Deeply Good Fakes

Tristan Breetz
Dr. Junxiu Zhou, Dr. Nicholas Caporusso

Abstract

"Deep Fakes" technology focuses on the creation and detection of algorithmically generated realistic images, voice, and text. This information can be used to create realistic video memories to help relatives and close friends cope with absence or loss of a beloved one. In this work, we collect a set of videos to train a deep neural network. Using this trained network, we are able to generate unseen videos. These new videos will serve well reinforcing users' physical presence, remembering the subject after they passed away. Therefore this work is able to achieve the intended memory generation task.

Problem

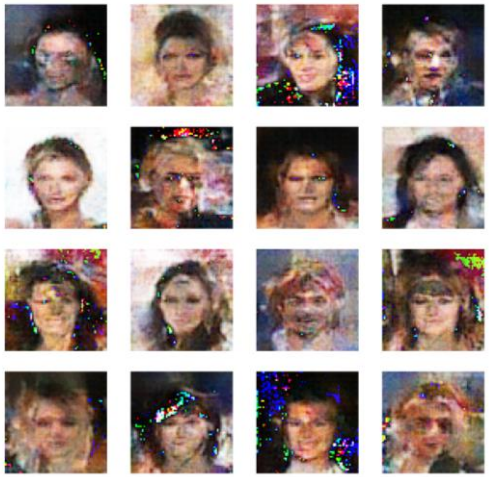
A substantial amount of current machine learning research focuses on the synthesis of novel media from large, already available datasets. The synthesis of this fake, realistic data is referred to as "deep fake" generation. The ability to create photorealistic imagery or voice data comes with some obvious ethical considerations. For this reason, the driving force behind this project is to showcase a benevolent use case of the technology; to create realistic memories in order to help relatives and close friends cope with absence or loss of a beloved one.

Methods

The model developed for this project is an implementation of DCGAN (Radford, Metz, & Chintala, 2016). In comparison with results shown in the original DCGAN paper, the results here are underwhelming. This is subject to investigation, as there is a good likelihood that one or more aspects of the current implementation are not ideal for the dataset used here, the CelebFaces Attributes (CelebA) Dataset. The dataset contains just over 200,000 images which were used for training. The dataset used in the original paper contained over 3 million images. It is suspected that this is a significant cause of the non-photorealistic results after 50 epochs of training. Perhaps more training would help some; but more data would be more helpful.

Results

While current results for this project are not yet good enough to serve as video memories, successful depictions of humans have emerged from the created generative adversarial network. This gives good confidence that results can be improved by tweaking the network structure and its parameters. Additionally, results from off-the-shelf models such as DeepFaceLab or Faceswap clearly show the project's goal attainability in terms of life-like recreation of human features. Shown are results from our work, as well as results from StyleGAN (Karras, et al, 2018), a Deep Neural Network with a different and more complex architecture developed by the NVIDIA research group. Results of StyleGAN caliber are easily good enough to use in video memories.



a) Our results after 50 epochs of training

Future Work

Of course, generating still images is only part of the solution to creating fully immersive video memories of loved ones. Creating moving images and accompanying voice generation that also sounds realistic adds another few layers of complication. That being said, the technology to implement these things already exists and is just waiting to be brought to life.
Work will also be made in enhancing the overall structure of the network, distinguishing it from being a straightforward implementation of the DCGAN architecture. Currently, the addition of a second GAN which takes the output of the current model and upscales it to a desirable definition (e.g. 1920x1980 pixels) is being considered. This secondary network would be of very similar structure to the current one and would easily be trained from the currently used dataset, as well as the newly generated dataset. A drawback to this design is that training time will of course, increase. This is not a significant problem as training time is not overly long as is; and once the model is trained it to satisfaction, it does not need to be repeated. The use of an upscaling network is not a novel concept and has been shown in previous research (Wang, et al., 2018).

b) Results of StyleGAN by NVIDIA Research.

Reference

Radford, Metz, & Chintala. (2016). Unsupervized Representation Learning with Deep Convolution Generative Adversarial Networks. indico Research, Facebook AI Research (ICLR Conference Paper).

Wang, Perazzi, McWilliams, Sorkine-Hornung, Sorkine-Hornung, & Schroers. (2018). A Fully Progressive Approach to Single-Image Super-Resolution. ETH Zurich, Disney Research.

Tero Karras, et al. "A Style-Based Generator Architecture for Generative Adversarial Networks." (2018).