SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

NEURAL INFORMATION PROCESSING SYSTEMS

Project Page

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)

* Equal Contribution

Background

Training GANs requires large datasets

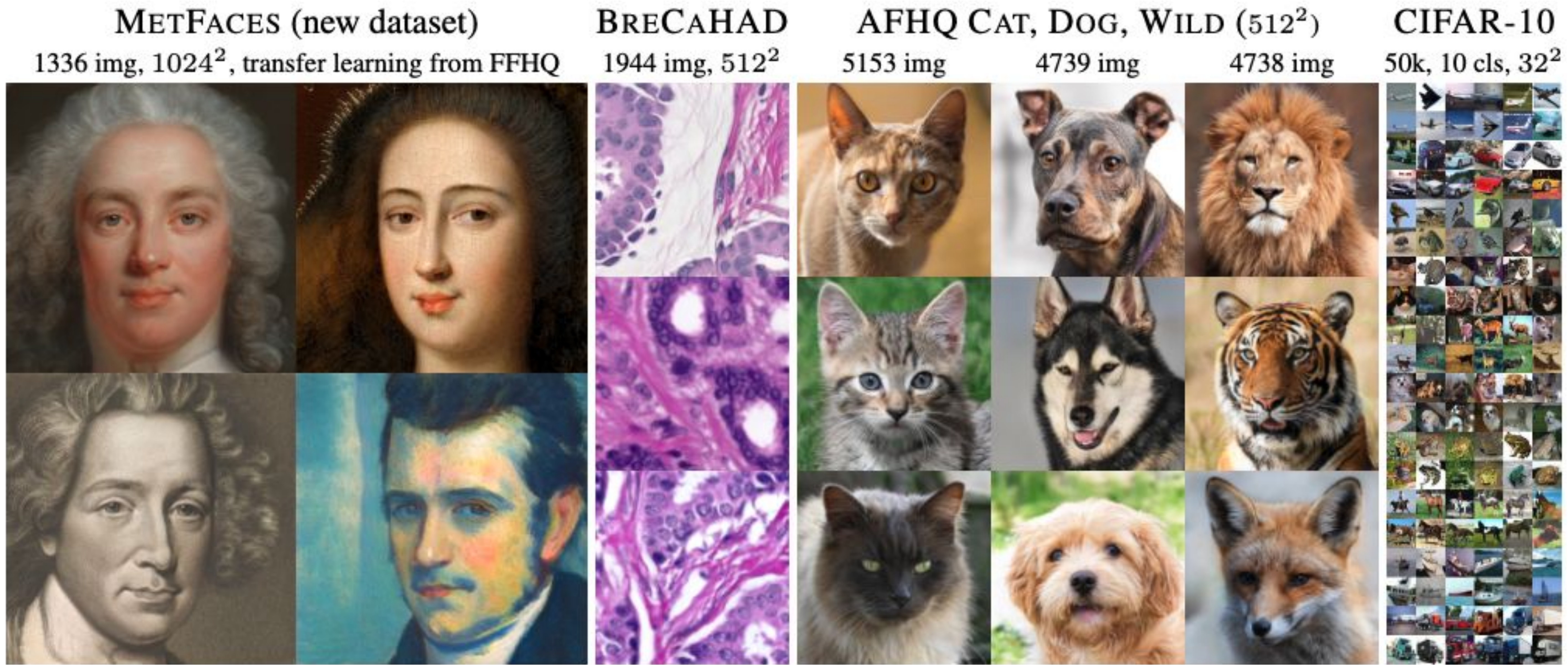
FFHQ dataset (credit)

ImageNet Dataset (credit)

1. Training GANs requires abundant training data (e.g., **FFHQ-70k** (left), **ImageNet** (right))
2. In real world, collecting images could be expensive and difficult (e.g., rare birds species)

