

# FEW-SHOT ADAPTATION OF GENERATIVE ADVERSARIAL NETWORKS

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

Generative Adversarial Networks (GANs) have shown remarkable performance in image synthesis tasks, but typically require a large number of training samples to achieve high-quality synthesis. This paper proposes a simple and effective method, Few-Shot GAN (FSGAN), for adapting GANs in few-shot settings (less than 100 images). FSGAN repurposes component analysis techniques and learns to adapt the singular values of the pre-trained weights while freezing the corresponding singular vectors. This provides a highly expressive parameter space for adaptation while constraining changes to the pretrained weights. We validate our method in a challenging few-shot setting of 5-100 images in the target domain. We show that our method has significant visual quality gains compared with existing GAN adaptation methods. We report extensive qualitative and quantitative results showing the effectiveness of our method. We additionally highlight a problem for few-shot synthesis in the standard quantitative metric used by data-efficient image synthesis works.

## 1 INTRODUCTION

Recent years have witnessed rapid progress in Generative Adversarial Networks (GAN) (Goodfellow et al., 2014) with improvements in architecture designs (Radford et al., 2015; Karras et al., 2018; Zhang et al., 2019; Karras et al., 2019a), training techniques (Salimans et al., 2016; Karras et al., 2018; Miyato et al., 2018), and loss functions (Arjovsky et al., 2017; Gulrajani et al., 2017). Training these models, however, typically requires large, diverse datasets in a target visual domain. While there have been significant advancements in improving training stability (Karras et al., 2018; Miyato et al., 2018), adversarial optimization remains challenging because the optimal solutions lie at saddle points rather than a minimum of a loss function (Yadav et al., 2017). Additionally, GAN-based models may suffer from the inadequate generation of rare modes in the training data because they optimize a mode-seeking loss rather than the mode-covering loss of standard likelihood maximization (Poole et al., 2016). These difficulties of training GANs become even more severe when the number of training examples is scarce. In the low-data regime (e.g., less than 1,000 samples), GANs frequently suffer from memorization or instability, leading to a lack of diversity or poor visual quality.

Several recent efforts have been devoted to improving the sample efficiency of GANs through transfer learning. The most straightforward approaches are finetuning a pre-trained generator and discriminator on the samples in the target domain (Wang et al., 2018; Mo et al., 2020). When the number of training examples is severely limited, however, finetuning the network weights often leads to poor results, particularly when the source and target domains are distant. Instead of finetuning the entire network weights, the method in (Noguchi & Harada, 2019) focuses on adapting batch norm statistics, constraining the optimization problem to a smaller set of parameters. The authors report that this method achieves better results using MLE-based optimization but fails for GAN-based optimization. Although their quality outperforms GAN-based methods in the low-shot setting, the images are blurry and lack details due to maximum likelihood optimization. Invertible flow-based models have shown promising results in data-efficient adaptation (Gambardella et al., 2019), but require compute- and memory-intensive architectures with high-dimensional latent spaces.

In this paper, we propose a method for adapting a pre-trained GAN to generate novel, high-quality sample images with a small number of training images from a new target domain (Figure 1). To accomplish this, we restrict the space of trainable parameters to a small number of highly-expressive parameters that modulate orthogonal features of the pre-trained weight space. Our method first applies singular value decomposition (SVD) to the network weights of a pretrained GAN (generator + discriminator). We then adapts the singular values using GAN optimization on the target few-shot



Figure 1: **Few-shot image generation.** Our method generates novel and high-quality samples in a new domain with a small amount of training data. (*Top*) Diverse random samples from adapting a FFHQ-pretrained StyleGAN2 to toddler images from the CelebA dataset (with **only 30 images**) using our method. (*Bottom*) Smooth latent space interpolation between two random seeds shows that our method produces novel samples instead of simply memorizing the 30 images. Please see the supplementary video for more results.

domain, with *fixed* left/right singular vectors. We show that varying singular values in the weight space corresponds to semantically meaningful changes of the synthesized image while preserving natural structure. Compared with methods that finetune all weights of the GAN (Wang et al., 2018), individual layers (Mo et al., 2020), or only adapt batch norm statistics (Noguchi & Harada, 2019), our method demonstrates higher image quality after adaptation. We additionally highlight problems with the standard evaluation practice in the low-shot GAN setting.

## 2 BACKGROUND

**Generative Adversarial Networks (GANs)** GANs (Goodfellow et al., 2014) use adversarial training to learn a mapping of random noise to the distribution of an image dataset, allowing for sampling of novel images. GANs optimize a competitive objective where a generator  $G(Z)$  maximizes the classification error of a discriminator  $D(X)$  trained to distinguish real data  $p(X)$  from fake data  $G(Z)$ . The GAN (Goodfellow et al., 2014) objective is expressed formally as:

$$\max_G \min_D \mathbb{E}_{x \sim p(X)} [\log D(x)] - \mathbb{E}_{x \sim G(Z)} [1 - \log D(x)] \quad (1)$$

