WeditGAN: Few-Shot Image Generation via Latent Space Relocation

Yuxuan Duan[1], **Li Niu**[*1], **Yan Hong**[2], **Liqing Zhang**[*1]

[1]MoE Key Lab of Artificial Intelligence, Shanghai Jiao Tong University
[2]Tiansuan Lab, Ant Group
sjtudyx2016@sjtu.edu.cn, ustcnewly@sjtu.edu.cn, yanhong.sjtu@gmail.com, zhang-lq@cs.sjtu.edu.cn

Abstract

In few-shot image generation, directly training GAN models on just a handful of images faces the risk of overfitting. A popular solution is to transfer the models pretrained on large source domains to small target ones. In this work, we introduce WeditGAN, which realizes model transfer by editing the intermediate latent codes w in StyleGANs with learned constant offsets (Δw), discovering and constructing target latent spaces via simply relocating the distribution of source latent spaces. The established one-to-one mapping between latent spaces can naturally prevents mode collapse and overfitting. Besides, we also propose variants of WeditGAN to further enhance the relocation process by regularizing the direction or finetuning the intensity of Δw. Experiments on a collection of widely used source/target datasets manifest the capability of WeditGAN in generating realistic and diverse images, which is simple yet highly effective in the research area of few-shot image generation. Codes are available at https://github.com/Ldhlwh/WeditGAN.

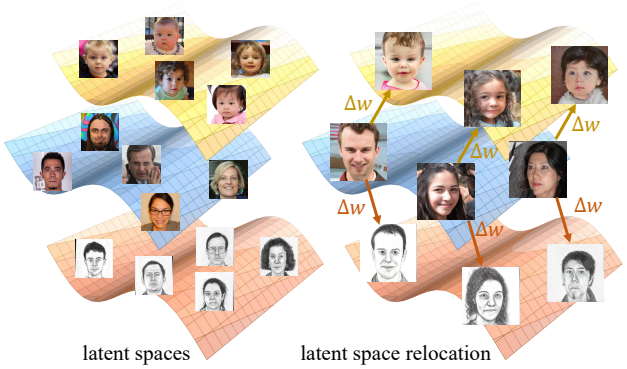

Figure 1: The core idea of latent space relocation with constant latent codes Δw, based on the fact that the latent spaces of related domains in the same generative model share similar shapes of manifolds.

1 Introduction

While achieving convincing performance in many tasks since proposed by Goodfellow et al. (2014), Generative Adversarial Networks (GANs) are renowned for the enormous data they require to possess satisfying fidelity and variety of the generated samples. Mainstream datasets for image generation methods typically cover 20k (CelebA) (Liu et al. 2015), 70k (FFHQ) (Karras, Laine, and Aila 2019) or 126k (LSUN church) (Yu et al. 2015) items. Some researches focus on efficient data usage by enhancing the training processes of GANs with differentiable augmentation (Zhao et al. 2020; Karras et al. 2020a). These methods lower the data threshold to hundreds or thousands, yet leaving training GANs on even fewer images (*e.g.* ten or five) unsolved. Another paradigm of solutions to **few-shot image generation (FSIG)** is model transfer, where GANs pretrained on source domains with large datasets are adapted to target domains with only a handful of images. Finetuning-based methods reduce the number of trainable parameters trying to alleviate the overfitting issue (Wang et al. 2018; Noguchi and Harada 2019; Robb et al. 2020; Mo, Cho, and Shin 2020; Wang et al. 2020), yet the effects are usually limited. Other methods propose regularization terms imposing penalties on feature/parameter changes during the transfer processes (Li et al. 2020; Ojha et al. 2021; Xiao et al. 2022; Zhao et al. 2022b; Hou et al. 2022). By minimally updating the model, these methods seek to keep some characteristics of the source images, hence inherit the diversity of the source domain. Nevertheless, regularization-based methods often face the dilemma of balancing the characteristics of the source/target domains, since sacrificing the characteristics of the target domain impairs the image fidelity.

StyleGANs (Karras, Laine, and Aila 2019; Karras et al. 2020b) are among the most popular GANs nowadays. They not only manifest strong generative abilities, but also construct an intermediate latent space $\boldsymbol{W}$. This latent space is highly disentangled and well aligned to dataset attributes, which enables a variety of latent space manipulation methods for image editing (Collins et al. 2020; Goetschalckx et al. 2019; Härkönen et al. 2020). Most of these methods focus on in-domain manipulation, such as editing the expression, appearance and/or facial pose of a face into another face. However, since some previous works discover that out-of-distribution latent codes may still render reasonable images (Abdal, Qin, and Wonka 2019, 2020; Tov et al. 2021),

*Corresponding authors.

we suppose that cross-domain model transfer in FSIG tasks can also be achieved by proper latent space manipulation.

According to Antoniou, Storkey, and Edwards (2017), the latent spaces of related domains (such as face photos and artistic portraits) in the same model have similar shapes of manifolds, such as the disentangled linear latent space $\boldsymbol{W}$ of StyleGANs (Shen et al. 2020). Previous work StyleCLIP (Patashnik et al. 2021) has validated that for in-domain manipulation, the editing directions are nearly identical when editing the same attribute on different facial images. In preliminary experiments of our work, we find that such identity can be extended to cross-domain scenarios, where the average pairwise cosine similarity of the adaptive editing offsets on different images is over 98% (refer to Appendix for details). Inspired by this idea, we propose WeditGAN (w-edit GAN), which relocates the latent space $\boldsymbol{W}$ of the target domain from that of the source domain by learning a constant latent offset Δw when transferring the pretrained model to the few-shot target domain, as shown in Figure 1. Although such constant Δw implies exactly identical shapes of manifolds between latent spaces, which seems restrictive and may not be absolutely true, we believe it is an adequately good approximation in FSIG, where learning extra modules giving adaptive Δw based on each source latent code are prone to overfitting (see Section 4.3). Instead, using constant Δw naturally establishes a one-to-one mapping between source/target latent codes, perfectly keeps the shape of the source latent space, and eventually prevents overfitting by inheriting the diversity from the source dataset.

