Ranti Dev Sharma(rds004@ucsd.edu)
Under guidance of Dr. Manmohan Chandraker

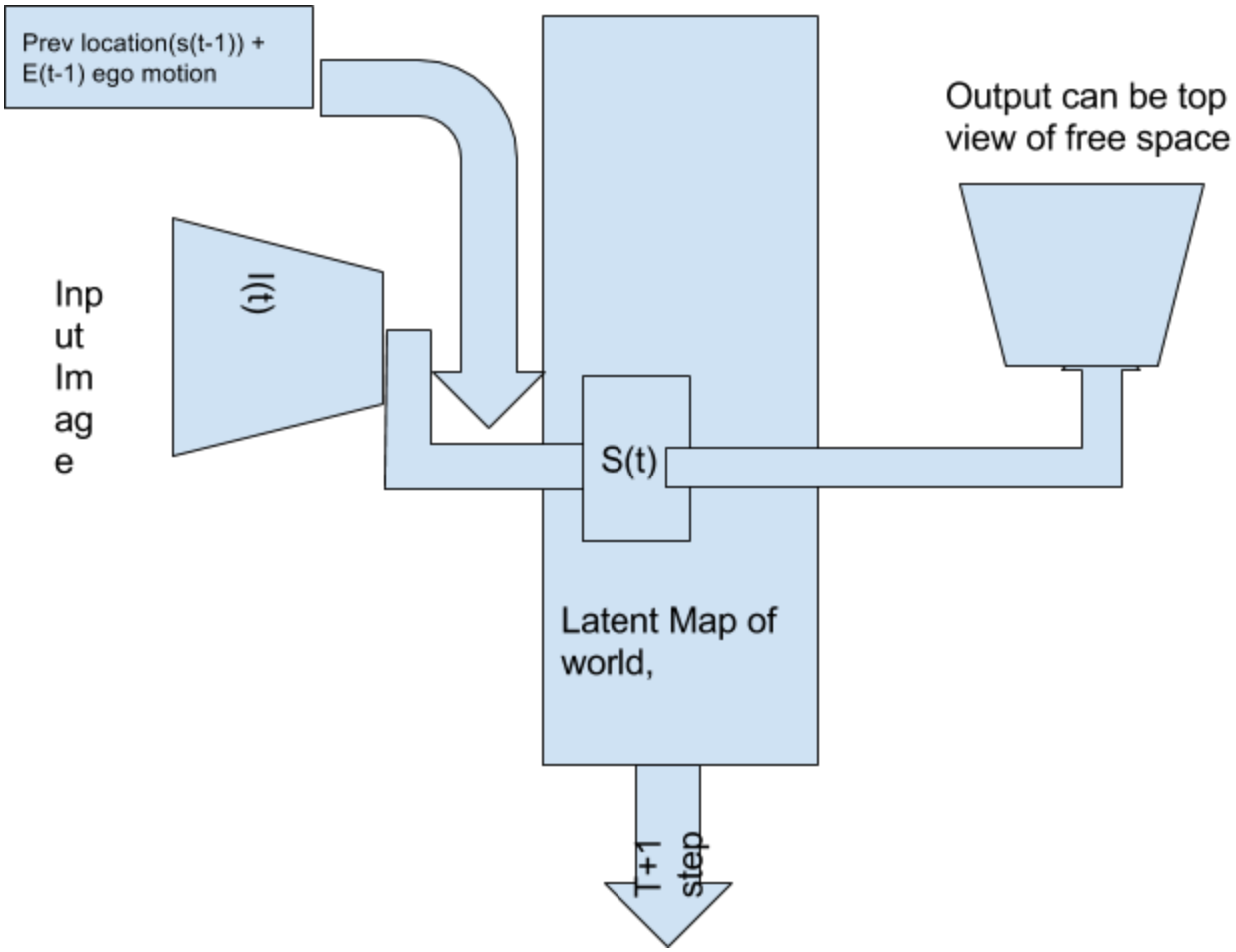
Latent Map Generation: Using Top View from Front View

Free space from top view gives localized map of world, so if we combine these localized views over time using ego motion we can get whole map of world. This problem is vision based SLAM.