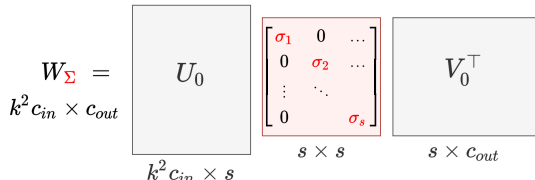
Recent research reformulated this objective to address instability problems (Arjovsky et al., 2017; Heusel et al., 2017a; Gulrajani et al., 2017). Improved architecture and training has led to remarkable performance in synthesis (Karras et al., 2020; Brock et al., 2019). Compared to pixel-reconstruction losses (Kingma & Welling, 2014; Higgins et al., 2017; Bojanowski et al., 2018) GANs typically produce sharper images, although strong priors over the latent space can offer competitive quality (Razavi et al., 2019). A high-quality generation has relied on large datasets of high-quality images (10K+) that may be expensive or infeasible to collect in many scenarios. Additionally, GANs can suffer from a lack of diversity, even when large training sets are used because the objective does not penalize the absence of outlier modes (Poole et al., 2016). Data-efficient GAN methods are, therefore, of great utility.

**Sample-efficient Image Synthesis** Sample-efficient image synthesis methods encourage diverse and high-quality generation in the low-data regime, most commonly through pretraining (Wang et al., 2018; Noguchi & Harada, 2019) or simultaneous training (Yamaguchi et al., 2019) on large image datasets. The main differences among these methods lie in the choice of learnable parameters used for adaptation. Examples include adapting all weights of the generator and discriminator Wang et al. (2018), freezing only lower layers of the discriminator Mo et al. (2020), or changing only channel-wise batch statistics Noguchi & Harada (2019). Flow-based methods (Gambardella et al., 2019) show promising results in few-shot adaptation, but their architecture is compute- and memory-intensive and requires latent space of the same dimensionality as the data. Our method uses a *smaller but more expressive set of parameters* (Figure 2), resulting in more natural adapted samples.

**One-shot Image Re-synthesis.** Recent work in one-shot image synthesis has demonstrated high-quality and diverse results by modeling the *internal* distribution of features from a single image

Method	Conv layer	#params	Count
Pretrain	$conv(x, W_0)$	–	–
TGAN	$conv(x, W)$	$k^2 c_{in} c_{out}$	59M
FreezeD	$conv(x, W)$	$k^2 c_{in} c_{out}$	47M
SSGAN	$conv(x, W_0) \cdot \gamma + \beta$	$2c_{out}$	23K
FSGAN (Ours)	$conv(x, W_\Sigma)$	$c_{out}$	16K

(a) Adaptation method formulations.



(b) FSGAN singular value adaptation.

Figure 2: **Comparing methods for GAN adaption.** Learnable parameters are denoted in red. (a) TransferGAN (TGAN for simplicity) (Wang et al., 2018) and FreezeD (Mo et al., 2020) retrain all weights  $W$  in a layer. SSGAN (Noguchi & Harada, 2019) and FSGAN train significantly fewer parameters per layer. Note FSGAN adapts both conv and FC layers, while SSGAN adapts only conv layers. *#params* is the number of learnable parameters per conv layer; *Count* gives parameter counts over the full StyleGAN2 generator and discriminator. (b) FSGAN (ours) adapts singular values  $\Sigma = \{\sigma_1, \dots, \sigma_s\}$  of pretrained weights  $W_0$  to obtain adapted weights  $W_\Sigma$ .

without pretraining (Shaham et al., 2019; Shocher et al., 2019). Our work differs as we transfer *external* knowledge from a pretrained GAN to a new domain and, therefore, can generate drastically more diverse samples.

**Singular Value Decomposition (SVD).** SVD factorizes any matrix  $M \in \mathbb{R}^{m \times n}$  into unitary matrices  $U \in \mathbb{R}^{m \times m}$ ,  $V \in \mathbb{R}^{n \times n}$  and diagonal matrix  $\Sigma$  such that  $M = U\Sigma V^\top$ , where  $U, V$  contain the left and right singular vectors respectively and  $\Sigma$  contains the singular values along the diagonal entries. SVD can be interpreted as a decomposition of a linear transformation  $x \rightarrow Mx$  into three separate transformations: a rotation/reflection  $U$ , followed by rescaling  $\Sigma$ , followed by another rotation/reflection  $V^\top$ . The transformation defined by the maximum singular value  $\sigma_0 = \Sigma^{(1,1)}$  and its corresponding normalized singular vectors represent the maximal axis of variation in the matrix  $M$ . This interpretation is commonly used in data science for dimensionality reduction via PCA (Kwak, 2008). PCA can be obtained via SVD on a column-normalized matrix (Golub & Reinsch, 1971). SVD is also used for a wide number of other applications, including regularization (Sedghi et al., 2018), and quantification of entanglement (Martyn et al., 2020), and has also been used to build theoretical background for semantic development in neural networks (Saxe et al., 2019). The work most closely related to ours is GANSpace (Härkönen et al., 2020) for image synthesis editing. GANSpace applies PCA within the *latent feature* space of a pretrained GAN to discover semantically-interpretable directions for image editing in the latent space. In contrast, our work performs SVD on the *weight* space of a GAN to discover meaningful directions for domain adaptation. Performing SVD on the weight space enables two critical differences between our work and Härkönen et al. (2020): (i) we edit the entire output *distribution* rather than one image, and (ii) rather than manual editing, we adapt the GAN to a new domain.

