learning has been applied to FSIG. For example, Transferring GAN [1] (**TGAN**) applies simple GAN loss [6] to fine-tune all parameters of both the generator and the discriminator. **FreezeD** [2] fixes a few high-resolution discriminator layers during fine-tuning. To augment and improve simple fine-tuning, more recent works have focused on preserving specific knowledge from the source models. Elastic weight consolidation (**EWC**) [3] identifies important weights for the *source* model and tries to preserve these weights. Cross-domain Correspondence (**CDC**) [4] preserves pair-wise distance of generated images from the source model to alleviate mode collapse. Dual Contrastive Learning (**DCL**) [5] applies mutual information maximization to preserve multi-level diversity of the generated images by the source model. In this work, we observe that these SOTA methods perform poorly when source and target domains are more apart. Therefore, their proposed source knowledge preservation criteria *may not* be generalizable. Based on our analysis, we propose an adaptation-aware knowledge selection which is more *generalizable* for source/target domains with different proximity.

3 Revisiting FSIG through the Lens of Source–Target Domain Proximity

In this section, we revisit existing FSIG methods (10-shot) [1, 2, 3, 4, 5] through the lens of source-target domain proximity. Specifically, we scrutinize the experimental setups of existing FSIG methods and observe that SOTA [3, 4, 5] largely focus on adapting to target domains that are (semantically)



Target Domain	Size	FID ↓	LPIPS ↓
FFHQ [32]	70.0K	-	-
Babies [32]	2.49K	147	0.274
Sunglasses [32]	2.68K	108	0.347
MetFaces [33]	1.33K	107	0.358
Cat [10]	5.15K	227	0.479
Dog [10]	4.74K	210	0.442
Wild [10]	4.74K	272	0.484

Figure 2: *Qualitative / Quantitative analysis of source-target domain proximity:* We use FFHQ [8] as the source domain. We show source-target domain proximity qualitatively by visualizing Inception-v3 (Left) [34] and LPIPS (Middle) [35] –using AlexNet [36] backbone– features, and quantitatively using FID / LPIPS metrics (Right). For feature visualization, we use t-SNE [37] and show centroids (△) for all domains. FID / LPIPS is measured with respect to FFHQ. There are 2 important observations: ① Common target domains used in existing FSIG works (Babies, Sunglasses, MetFaces) are notably proximal to the source domain (FFHQ). This can be observed from the feature visualization and verified by FID / LPIPS measurements. ② We clearly show using feature visualizations and FID / LPIPS measurements that additional setups – Cat [10], Dog [10] and Wild [10] – represent target domains that are distant from the source domain (FFHQ). We remark that large FID values in this analysis are reasonable due to the distance between the source (FFHQ) and different target domains as observed from centroid distance / feature variance. The effect of limited sample size (target domains) for FID / LPIPS measurements are minimal and we include rich supportive studies in Supplementary. Additional experiments and source/target setups in Supplementary to further support our analysis.

proximal to the source domain: Human Faces (FFHQ) → Baby Faces; Human Faces (FFHQ) → Sunglasses; Cars → Abandoned Cars; Church → Haunted Houses [3, 4, 5]. This raises the question as to whether existing source-target domain setups sufficiently represent *general* FSIG scenarios. Particularly, real-world FSIG applications may not contain target domains that are always proximal to the source domain (*e.g.*,: Human Faces (FFHQ) → Animal Faces). Motivated by this, we conduct an in-depth qualitative and quantitative analysis on source-target domain proximity where we introduce target domains that are distant from the source domain (Sec 3.1). Our analysis uncovers an important finding: **Under our additional setups where the assumption of close proximity between source and target domain is relaxed, existing SOTA FSIG methods [3, 4, 5] which consider only source domain/source task in knowledge preserving perform *no better* than a baseline fine-tuning method.** We show this is due to the strong focus of existing SOTA methods in preserving source domain knowledge, thereby not being able to adapt well to distant target domains (Sec 3.2) .

