==

GAN's operating in frequency domain

Alokendu Mazumder (20134,PhD,EE), Saurabh Shrivastava (19881,MTech,AI)

Abstract

It's been noticed that though GAN's are able to produce visually great images but one can easily distinguish between real and fake ones by looking into frequency do-main. Spectral discrepancy is being observed between generated images and real images as explained in (1).Mostly GAN's are unable to reproduce the high frequency components. Explanations for this phenomenon are controversial: While most works attribute the artifacts to the generator, other works point to the discriminator. Here, we will stick to our reference paper (1) and try to come up with a new architecture that may produce good results than the baseline. In a nutshell, (1)introduces a new discriminator architecture to overcome the gap in frequency domain between real and generated images, they construct two discriminators, one that looks into spatial domain and the other that looks into frequency domain.

We basically choose to do image dehazing, i.e minimizing the distance between distribution of hazy and clean images. The data-set is a custom made one with 15k samples of each clean and corresponding hazy images. I inherited the data-set when I was student at IIT Jammu. All images are $28x28x3$, but we converted them to gray-scale while working.

Proposed Framework

Approach 1

Here, we made multiple generators, one looks into spatial domain of hazy images, one looks into a masked low pass magnitude spectrum, one into masked high pass and the last one into masked band pass. The frequency response of the output of respective generator is clubbed and sent to a discriminator.

We have two discriminators, one which takes input from the generator that looks into spatial domain and clean image, another that took input from the clubbed version of the output of the generators working in frequency domain, it takes FFT of clean image as real sample.

Approach 2

Here, we made two generators, one looks into spatial domain of hazy images, one looks into the inverse fft of a masked high pass magnitude spectrum. The frequency response of the output of respective generator is clubbed and sent to a discriminator.

We have two discriminators, one which takes input from the generator that looks into spatial domain and clean image, another that took input from the output of the generator that looks into frequency domain, it takes FFT of clean image as real sample.

Also, each G_i and D_i share same architecture as shown in the figure.

Training and Output

The generator and discriminator that looks into spatial domain, only they are trainable, rest of the CNN modules (frequency domain generator-discriminator pair) only contributed towards giving output probability of discriminator. We took average of the output of the two discriminators and the gradients only flow back to the generator-discriminator pair that works in spatial domain.

The output of the discriminator also depends on the output of the pairs that looks into frequency domain, hence making the spatial-generator robust to learn the proper frequency response of generated images. The model is being trained for 10k epochs. **Plain adversarial loss of SGAN is used for training**.

Results

The results we got are not even up-to the mark. For the first approach, the generator after training is generating black images for all test inputs of hazy images. While for second approach we are getting some weird colorful images output from generator.

- I think the failure of first approach is due to the fact that we directly dealt with mask version of magnitude spectrum of images which contains majority of dark pixels due to the mask. Hence the CNN's are unable to learn from it.
- Also, for failure of both approach maybe the fact that data-set used is pretty small. In second approach although we get something bad output, it's almost same for all test hazy images with minor pixel intensity variation, mode collapse maybe a potential reason.
- Also, it seems from maxpooling layers employed in the network architecture, we are not getting desired output.

Further fact is, we have only taken magnitude response in account, phase response should NOT be overlooked if we want our generated images ti be consistent in frequency domain. The **FID** score of 1000 randomly sampled generated points and clean images is very poor, in order of 10^3

