A Turn Toward Better Alignment: Few-Shot Generative Adaptation with Equivariant Feature Rotation

Chenghao Xu[1], Qi Liu[2], Jiexi Yan[3], Muli Yang[4], Cheng Deng[1*]

[1] Hohai university, China,
[2] School of Electronic Engineering, Xidian University, China,
[3] School of Computer Science and Technology, Xidian University, China,
[4] Institute for Infocomm Research (I²R), A*STAR, Singapore

Abstract

Few-shot image generation aims to effectively adapt a source generative model to a target domain using very few training images. Most existing approaches introduce consistency constraints—typically through instance-level or distribution-level loss functions—to directly align the distribution patterns of source and target domains within their respective latent spaces. However, these strategies often fall short: overly strict constraints can amplify the negative effects of the domain gap, leading to distorted or uninformative content, while overly relaxed constraints may fail to leverage the source domain effectively. This limitation primarily stems from the inherent discrepancy in the underlying distribution structures of the source and target domains. The scarcity of target samples further compounds this issue by hindering accurate estimation of the target domain's distribution. To overcome these limitations, we propose Equivariant Feature Rotation (EFR), a novel adaptation strategy that aligns source and target domains at two complementary levels within a self-rotated proxy feature space. Specifically, we perform adaptive rotations within a parameterized Lie Group to transform both source and target features into an equivariant proxy space, where alignment is conducted. These learnable rotation matrices serve to bridge the domain gap by preserving intra-domain structural information without distortion, while the alignment optimization facilitates effective knowledge transfer from the source to the target domain. Comprehensive experiments on a variety of commonly used datasets demonstrate that our method significantly enhances the generative performance within the targeted domain.

1. Introduction

In recent years, there has been an exponential advancement in the field of generative vision tasks, particularly with the advent of deep generative models such as Generative Adversarial Networks (GANs). These models have proven to be remarkably successful in numerous tasks, including but not limited to, natural image synthesis [2, 12, 42], image editing [3, 51], and image inpainting [28, 38, 40]. The results yielded by GANs have been highly persuasive, demonstrating their capacity for realistic image generation. However, a significant challenge associated with GANs is their requirement for substantial volumes of data during the training process. For instance, popular datasets employed for GAN training, including FFHQ [12] and LSUN church [39], comprise 70,000 and 126,000 images, respectively. The lack of sufficient training data has been observed to lead to overfitting and collapse of generative models, thereby resulting in suboptimal performance. Consequently, the necessity for large quantities of training data presents a pivotal limitation of GANs, which necessitates prompt attention to enhance the versatility and practicality of these models in real-world applications.

In light of this, an increasing number of researchers [8, 11, 13, 31, 32, 37, 43] are striving to achieve robust image generation in the face of limited training data. A common strategy in this regard involves the adoption of few-shot generative models. This process entails fine-tuning a model that has been pre-trained on a comprehensive dataset from a source domain to accommodate a new domain characterized by limited target data. Through this approach, adaptation methods can produce diverse and realistic images for the target domain, even when faced with as few as ten training images.

The principal challenge in few-shot generative adaptation is to prevent overfitting while maintaining content consistency during the transfer from the source to the target domain. To address this, various loss functions, such as
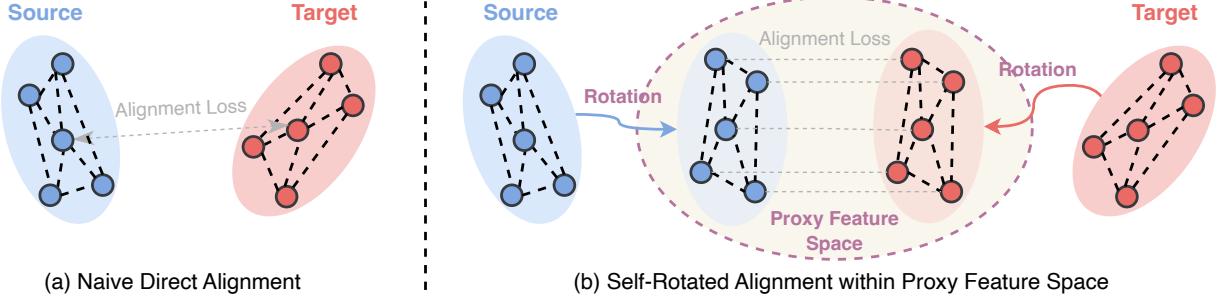
---
*Corresponding author

Figure 1. Comparison between the existing naive direct alignment and our self-rotated alignment.

IDC [25] and RSSA [36], have been proposed to enhance content preservation and structural coherence throughout the adaptation process. Although these approaches have led to notable improvements, directly imposing instance-level and distribution-level alignment between the source and target domains can compromise both training stability and accuracy. This is primarily due to the inherent discrepancy between the underlying distribution patterns of the source and target domains. Furthermore, the limited availability of target samples exacerbates this issue, making it more challenging to accurately represent the target domain's distribution during adaptation.

In this paper, we introduce a novel perspective based on a proxy feature space and propose the **Equivariant Feature Rotation** (EFR) method, a new approach designed to enable stable transfer of relevant content information from the source domain while preserving the model's capacity to acquire style characteristics from the target domain. This is accomplished by aligning the source and target domains at two complementary levels within a self-rotated proxy feature space. Specifically, we first propose an equivariant self-rotated proxy feature space projection strategy, which performs adaptive rotations within a parameterized Lie Group. Instead of directly aligning the source and target data, we transform the source and target domain samples into an equivariant proxy space and perform distribution pattern alignment between source and target domain within this proxy feature space. This strategy significantly improves training stability and preserves the structural integrity of the generated distribution, thereby alleviating the negative impact of domain discrepancies. Following this, we apply both instance-level and distribution-level alignment within the proxy feature space to enforce flexible identity consistency across domains for each instance. Recognizing that direct pairwise alignment may lead to excessive imposition of source domain content, we instead adopt a global distribution pattern alignment based on optimal transport theory. This more relaxed and principled alignment enables effective transfer of semantically meaningful content while avoiding the undesired entanglement of aggressive attribute

information from the source domain.