### 3 FEW-SHOT GAN

#### 3.1 OVERVIEW

Our goal is to improve GAN finetuning on small image domains by discovering a more effective and constrained parameter space for adapting the pretrained weights. We are inspired by prior work in GAN adaptation showing that constraining the space of trainable parameters can lead to improved performance on target domain (Rebuffi et al., 2017; Mo et al., 2020; Noguchi & Harada, 2019). In contrast to identifying the parameter space within the model architecture, we propose to discover a parameter space based on the pretrained weights. Specifically, we apply singular value decomposition to the pretrained weights and uncover a basis representing orthogonal directions of maximum variance in the weight space. To explore the interpretation of the SVD representation, we visualize the top three singular values of synthesis and style layers of StyleGAN2 (Karras et al., 2020). We observe that varying the singular values corresponds to natural and semantically-meaningful changes in the output image as shown in Figure 3. Changing the singular values can be interpreted as changing the entanglement between orthogonal factors of variation in the data (singular vectors), providing an

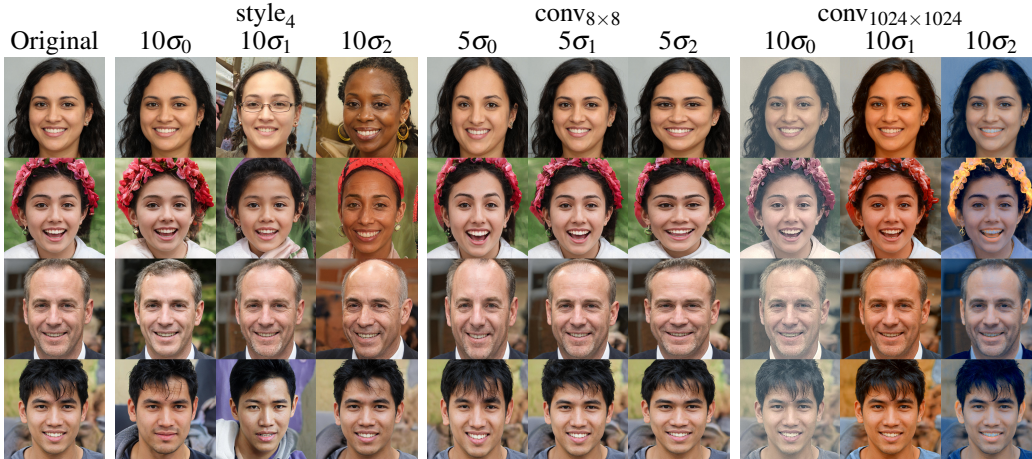


Figure 3: **Effects of singular values.** We visualize FSGAN’s adaptation space by magnifying the top 3 singular values  $\sigma_0, \sigma_1, \sigma_2$  from SVD performed on style and conv layers of a StyleGAN2 (Karras et al., 2019a; 2020) pretrained on FFHQ. In mapping layer 4 (style<sub>4</sub>), the leading  $\sigma$ s change the age, skin tone, and head pose. In synthesis layer 2 (conv<sub>8×8</sub>), face dimensions are modified in term of face height/size/width. In synthesis layer 9 (conv<sub>1024×1024</sub>), the face appearance changes in finer pixel stats such as saturation, contrast, and color balance.

expressive parameterization of the pretrained weights, which we leverage for adaptation as described in the following section.

### 3.2 ADAPTATION PROCEDURE

Our method first performs SVD on both the generator and discriminator of a pretrained GAN and adapts the singular values to a new domain using standard GAN training objectives. A generator layer  $G^{(\ell)}$  or a discriminator layer  $D^{(\ell)}$  may consist of either 2D ( $c_{in} \times c_{out}$ ) fully-connected weights or 4D ( $k \times k \times c_{in} \times c_{out}$ ) convolutional filter weights. We apply SVD separately at every layer of the generator  $G^{(\ell)}$  and discriminator  $D^{(\ell)}$ . Next, we describe the decomposition process for a single layer of pretrained weights  $W_0^{(\ell)}$ . For fully-connected layer  $W_0^{(\ell)}$ , we can apply SVD directly on the weight matrix. For 4D convolution weights  $W_0^{(\ell)} \in \mathbb{R}^{k \times k \times c_{in} \times c_{out}}$  this is not feasible because SVD operates only on a 2D matrix. We therefore reshape the 4D tensor by flattening across the spatial and input feature channels before performing SVD to obtain a 2D matrix  $W_0^{(\ell)} \in \mathbb{R}^{k^2 c_{in} \times c_{out}}$ . Our intuition is that the spatial-feature relationship in the pretrained model should be preserved during the adaptation. We apply SVD over each set of flattened convolutional weights or fully convolution weights to obtain the decomposition:

$$W_0^{(\ell)} = (U_0 \Sigma_0 V_0^\top)^{(\ell)}. \quad (2)$$

After decomposing the pretrained weights, we perform domain adaptation by freezing pretrained left/right singular vectors in  $(U_0, V_0)^{(\ell)}$  and optimizing the singular values  $\Sigma = \lambda \Sigma_0$  using a standard GAN objective to obtain transferred weights (Figure 2):

$$W_\Sigma^{(\ell)} = (U_0 \Sigma V_0^\top)^{(\ell)} \quad (3)$$

Effectively, our GAN domain adaptation aims to find a new set of singular values in each layer of a pretrained model so that the generated outputs match the distribution of the target domain.

