Domain Re-Modulation for Few-Shot Generative Domain Adaptation

Anonymous Author(s)
Affiliation
Address
email

A Supplementary Materials

A.1 Ablation Study

Effect of Target Mapping. In this section, we first investigate the significance of the target mapping module in our proposed DoRM. To evaluate the significance of the target mapping, we conduct an ablation study using the source mapping as a substitute for the target mapping, which remains frozen during the entire 10-shot generative domain adaptation training. As shown in Table 1, our results indicate a significant deterioration in the FID score without the target mapping, implying a considerable drop in the quality and diversity of the generated samples. Furthermore, Figure 1 illustrates that the generative domain adaptation barely occurs during training without the target mapping. This is because the target mapping plays a crucial role in capturing the representative attributes of the target domain and assisting in the acquisition of the domain shift during the adaptation process.

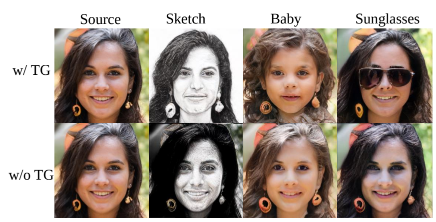

Figure 1: **Qualitative ablation study of the target mapping**. We compare the performance of our DoRM method, shown in the first row, with a variant of the method where the frozen source mapping is used as a substitute for the target mapping, depicted in the second row. We evaluate the performance of both methods on three different target domains: Sketches, FFHQ-Babies, and FFHQ-Sunglasses.

Table 1: **Quantitative ablation study of the target mapping**. The evaluation metric is FID (lower is better). We compare the performance of our DoRM method, shown in the first row, with a variant of the method where the frozen source mapping is used as a substitute for the target mapping, depicted in the second row. The source generator is pre-trained on FFHQ[2], and the target domains include FFHQ-Babies and FFHQ-Sunglasses.

	Babies	Sunglasses
DoRM	**30.31**	**17.31**
DoRM w/o Target Mapping	86.52	74.71

Effect of Re-Modulation Layers. Another crucial component of our proposed DoRM approach is the target affine module. To investigate the roles of the different target affines in DoRM, we

Submitted to 37th Conference on Neural Information Processing Systems (NeurIPS 2023). Do not distribute.

perform experiments where we drop the target affines in both the low-resolution and high-resolution feature maps. Specifically, we conduct 10-shot generative domain adaptation experiments to evaluate the Fréchet Inception Distance (FID) of the generated samples. For an image with a resolution of 256×256, the low-resolution feature maps include resolutions of 4×4, 8×8, 16×16, and 32×32. As presented in Table 2, our results demonstrate that all target affines are crucial for the performance of our DoRM approach and their removal leads to a significant drop in the quality and diversity of the generated samples.

Table 2: **Quantitative ablation study of the target affine layers**. The evaluation metric is FID (lower is better). We compare the FID score of our proposed DoRM method, shown in the first row, with two variants: one where the target affines are removed from the low-resolution feature maps, shown in the second row, and another where the target affines are removed from the high-resolution feature maps, shown in the third row. We use a source generator pre-trained on FFHQ [2] and evaluate all three methods on two different target domains: FFHQ-Babies and FFHQ-Sunglasses.

	babies	sunglasses
DoRM	**30.31**	**17.31**
DoRM w/o target affines in low resolution	93.28	92.42
DoRM w/o target affines in high resolution	37.16	20.81

Effect of Re-Modulation Weight. The re-modulation weight is a crucial parameter that controls the strength of the acquired domain shift in our proposed DoRM approach. A small re-modulation weight leads to a lower strength of the domain shift, resulting in more attributes of the source domain being preserved during generative domain adaptation. To investigate the impact of the re-modulation weight, we conduct 10-shot generative domain adaptation experiments using different re-modulation weights. The results are presented in Table 3. Our results demonstrate that different domain gaps have different optimal re-modulation weights, indicating that the selection of the re-modulation weight should be tailored to the specific target domain.

Table 3: **Quantitative ablation study of the re-modulation weight.** The evaluation metric is FID (lower is better). We conduct 10-shot generative domain adaptation experiments using our DoRM approach with varying re-modulation weights. We use a source generator pre-trained on FFHQ [2] and evaluate our method on two different target domains: FFHQ-Baby and FFHQ-Sunglasses.