Later in Section 4, we compare WeditGAN with several state-of-the-art works on a variety of source/target dataset pairs. Although WeditGAN is a finetuning-based method, experiment results show that it is actually good at maintaining the diversity from the pretrained model as the recent regularization-based ones. Besides, without strict penalties on parameter/feature changes during model transfer, WeditGAN as well outperforms recent methods in capturing characteristics of the target domain, producing images with high fidelity without loss of diversity. Last but not least, We also propose variants of WeditGAN, including a perpendicular regularization on Δw for a more precise relocation, finetuning the editing intensity for better quality, and equipping WeditGAN with contrastive loss to show the orthogonality of our work and the regularization-based methods.

Our contributions can be summarized as follows: (1) We design the simple yet highly effective WeditGAN which transfer the model from the source to the target domain by relocating the latent space with learned constant Δw; (2) We explore several variants of WeditGAN, which further enhance the performance and indicate possible research directions for future works; (3) Extensive experiments on commonly used pairs of source/target datasets for FSIG verify the state-of-the-art performance of WeditGAN.

2 Related Work

Latent Space Manipulation Manipulating the latent space of StyleGAN was mostly done on the extended latent space $\boldsymbol{W}^+$ (Abdal, Qin, and Wonka 2019) or style space $\boldsymbol{S}$ (Liu et al. 2020; Wu, Lischinski, and Shechtman 2021). GAN inversion methods (Abdal, Qin, and Wonka 2019, 2020; Tov et al. 2021; Richardson et al. 2021) provided latent codes of arbitrary in-domain image for editing. Among the manipulation methods, some fused two latent codes to combine certain attributes from the two images for semantic editing (Collins et al. 2020), style transfer (Yang et al. 2022) or face reenactment (Bounareli, Argyriou, and Tzimiropoulos 2022; Tewari et al. 2020). Other methods edited a single latent code along certain directions either supervised (Shen et al. 2020; Goetschalckx et al. 2019; Patashnik et al. 2021) or unsupervised (Härkönen et al. 2020). Nevertheless, most of the previous works focused on in-domain manipulation, leaving cross-domain manipulation rarely researched. Besides, some works extracted latent codes or offsets by training extra encoders (Yang et al. 2022; Tewari et al. 2020), which limited their applicability in few-shot scenarios since these encoders would render collapsed latent spaces given limited data. On the contrary, our WeditGAN proposes constant Δw to bridge the source/target latent spaces without collapse, performing cross-domain manipulation.

Few-shot Image Generation In FSIG, the paradigm of model transfer methods can be divided into two types.

The finetuning-based methods reduce the number of trainable parameters by finetuning a part of the model (Mo, Cho, and Shin 2020; Zhao, Cong, and Carin 2020), training additional parameters while fixing the main model (Noguchi and Harada 2019; Wang et al. 2020) or decomposing the parameters (Robb et al. 2020). Taking image quality and diversity into regard, earlier methods were generally not the most competitive. Nevertheless, two recent works (Zhao et al. 2022a, 2023) probed the important parameters using Fisher Information (Ly et al. 2017) to use different finetuning policies accordingly, reaching comparable results with the regularization-based methods in recent years.

The regularization-based methods usually finetune all the model parameters but introduce penalties on parameter/feature changes and encouraging feature distribution alignment. EWC (Li et al. 2020) decided the strength of penalties via Fisher Information (Ly et al. 2017). CDC (Ojha et al. 2021) maintained feature similarity among source/target samples with consistency loss. RSSA (Xiao et al. 2022) introduced spatial consistency losses to keep structural identity. DCL (Zhao et al. 2022b) proposed contrastive losses between source/target features of both generator and discriminator. DWSC (Hou et al. 2022) designed perceptual/contextual loss respectively for easy/hard-to-generate patches. These methods kept the characteristics of the source domain by imposing strong regularization thus inherited the diversity from the source dataset. Yet they were prone to lose characteristics of the target domain, especially when the characteristics of the two domains are in conflict (see Section 4.2).

Another paradigm of FSIG adopt the idea of *seen categories to unseen categories*. Methods of this paradigm trained an image-to-image model on some categories of a multi-class dataset, and then directly test the model on the other categories (Hong et al. 2020b,a, 2022; Gu et al. 2021; Ding et al. 2022). Besides, there are also some methods for specialized FSIG scenarios, such as font generation (Park

et al. 2021; Tang et al. 2022) and defect generation (Duan et al. 2023). Since the definition to the tasks and/or the experimental settings of these works significantly differ from ours thus less relevant, we will not detail them in this work.

Few-shot Domain Adaptation As a task related to FSIG, the essential distinction between the two tasks is that few-shot domain adaptation (FSDA) requires that the two images before/after adaptation ideally have the same content (*e.g.* depicting the same person), where in FSIG tasks the fidelity to the target domain and the diversity are the only metrics to evaluate the generated images. Recent works of FSDA usually rolled out in one-shot scenarios (Chong and Forsyth 2022; Zhang et al. 2022b; Zhu et al. 2021; Gal et al. 2021; Chefer et al. 2022; Zhang et al. 2022a; Kwon and Ye 2023). Their methods captured the characteristics of a single target image using encoders, instead of a target domain whose distribution is learned by discriminators in image generation tasks. Most of these methods cannot be directly adapted to few-shot scenarios either without non-trivial redesigns.

3 Method

WeditGAN follows the common noise-to-image pipeline of image generation tasks. First, we pretrain a StyleGAN model on a large source domain with abundant data. Then, we transfer the model via latent space relocation to a relevant small target domain containing no more than ten images.