Background

Training GANs with limited Data

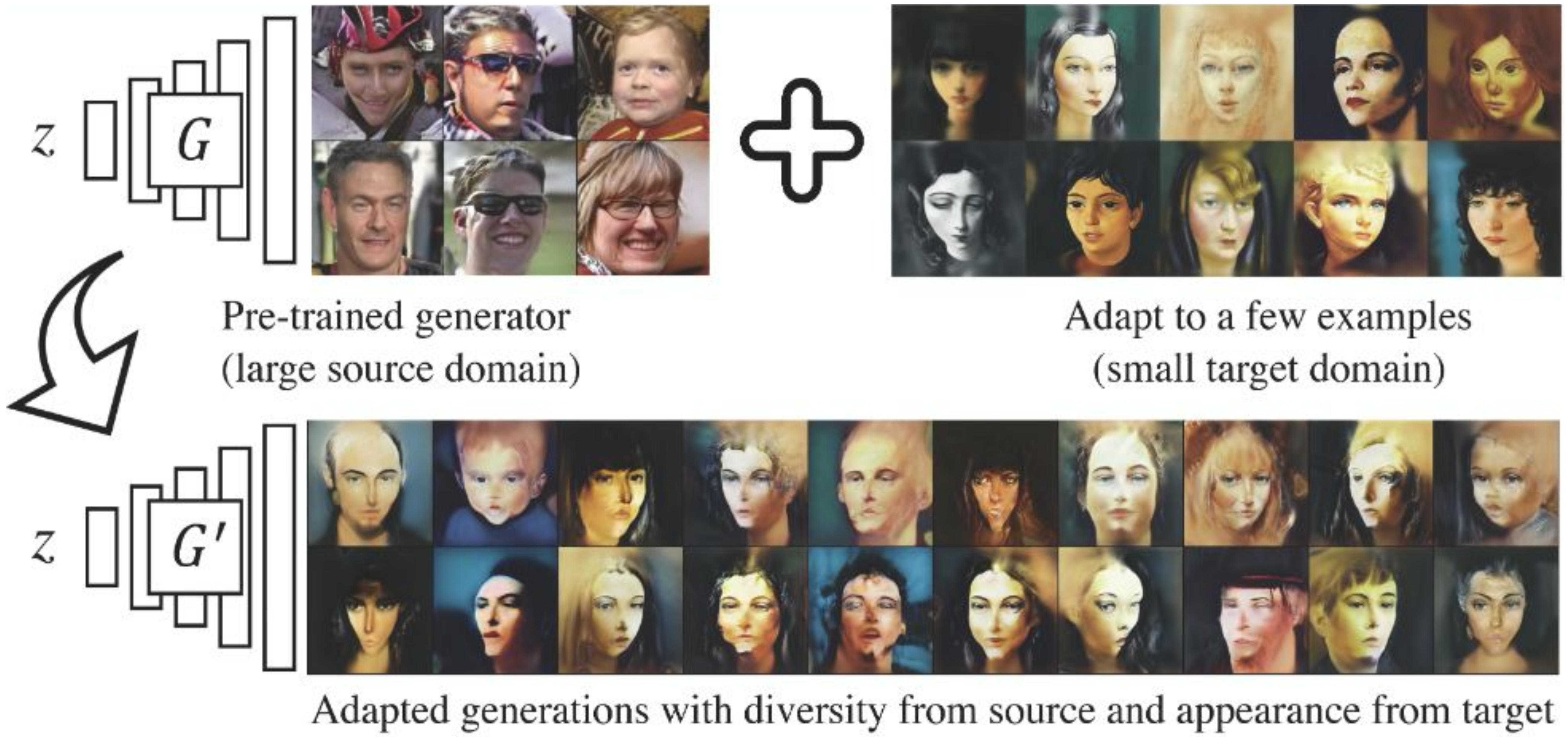


Small datasets used in ADA (Karras et al., credit)

1. Different works propose to training GAN with limited data (e.g., **1K ~ 5K images**)
2. **Approach:**
    a. Training from scratch (e.g., Karras et al., 2020)
    b. Transfer learning (e.g., Wang et al., 2018)

Wang, Yaxing, et al. "Transferring gans: generating images from limited data." *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
Karras, Tero, et al. "Training generative adversarial networks with limited data." Advances in Neural Information Processing Systems 33 (2020): 12104-12114.

Few-Shot Image Generation (FSIG)



FSIG setup: transfer learning (Li et al., 2020 credit)

Popular Approach for FSIG: Transfer Learning

1. Given a source GAN pretrained on a large dataset (e.g., **FFHQ**);
2. Adapt the source GAN to a small target domain (e.g., **10-shot**);
3. No access to source data during few-shot adaptation;

Li, Yijun, et al. "Few-shot Image Generation with Elastic Weight Consolidation." Advances in Neural Information Processing Systems 33 (2020): 15885-15896.

Related Works of FSIG

Ideas in state-of-the-art (SOTA) FSIG Methods:

- **EWC (Li et al., 2020):**
  - Penalize the changes of parameters important for source domain;
- **CDC (Ojha et al., 2021):**
  - Preserve the distance between generated images, before and after adaptation;
- **DCL (Zhao et al., 2022):**
  - Preserve multi-level source knowledge when adapting the source models to the target domain;

Observation:

- Central to SOTA methods is to preserve the source domain knowledge, or **"source-aware"**
- None of these work is **"Target/Adaptation - aware"**
  - i.e., their methods are target-domain agnostic

Li, Yijun, et al. "Few-shot Image Generation with Elastic Weight Consolidation." *Advances in Neural Information Processing Systems* 33 (2020): 15885-15896.
Ojha, Utkarsh, et al. "Few-shot image generation via cross-domain correspondence." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
Zhao, Yunqing, et al. "A Closer Look at Few-shot Image Generation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