Re-modulation weight α	0.5	0.2	0.05	0.005	0.001
FFHQ-Baby	37.9	36.1	34.0	**30.3**	32.3
FFHQ-Sunglasses	18.7	**17.3**	17.9	18.5	19.4

Effect of Target domain classifier. We investigate the effect of the target domain classifier on the performance of our proposed DoRM approach. Specifically, we experiment with different depths (*e.g.*, number of MLP layers) and initialization methods for the target domain classifier in 10-shot generative domain adaptation. As presented in Table 4, our results indicate that the two-layer MLP target domain classifier achieves the best performance. This is because the one-layer MLP lacks the ability to classify the target domain effectively, while the three-layer MLP is prone to overfitting due to the limited number of training images.

Table 4: **Quantitative ablation study of the target domain classifier**. The evaluation metric is FID (lower is better). We experiment with different depths of the target domain classifier, using various initialization methods. We use a source generator pre-trained on FFHQ [2] and evaluate our proposed DoRM approach on 10-shot generative domain adaptation tasks.

	FFHQ-Baby		FFHQ-Sunglasses	
MLP Depth	source initial	random initial	source initial	random initial
one layer	38.03	37.15	24.63	22.15
two layers	33.32	**30.31**	20.40	**17.31**
three layers	—	33.25	—	20.42

A.2 More Synthesis Results in 10-shot GDA

Qualitative results on 10-shot generative domain adaptation. We present qualitative results of our proposed DoRM approach on more target datasets, including face caricatures, face paintings by Raphael, face paintings by Amedeo Modigliani, and face paintings by Otto Dix [4]. All training images are shown in Figure 2, and the qualitative results are presented in Figure 3. Our results demonstrate that our proposed DoRM approach achieves appealing synthesis quality in various target domains.

Results of latent interpolation. We perform latent space interpolation to demonstrate that our DoRM is not harmful to the learned latent space. In Figure 4, the first and last columns show the generated images with two latent codes after 10-shot generative domain adaptation, while the remaining columns show the results obtained by linearly interpolating the two latent codes. Our results demonstrate that all intermediate synthesized images have high target-domain consistency and high cross-domain consistency. Moreover, the semantics of the generated images, such as gender, haircut, and pose, vary gradually throughout the interpolation, indicating that our proposed DoRM approach preserves the underlying semantic structure of the learned latent space.

A.3 Hybrid-domain Generation

In Figure 6, we present the results of generating hybrid domains using our proposed model. Our DoRM has a unique generator structure that is similar to the mechanism of the human brain. This structure endows the DoRM with two novel capabilities: memory and domain association. These capabilities enable the DoRM to not only retain knowledge from previously learned domains when generating images in new domains, but also integrate multiple learned domains and synthesize images in hybrid domains that were not encountered during training.

A.4 Experiments on one-shot Generative Domain Adaptation

Although our DoRM is mainly for few-shot generative domain adaptation, DoRM is also can be employed for one-shot generative domain adaptation. In the one-shot GDA, the training dataset is a single image, which is difficult for the backbone of discriminator to extract the main characters of the target domain because of the overfitting issue. In this case, we introduce a clip-based local-level adaptation loss L_{local} from [6] to help to acquire the local-level characters and styles of the target domain. Concretely, we extract the intermediate tokens of the adapted image I_B synthesized by DoRM and the single target image I_{tar} from the $k-th$ layer of CLIP image encoder. And align each of adapted token F_B with its closest target token from F_{tar}, where $F_B = F_B^1, ..., F_B^n$ and $F_{tar} = F_{tar}^1, ...F_{tar}^m$ are the extracted tokens. The clip-based local-level adaptation loss is defined as:

$$L_{local} = \max(\frac{1}{n}\sum_i \min_j C_{i,j}, \frac{1}{m}\sum_j \min_i C_{i,j}) \tag{1}$$

where C is calculated as:

$$C_{i,j} = 1 - \frac{F_B^i \cdot F_{tar}^j}{|F_B^i||F_{tar}^j|} \tag{2}$$

Furthermore, to better identify and maintain the domain-sharing attributes in one-shot generative domain adaptation, we also employ the inversion-based selective cross-modal consistency loss L_{scc} from [6]. Specifically, this loss function aims to identify and preserve domain-sharing attributes in the $W+$ space. The underlying assumption is that attributes that are similar in $W+$ space between the source and target domains during adaptation are more likely to be domain-sharing attributes. To achieve this, L_{scc} dynamically analyzes and retains these attributes. First, it inverts the source and corresponding target images into $W+$ latent codes, w_A and w_B, respectively, using a pre-trained inversion model such as pSp pr e4e, for each iteration. Next, it computes the difference Δw, between the centers of a source queue of $W+$ latent codes, X_A and the target queue of $W+$ latent codes, X_B, where X_A and X_B are dynamically updated with w_A and w_B during training. The loss function then encourages w_A and w_B to be consistent in channels with less difference, thereby facilitating the preservation of domain-sharing attributes. The inversion-based selective cross-modal consistency loss L_{scc} is defined as follows:

$$L_{scc} = ||mask(\Delta w, \alpha) \cdot (w_B - w_A)||_1 \tag{3}$$

where α represents the proportion of preserved attributes, and $mask(\Delta w, \alpha)$ determines which channels to retain. Specifically, let $|\Delta w_{s_{\alpha N}}|$ be the $\alpha N - th$ largest element of Δw. Then, each dimension of $mask(\Delta w, \alpha)$ is calculated as follows:

$$mask(\Delta w, \alpha)_i = \begin{cases} 1 & |\Delta w_i| \leq |\Delta w_{s_{\alpha N}}| \\ 0 & |\Delta w_i| \geq |\Delta w_{s_{\alpha N}}| \end{cases} \tag{4}$$

We compare our DoRM++ approach which denotes introducing the two new loss terms into training with state-of-the-art one-shot generative domain adaptation (GDA) methods, including JoJoGAN [1], Generalized One-shot Domain Adaptation [7], DynaGAN[3] and DiFa [6]. Figure 5 shows the comparison results. Our results indicate that JoJoGAN, DynaGAN and Generalized One-shot Domain Adaptation fail to achieve GDA when the target image is FFHQ-Baby and FFHQ-Sunglasses, and the synthesis quality of DynaGAN is limited. Similarly, DiFa also fails to achieve GDA when the target image is FFHQ-Sunglasses, and the synthesis diversity is unsatisfactory when the target image is FFHQ-Baby.

In contrast, our DoRM++ approach achieves one-shot GDA among all the reference images, resulting in high-quality and diverse synthesis, while maintaining appealing cross-domain consistency. Moreover, our DoRM++ generator has memory to realize multiple target domains' generation, which saves a significant amount of storage space. Our DoRM++ generator also has the ability to integrate the learned knowledge of multiple target domains to synthesize images in hybrid domains that are unseen in the target domains. As shown in Figure 6, the DoRM++ generator can synthesize high-quality and diverse images in hybrid domains while maintaining the domain-sharing attributes (e.g, pose, identity).

A.5 Experiments on Other Source Domains

In addition to the experiments on FFHQ, we conduct other 10-shot generative domain adaptation experiments to qualitatively evaluate the effectiveness of our proposed DoRM approach. Specifically, we pre-trained a StyleGAN2 on the LSUN-church [5] dataset and adapted the pre-trained GAN to generate haunted house images. The results of our experiments are presented in Figure 7.

Figure 2: **Training images in 10-shot generative domain adaptation experiments**.

Figure 3: **10-shot generative domain Adaptation on FFHQ**. We use a source generator pre-trained on the FFHQ [2] dataset and evaluate our proposed DoRM approach on various target domains, including FFHQ-Sunglasses, face caricatures, face paintings by Raphael, face paintings by Amedeo Modigliani, and face paintings by Otto Dix [4]. The training images are shown in Figure 2. Our results demonstrate that our proposed DoRM approach can maintain cross-domain consistency between the source domain and different target domains.

Figure 4: **Latent interpolation** using the generators adapted to different target domains in 10-shot generative domain adaptation (the first line is the source images). The first and last columns are the generated images with two latent codes after 10-shot generative domain adaptation. The remaining columns are the results by linearly interpolating the two latent codes. According to the figure, all the semantics (*e.g.*, the gender, the haircut and the pose) vary gradually.