3.1 Source–Target Domain Proximity Analysis

Introducing target domains with varying degrees of proximity to the source domain. In this section, we formally introduce source-target domain proximity with in-depth analysis to scrutinize existing FSIG methods under different degrees of source-target domain proximity. Following prior FSIG works [1, 2, 3, 4, 5], we use FFHQ [8] as the source domain in this analysis. We remark that existing works largely consider different types of human faces as target domain (*i.e.*: Babies [8], Sunglasses [8], MetFaces [33]), To relax close the proximity assumption and study *general* FSIG problems, we introduce more distant target domains namely Cat, Dog and Wild (from AFHQ [10], consisting of 15,000 high-quality animal face images at 512 × 512 resolution) for our analysis.

Characterizing source-target domain proximity. Given the wide success of deep neural network features in representing meaningful semantic concepts [35, 38, 39, 40], we visualize Inception-v3 and LPIPS features for source and target domains to qualitatively characterize domain proximity. Further, we use FID and LPIPS distance to quantitatively characterize source-target domain proximity. We remark that FID involves distribution estimation (first, second order moments) [41] and LPIPS computes pairwise distances (learned embeddings) [35] between source / target domains.

Analysis. Feature Visualization and FID/ LPIPS measurement results are shown in Figure 2. Our results both qualitatively (columns 1, 2) and quantitatively (column 3) show that target domains used in existing works (Babies [8], Sunglasses [8], MetFaces [33]) are notably proximal to the source domain (FFHQ), and our additionally introduced target domains (Dog, Cat and Wild [10]) are distant from the source domain thereby relaxing the close proximity assumption in existing FSIG works.

3.2 FSIG methods under Relaxation of Close Domain Proximity Assumption

Motivated by our analysis in Section 3.1, we investigate the performance of existing FSIG methods [1, 2, 3, 4, 5] by relaxing the close proximity assumption between source and target domains. We investigate the performance of these FSIG methods across target domains of different proximity to the source domain, which includes our additionally introduced target domains: Dog, Cat and Wild. The FID results for FFHQ → Cat are: TGAN (simple fine-tuning) [1]: 64.68, EWC [3]: 74.61, CDC [4]: 176.21, DCL [5]: 156.82. Full results can be found in Table 2. We emphasize that our investigation uncovers an important finding: *Under setups which the assumption of close proximity between source and target domain is relaxed (Dog, Cat, Wild), existing SOTA FSIG methods [3, 4, 5]* perform *no better* than a baseline fine-tuning method. This can be consistently observed in Table 2.

This finding is critical as it exposes a serious drawback of SOTA FSIG methods [3, 4, 5] when close domain proximity (between source and target) assumption is relaxed. We further analyse generated images from SOTA FSIG methods and observe that these methods are unable to adapt well to distant target domains due to *only considering source domain / task in knowledge preservation.* This can be clearly observed from Figure 3. We remark that TGAN (simple baseline) [1] also suffers from severe mode collapse. Given that our investigation uncovers an important problem in SOTA FSIG methods, we tackle this problem in Sec 4. Figure 3 (last row) shows a glimpse of our proposed method.

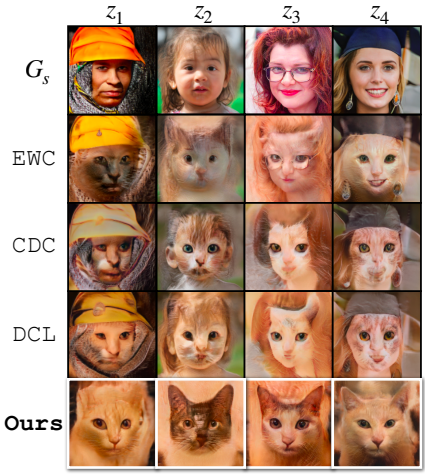


Figure 3: G_s is the Source Generator (FFHQ). Adapting from the source domain (FFHQ) to a distant target domain (Cat) using SOTA FSIG methods EWC [3], CDC [4], DCL [5] (rows 2, 3, 4) results in observable knowledge transfer that is not useful to the target domain. *i.e.*: Source task knowledge such as *Caps* (z_1, z_4)*, Hair styles/color – brown* (z_2)*, red-hair* (z_3) *–, Eye glasses* (z_3) *from FFHQ* are transferred to Cats during adaptation which is not appropriate. Our method (last row) can alleviate these issues.