Methodology:

Let the hidden layer be latent map of world which we will learn using weak supervision of detecting free space in top view problem or similar to that. So input will be Image I(t) and ego motion E(t-1) from previous time step (note: we can always estimate ego motion using simple ego motion detection neural net which takes current image and previous image, if ego motion is not given or accurate enough). So given ego motion with respect to previous step, We will move relative vector to S(t-1)+E(t-1) location on that latent map with some window size W and only update that part of map and produce output only form that locality of latent map

Latent Map Network:



Problem Division:

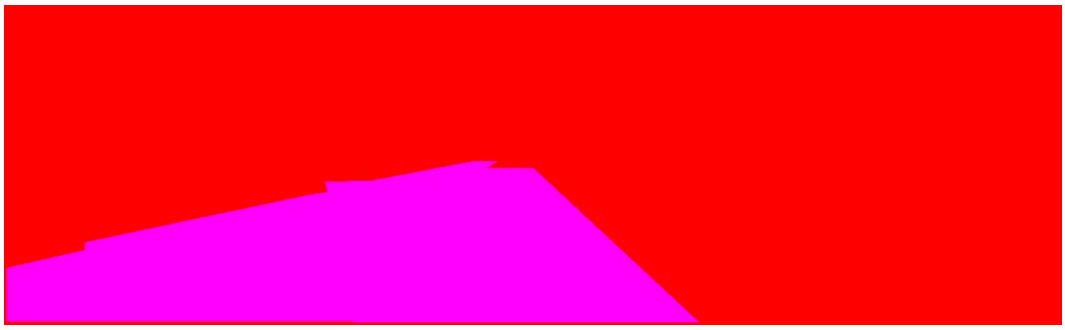
We divided the problem into three parts.

1) Data Collection.
2) Simple Front to Bird Eye View Generation.
3) Weakly Supervised Latent Map Generation.

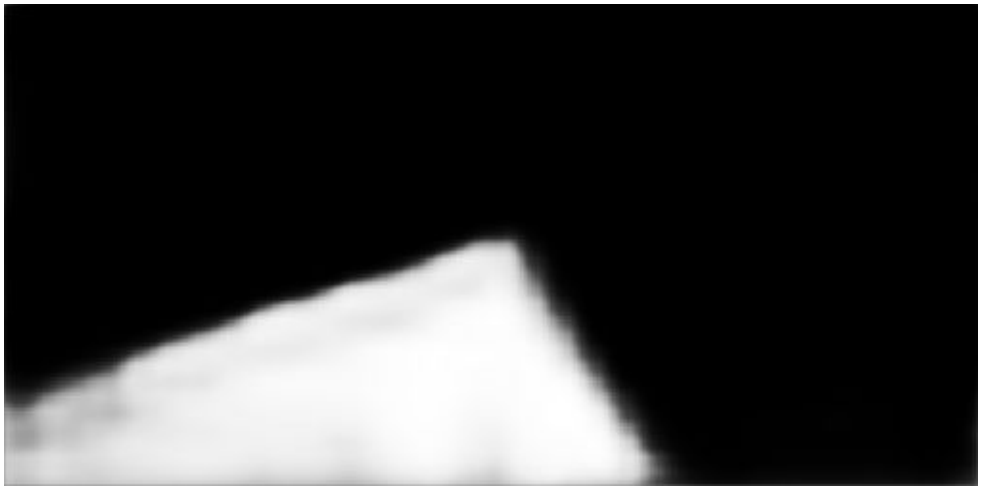
Part1: Data Collection:

As proposed problem is very novel. So we couldn't find any given datasets which can be directly used for this problem. We tried to use use KITTI datasets as they have GPS,3D point clouds and 3D bounding boxes associated with cars. We use pretrained Road Segmentation networks to detect road in image. We associate those road pixels with their corresponding 3d points. As mostly roads are horizontal we try to fit RANSAC plane to get a better road plane estimation and use inliers as points on road. After we have detected we use homography to project them to Bird Eye View.

