GAN Based State-of-Art Image Colorization

Gundla Kavya[1], Gudiboina Niharika[2], Paray Divya Sree[3], Dr. H. Balaji

[1, 2, 3] *Student, Department of Computer Science, SNIST, Hyderabad*
[4]*Project Guide, Associate Professor, Department of Computer Science, SNIST, Hyderabad*

Abstract: This project helps you to colorize the black and white pictures and it's done by using GANs(Generative Adversarial Networks) .GANs generally consists of a generator and a discriminator or critic which are competitive with each other. Force of GAN brings tones-generator applies tones to the perceived objects trained on, and discriminator attempts to scrutinize the color choice. They can reduce and replace the loss function with the help of the network and solves the problem of realistic colorization .It is further improvised and utilizes another sort of GAN training technique called NoGAN which is created to solve fundamental issues that seemed while trained using normal Generative Adversarial Networks composed of discriminator and a generator. The dataset used is ImageNet and the model depends on fast.ai library and also the series of evolution of improving GANs utilisation is explained.
Keywords: GAN(Generative Adversarial Networks),Generator, Discriminator, U-Net, No-GAN.

I. INTRODUCTION

For the real-world use of neural networks in our lives, deep learning models have an outstanding display. As mentioned, GANs can productively reduce and replace the hand coded loss function with network-generator/discriminator which learns it well by training and training. The generator and discriminator are trained separately for better results but let's just know more about the GANs.

A. Generative Adversarial Network

In simple words, GANs are of two models/networks-the generator and discriminator where the generator tries to improve itself by generating results that can fool the critic by convincing it is real and not generated and discriminator tries hard not to get fooled and by this adversarial dynamics the results get better.

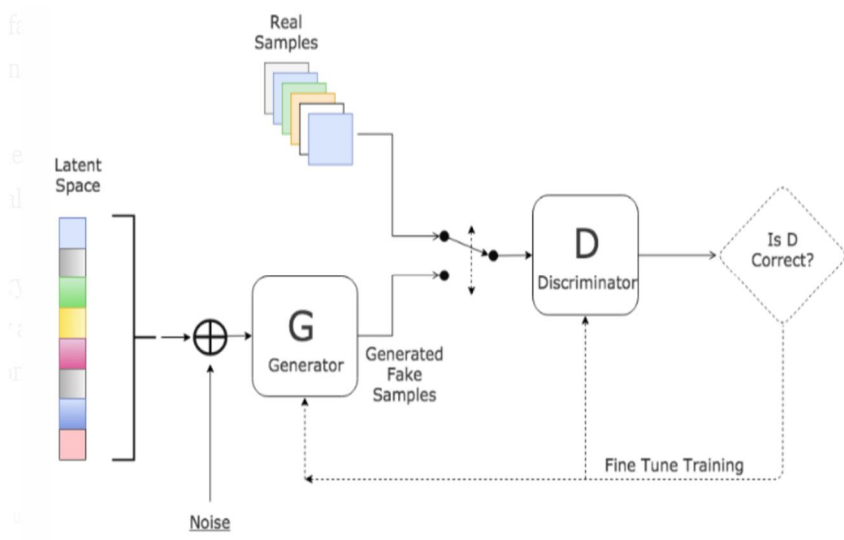The general GAN architecture would be like:

Figure 1: Generative Adversarial Network Architecture

Generally, the input to the generator model would be noise which is taken from the latent space and generate fake samples. It may fail first but it will eventually improves to generate indistinguishable images.
In this project, the two network trainings are done separately so it will be more effective unlike combining and training them. It mainly checks out two distinct parts, one is that the generated images have the grayscale features or not and the other is the generated image looks real or not which can also be defined as critic loss.

