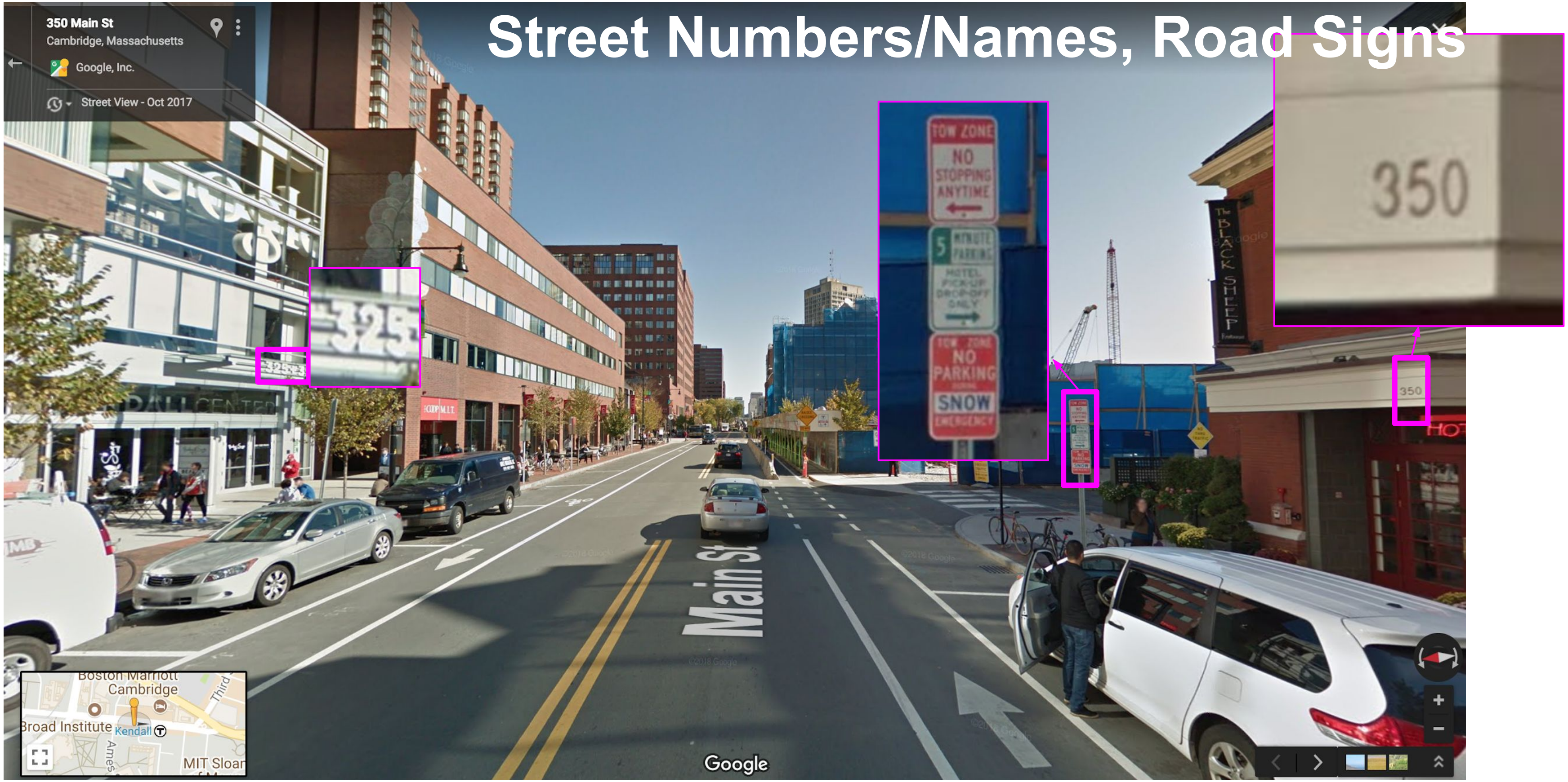
# Map Inference with In-Vehicle Cameras

# What can we infer from in-vehicle cameras?

# What can we infer from in-vehicle cameras?



# What can we infer from in-vehicle cameras?

# What can we infer from in-vehicle cameras?

# What can we infer from in-vehicle cameras?

# What can we infer from in-vehicle cameras?

- Road signs, street numbers and names, business names
- Detecting opening parking spaces
- Road Marker SLAM
- 3D Maps Pipeline