4 Adaptation-Aware Kernel Modulation

We focus on this question: *"Given a pretrained GAN on a source domain $\mathcal{D}_s$, and a few-samples from a target domain $\mathcal{D}_t$, which part of the source model's knowledge should be preserved, and which part should be updated, during the adaptation from $\mathcal{D}_s$ to $\mathcal{D}_t$?"* In contrast to SOTA FSIG methods [3, 4, 5], we propose an adaptation-aware FSIG that also considers the target domain / adaptation task in deciding which part of the source model's knowledge to be preserved. In a CNN, each *kernel* is responsible for a specific part of knowledge (*e.g.*, pattern or texture). Similar behaviour is also observed for both generator [42] and discriminator [43] in GANs. Therefore, in this work, we make this knowledge preservation decision at the kernel level, *i.e.*, ***casting the knowledge preservation to a decision problem of whether a kernel is important when adapting from $\mathcal{D}_s$ to $\mathcal{D}_t$.***

Our FSIG algorithm has two main steps: (i) a lightweight *importance probing* step, and (ii) *main adaptation* step. In the first step, *i.e.*, importance probing, we adapt the model using a parameter-efficient design to the target domain for a limited number of iterations, and during this adaptation, we measure the importance of each individual kernel for the *target domain*. The output of importance probing are decisions of importance / unimportance of individual kernels. Then, in the second step, *i.e.*, main adaptation, we preserve the knowledge of the important kernels and update the knowledge of the unimportant kernels. The overview of the proposed system is shown in Figure 1.

Proposed Importance Probing for FSIG. Our intuition for the proposed importance probing is: *"The source GAN kernels have different levels of importance for each target domain."* For example, different subsets of kernels could be important when adapting a pretrained GAN on FFHQ to Babies [32] compared to adapting the same GAN to Cat [10]. Therefore, we aim for a knowledge preservation criteria that is target domain/adaptation-aware (Table 1). In order to achieve adaptation-awareness, we propose a light-weight importance probing algorithm which considers adaptation from source to target domain. There are two important design considerations: probing under (i) extremely limited number of target data and (ii) low computation overhead.

As discussed, in this *importance probing* step, we adapt the source model to the target domain for a limited number of iterations and with a few available target samples. During this short adaptation step, we measure the importance of kernel for the adaptation task. To measure the importance, we use Fisher information (FI) which gives the *informative knowledge* of that kernel in handling adaptation task [44]. Then, based on FI measurement, we classify kernels into important / unimportant. These kernel-level importance decisions are then used in the next step, *i.e.*, main adaptation.

In the main adaptation step, we propose to apply *kernel modulation* to achieve restrained update for the important kernels, and *simple fine-tuning* for the unimportant kernels. As will be discussed, the modulation is rank-constrained and has restricted degree-of-freedom; therefore, it is capable to preserve knowledge of the important kernels. On the other hand, simple fine-tuning has large degree-of-freedom for updating knowledge of the unimportant kernels. Furthermore, the rank-constrained kernel modulation is parameter-efficient. Therefore, we also apply this rank-constrained kernel modulation *in the probing step* to determine the importance of kernels.

Kernel Modulation. The kernel modulation is used in the main adaptation step to preserve knowledge of important kernels into the adapted model. Furthermore, it is also used in the probing step as a parameter-efficient technique to determine importance of kernels. Specifically, we apply Kernel ModuLation (KML) which is proposed very recently [27]. In [27], KML is proposed for multimodal few-shot *classification* (FSC). In particular, in [27], KML has been found to be effective for knowledge transfer between different *classification* tasks of different modes under few-shot constraint. Therefore, in our work, we apply KML for knowledge transfer between different *generation* tasks of different domains under limited target domain samples.