3.1 StyleGAN Preliminary

WeditGAN adopts the most widely used StyleGAN2 (Karras et al. 2020b) as its base model. The generator of StyleGAN consists of a mapping network M and a synthesis network S. The mapping network maps a random noise code z to a latent code w in the intermediate latent space $\boldsymbol{W}$. Then the synthesis network, whose synthesis blocks are made up of convolutional layers, takes w and produces an image I. The whole process can be formulated as

$$z \sim \mathcal{N}(0, \mathbf{I}), \quad w = M(z), \quad I = S(w), \tag{1}$$

where w is transformed by the learned affine layers and modulates the weights of the filters of each convolutional layers in the synthesis network. We follow the normal procedure of pretraining StyleGANs on source datasets.

3.2 WeditGAN

During model transfer from source to target domains, WeditGAN mainly works on the latent space. It seeks an appropriate constant Δw which can relocate the latent space of the target domain from the source latent space by bridging the gap between these two distributions.

Extended Latent Space StyleGAN with default settings uses the same $w \in \boldsymbol{W} \subseteq \mathbb{R}^{512}$ to modulate every convolutional layer. However, to increase flexibility, we conduct latent space relocation on the extended latent space $\boldsymbol{W}^+ \subseteq \mathbb{R}^{n\times 512}$ (Abdal, Qin, and Wonka 2019), which uses separate latent codes for each of the n convolutional layers. Practically, WeditGAN makes n copies of the original w into $w_{\mathrm{src}} \in \boldsymbol{W}^+_{\mathrm{src}}$, and then edits w_{src} in a layer-wise manner, obtaining $w_{\mathrm{tgt}} \in \boldsymbol{W}^+_{\mathrm{tgt}}$ with possibly different latent codes modulating each layer respectively. Ablation study in Section 4.3 demonstrates the necessity of using $\boldsymbol{W}^+$.

Latent Space Relocation The fundamental idea of WeditGAN is to find the target latent space $\boldsymbol{W}^+_{\mathrm{tgt}}$ based on the source space $\boldsymbol{W}^+_{\mathrm{src}}$ with a learned constant latent offset $\Delta w \in \mathbb{R}^{n\times 512}$, so that the StyleGAN pretrained on the source domain can generate target domain images by merely taking target latent codes w_{tgt} instead of w_{src}. As shown in Figure 2, we introduce new parameters Δw to the generator. During the model transfer processes, we only update Δw by training on a few target images, yet fix both the mapping and the synthesis network. Since the parameters of these two networks remain unchanged, WeditGAN still keeps the ability of generating source domain images which previous works fail. WeditGAN is capable of producing paired source/target images using a single model by simply switching between w_{src} and w_{tgt}. The generation process can be formulated as

$$\begin{aligned} &z \sim \mathcal{N}(0, \mathbf{I}), \quad w = M(z), \quad w_{\mathrm{src}} = \mathrm{extend}(w), \\ &w_{\mathrm{tgt}} = w_{\mathrm{src}} + \Delta w, \quad I_{\mathrm{src}} = S(w_{\mathrm{src}}), \quad I_{\mathrm{tgt}} = S(w_{\mathrm{tgt}}). \end{aligned} \tag{2}$$

In total, only $512 \times n$ parameters need training. Since n is typically 20 for StyleGAN2 on 256^2, Δw is only 0.04% of the generator parameters. Compared with the previous methods which finetune the whole model or a large portion of its parameters, WeditGAN is less sensitive to data insufficiency. See Appendix for details.

Objective Besides the generator G, WeditGAN also finetunes the pretrained discriminator D without special strategy, following recent works (Li et al. 2020; Ojha et al. 2021; Xiao et al. 2022). Ablation study in Section 4.3 also manifests that the overfitting issue of the discriminator will be alleviated when the diversity of the generator is assured.

Similar to common GAN models, WeditGAN alternatively updates G and D via the original losses of StyleGAN:

$$L(G, D) = L_{\mathrm{adv}}(G, D) + L_{\mathrm{pl}}(G) + L_{\mathrm{R1}}(G, D), \tag{3}$$

where L_{adv}, L_{pl}, L_{R1} respectively represent adversarial loss, path length regularization encouraging smooth and disentangled latent spaces, and R1 regularization stabilizing the training process by adding gradient penalty. Refer to Karras, Laine, and Aila (2019) for details.

3.3 WeditGAN Variants

Perpendicular Regularization In WeditGAN, the latent offset Δw is learned in a fully data-driven manner. The discriminator provides supervision to Δw so that the relocated $\boldsymbol{W}^+_{\mathrm{tgt}}$ aligns with the target dataset. However, there might be a gap between the dataset distribution P_{data} and the domain distribution P_{domain} due to the potentially biased samples in few-shot datasets (Li et al. 2022). As a result, WeditGAN may relocate $\boldsymbol{W}^+_{\mathrm{tgt}}$ where the distribution of the generated images P_{gen} is closed to P_{data} instead of the expected P_{domain}, by unnecessarily editing in-domain attributes in the source latent space. In order to restrain such unnecessary edits and enhance diversity, we would like to make Δw perpendicular to the manifold of the source latent space. For