In summary, our contributions are as follows:

- Current losses enforce local pairwise alignment of generated samples in target and source domains, which does not resolve issue of distribution rotation. Therefore, we propose to autonomously rotate target domain's feature distribution to align with source domain within a proxy feature space, thereby ensuring consistency between domains while maintaining local pairwise alignment.
- Additionally, strong cross-domain alignment in existing losses may overemphasize content from source domain, potentially introducing extraneous information and diminishing generative performance. So, we adopt optimal distribution matching strategy based on OT theory.
- Extensive experiments on several widely used datasets demonstrate that our method effectively improves the generative quality of the target domain.

2. Related Work

2.1. Image Generation with Limited Training Data.

Efforts abound in the realm of instructing a generative model using a limited dataset. A number of previous methodologies have advocated for the application of data augmentation, aimed at curtailing the discriminator's propensity for overfitting. For instance, Zhang and Khoreva [41] introduced a progressive augmentation approach. An in-depth exploration of the effects of diverse augmentations during the training process was undertaken by y Zhao et al. [47], while Tran et al. [31] analyzed the theoretical underpinnings of several data augmentations. Moreover, Zhao et al. proposed the utilization of augmentations on both authentic and fabricated images in a differentiable fashion. Karras et al. [13] conceptualized an adaptive discriminator augmentation, specifically designed to preclude leakage and stabilize the training process. Adding to this, numerous regularizers have been introduced to provide supplementary supervision. For instance, Zhao et al. [48] incorporated consistency regularization for GANs, exhibiting competitive performance even with limited data. To

boost data efficiency, Yang et al. [37] integrated contrastive learning into the training framework as an auxiliary task.

2.2. Few-shot Generative Model Adaptation.

Transfer learning has emerged as a prevalent technique to enhance a model's generalization capabilities within a specific target domain [20, 26, 52], by capitalizing on knowledge acquired from pre-training on a separate source domain encompassing extensive data. Few-shot image generation [1, 6] entails adapting a pre-trained generative model to a novel target domain characterized by limited data. This established paradigm can be bifurcated into two categories: fine-tuning and regularization methods.

Fine-tuning methods address the issue of an overabundance of trainable parameters by updating only a portion of the model or incorporating additional parameters while maintaining the core model intact. Recent advancements in this domain are AdAM [44] and RICK [46], which employ Fisher Information to modulate critical kernels, thereby achieving results on par with regularization-based methodologies. Contrastingly, regularization methods fine-tune all model parameters, imposing penalties on parameter or feature alterations and advocating for the alignment of feature distributions during transfer. For instance, EWC [21] modifies penalties based on Fisher Information to sustain feature consistency between source and target samples. CDC [25] underscores consistency via a loss term. RSSA [36] incorporates spatial consistency losses to preserve structure. DCL [45] introduces contrastive losses for both generator and discriminator features. Lastly, DWSC [10] formulates perceptual and contextual losses for varying patch complexities.

3. Method

3.1. Preliminaries

Problem formulation. Existing few-shot image generation methods focus on a transfer learning paradigm that leverages a source generator $\mathcal{G}_s$ pre-trained on a large-scale dataset such as FFHQ [12] and then adapts it to a new domain with limited target images. During this adaptation procedure, we fine-tune the source GAN on the limited target data to derive a target generator $\mathcal{G}_t$ as follows:

$$
\begin{aligned}
\mathcal{L}_{G_t} &= -\mathbb{E}_{x \sim \mathcal{T}}[\log(\mathcal{D}_t(\mathcal{G}_t(z)))] \\
\mathcal{L}_D &= \mathbb{E}_{x \sim \mathcal{T}}[\log(1 - \mathcal{D}_t(x))] + \mathbb{E}_{z \sim p(z)}[\log(\mathcal{D}_t(\mathcal{G}_t(z)))],
\end{aligned}
\tag{1}
$$

where z denotes a noise vector sampled from noise distribution $p(z)$ including Gaussian distribution, and $\mathcal{T}$ is the target data distribution. Note that $\mathcal{D}_t$ is the discriminator corresponding to $\mathcal{G}_t$. The core goal of few-shot image generation is to capture the inaccessible $\mathcal{T}$.

Main Challenge and Contribution. To achieve better alignment between the source and target distributional patterns, we propose to impose equivariant feature rotation during optimization. The overall framework of the proposed EFR is shown in Figure 2. We first project the source and target data in a common proxy feature space, where the target distributional pattern is self-rotated for better alignment. And then we conduct instance-wise and distribution-level optimization. Specifically, we align the approximate locations of the distribution patterns in source and target domains, improving the training stability. Then, we match the distribution pattern in the target domain with the ones in the source domain from a global perspective, which can effectively maintain identity consistency while preventing undesirable information during generative adaptation.

3.2. Equivariant Self-Rotated Proxy Feature Space Projection

Existing few-shot generative adaptation methods primarily focus on preserving the distribution of generated images to match that of the source domain by employing various loss functions [25, 36]. This is commonly achieved through a combination of local sample-wise alignment and global distribution-level matching between the source and target domains. However, these strategies often suffer from the severe domain gap issue, and hence fail to leverage the source domain effectively, resulting in distorted or uninformative content.

