Image to Image Translation

Caroline Chandraguptharajah

***Abstract*—Image to Image translation is a class of vision learning problems, where the goal is to learn the mapping of images from one domain to another. This finds application in many areas like generating map from an aerial image of a place. In this project we use conditional adversarial networks to tackle this problem.**

***Index Terms*—ImageTranslation, Deep Learning, Conditional Adversarial Network**

I. Introduction

THE task of Image to Image Translation is a well researched topic. Most approaches require the design of a complex loss function that is specific to the application or image mapping the model is trying to learn. In our project we use the pix2pix model, which overcomes this issue, and learns a loss function that applies itself to a variety of image translation tasks. Through this we would no longer need to hand engineer complex loss functions.

Instead of engineering a loss function, we develop a neural network with the high level goal of generating images that are indistinguishable from the real ones. To this effect we have used the Conditional Generative adversarial neural network (CGANs). The Adversarial network learns a loss function based on being able to classify if the image was artificially generated or is a real image. This is done while the Generative network tries to generative images which are as close as possible to the real images. Since this loss function is data dependent, the same GAN architecture can be applied to a variety of image translation tasks.

The generator uses a "U-Net"-based architecture, and the discriminator uses convolutional "PatchGAN" classifier. The PatchGAN only penalizes structure at the scale of image patches.

II. Method

To understand the functioning of the network at a high level, consider the image shown in fig 1.

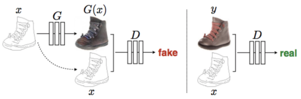

Fig. 1. High level functioning of network

The Conditional GAN network learns a mapping from input image x and a noise vector z to generate an output image y, G:(x,z) → y. The generator is trained to generate images that are indistinguishable from real images, while the Discriminator is trained to differentiate between real images and fake images generated by the generator.

A. The Loss function

The Conditional Adversarial loss function is given as below. The generator tries to maximize this loss function in adversary

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,z}[\log(1 - D(x,G(x,z))]$$

to the Discriminator which tries to minimize it. While an L1 loss function is given as below, Combining these two loss

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x,z)\|_1].$$

function, we get

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda\mathcal{L}_{L1}(G).$$

B. The Network Architecture

Both generator and discriminator use modules of the form convolution-BatchNorm-ReLu.

1) The Generator: The generator uses a U-Net architecture, compared to the encoder architecture used in other approaches. The U-Net architecture ensures that there is minimal loss in information, while encoding the input, which was present in other architectures and termed as the bottleneck issue. The U-Net overcomes this issue by using skip connection between layers. It is similar to the Res-Net Architecture in this aspect.

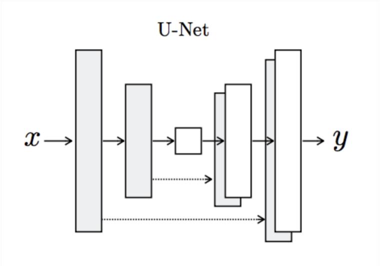

Fig. 2. U-Net Architecture

2) The Discriminator: The discriminator used is called a Markovian Discriminator or the PatchGAN Discriminator. This discriminator works by classifying individual (N x N) patches in the image as real vs. fake, opposed to classifying the entire image as real vs. fake. This enforces more constraints that encourage sharp high-frequency detail. Additionally, the PatchGAN has fewer parameters and runs faster than classifying the entire image

C. The Data set

The facades data set from Berkeley was used for training. Link. These image pairs are split vertically, and trained in a way such that given right image, we generate the left one.

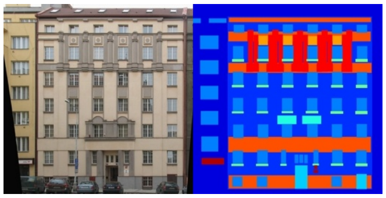

Fig. 3. Sample Image from data set

III. Results

A. Epoch 200

The loss values achieved after training the model for 200 Epoch are as below:
1. Discriminator loss : 0.810109
2. Generator loss GAN : 1.6185946
3. Generator loss L1 0.24270096

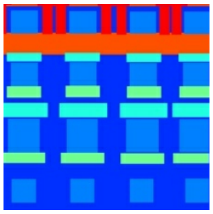

Fig. 4. Input Image

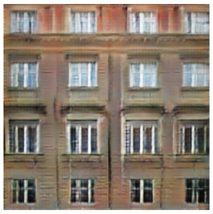

Fig. 5. Output Image

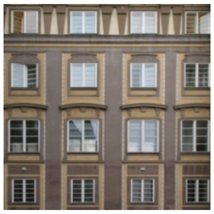

Fig. 6. Target Image

B. Epoch 300

The loss values achieved after training the model for 300 Epoch are as below:
1. Discriminator loss : 0.7933595
2. Generator loss GAN : 1.7311112
3. Generator loss L1 : 0.22838801

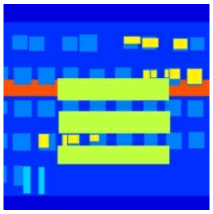

Fig. 7. Input Image

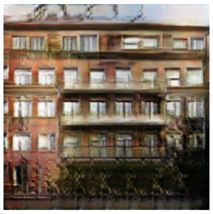

Fig. 8. Output Image

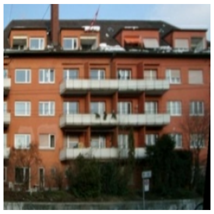

Fig. 9. Target Image

IV. Conclusion

The Conditional GAN architecture with L1 and adversarial loss, seems promising for the Image Translation task. The most advantageous point of this network is its ability to generalize to various image translation tasks with different data sets without requiring any manual engineering of loss function.

References

[1] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros . Image-to-Image Translation with Conditional Adversarial Networks. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1125-1134

[2] Yang, C., Kim, T., Wang, R., Peng, H. and Kuo, C. (2019). Show, Attend, and Translate: Unsupervised Image Translation With Self-Regularization and Attention. IEEE Transactions on Image Processing, 28(10), pp.4845-4856.