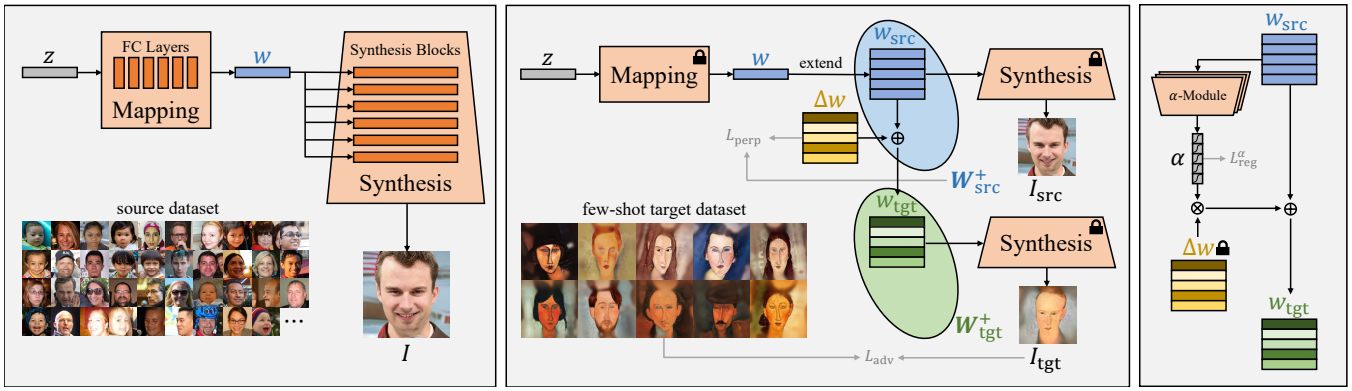

Figure 2: The procedure of WeditGAN. Left: A StyleGAN is first trained on a large source dataset. Middle: During the transfer process, the mapping and the synthesis network are both fixed. With the target latent code w_{tgt} constructed by summing up w_{src} and the only trainable parameters Δw, the synthesis network can generate images of the target domain. Right: After Δw is learned and fixed, WeditGAN trains a set of AlphaModules to finetune the editing intensity customized for each w_{src} (optional).

instance, the learned Δw in WeditGAN transferring from facial photos to sketches should be perpendicular to the editing offset from male to female in the source latent space, as altering the gender is not necessary between the two domains.

It is non-trivial to obtain the closed-form representation of $\boldsymbol{W}^+_{\text{src}}$, and we can only sample w_{src} by random input z and the mapping network M. Since M is a continuous function, the perpendicular regularization between Δw and $\boldsymbol{W}^+_{\text{src}}$ can be formulated as below with a small $\sigma \to 0$:

$$L_{\text{perp}} = \mathbb{E}_{z\sim\mathcal{N}(0,\mathbf{I}),\varepsilon\sim\mathcal{N}(0,\sigma^2\mathbf{I})}[\langle \Delta w, M(z+\varepsilon) - M(z)\rangle^2]. \quad (4)$$

For brevity, Eq. (4) can be approximated by relaxing the neighborhood requirement and using pairs of random w_{src} since $\boldsymbol{W}^+_{\text{src}}$ is a highly disentangled and linear manifold:

$$L_{\text{perp}} = \mathbb{E}_{w^1_{\text{src}}, w^2_{\text{src}}\sim P(w_{\text{src}})}[\langle \Delta w, w^1_{\text{src}} - w^2_{\text{src}}\rangle^2]. \quad (5)$$

During model transfer, for a batch $\{w^i_{\text{src}}\}_{i=1}^m$, the perpendicular loss in Eq. (5) can be computed on-the-fly by

$$L_{\text{perp}} = \sum_{1\le i,j\le m, i\ne j} \langle \Delta w, w^i_{\text{src}} - w^j_{\text{src}}\rangle^2. \quad (6)$$

We choose the inner product $\langle\cdot,\cdot\rangle$ instead of $\cos(\cdot,\cdot)$ because the former can also decrease the magnitude of Δw to avoid excessive edits. The perpendicular loss in Eq. (6) are appended to the original objective function Eq. (3) with a hyperparameter weight λ_{perp}.

Editing Intensity Finetuning WeditGAN relocates the whole latent space with a constant Δw, which is learned to render good images for most $w_{\text{src}} \in \boldsymbol{W}^+_{\text{src}}$. Nevertheless, some w_{src} at the margin of the distribution of $\boldsymbol{W}^+_{\text{src}}$ may still be relocated to suboptimal target images. For example, when transferring the facial photo domain to babies, source latent codes w_{src} corresponding to the elders/youths need more/less edits along the constant Δw learned for the whole source domain. Hence, a customized editing intensity α for each w_{src} may further improve the overall image quality.

After the constant Δw is learned and fixed, an additional lightweighted *AlphaModule* with two FC layers is attached to each synthesis block. These modules are learned to finetune the editing intensity based on each w_{src}. The latent space relocation in Eq. (2) is now formulated as

$$w_{\text{tgt}} = w_{\text{src}} + [1 + \text{AlphaModule}(w_{\text{src}})] \cdot \Delta w, \quad (7)$$

where the output of AlphaModules are activated by $\tanh$ to provide intensity residuals α in $(-1, 1)$ (weakened to strengthened), resulting in customized editing intensity in $(0, 2)$. We add L2 regularization (weighted by hyperparameter $\lambda^\alpha_{\text{reg}}$) to discourage significantly deviated intensities.

Orthogonality to Regularization-based Methods As a finetuning-based method, WeditGAN is orthogonal to the regularization-based methods adding penalties on feature changes. To verify that WeditGAN can be combined with such methods, we introduce another variant equipped with multilayer feature contrastive losses on both G and D, which pull closer the features generated by the same z and push apart those generated by different z. Akin to DCL (Zhao et al. 2022b), with a batch of latent codes $\{w^i_{\text{src}}, w^i_{\text{tgt}}\}_{i=1}^m$ and the corresponding generated images $\{I^i_{\text{src}}, I^i_{\text{tgt}}\}_{i=1}^m$, the contrastive losses can be formulated as:

$$\begin{aligned} L^G_{\text{CL}} &= \sum_l \sum_{i=1}^m -\log \frac{\phi[S^l(w^i_{\text{tgt}}), S^l(w^i_{\text{src}})]}{\sum_{j=1}^m \phi[S^l(w^i_{\text{tgt}}), S^l(w^j_{\text{src}})]}, \\ L^D_{\text{CL}} &= \sum_l \sum_{i=1}^m -\log \frac{\phi[D^l(I^i_{\text{tgt}}), D^l(I^i_{\text{src}})]}{\sum_{j=1}^m \phi[D^l(I^i_{\text{tgt}}), D^l(I^j_{\text{src}})]}, \end{aligned} \quad (8)$$

where $S^l(\cdot)$, $D^l(\cdot)$ is the feature of the l-th synthesis/discriminator block, and $\phi(\cdot,\cdot) = \exp(\text{CosineSimilarity}(\cdot,\cdot))$. These contrastive losses are appended to the original objective function Eq. (3) with a hyperparameter weight λ_{CL}.

