Road Segmentation

Faisal Maqbool, Syed Shahbaz Hussain
msds17027@itu.edu.pk, msds17016@itu.edu.pk

Abstract

Extracting road maps from high resolution optical remote sensing or satellite images have received much attention recently. We have developed a deep learning based system which, given an aerial image can output a binary mask for the input image showing for each pixel if it belongs to a road or not. Our approach is based on **Massachusetts Roads Dataset (Mnih)** containing total **1438** high resolution images along with their road masks.

Problem Statement

- In remote sensing analysis, automatic extraction of road network from satellite or aerial images can be a most needed approach for efficient road database creation, refinement, and updating.
- The challenge, now, is extracting the road map (mask) from a given aerial image.

Fig 1. Satellite Image

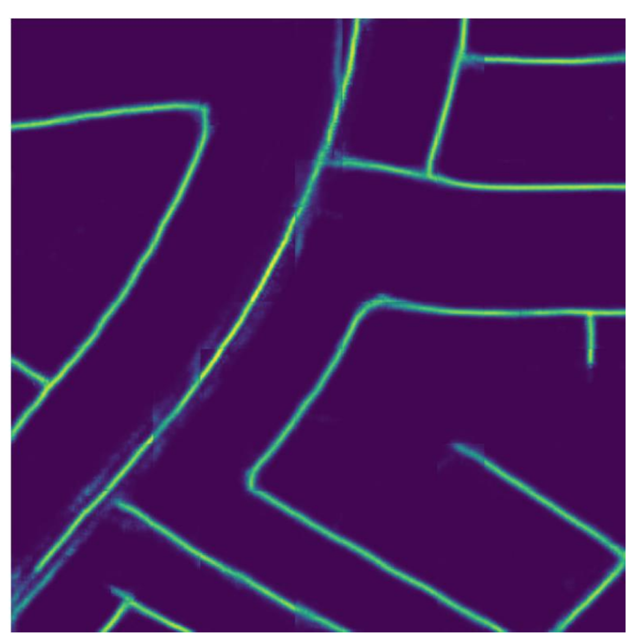


Fig 2. Extracted Road Map

Applications/Project Promise

The project falls into the domain of rapidly evolving field of remote sensing (Semantic segmentation) and its potential applications include:

