Refining Synthetic Face Images using Generative Adversarial Networks

Introduction

Training neural networks for Computer Vision tasks, such as face recognition or face detection, requires large amounts of data that can not always meet all the criteria that we are looking for, such as having very diverse complexitions or face orientations. Producing synthetic face images using 3D modeling tools is quite straightforward nowadays. However, these images can not be used as they are for the tasks aforementioned, as they do not provide enough resemblance to real face images.
In order to be able to use the synthetic face images, we need to find a mechanism of making them more realistic. We approach this problem from a Machine Learning point of view, making use of Generative Adversarial Networks (Goodfellow et al.).

Background

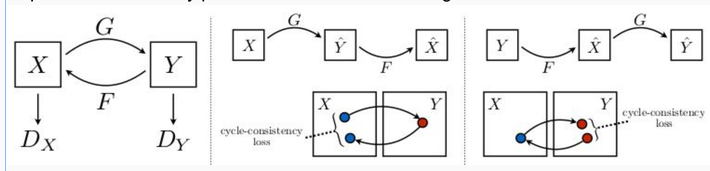
CycleGAN (Zhu et al.) is a type of conditional GAN that receives an image and tries to change characteristics of it to adapt it to another kind of image. A typical example is transforming horse images into zebra images. It is based on using two GANs, one for the direct transformation of the image and one for inverse transformation, therefore the name of CycleGAN. This process ensures that the annotations of the original image are preserved. The loss function can be written as:

$$L(G,F,D_X,D_Y) = L_{GAN}(G,D_Y,X,Y) + L_{GAN}(G,D_X,Y,X) + \lambda L_{cyc}(G,F)$$

$$L_{GAN}(G,D_Y,X,Y) = \mathbb{E}_{y\sim p(y)}[\log D_Y(y)] + \mathbb{E}_{x\sim p(x)}[\log (1 - D_Y(G(x)))]$$

$$L_{cyc}(G,F) = \mathbb{E}_{x\sim p(x)}[||F(G(x)) - x||_1] + \mathbb{E}_{y\sim p(y)}[||F(G(y)) - y||_1]$$

CycleGAN could be used to map human like features to the synthetic images. The experimental instability pointed us towards another algorithm.



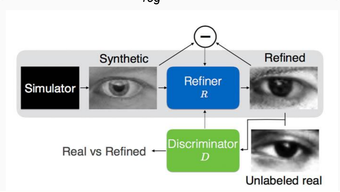
SimGAN (Shrivastava et al.) is another conditional GAN, which uses a so called “refiner” network in place of the generator to make synthetic images more realistic. The annotations of the synthetic images are protected, so SimGAN would be the perfect candidate for generating realistic face images that present a set of desired features. Given the unlabeled dataset of real images $\mathcal{Y}$, the loss functions for the refiner and the discriminator are as follows:

$$L_R(x;\theta) = L_{real}(x,\mathcal{Y};\theta) + \lambda L_{reg}(x;\theta)$$

$$L_{real}(x,\mathcal{Y};\theta) = -\log(1 - D_\varphi(R_\theta(x)))$$

$$L_D(x,y;\varphi) = -\log(D_\varphi(R(x))) - \log(1 - D_\varphi(y))$$

The realism loss L_{real} is used to make the input image similar to the ones in the dataset, while the regularization loss L_{reg} is used for conserving the annotations.



Our solution

Our solution uses the SimGAN algorithm, in which we tried using several network architectures and regularization techniques. The original algorithm was tested on refinement of eyes and hands images and we wanted to extend it to the case of synthetic face images that were generated using internal company software.
The first step was adapting the dataset loader to support face images in both grayscale and color format.
We then started testing with grayscale images. The refinements made on these images were hardly identifiable, the output images having a clear synthetic look. The use of color images seemed to improve the performance, as the network had more information and could therefore start changing skin tones and add shades.
The original article suggests using the mean of the RGB channels for computing the regularization loss, though this does not take into account the information that each of the channels brings. We therefore opted for a mean of the differences between each channel. A typical regularization loss looks as follows:

$$L_{reg}(x;\theta) = ||x - R_\theta(x)||_1 / n,$$

where n is the number of channels in the input image.
We tested both L1 and L2 losses as well as using feature extractors, such as ResNet networks, from which we would use intermediate feature maps, but losses computed over output images proved to be the most useful.

Results

We have used our ideas for refining grayscale and color face images.
Whereas grayscale images do not show real improvements, color images exhibit some kind of adaptation of skin tones towards a more realistic appearance, as it can be seen below. Little to no changes were observed regarding the change of face anatomy, pointing to a possibly excessive annotation preservation effect.

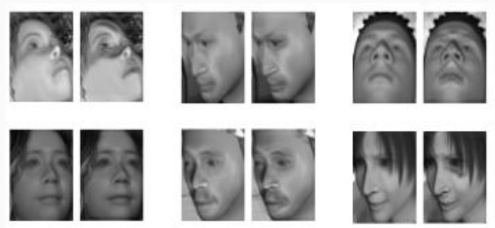


Figure 1. Refinement process on grayscale faces

Figure 2. Refinement process on color faces

Analysis

A normal training of our SimGAN adaptation leads to loss functions evolving like below. We see that discriminator follows a quite normal descending path towards convergence. The refiner however stagnates in terms of both the realism loss and the regularization loss, the former one being of a greater concern.

Conclusions

Making synthetic data more realistic is a hard task, as the algorithm needs adaptations according to the data we need to modify. The refinement of synthetic face images, and not only, is of utmost importance for cheaply obtaining datasets for other kinds of tasks.
We presented a possible approach to the problem of face image refinement, taking advantage of the SimGAN algorithm, which was already successful for similar problems: eye and hands image refinement.
Our solution still needs serious improvements and, among the things to try next, we thought about introducing another cost term similar to the one in CycleGAN, which would serve as an annotation preservation technique, as well as applying refinements to specific face parts, an idea inspired by the TP-GAN article (Huang et al.), which could reduce the complexity of refining an entire face image.

Acknowledgements and Bibliography

I thank my internship mentors: Liviu-Cristian Duțu and Oana Pârvan, from Fotonation, who have coordinated me and introduced me to the wonderful world of Machine Learning and its Computer Vision applications in particular.

[1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. Generative Adversarial Nets, arXiv:1406.2661v1, 2014
[2] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, arXiv:1703.10593v1, 2017
[3] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Josh Susskind, Wenda Wang, Russ Webb. Learning from Simulated and Unsupervised Images through Adversarial Training, arXiv:1612.07828v2, 2017
[4] Rui Huang, Shu Zhang, Tianyu Li, Ran He. Beyond Face Rotation: Global and Local Perception GAN for Photorealistic and Identity Preserving Frontal View Synthesis, arXiv:1704.04086v2, 2017