‘

Indian Institute of Information Technology Vadodara (Gandhinagar Campus)

Design Project Report-2021

on

Bringing Old Photos Back to Life

Submitted by

Chinigepalli Chakradhar Srinivas - 201951048
Malasani Raykanth Reddy - 201951088
Mudadla Siva Sai - 201951096
Penubala Deeraj - 201951113

under the supervision of

(Name of the Supervisor and Signature)
Dr. Ashish Phophalia

16/11/21

Abstract- The advance of computational power and big datasets brings the opportunity of using deep learning methods to do image processing. Generative adversarial network (GAN) has become a popular tool for image generation and restoration. GAN allows one to do things efficiently with less time consumption. Generally, their excellent performance is imputed to their ability to learn realistic image priors from a large number of example images. In this report, we show that, on the contrary, the structure of a GAN is sufficient to capture many low-level image statistics prior to any learning. In order to do so, we show a neural network that can be used as a handcrafted prior with excellent results in standard inverse problems such as denoising, superresolution, and inpainting. Furthermore, the same prior can be used to restore damaged parts of images. Apart from its diverse applications, our approach highlights the inductive bias captured by standard generator network architectures.

I. Introduction

We see many beautiful sceneries in our lifetime and we also spend a lot of cherishing and memorable events in our life. Whatever we see in our entire life with our own eyes is captured in the world’s largest memory cell that we often call the Human Brain. As said above there will be many things that we want to remember but we cannot. When we humans are aging as years are passing, our ultimate hard disk will slowly deteriorate, and we fail to remember many important things in life. But many decades ago, technology invented something to capture our important moments in life, which is indeed called the camera.

We humans have many important things in our life like marriage, a vacation trip with family or friends, gathering with friends after a long time, the first smile of your child when he/she is born, and many more. It would be always nice if we have them in our hands to remember our old memories back. His thought is also great, Camera is the one that makes this possible by capturing events. People are very much interested in this gadget and started capturing many events of life. Nowadays we save hundreds or thousands of images every year. We wish to see them and recollect those precious times even though it was taken a long time ago.

No one knows what Tajmahal will look like, unless either his image is captured. Even though books tell us many things but those are not in the form of pictures they are in form of words or text. It would be nice to have art drawn by the artist at their times that gives life to our imagination. By seeing that art, we could be able to imagine the living style or the habitat during those periods. Not only scientists, their inventions, and our ancestors, so we should preserve them in order to maintain the heritage of our nation or family.

Now imagine what happens when these images that are most beloved to us are damaged by any means like scratches in the image, partial images were torn, some regions got faded, and so on. How many times we would have thought if there is anything that could repair my damaged image which makes our day beautiful. Not only old images there is something beyond what this can do.

We have a good solution for all those different types of scenarios we discussed previously, and that is Image Inpainting and virtual enhancement. Let us now give an explanation of what image inpainting and virtual enhancement are. The art or process of inpainting involves filling in missing information in an image or removing damaged or unwanted parts of an image by using realistic techniques, showing off the image's color, and enhancing it to remove any dark areas and make it appear more vivid and clear.

We propose using the generative adversarial network (GAN) for image inpainting. It produces competitive results and can in some cases outperform them. Moreover, we were also able to analyze the performance of our model on different tasks and different datasets. MIRNet is the model used for the enhancement of images. It calculates a complementary set of features across multiple spatial scales while maintaining the original high-resolution features to preserve precise spatial details.

II. LITERATURE SURVEY

The article that was published on this domain stated that for restoration of a degraded image has a solution called Image Inpainting. In the process of its implementation, Inpainting uses the latest deep learning architectural methods. for inpainting the images using deep learning they used GAN methodology. GAN is a machine learning (ML) model in which two neural networks compete with each other to become more accurate in their predictions.GAN's generally have two major components. They stated that the actual reason for using GAN methodology is that it generates data that is similar to real data not only that but also GAN models are Faster than CNN models and also they do not need more pre-processing. But this model uses more time and space complexity than other models like CNN, RNN, and the GAN model is difficult to train and is quite unstable compared with other models.
Not only GAN methodology there are other Inpainting algorithms that yield better results with realistic outputs such as Context Encoder algorithm, Pluralistic Image Completion and etc…

1. **Context Encoder algorithm:** It inpaints based on context-based pixel prediction. By analogy with auto-encoders, we propose Context Encoders. When training context encoders, we have experimented with both a standard pixel-wise reconstruction loss, as well as a reconstruction plus an adversarial loss. The latter produces much sharper results because it can better handle multiple modes in the output. the complexity of training will be less, and the training will be performed quickly [1].
2. **Pluralistic Image completion** is one of the recent good approaches among the other image inpainting techniques which was released in recent years. In most image inpainting solutions, the algorithm produces only one result for the single mask image but through this approach, it generates multiple and diverse plausible solutions for image completion. it uses the framework with two parallel pipelines working simultaneously. This will be significantly less accurate than the inference models that we discussed above. However, it is computing through pipeline technique that completes quickly [2].

