Supplementary of "ACL-GAN"

July 16, 2020

1 Implementation Details

1.1 Data Augmentation

We flipped the images horizontally with a probability of 0.5. Due to the small number of images in selfie2anime [3] dataset, we also applied colour jittering with up to $hue = 0.15$, random grayscale with a probability of 0.25, random rotation with up to $35°$, random translation of up to 0.1 of the image, and random perspective with distortion scale of 0.35 with a probability of 0.5. We trained on the original images, without data augmentation, on the last 100K iterations [7].

Model	glasses removal		Model	male to female		Model	selfie to anime	
	FID	KID		FID	KID		FID	KID
CycleGAN	48.71	0.043 ± 0.0011	CycleGAN	21.30	0.021 ± 0.0003	CycleGAN	102.92	0.042 ± 0.0019
CycleGAN-5	44.51	0.040 ± 0.0008	CycleGAN-5	22.10	0.021 ± 0.0004	CycleGAN-5	100.41	0.041 ± 0.0024
CycleGAN-1	42.08	0.038 ± 0.0007	CycleGAN-1	33.57	0.037 ± 0.0005	CycleGAN-1	99.39	0.035 ± 0.032
DiscoGAN	58.14	0.054 ± 0.0010	DiscoGAN	58.77	0.065 ± 0.0005	DiscoGAN	155.20	0.120 ± 0.0063
MUNIT	28.58	0.026 ± 0.0009	MUNIT	19.02	0.019 ± 0.0004	MUNIT	101.30	0.043 ± 0.0041
DRIT++	33.06	0.026 ± 0.0006	DRIT++	24.61	0.023 ± 0.0002	DRIT++	104.40	0.050 ± 0.0028
Fixed-Point GAN	44.22	0.038 ± 0.0009	StarGAN	36.17	0.034 ± 0.0005	U-GAT-IT	99.15	0.039 ± 0.0030
CouncilGAN	27.77	0.025 ± 0.0011	CouncilGAN	18.10	0.017 ± 0.0004	CouncilGAN	98.87	0.042 ± 0.0047
ACL-GAN	**23.72**	**0.020** $\pm$ **0.0010**	ACL-GAN	**16.63**	**0.015** $\pm$ **0.0003**	ACL-GAN	**93.58**	**0.037** $\pm$ **0.0036**
ACL-GAN-0.2	**23.72**	**0.020** $\pm$ **0.0010**	ACL-GAN-0.2	**16.63**	**0.015** $\pm$ **0.0003**	ACL-GAN-0.2	**95.43**	**0.038** $\pm$ **0.0020**

Table 1: **Quantitative results of glasses removal, male-to-female translation, and selfie-to-anime translation.** For KID, mean and standard deviation are listed. A lower score means better performance. U-GAT-IT [3] is in light mode. Our method outperforms all other baselines in all applications.

1.2 Hyperparameters

In our experiment, λ_{idt} is fixed to be 1 for all tasks. We test λ_{ACL} with 0.2 on selfie2anime dataset, represented as ACL-GAN-0.2 and we find that λ_{ACL} of 0.2 works for all tasks. Besides, the hyperparameters related to bounded focus mask, have certain meanings and can be easily set according to different tasks. We run CycleGAN with smaller λ_{cycle} 5 and 1, represented as CycleGAN-5 and CycleGAN-1 respectively. The results are shown in Table 1, which are worse than ours.

2 Additional Experimental Results

2.1 Ablation Studies

In addition to the male-to-female translation, this section evaluates different ablations for both glasses removal and selfie-to-anime translation.

