International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

Realistic Image Synthesis using GAN

Shreeraj Suryavanshi[1], Rugveda Parab[2], Radhika Machale[3], Urvashi Kohale[4], Mrs. S. P. Khedkar[5]

[1,2,3,4] *Final Year (B.E.) Students,*
M.E.S. College of Engineering, Pune
[5] *Assistant Professor, Dept. of Computer Engineering, M.E.S. College of Engineering, Pune*

---***---

Abstract – *In today's day an age, artificial intelligence is at it's prime. Many companies are exploring different applications of AI. In recent years, a part of AI called GAN (Generative Adversarial Network) has gained significant traction. GAN is a generative model because it learns to copy data distribution of the data you gave it and generate novel outputs that look alike. The word Adversarial in GAN is in consideration with two different networks who are competing with each other, that's it tries to outwit each other. GANs is often linked to the analogy of a policeman(discriminator) and a counterfeiter (Generator).*

GAN was discovered by Ian Goodfellow [1]. GAN is generally used to generate output images or process them. The generative network generates candidates while the discriminative network evaluates them. GAN learns to map from a latent space to a data distribution of interest and distinguish the generated candidates from the true data distribution. It requires a big dataset for the training of generator and discriminator to produce acceptable result [6].

In this paper we will take a look at different components to build and train a GAN network which would output a realistic image from the input labelled map of the scenery. The input will consist of labelled maps of objects like mountains, clouds, grass, wall, buildings, windows, etc [5]. The goal here for the GAN is to generate realistic image fitting the boundaries of the labelled maps.

***Key Words*: Generator, Discriminator, Neural Network, Label map, Image processing**

1. INTRODUCTION

The task of generating realistic images from label maps is relatively new in the field of computer vision. The task consists of converting a basic colour coded image map to a photorealistic version of the same image following the same map boundaries. The image map is used as a reference to the boundaries of the objects in the realistic image.

Fig 1: Image synthesis

Synthesizing realistic images like this can become very useful in various fields. Artists can use it to create background settings, while graphic designers can dynamically create the image, they want without doing it themselves. It can also be used in photoshops of image.

However, it is not an easy task to generate a good enough photorealistic image based on an image map. The shape of a object is provided by the image map, poses the challenge of recognizing the scale of things and how the object texture would fit into it while keeping the overall image believable.

For determining the appropriate texture and scaling the texture requires the huge number of training dataset. Also mapping the effects of on object on other need to be known to gradually transition over certain objects.

1.1 Generative Adversarial Networks

GANs are called as generative adversarial networks. GANs consists of two neural networks namely generator and discriminator. These two networks are made to compete against each other. The generator tries to imitate the required result and the discriminator verifies the validity of the generated results.

What makes the GAN model different from other models is the ability of learning to create the input with certain variations. GANs make it possible to create a novel output with sufficient accuracy. The discriminator network makes sure that the generator output is tailored to the input requirements also if it is up to the mark by providing the generator with pointers to improve upon.

The generator network is a deep convolutional network, it is provided with the pixel values of the input image. The generator is encoded with all the necessary parameters. Using these parameters, the generator creates an image satisfying the constraints. The loss to the generator is provided by the discriminator which fine tunes the parameters of the generator.

The discriminator network is a simple classifier network. It is deconvolutional network. The task of discriminator is to classify whether the generator output is valid or not. The discriminator calculates a loss value of the obtained image and expected image and provide it to the generator.

The generator G and discriminator D in the generative adversarial networks compete against each other in a game of minimizing and maximizing the loss function. The discriminator D tries to differentiate between the synthesized image and actual image. While the generator tries to fool the discriminator by synthesizing an image as close to the actual image as possible.

If the loss function is denoted by L and generator, discriminator as G and D respectively the loss function can be defined as,

$$[1]\ \mathrm{Min}^{G}\mathrm{Max}^{D} L(D, G) = E_x{\sim}p_{data}(x)[\log D(x)] + E_x{\sim}p_z(z)[\log(1-D(G(Z)))]$$

Ian Goodfellow in 2004 showed that this loss function has a global optimum when the p_g approaches p_{data}.