During forward propagation, we reconstruct weights  $W_\Sigma$  using the finetuned singular values at each convolution or fully-connected layer of the generator and discriminator before applying the operation.

### 3.3 TRAINING & INFERENCE

Our experiments use the StyleGAN2 (Karras et al., 2020) training framework, which optimizes a logistic GAN loss (Equation 1) with latent space gradient regularization and a discriminator gradient








t	FID	Interpolation
0	121.21	
20	154.25	
40	134.22	
80	102.87	
120	93.65	
180	<b>92.94</b>	
Train set (10-shot):		

Figure 4: **Problem with FID as a few-shot metric.** TGAN (Wang et al., 2018) adaptation from English characters to 10-shot Kannada characters (*Bottom*) (De Campos et al., 2009). The adaptation process is illustrated by interpolating two random latent vectors at different timesteps ( $t=20$  means 20K images seen during training). We measure FID against a 2K-image Kannada set, from which the 10 images was sampled. The interpolation shows larger timesteps ( $t$ ) tend to memorize the 10-image training set while yielding lower FID, revealing that FID favors overfitting and is not suitable for the few-shot setting.

penalty. We retrain the singular values  $\Sigma$  for a fixed number of timesteps (20K images or 16K for 5-shot). We find limiting the training time is essential for quality and diversity in the low-shot setting, as longer training often leads to overfitting or quality degradation (examples in Figure 4 & 7). Like Noguchi & Harada (2019), we use the truncation trick (Brock et al., 2019) during inference, but our method works with a less-restrictive truncation parameter of  $\psi = 0.8$ , which enables more diversity in the generated images.

### 3.4 EVALUATION IN FEW-SHOT SYNTHESIS

A common adverse outcome in few-shot image generation is overfitting to the target set, such that all generated images look similar to the training data. Evaluation metrics should reflect the diversity of generated images, so that memorization is penalized. The standard evaluation practice used in prior low-shot GAN adaptation work (Wang et al., 2018; Noguchi & Harada, 2019; Mo et al., 2020) is to estimate FID (Heusel et al., 2017b) using a large held-out *test set* with 1K+ images, from which the low-shot *training set* was sampled. Standard GAN evaluation typically measures FID with respect to the *training set*, but in the low-shot setting, this is not desirable because the generator may simply memorize the training set. However, we find that even when measuring FID against a held-out test set, this evaluation still favors overfitted or poor-quality models, as shown in Figure 4. FID between real and fake images is calculated as the Frechet distance between perceptual features  $p_r(X)$  and  $p_f(Z)$ :

$$\|\mu_r - \mu_f\|^2 + \text{Tr}(C_r + C_f - 2\sqrt{C_r C_f}). \quad (4)$$

where it is assumed features are Gaussian i.e.,  $p_f(Z) = N(\mu_f, C_f)$  and  $p_r(X) = N(\mu_r, C_r)$ . In the few-shot setting, our  $n$ -shot training set  $T = (x_1, x_2, \dots, x_n)$  is sampled from our test set  $p_r(X)$ . Assuming  $T$  is chosen at random, its sample mean and variance  $\hat{\mu}, \hat{\sigma}^2$  are unbiased estimators of  $\mu_r, C_r$ . Therefore if the generator *memorizes*  $T$ , its statistics approximate  $\mu_r, C_r$ . This artificially decreases the FID of an overfit model (Figure 4). Consequently, we suggest that FID should be supplemented with additional metrics and extensive qualitative results in the low-shot setting. In high-data settings, a very large number of parameters would be required to memorize the images, so this problem is less likely to occur. Based on these observations, throughout our evaluation, we limit training timesteps rather than select the step with the best FID as we find the latter approach gives more inferior qualitative results. To address the limitations of standard metrics for GAN evaluation, we also report sharpness (Kumar et al., 2012) and face quality index (Hernandez-Ortega et al., 2019) for human face transfer.

## 4 EXPERIMENTS

### 4.1 SETTINGS

We adapt a pretrained model to a new target domain using only 5-100 target images, as we focus on scenarios with 1-2 orders fewer number of training samples than standard data-efficient GAN adaptation methods (Wang et al., 2018; Mo et al., 2020; Noguchi & Harada, 2019). As discussed in Section 3.4, we find that the FID score is unsuitable in the low-shot regime due to overfitting bias. However, we still report the FID scores of our experiments for completeness. In addition, we report additional quality metrics and extensive qualitative results.

