# A Monocular Local Mapper for Urban Scenes

Names

October 28, 2021

## 1 Proposed Goal

We propose a project that designs a visual system that is able to produce a forward-facing map with individual RGB images that is useful for downstream tasks such as motion planning and decision making.

Specifically, we would like to 1) solve the semantic segmentation, object detection, and depth estimation tasks on the perspective mapped level, that is, images, and 2) map the result to 3D, which would require information about specific camera intrinsic and extrinsic. Based on our progress we may choose to limit our goal to what is feasible.

## 2 Significance

Our proposed project can be seen as a upstream result which is valuable for numerous tasks like self-driving cars. To safely navigate through urban scenes, a system should first be able to differentiate drivable part of the road from the alternative. This simple capability demands successful detection of obstacles like vehicles, segmentation of roads, as well as estimation of distance from various entities.

## 3 Proposed Method

For all of the three tasks we will employ deep learning based methods like Masked-RCNN, U-Net, and YOLO. After each individual task is done, we would possibly try to merge the network into a single one which is a multitask learning problem.