Super Resolution of occluded or unclear faces using Generative Adversarial Networks

Project Proposal | DS 5500 | Fall 2019

Akhil Nair | Mounica Subramani | Mrinal Soni | SuRui Yang

1 Summary

Image or video data contain plenty of information and have a wide range of applications in the field of research and development. Many applications require zooming of a specific area of interest in the image where in high resolution becomes essential, e.g. tumors diagnosis, visual surveillance and autonomous vehicle navigation. However, no matter how well an upscaling algorithm performs, there will still be some amount of information lost and drop in image quality. An efficient fix for certain situation deserves high attention.

Estimating a high-resolution (HR) image from its low-resolution (LR) counterpart is referred to as super-resolution (SR). In this project, we aim to explore image super-resolution using generative adversarial networks, trying to minimize detail lost during the upscaling procedure. Based on the results of the experiments we would like to extend our methods and algorithms towards video footage Super Resolution, in the second phase of the project.

Generative Adversarial Network's are a relatively new area in the field of deep learning proposed by Goodfellow, et al.. GAN's focus on generating data by learning to mimic distribution of data. The way a GAN works is by this architecture where you have two networks behaving like adversaries i.e. generator tries to fool the discriminator.

(a) Discriminator: A discriminator classifier that decides whether an image is real or fake

(b) Generator: A generator that generates an image from random input noise

2 Proposed plan of research

We adopt a variation of GAN called SRGAN that employs a deep learning network inspired from ResNet and diverges from the traditional Mean Square Error (MSE) loss function to a perceptual loss function.

Our work flow would start by creating a new set(X) of images by down-sampling the images from our dataset(Y). The images from set X would be called a low resolution(LR) image whereas the ones from set Y would be called a high resolution image(HR). We would then feed to our generator LR images which then outputs an image that it considers to be of high resolution, we call it fake HR. The discriminator network is fed images from set Y(HR) and images generated from generator(fake HR) which then classifies the authenticity of the image. As training proceeds the discriminator becomes better at distinguishing HR,fake HR images and generator becomes better at generating indistinguishable images.

The learning is enabled by a double feedback loop, one between the HR image and discriminator, the other between discriminator and generator. The feedback loop utilizes a perceptual loss function which is a combination of Content and Adversarial functions that keep perceptual similarity between true and generated images instead of pixel wise similarity (traditional MSE), allowing us to recover photo-realistic textures from heavily down sampled images

The implementation of the model is done using tensorflow's TFGAN library and trained using TPU's available on google cloud platform.

3 Preliminary results and datasets

Figure 1: From left to right: High resolution (128 X 128), 2X downsampling, 4X downsampling

The super resolution models are experimented on a widely used benchmark dataset Celeb-A. Celeb-A is a large-scale facial attributes dataset with 202,599 face images of 10,177 unique identities. The images are mostly frontal images and less occluded which might create a bias in the model.
Due to this assumption, we in the later part of the project would experiment the model on Indian Movie Face database (IMFDB). IMFDB is a large unconstrained face database consisting of 34512 images of 100 Indian actors collected from more than 100 videos. Unlike the Celeb-A dataset the faces in IMFDB are collected from videos collected from the last two decades by manual selection and cropping of video frames resulting in diversity in age, poses, dress patterns, expressions etc.

The low resolution images are obtained by downsampling the high resolution images from the dataset. There are many resampling algorithms available, we use a Lanczos kernel to downsample the images with the help of pythons PIL library. A sample of the Lanczos downsampling can be seen in Figure 1

4 Related Work

A variety of work has been done in the field of image super resolution. Convolutional neural network(CNN) based super resolution models have shown great perfomances. A CNN class GAN called Deep Convolutional Generative Adversarial Network was proposed by Radford, et al..

References

[1] C. Ledig, L. Theis, F. Husz ar, J. Caballero, A. Cunningham,A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al.Photo-realistic single image super-resolution using a gener-ative adversarial network.arXiv preprint arXiv:1609.04802,2016

[2] A. Radford, L. Metz, and S. Chintala. Unsupervised repre-sentation learning with deep convolutional generative adver-sarial networks.arXiv preprint arXiv:1511.06434, 2015.

[3] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. InTheIEEE Conference on Computer Vision and Pattern Recogni-tion (CVPR), June 2016.

[4] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning faceattributes in the wild. InProceedings of International Con-ference on Computer Vision (ICCV), 2015.

[5] S. Setty, M. Husain, P. Beham, J. Gudavalli, M. Kandasamy, R. Vaddi, V. Hemadri, J C Karure, R. Raju, Rajan, V. Kumar and C V Jawahar. Indian Movie Face Database: A Benchmark for Face Recognition Under Wide Variations, National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2013.