Specifically, in each convolutional layer of a CNN, the ith kernel of that layer $\mathbf{W}_i \in \mathbb{R}^{c_{in} \times k \times k}$ is convolved with the input feature $\mathbf{X} \in \mathbb{R}^{c_{in} \times h \times w}$ to the layer to produce the ith output channel (feature map) $\mathbf{Y}_i \in \mathbb{R}^{h' \times w'}$, *i.e.*, $\mathbf{Y}_i = \mathbf{W}_i * \mathbf{X} + b_i$, where $b_i \in \mathbb{R}$ denotes the bias term. Then, KML [27] modulates $\mathbf{W}_i$ by multiplying it with the modulation matrix $\mathbf{M}_i \in \mathbb{R}^{c_{in} \times k \times k}$ plus an all-ones matrix $\mathbf{J} \in \mathbb{R}^{c_{in} \times k \times k}$:

$$\hat{\mathbf{W}}_i = \mathbf{W}_i \odot (\mathbf{J} + \mathbf{M}_i) \tag{1}$$

where $\odot$ denotes the Hadamard multiplication. In Eqn. 1, using $\mathbf{J}$ allows to learn the modulation matrix in a residual format. Therefore, the modulation weights are learned as perturbations around the pretrained kernels which helps to preserve source knowledge. The exact pretrained kernel can also be transferred to the target model if it is optimal. There are some important differences between discriminative version of KML in [27] and our version, please see Supplementary for details.

This baseline KML learns an individual modulation parameter for each coefficient of the kernel. Therefore, it could suffer from *parameter explosion* when using in recent GAN architectures (*e.g.*, more than 58M parameters in StyleGAN-V2 [8][1]). To address this issue, instead of learning the modulation matrix, we learn a *low-rank* version of it [27, 45]. More specifically, for a Conv layer within CNN, with a total number of d_{out} kernels to be modulated, instead of learning $\mathbf{M} = \{\mathbf{M}_i\}_{i=1}^{d_{out}}$, we learn two proxy vectors $\mathbf{m}_1 \in \mathbb{R}^{d_{out}}$, and $\mathbf{m}_2 \in \mathbb{R}^{(c_{in} \times k \times k)}$, and construct the modulation matrix using the outer product of these vectors, *i.e.*, $\mathbf{M} = \mathbf{m}_1 \otimes \mathbf{m}_2$. Furthermore, as we are using KML for adaptable knowledge preservation, we *freeze* the base kernel $\mathbf{W}_i$ during adaptation. Therefore, trainable parameters are $\mathbf{m}_1, \mathbf{m}_2$. This reduces the number of trainable parameters significantly, and has better performance on restraining the update of important kernels (see Supplementary). As it

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

will be discussed later, the value of d_{out} equals to the total number of kernels in a layer (c_{out}) during probing, and for main adaptation, it is determined by the output of our probing method ($d_{out} \leqslant c_{out}$).

Importance Measurement. Recall our FSIG has two main steps: (i) importance probing step, and (ii) main adaptation step. In probing, we also apply KML as a parameter-efficient technique to determine importance of individual kernels. In particular, for probing, we propose to apply KML to all kernels (in both generator and discriminator) to identify which of the *modulated* kernels are important for the adaptation task. To measure the importance of the modulated kernels, we apply Fisher information (FI) to the modulation parameters. In our FSIG setup, for a modulated GAN with parameters Θ, Fisher information $\mathcal{F}$ can be computed as:

$$\mathcal{F}(\Theta) = \mathbb{E}\big[-\frac{\partial^2}{\partial \mathbf{\Theta}^2}\mathcal{L}(x|\Theta)\big] \tag{2}$$