**Adaptation Methods.** We compare the proposed FSGAN with Transfer GAN (TGAN) (Wang et al., 2018), FreezeD (FD) (Mo et al., 2020), and the Scale & Shift GAN (SSGAN) baseline of Noguchi & Harada (2019). For a fair comparison in the GAN setting, we choose the GAN baseline of SSGAN (Noguchi & Harada, 2019) instead of their GLO-based variant. We implement all methods using the

StyleGAN2 (Karras et al., 2020) codebase.<sup>1</sup> We follow the training setting of StyleGAN, but change the learning rate to 0.003 to stabilize training and reduce the number of training steps to prevent overfitting in the low-shot setting. Figures 4, 7 show comparisons of different training times.

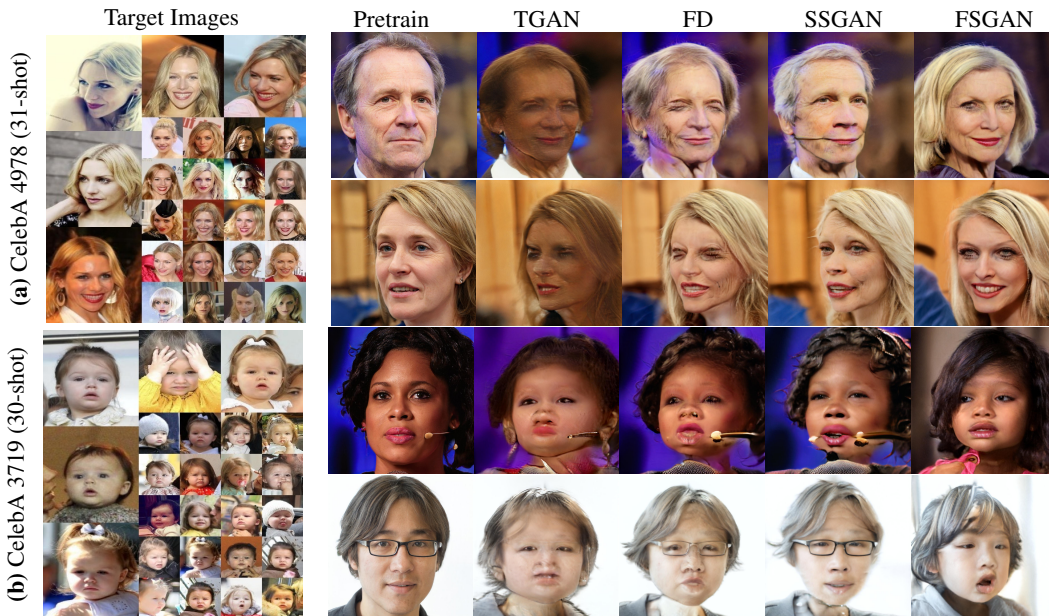


Figure 5: **Close-domain adaptation** (FFHQ→CelebA). Models adapted from a pretrained StyleGAN2 using  $\sim 30$  target images (left-most column) of (a) CelebA ID 4978 and (b) CelebA ID 3719. The proposed FSGAN generates more natural face images without noticeable artifacts. Comparison methods include TGAN (Wang et al., 2018), FD (Mo et al., 2020), SSGAN (Noguchi & Harada, 2019), trained with a limited number of timesteps to prevent overfitting or quality degradation.

Table 1: **Quantitative comparisons** in three metrics: FID (Heusel et al., 2017b), Face Quality Index (FQI) (Hernandez-Ortega et al., 2019), and sharpness (Kumar et al., 2012). See Fig 5 for illustrations. FQI and Sharpness are evaluated on 1,000 images randomly generated with the same set of seeds. Bracketed/bold numbers indicated the best/second best results, respectively.

Method	CelebA 4978			CelebA 3719		
	FID	FQI	Sharpness	FID	FQI	Sharpness
Pretrain	–	0.40±0.11	0.91±0.06	–	0.37±0.12	0.92±0.06
TransferGAN	75.41	0.30±0.07	0.61±0.05	178.31	0.26±0.09	<b>0.61±0.04</b>
FreezeD	75.30	<b>0.33±0.09</b>	0.58±0.04	143.83	<b>0.27±0.09</b>	0.56±0.05
SSGAN	87.79	0.32±0.08	<b>[0.67±0.05]</b>	147.14	0.27±0.10	0.58±0.05
FSGAN (ours)	78.90	<b>[0.36±0.07]</b>	<b>0.65±0.05</b>	170.00	<b>0.27±0.08</b>	<b>[0.68±0.07]</b>

**Datasets.** We used FFHQ (Karras et al., 2019a) and LSUN Churches (Yu et al., 2015) pretrained checkpoints from StyleGAN2 (Karras et al., 2019b), and transferred to few-shot single-ID CelebA (30 or 31 images) (Liu et al., 2015), Portraits (5-100 images) (Lee et al., 2018), Anime ID “Rem” (25 images)<sup>2</sup>, and Van Gogh landscapes (25 images) (Zhu et al., 2017). We evaluate FID against a large test set (10K for CelebA) following the evaluation method of Wang et al. (2018). We also evaluate face quality index (Hernandez-Ortega et al., 2019) and image sharpness (Kumar et al., 2012) for face domain adaptation, using 1000 images from each method generated using identical seeds. Full few-shot target sets are shown in Figures 5 & 6, and we will make all few-shot sets available online.

