Exploring Incompatible Knowledge Transfer in Few-shot Image Generation

Yunqing Zhao[1] Chao Du[2] Milad Abdollahzadeh[1] Tianyu Pang[2]
Min Lin[2] Shuicheng Yan[2] Ngai-Man Cheung[1]
[1] Singapore University of Technology and Design (SUTD) [2] Sea AI Lab
yunqing_zhao@mymail.sutd.edu.sg, ngaiman_cheung@sutd.edu.sg

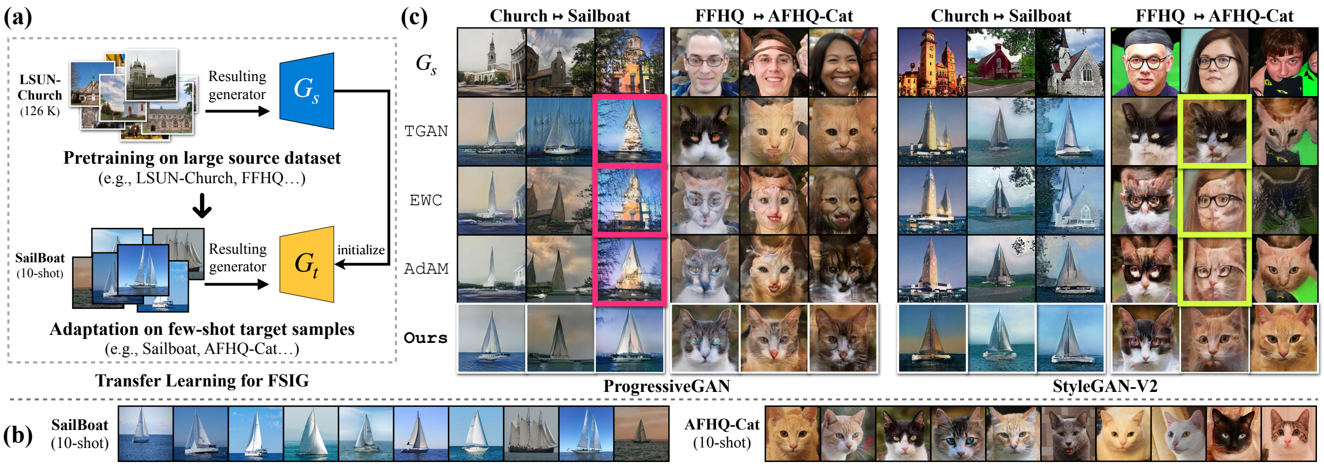


Figure 1. **In FSIG, transfer of incompatible knowledge would cause significant degradation of the realisticness of synthetic samples, and existing SOTA methods fail to address this issue.** **(a)** Prior knowledge of the source GAN generator G_s is selected, preserved, and transferred to learn the target generator G_t, given very limited target domain training samples (see 10-shot in **(b)**). In this work, we consider challenging setups where the target domain could be *semantically distant* from the source domain, *e.g.*, (i) `Church` → `Sailboat` (ii) `FFHQ` → `AFHQ-Cat`. **(c)** Synthetic images from G_s, TGAN [65]; and SOTA methods: EWC [41] and AdAM [75]. Images in each column are from the same noise input. **Critically**, existing FSIG methods focus on *knowledge preservation*, while failing to prevent the transfer of knowledge that is incompatible with the target domain. This gives rise to unrealistic samples generated by G_t, *e.g.*, "Trees/Buildings on the Sea" (**Red** frames), or "Cats with Glasses" (**Green** frames). In contrast, our method (row-5 in **(c)**) based on *knowledge truncation* can effectively prevent the transfer of incompatible knowledge. See our detailed analysis in Sec. 4. **Best viewed in color with zooming in.**

Abstract

Few-shot image generation (FSIG) learns to generate diverse and high-fidelity images from a target domain using a few (e.g., *10) reference samples. Existing FSIG methods select, preserve and transfer prior knowledge from a source generator (pretrained on a related domain) to learn the target generator. In this work, we investigate an underexplored issue in FSIG, dubbed as incompatible knowledge transfer, which would significantly degrade the realisticness of synthetic samples. Empirical observations show that the issue stems from the least significant filters from the source generator. To this end, we propose knowledge truncation to mitigate this issue in FSIG, which is a complementary operation to knowledge preservation and is implemented by a lightweight pruning-based method. Extensive experiments show that knowledge truncation is simple and effective, consistently achieving state-of-the-art performance, including challenging setups where the source and target domains are more distant. Project Page: yunqing-me.github.io/RICK.*

1. Introduction

Over recent years, deep generative models [14,15,21,30,52,78] have made tremendous progress, enabling many intriguing tasks such as image generation [5,27–29], image editing [45,63,68,80], and data augmentation [6,12,59]. In spite of their remarkable success, most research on generative models has been focusing on setups with sizeable training datasets [24,26,59,60,69,74], limiting its applications in many domains where data collection is difficult or expensive [12,13,46] (*e.g.*, medicine).

To address such problems, FSIG has been proposed recently [41,50], which learns to generate images with extremely few reference samples (*e.g.*, 10 samples) from a target domain. In these regimes, learning a generator to capture the underlying target distribution is an undetermined problem that requires some prior knowledge. The majority of existing state-of-the-art (SOTA) FSIG methods rely on transfer learning approaches [2, 23, 65, 71] to exploit prior knowledge (*e.g.*, a source generator) learned from abundant data of a different but related source domain, and then transfer suitable source knowledge to learn the target generator with fine-tuning [9, 26, 41, 48–50, 64, 65, 74, 75, 77]. Different techniques have been proposed to effectively preserve useful source knowledge, such as freezing [48], regularization [41, 50, 77] and modulation [75] (details in Sec. 2).

Incompatible knowledge transfer. Despite the impressive improvement achieved by different knowledge preservation approaches [41, 48, 50, 75, 77], in this work, we argue that preventing *incompatible knowledge transfer* is equally crucial. This is revealed through a carefully designed investigation, where such incompatible knowledge transfer is manifested in the presence of unexpected semantic features. These features are inconsistent with the target domain, thereby degrading the realisticness of synthetic samples. As illustrated in Figure 1, trees and buildings are incompatible with the domain of `Sailboat` (as can be observed by inspecting the 10 reference samples). However, they appear in the synthetic images when applying the existing SOTA methods [41, 75] with a source generator trained on `Church`. This shows that the existing methods cannot effectively prevent the transfer of incompatible knowledge.