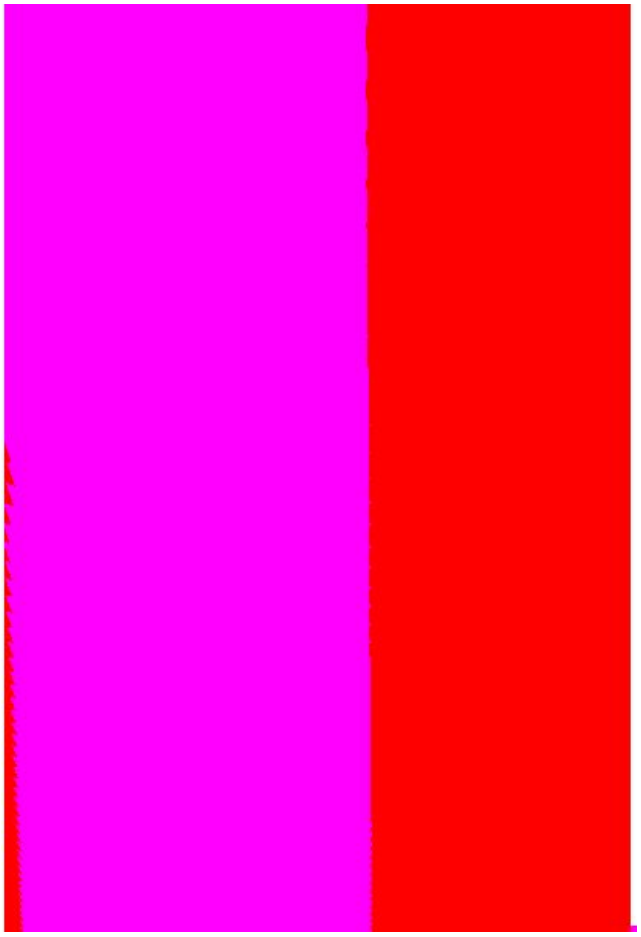
Road Segmentation:



RANSAC Plane fitting of 3D points:



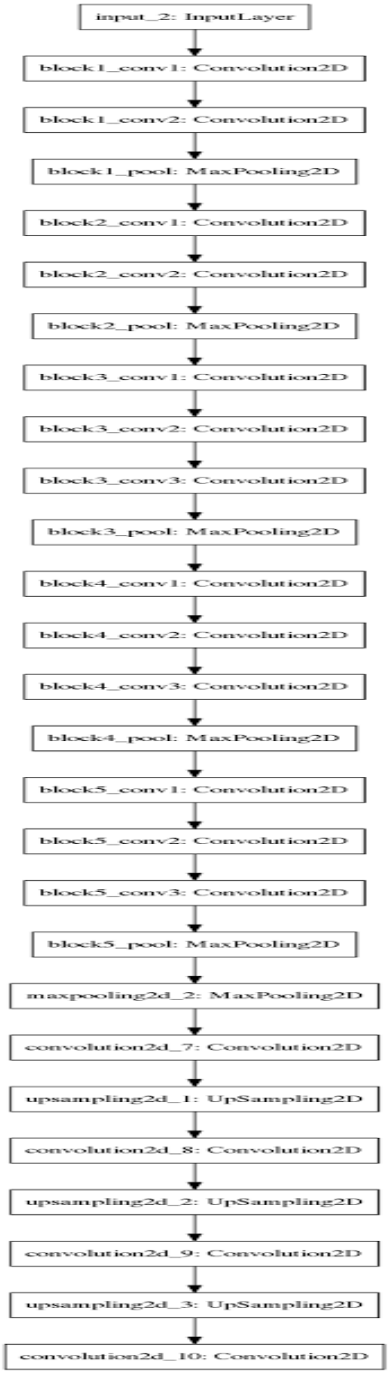
Using Homography to project in Bird Eye View:



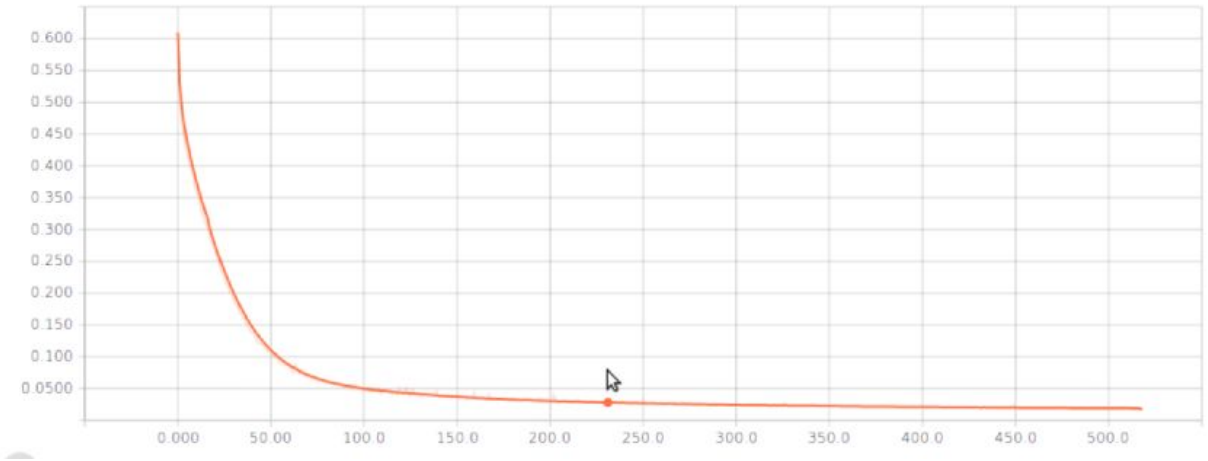
Part 2: Simple Front to Top View Generation:

We used a simple Convolution based Encoder-Decoder network. Where Encoder part was pretrained using VGG-16 weights. It was trained end to end to produce simple top view of road from given front view of road.

Architecture:

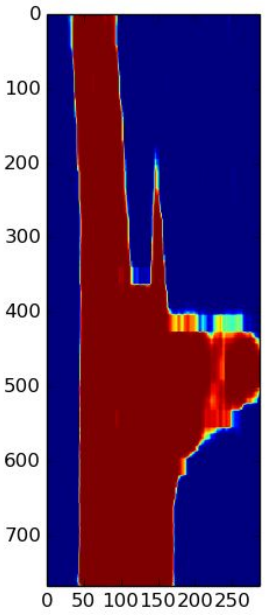


Training and Loss:
Cross entropy loss was used to output road in Bird Eye View. we trained our network with 500 epochs.
Validation loss curve:



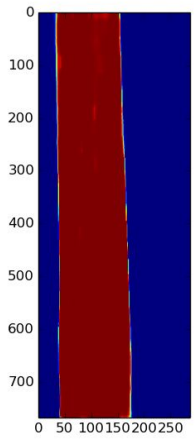
Some of the best results:

Input Image:



Output:

Input Image:



Output:

Part 3: Weakly Supervised Latent Map Generation:

For this part latent map layer will be custom made. Its architecture will be Locally Spatial Convolutional LSTM/GRU. It's input and output region of updation will be based on ego-motion or GPS of vehicle. Its area of activity is selected such that it only tries to update area surrounding vehicle which is currently in top view.

