VIP-IPA: Lane Detection

Authors: William Stevens, Mykhailo (Misha) Tsysin, Xingyan (Ian) Li
Mentors: Prof. Edward J.Delp, Prof. Carla Zoltowski

PURDUE UNIVERSITY

INTRODUCTION

With the prevalence of autonomous vehicles, the computer vision algorithms utilized for autonomous driving must be robust and accurate to assess road features through images captured in real time. In our previous work, we designed a multi-task neural network model according to the YOLO architecture and trained on the BDD100k dataset, to give detections and classifications of objects, as well as segmentations and classifications of lane lines, in an image.

While last semester's model was successful, there was room for improvement. So, we wanted to explore potential optimizations to the network to achieve better results. This process included the research, development, and testing of alternative strategies for the many components of our multi-task model. This will be achieved through modifying last semester's model according to the fourth edition of the YOLO architecture. The modifications to the model will include implementing an augmented feature extractor and an improved feature aggregation network through the works done in Cross-Stage Partial Networks and Path Aggregation Network. The feature extractor is designed to greatly reduce the number of duplicate gradient computations, and the improved feature aggregator prevents the significant loss of spatial information that occurred in our previous implementation, while still allowing the model to effectively merge shallow and deep information to achieve better predictions.

These developments will allow our multi-task model to perform the three vision tasks for autonomous driving at a lower computation cost while also achieving higher accuracy, thus creating an improved version of the model for future work to build upon. See Figure I for a diagram of the complete model.

DATA

We are continuing the use of the Berkeley Deep Drive (BDD100k) dataset from previous semesters because it contains all the necessary labels needed for the vision tasks we are performing. The dataset contains 100,000 driving videos and labeled images collected from more than 50,000 rides covering New York City and San Francisco Bay Area year-round with multiple weather and lighting conditions. Due to the diversity of geography, environment, and weather, training on the dataset results in a robust and generalizable model for encountering new environments or changes of seasons.

Containing thirteen different object classes shown in Table I and nine different lane classes shown in Table II, the BDD100k dataset encapsulates many of the classes that the model could be tasked to analyze. Another component of the dataset that was utilized for our purposes is the labeling of drivable areas, which is divided into two classes: direct and alternative. Direct drivable area is defined as the lane the vehicle is currently driving on and thus the region where the driver has priority over surrounding cars. Alternative drivable area is defined as any lane that the vehicle is currently not driving on, but has the ability to do so through changing lanes.

The BDD100k dataset with its large size, wide variety, paired and highly detailed labeling, makes a perfect fit for our training purposes.

Table I: BDD100k Object Classes
Pedestrian, Rider, Car, Truck, Bus, Train, Motorcycle, Bicycle, Traffic light, Traffic sign, Other vehicle, Trailer, Other person

Table II: BDD100k Lane Classes
Crosswalk, Double other, Double white, Double yellow, Road curb, Single other, Single white, Single yellow, Background