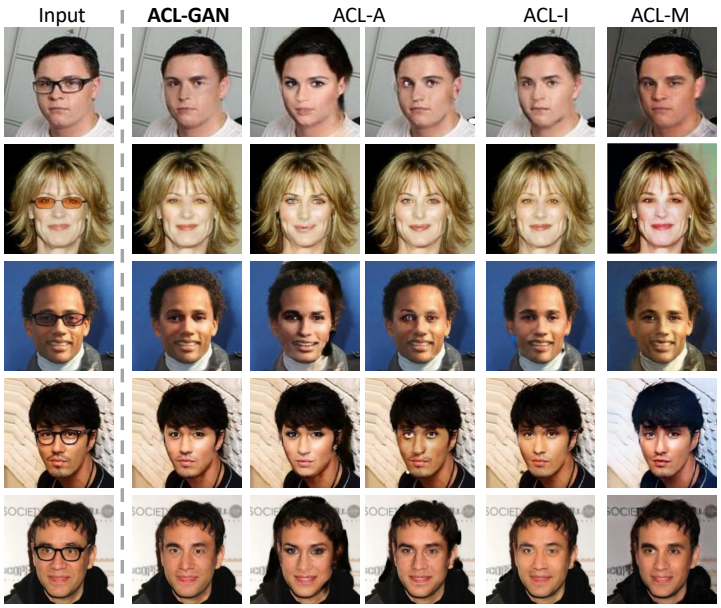

Figure 1: **Qualitative results for ablation studies on glasses removal.** From left to right: input, ACL-GAN (with total loss), ACL-A (without $\mathcal{L}_{acl}$), ACL-I (without $\mathcal{L}_{idt}$), ACL-M (without $\mathcal{L}_{mask}$).

Model	$\mathcal{L}_{acl}$	$\mathcal{L}_{idt}$	$\mathcal{L}_{mask}$	**FID**	**KID**
ACL-A	-	✓	✓	26.79	0.025 ± 0.0011
ACL-I	✓	-	✓	26.66	0.025 ± 0.0011
ACL-M	✓	✓	-	24.95	0.021 ± 0.0009
ACL-GAN	✓	✓	✓	**23.72**	$\mathbf{0.020 \pm 0.0010}$

Table 2: Quantitative results for ablation studies on glasses removal.

Specifically, for glasses removal, we conduct the same four settings as the male-to-female translation. The qualitative results are shown in Fig. 1. Due to relatively deterministic translation results, we only show one translated image of ACL-GAN, ACL-I, and ACL-M for each input image. However, without adversarial-consistency loss $\mathcal{L}_{acl}$, the results of ACL-A are inconsistent with the input images, *e.g.* the results are more feminine because of the imbalance of the dataset and the eye shapes are different with those in the input images.

Model	$\mathcal{L}_{acl}$	$\mathcal{L}_{idt}$	$\mathcal{L}_{mask}$	**FID**	**KID**
ACL-A	-	✓	-	101.38	0.044 ± 0.0026
ACL-I	✓	-	-	95.81	0.038 ± 0.0039
ACL-GAN	✓	✓	-	**93.58**	$\mathbf{0.037 \pm 0.0036}$

Table 3: Quantitative results for ablation studies on selfie-to-anime translation.

For selfie-to-anime translation, the style of selfies should be changed. Therefore, we did not use bounded focus mask and we compared three ablation settings, ACL-GAN (with total loss), ACL-A (without $\mathcal{L}_{acl}$), and ACL-I (without $\mathcal{L}_{idt}$). Two results are shown in Fig. 2 for each model and each input. The generated images of ACL-GAN successfully preserve the important features of the input, compared with ACL-A.

The quantitative results are shown in Table 2 and Table 3. The results are consistent with those of male-to-female translation and they show the effectiveness of adversarial-consistency loss, identity loss and bounded focus mask.

2.2 Additional Qualitative Results

To further demonstrate the effectiveness of ACL-GAN, we show the translated images along with the generated bounded focus masks in Fig. 3, 4 and 5. For glasses removal and male-to-female translation, we show one results and its bounded focus mask of ACL-GAN for each input. For selfie-to-anime, the bounded focus mask is not used and two results of ACL-GAN are exhibited for each input. We further test DiscoGAN [4] with the same setting of our paper. The results are shown in Table 1 which are worse than ours and show that smaller bottleneck is not sufficient to overcome the drawbacks of cycle loss.

Figure 2: **Qualitative results for ablation studies on selfie-to-anime translation.** From left to right: input, ACL-GAN (with total loss), ACL-A (without $\mathcal{L}_{acl}$), ACL-I (without $\mathcal{L}_{idt}$). Bounded focus mask and $\mathcal{L}_{mask}$ are not used for all models on selfie-to-anime translation.

