Heterogeneous Face Recognition from Facial Sketches

Ivan Gruber
UWB, Faculty of Applied Sciences, DEPT of Cybernetics, Pilsen, Czech Republic
E-mail: grubiv@kky.zcu.cu

Thesis summary

Keywords: generative adversarial networks, image translation, neural networks, machine learning

Received: September 21, 2020

This paper presents a short summary of a dissertation thesis [1]. The thesis presents a novel approach named X-Bridge for image-to-sketch translation for automatic heterogeneous face recognition. X-Bridge is based on a conditional adversarial network with an additional reconstruction path and a shared-latent space assumption between the original and the reconstruction path. With these modifications, the results provided by X-Bridge overcome other tested state-of-the-art methods. Code is available at: https://github.com/YvanG/Cross-modal-Bridge.

Povzetek: Povzetek doktorske disertacije na kratko opisuje prepoznavanje obrazov iz skic.

1 Introduction

Generative adversarial networks (GANs) [2] have gained a huge amount of popularity in recent years thanks to their ability to generate photo-realistic results and the ability to capture important image details during image-to-image translation task. In this paper, a novel approach called X-Bridge is presented. X-Bridge is designed specifically as a cross-modal bridge in the heterogeneous face recognition task. X-Bridge is a supervised method and its structure is based on a conditional GAN, however, it also assumes shared-latent space across two different domains. To fully demonstrate the abilities of the X-Bridge approach, we test it on the arguably very challenging task of facial sketch-to-image translation using CUFSF dataset [3].

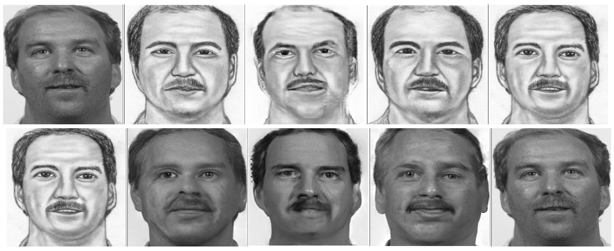

Figure 1: Cross-domain translation comparison. There are input images (the first column), translated corresponding sketches using Pix2pix (the second column), MUNIT (the third column), X-Bridge (the fourth column), and ground-truth outputs (the last column).

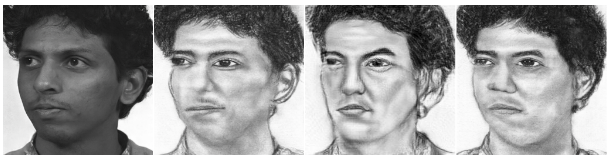

Figure 2: Cross-domain translation comparison for non-frontal view. The order of methods is the same as in Fig. 1.

2 X-Bridge method

X-Bridge contains two main paths based: translation path, and reconstruction path. These paths can be imagined as two separate GANs. They both have their own generator and discriminator, whereas both of them share one shared encoder. Each path has its own specific task. The task of the translation path, based on the conditional GAN, is to translate an input image from its domain into the other domain. On the other hand, the task of the reconstruction path, based on vanilla GAN, is to reconstruct the original input. Via this process, the reconstruction path motivates the shared encoder to preserve important features, to generalize better, and to learn important regularities. To further improve features propagation through the networks, skip connections in the form of channel concatenation between the last four layers of the encoder and the first four layers of the generators are added. Both of the paths utilize traditional adversarial loss, which can be for the domain translation problem expressed as follows:

$$L(EG, D) = \mathbb{E}_{\hat{x}}[\log D(\hat{x})]+$$
$$+\mathbb{E}_{x,z}[\log(1 - D(EG(x, z)))],$$

where x is a real image from the first domain, $\hat{x}$ is a real image from the second domain, the encoder, and the generator are together denoted as EG and D stands for the discriminator, and z is a vector from the shared-latent space.

Moreover, both paths utilize L_1 distance defined as follows:

$$L_1(EG) = \mathbb{E}_{x,\hat{x},z} \left[\|\hat{x} - EG_1(x,z)\|_1 \right].$$

The final loss is then defined as a sum of both losses for both paths.

3 Experiments

Several deep generative models were proposed for image-to-image translation in recent years. Most of existing approaches are based on supervised learning, however, models based on unsupervised learning became very popular lately. To benchmark X-Bridge, we decide to use two significant methods, one from each group. We compare it with the Pix2pix approach [4] and the MUNIT approach [5]. Both of these methods provide state-of-the-art results in image-to-image translation tasks, where Pix2pix is a supervised method the same as X-Bridge, whereas MUNIT is unsupervised. All the methods were trained and tested on the CUFSF dataset containing 1194 facial photo-sketch pairs.

In the first experiment, we test the translation of frontal views, see Fig. 1. All methods provide very realistic and precise results, however, we argue that both supervised models outperform the MUNIT approach, which has problems generating sharp images and therefore also small details. Pix2pix and X-Bridge reach comparable results. In the second experiment, where we test non-frontal cases, X-Bridge over-performed other methods, see Fig. 2 in terms of generalization, precision, and facial features preservation. All the methods were also tested on the IIIT-D Sketch dataset with comparable results.

4 Conclusion

We argue that qualitative results provided by X-Bridge overcome other tested methods. In our future research, firstly, we would like to propose a suitable metric to objectively compare the performance of methods in the image-to-image translation tasks, which is, to this day, non-existent. Secondly, we would like to address the ambiguity issue, which is a critical problem in the heterogeneous face recognition task.

Acknowledgement

This work was supported by the Ministry of Education of the Czech Republic, project No. LTARF18017.

References

[1] I. Gruber, "Heterogeneous face recognition from facial sketches," 2019.

[2] I. J. Goodfellow and al., "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, ser. NIPS'14. Cambridge, MA, USA: MIT Press, 2014, pp. 2672–2680.

[3] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[4] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *CoRR*, vol. abs/1611.07004, 2016.

[5] X. Huang, M. Liu, S. J. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," *CoRR*, vol. abs/1804.04732, 2018.