Data source: https://bdd-data.berkeley.edu/

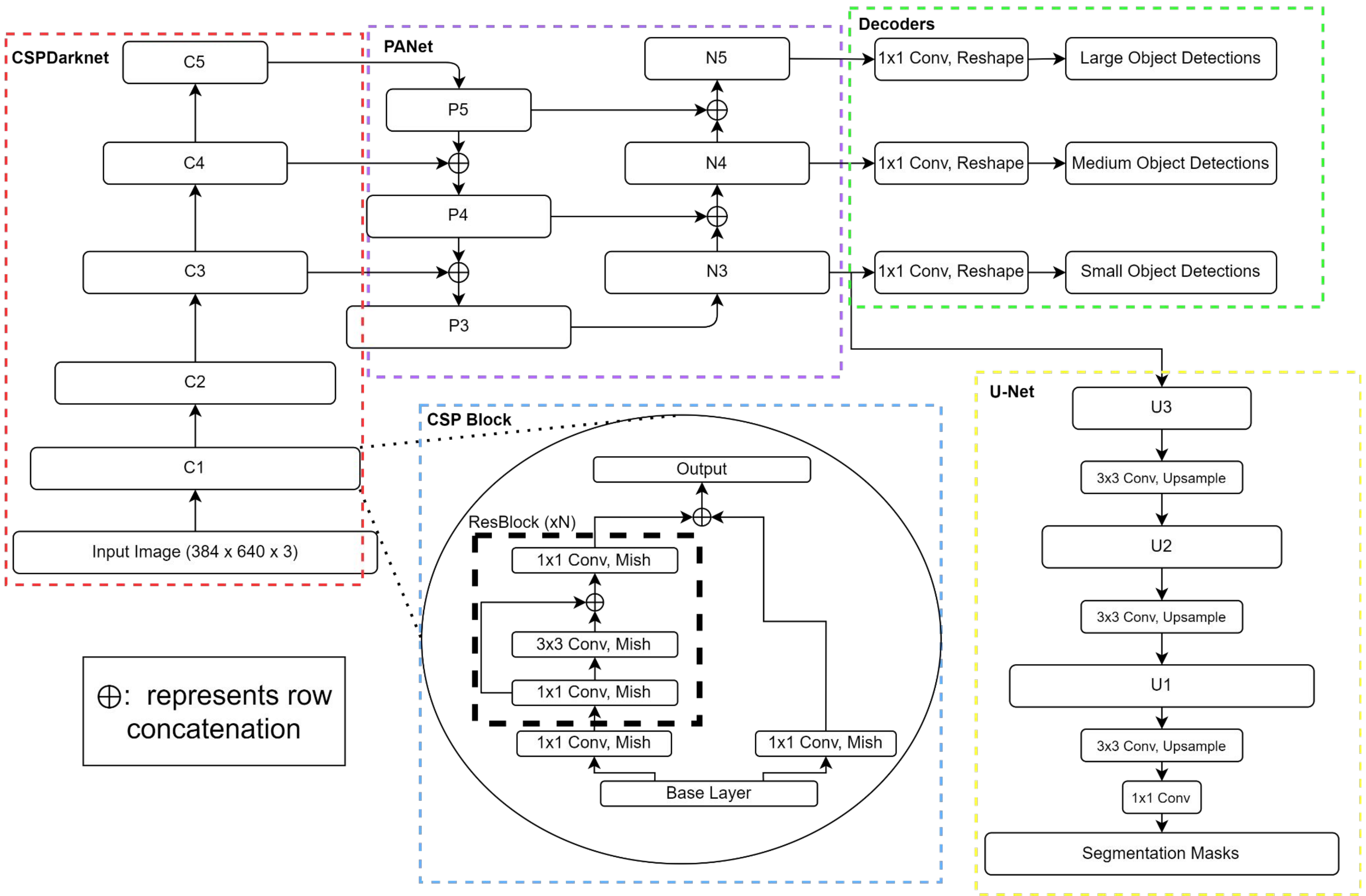

Figure I: Complete Model Architecture

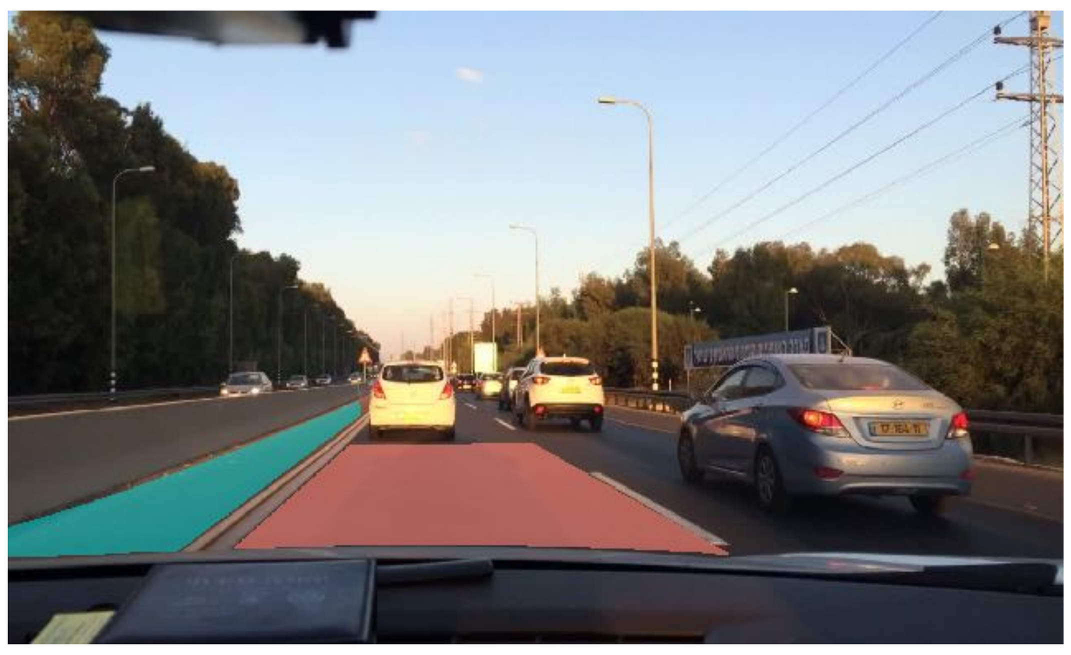

Figure II: Drivable Area Groundtruth Example

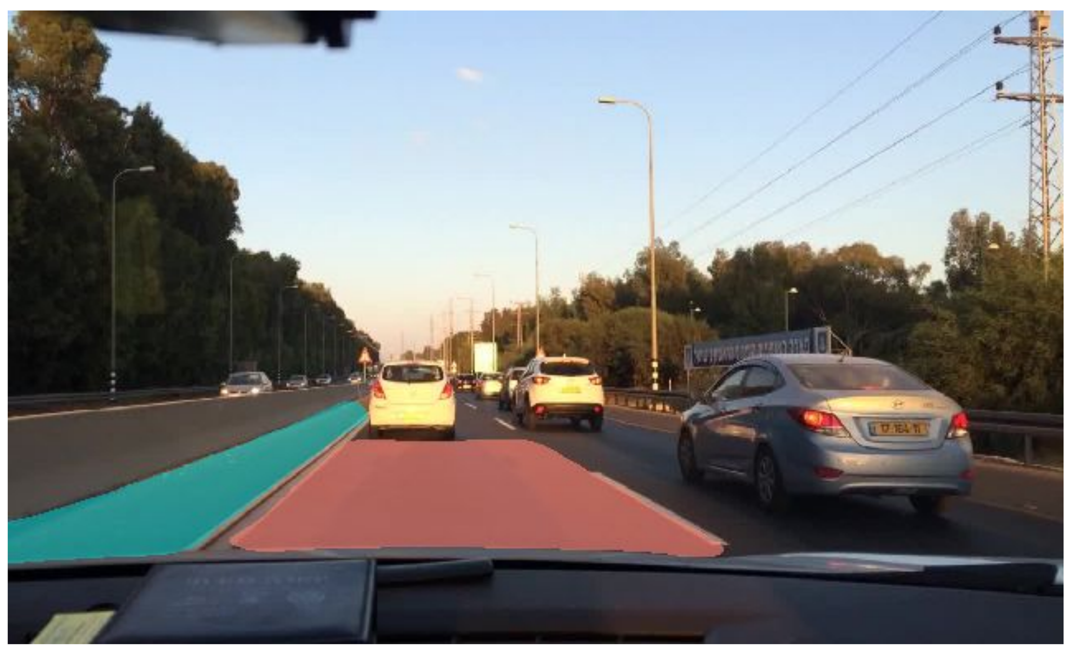

Figure III: Drivable Area Segmentation Result

METHODS

- **CSPDarknet:** The Cross-Stage Partial Darknet serves as the feature extractor. It carries out a series of convolution operations, batch normalizations, and activations, grouped together in CSP Blocks to produce feature maps.
- **PANet:** The Path Aggregation Network serves as the feature aggregator for our model. It utilizes the CSPDarknet features maps at different resolutions to pool together strong semantic information with important spatial information.
- **Mish Activation**: The Mish Activation function is used to introduce non-linearity to our network. It is an augmentation of the previously used Rectified Linear Unit (ReLU), through its properties of continuous differentiability, non-monotonicity, unboundedness above, and boundedness below.
- **U-Net**: Portions of the U-Net framework used in Spring 2022 is re-used in our model to transform the feature maps into segmentation masks for detecting lane lines and drivable areas

RESULTS

The team has managed to implement and test all of the mentioned modifications to last semester's model. Drivable area segmentation (Figure II,III) is visually acceptable. Lane line segmentation and object detection require more work and testing before we can evaluate the performance. We have also implemented data augmentation methods to improve the robustness and generalizability of the model's learning, which will be incorporated into the training in the future.

It is difficult to judge the accuracy of our model before all three implementation tasks (drivable area segmentation, lane line segmentation, and object detection) are complete and fully tested. However, our preliminary results in drivable area segmentation are promising. Our next steps will be finalizing the lane line segmentation and object detection tasks, and then acquiring precise evaluation metrics to compare the accuracy of our model with the metrics obtained from last semester's model.