

Machine Learning Project Proposal Sat-to-Map



Shagun Uppal and Sarthak Bhagat

2016088 | 2018189

shagun16088@iitd.ac.in | sarthak16189@iitd.ac.in

Motivation and Problem Statement

Learning mappings from one image to another has been a problem of great interest since recent times in the field of image analysis, computer vision and learning models. With this project, we aim to develop an image-to-image learning model, which will be able to synthesis a given image representation to any other target scene representation.

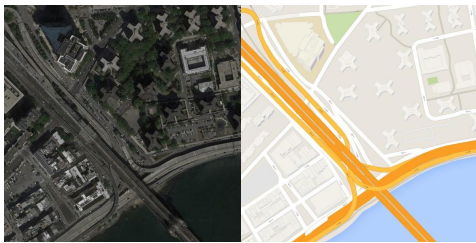
To address this problem, we are specifically focusing on generating maps from the corresponding satellite images for cities. Such kind of a problem seeks great attention as well as application in real life situations for the purpose of navigation and generating city maps. The data-set to be used for the same consists of satellite images of cities along with their corresponding generated maps as the ground truth.

Data-sets

Data Acquisition The data-set containing satellite image of cities obtained via web scraping from Google Maps was available publicly provided by UC Berkeley's Electrical and Computer Science Dept. The data is available as a concatenated form of satellite and map images. The details of data-set is as follows:

- Number images in data-set (train + validation): 2194
- Size of each combined image (the satellite and the map image) in data-set: 1200 x 600 so that means each image (map or satellite image) is of size 600 x 600.
- Size of complete data-set: 257.7 MB

Preprocessing Techniques The fact that the images are coupled together, requires us to separate the two images of each scene by simply cropping to retrieve the satellite image as our training datapoint and the corresponding map images becomes the ground truth data for the same, required for computing the loss between our predicted and the desired images.



Individual Contributions

Shagun : Data preprocessing
Training a GAN

Sarthak : Training a VAE
Training a CNN

References

- [1] Phillip Isola and Jun-Yan Zhu and Tinghui Zhou and Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks.
- [2] R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization.ECCV, 2016.

Model Description

Learning Techniques We intend to compare the performance on the task of map generation from a given satellite image on a trivial neural network as well as simple and compound generative models that include the following:

- **Baseline - Convolutional Neural Network (CNN)** : This is a simple neural network consisting of convolution layers. CNNs learn on the basis of learning the loss function to enhance the quality of the results, the loss function being manually defined as per the task taken into consideration.
- **Advanced Technique - Variational Autoencoder (VAE)** : This is an autoencoder consisting of an encoder and decoder and a gaussian prior imposed on the latent space. The training approach is based on reducing the reconstruction loss between the predicted and ground truth images.
- **Advanced Technique - Generative Adversarial Network (GAN)** : This is a complex model consisting of a generator capable of producing fake data samples (images in our case) and a discriminator whose job is to differentiate between the real and the fake samples. The learning is influenced by the min-max tradeoff between the generator and the discriminator.

Strategy for Model Selection We select kernel based approach to use convolution filters. Also, in the neural network architecture, both linear and non-linear approaches are also embedded within in the form of fully connected components as well as activation layers while training models like the VAE.

In order to tune the hyperparameters like learning rate and gradient descent approaches, we use the holdout technique. Based on the different values of the hyperparameters and hence, different configurations of the model, we compare the different evaluation metrics and choose the one which guarantees the highest performance measure.

Training Approaches

For the task of training a VAE and a CNN, we propose to use a Stochastic Gradient Descent which is relatively faster than the Batch Gradient Descent.

We intend to explore and compare the performance of two different types of gradient descent approaches - Batch GD and Stochastic GD for training our models on GANs. We iteratively alternate between one gradient descent step on the discriminator to one on the generator so as to minimise the loss between the ground truth image and the output image of the network. We intend to use Adam's optimiser for mini-batch gradient descent with appropriate learning rates and momentum parameters.

Evaluation Metric

In order to estimate an evaluation metric for all the three CNN, VAE and GAN based approaches, we consider taking the euclidean pixel to pixel difference between the ground truth and the predicted image as the desired measure. However, for the GAN based approach, the idea is not only to get the predicted synthesized image closer to the ground truth image but also at the same time generated the predicted image no far from reality.

So, in order to cover up this high-level goal of "real v/s fake" to make the output more genuine to human vision and observation, we aim to conduct a user survey. For each person appearing for the survey, he is shown a set of images, each for a time duration of 1 second and then given enough time to classify it as a real or a fake image. After certain such trials, he is given some feedback accordingly. When evaluating the final performance, the evaluator is given real and fake images corresponding to different set of inputs in order to maintain the complication of the task.

Both these evaluation metrics help us in finally commenting on the performance of various techniques and algorithms implemented in order to tackle the problem of consideration.