<sup>1</sup><https://github.com/NVlabs/stylegan2>

<sup>2</sup><https://www.gwern.net/Danbooru2019>

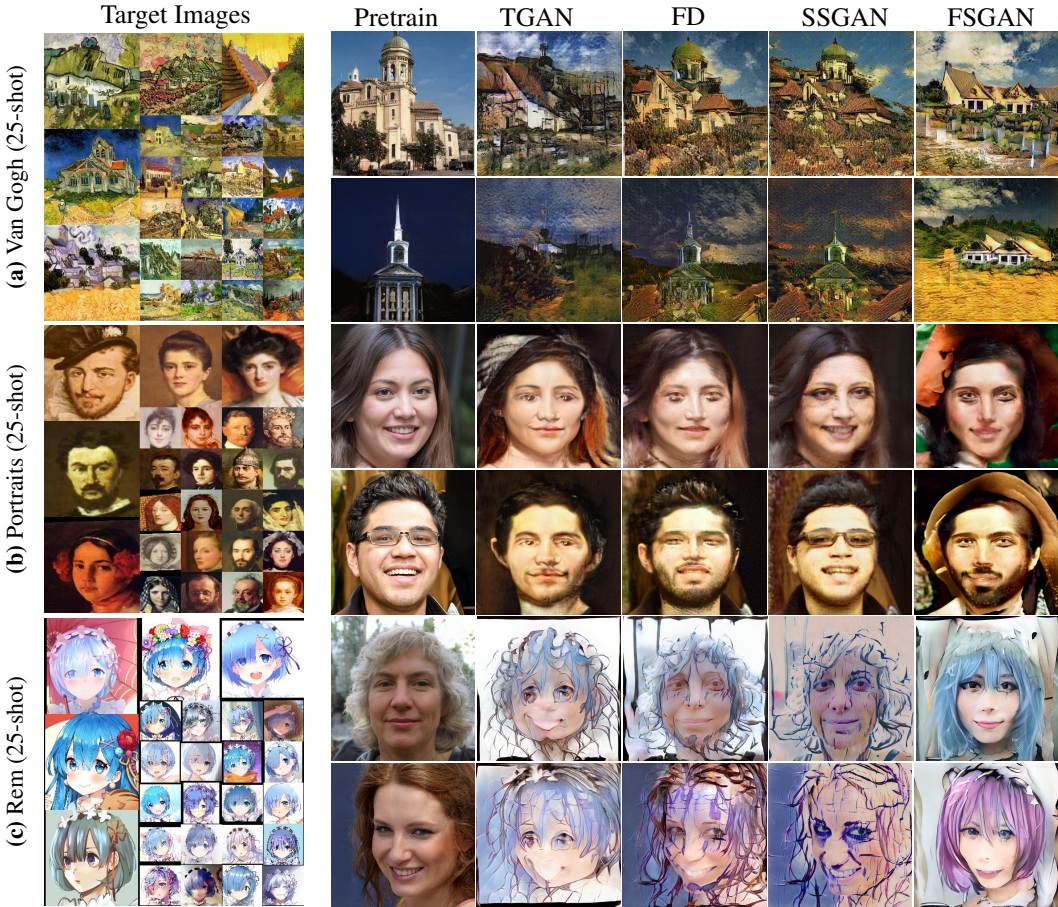


Figure 6: **Far-domain adaptation** (Photo→Art). Comparing FSGAN with alternative GAN adaptation methods in the photo-to-art setting. (a) FSGAN more effectively alters building layouts and adds landscape in the foreground to match the Van Gogh paintings, maintaining better spatial coherency. (b) FSGAN adopts features from the Portraits dataset (hats, beards, artistic backgrounds), while other methods primarily alter image textures. (c) FSGAN transforms natural hair and facial features to imitate the anime target while retaining spatial consistency. Note the occurrence of pink hair in our generated images, which does not exist in the few-shot target but is visually consistent.

#### 4.2 NEAR-DOMAIN ADAPTATION

We first show a *near domain* transfer setting (adapting FFHQ to single-ID CelebA dataset (Liu et al., 2015)). As both source and target domains contain faces, the pretrained model has useful features for the transfer domain. Figure 5 shows that existing GAN adaptation methods produce artifacts around the eyes/chin and low overall structural consistency. In contrast, our method generates more natural face images with characteristics similar to the training samples (e.g., the head size, position of the faces). Comparing Figure 5 and Table 1 shows that the FID correlates poorly with qualitative evaluation for this setting. In light of this, we report additional metrics of face quality (Hernandez-Ortega et al., 2019) and sharpness (Kumar et al., 2012). On these metrics, our method achieves competitive performance across adaptation settings.