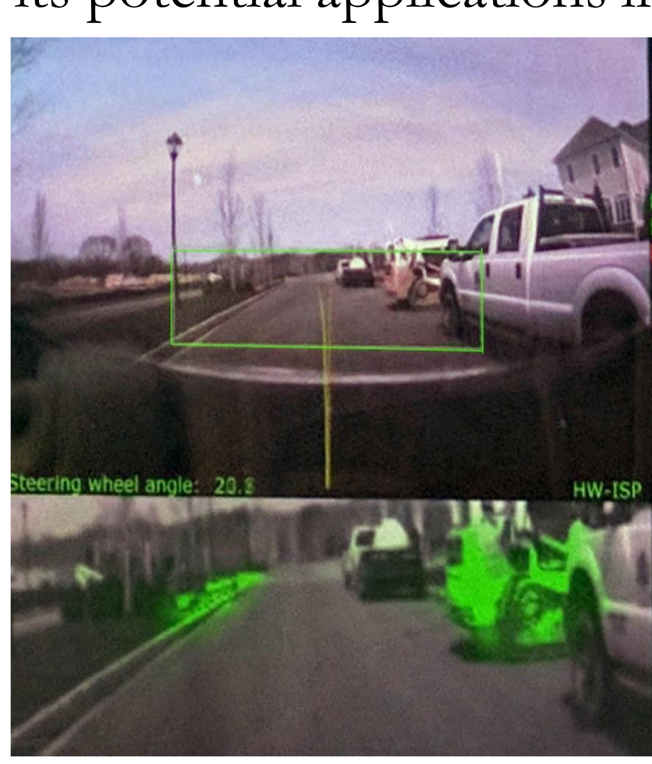


Fig 3. Road Navigation

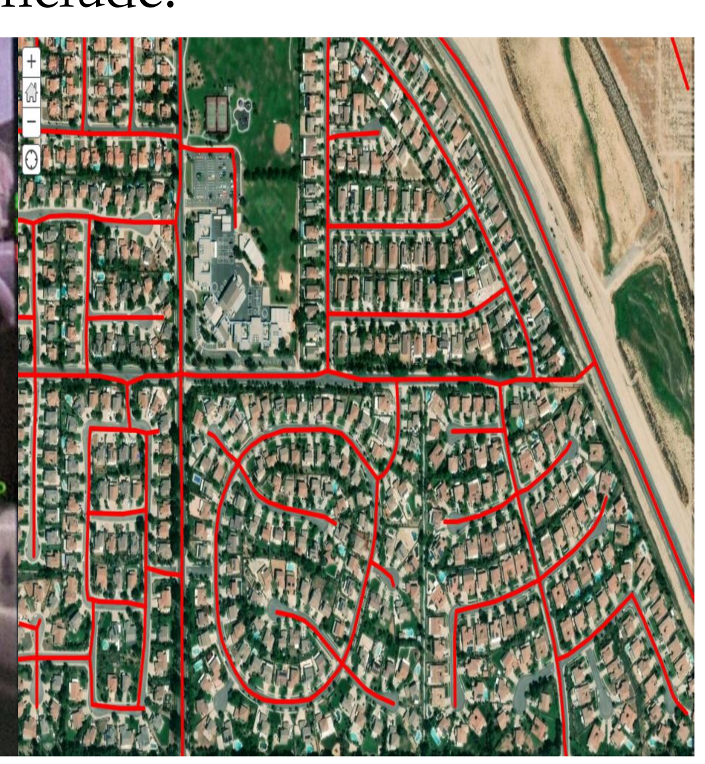
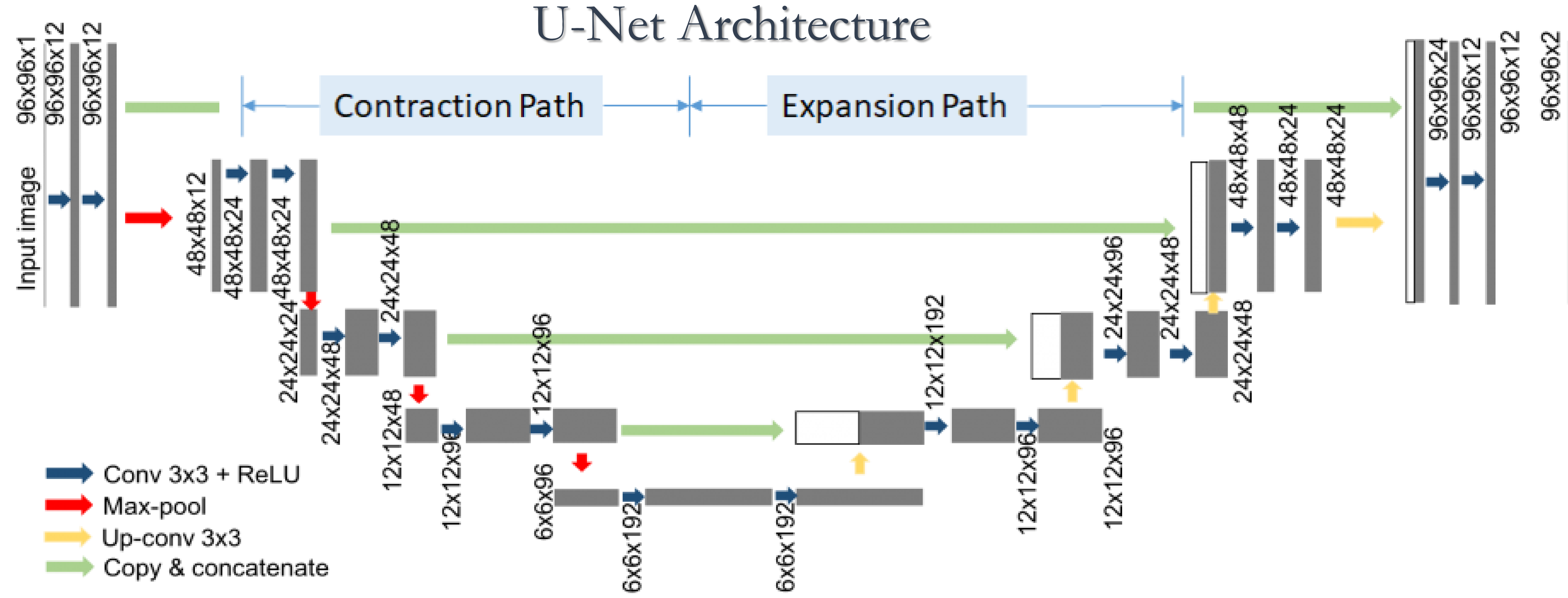