To address this limitation, we propose an *equivariant self-rotated proxy feature space projection* strategy based on a parameterized Lie group. This approach autonomously rotates the target domain's feature distribution within a proxy feature space to align with that of the source domain, thereby enhancing consistency between the two domains. Specifically, we enforce an equivariant rotation parameterized by an orthogonal matrix $\boldsymbol{R}^*$ in the proxy feature space, which facilitates better preservation of structural consistency in the distributional patterns across domains. To better perform orthogonal rotation, we exploit a rotation matrix that lies in a particular orthogonal group, *i.e.*, Lie Group $SO(d)$, which is defined as follows:

$$
SO(d) = \{\boldsymbol{R} \in \mathbb{R}^{d \times d} \,|\, \boldsymbol{R}^\top \boldsymbol{R} = \boldsymbol{I}, \det \boldsymbol{A} = 1\}. \tag{2}
$$

Note that the standard SGD can not assure that $\boldsymbol{R}^*$ always be in $SO(d)$ during training. We tend to address this issue in an algebra way [19], where the Lie Algebra $\mathfrak{so}(d)$ is formed by skew-matrices:

$$
\mathfrak{so}(d) = \{\boldsymbol{R} \in \mathbb{R}^{d \times d} \,|\, \boldsymbol{R} + \boldsymbol{R}^\top = 0\}. \tag{3}
$$

There exists a mapping $\exp(\cdot) : \mathfrak{so}(d) \to SO(d)$ defined as:

$$
\exp(\boldsymbol{R}) = \boldsymbol{I} + \boldsymbol{R} + \frac{\boldsymbol{R}^2}{2} + \cdots \tag{4}
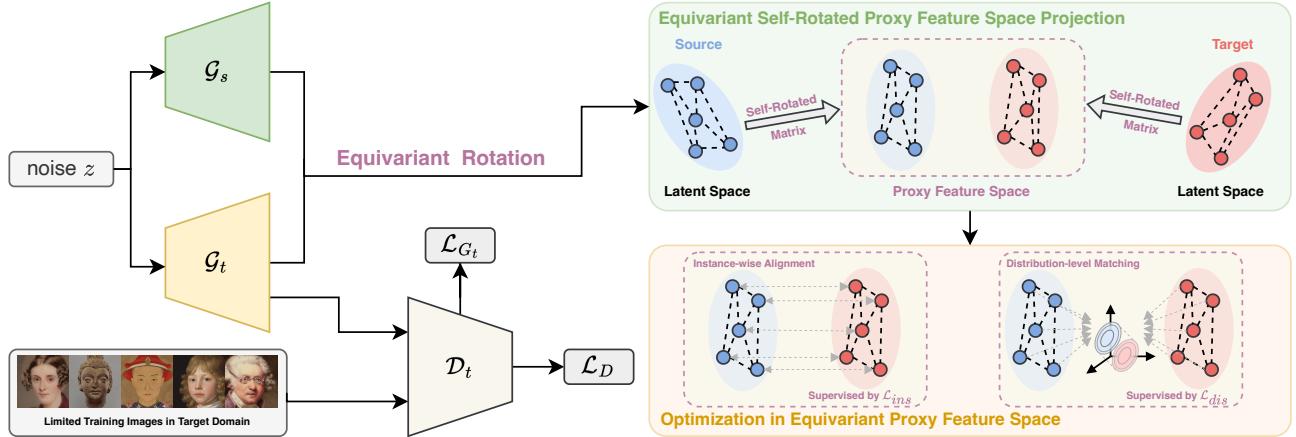$$

Figure 2. The overall framework of our proposed method.

Therefore, the optimization in $SO(d)$ could be transformed into optimization in $\mathfrak{so}(d)$. Furthermore, the Lie Algebra $\mathfrak{so}(d)$ is isomorphic to a linear space. The isomorphism mapping is given by $\boldsymbol{R} \rightarrow \boldsymbol{R} - \boldsymbol{R}^{\top}$. Consequently, the optimization with orthogonal constraint is transformed into the optimization in $\mathbf{R}^{d \times d}$ as follows:

$$\min_{\boldsymbol{R} \in SO(d)} \mathcal{L}(\boldsymbol{R}) \iff \min_{\hat{\boldsymbol{R}} \in \mathbb{R}^{d \times d}} \mathcal{L}(\exp(\hat{\boldsymbol{R}} - \hat{\boldsymbol{R}}^{\top})) \quad (5)$$

In the above formulation, the optimization with orthogonal constraint is transformed into the optimization in $\mathbb{R}^{d \times d}$. Therefore, we can straightly adopt standard optimization techniques such as SGD [29] and Adam [17] for the right side of Eq.(5).

During optimization, simultaneously rotating both the source and target domains is equivalent to rotating only a single domain. Furthermore, rotating a single domain is computationally more efficient. Therefore, in the subsequent optimization process, we apply rotation exclusively to the target domain. Additional implementation details and mathematical proof are provided in the supplementary material.

3.3. Optimization in Equivariant Proxy Feature Space

Instance-wise Alignment. Here, we adopt a contrastive loss for local instance-wise optimization, which tends to produce a smooth content information transition from the source domain to the target domain within the equivariant proxy feature space. Specifically, given a batch of N noise input $\{z_i\}_{i=1}^{N}$, we could obtain the corresponding intermediate feature maps $\boldsymbol{I}_i^s$ and $\boldsymbol{I}_i^t$ in the source and target generators $\mathcal{G}_s(\cdot)$ and $\mathcal{G}_t(\cdot)$, respectively. The self-rotated instance-

wise alignment loss is defined as follows:

$$\mathcal{L}_{ins} = \sum_{i=1}^{N} -\log \frac{\exp(\text{sim}(\boldsymbol{I}_i^s, \exp(\hat{\boldsymbol{R}} - \hat{\boldsymbol{R}}^{\top})\boldsymbol{I}_i^t)/\tau)}{\sum_{j=1}^{N} \exp(\text{sim}(\boldsymbol{I}_i^s, \exp(\hat{\boldsymbol{R}} - \hat{\boldsymbol{R}}^{\top})\boldsymbol{I}_j^t)/\tau)},$$
(6)

where $\text{sim}(\cdot, \cdot)$ denotes the cosine similarity, and τ is the temperature hyperparameter.

