Few-shot Image Generation via Adaptation-Aware Kernel Modulation

Yunqing Zhao* Keshigeyan Chandrasegaran* Milad Abdollahzadeh* Ngai-Man Cheung

Singapore University of Technology and Design (SUTD)

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

NEURAL INFORMATION PROCESSING SYSTEMS

Project Page

Few-shot Image Generation (FSIG)

Problem setup: Training GANs with only small amount of data (e.g., 10 - shot image)

Datasets: Babies, Sunglasses, Sketches, Art paintings, LSUN Church/Cars, Haunted House, etc.

Approach: Often tackled with Transfer learning, i.e., leveraging a GAN pretrained on a large dataset and adapt it to the small target domain.

Transfer learning Approach for FSIG

Knowledge Preservation of FSIG Methods:

Baselines:

1. **TGAN**: simple fine-tune methods; no knowledge preserve;
2. **FreezeD**: freeze low-level layers of the discriminator.

State-of-the-art (SOTA):

1. **EWC** (Li et al.): Penalize change of weights important on *source* domain;
2. **CDC** (Ojha et al.): Preserve distance between generated images *before and after* adaptation;
3. **DCL** (Zhao et al.): Maximize the mutual information of generated images before and after adaptation on multi-level features.

Observation:

1. Current SOTA methods consider only source domain in knowledge preservation for adaptation, i.e., **source-aware**;
2. None of them is "**target - aware**"

Source/Target domain Proximity Analysis

Main observation:

- SOTA methods often focus on source and target domains with close proximity, which share the similar semantic features (e.g., *real human face* to *baby face*).

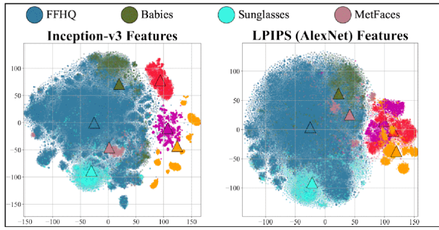
Analysis to confirm this observation:

- Qualitatively, we use Inception-V3 features and LPIPS features to project the entire dataset into feature space to visualize the source/target proximity;
- Quantitatively, we compute the FID and LPIPS as distance measurement between source/target datasets.

Datasets we used:

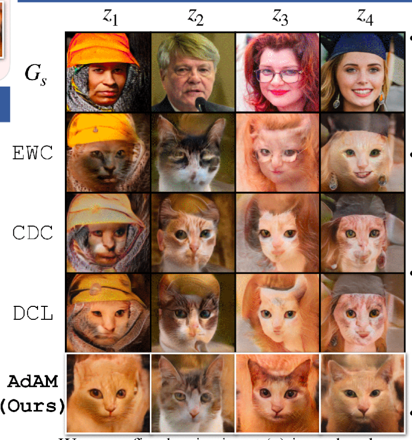
- **Source:** FFHQ - 70K,
- **Target - (A):** Babies, Sunglasses, MetFaces
- **Target - (B):** AFHQ-Cat, AFHQ-Dog, AFHQ-Wild

Results: Our introduced datasets (B) are more apart to the source.



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

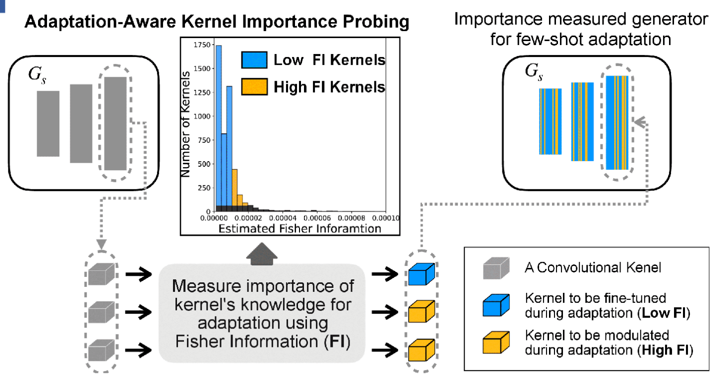
SOTA methods with relaxed source/target proximity



We use a fixed noise input (z) in each column

- In practice, we cannot always find the few-shot target domains with close proximity to the source.
- Therefore, we introduce new setup where the target and the source are more apart: e.g., FFHQ to AFHQ-Cat
- We show that, SOTA methods that aim to preserve source knowledge which is **non-target-aware,** will transfer unimportant features to the target.
- Our method **AdAM** (last row) can address this issue

Adaptation-Aware Kernel Modulation (AdAM)



Propose method: AdAM

- **Importance Probing (IP)**: We measure the importance of pretrained filters of the target domain as knowledge selection criterion. Specifically, we modulate the filter on the target domains for a few iterations. We then use Fisher Information (FI) to measure the importance of each filter. The IP step is applied to both generator and discriminator.
  - The computation cost of IP is lightweight. In practice, we only need ~8 mins for IP.
- **Main Adaptation**: If a filter is with *high FI*, we freeze this filter and only modulate it during adaptation. If a filter is with *low FI*, we instead fine-tune it on the target domain.
  - We use a threshold to determine whether a filter is important or unimportant.

Experiments: FSIG with close/distant target domains

	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 (Ours)	**48.83**	**0.590**
G_s (Pretrained)		

	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 (Ours)	**58.07**	0.557
G_s (Pretrained)		

We use a fixed noise input (z) to the generator in each column for a comprehensive comparison

Conclusion

- State-of-the-art FSIG methods fail when source/target domains are more apart because they are non-target-aware;
- We propose AdAM, a target-aware method to selectively preserve source knowledge useful for the target domains.
- Our proposed AdAM achieves SOTA on target domains with different proximity, both visually and quantitatively.

Reference

Li, Yijun, et al. "Few-shot Image Generation with Elastic Weight Consolidation." *NeurIPS. 2020.*

Ojha, Utkarsh, et al. "Few-shot image generation via cross-domain correspondence." *CVPR*. 2021.

Zhao, Yunqing, et al. "A Closer Look at Few-shot Image Generation." *CVPR*. 2022.