Image Super Resolution using General Adversarial Network
CSCI-B657: Computer Vision
Spring 2019
Interim Project Report

Darshan Shinde[1], Virendra Wali[1], Pei-Yi Cheng[1]

I. INTRODUCTION

In many crime incidents, we have some security camera footage. But, quality of footage might be very poor which makes difficult to get information about the criminal. To resolve this problem, there exists many image quality improving algorithms [4], [2] (and many more). Out of which we are trying to re-implement the idea of increasing the resolution of low-resolution images using GAN proposed by [1].

Author claims that the low-resolution images which are artificially generated using bi-linear/bi-cubic down sampling or blurring using several different sized blur kernels used in other image super resolution techniques does not resembles to real-world low-resolution images properly (ignores the degradation process e.g., motion blur, compression artifacts, etc.) and hence leads to poor performance during test time. To alleviate this, author has proposed process of producing low-resolution images using High-to-Low GAN model. He concluded that this process of generating low-resolution image and using this for learning GAN model for generating higher resolution images outperforms all other previous technique. To support that he has provided numeric result of some experiment in comparison with other approaches. In this project, we are trying to perform same operation over different data sets and check whether this approach really outperforms other approaches or not. In addition to that we will try to improve the performance this approach by combining it with some other existing techniques.

II. BACKGROUND & RELATED WORKS

The vast majority of work is done towards solving the problem of low resolution to high resolution. Prior to this, majority of super-resolution methods used low resolution image created by bi-linear down-sampling. bi-linear down-sampling followed by blurring or GAN with l2-pixel loss. The results of those methods are noisier and blurrier. To overcome the flaws of bi-linear down-sampling methods, GAN + L2 loss is used but L2 loss alone failed to de-noise the input and good out output. So, to solve this problem, a High-to-Low network GAN is used to learn how to down-sample high-resolution images first, and another Low-to-High GAN is used for super-resolving images into high-resolution ones.

[1]Darshan Shinde, Virendra Wali and Pei Yi Cheng are data science graduate students at the School of Informatics, Computing, and Engineering, Indiana University, Bloomington.

In addition to this, Dong et al. propose a Convolutional Neural Network model to jointly optimize three operations for super-resolution: extract feature maps of low-resolution image, non linearly map low-resolution feature maps to high-resolution patch representation, and reconstruct high-resolution image by combining the predictions within a spatial neighborhood. Mean Squared Error is used as the loss function. The results show that increasing the number of filters and using reasonably larger filter size would improve performance at cost of running time. However, deeper structure is not always a good choice for super-resolution.

III. PROGRESS SO FAR

Our model of image super resolution has 2 GAN models and for training them we need 2 datasets: dataset of high resolution images and dataset of low resolution images.

We almost completed the 1st sub-model of high-to-low GAN. Listing of the task done till now is as follows:

- Preprocessing: Re-scaled all images to same dimensions by keeping pixel density constant.
- Implemented basic GAN model to understand working of general adversarial network in detail for mnist dataset.
- Designed and implemented 1st sub-model of high-to-low GAN model to generate paired dataset of high resolution images and respective low resolution images with natural looking noise pattern.
- Trained designed model for high-to-low part on CelebA (high resolution images) dataset and Widerface (low resolution images) dataset.

We are trying to use network structure for generator and discriminator as mentioned in original paper. But, we are facing some computational resource limitation issue. The model take long time to train. We are trying to figure out way to keep our code running on silo even if we disconnected from the network.

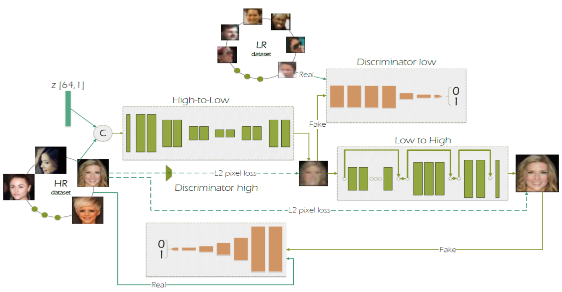
IV. METHODOLOGY & IMPLEMENTATION

Our system (the paper which we are re-implementing) uses General Adversarial Network to super-resolve the low-resolution image of size 16x16 into high-resolution of size 64x64. This network is trained on paired dataset of low-resolution images and high-resolution images. The main difference in implementation of this paper differs from past related work in sense of how it generated the paired dataset (in-short how it generates low-resolution images from high-resolution images). Many prior papers use bilinear downsampling process to generate low-resolution images from corresponding (original) high-resolution images which are flawed in some sense, which completely ignores the degradation effects like as motion blur and many compression artefacts which are very common in real world low-resolution images. Ignoring such important aspect results into less accurate high-resolution image. This paper is inspired from this point, the author of this paper has proposed a new architecture to learn this degradation effects.

The main architecture proposed in this paper has two main sub components (GANs), out of which first one (High-to-Low) is for learning the downsampeling and another (Low-to-High) is for the main purpose of improving quality of low-resolution image to high-resolution. High-to-Low is trained on two disjoint and unpaired datasets which can be used effectively to simulate the image degradation process. The first dataset contains the high-resolution facial images. The second dataset contains the blurry and low-quality low-resolution images. For our experiments, we are using Celeb-A dataset (High resolution) and Widerface dataset (low resolution) which is used to contaminate the high resolution images with noise and artefact to create low resolution images.