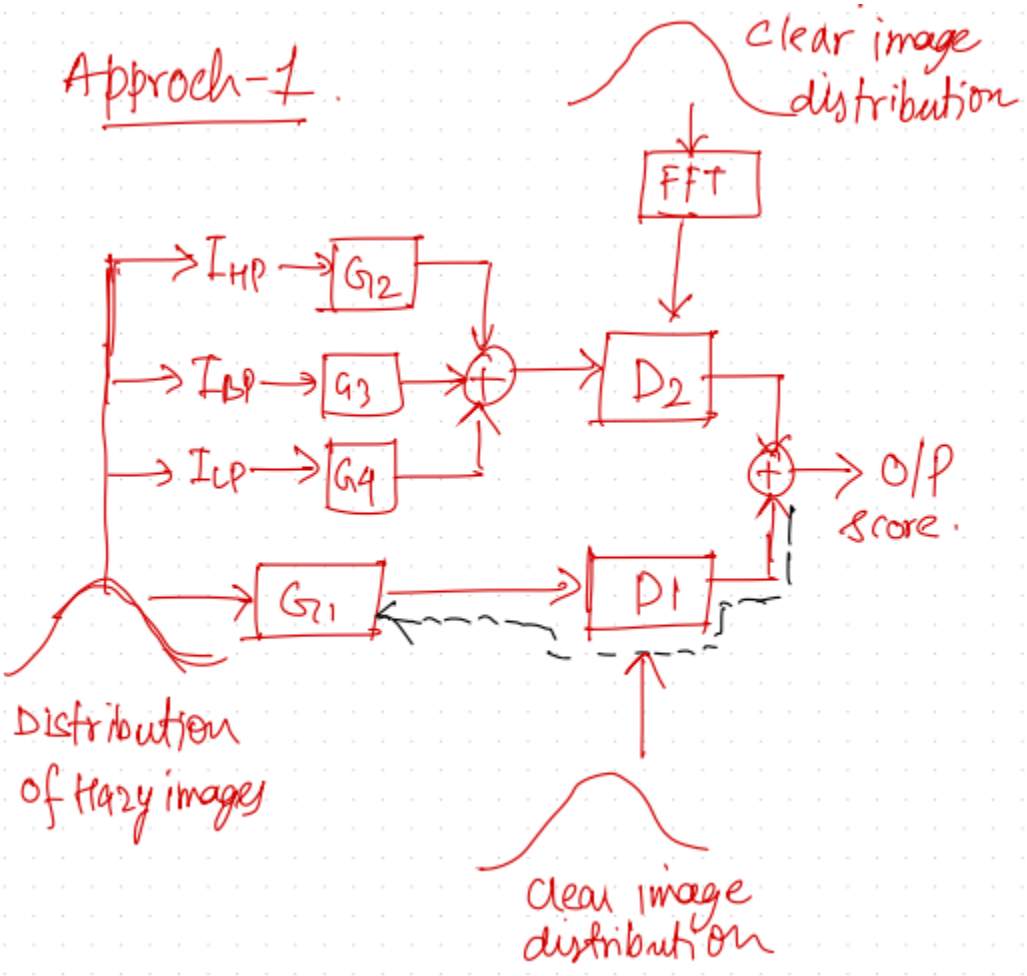


Fig. 1.—: Approach 1