Knowledge truncation. Based on our observations, we propose **R**emoving **I**n-**C**ompatible **K**nowledge (RICK), a lightweight filter-pruning based method to remove filters that encode incompatible knowledge (*i.e.*, filters with least estimated importance for adaptation) during FSIG adaptation. While filter pruning has been applied extensively to achieve compact deep networks with reduced computation [19, 57, 66], its application to prevent transfer of incompatible knowledge is underexplored. We note that our proposed knowledge truncation and pruning of incompatible filters are orthogonal and complementary with existing knowledge preservation methods in FSIG. In this way, our method effectively removes the incompatible knowledge compared to prior works, and achieves noticeably improved quality (*e.g.*, FID [20]) of generated images.

Our contributions can be summarized as follows:

- We explore the incompatible knowledge transfer for FSIG, reveal that SOTA methods fail in handling this issue, investigate the underlying causation, and disclose the inadequacy of fine-tuning in removal of incompatible knowledge.
- We propose knowledge truncation to alleviate incompatible knowledge transfer, and realize it with a lightweight filter-pruning based method.
- Extensive experiments show that our method effectively removes incompatible knowledge and consistently improves the generative quality, including challenging setups where source and target domains are dissimilar.

2. Related Works

Transfer learning (TL) [33, 51, 81] is a widely used approach to improve the performance of a model in a data-limited target domain by leveraging the knowledge of a model pretrained on a data-rich source domain [10, 38, 79]. Conventionally, TL has been applied to predictors [35–37], including image classifiers [2, 7, 23, 58, 71, 76], or object detectors [16, 34, 54]. The primary focus of TL has been on selecting and preserving useful knowledge of the source model into the target model [51, 81]. For example, [32, 40, 54, 73] preserve and transfer generalizable layers of a source network into the target network. Recent work searches for useful features in the entire source network, preserves them in the target network, and trains a linear classification head on top of the preserved features [11].

Besides predictive models, TL has been applied to generative models recently for FSIG [39, 48, 50, 75, 77], where they adapt a GAN pretrained on a large source dataset as initialization, and perform adaptation on very limited target training domains. For example, baseline methods such as TGAN [65] simply fine-tune the pretrained model via GAN loss (see Eqn. (1)). Recent state-of-the-art methods propose to preserve some knowledge for adaptation. For example, FreezeD [48] fixes some low-level layers of the discriminator for adaptation; EWC [41] identifies the important parameter of a source task and penalizes the weights change; CDC [50] aims to preserve the consistency of distance between generated images before and after adaptation; DCL [77] maximizes the mutual information between generated images on source and target from the same input latent code to preserve the knowledge. More recently, AdAM [75] proposes a modulation based method to identify the source knowledge important for the target domain, and preserves the knowledge for adaptation.

3. Preliminaries

Existing FSIG methods adopt TL approach and leverage a source GAN pretrained on a large source dataset. We denote the source generator as G_s (source discriminator as D_s). During adaptation, the target generator G_t (target discriminator as D_t) is obtained by *fine-tuning* the source GAN on few-shot target images via adversarial loss $\mathcal{L}_{adv}$ [15]:

$$\min_{G_t} \max_{D_t} \mathcal{L}_{adv} = \mathbb{E}_{x\sim p_{data}(x)}[\log D_t(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D_t(G_t(z)))], \quad (1)$$

where z is a 1-D latent code sampled from noise distribution $p_z(z)$ (*e.g.*, Gaussian), and $p_{data}(x)$ denotes the few-shot target data distribution. Note that source data is inaccessible. In fine-tuning, the weights of G_s (and D_s) are used to initialize G_t (and D_t). See Figure 1(a). The main goal of FSIG is to learn G_t to capture $p_{data}(x)$.

To alleviate mode collapse due to very limited target samples, recent methods augment fine-tuning with knowledge preservation to carefully select and preserve subset of source knowledge during adaptation, *e.g.*, freezing [48, 70], regularization [41, 50, 77] and modulation [75] based methods. The aim of these methods is to preserve knowledge that is deemed to be useful for target generator, *e.g.*, improving the diversity of target sample generation [50]. For knowledge that is deemed to be less useful, fine-tuning using Eqn. (1) is applied as a common practice to update such knowledge during adaptation.

4. Incompatible Knowledge Transfer in FSIG

In this section, as our first contribution, we observe and identify the unnoticed issue of incompatible knowledge transfer in existing FSIG methods, and reveal that fine-tuning based knowledge update is inadequate to remove incompatible knowledge after adaptation.

To support our claim and figure out the root cause of the observed incompatible knowledge transfer, we apply GAN dissection [3, 4], a framework that can identify the correspondence between filters and the semantic segmentation of a particular object class (*e.g.*, tree) across different images, to disclose filters that retain incompatible knowledge after fine-tuning. Overall, our main findings establish that existing SOTA methods fail to address this issue and uncover the root cause of incompatible knowledge transfer.

4.1. Investigating incompatible knowledge

Prior SOTA FSIG methods [41, 50, 75, 77] propose different knowledge preservation criteria to select pretrained source knowledge for few-shot adaptation. The adaptation is typically done by fine-tuning the source generator (via Eqn. (1)) with the few-shot target samples. An assumption in these methods is that fine-tuning can adapt the source generator to the target one such that the irrelevant and incompatible source knowledge can be dropped or updated.

In this work, we show that the assumption becomes invalid in the cases where the source and target domains are semantically distant (*e.g.*, `Human Face` → `Cat Face` in Figure 1), where the incompatible knowledge transfer severely hurts the realisticness of the generated images. We note that this has not been well studied in prior SOTA FSIG works as they mainly focus on *knowledge preservation* from the source (see Sec. 2), while little attention has been paid to incompatible knowledge transfer with fine-tuning based *knowledge update*.

In convolutional neural networks, each filter can be viewed as an encoding of a specific part of knowledge [3,4]. Intuitively, in generative models, such knowledge could be either low-level textures (*e.g.*, fur) or high-level human-interpretable concepts (*e.g.*, eyes). Therefore, we hypothesize that clues for incompatible knowledge transfer could be found by attending to filters of the generator. Recently, AdAM [75] proposes an importance probing (IP) method to determine if a source GAN filter is important for adaptation and achieves impressive performance. We adopt IP to evaluate the source generator filter importance for target domain adaptation in our analysis (we include a brief introduction of IP in Supplement). We propose two experiments at different granularity:

Exp-1: Generate images with fixed generator input. We visualize the generated images via different methods. To understand the knowledge transference before and after adaptation, we use the *same noise* as input to source and target generator. Conceptually, this provides us an intuitive and direct comparison of knowledge transference.

