Development of domain adaptation methods for generative models

Testing protocol and metrics

Kirill Korolev

1 Testing protocol

For comparison with other methods it is essential to fix the testing protocol. For evaluation 100 face images from FFHQ dataset were sampled and 15 domains were selected (10 style images from AAHQ dataset, 'Ariel', 'Dumbledore', 'Trump' domains from TargetCLIP and 'Sketch', 'Anastasia' domains from StyleDomain). Inspired by [2] and [1], the semantic score is used to estimate how close the adapted images to the reference ones. For a given style or domain define

$$\text{Semantic-Score} = \frac{1}{|\mathcal{S}|} \sum_{I_{src} \in \mathcal{S}} \langle E_{CLIP}(I_{domain}), E_{CLIP}(I_{src,domain}) \rangle \tag{1}$$

where $E_{CLIP}(\cdot)$ is a CLIP visual-embedding and $I_{src,domain}$ is an adapted image.

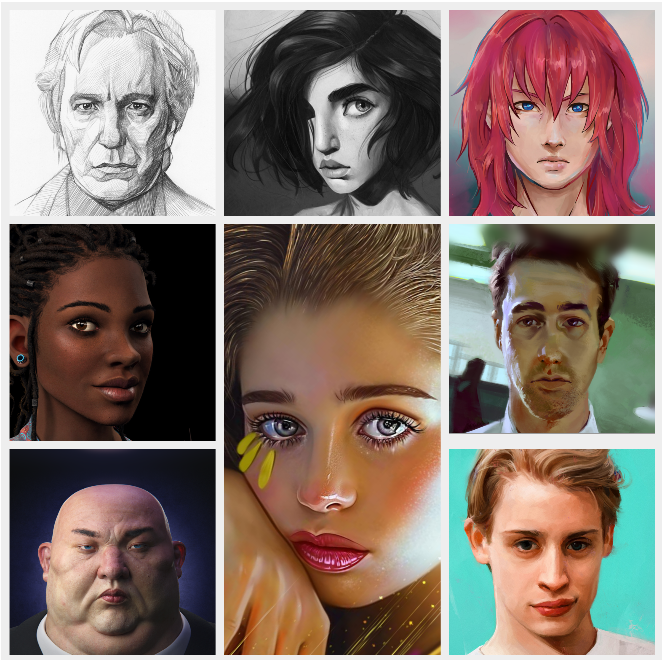

Figure 1.1: Examples of style images / domains for evaluation.

To evaluate the diversity of the generated images for a specific domain the following metric is calculated from [1]

$$Diversity = \frac{1}{\binom{|\mathcal{S}|}{2}} \sum_{src_1 < src_2} (1 - \langle E_{CLIP}(I_{src_1,domain}), E_{CLIP}(I_{src_2,domain}) \rangle) \tag{2}$$

Eventually, both Semantic-Score and Diversity metrics are averaged across 10 styles.

Method	Semantic-Score	Diversity
BlendGAN	0.663 ± 0.0685	0.187 ± 0.0222
TargetCLIP	0.702 ± 0.036	0.248 ± 0.008
StyleDomain	0.617 ± 0.04	0.267 ± 0.035

Table 1.1: Evaluation results for several methods.

2 BlendGAN

For reproducing BlendGAN results framework's training code is adapted from an unofficial PyTorch implementation of StyleGAN, maintaining most hyper-parameters without alteration. The Adam optimizer is employed with $\beta_1 = 0, \beta_2 = 0.99$, and a learning rate of 0.002. The style encoder E_{style} utilizes the AAHQ dataset downscaled to 256×256 resolution. The two MLP modules consist of FC layers with equalized learning rates and leaky ReLU activations ($\alpha = 0.2$). Affine transformations are applied during augmentation to preserve image style, while color transformations are omitted. For the generator and three discriminators, FFHQ and AAHQ datasets with 1024×1024 resolution are utilized, and the minibatch size is set to 256. The weights of the generator G and face discriminator D_{face} are loaded from the official pretrained StyleGAN2 model to expedite training, whereas those of the style discriminator D_{style} and style latent discriminator D_{style_latent} are initialized randomly. Joint optimization of the generator and discriminators involves non-saturating logistic loss and path length regularization, with training conducted using a minibatch size of 8.

