ICANN19

28th International Conference on Artificial Neural Networks

Self-attention StarGAN for Multi-domain Image-to-image Translation

Ziliang He[1], Zhenguo Yang[1]✉, Xudong Mao[2], Jianming Lv[3], Qing Li[2], Wenyin Liu1[1] ✉

School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China[1]

Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China[2]

School of Computer Science and Engineering, South China University of Technology, Guang−zhou, China[3]

Contact Author: ziliangbufer@gmail.com, zhengyang5-c@my.cityu.edu.hk, liuwy@gdut.edu.cn

Wechat ID: a13719037726

Introduction

We propose a Self-attention StarGAN by introducing the self-attention mechanism into StarGAN [1] to deal with multi-domain image-to-image translation, aiming to generate images with high-quality details and obtain consistent backgrounds. The self-attention mechanism [2] models the long-range dependencies among the feature maps at all positions, which is not limited to the local image regions. Simultaneously, we take the advantage of batch normalization to reduce reconstruction error and generate fine-grained texture details. We adopt spectral normalization in the network to stabilize the training of Self-attention StarGAN.

Methods

The overall framework of the proposed Self-attention StarGAN is shown in Fig. 1. The Self-attention StarGAN contains a Self-attention Generator as and a Self-attention Discriminator as shown in Fig. 2, which are shared for the source and target domains. We introduce the self-attention module to both generator and discriminator in order to model the global dependencies of images, the structure of the module is shown in Fig. 3. Real images and their target domain labels are fed into the Self-attention Generator to generate fake images. On the other hand, both fake images and real images are fed into Self-attention Discriminator, which tries to recognize whether they are real or not, and outputs their domain labels. Finally, the original domain labels and the fake images are fed into the shared Self-attention Generator again to obtain reconstructed images, based on which the reconstruction loss can be calculated by comparing the reconstructed images with the original images.