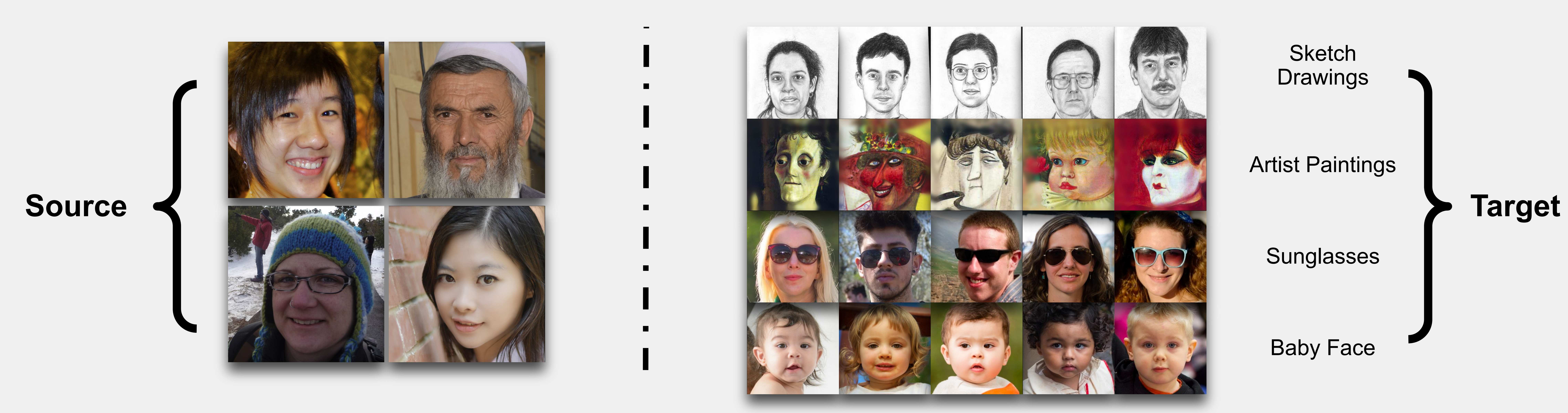
NEURAL INFORMATION PROCESSING SYSTEMS

Related Works of FSIG

Dataset / Domains in SOTA methods

Existing works focus on target domains with **close proximity** to the source:

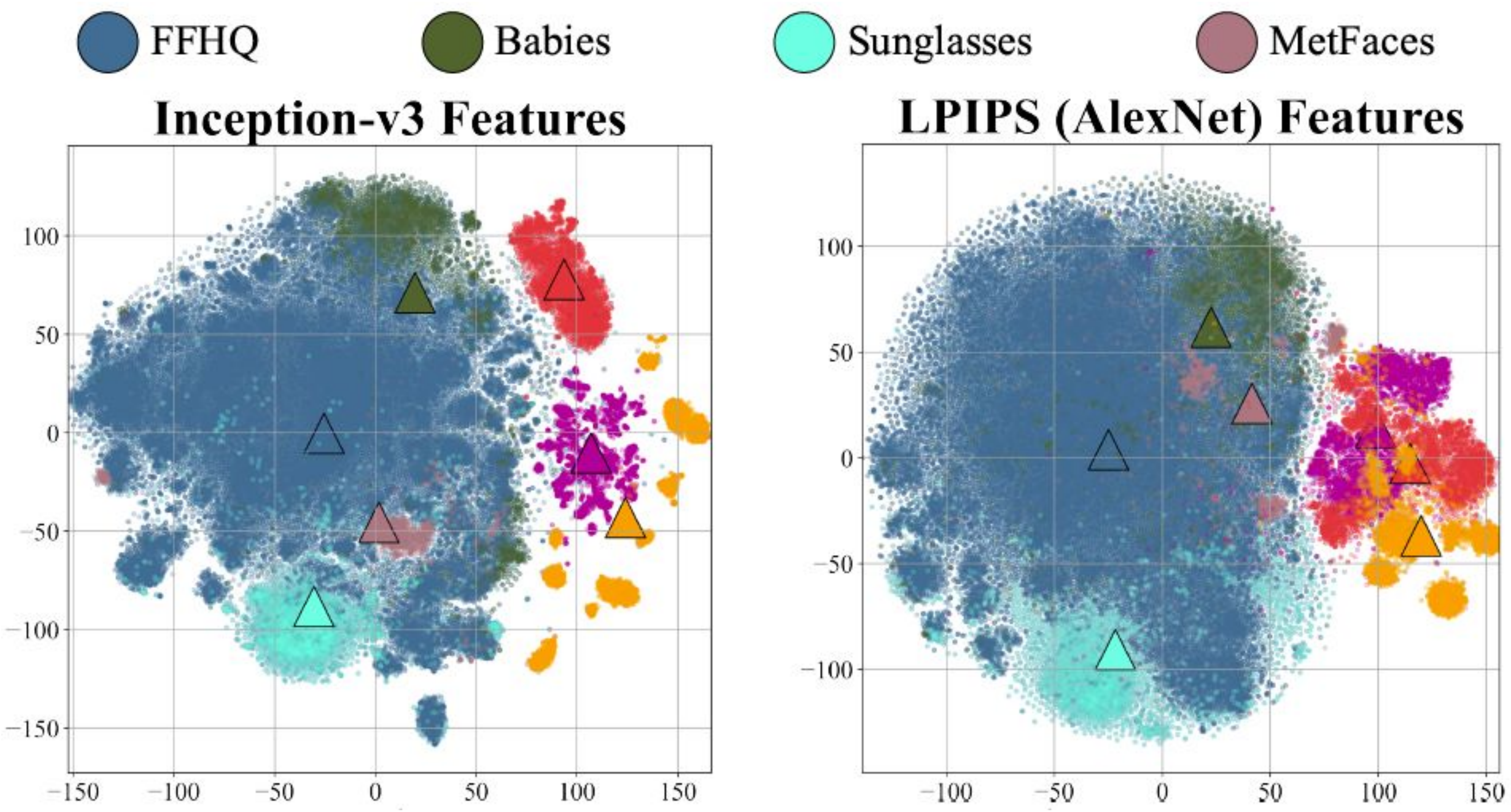
- E.g., **Source**: FFHQ (human face, **left**), **Target**: different Face paintings (**right**)



FFHQ (real Photos) **Different target domains**

Source/Target Domain Proximity Analysis

Dataset / Domains in SOTA methods



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

Close domains

Distant domains

(Left) Visualization of Inception-V3 and AlexNet Features of different domains. (Right) We also provide FID/LPIPS values for numerical analysis.