Exp-2: Dissect pretrained and adapted generator. To find the filters that are mostly correlated to a specific type of knowledge across different images (*e.g.*, source features that are incompatible to the target) and track their transference before and after adaptation, we label G_s filters with the estimated importance (via IP [75]) and apply GAN dissection [4] to visualize the semantic features corresponding to the *same filters* to G_s and G_t.

These experiments could help us understand the knowledge transference before and after adaptation at both gross granularity (visualization of generated images in pixel space) and fine granularity (dissection of G_s and G_t in filter space). Next, we discuss the setups and results.

4.2. Experiment setups

Model and dataset: In Exp-1, we use ProgressiveGAN ("ProGAN") [25] and StyleGAN-V2 [29] as GAN architectures. FFHQ [28], LSUN-Church [72] are source domains; 10-shot AFHQ-Cat [8] and Sailboat are target domains. Since Exp-2 is limited by the GAN architecture and segmentation model/dataset used in the original dissection work [4], that are more suitable for scene-based scenarios [72], we only use ProGAN with LSUN-Church [72] as the source domain and 10-shot Sailboat as target domain. Nevertheless, we remark that our analysis applies to a wide range of GAN architectures and domains. In all experiments, we use resolution 256 x 256 for adaptation.

Evaluation methods: G_s is the source generator. In Exp-1, we evaluate the baseline method, TGAN [65], and recent SOTA, EWC [41], AdAM [75]. In Exp-2, we use AdAM [75] for dissection (since we use IP to evaluate source filter importance for target adaptation). We follow AdAM [75] to use Fisher Information (FI) [47] as importance measurement in IP. The dissection results of other methods, *e.g.*, EWC [41], are in Supplement.

4.3. Results and analysis

We reveal that existing SOTA FSIG methods with a focus on source knowledge preservation lead to the transfer of in-

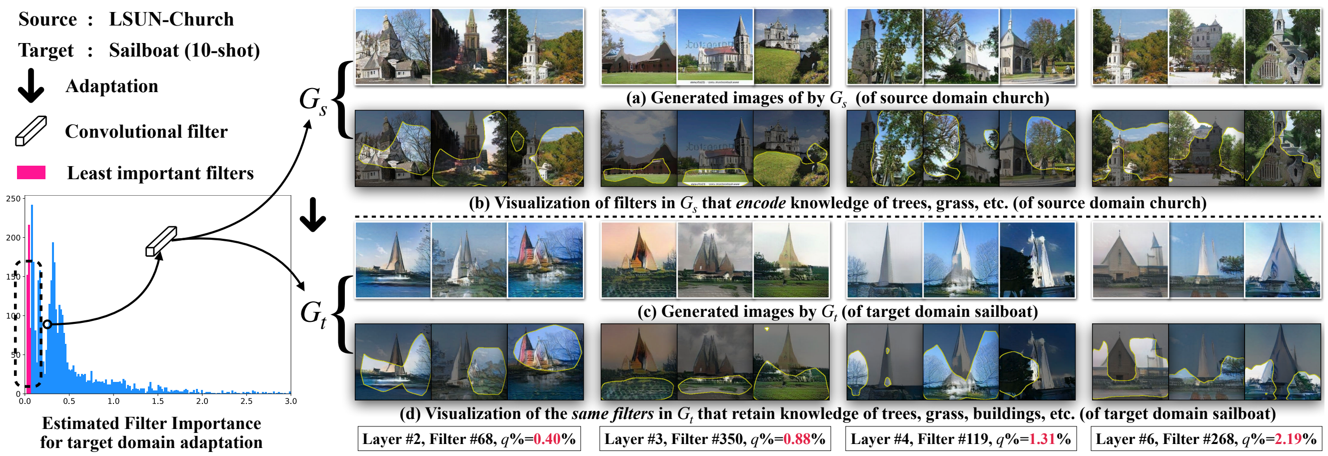


Figure 2. **Investigating the cause of incompatible knowledge transfer.** Since in Figure 1 we observe that *fine-tuning* commonly used in SOTA methods [41,75] is inadequate for preventing incompatible knowledge transfer, we apply GAN dissection [4] to identify interpretable filters whose feature maps are highly correlated to the region of an object class (*e.g.*, trees) across different images. We discover that the incompatible knowledge (*e.g.*, trees, grass and buildings to Sailboat domain in this example) is correlated to the filters of G_s that are deemed to be unimportant/irrelevant for target domain (estimated via IP [75]). We use the quantile value $q\%$ to indicates the filter importance compared to all filters in G_s. In [75], fine-tuning is applied during adaptation to update these low importance filters. Surprisingly, we observe that similar knowledge (tree, building, grass) remains in the *same filters* in G_t after fine-tuning. As this knowledge is incompatible to the target domain, it degrades the realisticness of synthetic images substantially. Additional examples are in Supplement.

compatible knowledge. More importantly, the *root cause* of such incompatible knowledge transfer is the least important filters in G_s determined to be irrelevant to the target domain adaptation, and fine-tuning is not adequate for removing the incompatible knowledge after adaptation. Specifically, we summarize our observations in Figure 1 and Figure 2:

Observation 1: In Figure 1 **(c)**, we visualize the generated images by different methods with fixed noise input. Interestingly, features that are incompatible for the target domain are indeed transferred after adaptation with different knowledge preservation criteria, *e.g.*, "tree on sea" where "tree" is from Church domain, and "Cat with glasses" where "glasses" is from FFHQ domain. All these incompatible source features severely curtail the realism of generated target images. Similar observation can be made on TGAN [65], *i.e.*, simple fine-tuning based method without explicit knowledge preservation. In contrast, our method (we discuss it in Sec. 5) can address this issue.

Observation 2: In Figure 2, we dissect and visualize the incompatible features observed in Figure 1, and find their mostly correlated filters in G_s and G_t. Surprisingly, we find that the filters in G_s identified with *least* importance for target domain are *mostly* relevant to incompatible features transferred from the source, which is the root cause of degradation of realism of generated images. After adaptation, the same filters will still cause the same type of incompatible features, and fine-tuning for knowledge update cannot effectively address this issue. When the target domains become distant, this observation is more obvious.

Our analysis uncovers an important issue of existing SOTA FSIG methods: source domain features that are incompatible to the target domain are transferred to the target generator, with various knowledge preservation criteria. The incompatible knowledge from the source is highly correlated to filters that are deemed to be irrelevant to the target domain, and fine-tuning based knowledge update cannot effectively address this issue. This motivates us to remove the incompatible source features during FSIG adaptation.