4 Experiment

To validate the ability of WeditGAN, we conduct experiments on an extensive set of eight commonly used source/-target dataset pairs. We detail the experimental settings in Section 4.1, illustrate and analyze results in Section 4.2, and investigate WeditGAN through ablation study in Section 4.3.

4.1 Experimental Setting

Dataset Following previous works, we mainly focus on model transfer from face photos to artistic portraits, including **FFHQ** $\rightarrow$ **Sketches** (Wang and Tang 2009), **Babies**, **Sunglasses**, paintings by **Amedeo** Modigliani, **Raphael**, and **Otto** Dix (Yaniv, Newman, and Shamir 2019). We also test WeditGAN on **LSUN Church** $\rightarrow$ **Haunted** houses, and **LSUN Car** $\rightarrow$ **Wrecked** cars. All the target datasets contain only ten training images, with resolution of 256^2.

Baseline We include the regularization-based methods CDC (Ojha et al. 2021), DCL (Zhao et al. 2022b), RSSA (Xiao et al. 2022), and DWSC (Hou et al. 2022). We also include the latest finetuning-based methods AdAM (Zhao et al. 2022a) and RICK (Zhao et al. 2023). Our experiments ensure a fair comparison by transferring from the same pretrained StyleGAN models. See Appendix.

Metric We evaluate both quality and diversity of the generated images. For the three domains (Sketches, Babies, Sunglasses) sampled from larger full datasets, we calculate FID (Heusel et al. 2017) between 5,000 generated samples and the full datasets. For other domains only containing ten images, we compute KID (Bińkowski et al. 2021) instead, which is more precise than FID in few-shot cases (Karras et al. 2020a). Nevertheless, since ten images are probably insufficient to perfectly represent the target domains, such KID scores are shown for reference only.

We also report the intra-cluster version of LPIPS (Zhang et al. 2018) of 1,000 samples as a standalone diversity metric. For details, refer to Ojha et al. (2021) where it originates.

4.2 Few-shot Image Generation

We conduct the experiments with **WeditGAN** and its three variants. **WeditGAN perp** impose perpendicular regularization, with $\lambda_{\text{perp}} = 10^{-4}$. **WeditGAN alpha** finetunes the editing intensity after learning Δw, with $\lambda^{\alpha}_{\text{reg}} = 1/0.1/0.01$ for different cases. **WeditGAN CL** appends contrastive losses on feature changes for multiple layers in both the generator and the discriminator, with $\lambda_{\text{CL}} = 0.5$. Besides 10-shot, we also provide results of FSIG experiments in extreme 5-shot/1-shot settings in Appendix.

Quantitative Result The quantitative results are listed in Tables 1 and 2. Although our baselines consist of the most competitive recent methods, our WeditGAN and its variants still achieve state-of-the-art performance on both metrics. A possible reason is that finetuning the pretrained generator on few-shot datasets may inevitably harm the image quality and cause overfitting. Even though the previous works design specialized strategies trying to counteract such effect, they are still outperformed by WeditGAN which keeps the pretrained generator completely intact.

Qualitative Result We depict the generated samples of Sketches and Babies in Figure 3. See Appendix for the other domains, and Section 4.3 for visual comparisons between WeditGAN and its variants.

As the generated samples show, WeditGAN achieves good balance between inheriting the characteristics of the source domains and capturing those of the target domains. The generated images not only roughly share similar attributes (*e.g.* poses, appearances, expressions) with their corresponding source images from w_{src}, but also present the symbolic attributes of the target domains. In Sketches, our images have similar artistic styles of strokes, shadows, and other facial details with the dataset images. In Babies, WeditGAN learns to produce round faces with large eyes and small noses look natural to infants. With these iconic characteristics covered, WeditGAN attains competitive image fidelity, while still keeping high diversity. Such observation matches the quantitative results in Table 1.

	Sketches		Babies		Sunglasses	
Method	FID↓	LPIPS↑	FID↓	LPIPS↑	FID↓	LPIPS↑
CDC	70.65	0.4412	43.99	0.5859	34.77	0.5873
DCL	57.72	**0.4477**	46.57	0.5833	31.37	0.5844
RSSA	66.97	**0.4448**	55.50	0.5786	27.40	0.5748
DWSC	61.03	0.4095	39.00	0.5604	31.20	0.5799
AdAM	38.11	0.4446	42.44	0.5837	26.98	0.5957
RICK	40.52	0.4310	39.41	0.5709	25.09	0.5992
WeditGAN	**35.41**	0.4339	**38.97**	0.6174	21.72	**0.6254**
+ perp	37.12	**0.4504**	**37.78**	**0.6386**	**19.54**	**0.6424**
+ alpha	**34.13**	0.4250	**36.19**	**0.6296**	**17.06**	0.6223
+ CL	**36.13**	0.4176	40.22	**0.6388**	**19.59**	**0.6472**

Table 1: The results of FSIG on FFHQ $\rightarrow$ Sketches, Babies, and Sunglasses. We report FID@5k with the full datasets, and Intra-cluster LPIPS@1k. Top-3 results are in bold.

Among the baselines, many images rendered by the recent finetuning-based methods AdAM and RICK show unnatural distortions and artifacts. As for the regularization-based methods CDC, DCL, RSSA and DWSC, their images manifest low diversity in local semantic regions (*e.g.* eyes, noses, mouths) possibly due to the usage of patch discriminators.

Sometimes the samples generated by regularization-based methods seem more similar to the corresponding source images than the finetuning-based methods (including WeditGAN). The direct reason is that they impose strong regularization terms penalizing feature changes, which encourage preserving the attributes of the source images. However, such regularization may hinder the model from capturing target characteristics when such characteristics conflict with the source domain. For example, these methods try to produce small and round baby faces while keeping larger and longer adult faces, making the faces obscure and unnatural. Since the ultimate goal of FSIG is to generate images faithful to the target domain, high similarity between source/target images should not be overemphasized. Table 1 have manifested the superiority of WeditGAN over these regularization-based methods with quantitative evaluation objectively without the involvement of the source images.

