Road Surface Segmentation and Classification

CSE 509 : Digital Video Processing

Group 8:
Krutik Parmar, Janam Vaidya, Vedant Parikh, Rishabh Pandat

Arizona State University

Problem Statement

- In the area of vehicle and robotic navigation, a significant unresolved challenge is that of detection of obstacles and finding a reliable path.
- The work done so far in this area has focused on both path detection and obstacle detection, but not on relevant factors for vehicular navigation like pavement of the road surface, speed-bumps and markings of the road.
- The dataset that has been used are mostly based on images from developed countries, from Europe or North America, with well-maintained roads, having few examples of damaged roads, nor dealing with variations in terrain type.
- It is our understanding that, in dealing with common problems found on roads in emerging countries, but which can also occur in developed countries and are of utmost importance to vehicle behavior, both for the sake of vehicle preservation and especially for safety, we are advancing the state-of-the art of path detection.

Overview

- Our goal with this work is to perform road detection with the differentiation of surface variations, in addition to a concomitant surface damage detection.
- We also aim to show that it is possible to use passive vision (dashboard cameras) to detect road damages.
- We used two methods for our problem and compared their results.
- We believe that the contributions and differentials of our approach are:
    - Detect and recognition of the different types of surfaces (asphalt, other pavement and unpaved);
    - Detect potholes and water-puddles on the road, even on different types of surfaces;
    - In conjunction with the previous points, detect other damage and patterns on the road, all with the same approach

Dataset

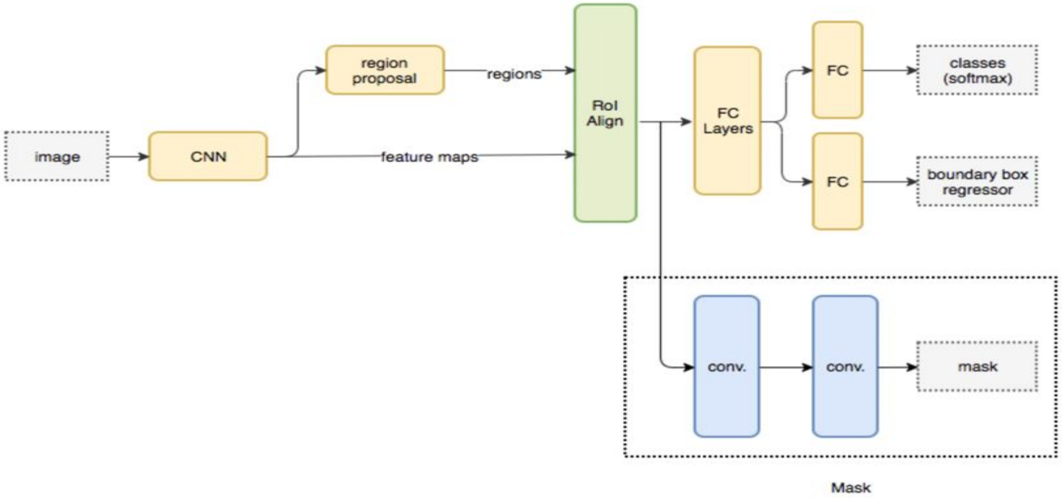
- We used a subset of Road Traversing Knowledge (RTK) dataset with instance segmentation ground truth masks
- The dataset consists of 701 images of 12 different classes.

Original Image with Paved Road and its ground truth mask

Original Image with Asphalt Road, Road markings,water puddle and its ground truth mask