Not only these models there are other models too for Inpainting. Among the above-discussed models, the Context Encoder algorithm and Pluralistic Image completion algorithm give us an accurate image with utmost precision whereas the GAN produces images that are also accurate. This is due to most Context Encoder algorithms and Pluralistic Image completion algorithms being trained with synthetic datasets. When we have accuracy as a concern, the Context Encoder algorithm and Pluralistic Image completion algorithm are faster than GAN models in computation.

III. THE PRESENT INVESTIGATION

The restoration of old photos is more challenging than conventional image restoration. Unstructured defects such as film noise, blurriness and color fading, etc. can be restored with spatially homogeneous filters by making use of surrounding pixels within the local patch structured defects such as scratches and blotches, on the other hand, should be inpainted by considering the global context to ensure the structural consistency. The following sections propose solutions to the aforementioned generalization and mixed degradation issues.

The problem of restoration is broken down into two subroutine problems, Primarily the key part in restoring the damaged parts that are done by inpainting the damaged part using a generative adversarial network (GAN). Secondly Enhancing the dark parts of the image one should be able to clearly see the facial components by using the MIRNet model.

Generative adversarial network

The high-level architecture of the deep learning model is to use a generative adversarial network (GAN) proposed by Goodfellow et al[3]. Generative adversarial networks (GANs) are an interesting recent innovation in machine learning. GANs are generative models that create new data instances that resemble your training data. For example, GANs can create images that look like photographs of human faces, even though the faces don't belong to any real person.

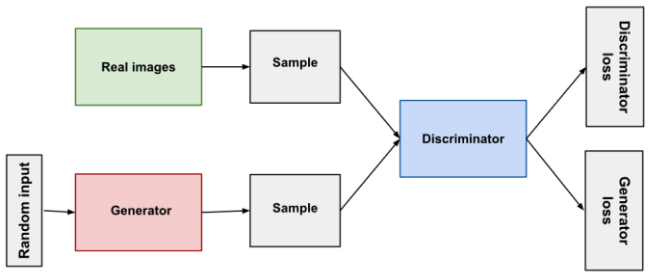

Figure 1: **GAN Architecture**

A generative adversarial network (GAN) consists of primarily two components

- The generator learns how to generate plausible data.
- The discriminator learns how to distinguish the generator's fake data from real data.

Both the generator and the discriminator are neural networks. The generator output is connected directly to the discriminator input. Through backpropagation, the discriminator's classification provides a signal that the generator uses to update its weights. Let's know more about the generator.

Generator

The generator part of a GAN learns to create fake data by incorporating feedback from the discriminator. It learns to make the discriminator classify its output as real.

Generator training is a tough task and is dependent on many other factors, the training of this component is the most important one as the image produced by this needs to be treated as real which is a very tough challenge, The portion of the GAN that trains the generator includes:

- Random input
- Discriminator output
- Generator loss

As said earlier the training of the Generator is the toughest part of the whole model as it introduces many new challenges. To produce an output that satisfies the Discriminator, the generator must be trained using various attributes which results in a realistic trained model and can generate outputs accurately.

So the procedure followed by us in training our model is the standard procedure mentioned in the origin research papers, which follows the below steps

- Sample random noise.
- Produce generator output from sampled random noise.
- Get discriminator "Real" or "Fake" classification for generator output.
- Calculate loss from discriminator classification.
- Backpropagate through both the discriminator and generator to obtain gradients.
- Use gradients to change only the generator weights.

Discriminator

The discriminator in a GAN is simply a classifier. It tries to distinguish real data from the data created by the generator. It could use any network architecture appropriate to the type of data it's classifying.

Now coming to the training part of Discriminator as the discriminator takes two inputs which act as training data to the Discriminator, They are

- Real data instances, such as real pictures of people. The discriminator uses these instances as positive examples during training.
- Fake data instances created by the generator. The discriminator uses these instances as negative examples during training.