5. Proposed Method

Based on our analysis and observation in Sec. 4, we argue that **R**emoving **I**n-**C**ompatible **K**nowledge (**RICK**) for the target domain is similarly important as preserving knowledge useful for the target domain to achieve improved quality of generated samples. In contrast to most prior works that only propose different knowledge preservation criteria, *e.g.*, freezing [48], regularization [41, 50, 77] or modulation [75], we propose knowledge truncation, a novel and complementary concept in existing methods for FSIG. Overall, our proposed method, named RICK, is summarized in Figure 3, and Algorithm 1 in Supplement.

5.1. Knowledge truncation via network pruning

Pruning [17,22,57,66] has been one of the useful tools to achieve a compact neural network with comparable performance to a larger, entire model. Early efforts on compacting networks focus on model acceleration [18,19], inference efficiency [42], and deployment [44, 62], which target at the discriminative tasks, *e.g.*, image classification [55] and ma-

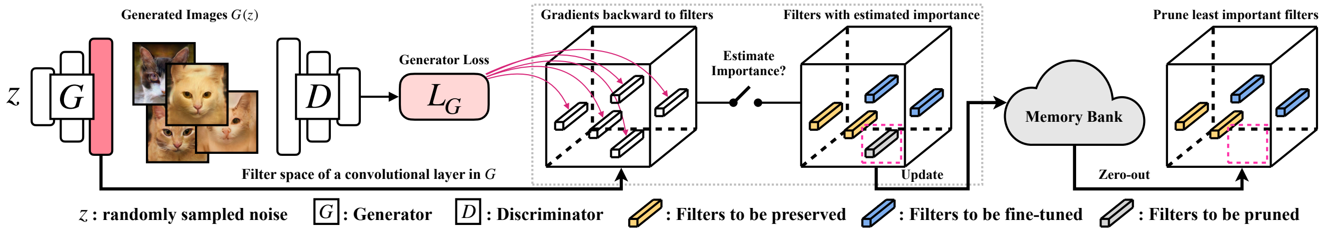


Figure 3. **Overview of the proposed method.** To **R**emove **I**n-**C**ompatible **K**nowledge during target adaptation, we propose *knowledge truncation*, a novel concept for FSIG via pruning filters that are deemed with least importance for target domain adaptation. During training, we estimate the importance of filters for the target domain every certain iterations. After that, we apply a fixed threshold to determine whether a filter should be pruned, and such decision will be maintained in a lightweight memory bank that is updated regularly upon importance estimation. Similar to some prior works [41, 75, 77] that preserve useful source knowledge for adaptation, we preserve filters that are deemed to be important to the target domain by freezing them, and we fine-tune the rest of filters via Eqn. (1) for adaptation.

chine translation [43], often by removing least important neurons (the definition of importance could be various and is discussed in Sec. 5.2). In contrast to prior works on network pruning that pursue the sparsity in the model, we aim to improve the quality of generated images via *removing least important filters relevant to incompatible knowledge to the target domain*, particularly, in FSIG tasks.

Our proposed method contains two major steps: 1) a lightweight filter importance estimation on-the-fly during adaptation; and 2) determine operations to filters based on their estimated importance. In step 1), we leverage the gradient information during adaptation to evaluate the filter importance for target adaptation every certain iterations. Then in step 2), based on the estimated filter importance, we prune the filters with least importance, which are deemed to be irrelevant to the target domain to remove the incompatible knowledge for adaptation. Meanwhile, we preserve the filters with high importance to achieve knowledge preservation in FSIG, and fine-tune the rest of filters to let the source generator adapt to the target domain.

Proposed filter importance estimation. We estimate the importance of each filter by leveraging the on-the-fly gradient information during FSIG adaptation. We denote a filter as $\mathbf{W} \in \mathbb{R}^{c^{in} \times k \times k}$, where k is the spatial size of the filter and c^{in} is the dimension (number) of the input feature maps. We use Fisher Information (FI) [47] as importance estimator for each filter $\mathcal{F}(\mathbf{W})$ (to be further discussed in Sec. 5.2) that could tell quantitative information of compatibility between filter weights and the FSIG task [1, 75]:

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

where $\mathcal{L}_G$ is the binary cross-entropy loss computed with output from the discriminator. x denotes a set of generated images. In practice, we use first-order approximation of FI [1] to lower the computational cost.

Our filter importance estimation for knowledge selection is lightweight and highly efficient: compared to prior SOTA methods that propose different knowledge selection criteria (though they only focus on knowledge preservation), our method does not require external models to provide additional information during adaptation [50, 77], nor introduce additional learnable parameters and pre-adaptation iterations for importance estimation [75], and it takes benefits from the output of G_t and D_t during training.

Proposed knowledge truncation via filter pruning. In Sec. 4, we have shown rich evidence that least important filters are relevant to semantic features incompatible to the target domain (*e.g.* “Tree on sea” or “Building structure on sea”). Importantly, given different knowledge preservation criteria, fine-tuning based knowledge update cannot properly remove incompatible knowledge after adaptation. Therefore, we propose a simple and novel method for knowledge truncation via pruning (zeroing-out) the filters with least importance for adaptation.

Specifically, after the estimation of filer importance in step 1), for the i-th filter $\mathbf{W}^i$ in the network, we apply a threshold ($q\%$, *i.e.*, the quantile of its importance compared to all filters) to determine whether $\mathbf{W}^i$ should be pruned :

$$\mathbf{W}^i \leftarrow \mathbf{0}, \text{ if } \mathcal{F}(\mathbf{W}^i) < q\% \tag{3}$$

We remark that, once a filter is determined to be pruned, it will no longer be involved in training/inference and will not be recovered in the rest of training iterations. The knowledge truncation is applied to both generator and discriminator, and we use separate thresholds to G_t and D_t. Since we regularly estimate the filter importance during adaptation and the “non-recoverable” attribute of pruned filters, the amount of zeroized filters using Eqn. (3) will accumulate to a specific value $p\%$, at the end of the adaptation.

Similar to prior works that focus on knowledge preservation [41, 50, 75, 77] and propose different knowledge selection criteria, we preserve the filters with high estimated importance for adaptation via freezing the filters during training. For the rest of filters, we simply let them fine-tune using Eqn. (1). Whether the filter needs to be fine-tuned or preserved depends dynamically on it’s importance for target. We discuss the effect of selecting high importance filters in Supplement. Since we estimate the filter importance multiple times during adaptation, the operations to a specific filter may change after different evaluations, except the case that the filter is pruned and will not be recovered.