where $\mathcal{L}(x|\Theta)$ is the binary cross-entropy loss computed using the output of the discriminator, and x includes few-shot target samples, and fake samples generated by GAN. Then, FI for a modulation matrix $\mathcal{F}(\mathbf{M}_i)$ can be computed by averaging over FI values of parameters within that matrix. As we are using the low-rank estimation to construct the modulation matrix, we can estimate $\mathcal{F}(\mathbf{M}_i)$ by FI values of the proxy vectors. In particular, considering the outer product in low-rank approximation, we have $\mathbf{M}_i = reshape([\mathbf{m}_1(i)\mathbf{m}_2(1), \ldots, \mathbf{m}_1(i)\mathbf{m}_2(k)])$, where $k = |\mathbf{m}_2|$. Then we use the unweighted average of FI for parameters of $\mathbf{m}_1$ and $\mathbf{m}_2$, proportional to their occurrence frequency in calculation of $\mathbf{M}_i$, as an estimate of $\mathcal{F}(\mathbf{M}_i)$ (details in Supplementary):

$$\hat{\mathcal{F}}(\mathbf{M}_i) = \mathcal{F}(\mathbf{m}_1(i)) + \frac{1}{|\mathbf{m}_2|}\sum_{j=1}^{|\mathbf{m}_2|}\mathcal{F}(\mathbf{m}_2(j)) \tag{3}$$

After calculating $\hat{\mathcal{F}}(\mathbf{M}_i)$ for all modulation matrices in both generator and discriminator, we use the $t\%$ quantile of these values as a threshold (separately for generator and discriminator) to decide whether modulation of a kernel is important or unimportant for adaptation to the target domain. If the modulation of a kernel is determined to be important (during probing), the kernel is modulated using KML during main adaptation step; otherwise, the kernel is updated using simple fine-tuning during main adaptation. In all setups, we perform probing for 500 iterations and use a fixed $t = 75$. We remark that in probing only modulation parameters $\mathbf{m}_1, \mathbf{m}_2$ are trainable, and FI is only computed on them, the probing is a very lightweight step and can be performed with minimal overhead (details in Supplementary). The output of probing step are the decisions to apply kernel modulation or simple fine-tuning on individual kernels. Then, based on these decisions, the main adaptation is performed.

5 Empirical Studies

5.1 Experiments / Results

Experiment Details. For fair comparison, we strictly follow prior works [1, 2, 3, 4, 5] in the choice of GAN architecture, source-target adaptation setups and hyper-parameters. We use StyleGAN-V2 [8] as the GAN architecture and FFHQ as the source domain. Our experiments include setups with different source-target proximity: Babies/Sunglasses [4], MetFaces [33] and Cat/Dog/Wild (AFHQ) [10] (See Sec. 3). We use 256 x 256 resolution for adaptation. Adaptation is performed with batch size 4 on a single Tesla V100 GPU. We apply importance probing and modulation on base kernels of both generator and discriminator. We focus on 10-shot target adaptation setup in the main paper.

Qualitative Results. We show generated images with our proposed adaptation-aware FSIG method along Baseline [1, 2] and SOTA FSIG methods [3, 4, 5] for 2 target domains, Babies and Cats with different degrees of proximity to FFHQ, before and after adaptation. The results are shown in Figure 4 top and bottom respectively. By preserving source domain knowledge that is important for target domain, our proposed adaptation-aware FSIG method can generate substantially high quality images with high diversity for both Babies and Cat domains. We also include FID and Intra-LPIPS metrics to quantitatively show that our proposed method outperforms SOTA FSIG methods [3, 4, 5]. We show more generated samples in Supplementary.

Figure 4: Qualitative and quantitative comparison of 10-shot image generation with different FSIG methods. Images of each column are from the same noise input (Last row). **Left**: 10 real target images for few-shot adaptation. **Mid, Right**: For target domain with close proximity (*e.g.*Babies, top), our method can generate high quality images with more refined details and diversity knowledge, achieving best FID and Intra-LPIPS socre. For target domain which is distance (*e.g.*, Cat, bottom), TGAN/FreezeD overfit to the 10-shot samples and others fail. In contrast, our method preserves meaningful semantic features at different levels (*e.g.*, posture and color) from source, achieving a good trade off between quality and diversity. In particular, our Intra-LPIPS approaches that of EWC, while our generated images have much better quality qualitatively and quantitatively.