Analysis of Related Works

Problem?

1. **Observation: strong assumption of existing FSIG**
   - A. The source/target domains are of close proximity;
   - B. Therefore, SOTA methods aims to preserve **all information** of pretrained models.

2. **Issue:**
   - C. In real world, the target samples are **not necessarily** similar to the source;
   - D. If this close proximity assumption is relaxed, SOTA methods may fail.

Analysis of Related Works

Problem?

④

Each column is a fixed noise input (z) to the generator.

①

Target:
10-shot real cat

②

These SOTA methods are not target-aware, introducing source domain information that is **improper to target dataset.**

E.g., hair, specs from human face domains, these features are no-good for cat domain.

→ ③

A glimpse of ours compared to SOTA on <FFHQ → AFHQ-Cat>

Our Ideas

1. In CNN, each filter is responsible for a specific part of knowledge (texture/pattern);
2. For different target domains, not all knowledge should be preserved;
3. We selectively preserve part of the source knowledge
   A. If the knowledge is important for adaptation, preserve it;
   B. If the knowledge is unimportant for adaptation, update it (via GAN loss).

In short: Casting the **knowledge preservation proble**m to a **decision problem** of whether a kernel is important when adapting from source to target

Our Ideas

1. In CNN, each filter is responsible for a specific part of knowledge (texture/pattern);
2. For different target domains, not all knowledge should be preserved;
3. We selectively preserve part of the source knowledge
   A. If the knowledge is important for adaptation, preserve it;
   B. If the knowledge is unimportant for adaptation, update it (via GAN loss).

Question (We answer them in the next):

1. **How to determine the important filters?**
2. **How to preserve the knowledge?**

Proposed Method (Overview)

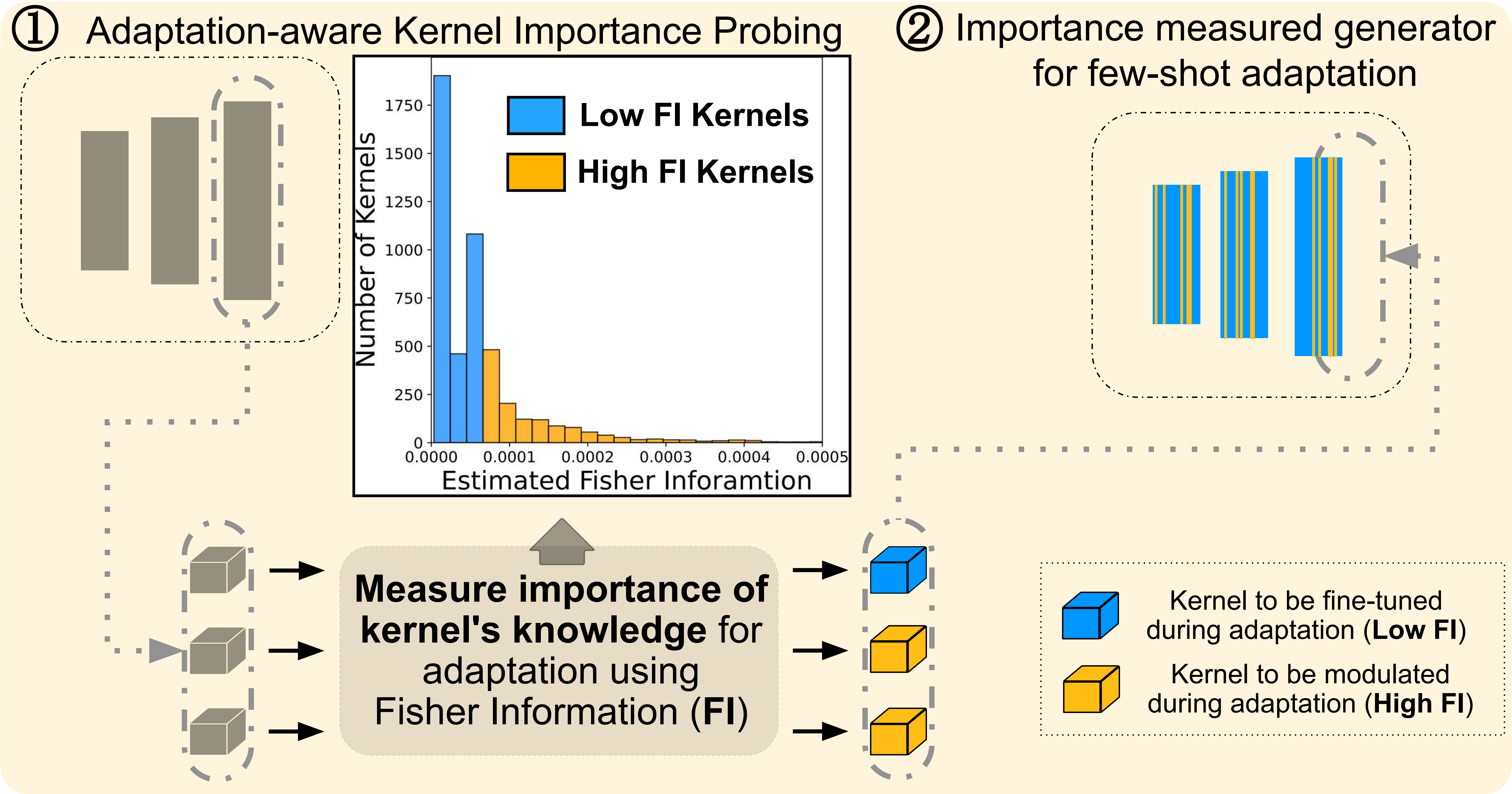
Adaptation-Aware Kernel Modulation (AdAM) for FSIG