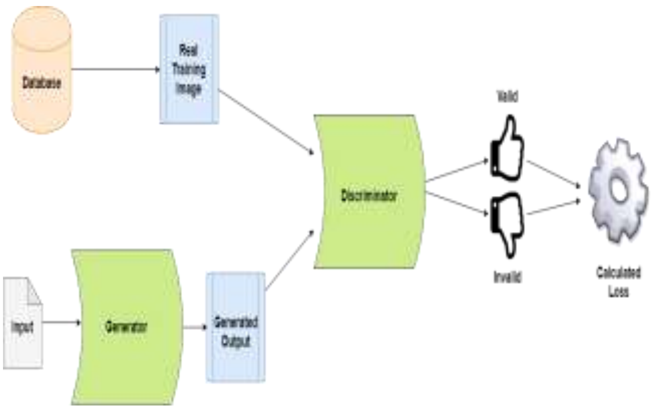

Fig 2: GAN Architecture

1.2 Literature Review

Related works done on this topic is building GAN networks to generate images[1], creating a stable architecture for Training GANs [6] for generation of high resolution images, building text to image generators using GAN[7].Ian Goodfellow's concept of pitting 2 networks against each other leaves the process of feature extraction and learning them, finding patterns in images to the discriminator and proved the viability of the adversarial network. Various endeavors have been made in GAN to improve the overall quality of GANs for image synthesis like large scale GAN training for high fidelity natural image synthesis[7] and photographic image synthesis with cascaded refinement networks[4].To get better results many different approaches are applied Conditional GANs[3] and DC-GANs.

The main goal is to develop a stable and reliable framework of GAN network which can produce consistent high-quality images. This task can extend its applications in variety of fields like medical imaging and other computer vision problems.

2. PROBLEM STATEMENT

The idea behind this project is to build an application which generates realistic images based on the input labels. For the generation of images, a network called GAN Generative adversarial network is used.

In a GAN network there are 2 embedded networks one is generator network, which will learn to generate realistic images based on input label maps and another is discriminator network, which will classify whether

the generated image is realistic or not. Both of the networks compete with each other as generator generates more and more realistic images and discriminator becomes more and more effective in recognizing fake generated and realistic images. The point at which the discriminator can no longer discriminate between the real and the fake generated image we take that image as an effective output.

3. IMPLEMENTATION

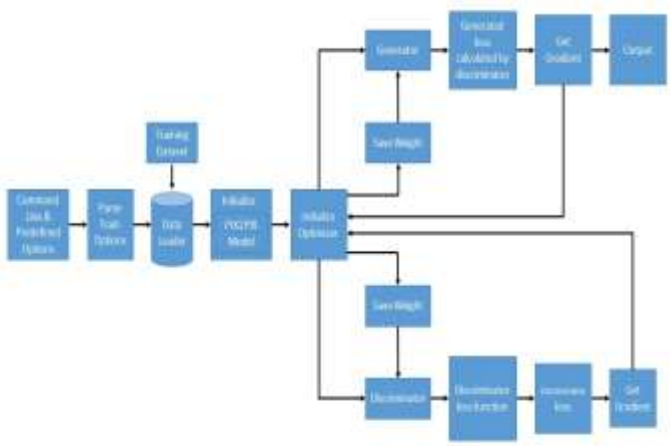

Fig 3: Training Architecture

A] Pix2pix model

To supervise the generator and discriminator and conditionalize the network for the task of realistic generation the pix2pix model used. This model provides a general-purpose implementation of conditional GAN which can be modified for the various purposes.

CNNs are common neural networks used in variety of image processing or prediction problems. In problems like this the CNNs provide a better approach in handling the complicated task of minimizing the loss function.

B] Pre-processing Images

The input segmentation map first needs to be scaled to a standard resolution, for consistency in training data. The Having a fixed number of pixels for the training image contributes towards stability. To avoid dealing with various aspect ratios of images the images must be cropped to a simple 1:1 aspect ratio after being scaled to the required size.

C] Generator Network

The input layer of the generator is a simple 2d Convolutional layer with the number of pixels as number of neurons in the layer. The hidden layers will consist of n numbers of downscaling convolutional layers. Then a resnet block [8] with skip connections and again an upscaling convolutional layer. The final output layer will also be a 2d-convolutional layer with same numbers of neurons as input layer. The activation function used will be tanH. Each layer will have an instance Normalization layer.

The generator will be optimized by the discriminator output and ground truth image using Adam optimizer.

