UNPAIRED IMAGE-TO-IMAGE TRANSLATION FROM SHARED DEEP SPACE

Xuehui Wu, Jie Shao*, Lianli Gao, Heng Tao Shen

Center for Future Media, School of Computer Science and Engineering
University of Electronic Science and Technology of China

Motivation

- Pixel-level representation cannot sufficiently express the semantic information of images, so pixel-to-pixel translation basically makes generators just recolor but not do enough texture translation.
- The feature maps of different layers in pre-trained VGG19 on Imagenet can provide a hierarchical representation, which expresses an image from low-level to high-level.

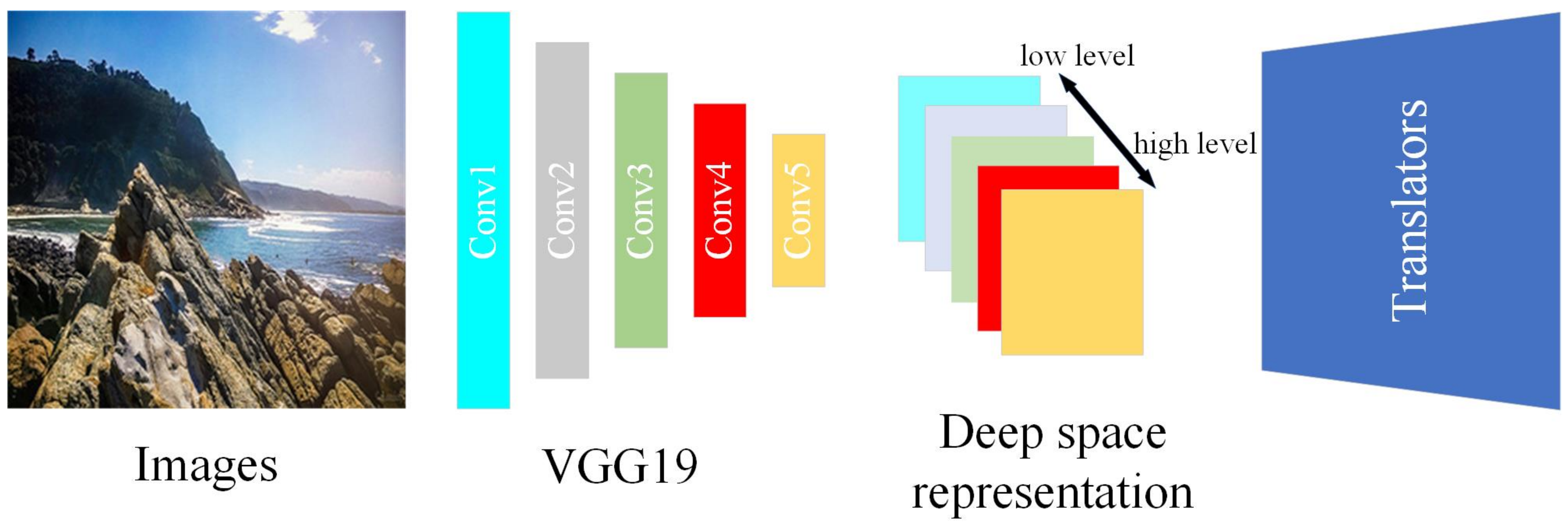

Figure #1: Image-to-image translation from shared deep space. Previous works usually translate images in pixel level, and they put the raw pixel data into translator to tranlate images from one domain to another domain. Our method applies a pre-trained VGG19 network on Imagenet to encode images into a shared deep space. The deep space representation expresses images from low-level to hight-level, and such a hierarchical representation can be separatly input into the layers of traslators, which can effectively fuse the features of original and target domains during cross-domain translation.