Quantitative Results. We show complete FID scores in Table 2. Our proposed adaptation-aware FSIG achieves SOTA results across all target domains of varying proximity to the source (FFHQ). We emphasize that it is achieved by preserving source domain knowledge that is important for target domain adaptation (Sec 4). We also report Intra-LPIPS (↑) as an indicator of diversity, as Figure 4.

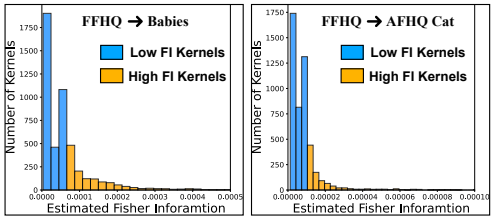
5.2 Analysis

Ablation study of Importance Probing. The goal of importance probing (denoted as "IP") is to identify kernels that are important for *few-shot target adaptation* as shown in Figure 5 (Top). To justify the effectiveness of our design choice, we perform an ablation study that discards the IP stage and regard all kernels as *equally important* for target adaptation. Therefore, we simply modulate all kernels *without any knowledge selection*. As one can observe from Figure 5 (Bottom), knowledge selection plays a vital role in adaptation performance. Specifically, the significance of knowledge preservation is more evident when the target domains are distant from the source domain.

Number of target samples (shots). The number of target domain training samples is an important factor that can impact the FSIG performance. In general, more target domain samples can allow better estimation of target distribution. We study the efficacy of our proposed method under different

Table 2: FSIG (10-shot) results: We report FID scores (↓) of our proposed *adaptation-aware* FSIG and compare with existing FSIG methods. We emphasize that Cat, Dog and Wild target domains are additional experiments included in this work. (Sec 3.1). Our experiment results show two important findings: **1)** Under setups which assumption of close proximity between source and target domains is relaxed (Cat, Dog, Wild), SOTA FSIG methods – EWC, CDC, DCL – which consider only source domain in knowledge preserving perform *no better* than a baseline fine-tuning method (TGAN). **2)** Our proposed adaptation-aware FSIG achieves SOTA performance in *all* target domains due to preserving source domain knowledge that is important for few-shot target domain adaptation. We generate 5,000 images using the adapted generator to evaluate FID on the whole target domain. We also report the corresponding KID, Intra-LPIPS and standard deviations in Supplementary.

Target Domain	**Babies** [8]	**Sunglasses** [8]	**MetFaces** [8]	**Cat** [10]	**Dog** [10]	**Wild** [10]
TGAN [1]	101.58	55.97	76.81	64.68	151.46	81.30
FreezeD [2]	96.25	46.95	73.33	63.60	157.98	77.18
TGAN+ADA [33]	97.91	53.64	75.82	80.16	162.63	81.55
EWC [3]	79.93	49.41	62.67	74.61	158.78	92.83
CDC [4]	69.13	41.45	65.45	176.21	170.95	135.13
DCL [5]	56.48	37.66	62.35	156.82	171.42	115.93
Ours	**48.83**	**28.03**	**51.34**	**58.07**	**100.91**	**36.87**



Target Domain	**Babies**	**Sunglasses**	**MetFaces**	**Cat**	**Dog**	**Wild**
Ours (w/o probing)	54.46	33.66	60.43	82.41	160.87	81.24
Ours	**48.83**	**28.03**	**51.34**	**58.07**	**100.91**	**36.87**

Figure 5: **(Top)** Our proposed IP identifies and preserves source kernels important (high FI) for target adaptation. **(Bottom)** We validate the effectiveness of IP by modulating all kernels without IP. On the other hand, if we fine-tune all parameters without IP and modulation (TGAN), it suffers mode collapse (Table 2 and Figure 4).