Distribution-level Matching. For the existing few-shot generative adaptation losses [36], the direct strong alignment of cross-domain pairwise relationships may result in an overemphasis of content information from the source domain. This over-saturation can introduce unwelcome information, thereby deteriorating the performance of the generation. To address this challenge, we shift our approach to embrace a global-scale optimal distribution matching strategy, grounded in optimal transport theory [27, 30]. Specifically, we propose a distribution-level variational regularizer that penalizes the inter-domain distribution discrepancy via the intra-domain variations. We first calculate the similarity graphs of intermediate features in the source and target domains $G_s = \{\text{sim}(\boldsymbol{I}_i^s, \boldsymbol{I}_j^s)\}_{i,j=1}^{N}$ and $G_t = \{\text{sim}(\boldsymbol{I}_k^t, \boldsymbol{I}_l^t)\}_{k,l=1}^{N}$, respectively. Then, we can indicate the inter-domain distribution discrepancy between the intermediate features in the source and target domains according to the discrepancy between intra-domain variations G_s and G_t. In this procedure, we adopt the discrete optimal transport to measure the inter-domain discrepancy since it can effectively induce the intrinsic geometries of distributions. The corresponding Gromov Wasserstein distance between distributions $\mathcal{P}_s$ and $\mathcal{P}_t$ is formulated as follows:

$$\mathcal{W}(\mathcal{P}_s, \mathcal{P}_t) = \sum_{i,j,k,l} |\text{sim}(\boldsymbol{I}_i^s, \boldsymbol{I}_j^s) - \text{sim}(\exp(\hat{\boldsymbol{R}} - \hat{\boldsymbol{R}}^{\top})\boldsymbol{I}_k^t,$$
$$\exp(\hat{\boldsymbol{R}} - \hat{\boldsymbol{R}}^{\top})\boldsymbol{I}_l^t)|^2 \Lambda_{ik} \Lambda_{jl},$$
(7)

where $|\cdot|$ denotes the ℓ_1 norm, Λ_{ik} and Λ_{jl} are the corresponding items of coupling matrix $\Lambda \in \mathbb{R}^{N \times N}$ that is constrained to satisfy $\Lambda \mathbb{1}_N = \rho$ and $\Lambda^\top \mathbb{1}_N = \varrho$, where $\mathbb{1}_N$ denotes a n-dimensional all-one vector and ρ, ϱ are weight vectors associated with $\frac{s}{i}, I_k^t$. In this paper, we set $\rho_i = 1/N, \varrho_k = 1/N, i, k \in [1, 2, c \ldots, N]$.

Given that the solution for the distribution equalization, as described in Eq.(7), poses a non-convex optimization problem, we employ the sliced Gromov Wasserstein distance for its resolution. Specifically, we project the learned metric space into various one-dimensional spaces using random directions. Through this approach, the sliced Gromov-Wasserstein distance can be effectively approximated by capturing sample observations from the distributions. Formally, the sliced Gromov Wasserstein distance with T projection vectors $\{\pi_t\}_{t=1}^T$ is easy to calculated as follows:

$$
\begin{aligned}
\mathcal{L}_{dis} = \frac{1}{T} \sum_{t=1}^T \sum_{i,j,k,l} &|\mathrm{sim}\left(\langle I_i^s, \pi_t \rangle, \langle I_j^s, \pi_t \rangle\right) \\
&- \mathrm{sim}(< \exp(\hat{R} - \hat{R}^\top)I_k^t, \pi_t >, \\
&< \exp(\hat{R} - \hat{R}^\top)I_l^t, \pi_t >)|^2 \Lambda_{ik}\Lambda_{jl}.
\end{aligned}
\tag{8}
$$

Overall. The eventual loss function for optimization can be summarized as follows:

$$
\mathcal{L} = \mathcal{L}_{G_t} + \mathcal{L}_D + \lambda_1 \mathcal{L}_{ins} + \lambda_2 \mathcal{L}_{dis},
\tag{9}
$$

where λ_1, λ_2 are the hyperparameters.

4. Experiments

In this section, we evaluate our proposed SeAM and analyze its essential characteristics.

4.1. Experimental Settings

Experimental Implementation. For a fair comparison, we follow the standard experimental protocol [25, 44] as previous works and explore different $source \rightarrow target$ adaptation settings to analyze the effectiveness of our method. Model adaptations are done in a 10-shot setting. In all experiments, StyleGAN-V2 [16] is employed as the GAN architecture for pre-training and fine-tuning. We train our models on an NVIDIA A6000. We operate on 256 × 256 resolution images for both the source and target domains. We train the generator and discriminator by Adam optimizer [17] with the same hyperparameters (learning rate, β_1, and β_2 are set as 0.002, 0, and 0.99, respectively) as previous works. We set λ_1 and λ_2 to 0.6 and 0.4 in all the experiments. We set the size of the mini-batch as 8 and trained for about 1000 iterations. To comprehensively evaluate the effectiveness of the proposed model, we compare it with various state-of-the-art methods based on different backbones. Specifically, we include

methods built upon Diffusion Models (DMs): Domain-Gallery [7], DDPM-PA [49], CRDI [4], and DomainStudio [50]; a method based on StyleGAN-ADA [14]: Wedit-GAN [8]; as well as methods based on the StyleGAN2 [15] backbone: TGAN [35], TGAN+ADA [13], FreezeD [23], EWC [21], MineGAN [34], CDC [25], RSSA [36], So-LAD [24], AdAM [44], PIP [22], and RICK [46].

Datasets. Following previous works, we use the GAN pre-trained on a large-scale image dataset FFHQ [12]. For few-shot adaptations, we select several target domains that have different proximity to the source dataset: 1) Semantic-related domains: Babies(B) [25], Sunglasses(S) [25], Sketches(Skt) [33], and MetFaces(MF) [13]; 2) Distant domains: AFHQ-Cat(AC) [5], AFHQ-Dog(AD) [5], and AFHQ-Wild(AW) [5].

To augment the evaluation of the efficacy of the sophisticated SeAM methodology, we incorporate two supplementary experimental trials for alternate source domains. We select LSUN Churches [39] and LSUN-Stanford car [18] datasets as source domain datasets. Subsequently, we correspondingly adapt them to the haunted house [25] and wrecked cars [25] datasets.