Approach

- Assume that there are two image domains *A* and *B*. x_a is an image of domain *A*, x_b is an image of domain *B*. There are two encode-decode networks: $ED_A = \{VGG, G_1\}$, $ED_B = \{VGG, G_2\}$ and two generative adversarial networks: $GAN_A = \{\{VGG, G_1\}, D_1\}$, $GAN_B = \{\{VGG, G_2\}, D_2\}$. Our method consists of three processes.
- In self-reconstruction process, x_a and x_b pass throught encode-decode networks to get the reconstructive images r_a and r_b.
- In cross-domain process, x_a and x_b first pass throught cross-domain encode-decode networks to get the "fake" images f_a and f_b, then f_a and f_b go back to get the cycle images c_a and c_b.
- In adversarial process, the "real" domain images x_a, x_b and the "fake" domain images are put into discriminators to get adversarial loss.

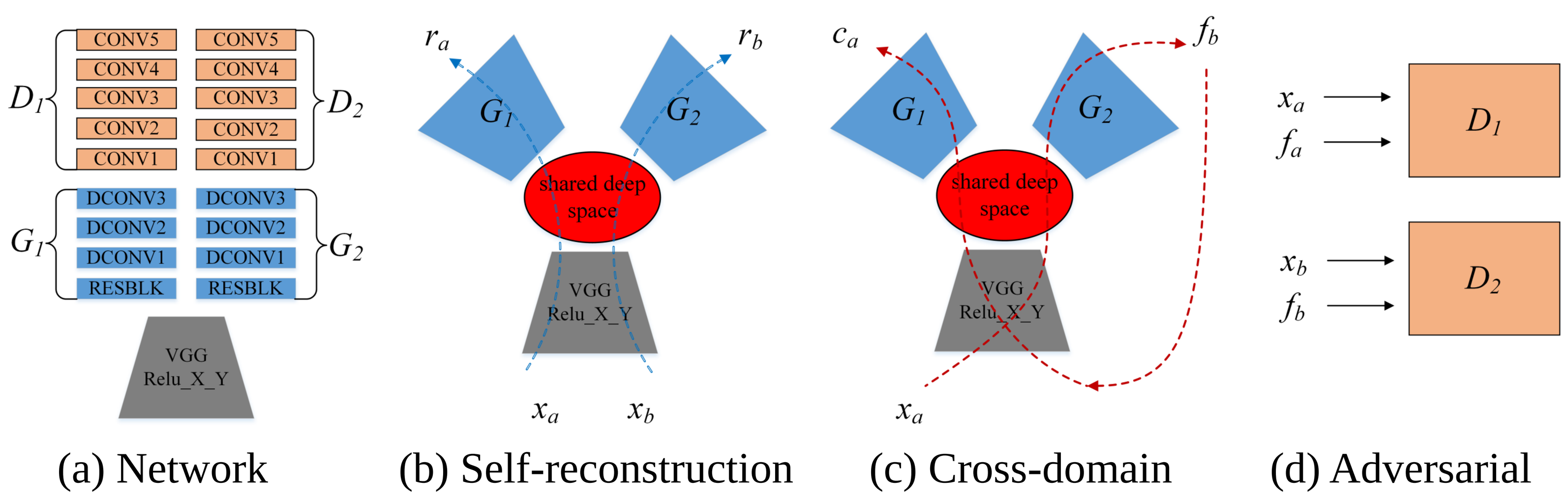

Figure #2: Overview of our architecture.

Evaluation

- We compare with some state-of-the-art methods, CoGAN[1], BiGAN/ALI[2,3], SimGAN[4] and CycleGAN[5]. Tables #1, #2 and Figures #1, #2 present the results. Our method achieves both numerical and perceptual superiorities to existing methods.

Method	Map → Photo Turkers labeled real	Photo → Map Turkers labeled real
CoGAN	0.6%±0.5%	0.9%±0.5%
BiGAN/ALI	2.1%±1.0%	1.9%±0.9%
SimGAN	0.7%±0.5%	2.6%±1.1%
CycleGAN	26.8%±2.8%	23.2%±3.4%
SDSGAN	**35.6%±3.1%**	**33.5%±2.8%**

Table #1: AMT real vs fake test on maps↔aerial photos at 256×256 resolution.

- The performance on AMT real vs fake and FCN scores shows our score is the highest, which indicates that SDSGAN has better image translation quality.