Figure 5: **Qualitative comparison on one-shot GDA**. The source domain is FFHQ, and the target domains include different reference images, as shown in the first row of the figure. We compare our method with JoJoGAN [1], Generalized One-shot Domain Adaption [7], DynaGAN[3] and DiFa[6]. Our method not only achieves better synthesis quality and diversity but also maintains higher cross-domain consistency than other methods.

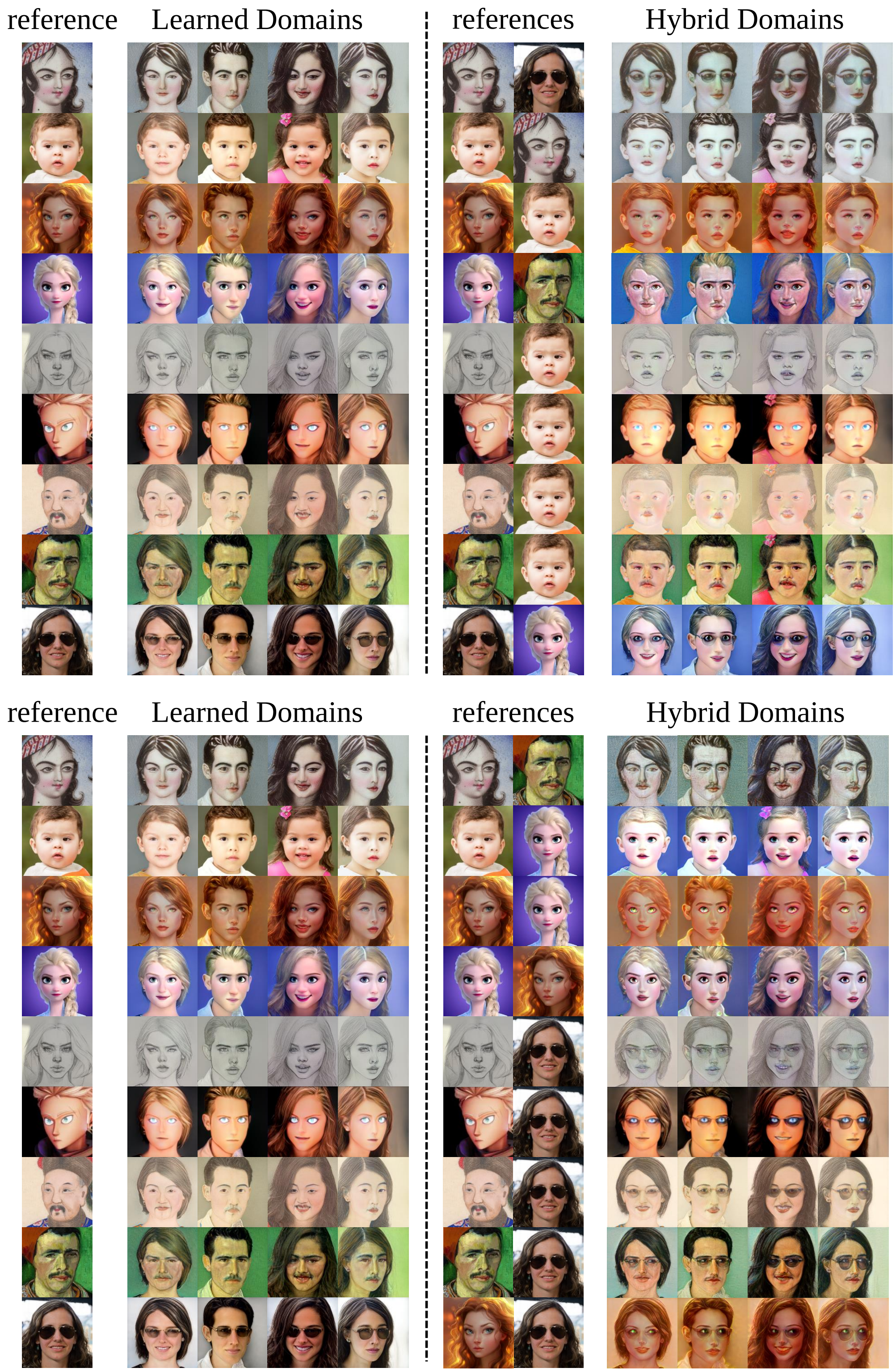

Figure 6: **One-shot generative domain adaptation and domain association on FFHQ**. The source domain is FFHQ, and the target domains include different reference images, as shown in the first column of the figure. Once our DoRM++ generator learns to synthesize images in multiple target domains, it can integrate the knowledge from the learned multiple domains and synthesize images in hybrid domains which are unseen in the target domains.