4.3 Ablation Study

WeditGAN vs. Variants In Figure 4, we provide visual comparisons between WeditGAN and its variants. For WeditGAN perp on Sunglasses, the perpendicular regularization reduces unnecessary edits on facial appearances, hairstyles or clothes. By relocating the latent space according to

Figure 3: The 10-shot datasets (left), the generated samples of source domain FFHQ (top), target domain Sketches (middle) and Babies (bottom). Generated samples in each column are generated with the same random input z.

Method	Amedeo		Raphael		Otto		Haunted		Wrecked	
	KID↓	LPIPS↑	KID↓	LPIPS↑	KID↓	LPIPS↑	KID↓	LPIPS↑	KID↓	LPIPS↑
CDC	23.33	0.5860	8.58	0.5711	16.54	0.6579	23.96	0.6075	25.89	0.4571
DCL	14.64	0.5756	3.66	0.5500	13.66	0.6480	27.69	0.6137	31.54	0.3928
RSSA	20.24	**0.6150**	3.46	0.5545	18.74	0.6308	27.88	0.6011	27.13	0.3739
DWSC	19.79	0.5515	22.63	0.5162	32.68	0.5924	24.45	0.5666	21.08	**0.4883**
AdAM	**12.73**	0.5492	4.70	0.5744	**9.89**	0.6440	24.54	0.6228	**16.34**	0.3968
RICK	18.12	0.5765	4.49	0.5569	11.24	0.6326	21.02	0.6201	35.73	0.3961
WeditGAN	**12.07**	0.5874	**1.68**	**0.6120**	10.78	**0.6834**	**17.40**	0.6427	**20.16**	0.3999
+ perp	17.07	**0.5950**	3.17	**0.6127**	13.85	**0.6943**	**18.65**	**0.6562**	**17.07**	0.4540
+ alpha	**12.67**	0.5895	**0.81**	**0.6016**	**10.55**	0.6832	**14.96**	**0.6440**	27.03	**0.4916**
+ CL	16.86	**0.6150**	**2.16**	0.6011	**7.72**	**0.6924**	19.45	**0.6644**	22.59	**0.4840**

Table 2: The results of FSIG on FFHQ → Amedeo, Raphael, Otto; LSUN Church → Haunted; and LSUN Car → Wrecked. We report KID×10^3@5k, and Intra-cluster LPIPS@1k. Top-3 results are in bold.

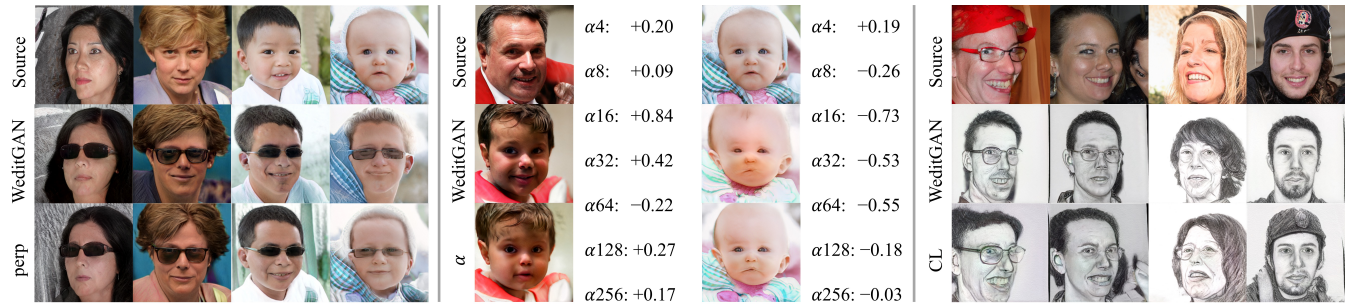

Figure 4: Visual comparisons between WeditGAN and its three variants. Left: WeditGAN perp on Sunglasses. Middle: WeditGAN alpha on Babies. $\alpha 4, \alpha 8, \ldots, \alpha 256$ are the intensity residuals corresponding to the synthesis blocks at resolution $4, 8, \ldots, 256$. Right: WeditGAN CL on Sketches.

Method	Sketches	
	FID↓	LPIPS↑
WeditGAN	35.41	0.4339
w/ unified const Δw	88.73	0.4514
w/ adaptive Δw	37.70	0.2515
+ FreezeD 13	46.21	0.3806
+ FreezeD 19	43.74	0.3837

Table 3: The results of WeditGAN with other relocation methods or partially freezed discriminators on Sketches.

P_{domain} rather than P_{data}, WeditGAN perp generally improves the diversity in Tables 1 and 2. For WeditGAN alpha on Babies, the AlphaModules strengthen the editing intensities for elders and weaken those for children, respectively preventing under-editing or over-editing. Therefore, WeditGAN alpha increases the image fidelity for most cases in Tables 1 and 2. For WeditGAN CL equipped with contrastive losses, the model preserves more source characteristics, as source/target image pairs possess higher resemblance.

WeditGAN Designs To verify that constant Δw in the extended latent space $\boldsymbol{W}^+$ is the best choice, we try other relocation designs. Firstly we use a unified constant $\Delta w \in \boldsymbol{W}$ for all convolutional layers instead of separate ones. Table 3 shows that such unified Δw is not flexible enough to construct the target latent space, thus rendering low quality.

Secondly, we replace constant Δw with adaptive Δw learned from w_{src} via small networks of two FC layers. As in Table 3, adaptive Δw faces overfitting due to data insufficiency. However, when adaptive Δw is used in transferring to the relatively adequate full dataset of Sketches without obvious overfitting, the near-to-one cosine similarities among adaptive Δw have inspired our final design of constant Δw, see Appendix for further investigations.