5.2. Design choice

Here we discuss the design choice of our proposed method and adopted importance measurement. Since we dynamically evaluate the filter importance every certain iterations, we need to maintain the operation to each filter (could be "preserve", "fine-tune" or "prune") until the next estimation. To lower the compute cost, we maintain the determination of operation to each filter (obtained via estimated filter importance) in a lightweight memory bank $\mathcal{M}$: for each high dimension filter $\mathbf{W} \in \mathbb{R}^{c^{in} \times k \times k}$, we only need a single character to record the corresponding operation in $\mathcal{M}$. For example, for StyleGAN-V2 [29] used in main experiments whose generator contains ~30M parameters [1], $\mathcal{M}$ is a one dimension array with $\mathsf{size}(\mathcal{M}) \sim$ 5,000.

Similar to prior works [1, 41, 75], we use fisher information [31] as importance measurement to estimate how well a network parameter (filter in our work) does on the adaptation task [1]. We note that there are alternative measurements to estimate the filter importance for adaptation, *e.g.*, class salience [56] or reconstruction loss [41]. In Supplement, we conduct a study and empirically find that we can achieve similar performance as FI. Moreover, in Sec. 6.2, we surprisingly find that even without pruning (*i.e.*, a filter can only be preserved or fine-tuned), our proposed method can still achieve competitive performance compared to SOTA methods, which implies the effectiveness of the proposed dynamic importance estimator.

6. Experiments

Basic setups. For fair comparison, we strictly follow the experiment setups as previous works [41, 50, 77], *e.g.* in the choice of source and target domains and few-shot target samples. We employ StyleGAN-V2 [29] as the GAN architecture for pretraining and adaptation in main experiments, similar to previous works [50, 67, 70, 75, 77]. We train our models with on an NVIDIA A100 PCIe 40GB. We include more implementation details in Supplement.

Datasets and baseline methods. We use the GAN pretrained on FFHQ [28] and the target domains that have different proximity to the source dataset: semantically related domains include Babies [50], Sunglasses [50] and MetFaces [26] (we note that this is the common setup in prior works for FSIG); distant domains include AFHQ-Cat, AFHQ-Dog and AFHQ-Wild [8] (we note that this setup is more challenging compared to most prior works [41, 50, 77]). Adaptation on more target domains are in Supplement for comprehensive analysis. We use images with resolution 256 x 256 for adaptation. Baseline methods used for comparison are introduced in Sec. 2. We additionally compare with ADA [26], a popular baseline approach that applies data augmentation during adaptation.

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

6.1. Performance evaluation and comparison

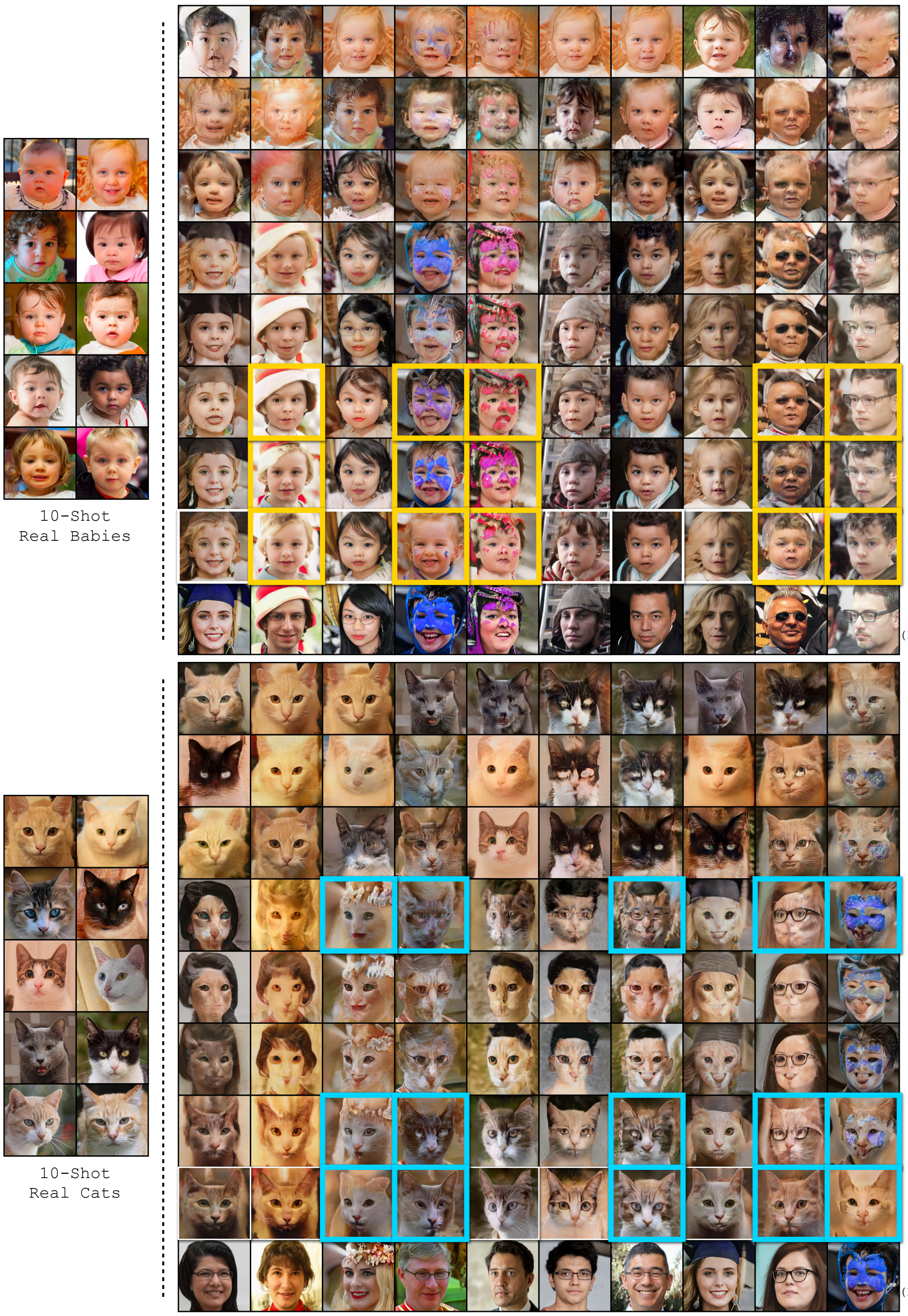
Evaluation measurements. We adopt three types of evaluation methods in the main paper. The popular metric Fréchet Inception Distance (FID ↓) [20] measures the distance between the fitted Gaussian distribution of the real and generated data. We also evaluate the diversity of the generated images using intra-LPIPS (↑) [50], the perceptual distance of generated images to few-shot target samples (we include the pseudo-code in Supplement). Finally, we visualize the generated images and compare with different methods, often with a fixed noise input for fair comparison.