Stage-① : Importance Probing:

- **Goal:** identify filters important for few-shot adaptation process.

Stage-②: Hybrid Training

- **Important filters:** Preserve knowledge;
- **Unimportant filters:** Fine-tuning



Proposed Method

Adaptation-Aware Kernel Modulation (AdAM) for FSIG

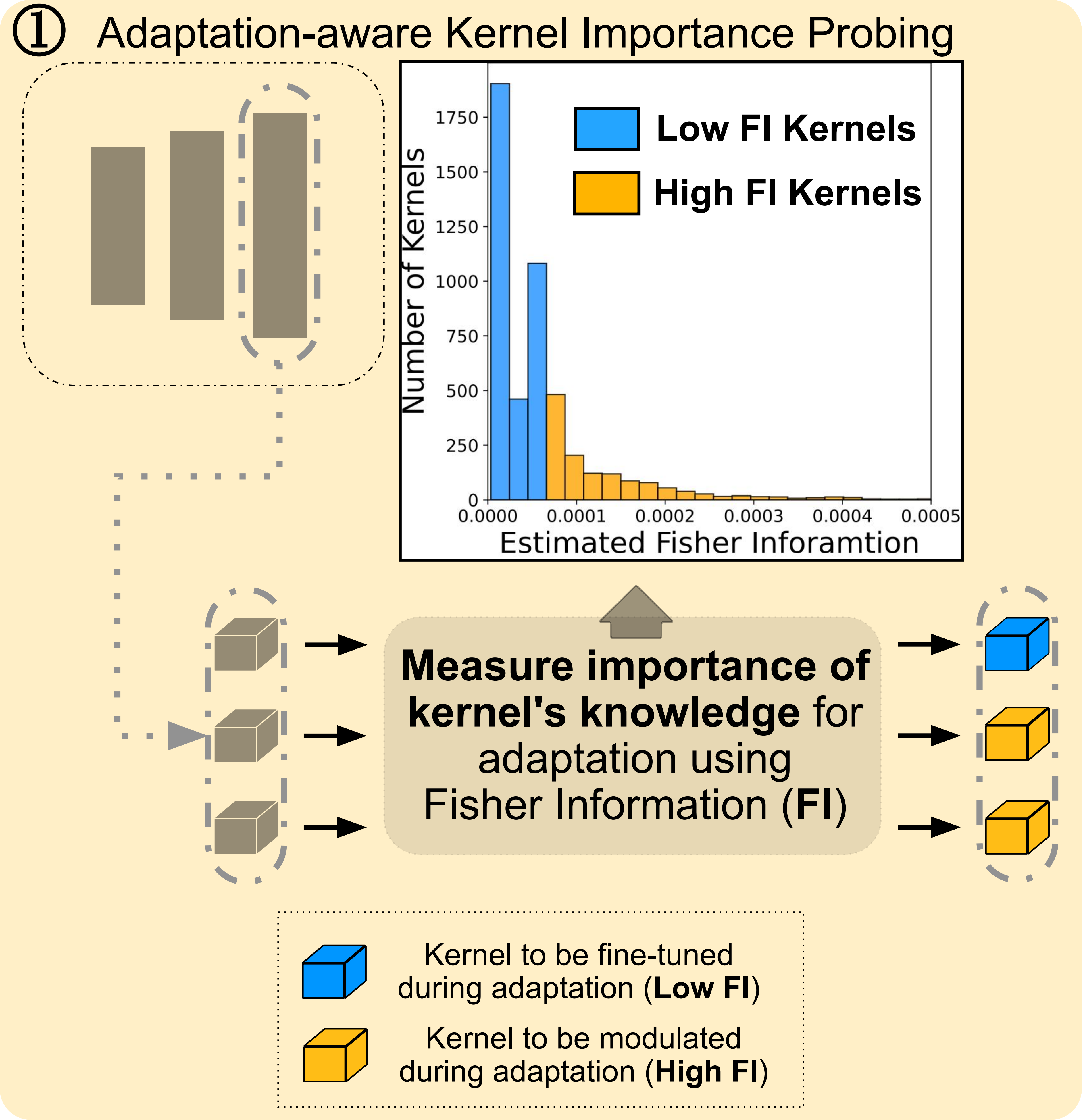
Stage-① : Importance Probing:

Steps:

1. A parameter-efficient adaptation on target domain for a few iterations;
2. Measure the importance for all filters for the target.