We also justify that finetuning the whole discriminator will not cause diversity degradation in WeditGAN. We combine WeditGAN with FreezeD (Mo, Cho, and Shin 2020), a popular strategy to finetune discriminators. FreezeD 13/19 fixes the lowest 13/19 layers of the discriminiator, and only the remaining layers are trainable.Table 3 shows that finetuning the full discriminator does not result in overfitting.

5 Conclusion

We propose WeditGAN, a finetuning-based FSIG method. By relocating the latent spaces with learned constant latent offsets, WeditGAN is able to transfer the model pretrained on large source domain to few-shot target ones. Compared to previous works, our method balances well between source/-target domains, generating images both inheriting the diversity of the source domain and faithful to the target domain. The experimental results of WeditGAN and its variants also manifest their capability as a simple yet highly effective method. Possible limitations and future works are discussed in Appendix.

Ethics Statement

Depending on the specific applications, possible societal harms of the few-shot image generation method proposed in this work can be (1) generating fake images for misuse; and (2) copyright violation. The authors hereby solicit proper usage of this work.

Acknowledgements

This work was supported by the Shanghai Municipal Science and Technology Major/Key Project, China (Grant No. 2021SHZDZX0102, Grant No. 20511100300) and the National Natural Science Foundation of China (Grant No. 62076162).

References

Abdal, R.; Qin, Y.; and Wonka, P. 2019. Image2StyleGAN: How to Embed Images Into the StyleGAN Latent Space? In *ICCV*.

Abdal, R.; Qin, Y.; and Wonka, P. 2020. Image2StyleGAN++: How to Edit the Embedded Images? In *CVPR*.

Antoniou, A.; Storkey, A. J.; and Edwards, H. 2017. Data Augmentation Generative Adversarial Networks. *arXiv preprint arXiv:1711.04340*.

Bińkowski, M.; Sutherland, D. J.; Arbel, M.; and Gretton, A. 2021. Demystifying MMD GANs. In *ICLR*.

Bounareli, S.; Argyriou, V.; and Tzimiropoulos, G. 2022. Finding Directions in GAN's Latent Space for Neural Face Reenactment. In *BMVC*.

Chefer, H.; Benaim, S.; Paiss, R.; and Wolf, L. 2022. Image-Based CLIP-Guided Essence Transfer. In *ECCV*.

Chong, M. J.; and Forsyth, D. 2022. JoJoGAN: One Shot Face Stylization. In *ECCV*.

Collins, E.; Bala, R.; Price, B.; and Süsstrunk, S. 2020. Editing in Style: Uncovering the Local Semantics of GANs. In *CVPR*.

Ding, G.; Han, X.; Wang, S.; Wu, S.; Jin, X.; Tu, D.; and Huang, Q. 2022. Attribute Group Editing for Reliable Few-shot Image Generation. In *CVPR*.

Duan, Y.; Hong, Y.; Niu, L.; and Zhang, L. 2023. Few-Shot Defect Image Generation via Defect-Aware Feature Manipulation. In *AAAI*.

Gal, R.; Patashnik, O.; Maron, H.; Chechik, G.; and Cohen-Or, D. 2021. StyleGAN-NADA: CLIP-Guided Domain Adaptation of Image Generators. *arXiv preprint arXiv:2108.00946*.

Goetschalckx, L.; Andonian, A.; Oliva, A.; and Isola, P. 2019. GANalyze: Toward Visual Definitions of Cognitive Image Properties. In *ICCV*.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Nets. In *NeurIPS*.

Gu, Z.; Li, W.; Huo, J.; Wang, L.; and Gao, Y. 2021. LoFGAN: Fusing Local Representations for Few-Shot Image Generation. In *ICCV*.

Härkönen, E.; Hertzmann, A.; Lehtinen, J.; and Paris, S. 2020. GANSpace: Discovering Interpretable GAN Controls. In *NeurIPS*.

Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; and Hochreiter, S. 2017. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In *NeurIPS*.

Hong, Y.; Niu, L.; Zhang, J.; Liang, J.; and Zhang, L. 2022. Deltagan: Towards diverse few-shot image generation with sample-specific delta. In *ECCV*.

Hong, Y.; Niu, L.; Zhang, J.; and Zhang, L. 2020a. MatchingGAN: Matching-based Few-shot Image Generation. In *ICME*.

Hong, Y.; Niu, L.; Zhang, J.; Zhao, W.; Fu, C.; and Zhang, L. 2020b. F2GAN: Fusing-and-Filling GAN for Few-Shot Image Generation. In *ACM MM*.

Hou, X.; Liu, B.; Zhang, S.; Shi, L.; Jiang, Z.; and You, H. 2022. Dynamic Weighted Semantic Correspondence for Few-Shot Image Generative Adaptation. In *ACM MM*.

Karras, T.; Aittala, M.; Hellsten, J.; Laine, S.; Lehtinen, J.; and Aila, T. 2020a. Training Generative Adversarial Networks with Limited Data. In *NeurIPS*.

Karras, T.; Laine, S.; and Aila, T. 2019. A Style-Based Generator Architecture for Generative Adversarial Networks. In *CVPR*.

Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; and Aila, T. 2020b. Analyzing and Improving the Image Quality of StyleGAN. In *CVPR*.

Kwon, G.; and Ye, J. C. 2023. One-Shot Adaptation of GAN in Just One CLIP. *arXiv preprint arXiv:2203.09301*.

Li, Y.; Zhang, R.; Lu, J. C.; and Shechtman, E. 2020. Few-shot Image Generation with Elastic Weight Consolidation. In *NeurIPS*.

Li, Z.; Xia, B.; Zhang, J.; Wang, C.; and Li, B. 2022. A Comprehensive Survey on Data-Efficient GANs in Image Generation. *arXiv preprint arXiv:2204.08329*.

Liu, Y.; Li, Q.; Sun, Z.; and Tan, T. 2020. Style Intervention: How to Achieve Spatial Disentanglement with Style-based Generators? *arXiv preprint arXiv:2011.09699*.

Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015. Deep Learning Face Attributes in the Wild. In *ICCV*.

Ly, A.; Marsman, M.; Verhagen, J.; Grasman, R. P.; and Wagenmakers, E.-J. 2017. A Tutorial on Fisher information. *Journal of Mathematical Psychology*, 80: 40–55.

Mo, S.; Cho, M.; and Shin, J. 2020. Freeze the Discriminator: a Simple Baseline for Fine-Tuning GANs. In *CVPR Workshop*.

Noguchi, A.; and Harada, T. 2019. Image Generation From Small Datasets via Batch Statistics Adaptation. In *ICCV*.

Ojha, U.; Li, Y.; Lu, J.; Efros, A. A.; Lee, Y. J.; Shechtman, E.; and Zhang, R. 2021. Few-Shot Image Generation via Cross-Domain Correspondence. In *CVPR*.

Park, S.; Chun, S.; Cha, J.; Lee, B.; and Shim, H. 2021. Multiple Heads are Better than One: Few-shot Font Generation with Multiple Localized Experts. In *ICCV*.

Patashnik, O.; Wu, Z.; Shechtman, E.; Cohen-Or, D.; and Lischinski, D. 2021. StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery. In *ICCV*.

Richardson, E.; Alaluf, Y.; Patashnik, O.; Nitzan, Y.; Azar, Y.; Shapiro, S.; and Cohen-Or, D. 2021. Encoding in Style: a StyleGAN Encoder for Image-to-Image Translation. In *CVPR*.

Robb, E.; Chu, W.; Kumar, A.; and Huang, J. 2020. Few-Shot Adaptation of Generative Adversarial Networks. *arXiv preprint arXiv:2010.11943*.

Shen, Y.; Gu, J.; Tang, X.; and Zhou, B. 2020. Interpreting the Latent Space of GANs for Semantic Face Editing. In *CVPR*.

Tang, L.; Cai, Y.; Liu, J.; Hong, Z.; Gong, M.; Fan, M.; Han, J.; Liu, J.; Ding, E.; and Wang, J. 2022. Few-Shot Font Generation by Learning Fine-Grained Local Styles. In *CVPR*.

Tewari, A.; Elgharib, M.; Bharaj, G.; Bernard, F.; Seidel, H.-P.; Pérez, P.; Zollhöfer, M.; and Theobalt, C. 2020. StyleRig: Rigging StyleGAN for 3D Control over Portrait Images. In *CVPR*.

Tov, O.; Alaluf, Y.; Nitzan, Y.; Patashnik, O.; and Cohen-Or, D. 2021. Designing an Encoder for StyleGAN Image Manipulation. *arXiv preprint arXiv:2102.02766*.

Wang, X.; and Tang, X. 2009. Face Photo-Sketch Synthesis and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11): 1955–1967.

Wang, Y.; Gonzalez-Garcia, A.; Berga, D.; Herranz, L.; Khan, F. S.; and van de Weijer, J. 2020. MineGAN: Effective Knowledge Transfer From GANs to Target Domains With Few Images. In *CVPR*.

Wang, Y.; Wu, C.; Herranz, L.; van de Weijer, J.; Gonzalez-Garcia, A.; and Raducanu, B. 2018. Transferring GANs: generating images from limited data. In *ECCV*.

Wu, Z.; Lischinski, D.; and Shechtman, E. 2021. StyleSpace Analysis: Disentangled Controls for StyleGAN Image Generation. In *CVPR*.

Xiao, J.; Li, L.; Wang, C.; Zha, Z.-J.; and Huang, Q. 2022. Few Shot Generative Model Adaption via Relaxed Spatial Structural Alignment. In *CVPR*.

Yang, S.; Jiang, L.; Liu, Z.; and Loy, C. C. 2022. Pastiche Master: Exemplar-Based High-Resolution Portrait Style Transfer. In *CVPR*.

Yaniv, J.; Newman, Y.; and Shamir, A. 2019. The Face of Art: Landmark Detection and Geometric Style in Portraits. *ACM Transactions on Graphics*, 38(4).

Yu, F.; Zhang, Y.; Song, S.; Seff, A.; and Xiao, J. 2015. LSUN: Construction of a Large-scale Image Dataset using Deep Learning with Humans in the Loop. *arXiv preprint arXiv:1506.03365*.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *CVPR*.

Zhang, Y.; Yao, M.; Wei, Y.; Ji, Z.; Bai, J.; and Zuo, W. 2022a. Towards Diverse and Faithful One-shot Adaption of Generative Adversarial Networks. In *NeurIPS*.

Zhang, Z.; Liu, Y.; Han, C.; Guo, T.; Yao, T.; and Mei, T. 2022b. Generalized One-shot Domain Adaptation of Generative Adversarial Networks. In *NeurIPS*.

Zhao, M.; Cong, Y.; and Carin, L. 2020. On Leveraging Pretrained GANs for Generation with Limited Data. In *ICML*.

Zhao, S.; Liu, Z.; Lin, J.; Zhu, J.-Y.; and Han, S. 2020. Differentiable Augmentation for Data-Efficient GAN Training. In *NeurIPS*.

Zhao, Y.; Chandrasegaran, K.; Abdollahzadeh, M.; and Cheung, N.-M. 2022a. Few-shot Image Generation via Adaptation-Aware Kernel Modulation. In *NeurIPS*.

Zhao, Y.; Ding, H.; Huang, H.; and Cheung, N.-M. 2022b. A Closer Look at Few-Shot Image Generation. In *CVPR*.

Zhao, Y.; Du, C.; Abdollahzadeh, M.; Pang, T.; Lin, M.; Yan, S.; and Cheung, N.-M. 2023. Exploring Incompatible Knowledge Transfer in Few-shot Image Generation. In *CVPR*.

Zhu, P.; Abdal, R.; Femiani, J.; and Wonka, P. 2021. Mind the Gap: Domain Gap Control for Single Shot Domain Adaptation for Generative Adversarial Networks. In *ICLR*.