Evaluation Metrics. For a more robust demonstration of the effectiveness of our proposed methodology, we utilize three separate evaluation methods to measure not only the quality but also the diversity of the images generated by our technique in conjunction with those created by the comparative baselines. We employ the Fréchet Inception Distance (FID) [9], a widely accepted metric, to quantify the divergence between the fitted Gaussian distribution of authentic and generated samples. Additionally, we utilize the intra-cluster variant of the Learned Perceptual Image Patch Similarity (intra-LPIPS) to ascertain the variance within the collection of images generated by our method [25]. As a final point, we visually display the generated images to provide a more instinctive comparison.

4.2. Evaluation Results

Qualitative Result. The generated samples on FFHQ $\rightarrow$ Sketches and FFHQ $\rightarrow$ AFHQ-Cats, employing varying few-shot generative adaption methods are shown in Figure 4. In the adaptation FFHQ $\rightarrow$ Sketches, MineGAN demonstrates a significant overfitting to the samples derived from the target domain. When juxtaposed with MineGAN, AdAM does not yield any enhancement in the outcomes. In the adaptation process FFHQ $\rightarrow$ AFHQ-Cats, both RSSA and CDC exhibit underfitting towards the target source, implying that the generated feline images via these methods still maintain certain human traits. As evidenced by the generated images, our proposed SeAM method delivers superior generation performance due to its ability to adeptly inherit the characteristics while effectively encapsulating those from the target domains. This underscores the effi-

Table 1. The results of FID (↓) of the few-shot image generation experiments on seven different target domains. The best results are indicated as **Bold**, and the second ones are indicated as <u>Underline</u>.

	Backbone	B	S	MF	Skt	AC	AD	AW	Mean
TGAN	StyleGAN2	101.58	55.97	76.81	53.42	64.68	151.46	81.30	83.60
TGAN+ADA	StyleGAN2	97.91	53.64	75.82	66.99	80.16	162.63	81.55	88.39
FreezeD	StyleGAN2	96.25	46.95	73.33	46.54	63.60	157.98	77.18	80.26
EWC	StyleGAN2	79.93	49.41	62.67	64.55	74.61	158.78	92.83	83.25
CDC	StyleGAN2	69.13	41.45	65.45	47.62	176.21	170.95	135.13	100.85
RSSA	StyleGAN2	66.81	42.03	63.97	69.51	159.54	169.84	100.40	96.01
SoLAD	StyleGAN2	52.01	33.05	54.64	37.23	61.35	112.91	55.27	43.78
AdAM	StyleGAN2	48.83	28.03	51.34	42.64	58.07	100.91	36.87	52.38
PIP	StyleGAN2	-	29.28	-	37.40	-	-	-	-
RICK	StyleGAN2	<u>39.39</u>	<u>25.22</u>	<u>48.53</u>	<u>35.66</u>	**53.27**	<u>98.71</u>	**33.02**	<u>47.69</u>
Ours	StyleGAN2	**37.16**	**24.98**	**46.03**	**33.05**	<u>53.91</u>	**97.22**	<u>34.31</u>	**46.67**
DomainGallery	SD1.4	58.86	43.10	60.38	44.86	77.15	123.54	65.32	78.22
DDPM-PA	DDPM	48.92	34.75	55.39	39.68	69.22	58.27	60.24	52.35
CRDI	DDPM	48.52	24.62	51.28	36.59	65.30	54.35	68.31	49.85
DomainStudio	DDPM	33.26	**21.92**	-	-	-	-	-	-
Ours	DDPM	**32.65**	22.98	**31.44**	**26.67**	<u>43.56</u>	**77.52**	<u>28.15</u>	**37.57**
RICK*	StyleGAN-ADA	52.01	33.05	54.64	37.63	61.35	12.91	55.27	43.78
WeditGAN	StyleGAN-ADA	48.83	28.03	51.34	35.41	58.07	100.91	36.87	52.38
Ours	StyleGAN-ADA	**32.74**	**16.35**	**40.02**	**33.05**	<u>53.91</u>	**97.22**	<u>34.31</u>	**46.67**

Figure 3. Generative adaptation results of our method on `Cars → Wrecked cars` and `Church → Haunted house` (10-shot). Best zoomed in and viewed in color.

cacy of our unified knowledge embedding strategy.

To further substantiate the universal applicability of our methodology, we executed experiments in a 10-shot setting across three distinct scenarios: `Cars → Wrecked cars` `Church → Haunted house`. The empirical in Figure 3 outcomes indicate that our method exhibits exceptional performance irrespective of the pre-training models utilized.

To provide a comprehensive demonstration of the efficacy of our Equivariant Self-Rotated Proxy Feature Space Projection strategy, we also undertake a comparison of latent space interpolation. We subtly adjust the metric by selecting 10 subintervals between any two latent vectors, thereby generating corresponding images by source model. Concurrently, we generate images using the target model with the same latent vectors. As shown in Figure 5, it becomes evident that the data distribution within the target domain aligns with the source domain. For instance, the rotation of a person's head in the source domain corresponds to the angle in the target domain. This demonstrates that the proposed proxy space enables a smooth and stable alignment of spatial distributions.

To more comprehensively validate the effectiveness of our self-adaptive rotation strategy, we visualize the generative outcomes of the rotated features, as illustrated in Figure 7. Specifically, subfigure (a) depicts the image generated from the source feature using the source generator, $\mathcal{G}_s(I^s)$; (b) shows the output $\mathcal{G}_s(R^*I^t)$, where the target feature is first rotated; (c) presents the generated image $\mathcal{G}_t(I^t)$ from the target generator; and (d) illustrates $\mathcal{G}_t((R^*)^\top I^s)$, where the source feature is inversely rotated. The visual comparison indicates that the rotation matrix primarily transfers domain-specific characteristics without introducing additional semantic content. This supports our claim that the self-rotated proxy feature space effectively bridges the domain gap while preserving the intrinsic content of the original representations.