Method	Per-pixel acc.	Per-class acc.	Class IOU
CoGAN	0.40	0.10	0.06
BiGAN/ALI	0.19	0.06	0.02
SimGAN	0.20	0.10	0.04
CycleGAN	0.52	0.17	0.11
SDSGAN	**0.58**	**0.19**	**0.14**

Table #2: FCN scores for different methods, evaluated on Cityscapes labels→photos.

- In smiling to not smiling translation, SDSGAN can translate the emotion attribute while keeping the shape of face unchanged.

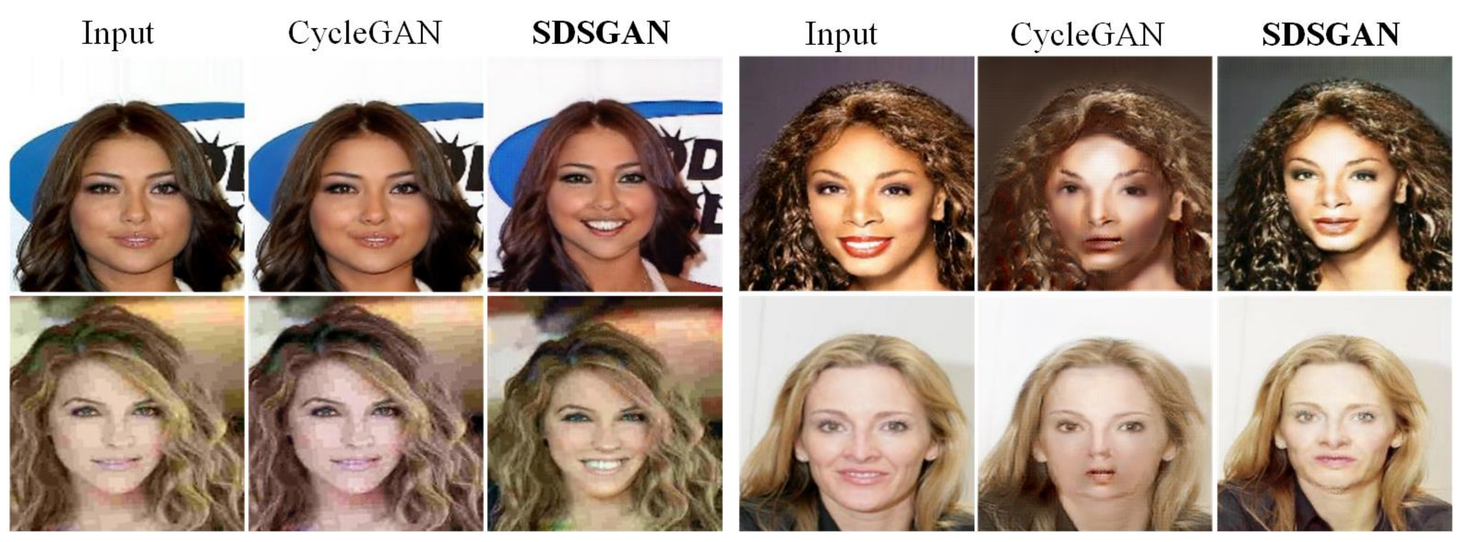

Figure #3: Generated images for not smiling to smiling (left) and smiling to not smiling (right).

- Performance on artistic style transfer shows SDSGAN performs better in texture translation than others.