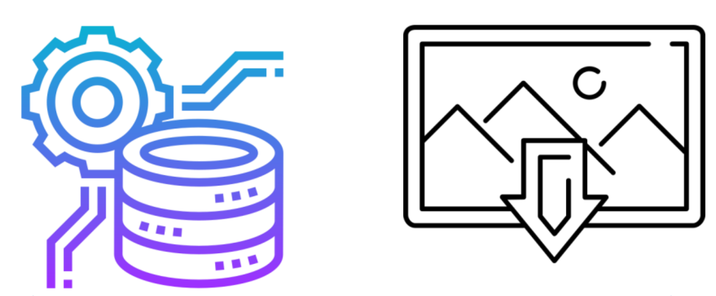
Fig 4. Urban Planning

U-Net Architecture



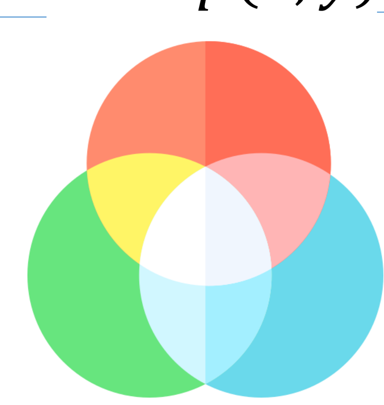
Methodology

- For extraction of data we wrote a scrapper and downloaded training, testing and validation dataset.
- A visual depiction pre-processing applied for both DNN and U-Net [1] is as follow:



CSV Files

L Pixel $p(x, y)$ L

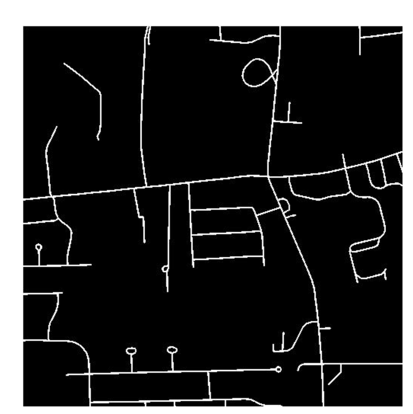


$FeatureVector = 3 * L * L$

Containing RGB values

1500 x 1500

256 x 256

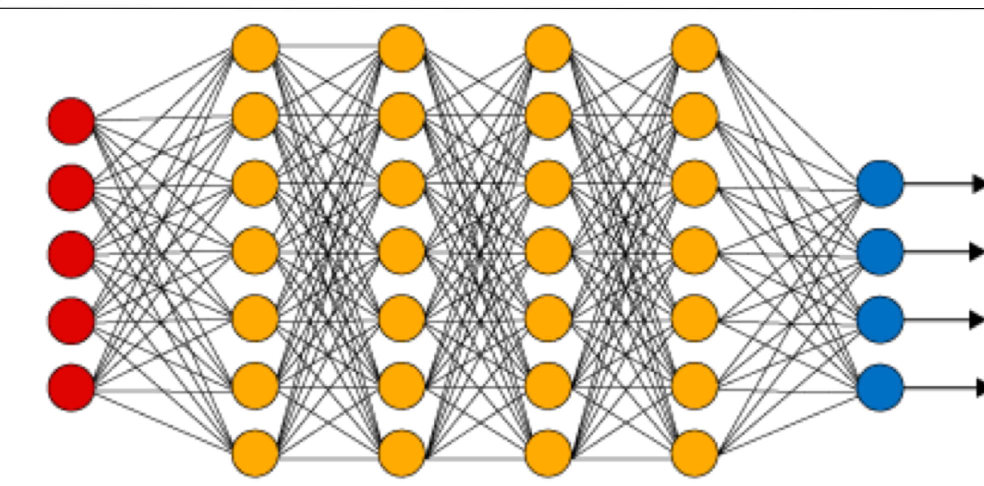


Ground Truths are binarized version of input

Test, Train & Validation set saved in h5 format

$Per - Pixel\ Technique$ $U - Net$

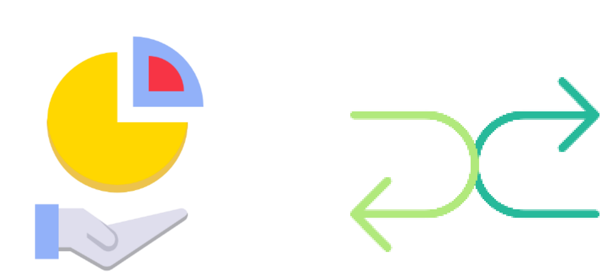
Fig 5. Pre Processing



- For per-pixel technique, we used **Tensorflow** for a DNN with 4 hidden layers of sizes **100, 150, 100, 50**
- For second method, we used **U-Net** model
- with **Keras** which is encoder-decoder type network where feature maps from convolution part in down sampling step are fed to the up-convolution part in up-sampling step

Class Imbalance
Pixel wise classification

- Ratio between road pixels and non-road pixels in an image is very large
- To handle that we performed dropout by randomly shuffle and for each road pixel took two non –road pixel to be 1:2
- This is done for training images, testing images and validation images to generate the three files.