Qualitative results. In Figure 4, we visualize the generated images by different methods before and after adaptation for comparison. In each column, images are generated from the same noise input. We use FFHQ [28] as the source domain. Babies [50] and AFHQ-Cat [8] are target domains with different semantic proximity to the source. We show that, our proposed method, while preserving useful source knowledge, reliably removes the incompatible knowledge to the target, and therefore achieves improved quality of generated images. More results are in Supplement.

Quantitative results. Considering that the whole target dataset often contains about 5,000 images (*e.g.*, AFHQ-Cat [8]), following prior works [50, 75, 77] we randomly generate 5,000 images using the adapted generator and compare with the entire target dataset to compute FID (↓). In Table 1, we show complete FID results over six benchmark datasets. In Figure 4, we also compute intra-LPIPS (↑) as diversity measurement over 10-shot target samples and we report the FID using the same checkpoint. All these results show the effectiveness of our proposed method.

6.2. Discussion

Knowledge truncation with different methods. Ideally, our proposed concept of knowledge truncation for FSIG can be applied to different methods, as long as we can estimate the parameter importance (*e.g.*, filter importance in our method). In literature, EWC [41] and AdAM [75] proposed different methods to evaluate the parameter importance: EWC directly estimates the parameter importance on the source dataset of G_s, while AdAM uses a modulation based method to estimate the G_s parameter importance on the target dataset. Therefore, in Table 1, we also show the results of applying our proposed knowledge truncation to EWC and AdAM. As our method can effectively remove the incompatible knowledge by pruning least important filters, we can achieve consistent improved performance on



Method	FID (↓)	Intra-LPIPS (↑)
TGAN	101.58	0.517
TGAN + ADA	97.91	0.511
FreezeD	96.25	0.518
EWC	79.93	0.521
CDC	69.13	0.578
DCL	56.48	0.580
AdAM	48.83	0.590
Ours	**39.39**	**0.608**
G_s (Pretrained)		

Method	FID (↓)	Intra-LPIPS (↑)
TGAN	64.68	0.490
TGAN + ADA	80.16	0.513
FreezeD	63.60	0.492
EWC	74.61	0.587
CDC	176.21	0.629
DCL	156.82	0.616
AdAM	58.07	0.557
Ours	**53.27**	0.569
G_s (Pretrained)		

Figure 4. Experiment results of FSIG on few-shot target samples. FFHQ [28] is the source domain. **Left**: 10-shot real target samples for adaptation. **Mid**: We visualize the generated images using the adapted generator G_t with different methods. Images in each column are from the same noise input. It is noticeable that our method while preserving the source knowledge useful for the target domain, reliably removes the incompatible source knowledge. For example, in SOTA methods, (**Top** orange frames) hat, doodle, sunglasses and beard are transferred and lead to generated babies with degraded realism; (**Bottom** blue frames) hat, glasses, human face texture and artifacts are transferred and lead to generated cats with low realism. In contrast, our method can address this issue in different setups. **Right**: Quantitatively, we measure the quality and diversity of generated images via FID (↓) [20] and intra-LPIPS (↑) [50]. See details in Sec. 6.

Table 1. We report FID (↓) as quantitative results for FSIG (10-shot), FFHQ is the source domain. We compare our proposed with other baseline and SOTA methods over six target datasets, including the challenging setups that target domains are distant to the source (*e.g.* AFHQ datasets). We emphasize that, for SOTA methods that focus only on knowledge preservation (*e.g.*, EWC [41], CDC [50], DCL [77], AdAM [75]), incompatible source knowledge is still transferred and therefore it curtails the quality of generated images. In contrast, our methods can remove the knowledge incompatible for the target and preserve the knowledge important for the target, therefore achieve improved quality of generated images.

Target Domain	**Babies** [50]	**Sunglasses** [50]	**MetFaces** [26]	**AFHQ-Cat** [8]	**AFHQ-Dog** [8]	**AFHQ-Wild** [8]
TGAN [65]	101.58	55.97	76.81	64.68	151.46	81.30
TGAN+ADA [26]	97.91	53.64	75.82	80.16	162.63	81.55
FreezeD [48]	96.25	46.95	73.33	63.60	157.98	77.18
CDC [50]	69.13	41.45	65.45	176.21	170.95	135.13
DCL [77]	56.48	37.66	62.35	156.82	171.42	115.93
EWC [41]	79.93	49.41	62.67	74.61	158.78	92.83
EWC + RICK (**Ours**)	68.22(**−11.71**)	39.53(**−9.88**)	54.7(**−7.97**)	64.35(**−10.26**)	124.50(**−34.28**)	56.83(**−36.00**)
AdAM [75]	48.83	28.03	51.34	58.07	100.91	36.87
AdAM + RICK (**Ours**)	43.12(**−5.71**)	26.25(**−1.78**)	49.47(**−1.87**)	53.94(**−4.13**)	100.35(**−0.56**)	35.54(**−1.33**)
Ours	**39.39**	**25.22**	**48.53**	**53.27**	**98.71**	**33.02**

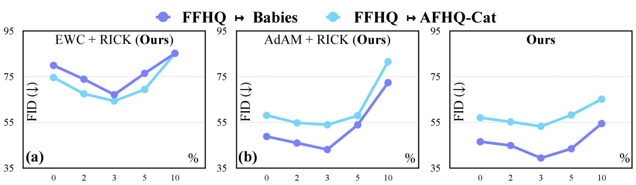


Figure 5. FID (↓) by pruning different percent (%) of filters, which importance are estimated in different ways. We note that prune 0% filters in (**a**-**b**) indicates the original EWC [41] and AdAM [75].

different datasets. More details are in Supplement.

Prune different percent of filters. We empirically study the impact of pruning different percent of filters. According to the results in Figure 5, we prune different amounts of filters on three different methods. Ideally, if we prune more filters, some important knowledge will be removed and the performance will be degraded accordingly. Therefore, we prune 3% (*i.e.*, p=3 in Sec. 5.1) filters in different setups, which can achieve considerable and stable improvements.