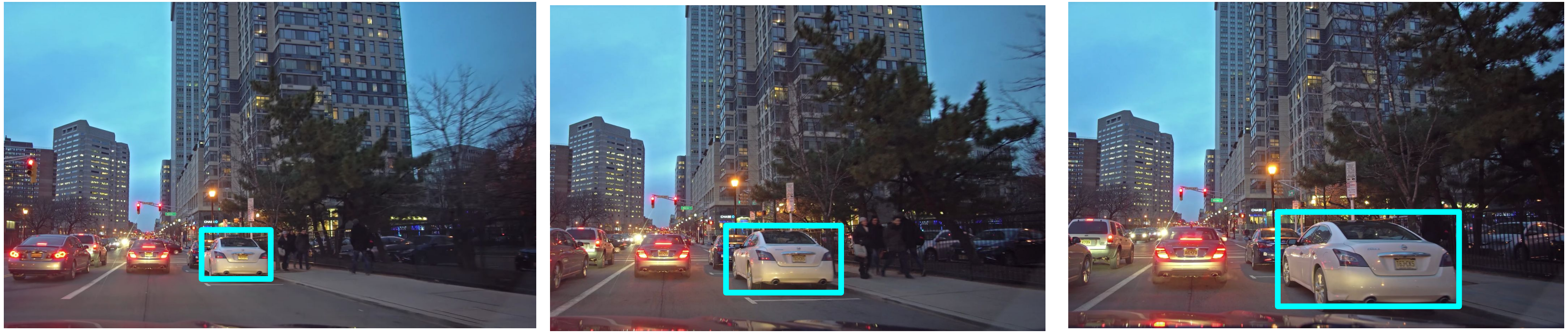
# Detecting Open Parking Spaces


Goal: Counting the number of cars parked on each road.

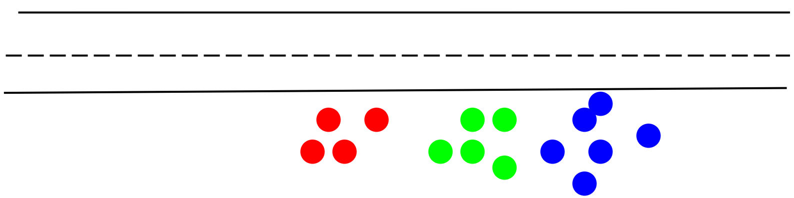
A car may appear in several consequent frames of the video.

ParkMaster: An In-Vehicle, Edge-Based Video Analytics Service for Detecting Open Parking Spaces in Urban Environments.
Giulio Grassi, Paramvir Bahl, Kyle Jamieson, Giovanni Pau.

