Machine Learning Nanodegree

Capstone Proposal

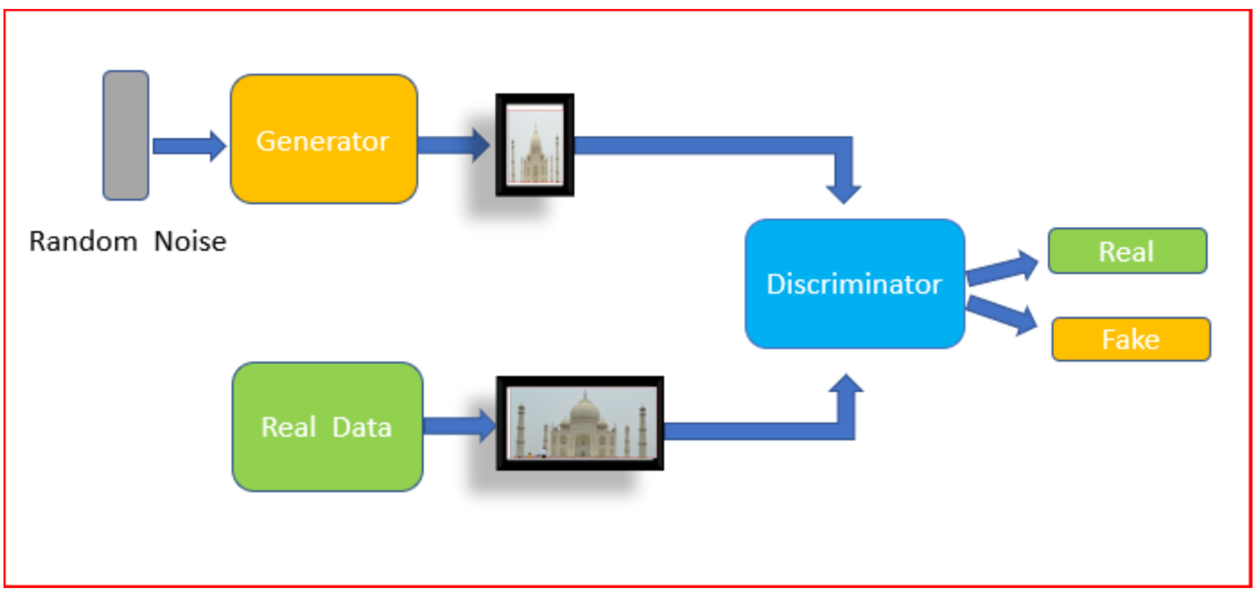
By Chingis Oinar

August 11th, 2021

Domain Background

Many problems in image processing and computer vision can be viewed as an image-to-image translation where input is translated from one possible representation into another. The community has made significant steps in this direction by utilizing Convolutional Neural Networks (CNNs) which subsequently became the common workhorse for numerous image prediction problems. CNNs learn to minimize a loss/objective function; however, there have been a lot of attempts of designing effective losses. Thus, CNNs still need a proper loss function depending on a given task. For example, if we use Euclidean distance to measure the difference between predicted and ground truth pixels, it most likely produces blurry results (Isola, 2017). However, coming up with an effective loss function that would encourage CNN to produce very realistic images is extremely difficult.

Fortunately, Generative Adversarial Networks (GANs) have recently achieved impressive results in the field. Specifically, by incorporating two separate networks, generator and discriminator, GAN learns a loss function due to which it is able to produce highly realistic images (Khandelwal, 2019).



The figure attached above demonstrates how GAN works. As seen, Generator's objective is to generate data that is indistinguishable from the real data, whereas the Discriminator takes both real and generated data and tries to classify them correctly. Therefore, blurry images are most likely to be classified as fake.

Furthermore, considering that GAN learns an objective that adapts to the training data, they have been applied to a wide variety of tasks.
Recently, GANs have been employed for (Brownlee, 2019):

- Image Restoration
- Text-to-Image Translation
- Image-to-Image Translation
- Face Frontal View Generation
- Cartoon Characters Generation
- Style Transfer
- Face Aging
- 3D Object Generation

As seen in the figure above, GANs have done staggering improvements in terms of image generation. Therefore, they have been receiving massive attention for a variety of tasks.

Problem Statement

In this capstone project, I would like to tackle a satellite to map image translation problem using GAN. For the chosen task, I will be using a Pix2Pix GAN that is proposed by Philip Isola, et al. in the paper titled "Image-to-Image Translation with Conditional Adversarial Networks" and presented at CVPR in 2017. The Pix2Pix model is a type of conditional GAN where the generator takes a source image as a conditional input. The discriminator is provided with both a source and the target/generated as input and predicts whether it's fake or real (Brownlee, 2021). Thus, inputs are aerial photos obtained from Google Maps and the goal is to convert them into user-friendly Google maps format.

Dataset and Inputs

The dataset is provided on the pix2pix website, hence it is publicly available and can be easily accessed. The train set contains 1,097 images of satellite images of New York and the corresponding Google maps pages, whereas the validation dataset had 1,099 images.

Each image is 1,200 pixels wide and 600 pixels tall. An example of the images in the dataset is attached below, where the target and the source images are shown side by side.

Solution Statement

The solution is predicted Google Maps-like images. As has been mentioned, I will use Pix2Pix GAN. Previous approaches have observed that it is beneficial to mix the GAN objective with a more conventional loss, such as L2 distance (Isola, 2017). However, the authors use L1 distance instead as they claim it produces less blurry images. Some of the other related works include Style Loss as well (Ganguli, 2019).

Benchmark Model

For this project, the benchmark model is Pix2Pix trained without any additional loss functions. I will try different traditional loss functions and see how they affect image quality.

Evaluation Metric

It is quite complicated to evaluate GANs comparing to other traditional networks. For evaluation, I can do cross-validation by utilizing different discriminators and share the classification accuracies of each one. Thus, the generator that is able to fool discriminators the most is most likely to be the best one. Additionally, I will inspect the generated images myself and report the results, including high and low-quality images.

Project Design

Before I start working on my project, I will study the dataset and try to provide some visualizations. Then, I will develop my model and set the hyperparameters as stated in the papers. I will run the experiments and perform my evaluation procedure. I plan to train 3-4 models using different loss functions and compare image qualities. Finally, I will report my observations as well as the necessary quantitative results.

I expect to spend 10% of the time on data preparation, 60% of the time on training all the models and the remaining 30% on evaluation. Finally, all the experiments will be done using PyTorch.

Reference

Brownlee, J. (2019, July 12). *18 impressive applications of Generative adversarial Networks (GANs)*. Machine Learning Mastery. https://machinelearningmastery.com/impressive-applications-of-generative-adversarial-networks/.

Brownlee, J. (2021, January 18). *How to develop a pix2pix gan for image-to-image translation*. Machine Learning Mastery. https://machinelearningmastery.com/how-to-develop-a-pix2pix-gan-for-image-to-image-translation/.

Ganguli, Swetava, Pedro Garzon, and Noa Glaser. "GeoGAN: A conditional GAN with reconstruction and style loss to generate standard layer of maps from satellite images." *arXiv preprint arXiv:1902.05611* (2019).

Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

Khandelwal, R. (2019, November 17). *Generative adversarial Network(GAN) using Keras*. Medium. https://medium.datadriveninvestor.com/generative-adversarial-network-gan-using-keras-ce1c05cfdfd3.