#### 4.3 FAR-DOMAIN ADAPTATION

We show *far-domain* 25-shot transfer, where we define “far” as differing significantly in the distribution of image features such as textures, proportions, and semantics. 1) *LSUN Churches*→*Van Gogh paintings*: The two domains differ in the foreground, building shapes, and textural styles. 2) *FFHQ*→*Art portraits*: The main differences between the two domains are low-level styles and facial features. 3) *FFHQ*→*Anime Rem ID*: A challenging setting with exaggerated facial proportions and lack of texture details. Figure 6 shows visual comparisons with three state-of-the-art methods. We

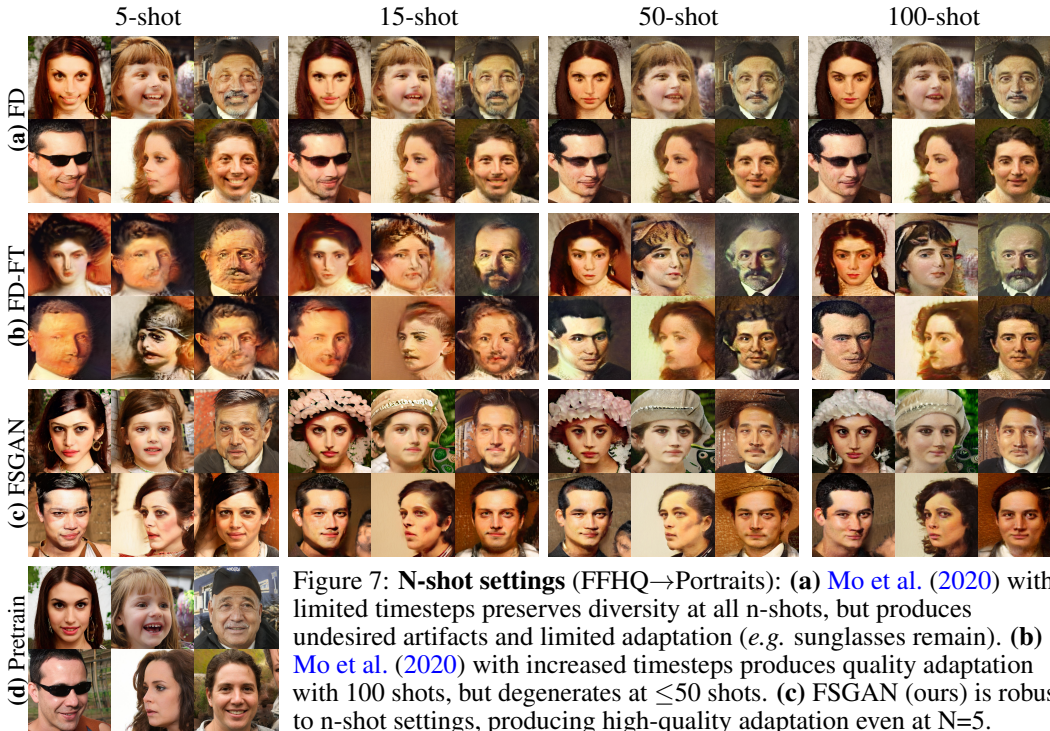


Figure 7: **N-shot settings** (FFHQ→Portraits): **(a)** Mo et al. (2020) with limited timesteps preserves diversity at all n-shots, but produces undesired artifacts and limited adaptation (e.g. sunglasses remain). **(b)** Mo et al. (2020) with increased timesteps produces quality adaptation with 100 shots, but degenerates at  $\leq 50$  shots. **(c)** FSGAN (ours) is robust to n-shot settings, producing high-quality adaptation even at  $N=5$ . **(d)** Pretrained FFHQ images.

find that the proposed FSGAN can adapt the model to produce more dramatic changes to match the target distributions in terms of semantics, proportions, and textures while maintaining high image quality.

#### 4.4 N-SHOT SETTINGS

We test the sensitivity of both FSGAN (ours) and FreezeD (Mo et al., 2020) to differing n-shot settings and show the results in Figure 7. We find that FSGAN is more robust to n-shot setting compared to FreezeD. To show this better, we compare two variations of FreezeD. The first FreezeD variant (FD) is limited in timesteps (20K images / 16K on 5-shot) to match FSGAN and the results reported in Figures 5 & 6. Limiting timesteps prevents degradation that occurs at later iterations in the few-shot settings. However, the time-limited FD produces low quality and limited adaptation of textures and semantic features. The second FreezeD variant (FD-FT) is trained for longer (60K images) to demonstrate (1) degradation in fewer n-shot and (2) improvements in quality/adaptation in higher n-shot as seen in (Mo et al., 2020). In contrast, our method (FSGAN) effectively transfers semantic features while preserving quality across all n-shot settings tested in Figure 7. We note variance across n-shot settings for all methods as the data distribution changes.