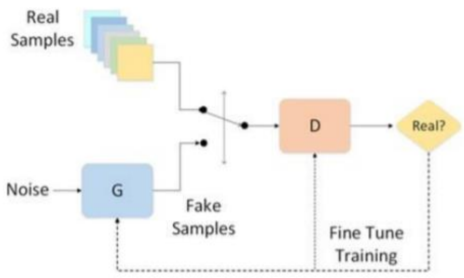

Figure 2: **GAN Framework**

Training of GAN

As GAN contains two separately trained networks, They have two different training processes, and thus to maintain equality the networks run in alternating periods i.e the discriminator trains for 1-2 epochs and now generator trains for 1-2 epochs and this cycle goes on continuing.

Loss function

In any model loss function plays a significant role as bearing minimum loss is one of the primary goals of any model and the loss functions discussed here are MINIMAX LOSS which is used in the paper that introduced GANs.

The discriminator seeks to maximize the average of the log probability of real images and the log of the inverse probability for fake images.

Discriminator: maximize $\log D(x) + \log(1 - D(G(z)))$ (1)

The generator seeks to minimize the log of the inverse probability predicted by the discriminator for fake images. This has the effect of encouraging the generator to generate samples that have a low probability of being fake.

Generator: minimize $\log(1 - D(G(z)))$ (2)

Where

- $D(x)$ is the discriminator's estimate of the probability that real data instance x is real.
- $G(z)$ is the generator's output when given noise z.
- $D(G(z))$ is the discriminator's estimate of the probability that a fake instance is real.

In this model to reduce the loss, the loss value is calculated for both the generator and discriminator model for each epoch and then the model adjusts the weight based on the backpropagated error value

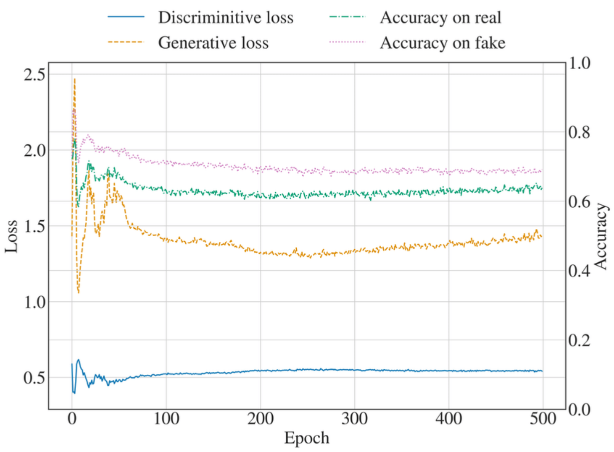

Figure 3: **Loss and Accuracy curve for GAN model**[4]

The above graph is between no of epochs and loss of both discriminator and generator. One can notice that the Discriminator works with very less loss when compared to Generator due to its functionality and upon increasing epochs, one can notice the loss is decreasing and becoming constant

Now if we take attributes of our project into consideration, there are some that need to be specified explicitly,

Data set and evaluation measurements

The dataset used by us is “Paris Street View”. The Paris Street View dataset was created by the authors of “Context Encoders”.The Paris Street View dataset, consisting of 14900 training images. In order to determine the relationship between epochs and image clarity, we have taken three sets of three images each of which is running at various levels of epochs.

The image below provides a comparative analysis output of 50 and 200 epochs. They yield low-level images due to high Losses at 50 epochs, As the number of epochs is very less the model is not well trained and unable to yield accurate results. Now the right set of images are the ones that are obtained after the completion of 200 epochs due to which the images are recognizable and the machine is also well trained as it indicates the image with more accuracy than before.

Figure 4: **Resolution of Images at various epochs.** The figure shows how the output differs at different levels of the epoch.

This set of images is something which is not regular as they are not possible in our machines, these images are a result of 5000 epochs which seems to be very natural and one cannot differentiate them.

Figure 5: **Images at 5000 epochs.** This figure shows the naturality in output at 5000 epochs simulated by Research Institute[5]

To enhance the image virtually so that the darkened part of the image becomes vivid and Oftentimes, cameras generate images that are less vivid and lack contrast. A number of factors contribute to the low quality of images, including unsuitable lighting conditions and physical limitations of camera devices. For image enhancement, histogram equalization is the most commonly used approach. However, it frequently produces under- or over-enhanced images.

MIRNet

MIRNet model is the one which is used extensively in current days to enhance the image quality and to bring up low light images, the characteristics of the MIRNet model are

- A feature extraction model that computes a complementary set of features across multiple spatial scales, while maintaining the original high-resolution features to preserve precise spatial details.
- A regularly repeated mechanism for information exchange, where the features across multi-resolution branches are progressively fused for improved representation learning.
- A new approach to fuse multi-scale features using a selective kernel network that dynamically combines variable receptive fields and faithfully preserves the original feature information at each spatial resolution.
- A recursive residual design that progressively breaks down the input signal to simplify the overall learning process, and allows the construction of very deep networks.[6]

Training & loss function

We trained the MIRNet model using Charbonnier Loss as the loss function and Adam Optimizer.