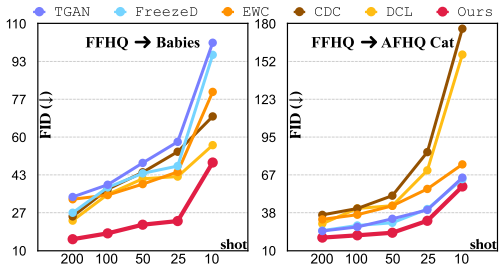


Figure 6: We evaluate the performance of different number of shots (10, 25, 50, 100, 200) on Babies **(Left)** and Cat **(Right)**. We show that our method consistently outperforms other FSIG methods in all setups. In Supplementary, we also show the generated images given different number of shots on more target domains.

number of target domain samples. The results are shown in Figure 6, and we show that our proposed adaptation-aware FSIG method consistently outperforms existing FSIG methods in all setups.

6 Discussion

Conclusions. Focusing on FSIG, we make two contributions. First, we re-visit current SOTA methods and their experiments. We discover that SOTA performs poorly in setups when source and target domains are more distant, as existing methods consider only source domain/task for knowledge preservation. Second, we propose a new FSIG method which is target/adaptation-aware. Our proposed method outperforms previous work across all setups of different source-target domain proximity.

Broader Impact. Our work makes contribution to generation of synthetic data in applications where sample collection is challenging, e.g., photos of rare animal species. This is an important contribution to many data-centric applications. Furthermore, transfer learning of generative models using a few data sample enables data and computation-efficient model development. Our work has positive impact on environmental sustainability and reduction of greenhouse gas emission.

Limitations. While our experiments are extensive compared to previous works, in practical applications, there are many possible target domains which cannot be included in our experiments. However, as our method is target/adaptation aware, we believe our method can generalize better than existing SOTA which are target-agnostic.

References

[1] 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.

[2] Sangwoo Mo, Minsu Cho, and Jinwoo Shin. Freeze the discriminator: a simple baseline for fine-tuning gans. In *CVPR AI for Content Creation Workshop*, 2020.

[3] Yijun Li, Richard Zhang, Jingwan (Cynthia) Lu, and Eli Shechtman. Few-shot image generation with elastic weight consolidation. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 15885–15896. Curran Associates, Inc., 2020.

[4] 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.

[5] Yunqing Zhao, Henghui Ding, Houjing Huang, and Ngai-Man Cheung. A closer look at few-shot image generation. In *CVPR*, 2022.

[6] 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.

[7] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *International Conference on Learning Representations*, 2018.

[8] 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.

[9] 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*, pages 4401–4410, 2019.

[10] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.

[11] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[12] Ming-Yu Liu, Xun Huang, Arun Mallya, Tero Karras, Timo Aila, Jaakko Lehtinen, and Jan Kautz. Few-shot unsupervised image-to-image translation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10551–10560, 2019.

[13] Ji Lin, Richard Zhang, Frieder Ganz, Song Han, and Jun-Yan Zhu. Anycost gans for interactive image synthesis and editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14986–14996, 2021.

[14] Hongyu Liu, Ziyu Wan, Wei Huang, Yibing Song, Xintong Han, Jing Liao, Bin Jiang, and Wei Liu. Deflocnet: Deep image editing via flexible low-level controls. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10765–10774, 2021.

[15] Ngoc-Trung Tran, Viet-Hung Tran, Ngoc-Bao Nguyen, Trung-Kien Nguyen, and Ngai-Man Cheung. On data augmentation for gan training. *IEEE Transactions on Image Processing*, 30:1882–1897, 2021.

[16] Antreas Antoniou, Amos Storkey, and Harrison Edwards. Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340*, 2017.

[17] Lucy Chai, Jun-Yan Zhu, Eli Shechtman, Phillip Isola, and Richard Zhang. Ensembling with deep generative views. In *CVPR*, 2021.

[18] Qianli Feng, Chenqi Guo, Fabian Benitez-Quiroz, and Aleix M Martinez. When do gans replicate? on the choice of dataset size. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6701–6710, 2021.

[19] 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.

[20] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[21] 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.