D] Discriminator Network

The discriminator network will be a multiscale discriminator. Discriminator will discriminate between 3 different scaled versions of the generated and actual image. The discriminator network will consist of n layers of standard 2d-convolutional layers with LeakyRelu activation. The output layer will be a single out channel which will tell the given image is real or not. All the n different Discriminator networks will have different numbers of input channel based of the scale at which they are discriminating.

The discriminator network will be used for calculating loss for the generator as well as the discriminator itself. Using multiscale discriminators, the gradients of the lower scaled images can be brought closer to the actual images and slowly building upwards [2][4]. The generator will learn how to make good lower resolution images, and then use it to improve higher resolution images.

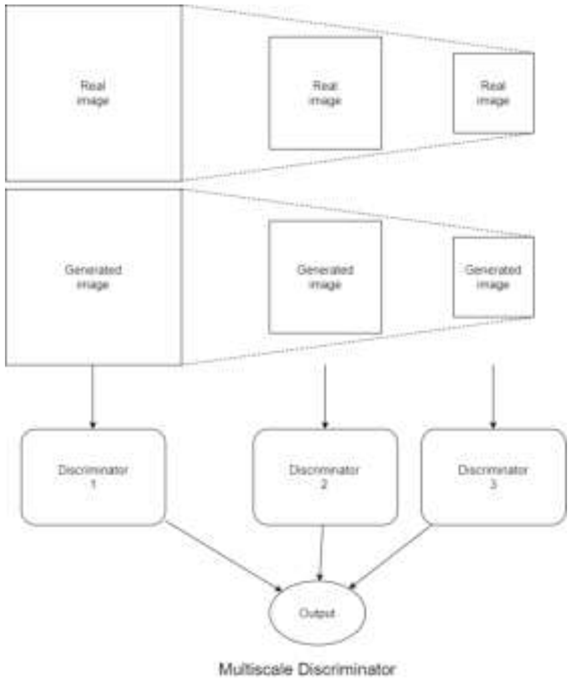

Fig 4: Multiscale Discriminator

E] Loss

The loss calculated will be MSE loss between the predicted and the target tensors. All the losses at different scales will be mapped to the generator network. The optimizer for generator just uses the loss to optimize the network to decrease the loss. While the optimizer for the discriminator will optimize the network so that the loss is increased.

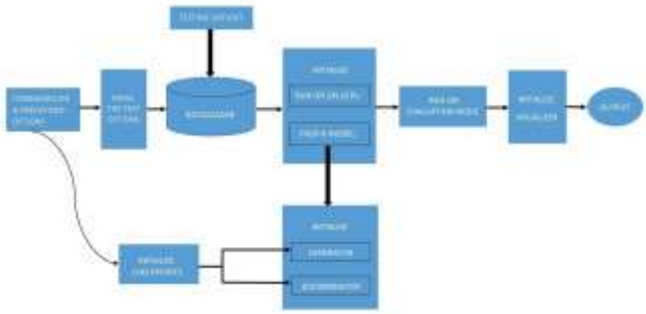

Fig 5: Testing Architecture

4. RESULTS

5. CONCLUSION

In conclusion, the proposed GAN model can generate acceptably realistic images from input semantic maps which are very close to the actual images. The model generates consistent good results with sufficient training. The model provides stable training and better results than most image generation models. This model can allow many applications in various fields and be used in domains where realistic image generation is needed, for example photoshop, graphic designing etc.

REFERENCES

[1] 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, 2014.

[2] T. C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional GANs. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image to image translation with conditional adversarial networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[4] Q. Chen and V. Koltun. Photographic image synthesis with cascaded refinement networks. In IEEE International Conference on Computer Vision (ICCV), 2017.

[5] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba. Scene parsing through ADE20K dataset. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[6] Andrew Brock, Jeff Donahue, Karen Simonyan. Large Scale GAN Training for High Fidelity Natural Image Synthesis. Published as a conference paper at ICLR 2019

[7] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran REEDSCOT1, AKATA2, XCYAN1 , LLAJAN1 Bernt Schiele, Honglak Lee. Generative Adversarial Text to Image Synthesis. arXiv cs.NE 2016

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition arXiv:1512.03385v1 [cs.CV] 10 Dec 2015