Can we train longer to remove the incompatible knowledge? Ideally, an intuitive and potentially useful way to remove the incompatible knowledge is to simply train longer iterations. In Supplement, however, we conduct a study and show that for existing FSIG methods, as the target set contains only 10-shot training images, training longer iterations will let the generator overfit and tend to replicate the few-shot target samples, such that it can fool the discriminator. The diversity of generated images is degraded significantly. Therefore, it is important to remove incompatible knowledge before the overfitting becomes severe.

7. Conclusion

We tackle few-shot image generation (FSIG) in this work. As the first contribution, we uncover the unnoticed issue of incompatible knowledge transfer of existing SOTA methods, which leads to significant degradation of generated image realism. Surprisingly, we discover that the root cause of such incompatible knowledge transfer is the filters that are deemed to have least importance for target adaptation, and fine-tuning based SOTA methods cannot properly address this issue. We therefore propose a novel concept, *knowledge truncation* for FSIG, which aims to eradicate the incompatible knowledge via pruning filters with least importance for adaptation. Our proposed filter importance estimation takes benefits of the gradient information from dynamic training process, and it is lightweight on computational cost. Through extensive experiments, we show that our proposed method can be applied to a wide range of adaptation setups with different GAN architectures. We achieve new state-of-the-art performance, including visually pleasant generated images without much incompatible knowledge transferred, and improved quantitative results.

Limitation and ethical concerns. The scale of our experiments is comparable to prior works. Nevertheless, extensions of our knowledge truncation method, additional datasets and generative models beyond GANs (*e.g.*, Variational Auto-Encoders [61] or Diffusion Models [53]) can be considered as future work. Our proposed FSIG methods may have negative societal impacts if it is used by malicious users. However, our work contributes to increased awareness about image generation with limited data.

Acknowledgment

This research is supported by the National Research Foundation, Singapore under its AI Singapore Programmes (AISG Award No.: AISG2-RP-2021-021; AISG Award No.: AISG-100E2018-005). This project is also supported by SUTD project PIE-SGP-AI-2018-01. We thank anonymous reviewers for their insightful comments.

References

[1] 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. 5, 6

[2] Arman Afrasiyabi, Hugo Larochelle, Jean-François Lalonde, and Christian Gagné. Matching feature sets for few-shot image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9014–9024, June 2022. 2

[3] 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. 3

[4] 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. 3, 4

[5] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *International Conference on Learning Representations*, 2018. 1

[6] Lucy Chai, Jun-Yan Zhu, Eli Shechtman, Phillip Isola, and Richard Zhang. Ensembling with deep generative views. In *CVPR*, 2021. 1

[7] Keshigeyan Chandrasegaran, Ngoc-Trung Tran, Yunqing Zhao, and Ngai-Man Cheung. Revisiting label smoothing and knowledge distillation compatibility: What was missing? In *International Conference on Machine Learning*, pages 2890–2916. PMLR, 2022. 2

[8] 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. 3, 6, 8

[9] 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

[10] 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. 2

[11] Utku Evci, Vincent Dumoulin, Hugo Larochelle, and Michael C. Mozer. Head2toe: Utilizing intermediate representations for better transfer learning. In *ICML*, 2022. 2

[12] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 321:321–331, 2018. 1

[13] Jia Gong, Zhipeng Fan, Qiuhong Ke, Hossein Rahmani, and Jun Liu. Meta agent teaming active learning for pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11079–11089, 2022. 1

[14] Jia Gong, Foo Lin Geng, Zhipeng Fan, and Jun Liu. Diffpose: Toward more reliable 3d pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 1

[15] 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. 1, 2

[16] Akshita Gupta, Sanath Narayan, K J Joseph, Salman Khan, Fahad Shahbaz Khan, and Mubarak Shah. Ow-detr: Open-world detection transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9235–9244, June 2022. 2

[17] Song Han, Jeff Pool, John Tran, and William Dally. Learning both weights and connections for efficient neural network. *Advances in neural information processing systems*, 28, 2015. 4

[18] Yang He, Yuhang Ding, Ping Liu, Linchao Zhu, Hanwang Zhang, and Yi Yang. Learning filter pruning criteria for deep convolutional neural networks acceleration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2009–2018, 2020. 4

[19] Yang He, Ping Liu, Ziwei Wang, Zhilan Hu, and Yi Yang. Filter pruning via geometric median for deep convolutional neural networks acceleration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4340–4349, 2019. 2, 4

[20] 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, 6, 7

[21] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 1

[22] Torsten Hoefler, Dan Alistarh, Tal Ben-Nun, Nikoli Dryden, and Alexandra Peste. Sparsity in deep learning: Pruning and growth for efficient inference and training in neural networks. *J. Mach. Learn. Res.*, 22(241):1–124, 2021. 4

[23] 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. 2

[24] Jiaxing Huang, Kaiwen Cui, Dayan Guan, Aoran Xiao, Fangneng Zhan, Shijian Lu, Shengcai Liao, and Eric Xing. Masked generative adversarial networks are data-efficient generation learners. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. 1

[25] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017. 3

[26] 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. 1, 2, 6, 8

[27] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. In *Proc. NeurIPS*, 2021. 1

[28] 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. 1, 3, 6, 7

[29] 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, 3, 6

[30] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. 1

[31] 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. 6

[32] Ananya Kumar, Aditi Raghunathan, Robbie Jones, Tengyu Ma, and Percy Liang. Fine-tuning can distort pretrained features and underperform out-of-distribution. In *ICLR*, 2022. 2

[33] Honglin Li, Chenglu Zhu, Yunlong Zhang, Yuxuan Sun, Zhongyi Shui, Wenwei Kuang, Sunyi Zheng, and Lin Yang. Task-specific fine-tuning via variational information bottleneck for weakly-supervised pathology whole slide image classification. *arXiv preprint arXiv:2303.08446*, 2023. 2

[34] Shuaifeng Li, Mao Ye, Xiatian Zhu, Lihua Zhou, and Lin Xiong. Source-free object detection by learning to overlook domain style. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8014–8023, 2022. 2

[35] Tianjiao Li, Lin Geng Foo, Qiuhong Ke, Hossein Rahmani, Anran Wang, Jinghua Wang, and Jun Liu. Dynamic spatio-temporal specialization learning for fine-grained action recognition. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part IV*, pages 386–403. Springer, 2022. 2

[36] Tianjiao Li, Qiuhong Ke, Hossein Rahmani, Rui En Ho, Henghui Ding, and Jun Liu. Else-net: Elastic semantic network for continual action recognition from skeleton data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13434–13443, 2021. 2