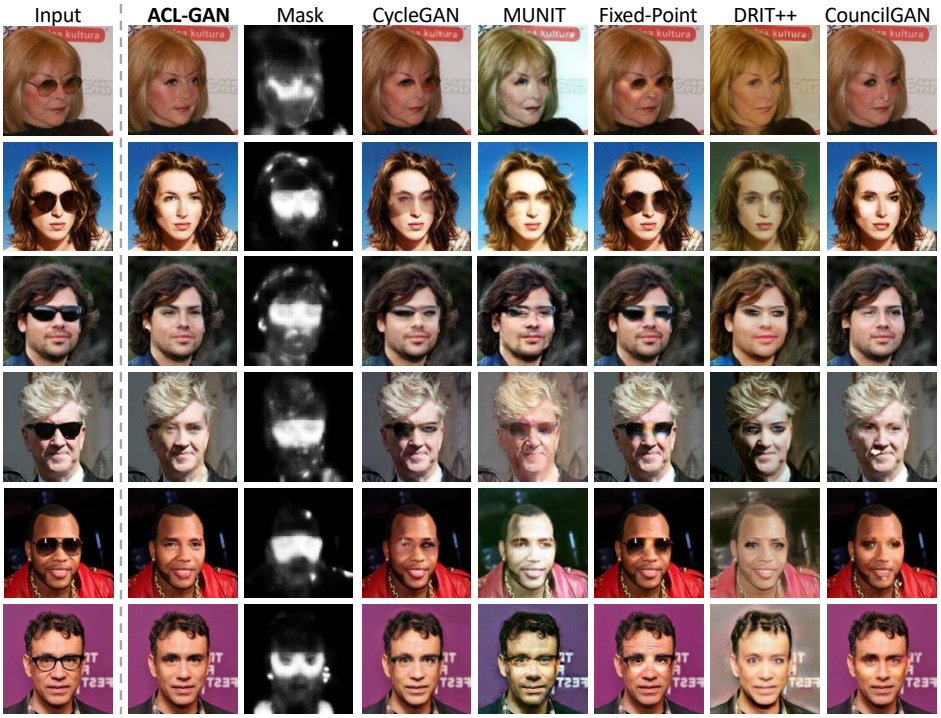

Figure 3: **Additional qualitative results on glasses removal.** From left to right: input, our ACL-GAN, mask of ACL-GAN, CycleGAN [9], MUNIT [2], Fixed-Point GAN [8], DRIT++ [6, 5], and CouncilGAN [7].

Acknowledgements

The authors would like to thanks Jie Fu, Shuang Hu and Haoqi Yuan for helpful discussions. This work was supported by the start-up research funds from Peking University (7100602564) and the Center on Frontiers of Computing Studies (7100602567). We would also like to thank Imperial Institute of Advanced Technology for GPU supports.

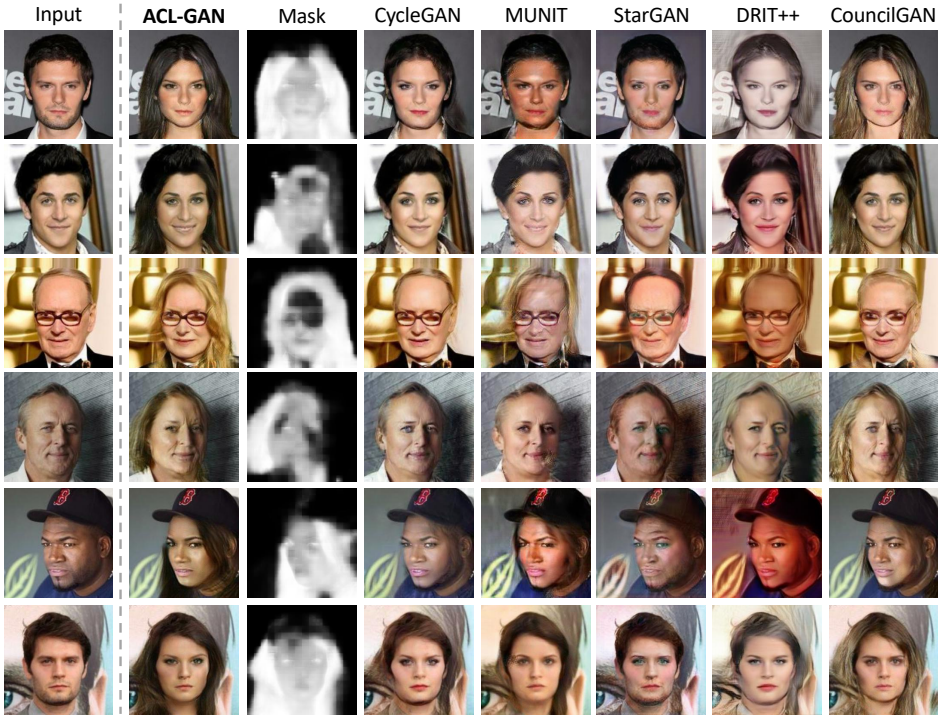

Figure 4: **Comparison against baselines on male-to-female translation.** From left to right: input, our ACL-GAN, mask of ACL-GAN, CycleGAN [9], MUNIT [2], StarGAN [1], DRIT++ [6, 5], and CouncilGAN [7].