II. LITERATURE SURVEY

Emil Wallner [1] developed neural networks of three versions alpha, beta and full for colorizing black and white images but the input images should be only 400x400 for alpha and 256x256 for remaining versions. Olaf Ronneberger [2] used U-net for Biomedical image segmentation, but this mechanism takes more time for training. Kaiming He [3] implemented deep residual learning for image recognition. Alec Radford [4] from this paper we understand the implementation of DCGANs(Deep Convolutional Generative Adversarial Networks) and their consequences. Martin Heusel [5] in this paper GANs are trained by a Two Time-Scale update rule,but the output images are not upto the mark and consists of glitches. Tero Karras [6] ,in this paper the key idea is to grow both the generator and discriminator progressively used this approach for better quality of images. Han Zhang [7] ,in this paper the authors used Self-Attention Generative Adversarial Networks(SAGANs).

III. PROPOSED WORK

The improvised or new model is called "NoGAN" which spends less time on the training on GAN architecture but also gives the same benefits like the usual GAN training. Using a normal loss function, the generator was pre-trained to keep it efficient, fast, and dependable. Glitches and artifacts are almost entirely eliminated and most highly detailed & photorealistic renders are achieved.

IV. METHODOLOGY

The deep learning techniques involved to train the generator and discriminator models are the convolutional neural network architectures which can classify and identify the data.

A. Generator

Generally, the generator is a common U-Net [2], to be specific a U-net is an architecture which   basically divided into two parts where one parts recognizes the data or images and the other part outputs an image according to the recognized features . In this case, the U-Net used is Resnet34 [3] which is a pretrained model and also trained on ImageNet dataset. From the start, this model gets the capacity to recognize things in photos, and the second part of the model tests out what the backbone identifies and then calculates what color to use based on that, effectively resulting in a color image.
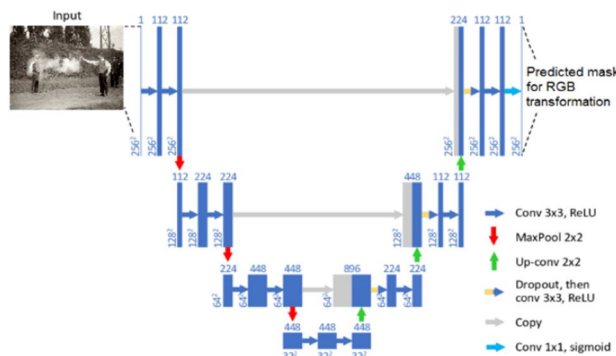
The U-Net looks like this:

Figure 2: A U-Net Architecture

B. Discriminator

The network used for discriminator is simple convolutional network from DC-GAN [4] but some modifications are done like the batchnorm is removed and the output layer is convolutional instead of a linear layer ,it looks big but simple and model learns by taking input images and assigns a single value like 0 or 1 to them for how they realistic they look.

There are some special techniques that are added to the models. Some of them are adapted from the Self-Attention GANs [7] where a new attention layer is added (which is proposed by author) in both the critic and the generator and also added special normalization. For discriminator vs generator, Hinge loss and various learning levels from Two Time-Scale Update Rule [5] are utilized. These modifications improved the training more stable, and the attention layers made a significant impact in color consistency and overall quality.

C. NoGAN

The improvised or new method is called as "NoGAN" which spends less time for the training of GAN architecture but also gives the same benefits as the usual GAN training. Here, the generator is pre-trained to make it powerful, fast and reliable using a regular loss function.

The architecture is same as the previous GAN like the generator is trained by using a regular deep neural network's architecture such as ResNet so the model is already prepared and good at colorizing an image even before the whole architecture is trained. Then, only a short amount of time will be required to optimize or improvise the realism of the generated images by this typical generator-discriminator GAN training.

One more key feature about this new GAN is that we can repeat the pretraining of discriminator on generated images even after the initial training and then continuing the training and by this we can achieve colorful results.

Same as in the previous model, Self-attention GANs are used in NoGAN but with a slight modification of having spectral normalization and also the two time-scale update rule also included but there is no longer usage of progressive growing of GANs because the results are better with the newly introduced model .