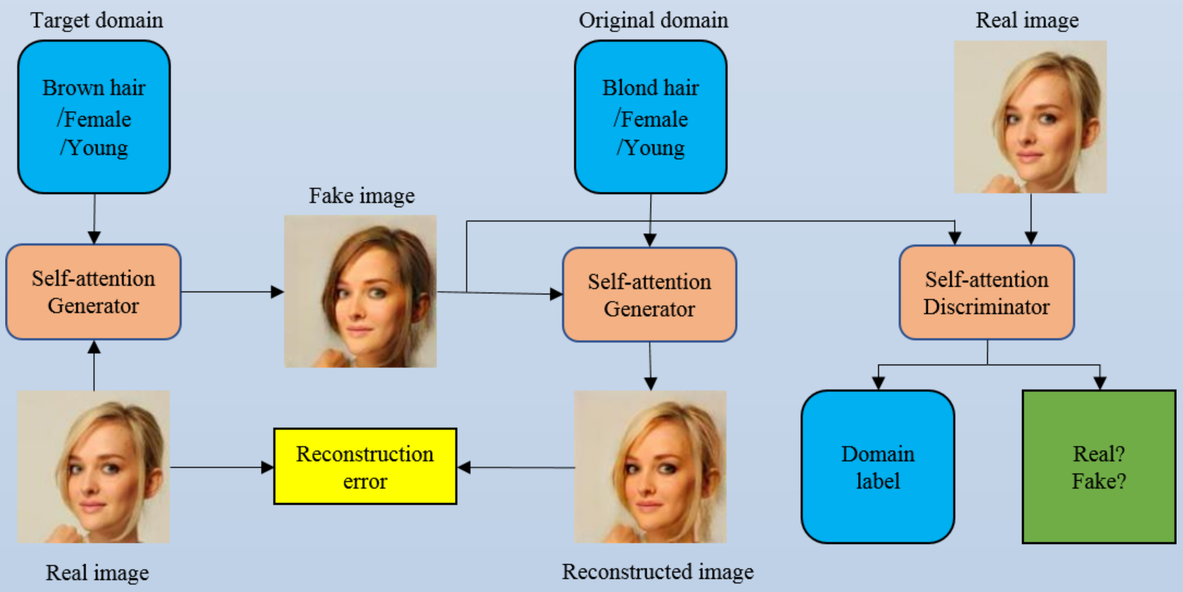

Fig. 1. Overview of the framework

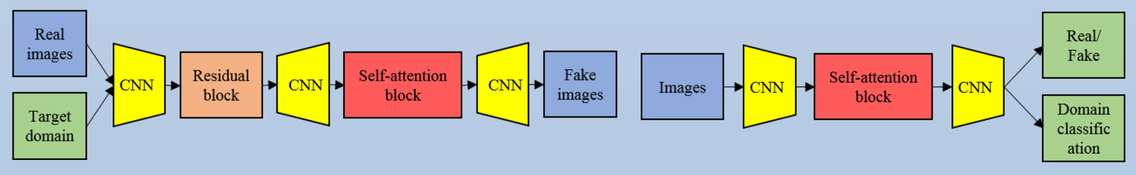

Fig. 2. Self-attention Generator and Self-attention Discriminator

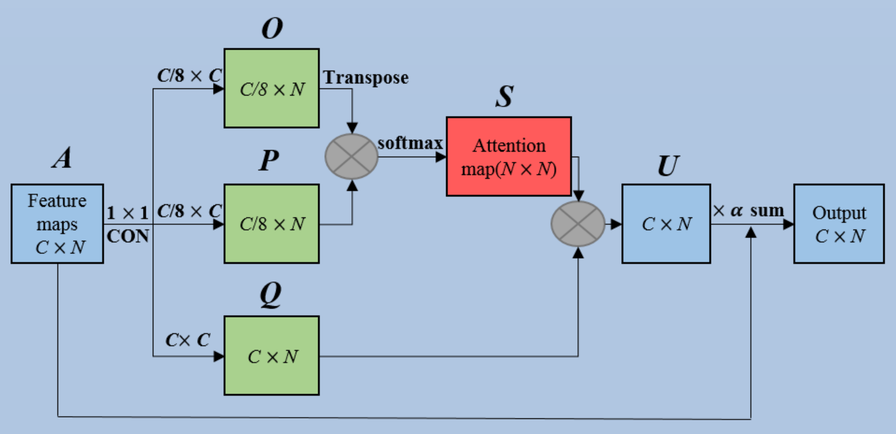

Fig. 3. Self-attention block, where C represents the number of channels and N represents the size of the feature map

Experiments

We compare our method against StarGAN on the task of facial attribute translation. We show some transfer results of facial attribute on CelebA [3] in Fig. 4. Furthermore, We have conducted a survey on Amazon Mechanical Turk (AMT) to evaluate the quality of face attribute transfer as shown in table 1. We can observe that the backgrounds of the images generated by our method have stronger robust-ness and are more consistent with the original images than the baseline model and 68.1% of 1,000 Tuckers agree that the images generated by our model is more consistent with the original images.

Fig. 4. Transfer results of facial attribute on the CelebA dataset

Table 1. AMT perceptual evaluation for different models

Method	Black hair	Blond hair	Brown hair	Gender	Age	Background consistency
StarGAN	46.1%	**55.7%**	48.7%	**54.0%**	47.0%	32.9%
Our method	**53.9%**	44.3%	**51.3%**	46.0%	**53.0%**	**68.1%**

In Fig. 7, we can observe that there are uneven color blocks in the images generated by StarGAN and the images generated by our method with fine-grained details, which proves that our proposed Self-attention StarGAN benefits from both the self-attention mechanism and BN strategy, improving the quality of generated images.

Fig. 5. Visualizations by magnifying the generated images partially

References:

1. Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: Stargan: Unified generative ad-versarial networks for multi-domain image-to-image translation. In: Computer Vision and Pattern Recognition (CVPR) (2018)
2. Zhang, H., Goodfellow, I., Metaxas, D., Odena, A.: Self-Attention Generative Adversarial Networks. In: Machine Learning (2018)
3. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceed-ings of the IEEE International Conference on Computer Vision (ICCV) (2015).

ManuscriptID: 062