Quantitative Result. In order to further delineate the caliber and heterogeneity of the synthesized images, the
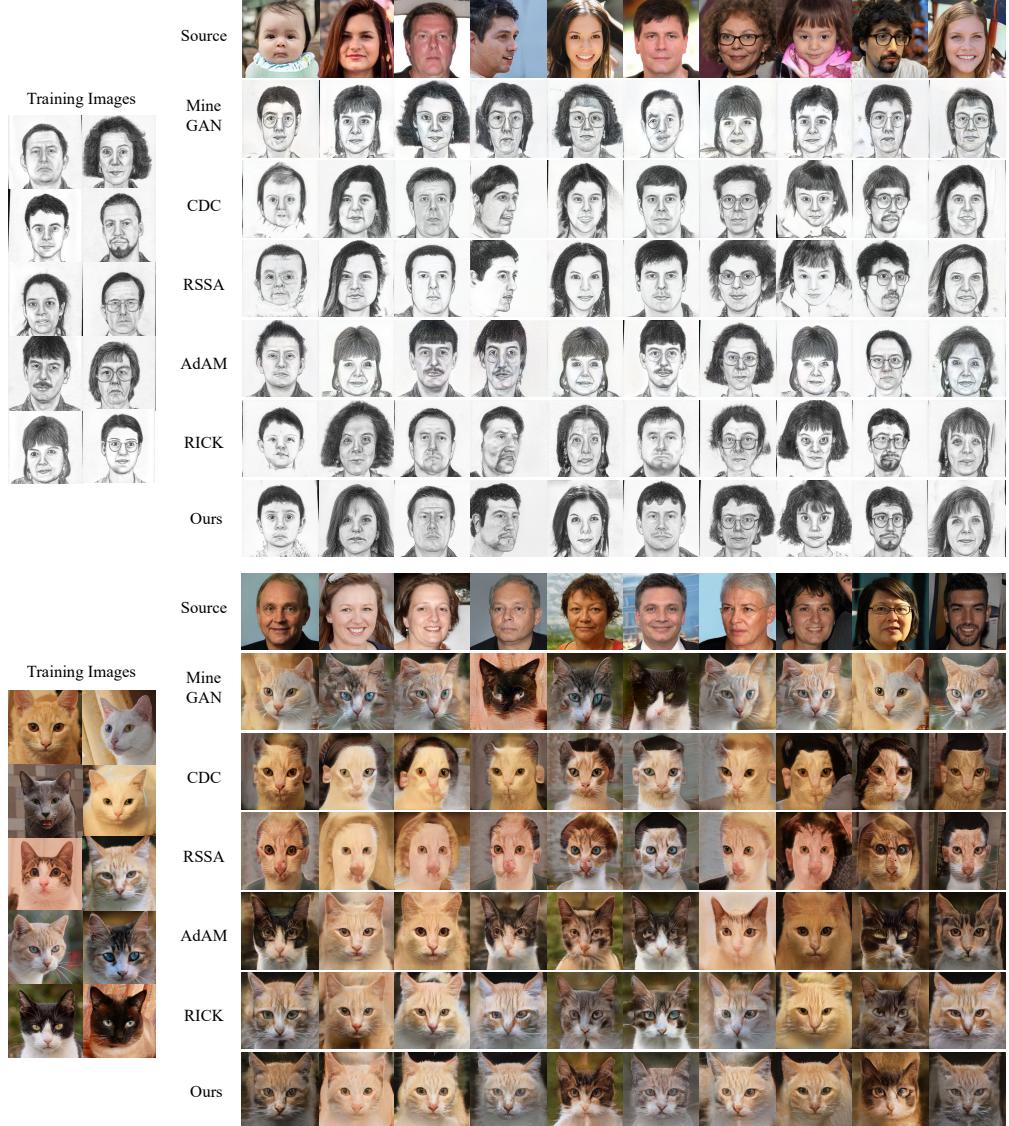
Figure 4. Visualized comparison results with different methods on FFHQ → Sketches and FFHQ → AFHQ-Cat adaptations. The target training data is under the 10-shot setting. Synthesized samples in each column are generated with the same random input **z**. Best zoomed in and viewed in color. Best zoomed in and viewed in color.
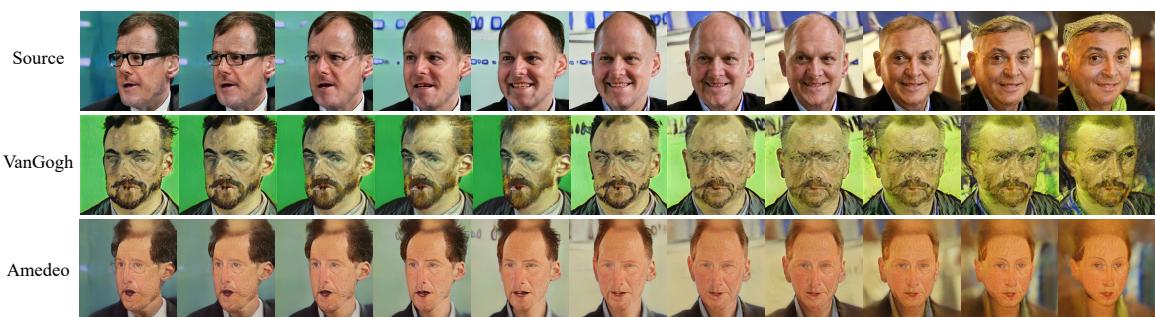

Figure 5. Latent space interpolation results on FFHQ→VanGoGh and FFHQ→Amedeo

Table 2. The results of intra-LPIPS (↑) on three target domains.

	B	AC	Skt
TGAN	0.517	0.490	0.386
TGAN+ADA	0.511	0.513	0.344
FreezeD	0.518	0.492	0.351
EWC	0.521	0.587	0.423
CDC	0.578	**0.629**	0.418
RSSA	0.582	<u>0.612</u>	0.478
SoLAD	0.587	0.601	0.483
AdAM	0.590	0.557	0.482
RICK	<u>0.608</u>	0.569	0.493
DomainStudio	0.599	-	<u>0.495</u>
Ours	**0.613**	<u>0.612</u>	**0.511**

Table 3. The ablation study of different losses and the effect of rotation for EFR. R^* indicates whether the rotation strategy is applied. Results are evaluated using FID (↓).