[22] Jordan Yaniv, Yael Newman, and Ariel Shamir. The face of art: landmark detection and geometric style in portraits. *ACM Transactions on graphics (TOG)*, 38(4):1–15, 2019.

[23] Alexander Ly, Maarten Marsman, Josine Verhagen, Raoul PPP Grasman, and Eric-Jan Wagenmakers. A tutorial on fisher information. *Journal of Mathematical Psychology*, 80:40–55, 2017.

[24] 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.

[25] Li Fei-Fei, Rob Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence*, 28(4):594–611, 2006.

[26] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in neural information processing systems*, pages 4077–4087, 2017.

[27] Milad Abdollahzadeh, Touba Malekzadeh, and Ngai-Man Man Cheung. Revisit multimodal meta-learning through the lens of multi-task learning. *Advances in Neural Information Processing Systems*, 35, 2021.

[28] Weide Liu, Chi Zhang, Guosheng Lin, and Fayao Liu. Crnet: Cross-reference networks for few-shot segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4165–4173, 2020.

[29] Malik Boudiaf, Hoel Kervadec, Ziko Imtiaz Masud, Pablo Piantanida, Ismail Ben Ayed, and Jose Dolz. Few-shot segmentation without meta-learning: A good transductive inference is all you need? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13979–13988, 2021.

[30] Gongjie Zhang, Kaiwen Cui, Rongliang Wu, Shijian Lu, and Yonghong Tian. Pnpdet: Efficient few-shot detection without forgetting via plug-and-play sub-networks. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3823–3832, 2021.

[31] Zhibo Fan, Yuchen Ma, Zeming Li, and Jian Sun. Generalized few-shot object detection without forgetting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4527–4536, 2021.

[32] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *Advances in Neural Information Processing Systems*, 33:12104–12114, 2020.

[33] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *Advances in Neural Information Processing Systems*, 33:12104–12114, 2020.

[34] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[35] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.

[36] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[37] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.

[38] Hossein Talebi and Peyman Milanfar. Learned perceptual image enhancement. In *2018 IEEE international conference on computational photography (ICCP)*, pages 1–13. IEEE, 2018.

[39] Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. *IEEE transactions on image processing*, 27(8):3998–4011, 2018.

[40] Stanislav Morozov, Andrey Voynov, and Artem Babenko. On self-supervised image representations for {gan} evaluation. In *International Conference on Learning Representations*, 2021.

[41] 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.

[42] David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B Tenenbaum, William T Freeman, and Antonio Torralba. Gan dissection: Visualizing and understanding generative adversarial networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2019.

[43] David Bau, Bolei Zhou, Aditya Khosla, Aude Oliva, and Antonio Torralba. Network dissection: Quantifying interpretability of deep visual representations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6541–6549, 2017.

[44] Alessandro Achille, Michael Lam, Rahul Tewari, Avinash Ravichandran, Subhransu Maji, Charless C Fowlkes, Stefano Soatto, and Pietro Perona. Task2vec: Task embedding for meta-learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6430–6439, 2019.

[45] Christian Simon, Piotr Koniusz, Richard Nock, and Mehrtash Harandi. On modulating the gradient for meta-learning. In *European Conference on Computer Vision*, pages 556–572. Springer, 2020.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] *See Sections 3, 4, 5 and more details in Supplementary.*
   (b) Did you describe the limitations of your work? [Yes] *See Section 6*
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] *See Supplementary*
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] *We include code, data URL, pre-trained GAN models and reproducibility details in Supplementary.*
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] *We specify all training details in Supplementary.*
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] *We report standard deviations in Supplementary.*
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] *We include all compute details in Supplementary.*
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] *We have cited all original papers and included URLs for the code and datasets used.*
   (b) Did you mention the license of the assets? [Yes] *We mention the license of the assets in Supplementary.*
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] *We are releasing code and pre-trained GAN models. The URL details are included in Supplementary.*

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A] *All assets are publicly available and contain no personally identifiable information or offensive content.*

5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]