УДК 004.89

Learning disentangled representation of identity and head pose using generative adversarial networks

Bassel H. Zeno
(ITMO University, St. Petersburg, Russia)
Supervisor – D.Sc., Professor Yuri N. Matveev
(ITMO University, St. Petersburg, Russia)

The research were carried out at the expense of start-up financing of the ITMO University as part of the research project No. 618278 “Synthesis of Emotional Speech Based on Generative Adversarial Networks”.

Benefiting from the convolutional neural networks trained on large-scale face databases [1, 2], the performance of face recognition systems has been significantly improved over the past few years. However, pose variations are still the bottleneck for many real-world face recognition scenarios.

Existing methods that address pose variations can be divided into two categories. One category tries to adopt hand-crafted or learned pose-invariant features [1, 2] while the other resorts to synthesis techniques to synthesize face images of a specific person. For example, TP-GAN [3] and FF-GAN [4] attempt to recover a frontal view image from any face image with large pose. DR-GAN [5] can change the pose of an input face image. However, these methods can only manipulate limited poses of a face image. In additions they also require full annotation of attributes to train the models. Some of these GAN-based methods [4, 5, 6] usually have a single-pathway design: an encoder-decoder network is followed by a discriminator network. While the others [3] have two-pathways design. The encoder (E) maps input images into a latent space (Z), and then fed into the decoder (G) with a pose vector to generate novel views.

In practice, in order to learn a meaningful representation of the training dataset without additional constraints, many different GAN frameworks were proposed. They learn interpretable and meaningful latent representations in an unsupervised setting, such as InfoGAN, BiGAN, or in supervised setting. Despite all the effort in this area, these approaches ignore one of the most fundamental principles of face image generation, which is the disentanglement of the face’s identity and pose.

Learning disentangled identity and pose codes is a challenging task since 1) computers need to “imagine” what a given object would look like after a 3D rotation is applied; 2) the multi-view generations should preserve the same “identity”; and 3) the developed algorithms for solving this class of problems train GAN in a supervised fashion, or in its conditional setting.

Despite all the effort in this area, the field still lacks of a coherent framework for unsupervised disentangled pose and identity representations learning. From this consideration, in this article, we propose an end-to-end framework to learn a pair of disentangled identity and pose codes for a given dataset.

Our framework consists of a single generator, in contrast to [3]. The generator has several upsampling, convolutional layers, and a set of residual blocks. The input of the generator is an identity code, while the desired pose code is fed into a multilayer perceptron (MLP) in order to generate parameters of AdaIn layers (Adaptive Instance Normalization) which they were injected in each residual block of generator.

References

1. D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In CVPR, 2013.
2. F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In CVPR, 2015.
3. R. Huang, S. Zhang, T. Li, and R. He. Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. arXiv preprint arXiv:1704.04086, 2017.
4. X. Yin, X. Yu, K. Sohn, X. Liu, and M. Chandraker. Towards large-pose face frontalization in the wild. arXiv preprint arXiv:1704.06244, 2017.
5. L. Tran, X. Yin, and X. Liu. Disentangled representation learning gan for pose-invariant face recognition. In CVPR, volume 4, page 7, 2017.
6. J. Bao, D. Chen, F. Wen, H. Li, G. Hua. Towards Open-Set Identity Preserving Face Synthesis 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018).