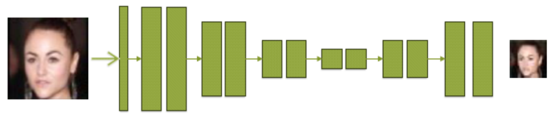
We are using this High-to-Low network to generate paired dataset of high-resolution images and their corresponding low-resolution images which we will use to train Low-to-High network. As a final model, both networks are combined to generate a final single network. The first generator module of this single network takes high-resolution image of size 64x64 along with random noise of size 64x1 as input and generates a low-resolution image of size 16x16. This low-resolution image is further passed to first discriminator module (which is trained on low-resolution image dataset) and judge the generated image is fake or real. Then this generated low-resolution image is passed to second generator module (which is trained on pair dataset of high-resolution images and low-resolution images) and converts it into high-resolution image of size 64x64 which is further passed to second discriminator module which classifies the ground truth image and fake images. This is can be easily understood from the Fig 1.

Fig. 1. Overall architecture and training pipeline



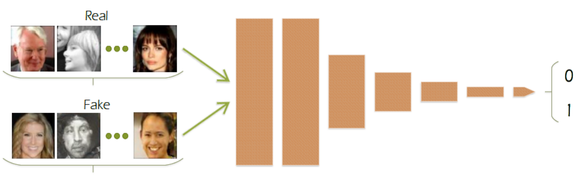
For High-to-Low generator, the first layer is a fully connected layer that takes high-resolution image concatenated with noise as model input. Then an encoder-decoder structure contains 6 groups of residual blocks (2 blocks in each group) so that image resolution is reduced from size 64x64 to size 4x4 after 4 times resolution degradation using pooling layers, and image resolution is increased from size 4x4 to size 16x16 after twice resolution increasing by using pixel shuffle layers. A residual block here consists of batch normalization, ReLU, Conv3x3, batch normalization, ReLU, and Conv3x3. Each layer feeds into the next layer while utilizing skip connection to jump over some layers. Fig 2 shows the architecture of generator model.

Fig. 2. High-to-Low generator



High-to-Low discriminator contains 6 residual blocks but without using batch normalization. Then a fully connected layer is used. The input size 16x16 resolution is reduced at the last two blocks by using max-pooling. Fig 3 shows the architecture of discriminator model.

Fig. 3. High-to-Low discriminator



Structure of residual blocks used in generator model and discriminator model shown in Fig 4 and 5.

Fig. 4. Residual block used for generator

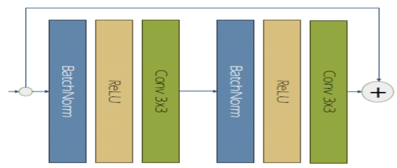
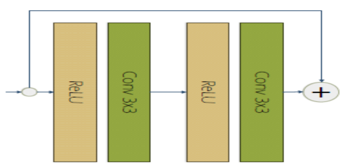


Fig. 5. Residual block used for discriminator



V. REVISED PROJECT TIMELINE

- Week 1:
  - Try some different network structures for high-to-low GAN model and train it to the sweet spot to get more accurate low resolution images. - Pie-Yi Cheng
  - Generate paired data set of high resolution and low resolution images using high-to-low GAN model. - Pie-Yi Cheng
- Week 2:
  - Design low-to-high GAN model and train it on paired data set. - Darshan Shinde
- Week 3:
  - Combine both GAN models (high-to-low and low-to-high). Train and test end-to-end model. - Virendra Wali
  - Implement FID [3] metric module which we are thinking to use for measure of accuracy for our model. - Darshan Shinde, Virendra Wali
- Week 4:
  - Try to use model for other data sets. -Darshan, Virendra, Pei-Yi Cheng
  - Write report and design poster for presentation. - Darshan, Virendra, Pei-Yi Cheng

VI. RESULTS

Low resolution images generated by high-to-low resolution GAN

Fig. 6. After 1400 epochs

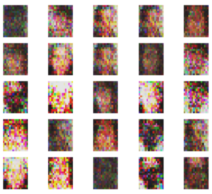


Fig. 7. After 2200 epochs

Fig. 8. After 2800 epochs

We are still trying some different network structure to get more better results.

REFERENCES

[1] Yang J. Bulat A. and Tzimiropoulos G. "To learn image super-resolution, use a GAN to learn how to do image degradation first". In: *"ECCV"* ("2018"). URL: http://arxiv.org/abs/1807.11458.

[2] Zhenan Sun Huaibo Huang Ran He. "Wavelet-SRNet: A Wavelet-Based CNN for Multi-scale Face Super Resolution". In: *"ICCV"* ("2017"). DOI: 10.1109/ICCV.2017.187.

[3] "Heusel M." "Unterthiner T." "Ramsauer H.". "Gans trained by a two time-scale update rule converge to a local nash equilibrium". In: *"NPIS"* ("2017"). URL: http://arxiv.org/abs/1706.08500.

[4] Isola P. Zhu J.Y. Park T. "Unpaired image-to-image translation using cycle-consistent adversarial networks". In: *"ICCV"* ("2017"). URL: https://arxiv.org/abs/1703.10593.