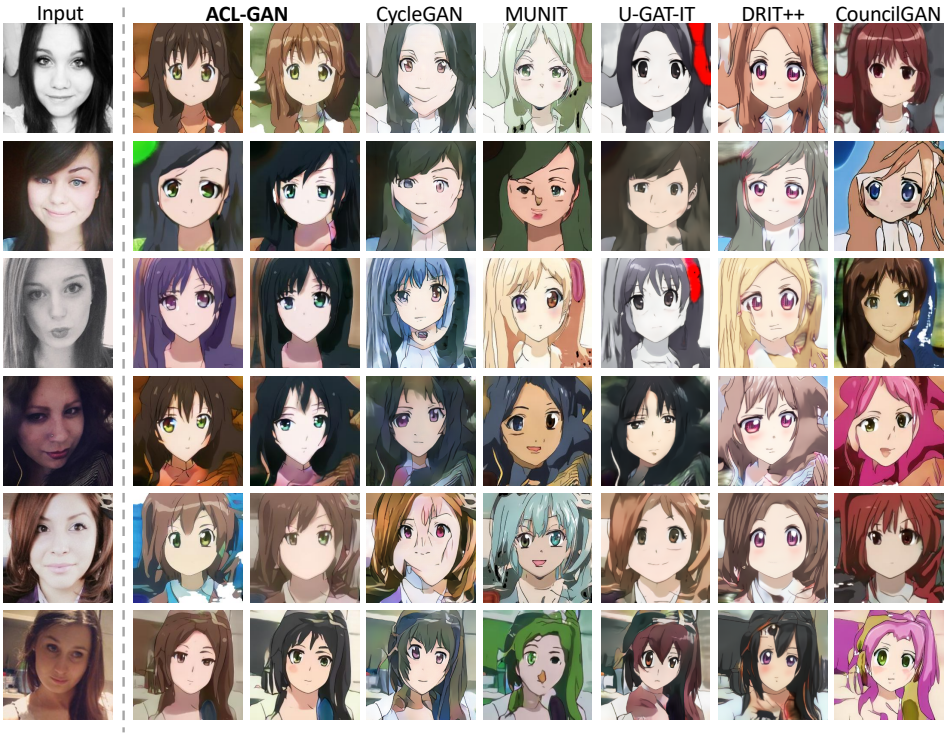

Figure 5: **Comparison against baselines on selfie-to-anime translation.** From left to right: input, our ACL-GAN, CycleGAN [9], MUNIT [2], U-GAT-IT [3], DRIT++ [6, 5], and CouncilGAN [7].

References

[1] Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)

[2] Huang, X., Liu, M.Y., Belongie, S., Kautz, J.: Multimodal unsupervised image-to-image translation. In: European Conference on Computer Vision (2018)

[3] Kim, J., Kim, M., Kang, H., Lee, K.H.: U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. In: International Conference on Learning Representations (2020)

[4] Kim, T., Cha, M., Kim, H., Lee, J.K., Kim, J.: Learning to discover cross-domain relations with generative adversarial networks. In: International Conference on Machine Learning (2017)

[5] Lee, H.Y., Tseng, H.Y., Mao, Q., Huang, J.B., Lu, Y.D., Singh, M., Yang, M.H.: Drit++: Diverse image-to-image translation via disentangled representations. International Journal of Computer Vision (2020)

[6] Lee, H., Tseng, H., Huang, J., Singh, M., Yang, M.: Diverse image-to-image translation via disentangled representations. In: European Conference on Computer Vision (2018)

[7] Nizan, O., Tal, A.: Breaking the cycle – colleagues are all you need. In: arXiv preprint arXiv 1911.10538 (2019)

[8] Siddiquee, M.M.R., Zhou, Z., Tajbakhsh, N., Feng, R., Gotway, M.B., Bengio, Y., Liang, J.: Learning fixed points in generative adversarial networks: From image-to-image translation to disease detection and localization. In: IEEE International Conference on Computer Vision (2019)

[9] Zhu, J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE International Conference on Computer Vision (2017)