Generating Facial Images using VAEGAN

Ivan Gruber[1]

1 Introduction

A generative modeling is a powerful way how to learn any kind of data distribution using unsupervised learning. The main advantages of generative models are as follows: (1) they generally don't need any labeling during training; (2) they are able to generate new data similar to existing data. Thanks to these features, generative models become very popular during last three years with a vast field of usage. In this work, I introduce an experiment of generating facial images by using VAEGAN. For the training of my model, I used Casia WebFace database (Yi et al. (2014)) and CelebFaces Attributes dataset (Liu et al. (2015)).

2 VAEGAN

In 2014, Goodfellow et al. (2014) proposed a new approach for estimating generative models via an adversarial process. They called it generative adversarial network (GAN). However, there were a couple of downsides to using only a plain GAN. First, the images are generated off some arbitrary noise, so there is no way to determine which initial noise to use if you wanted to generate a picture with specific features. Second, a generative adversarial model only discriminate between "real" and "fake" images, i.e. there are no constraints on how the image should look like. These two problems solve variational autoencoder (VAE) (Kingma and Welling (2014)), which is an artificial neural network used for unsupervised encoding. In 2015, Larsen et al. (2015) combined these two approaches together and proposed a new approach called VAEGAN. This model is composed of three parts - encoder, decoder, and discriminator, see Figure 1. In my article, all three of them are represented by a neural network.

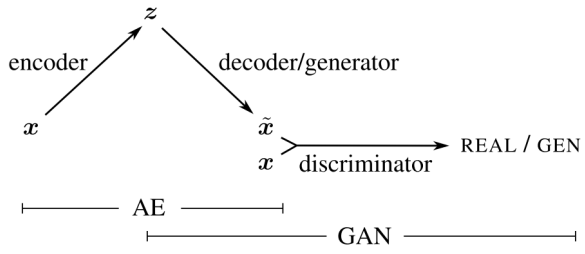

Figure 1: Overview of VAEGAN approach - combining a VAE with a GAN.

[1] Ph.D student, University of West Bohemia, Faculty of Applied Sciences, Department of Cybernetics, Computer Vision specialization & Ph.D student, ITMO University, Saint Petersburg, email: grubiv@kky.zcu.cz

3 Experiment Setup

For the training, I used data from two famous facial datasets - Casia WebFace database, and CelebFaces Attributes dataset. I did not make any augmentations, I just resize all images to the same size and then cropped the area of size 64×64 pixels around the center of each image. Such modified images, I used to train my neural networks. I utilize pretty standard architecture for each of the parts of VAEGAN. The neural network was trained with 50 epochs with mini-batch size 64 and initial learning rate = 3×10^{-3}. For updating network parameters RMSProp optimization was used. VAEGAN was implemented in Python using Tensorflow machine learning framework. The examples of results can be found in Figure 2.

Figure 2: The examples of reconstruction. There are original images in the first line. In the second line, there are their reconstructions.

4 Future Work

In my future work, I would like to train VAEGAN to generate facial sketches from facial images. This approach has a potential to be used as a cross-modal bridge in a task of heterogeneous face recognition. It should be noted, that the cross-modal bridge is the weakest part of modern heterogeneous face recognition algorithms.

Acknowledgement

This work is supported by grant of the University of West Bohemia, project No. SVK1-2018-024.

References

Larsen, B.L., Sonderby, S.K., Winther, O., (2015) Autoencoding beyond pixels using a learned similarity metric. CoRR.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2014) Generative Adversarial Nets. Curran Associates, Inc. pp. 2672–2680.

Kingma, D. P., Welling, M. (2014) Auto-Encoding Variational Bayes. Proceedings of the 2nd International Conference on Learning Representations (ICLR).

Yi, D., Lei, Z., Liao S., Li, S.Z. (2014) Learning Face Representation from Scratch.

Liu, Z., Luo, P., Wang, X., Tang, X. (2015) Deep Learning Face Attributes in the Wild. Proceedings of International Conference on Computer Vision (ICCV).