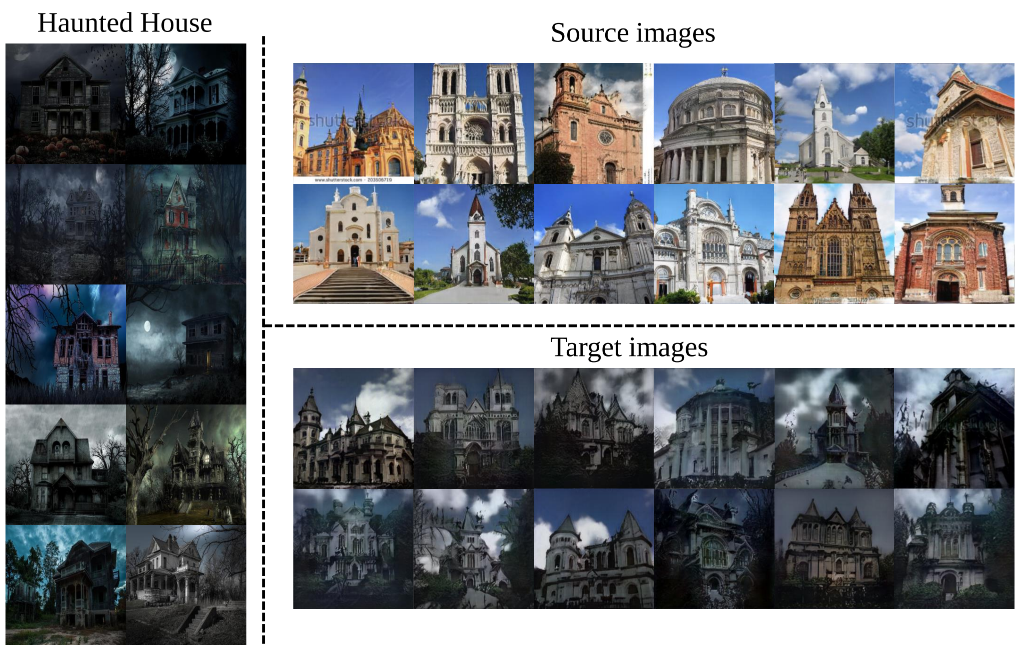

Figure 7: **10-shot generative domain adaptation on LSUN-Church[5]**. The source generator is pretrained on LSUN-Church[5]. The target domain is haunted house (10 training images are shown on the left side). The result shows that our method can maintain the cross-domain consistency between the source domain and the target domain.

References

[1] Min Jin Chong and David Forsyth. Jojogan: One shot face stylization. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVI*, pages 128–152. Springer, 2022.

[2] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.

[3] Seongtae Kim, Kyoungkook Kang, Geonung Kim, Seung-Hwan Baek, and Sunghyun Cho. Dynagan: Dynamic few-shot adaptation of gans to multiple domains. In *SIGGRAPH Asia 2022 Conference Papers*, pages 1–8, 2022.

[4] Jordan Yaniv, Yael Newman, and Ariel Shamir. The face of art: landmark detection and geometric style in portraits. *ACM Transactions on graphics (TOG)*, 38(4):1–15, 2019.

[5] 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.

[6] Yabo Zhang, mingshuai Yao, Yuxiang Wei, Zhilong Ji, Jinfeng Bai, and Wangmeng Zuo. Towards diverse and faithful one-shot adaption of generative adversarial networks. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022.

[7] Zicheng Zhang, Yinglu Liu, Congying Han, Tiande Guo, Ting Yao, and Tao Mei. Generalized one-shot domain adaption of generative adversarial networks. *arXiv preprint arXiv:2209.03665*, 2022.