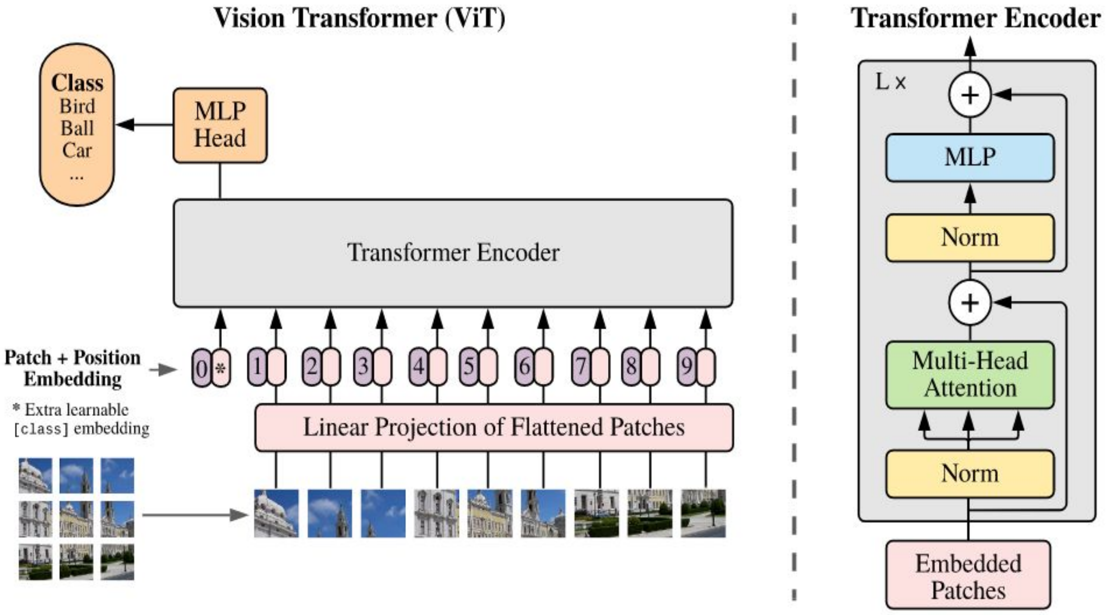
Approach

- Mask RCNN - Mask Region-Based Convolutional Neural Network
- Input data is images of road and their ground truth masks
- Output of mask rcnn is labels of the detected object, their masks and scores.



Approach 2

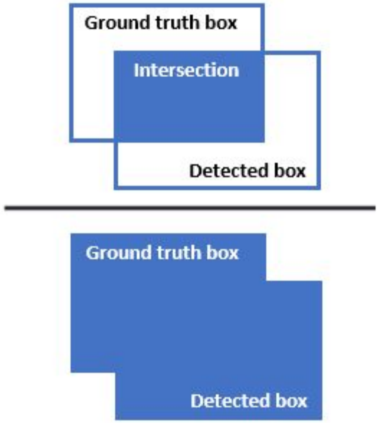
- While the Transformer architecture has become the de-facto standard for natural language processing tasks, its applications to computer vision remain limited.
- It takes input of 16x16 pixel patches of the input image and returns the mask, confidence score and classes.
- CNN counterparts like Resnet are computationally expensive as we can perform parallel computing in visual transformer.



