Image to Image Translation using C-GAN

Shubham Miglani Yati Katoch

I. Motivation

The problem that we are trying to investigate is finding a common framework for image to image translation. The baseline of the project comes from [1] in which image to image translation using conditional GANs is explored. The paper talks about problems in this area such as using special-purpose loss function formulation for each task. To tackle this problem, they introduce a framework using C-GAN. The advantage of using C-GAN is described in the paper as in comparison to GANs which learn a generative model of data, C-GAN learns a conditional generative model meaning that they condition to a specific input image and also learn a loss function to train this mapping. There have been many applications of GAN in image style transfer, super-resolution, etc. but they are all application-specific but the implementation in the paper is application-independent. We plan to implement the architecture used in [1] and exploring different architectures for generator including UNet, ResNet-6, ResNet-9, ResNet-50 and the Discriminator including ImageGAN, PatchGAN, ResNet and different applications as well.

II. Data

For this project, we plan to use the data provided by the pix2pix dataset [2] as the source of our data. The features of the data set include images to be used for Facades, Edges to Shoes, Cityscapes, Maps in Architecture to Labels, Edges to Photo, Day to Night, Aerial to Map applications.

Table 1
Dataset Description

Image Category	Training	Validation	Pixel Size
CityScapes	2975	500	256,256
Maps	1096	1098	600,600
Facades	400	100	256,256
Edges to Shoes	49825	200	256,256

We plan to work on the Facades database first as it is the least computationally intensive and then move on to other datasets. Also as noted in [1], even with few training examples, decent results can be obtained using C-GANs on this dataset.

III. Evaluation

The evaluation of C-GANS is a tricky task. We plan to do it by relating the output image to the ground truth by using the following criteria:

1) MSE

This will be defined as the pixel by pixel Mean Square Error between the generated images to the ground truth images. Although as mentioned in [1], this does not measure the very structure that structured losses aim to capture, it provides a good starting point.

2) Discriminator Score

The discriminator score can be used to determine how well the discriminator works to classify the generated images as fake and the ground truth image as real.

3) t-SNE

T-Distributed Stochastic Neighbor Embedding is a technique to visualize the distribution of the generated images. This can be used as a visualization tool to compare how different network architectures are performing for randomly generated images.

4) Visual Analysis

The generated images can be visually compared to explore how different architectures capture different features of the images.

IV. Methodology

The conditional GAN consists of two major parts generator and discriminator. The architecture used for the generator in [1] is U-Net which consists of 8 encoding and decoding layers having skip connections between layer i to layer n-i. We plan to implement a residual-based network consisting of 2 encoding and 2 decoding layers with 6 (ResNet-6) or 9(ResNet-9) residual blocks in between. For the discriminator, Patch GAN is used in [1] with a patch size of 70x70. We plan to experiment with different patch sizes according to which the discriminator architecture will change as well. Random jitter and random cropping were applied to the images in [1]. We plan to implement other data augmentation methods as well. For the activation function, Leaky Relu was used in the encoder and Relu for the decoder in [1]. We plan to experiment by changing the slope of the leaky Relu and with other activation functions such as mish, swish as well. Also, there was a recommendation to experiment with the training frequency of the generator and the discriminator which we are planning to explore as well.

V. Challenges

There are various challenges associated with this project. The first part of the project will be to utilize the baseline in [1] and to understand the source code to experiment with it to understand the effects of various parameters. The second part would be to add and implement different architectures to the existing baseline and compare the results using TensorFlow. Also, more evaluation criteria could be studied and used such as explored in [3].

VI. References

[1] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-Image translation with conditional adversarial networks, 2016.
[2]https://people.eecs.berkeley.edu/~tinghuiz/projects/pix2pix/datasets/
[3] Pros and Cons of GAN Evaluation Measures, Ali Borji