The loss function and accuracy of the model:
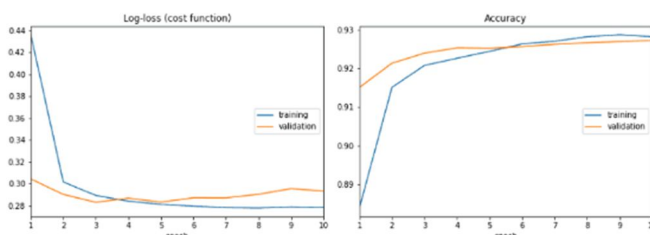
Further improvisations to the model are that gaussian noise is also randomly added to images so fake noise is generated during training. This can also be a type of data augmentation that can be performed on the training images which is to improve the results and the resistance to the noisy inputs which in turn same as the style transfer technique, where the noise added would be the style of the image and can be applied less or more to the transformation.

D. Other Technical Strategies

1) SA-GANs

Most of the GAN models for generation of images are constructed by using convolutional layers. But these layers process the information from the local data so by using these layers alone is computationally inefficient for training long-range dependencies in images. These self attention shows a better balance between the ability to train long range dependencies and computational efficiency .In simple words , self attention layers determines the amount of exposure of the values to the next layer. Self attention mechanisms are added to U-Net block.

2) Two Time-Scale Update Rule

The authors proved that GANs trained with two time-scale update rule converge to a stable level because in general implementations of GANs ,the discriminator is seemed to learn faster than generator even though they are trained on same level of learning rate.

So there will be no guarantee that generative model converge since the discriminator network serves as an objective to the generator thus these update rules are used for this particular reason.

And also one more technique involved is progressive growing of GANs, the training regime is modified by initially starting at 64x64 images, and to progress from there gradually – 96x96, 128x128, 196x196, and 256x256.This is because of the real problems that occurred when tried to train the model on just one resolution-the colorization of the photos had glitches. So progressive sizing training regime is introduced which is inspired by the progressive growing of GANs.

3) Progressive Growing of GANs

Higher resolution images makes it easier for the discriminator to tell the generated images apart from training images so amplifying the gradient problem and because of this the generation of high resolution images are difficult.

A solution was proposed by the authors [6] for training GANs for improvised quality, stability & variation. The main objective is to grow both critic & generator model's layers progressively which is by starting from low resolutions and then adding new layers to the model as the training progresses.

The results of the first stage of improvisation of GANs:

Figure 3: colorized image of Paris in 1880s

V.   PERFORMANCE AND EVALUATION

The results after latest improvisation of GANs,

Compare these pictures with the above pictures, we can notice that a perfect blend of colorization is applied to the latter improvisation of the model , many of the errors are reduced.

VI.   CONCLUSION

The previous model of GAN training is improvised and a lost of artifacts are removed by the new technique or a new GAN which is called as "NoGAN". This new type of training also combines benefits from usual GAN while removing side effects. The colorization looks more realistic in the new type of GAN training than the previous model ,its mainly because less time is spent on the direct GAN training and by using fast & reliable conventional methods . Also during only less amount of GAN training , the generator not only get the realistic colorization capabilities but it also results in less glitches and errors in the output.

VII.   FUTURE WORK

It seems like only a few modifications are done to the previous model but the results got better eventually. This improvisation can be applied to colorize photos and videos too. The future work may concentrate more on restoring old photos and videos and make this project more useful to the people.

VIII. ACKNOWLEDGMENT

REFERENCES

[1] Emil Wallner, Colorizing B&W Photos with Neural Networks
[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation
[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun , Deep Residual Learning for Image Recognition
[4] Alec Radford & Luke Metz, Soumith Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks
[5] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, Sepp Hochreiter , GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium
[6] Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen , Progressive Growing of GANs for Improved Quality, Stability, and Variation
[7] Han Zhang, Ian Goodfellow, Dimitris Metaxas, Augustus Odena , Self-Attention Generative Adversarial Networks

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ⊙ (24*7 Support on Whatsapp)