R^*	$\mathcal{L}_{ins}$	$\mathcal{L}_{dis}$	B	S	MF
		✓	54.26	41.77	65.37
✓		✓	43.91	37.41	60.32
		✓	51.35	40.38	64.18
✓		✓	42.88	34.69	62.14
	✓	✓	46.33	29.05	54.64
✓	✓	✓	**37.16**	**24.98**	**46.03**

Figure 6. The FID results with regard to λ_1 and λ_2 on different datasets.

Figure 7. The generative visualization of rotation.

Fréchet Inception Distance (FID) [9] and intra-cluster version of the Learned Perceptual Image Patch Similarity (intra-LPIPS) [25] scores are utilized as quantitative measures. The comprehensive numerical findings pertaining to several benchmark datasets are documented in Table 1 and Table 2. The target datasets comprise a significant volume of images, for instance, 5000 instances in the AFHQ-Cat dataset. It is feasible to employ the entire target dataset for evaluative purposes by using our refined generator to generate an equivalent quantity of images arbitrarily. Among the entirety of these outcomes, our methodology exhibits superior performance on both metrics, thereby evidencing the efficacy of our proposed technique.

4.3. Ablation Study

The Impact of Diverse Compositional Losses and Rotation. To systematically evaluate the influence of different loss components, we conduct an ablation study considering various combinations of $\mathcal{L}_{ins}$ and $\mathcal{L}_{dis}$. The resulting FID scores are summarized in Table 3. It is evident that both losses contribute positively to the performance of the generation model. To further assess the effectiveness of the rotation mechanism, additional ablation experiments are car-

ried out. As shown in Table 3, incorporating rotation consistently improves performance across all settings, demonstrating its beneficial impact on few-shot image generation.

The Impact of different weights for Losses. To meticulously analyze the influence of assorted λ_1 and λ_2 variables, we conduct ablation studies on the hyperparameters λ_1 and λ_2 using a grid search approach over the Babies and Sunglasses datasets. As shown in the corresponding Figure 6, the performance drops significantly when both λ_1 and λ_2 are set to zero, indicating the importance of the proposed losses. It is evident that the quantitative values of the hyperparameters exert minimal effect on the outcomes, implying that the $\mathcal{L}_{ins}$ and $\mathcal{L}_{dis}$ sensitivity are not pronounced. Based on the results, we select $\lambda_1 = 0.6$ and $\lambda_2 = 0.4$ as the final hyperparameters configuration.

5. Conclusion

This paper presents Equivariant Feature Rotation (EFR), a novel and effective strategy for few-shot generative adaptation. By introducing a self-rotated proxy feature space through learnable Lie Group-based rotations, EFR addresses the fundamental challenge of misaligned distribution patterns between source and target domains.

This approach not only preserves the structural integrity of intra-domain representations but also facilitates robust cross-domain alignment, enabling more stable and informative generation. Extensive empirical evaluations across multiple benchmarks confirm that EFR significantly mitigates overfitting and content degradation, outperforming existing methods in terms of both content fidelity and adaptation efficiency. These findings underscore the potential of equivariant transformations in advancing the capabilities of few-shot generative modeling.

References

[1] Stella Bounareli, Vasileios Argyriou, and Georgios Tzimiropoulos. Finding directions in gan's latent space for neural face reenactment. *BMVC*, 2022. 3

[2] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *ICLR*, 2018. 1

[3] Meng Cao, Haozhi Huang, Hao Wang, Xuan Wang, Li Shen, Sheng Wang, Linchao Bao, Zhifeng Li, and Jiebo Luo. Unifacegan: a unified framework for temporally consistent facial video editing. *IEEE TIP*, 2021. 1

[4] Yu Cao and Shaogang Gong. Few-shot image generation by conditional relaxing diffusion inversion. In *European Conference on Computer Vision*, pages 20–37. Springer, 2024. 5

[5] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *CVPR*, pages 8188–8197, 2020. 5

[6] Edo Collins, Raja Bala, Bob Price, and Sabine Susstrunk. Editing in style: Uncovering the local semantics of gans. In *CVPR*, 2020. 3

[7] Yuxuan Duan, Yan Hong, Bo Zhang, Huijia Zhu, Weiqiang Wang, Jianfu Zhang, Li Niu, Liqing Zhang, et al. Domaingallery: Few-shot domain-driven image generation by attribute-centric finetuning. *Advances in Neural Information Processing Systems*, 37:537–561, 2024. 5

[8] Yuxuan Duan, Li Niu, Yan Hong, and Liqing Zhang. Weditgan: Few-shot image generation via latent space relocation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1653–1661, 2024. 1, 5

[9] 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. *NeurIPS*, 2017. 5, 8

[10] Xingzhong Hou, Boxiao Liu, Shuai Zhang, Lulin Shi, Zite Jiang, and Haihang You. Dynamic weighted semantic correspondence for few-shot image generative adaptation. In *ACM MM*, 2022. 3

[11] Jiaxing Huang, Kaiwen Cui, Dayan Guan, Aoran Xiao, Fangneng Zhan, Shijian Lu, Shengcai Liao, and Eric Xing. Masked generative adversarial networks are data-efficient generation learners. *NeurIPS*, 2022. 1

[12] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019. 1, 3, 5

[13] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *NeurIPS*, 2020. 1, 2, 5

[14] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *Advances in neural information processing systems*, 33:12104–12114, 2020. 5

[15] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020. 5

[16] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *CVPR*, 2020. 5

[17] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4, 5

[18] Tin Kramberger and Božidar Potočnik. Lsun-stanford car dataset: enhancing large-scale car image datasets using deep learning for usage in gan training. *Applied Sciences*, 10(14): 4913, 2020. 5