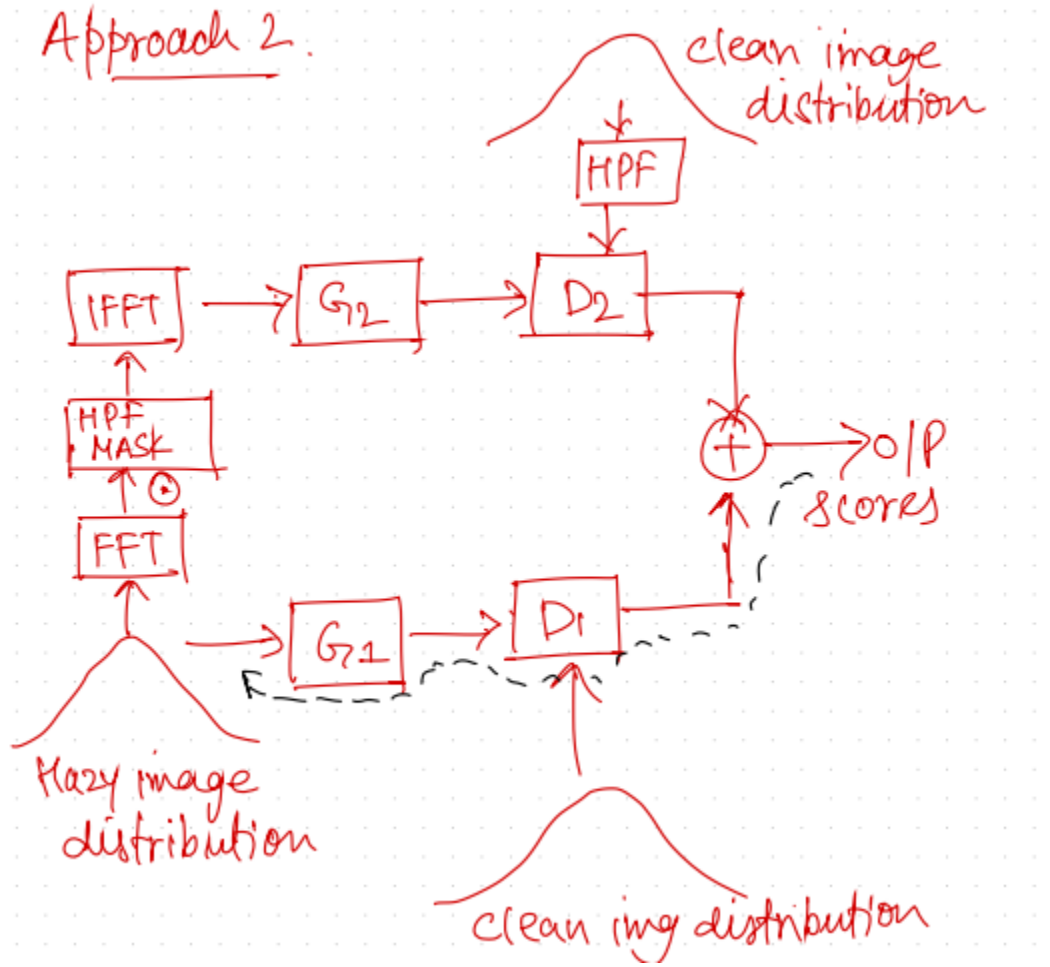
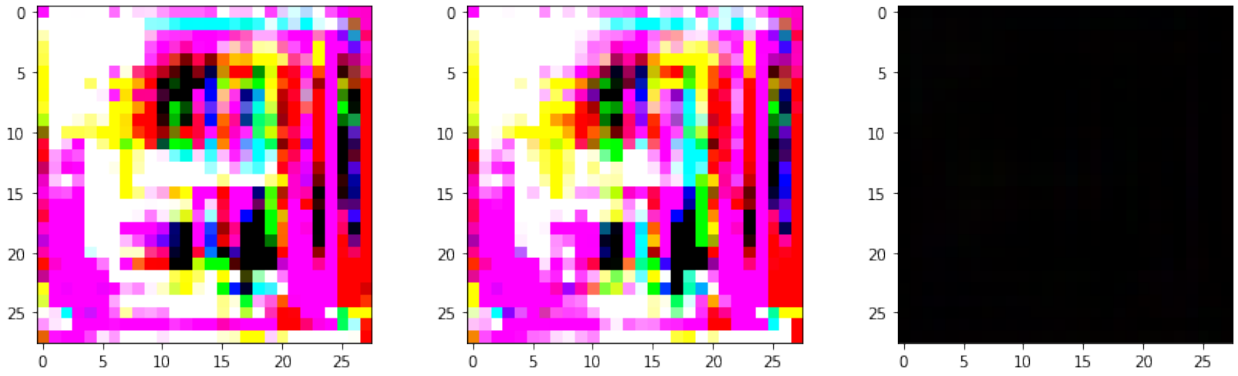