The below graph represents the train and validation losses vs no of Epochs for our MIRNet model

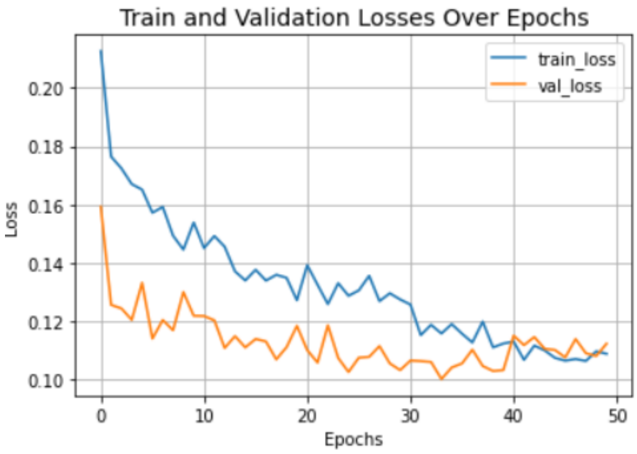

Figure 6: **Train and validation Losses Over Epochs** [7]

The above graph yields some predictable results which are the same with all models that we one can observe from the graph that the losses of both test and train are decreasing iteratively with the increase in no epochs as the machine gets more and more trained.

IV. Results and Discussions

The main objective of our project is to make old photos look a better way, which was done by repairing damaged parts using GAN, and to remove the aging effects, to enhance the image, to give it more colors we had opted for the use of most popular model MIRNet model which often deals with the enhancement of low light images, This can be seen through the above discussion on MIRNet.

The output results shown below are from our 200th epoch. In order to attain or visualize better results increase no of epochs which makes the model trained better and offers better results,

Figure 7: **Image after restoration**

This can be observed from the above images(for 50,200,5000 epochs) which show the results. But finally, as this is also a method that has some limitations Primarily this model consumes a huge amount of time, For Paris Street View dataset, consisting of 14900 training images, it took around 6 hours of training time in Google colab which has Nvidia K80 / T4 with 16GB Ram for 200 epochs. The execution time of the test image for inference is approx 5 seconds.

Last but not least, our findings contradict the narrative that deep learning performs better in image restoration because it can learn rather than construct priors; random networks are better at constructing priors, and learning builds on them. Furthermore, this validates the need to develop new deep learning architectures.

V. Conclusions and Future Work

Although there have been some image restoration procedures proposed in the literature, this problem is far from being solved satisfactorily. Image inpainting is an important task for computer vision applications, due to large modified data using image editing tools. From these applications, we can find image quality enhancement, image restoration,, and others. A brief image inpainting review is performed through the GAN model. GAN model is proposed and it is applied to inpaint the missing regions. The enhancement is also applied to the image, which in turn results in a high-resolution image. Our project demonstrates good performance in restoring degraded old photos.

The proposed Inpainting may be applied for removing the unwanted object from an image without affecting the background of an image. Extending the project towards the removal of blurring on the borders of coherent inpainted regions through adding extra functions or methods. In conclusion, there is no method that can inpaint all the types of distortion in images, but using learning techniques there are some promising results for each category of analyzed cases.

Acknowledgment

We would like to thank Dr. Ashish Phophalia sir for his support and for his constructive comments.

References

[1] "Context Encoders: Feature Learning by Inpainting." n.d. Accessed November 16,2021. https://www.cs.cmu.edu/~dpathak/context_encoder/.

[2] Zheng, Chuanxia, Tat-Jen Cham, and Jianfei Cai. 2019. "Pluralistic Image Completion." http://arxiv.org/abs/1903.04227.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Advances in neural information processing systems, pages 2672–2680, 2014.

[4] https://www.researchgate.net/figure/Plot-of-the-discriminator-and-generator-loss-and-accuracy-as-a-function-of-epochs-Early_fig1_349727792.

[5] Wan, Ziyu, Bo Zhang, Dongdong Chen, and Fang Wen. 2020. "Bringing Old Photos Back to Life." 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June, 2744–54.

[6] Keras Team. n.d. "Low-Light Image Enhancement Using MIRNet." Accessed November 16, 2021. https://keras.io/examples/vision/mirnet/.

[7] McGinn, J., C. Messenger, I. S. Heng, and Michael Joshua Williams. 2021. "Figure 2: Plot of the Discriminator and Generator Loss and Accuracy As." March 2, 2021.