## 5 CONCLUSIONS

We presented Few-shot GAN, a simple yet effective method for adapting a pre-trained GAN based model to a new target domain where the number of training images is scarce. Our core idea lies in factorizing the weights of convolutional/fully-connected layers in a pretrained model using SVD to identify a semantically meaningful parameter space for adaptation. Our strategy preserves the capability of a pre-trained model of generating diverse and realistic samples while provides the flexibility for adapting the model to a target domain with few examples. We demonstrate the effectiveness of the proposed method with close-domain and far-domain adaptation experiments and across various n-shot settings. We show favorable results compared with existing data-efficient GAN adaptation methods.



## REFERENCES

- Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein GAN. *arXiv preprint arXiv:1701.07875*, 2017.
- Piotr Bojanowski, Armand Joulin, David Lopez-Paz, and Arthur Szlam. Optimizing the latent space of generative networks. In *ICML*, 2018.
- Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *ICLR*, 2019.
- Teófilo Emídio De Campos, Bodla Rakesh Babu, and Manik Varma. Character recognition in natural images. *VISAPP (2)*, 7, 2009.
- Andrew Gambardella, Atılım Günes Baydin, and Philip H. S. Torr. TransFlow Learning: Repurposing flow models without retraining. In *arXiv*, 2019.
- Gene H Golub and Christian Reinsch. Singular value decomposition and least squares solutions. In *Linear Algebra*, pp. 134–151. 1971.
- Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, 2014.
- Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of Wasserstein GANs. In *NeurIPS*, pp. 5767–5777, 2017.
- Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. GANSpace: Discovering interpretable gan controls. In *ECCV*, 2020.
- Javier Hernandez-Ortega, Javier Galbally, Julian Fierrez, Rudolf Haraksim, and Laurent Beslay. Faceqnet: quality assessment for face recognition based on deep learning. In *2019 International Conference on Biometrics (ICB)*, pp. 1–8. IEEE, 2019.
- 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. In *NeurIPS*, 2017a.
- 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. In *NeurIPS*, 2017b.
- Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-VAE: Learning basic visual concepts with a constrained variational framework. *ICLR*, 2017.
- Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *ICLR*, 2018.
- Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019a.
- Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of StyleGAN. *CoRR*, 2019b.
- Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of StyleGAN. In *CVPR*, 2020.
- Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *ICLR*, 2014.
- Jayant Kumar, Francine Chen, and David Doermann. Sharpness estimation for document and scene images. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pp. 3292–3295. IEEE, 2012.
- Nojun Kwak. Principal component analysis based on l1-norm maximization. *TPAMI*, 30(9):1672–1680, 2008.
- Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *ECCV*, 2018.
- Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *ICCV*, 2015.
- John Martyn, Guifre Vidal, Chase Roberts, and Stefan Leichenauer. Entanglement and tensor networks for supervised image classification. *arXiv preprint arXiv:2007.06082*, 2020.

- Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *ICLR*, 2018.
- Sangwoo Mo, Minsu Cho, and Jinwoo Shin. Freeze Discriminator: A simple baseline for fine-tuning GANs. *arXiv preprint arXiv:2002.10964*, 2020.
- Atsuhiko Noguchi and Tatsuya Harada. Image generation from small datasets via batch statistics adaptation. In *ICCV*, 2019.
- Ben Poole, Alexander A Alemi, Jascha Sohl-Dickstein, and Anelia Angelova. Improved generator objectives for gans. *arXiv preprint arXiv:1612.02780*, 2016.
- Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *ICLR*, 2015.
- Ali Razavi, Aaron van den Oord, and Oriol Vinyals. Generating diverse high-fidelity images with vq-vae2. In *NeurIPS*, 2019.
- Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi. Learning multiple visual domains with residual adapters. In *NeurIPS*, 2017.
- Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training GANs. In *NeurIPS*, 2016.
- Andrew M Saxe, James L McClelland, and Surya Ganguli. A mathematical theory of semantic development in deep neural networks. *Proceedings of the National Academy of Sciences*, 116(23):11537–11546, 2019.
- Hanie Sedghi, Vineet Gupta, and Philip M Long. The singular values of convolutional layers. *arXiv preprint arXiv:1805.10408*, 2018.
- Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. SinGAN: Learning a generative model from a single natural image. In *ICCV*, 2019.
- Assaf Shocher, Shai Bagon, Phillip Isola, and Michal Irani. InGAN: Capturing and retargeting the “dna” of a natural image. In *ICCV*, 2019.
- Yaxing Wang, Chenshen Wu, Luis Herranz, Joost van de Weijer, Abel Gonzalez-Garcia, and Bogdan Raducanu. Transferring GANs: generating images from limited data. In *ECCV*, 2018.
- Abhay Yadav, Sohil Shah, Zheng Xu, David Jacobs, and Tom Goldstein. Stabilizing adversarial nets with prediction methods. *arXiv preprint arXiv:1705.07364*, 2017.
- Shin’ya Yamaguchi, Sekitoshi Kanai, and Takeharu Eda. Effective data augmentation with multi-domain learning GANs. *AAAI*, 2019.
- Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, 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.
- Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-attention generative adversarial networks. In *ICML*, 2019.
- Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017.