# Project Proposal

# Estimating Geo Location using Convolutional Neural Networks

Manan Shah
Arizona State University
Tempe, Arizona
Mpshah3@asu.edu

Aastha Khanna
Arizona State University
Tempe, Arizona
Aastha.khanna@asu.edu

## I. INTRODUCTION

While the human brain interestingly is able to find contextual information on seeing images and make intellligent decisions based on that information, it is relatively difficult to reach an comprehensive solution on seeing images . We aim to solve such a problem in the field of Deep Learning, where our main focus is going to be finding the location based on a given image. It solves a small part of a bigger problem which is making decsiion based on the inormatio extracted from the image.

The problem of finding the "geolocation" is one which has its wide uses in various problems and formes the basis for solving the bigger problem. It is a problem which aims to recognise several distinct features and attributes of an image, rather than just the identification of an object. Resolving this recognition based on geo location is essential for the development of sophisticated Artificial Intelligence Framework.

## II. RELATED WORK

There have been a couple of work done in the past in order to estimate geo location based on an image. Work done by Pedada and Hong use Convolution Neural Network(CNN) to train models from scratch and with trasfer learning. They have used NINImage Net Transfer Learning Model to achieve more accurancy and better performance.

The approach of finding geolocation based on landmarks has also been previously used in which a collection of images containeing the lnadmarks images is used to train the model. Using this method, the image which needs to be located is matched with its corresponding landmark image . It uses Support Vector Machine trained on Bag of Visual Words[3].

When image is retreived based on certain attributes and features like of Bag of Visual Words and Inverted Indices give better results and accuracy rather than training the model based on landmark images. The product Skyline2GPS uses the same approach and converts the view of the image taken from camera into segments and compares it with 3D model of the city[5][3].

The approach used by Zamir and Shah is to use Generalized Minimum Clinique Graph (GMCP) and Nearest Neighbour Algorithm to get suprisingly accurate results for geolocation estimation[3].

Given the scope of the project , we aim to extensivey use Convolution Neural Ntework to achieve better results in geolocation estimation problem and would strive to gain more knowledge and applications of the subject.

## III. APPROACH

Our initial approach is to train the model on google street view images obtained by Zamir and Shah[1]. The dataset contains 62,058 high quaThe input to our CNN are the image pixels and the target output is a one-hot vector encoding the cell containing the geotag of the image. Given a test image, the output of this model is a probability distribution over the world. The advantage of this formulation over a regression from pixels to latitude/longitude coordinates is that the model can express its uncertainty about an image by assigning each cell a confidence that the image was taken there. In contrast, a regression model would be forced to pinpoint a single location and would have no natural way of expressing uncertainty about its prediction, especially in the presence of multi-modal answers (as are expected in this task). Google Street View images. The images cover the downtown and neighbouring areas of Pittsburgh, PA; Orlando, FL and partially Manhattan, NY. Accurate GPS coordinates of the images and their compass direction are provided as well.

For each Street View placemark (i.e. each spot on one street), the 360° spherical view is broken down into 4 side views and 1 upward view. We are first going to study the data and find out ways to extract the features from the data. Before using the data, we will down size each image to approximately 256x256 pixels for each RGB colour.

We treat this task of image geo-location as the classification problem and use CNN for this purpose. The input to our CNN are the image pixels and the target output is a one-hot vector encoding the cell containing the geotag of the image.[2] Given a test image, the output of this model is a probability distribution over the world. The advantage of treating this as classification and considering the probability distribution is unlike regression we don't have to predict a certain value in this case a latitude and longitude. We can have the probability of each pixel to belong to some place in the real world.

We will decide the design of the image after training the model on subset of data, say only one region and see the accuracy of the model on validation set.

## IV. TIMELINE

The initial couple of weeks, we plan to go through different machine learning algorithms like KNN[1] which can also appropriate for classification in this problem statement and understand how CNN is better [3].Next we have to go through the data and play with it and find out ways to downsize the pixels and extract useful features. At the time of writing this proposal, both the authors have rudimentary skills in constructing networks in TensorFlow[4] but we wish to explore other frameworks, such as Theano or Torch, before we finalize on the framework we decide to use for development.

To start of we will first try to generate a model on very small set of training data using CNN. Then we would observe how the model learns and then tune the parameters. Once the model begins to learn, we hypothesize that training might take a significant portion of our project timeline. In order to minimize the training time we would try to train the model on GPU.

During the final set of the project we will try to use this model and test in on Flickr dataset and see how well it generalizes on the user generated location subject tags.

## V. CONTRIBUTION

We believe in together working towards a problem and would be contributing equally in all the phases of every task so that both us have a equal knowledge of all the tasks involved including the Research related to the topic as well as Development. We would divide the topics amongst us and would explain each other for effective team work.

## VI. CONCLUSION

As the part of this project we will learn different methods that can be used for the geo-location estimation based on images. We will work on developing a CNN and evaluate its performance for classification for such problem statement. We will try to generalize the model such that it classify the geo-location images from other data source like Flickr to certain degree of confidence.

## VII. REFERNECES


[1] http://crcv.ucf.edu/data/GMCP_Geolocalization/.

[2] https://arxiv.org/pdf/1602.05314.pdf

[3] http://cs231n.stanford.edu/reports/CS231N_Final_Report_amanivp_jamesh93.pdf

[4] TensorFlow, by Google: http://tensorflow.org/

[5] http://ieeexplore.ieee.org/document/5649105/


.