Fig. 2.—: Approach 2



(a) Output by Approach 2 (b) Output by Approach 2 (c) Output by Approach 1

Fig. 3.—: Results

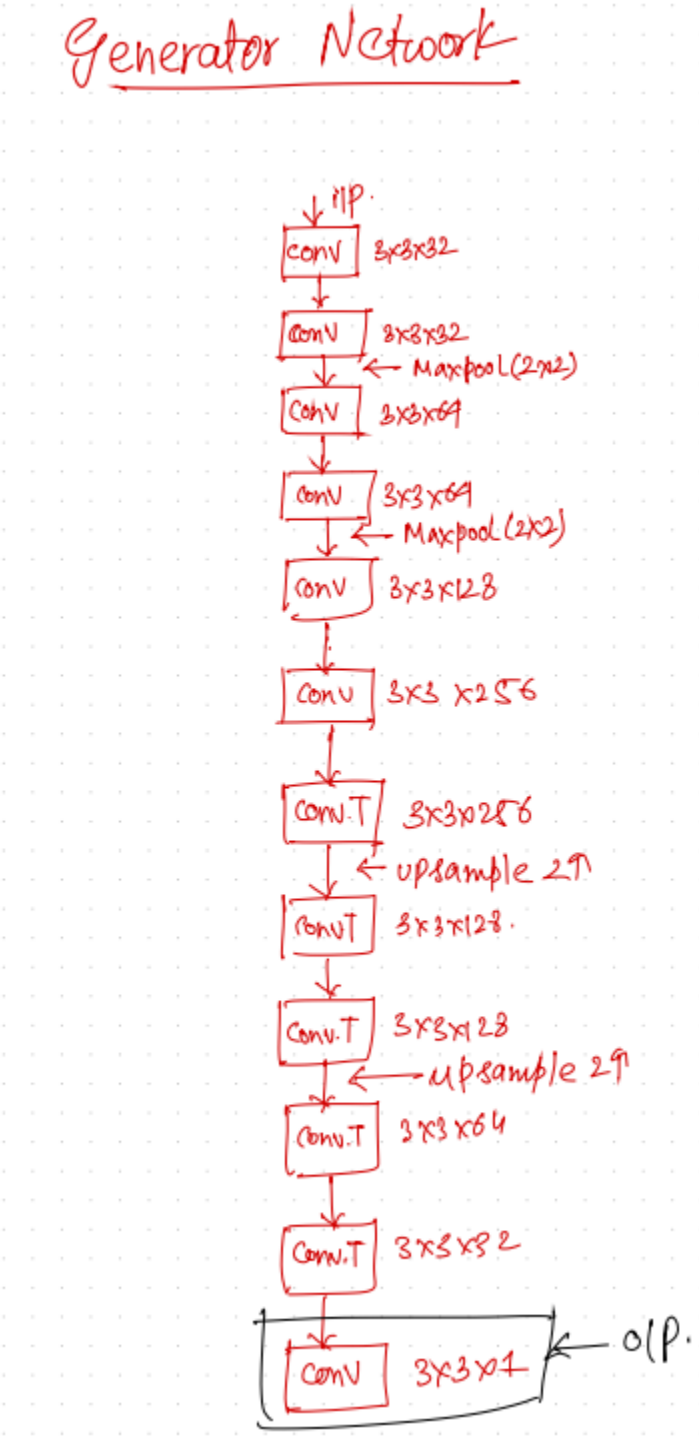


Fig. 4.—: Generator model common across all generators defined and across both approach

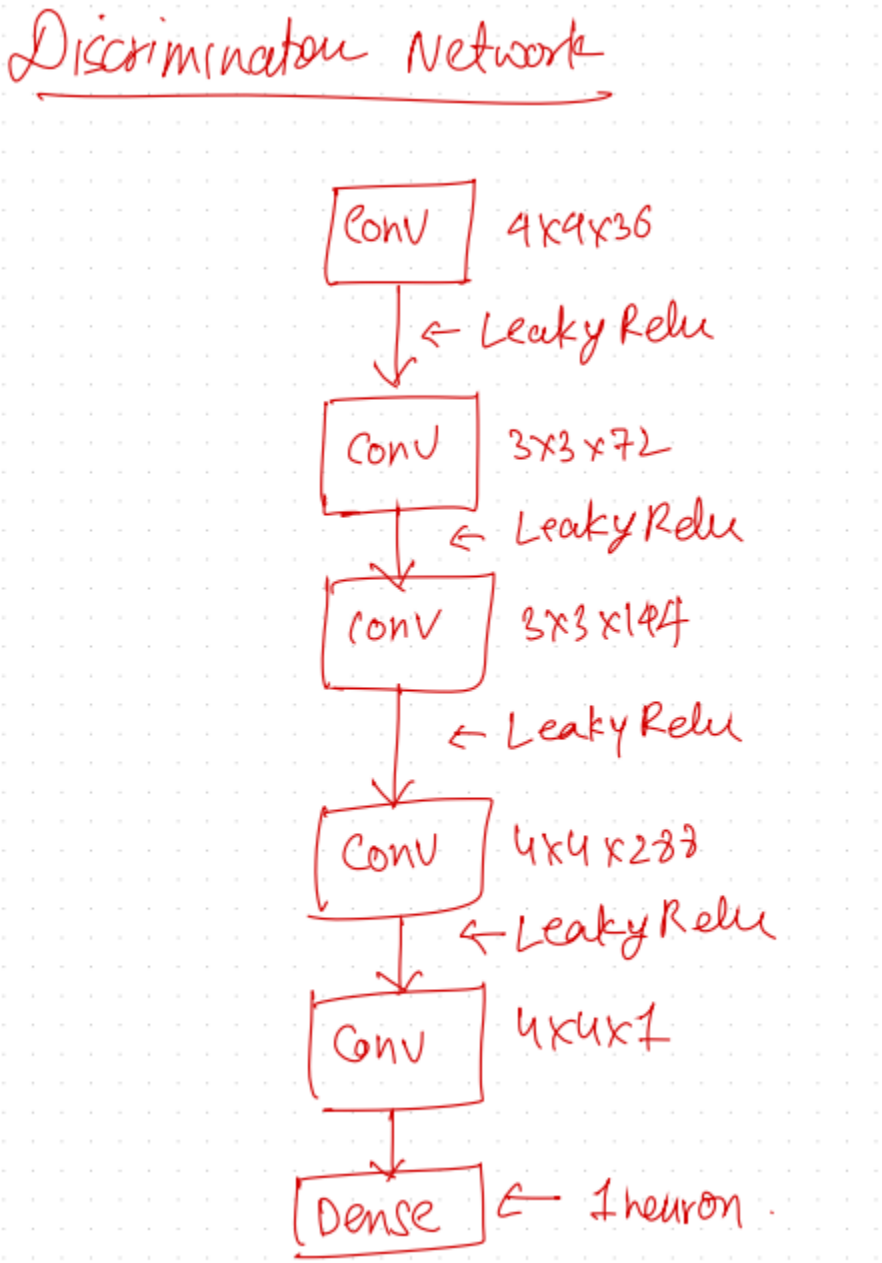


Fig. 5.—: Discriminator model common across all generators defined and across both approach

References

[1] Y. Chen, G. Li, C. jin, S. Liu and T. Li, “SSD-GAN: Measuring the Realness in the Spatial and Spectral Domains ,” AAAI 2021.