Output:

1. Decisions of Importance/unimportance to each individual filters.



Proposed Method

Adaptation-Aware Kernel Modulation (AdAM) for FSIG

Stage-① : Importance Probing:

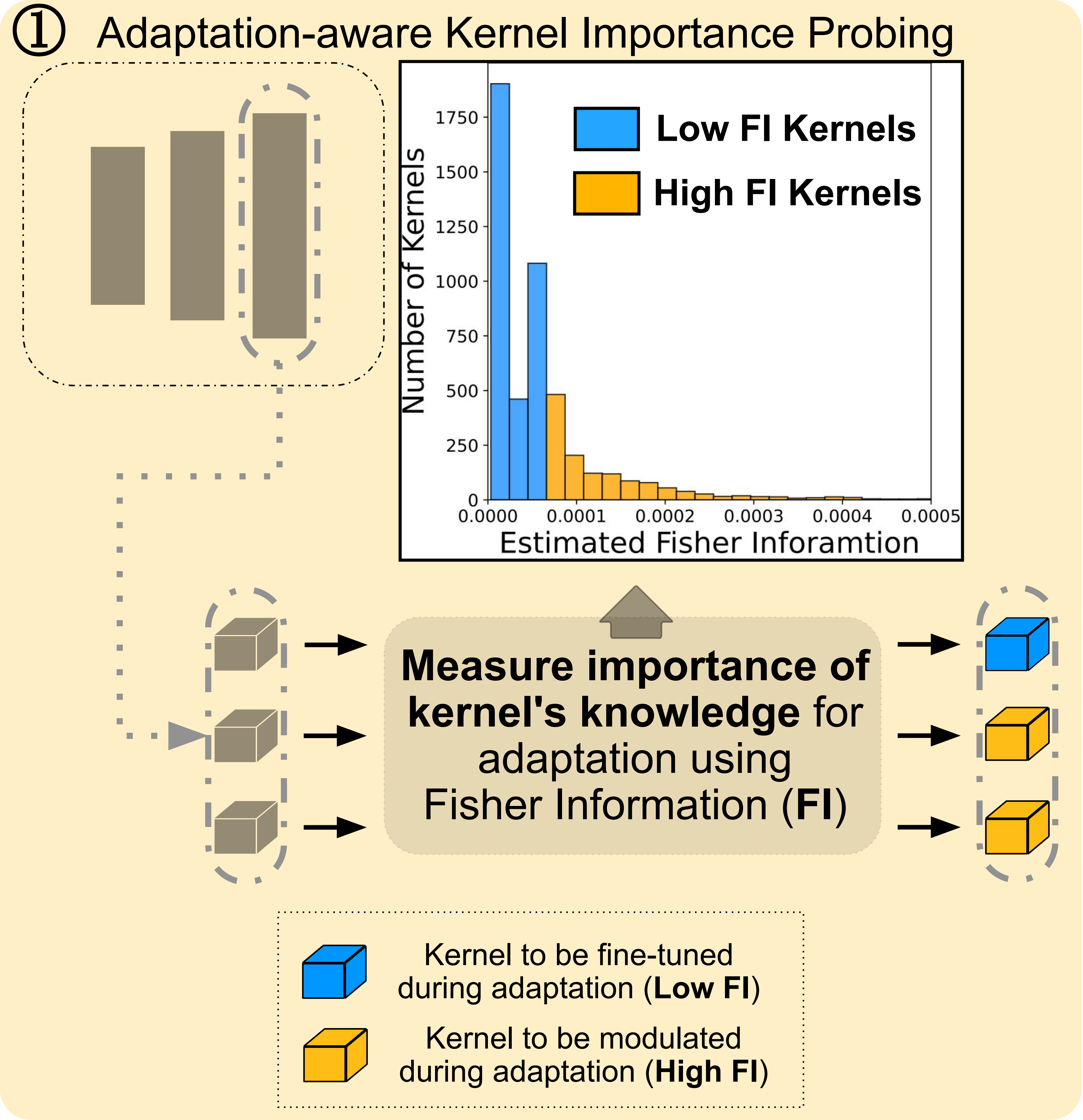
Importance measurement:

To measure the importance of the modulated kernels, we apply **Fisher Information (FI)** to modulation parameters.

$$\mathrm{FI}(\Theta) = \mathbb{E}\left[-\frac{\partial^2}{\partial \Theta^2} \mathcal{L}(x \mid \Theta) \right]$$

$$\begin{cases} \Theta: \text{parameters of G/D} \\ \mathcal{L}: \text{loss of G/D} \end{cases}$$

In practice, use the first-order approximation to estimate FI.



SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

NEURAL INFORMATION PROCESSING SYSTEMS

Proposed Method

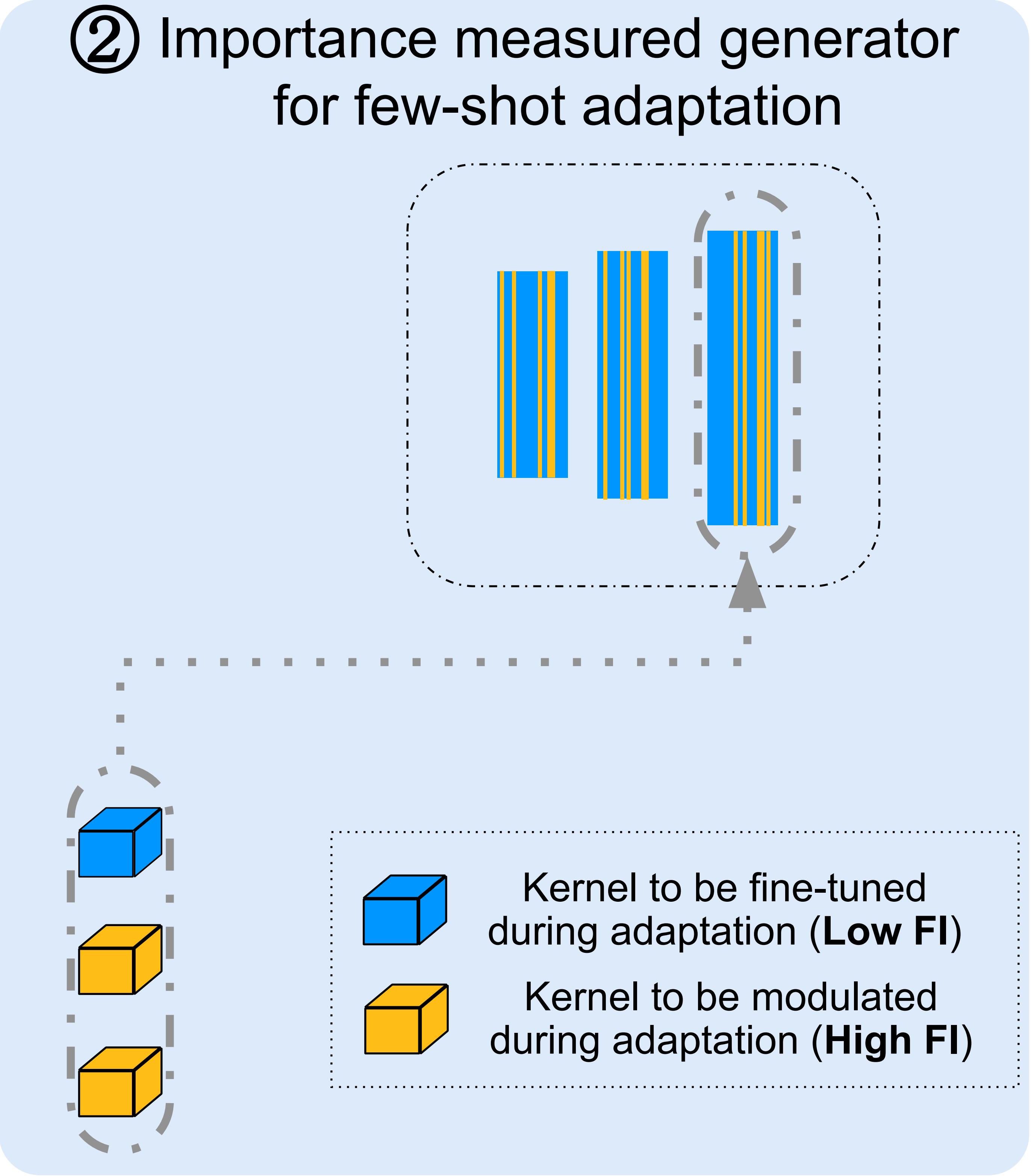
Adaptation-Aware Kernel Modulation (AdAM) for FSIG

Stage-②: Hybrid Training

- **For Important filters:**

We use KML to protect the pretrained knowledge from distorted [1].
E.g., if $i-th$ kernel is identified as important for adaptation:

$$\hat{W}_i = \underbrace{W_i}_{\text{Fixed}} \odot (\underbrace{J}_{\text{All-one Matrix}} + \underbrace{M_i}_{\text{Learnable}})$$



- **For Unimportant filters:** Simply Fine-tuning

[1] Kumar, Ananya, et al. "Fine-Tuning can Distort Pretrained Features and Underperform Out-of-Distribution." *ICLR*. 2022.

