Exploring Incompatible Knowledge Transfer in Few-shot Image Generation

Yunqing Zhao\textsuperscript{1}  Chao Du\textsuperscript{2}  Milad Abdollahzadeh\textsuperscript{1}  Tianyu Pang\textsuperscript{2}  
Min Lin\textsuperscript{2}  Shuicheng Yan\textsuperscript{2}  Ngai-Man Cheung\textsuperscript{1}  
\textsuperscript{1}Singapore University of Technology and Design (SUTD)  \textsuperscript{2}Sea AI Lab

yunqing.zhao@mymail.sutd.edu.sg, ngaiman_cheung@sutd.edu.sg

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 **Removing In-Compatible Knowledge** (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 $L_{adv}$ [15]:

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

(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 realism 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 $256 \times 256$ 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-
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.

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 removing In- Compatible Knowledge (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 Remove In- Compatible Knowledge 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.

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 \( W \in \mathbb{R}^{c'' \times k \times k} \), where \( k \) is the spatial size of the filter and \( c'' \) is the dimension (number) of the input feature maps. We use Fisher Information (FI) [47] as importance estimator for each filter \( \mathcal{F}(W) \) (to be further discussed in Sec. 5.2) that could tell quantitative information of compatibility between filter weights and the FSIG task [1, 75]:

\[
\mathcal{F}(W) = \mathbb{E} \left[ -\frac{\partial^2}{\partial W^2} \mathcal{L}_{G}(x|W) \right],
\]

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 filter importance in step 1), for the \( i \)-th filter \( W_i \) in the network, we apply a threshold \( (q\% \text{, i.e., the quantile of its importance compared to all filters}) \) to determine whether \( W_i \) should be pruned:

\[
W_i \leftarrow 0, \text{ if } \mathcal{F}(W_i) < q\%
\]

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 $M$: for each high dimension filter $W \in \mathbb{R}^{c \times k \times k}$, we only need a single character to record the corresponding operation in $M$. For example, for StyleGAN-V2 [29] used in main experiments whose generator contains $\sim 30$M parameters $^1$, $M$ is a one dimension array with size $|M| \sim 5,000$.

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 FL. 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, 75], 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 pre-trained 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.

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 ($\downarrow$) [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 ($\downarrow$). In Table 1, we show complete FID results over six benchmark datasets. In Figure 4, we also compute intra-LPIPS ($\downarrow$) 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

$^1$https://github.com/rosinality/stylegan2-pytorch
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 (\text{\textfraction}) 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.

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

Figure 5. FID (\text{\textfraction}) 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].

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., \text{\textfrac{p}\text{\textfrac{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


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