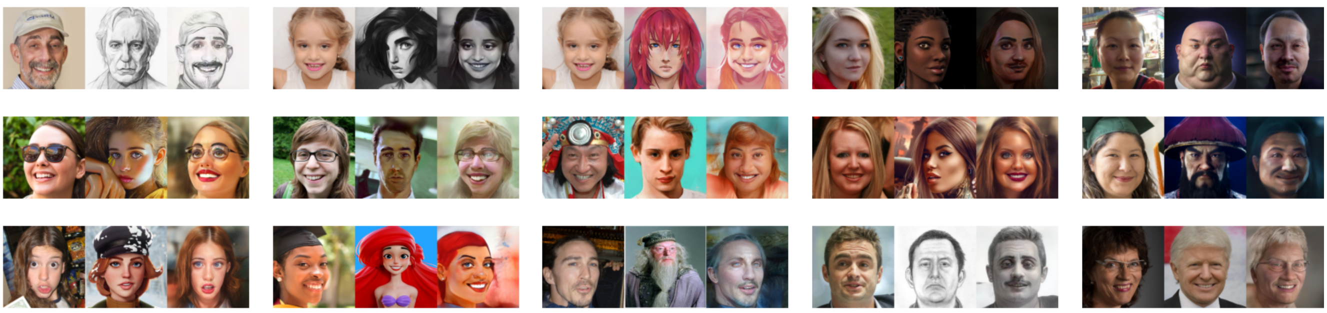

Figure 2.1: Adaptation using a BlendGAN of 15 different domains. The first image in a triplet is a source image, the second one presents style/domain images, the third one shows adapted images.

3 TargetCLIP

The initial approach involves employing the Adam optimizer for 1000 iterations with a learning rate of 0.2. Double additivity losses are calculated using only N=4 images. Fine-tuning of the encoder occurs on a small dataset consisting of 200 randomly selected images from the

CelebA-HQ dataset. The evaluation is conducted on 50 random images from the CelebAHQ test set, using a learning rate of 1e-4 for 3000 iterations. The batch size is set to 1 for the target image, and N=5 source images are utilized for double additivity losses. The objective and hyperparameters remain consistent with those employed in the optimization-based method.

Figure 3.1: Adaptation using a TargetCLIP of 'Ariel', 'Dumbledore', 'Trump' domains. The first column consists of source images, the second one presents style/domain images, the third one shows adapted images.

4 StyleDomain

For the one-shot image-based domain adaptation set loss coefficients as $\lambda_{cw} = 0.5, \lambda_{rc} = 30, \lambda_{rr} = 10$. Use batch of 4 and 300 number of iterations and utilize the Adam Optimizer with parameters from the following table.

Parameter Space	lr	betas	weight_decay
SyntConv	0.002	(0.0, 0.999)	0
Full	0.002	(0.0, 0.999)	0
Affine	0.01	(0.0, 0.999)	0
Mapping	0.3	(0.0, 0.999)	0
StyleSpace	0.05	(0.9, 0.999)	0

Figure 4.1: Optimizer hyperparameters from the paper for each setting.

Figure 4.2: Adaptation using a StyleDomain of ‘Anastasia‘ and ‘Sketch‘ domains. The first column consists of source images, the second one presents style/domain images, the third one shows adapted images.

References

[1] Aibek Alanov, Vadim Titov, and Dmitry P Vetrov. “Hyperdomainnet: Universal domain adaptation for generative adversarial networks”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 29414–29426.

[2] Hila Chefer, Sagie Benaim, Roni Paiss, and Lior Wolf. “Image-Based CLIP-Guided Essence Transfer”. In: *arXiv preprint arXiv: 2110.12427* (2021).