Super Resolution of occluded or unclear faces using Generative Adversarial Netowrks

Nair, Soni, Subramani, Yang

I. Summary

Image generation or super resolution of images always carry a big trade off between quantitative quality and perceptual quality of images. A lot of available approaches end up with great quantitative results but end up missing out on perceptual quality of images. Perceptual quality is defined as the ability to capture intricate details in features like texture, colour and shapes of objects in an image or video. Most image processing approaches leave the question of how to preserve perceptual quality unanswered. Our team aims to address and solve this issue using the concepts of Generative Adversarial Networks by using its variant SRGAN (Super Resolution Generative Adversarial Network), which aims to recover / preserve perceptual quality (the finer details) of images while super-resolving at large upscaling factors while also ensuring that it doesn't compromise on the quantitative performance.

Estimating a high-resolution (HR) image from its low-resolution (LR) counterpart is referred to as super-resolution (SR). Based on the results of the experiments we plan on extending our methods towards video footage Super Resolution for the second phase of our project.

1) Related Work: A variety of work has been done in the field of image super resolution. Convolutional neural net- work(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..

2) Data: 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. To make sure we eliminate this bias, we plan on implementing our current model on an 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 over the last two decades by manual selection and cropping video frames resulting a diversity in age, poses, dress patterns, expressions etc.

3) Architecture: Generative Adversarial Networks are a type of deep neural networks which are used to generate images with the help of two networks; generator and discriminator. Generator (as the name implies) attempts to generate images based on the input, and the discriminator tries to distinguish between the generated images and the true images to determine whether the images are real or fake. Both networks try to get the better of each other and are each others adversaries, hence the name Generative Adversarial Networks. Super resolution of a single image can be achieved by implementing GANs as it aims to refine and convert a low-resolution image to high resolution.

For the scope of our project, we adopt a variant of GAN called SRGAN (Super Resolved GAN), that employs a deep learning network inspired from ResNet which deviates from the traditional Mean Square Error (MSE) to a perceptual loss function; which is a combination of Adversarial Loss and Content Loss Function.

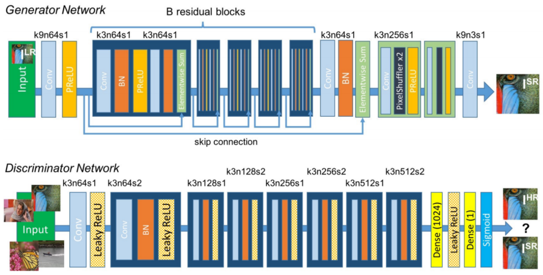


Fig. 1: Architecture

4) Loss Function: Traditionally GANs have focused on defining their loss functions based on pixel similarity in images by calculating the mean pixel value for a set of images, but that results in really smooth images where the model fails to capture the perceptual quality i.e. the texture of an image or the shape of an object in a particular image. Recent approaches have shown that defining a perceptual loss function is a better approach since it works towards preserving / recovering the features that might be lost in pixel based loss functions.

Perceptual loss function is defined as a combination of two functions Adversarial and Content, both being responsible for learning the weights of Discriminator and the Generator respectively. These two functions form a closed loop between the generator and discriminator where they effectively back-propagate after every iteration making the networks learn weights for better learning.

II. Methods

A. Data Preprocessing

To generate our training images, we down sampled our original images from 218x178 to 64x64, 32x32 i.e. high resolution and low resolution respectively. After obtaining the right set of LR and HR images we fed them into the network as features and labels respectively. To ensure that a one to one mapping between the LR and HR images were maintained, both the images were sorted in their corresponding lists.

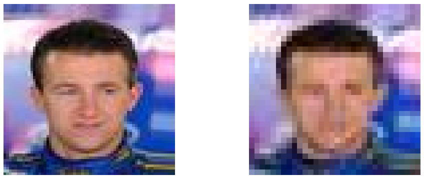
Once pre-processing was achieved, the next step was to train our GAN with the right set of loss functions and activation functions so that appropriate weights could be learned and model parameters could be tuned. LR images were converted to a bit map array to represent images in a quantifiable manner. We used TensorFlow's dataset input pipeline to read the images, decode them and then normalize it to values between [-1,1] as the activation functions that we use like 'tanh' tends to perform better in that range. Finally we zip the LR, HR images and then feed them in batches to our networks.