[37] Tianjiao Li, Jun Liu, Wei Zhang, and Lingyu Duan. Hardnet: Hardness-aware discrimination network for 3d early activity prediction. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 420–436. Springer, 2020. 2

[38] Tianjiao Li, Jun Liu, Wei Zhang, Yun Ni, Wenqian Wang, and Zhiheng Li. Uav-human: A large benchmark for human behavior understanding with unmanned aerial vehicles. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16266–16275, 2021. 2

[39] Tianjiao Li, Wei Zhang, Ran Song, Zhiheng Li, Jun Liu, Xiaolei Li, and Shijian Lu. Pot-gan: pose transform gan for person image synthesis. *IEEE Transactions on Image Processing*, 30:7677–7688, 2021. 2

[40] Xuhong LI, Yves GRANDVALET, and Franck DAVOINE. Explicit inductive bias for transfer learning with convolutional networks. In *ICML*, 2018. 2

[41] 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. 1, 2, 3, 4, 5, 6, 8

[42] Zhuohan Li, Eric Wallace, Sheng Shen, Kevin Lin, Kurt Keutzer, Dan Klein, and Joey Gonzalez. Train big, then compress: Rethinking model size for efficient training and inference of transformers. In *International Conference on Machine Learning*, pages 5958–5968. PMLR, 2020. 4

[43] Jianze Liang, Chengqi Zhao, Mingxuan Wang, Xipeng Qiu, and Lei Li. Finding sparse structures for domain specific neural machine translation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13333–13342, 2021. 5

[44] Ji Lin, Yongming Rao, Jiwen Lu, and Jie Zhou. Runtime neural pruning. *Advances in neural information processing systems*, 30, 2017. 4

[45] 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. 1

[46] Foo Lin Geng, Jia Gong, Zhipeng Fan, and Jun Liu. System-status-aware adaptive network for online streaming video understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 1

[47] Sebastian Mika, Gunnar Ratsch, Jason Weston, Bernhard Scholkopf, and Klaus-Robert Mullers. Fisher discriminant analysis with kernels. In *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468)*, pages 41–48. Ieee, 1999. 3, 5

[48] 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. 2, 3, 4, 8

[49] 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. 2

[50] 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, 3, 4, 5, 6, 7, 8

[51] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009. 2

[52] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015. 1

[53] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2021. 8

[54] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. In *ICLR*, 2014. 2

[55] Maying Shen, Pavlo Molchanov, Hongxu Yin, and Jose M Alvarez. When to prune? a policy towards early structural pruning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12247–12256, 2022. 4

[56] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013. 6

[57] Yang Sui, Miao Yin, Yi Xie, Huy Phan, Saman Aliari Zonouz, and Bo Yuan. Chip: Channel independence-based pruning for compact neural networks. *Advances in Neural Information Processing Systems*, 34:24604–24616, 2021. 2, 4

[58] Jiamei Sun, Sebastian Lapuschkin, Wojciech Samek, Yunqing Zhao, Ngai-Man Cheung, and Alexander Binder. Explanation-guided training for cross-domain few-shot classification. *arXiv preprint arXiv:2007.08790*, 2020. 2

[59] 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. 1

[60] Hung-Yu Tseng, Lu Jiang, Ce Liu, Ming-Hsuan Yang, and Weilong Yang. Regularizing generative adversarial networks under limited data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7921–7931, 2021. 1

[61] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017. 8

[62] Tianzhe Wang, Kuan Wang, Han Cai, Ji Lin, Zhijian Liu, Hanrui Wang, Yujun Lin, and Song Han. Apq: Joint search for network architecture, pruning and quantization policy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2078–2087, 2020. 4

[63] Tengfei Wang, Yong Zhang, Yanbo Fan, Jue Wang, and Qifeng Chen. High-fidelity gan inversion for image attribute editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11379–11388, 2022. 1

[64] Yaxing Wang, Abel Gonzalez-Garcia, David Berga, Luis Herranz, Fahad Shahbaz Khan, and Joost van de Weijer. Minegan: effective knowledge transfer from gans to target domains with few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9332–9341, 2020. 2

[65] 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. 1, 2, 3, 4, 8

[66] Zi Wang, Chengcheng Li, and Xiangyang Wang. Convolutional neural network pruning with structural redundancy reduction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14913–14922, June 2021. 2, 4

[67] 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. 6

[68] Yanbo Xu, Yueqin Yin, Liming Jiang, Qianyi Wu, Chengyao Zheng, Chen Change Loy, Bo Dai, and Wayne Wu. Transeditor: Transformer-based dual-space gan for highly controllable facial editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7683–7692, June 2022. 1

[69] Ceyuan Yang, Yujun Shen, Yinghao Xu, and Bolei Zhou. Data-efficient instance generation from instance discrimination. *Advances in Neural Information Processing Systems*, 34, 2021. 1

[70] Ceyuan Yang, Yujun Shen, Zhiyi Zhang, Yinghao Xu, Jiapeng Zhu, Zhirong Wu, and Bolei Zhou. One-shot generative domain adaptation. *arXiv preprint arXiv:2111.09876*, 2021. 3, 6

[71] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? In *NIPS*, 2014. 2

[72] Fisher Yu, Yinda Zhang, Shuran Song, Ari Seff, 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. 3

[73] Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. *CoRR*, abs/1311.2901, 2013. 2

[74] 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. 1, 2

[75] Yunqing Zhao, Keshigeyan Chandrasegaran, Milad Abdollahzadeh, and Ngai man Cheung. Few-shot image generation via adaptation-aware kernel modulation. In *Thirty-Sixth Conference on Neural Information Processing Systems*, 2022. 1, 2, 3, 4, 5, 6, 8

[76] Yunqing Zhao and Ngai-Man Cheung. Fs-ban: Born-again networks for domain generalization few-shot classification. *IEEE Transactions on Image Processing*, 2023. 2

[77] Yunqing Zhao, Henghui Ding, Houjing Huang, and Ngai-Man Cheung. A closer look at few-shot image generation. In *CVPR*, 2022. 2, 3, 4, 5, 6, 8

[78] Yunqing Zhao, Tianyu Pang, Chao Du, Xiao Yang, Ngai-Man Cheung, and Min Lin. A recipe for watermarking diffusion models. *arXiv preprint arXiv: 2303.10137*, 2023. 1

[79] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 40(6):1452–1464, 2017. 2

[80] Jiapeng Zhu, Yujun Shen, Deli Zhao, and Bolei Zhou. In-domain gan inversion for real image editing. In *European conference on computer vision*, pages 592–608. Springer, 2020. 1

[81] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2021. 2