[19] Mario Lezcano-Casado and David Martınez-Rubio. Cheap orthogonal constraints in neural networks: A simple parametrization of the orthogonal and unitary group. In *ICML*, pages 3794–3803. PMLR, 2019. 3

[20] Honglin Li, Chenglu Zhu, Yunlong Zhang, Yuxuan Sun, Zhongyi Shui, Wenwei Kuang, Sunyi Zheng, and Lin Yang. Task-specific fine-tuning via variational information bottleneck for weakly-supervised pathology whole slide image classification. In *CVPR*, 2023. 3

[21] Yijun Li, Richard Zhang, Jingwan Lu, and Eli Shechtman. Few-shot image generation with elastic weight consolidation. *arXiv*, 2020. 3, 5

[22] Ziqiang Li, Chaoyue Wang, Xue Rui, Chao Xue, Jiaxu Leng, Zhangjie Fu, and Bin Li. Peer is your pillar: A data-unbalanced conditional gans for few-shot image generation. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024. 5

[23] Sangwoo Mo, Minsu Cho, and Jinwoo Shin. Freeze the discriminator: a simple baseline for fine-tuning gans. *arXiv*, 2020. 5

[24] Arnab Kumar Mondal, Piyush Tiwary, Parag Singla, and AP Prathosh. Solad: Sampling over latent adapter for few shot generation. *IEEE Signal Processing Letters*, 2024. 5

[25] Utkarsh Ojha, Yijun Li, Jingwan Lu, Alexei A Efros, Yong Jae Lee, Eli Shechtman, and Richard Zhang. Few-shot image generation via cross-domain correspondence. In *CVPR*, 2021. 2, 3, 5, 8

[26] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE TKDE*, 2009. 3

[27] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. 4

[28] Weize Quan, Ruisong Zhang, Yong Zhang, Zhifeng Li, Jue Wang, and Dong-Ming Yan. Image inpainting with local and global refinement. *IEEE TIP*, 2022. 1

[29] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951. 4

[30] Justin Solomon. Optimal transport on discrete domains. *AMS Short Course on Discrete Differential Geometry*, 2018. 4

[31] Ngoc-Trung Tran, Viet-Hung Tran, Ngoc-Bao Nguyen, Trung-Kien Nguyen, and Ngai-Man Cheung. On data augmentation for gan training. *IEEE TIP*, 2021. 1, 2

[32] Hung-Yu Tseng, Lu Jiang, Ce Liu, Ming-Hsuan Yang, and Weilong Yang. Regularizing generative adversarial networks under limited data. In *CVPR*, 2021. 1

[33] Xiaogang Wang and Xiaoou Tang. Face photo-sketch synthesis and recognition. *IEEE PAMI*, 2008. 5

[34] Yaxing Wang, Abel Gonzalez-Garcia, David Berga, Luis Herranz, Fahad Shahbaz Khan, and Joost van de Weijer. Minegan: effective knowledge transfer from gans to target domains with few images. In *CVPR*, 2020. 5

[35] Zi Wang, Chengcheng Li, and Xiangyang Wang. Convolutional neural network pruning with structural redundancy reduction. In *CVPR*, 2021. 5

[36] Jiayu Xiao, Liang Li, Chaofei Wang, Zheng-Jun Zha, and Qingming Huang. Few shot generative model adaption via relaxed spatial structural alignment. In *CVPR*, 2022. 2, 3, 4, 5

[37] Ceyuan Yang, Yujun Shen, Yinghao Xu, and Bolei Zhou. Data-efficient instance generation from instance discrimination. *NeurIPS*, 2021. 1, 3

[38] Raymond A Yeh, Chen Chen, Teck Yian Lim, Alexander G Schwing, Mark Hasegawa-Johnson, and Minh N Do. Semantic image inpainting with deep generative models. In *CVPR*, 2017. 1

[39] 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*, 2015. 1, 5

[40] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *CVPR*, 2018. 1

[41] Dan Zhang and Anna Khoreva. Progressive augmentation of gans. *NeurIPS*, 2019. 2

[42] Ziqi Zhang, Zeyu Li, Kun Wei, Siduo Pan, and Cheng Deng. A survey on multimodal-guided visual content synthesis. *Neurocomputing*, 2022. 1

[43] Miaoyun Zhao, Yulai Cong, and Lawrence Carin. On leveraging pretrained gans for generation with limited data. In *ICML*, 2020. 1

[44] Yunqing Zhao, Keshigeyan Chandrasegaran, Milad Abdollahzadeh, and Ngai-Man Man Cheung. Few-shot image generation via adaptation-aware kernel modulation. *NeurIPS*, 2022. 3, 5

[45] Yunqing Zhao, Henghui Ding, Houjing Huang, and Ngai-Man Cheung. A closer look at few-shot image generation. In *CVPR*, 2022. 3

[46] Yunqing Zhao, Chao Du, Milad Abdollahzadeh, Tianyu Pang, Min Lin, Shuicheng Yan, and Ngai-Man Cheung. Exploring incompatible knowledge transfer in few-shot image generation. In *CVPR*, 2023. 3, 5

[47] Zhengli Zhao, Zizhao Zhang, Ting Chen, Sameer Singh, and Han Zhang. Image augmentations for gan training. *arXiv*, 2020. 2

[48] Zhengli Zhao, Sameer Singh, Honglak Lee, Zizhao Zhang, Augustus Odena, and Han Zhang. Improved consistency regularization for gans. In *AAAI*, 2021. 2

[49] Jingyuan Zhu, Huimin Ma, Jiansheng Chen, and Jian Yuan. Few-shot image generation with diffusion models. *arXiv preprint arXiv:2211.03264*, 2022. 5

[50] Jingyuan Zhu, Huimin Ma, Jiansheng Chen, and Jian Yuan. Domainstudio: Fine-tuning diffusion models for domain-driven image generation using limited data. *International Journal of Computer Vision*, 133(10):7012–7036, 2025. 5

[51] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017. 1

[52] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 2020. 3