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

NEURAL INFORMATION PROCESSING SYSTEMS

Experiment Results

1: Source/target domains are closely related

10-Shot Real Babies

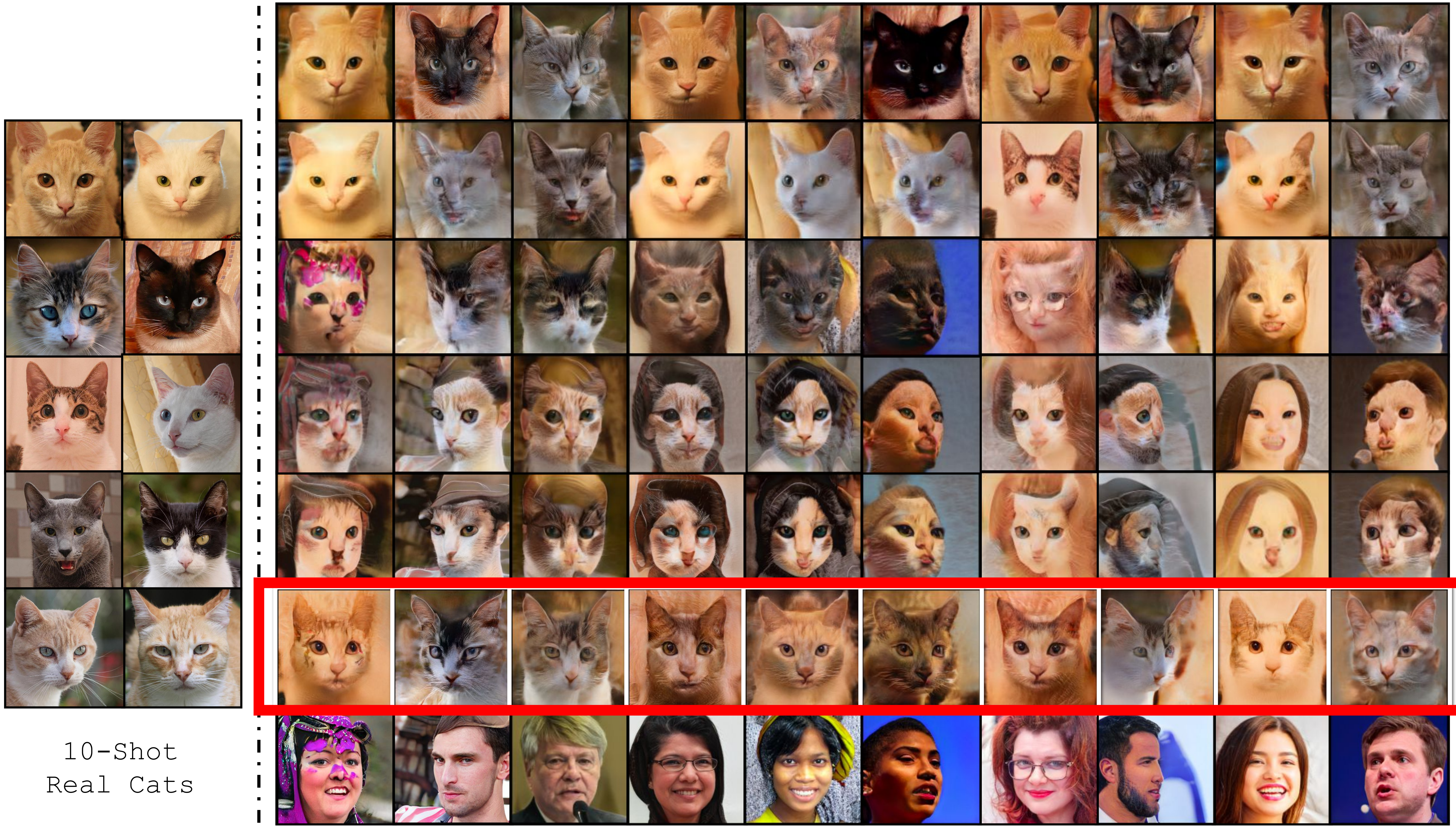
	FID (↓)	Intra-LPIPS (↑)
TGAN	101.58	0.517
FreezeD	96.25	0.518
EWC	79.93	0.523
CDC	69.13	0.573
DCL	56.48	0.582
AdAM (Ours)	**49.65**	**0.591**
(Pretrained)		

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

NEURAL INFORMATION PROCESSING SYSTEMS

Experiment Results

2: Source/target domains are distant



10-Shot Real Cats

	FID (↓)	Intra-LPIPS (↑)
TGAN	64.68	0.490
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
(Pretrained)		

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

NEURAL INFORMATION PROCESSING SYSTEMS

Application to another GAN architecture (ProGAN)

10-Shot Palace

Progressive-GAN: Church → Palace

	FID (↓)	Intra-LPIPS (↑)
TGAN	Insufficient real samples in the target domain for reliable FID computation.	0.593
TGAN + ADA		0.579
EWC		0.645
AdAM (Ours)		**0.679**
(Pretrained)		

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

NEURAL INFORMATION PROCESSING SYSTEMS

Application to another GAN architecture (ProGAN)

10-Shot Real Babies

Progressive-GAN: FFHQ → Babies

	FID (↓)	Intra-LPIPS (↑)
TGAN	86.91	0.507
TGAN + ADA	83.09	0.555
EWC	80.77	0.559
AdAM (Ours)	**78.33**	**0.575**
(Pretrained)		

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

NEURAL INFORMATION PROCESSING SYSTEMS

Conclusion

Our analysis & main findings

- Current SOTA FSIG methods focus on similar source/target domains;
- Therefore, they aim to preserve all source knowledge
  1. If source/target are **related**, the performance is good
  2. If source/target are **distant**, SOTA is no better than the baseline

Our contributions

- We propose to identify the filters important for target adaptation;

  Stage-1: Adaptation-Aware **importance probing;**

  Stage-2: Hybrid training (**Low-rank Kernel Modulation** and Fine-tuning)
- We achieve new SOTA when source/target **are either related or distant**

Few-shot Image Generation via Adaptation-Aware Kernel Modulation

Code, dataset and checkpoints:

Project Page

Thank you for watching

[We use materials compiled from various sources such as textbooks, lecture materials, conference talks, web resources and are shared for research purposes only. In the interest of brevity, every source is not cited. The compiler of these materials gratefully acknowledges all such sources. Please contact yunqing_zhao@mymail.sutd.edu.sg for any queries.]