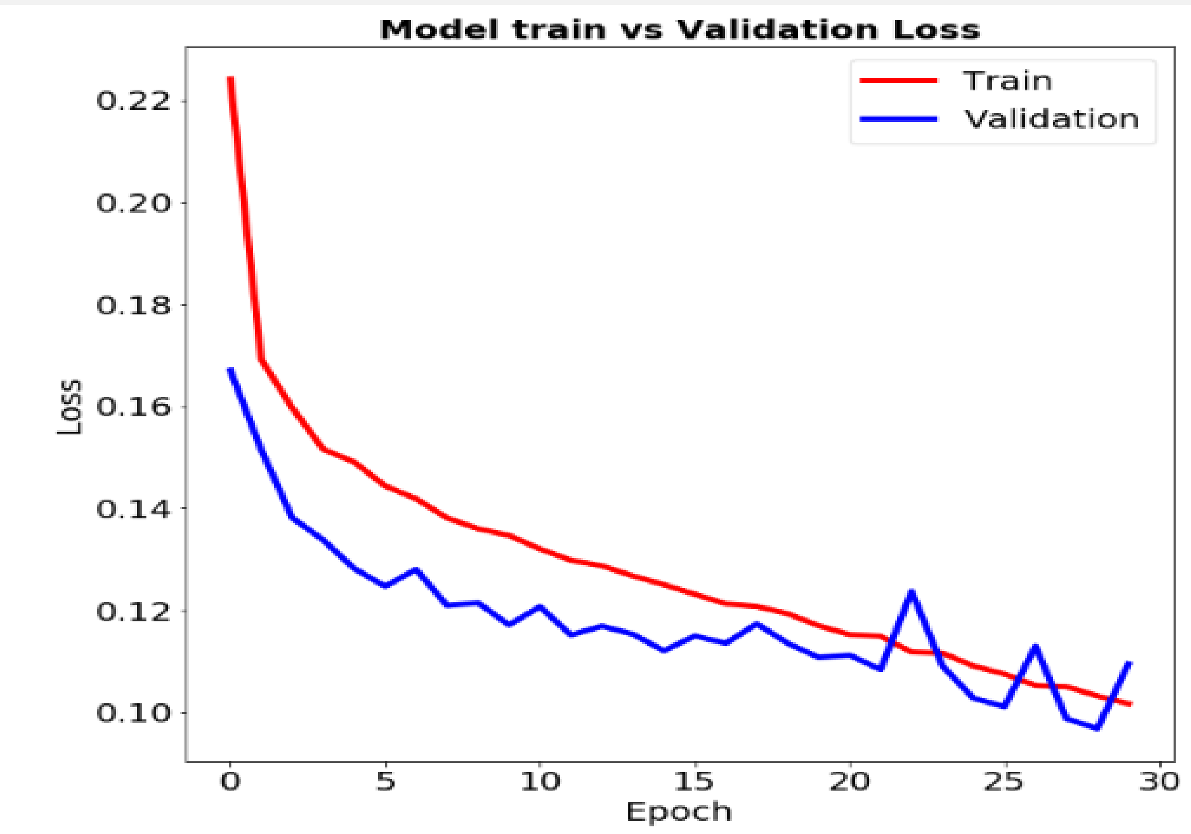
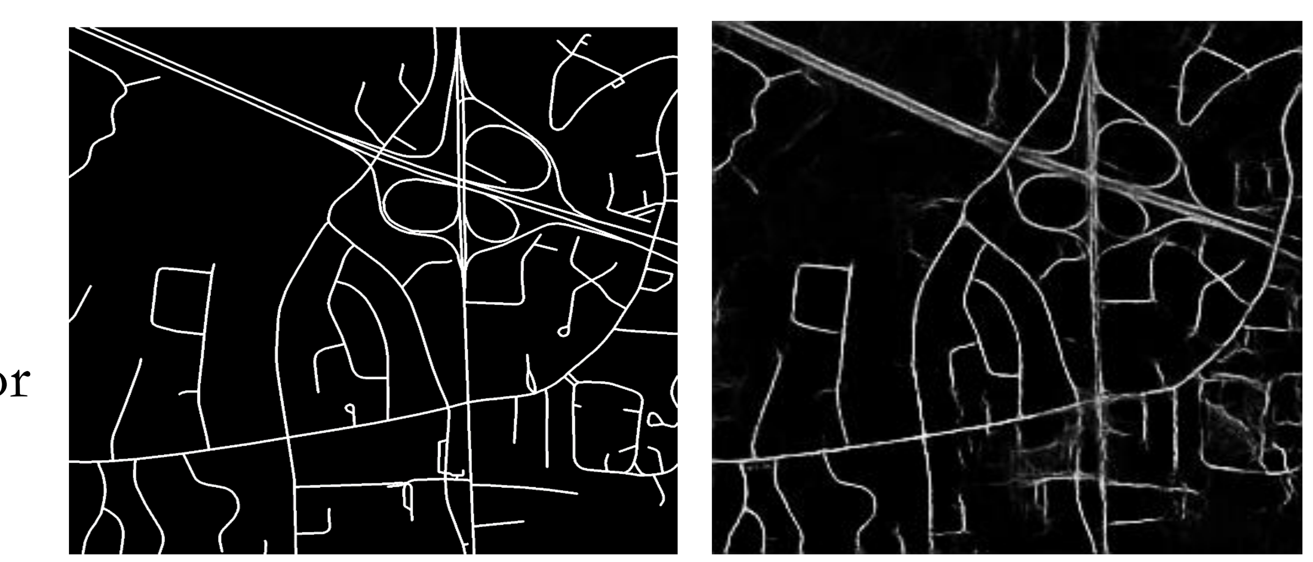


Figure 6. The model U-Net was trained for 30 epochs on colab GPU. After 30 epochs, the model training accuracy was about 97% and validation accuracy was about 97%

Main Result



Ground Truth Prediction

- Loaded model for **EPEL** dataset for predicting road masks

Testing Accuracy

Model	Input	Accuracy
DNN	3 * L * L RGB	81.7 %
U-Net	256 * 256 Image	96.7 %

References

[1] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Springer, Cham, 2015.