Evaluation

- We used mean IOU as evaluation metric

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} =$$

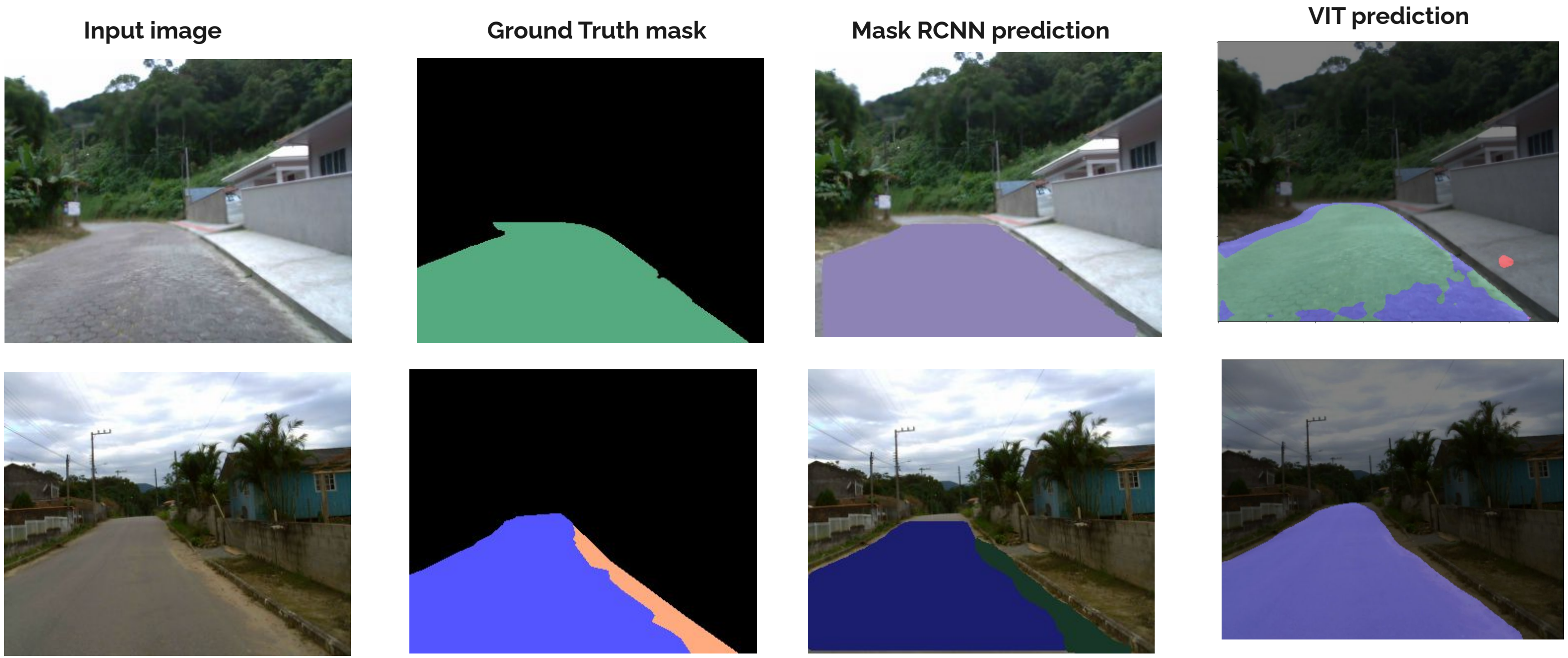


Results

Class	Mask RCNN	VIT
roadAsphalt	0.471	0.638
roadPaved	0.573	0.545
roadUnpaved	0.537	0.657
roadMarking	0.34	0.12
speedBump	0	0
catsEye	0.108	0

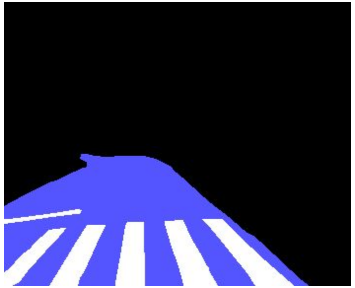
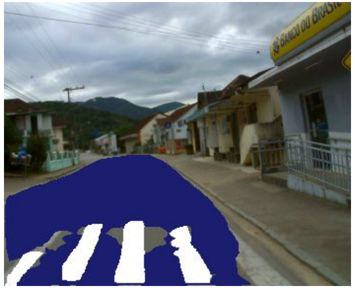
Class	Mask RCNN	VIT
stormDrain	0.67	0
manholeCover	0	0
patchs	0.036	0.08
waterPuddle	0.061	0
pothole	0.004	0.016
cracks	0.038	0.094

Results

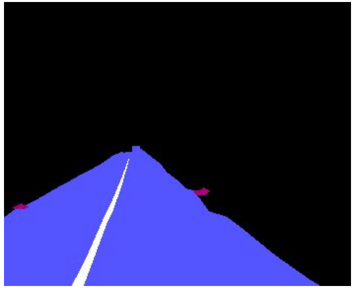

Input image
Ground Truth mask
Mask RCNN prediction
VIT prediction

Results

Input image | Ground Truth mask | Mask RCNN prediction | VIT prediction





Learnings

- We got very high mean IOU scores for *roadAsphalt*, *roadPaved* and *roadUnpaved* since most of the images contained these three classes and the area of these classes was higher than the rest of the classes. So the model learned these three classes better than others.
- VIT is data hungry therefore the classes with less area had a better mean IOU score with maskRCNN compared to VIT.

Thank You