There are many benefits of using latent map layer as we can input prior information of map of world into map which can be further used by network to update its world.

Architecture: please note that architecture may change in future

Current architectures encoder-decoder type architecture.

Encoder: We directly used VGG-16 as encoder for this network. Encoder gave us output as 25*25*512 as output.

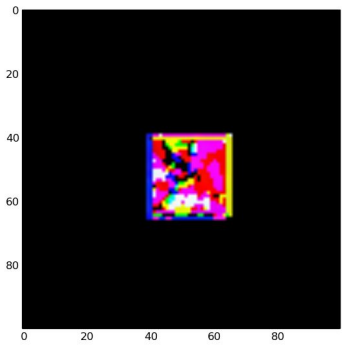
Latent Map ConvLSTM layer: Our main contribution to this is, We introduced a new LatentMap ConvLSTM layer which selectively updates its LSTM memory by taking selective spatial convolutions over input given by encoder and produces latent map of world.

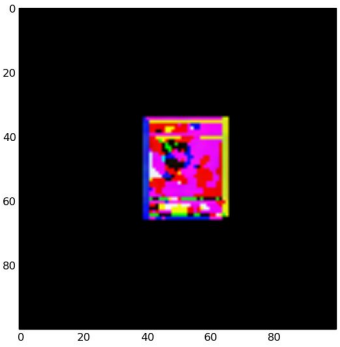
Decoder: Our decoder is also spatially selective it only takes that part of memory which can be used to project top view of world. Using Transpose Convolutional layers also known as Deconvolutional layers it Upsamples view from 25*25*3 to 400*400*1.

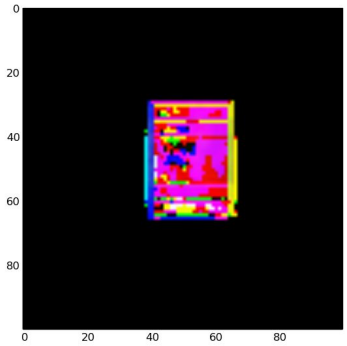
All these layers were coded in TensorFlow library.

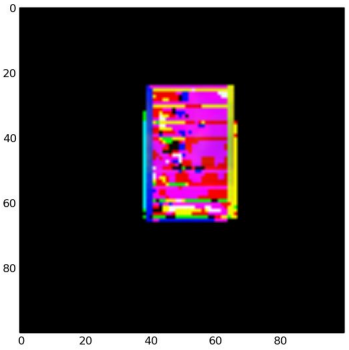
Some Results:

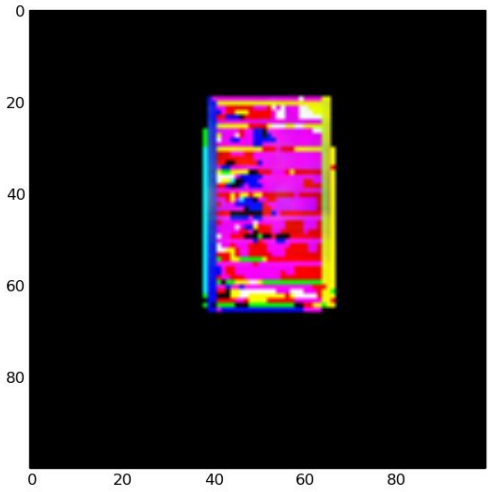
In these results a car is moving forward updating latent map.











Above FINAL MAP:
This map was generated for this below sequence:

Currently for this part work is in progress.

Conclusion:

We experimented with simple Bird Eye View generation from front view images and we were able to produce decent results. Also using similar approach we can learn various other type of homographies on the go using our CNN based neural networks. As this is a novel problem there are many research problems which can open up. We experimented with latent map model but we were not able to produce clean results. As that requires a lot of training but we trained on some small training data. In future we will try to improve latent map model with more training and more better architectures.