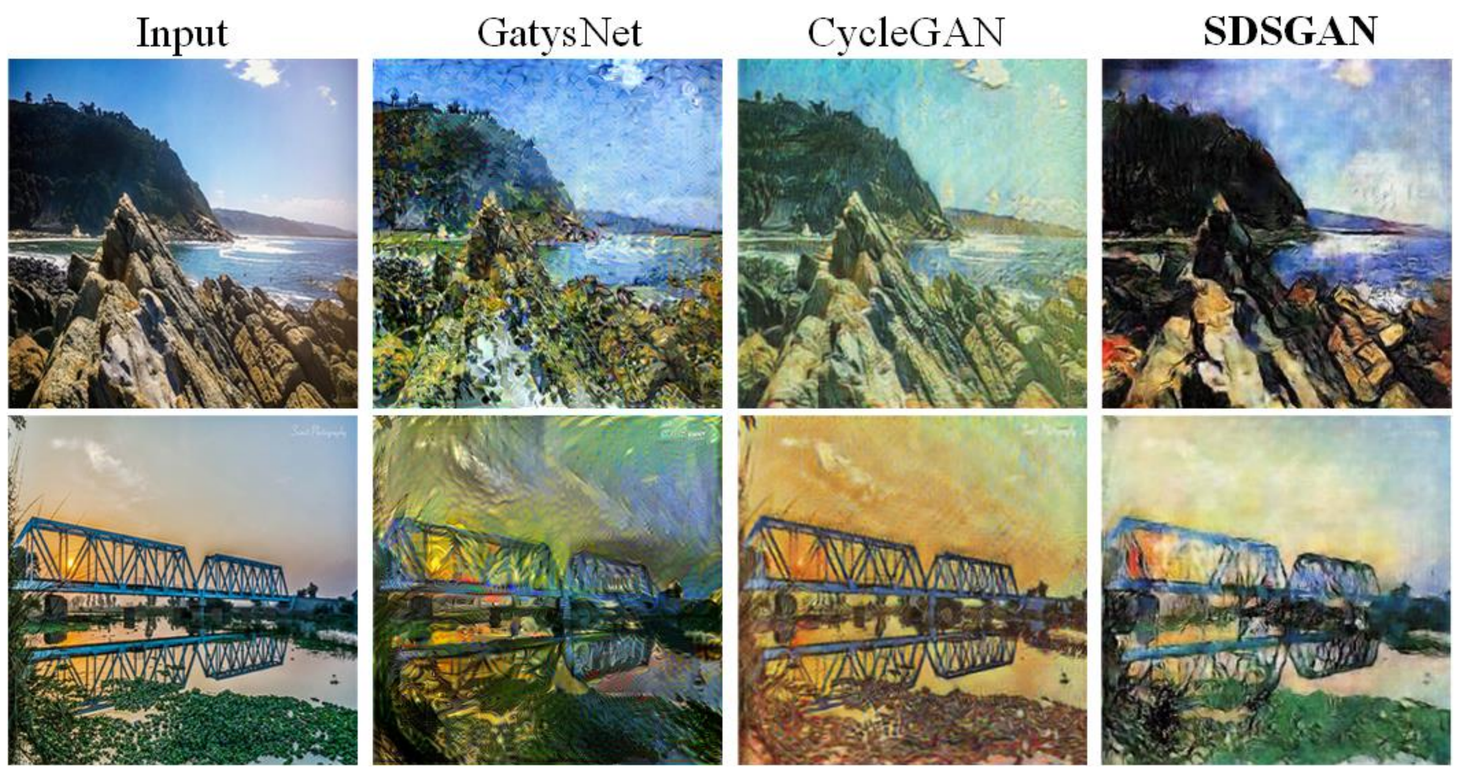

Figure #4: Generated images for artistic style transfer (image source: Vangogh dataset).

Conclusion

This paper proposes a novel framework for unpaired image-to-image translation using shared deep space. Both two images are encoded into a shared deep space through a pre-trained VGG-19 network, and then we use two decoders to convert them separately to corresponding image domains. In addition, we introduce skip-connection block and self-reconstruction loss to facilitate the mapping. Experimental results show that the proposed SDSGAN has both numerical and perceptual superiorities to existing methods.

References

1. Ming-Yu Liu, Oncel Tuzel: Coupled Generative Adversarial Networks. NIPS 2016: 469-477
2. Jeff Donahue, Philipp Krahenbuhl, Trevor Darrell:Adversarial Feature Learning. ICLR 2017
3. Vincent Dumoulin, Ishmael Belghazi, Ben Poole, Alex Lamb, Martín Arjovsky, Olivier Mastropietro, Aaron C. Courville: Adversarially Learned Inference. ICLR 2017
4. Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Joshua Susskind, Wenda Wang, Russell Webb:Learning from Simulated and Unsupervised Images through Adversarial Training. CVPR 2017: 2242-2251
5. Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros:Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. ICCV 2017: 2242-2251

电子科技大学
University of Electronic Science and Technology of China