B. Implementation

Once fed into our model the generator takes the low resolution images as input and tries to identify the shape, colour and texture of objects in our images, generating a new fake image (called SR) based on what was learned from the LR images. These fake / generated images were then fed to the discriminator which also takes in the actual HR images as an input. Based on their quantitative values, discriminator classifies each image as fake (close to 0) or real (close to 1). Initially the generator model is quite naive, but as training progresses the generator becomes better at generating fake images and discriminator gets worse at distinguishing between the generated images and the actual images. Over time, the feedback provided by discriminator becomes less meaningful making the convergence of GAN really unstable as it starts to take in random feedback from the discriminator.

III. Results

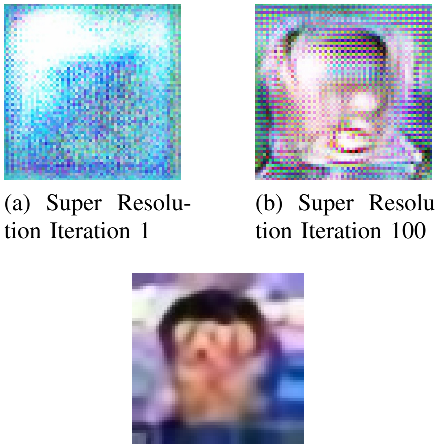
Once the models were built, we trained it for 500 iterations and observed the following results:



(a) High Resolution

(b) Low Resolution

Fig. 2: Input Images



(a) Super Resolution Iteration 1

(b) Super Resolution Iteration 100

(c) Super Resolution Iteration 500

Fig. 3: Result Images

As clear from the Figure 3, Iteration 1 produces a terrible image showing that the generator hasn't learned much from the input image and it still has to learn all the features of our input image. There is a definite improvement at the 100th iteration as now the generator finally starts to learn and recognize features like shape of the object and colour of that object in our image. At 500th iteration its clear that there is a significant improvement from the first 100 iterations as we see a face like structure with good amount of colours and a decent background, showing that the generator has finally started to learn the features and is getting better at generating images based on what is being fed. The quality of image generated at the 500th iteration is nowhere near the quality of the actual image which shows that the generator still has a long way to go, but was restricted due to the computational challenges faced while running the model on a local machine. This is further discussed in the next section.

IV. Discussions

There were a few challenges faced through the course of this project. It was a challenge to implement the network from scratch in TensorFlow, as it required a lot of research to get the architecture right. The next major challenge was the limitations in computational resources. The model being quite huge had 14 million trainable parameters in total, we had to considerably reduce the number of training images from 100K to 800 and the upscaling factor to about 2X to avoid running out of memory. The model was trained for 500 epochs totalling a runtime execution of around 50 hours. The first stage in the second phase of the project is to deploy the model in a cloud platform and benefit from its computational power.

We would then like to train the model on the entire dataset and optimize it further to make it scalable enough for super resolution of frames of a video.

V. Statement of Contributions

1) Akhil Nair Mrinal Soni: architecture implementation, presentation, report
2) Mounica Subramani Suri Yang: model architecture research, presentation, report, loss functions

Once individual tasks were completed, the team got together to combine the tasks and successfully implement the network of GAN. It was a collective effort to put this all together and achieve the results that we did.

VI. References

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

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

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.

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.

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.

Diederik P. Kingma, Jimmy Ba. Adam: A Method for Stochastic Optimization, arXiv:1412.6980, 2014.

Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, Jing-Hao Xue, Deep Learning for Single Image Super-Resolution: A Brief Review, arXiv:1808.03344, 2018.

Justin Johnson, Alexandre Alahi, Li Fei-Fei, Perceptual Losses for Real-Time Style Transfer and Super-Resolution, arXiv:1603.08155, 2016.