# Detecting Open Parking Spaces

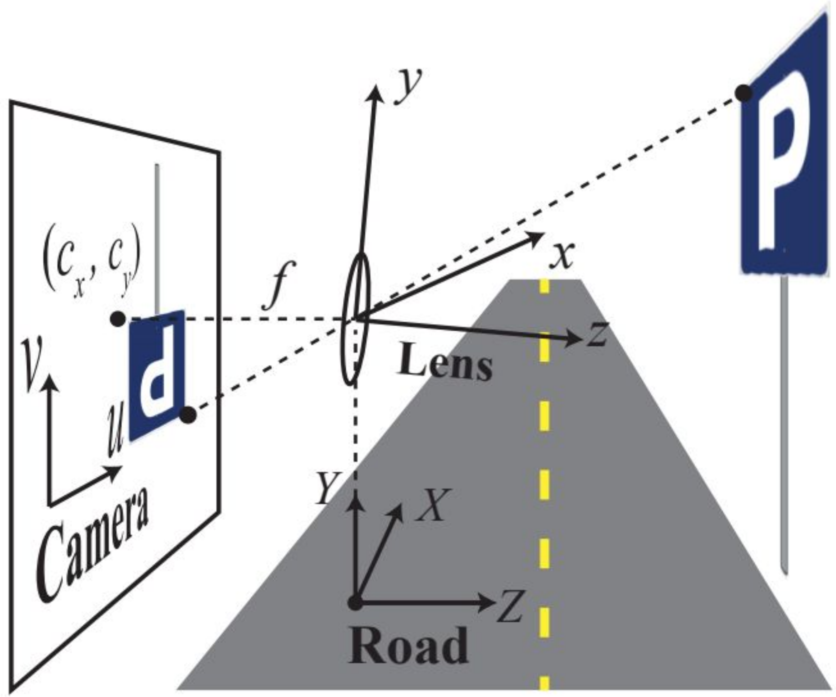


Localize each car in each frame. Map the location to lat-lon. Use clustering algorithm to infer the number of cars.



ParkMaster: An In-Vehicle, Edge-Based Video Analytics Service for Detecting Open Parking Spaces in Urban Environments.
Giulio Grassi, Paramvir Bahl, Kyle Jamieson, Giovanni Pau.

# Detecting Open Parking Spaces




Goal: Counting the number of cars parked on each road.

A car may appear in several consequent frames of the video.

Localize each car in each frame.
Use clustering algorithm to infer the number of cars.

ParkMaster: An In-Vehicle, Edge-Based Video Analytics Service for Detecting Open Parking Spaces in Urban Environments.
Giulio Grassi, Paramvir Bahl, Kyle Jamieson, Giovanni Pau.

# Street Numbers


Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks
Ian J. Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, Vinay Shet

# Street Numbers

We assume that the identities of the separate digits are independent from each other, so that the probability of a specific sequence $\mathbf{s} = s_1, \ldots, s_n$ is given by

$$P(\mathbf{S} = \mathbf{s}|X) = P(L = n \mid X)\Pi_{i=1}^{n} P(S_i = s_i \mid X).$$

Input Image    # of digits

Convolutional Neural Networks

P(S_i = s_i | X) for all i and s_i
P(L=n | X) for all n

$$\mathbf{s} = (l, s_1, \ldots, s_l) = \text{argmax}_{L, S_1, \ldots, S_L} \log P(S \mid X).$$

Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks
Ian J. Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, Vinay Shet

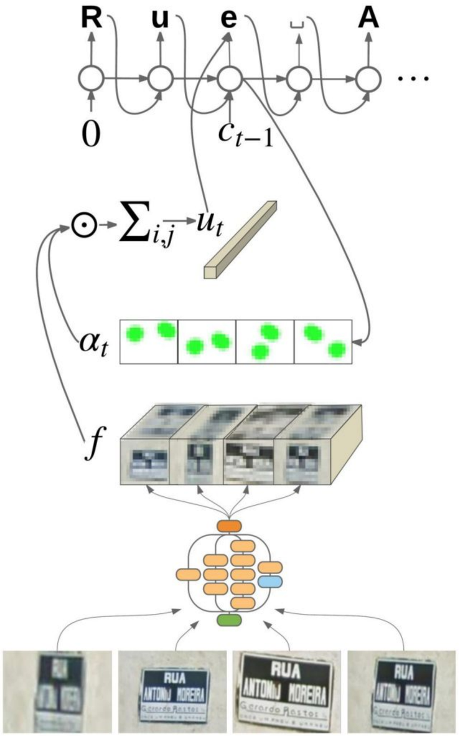
# Street Numbers

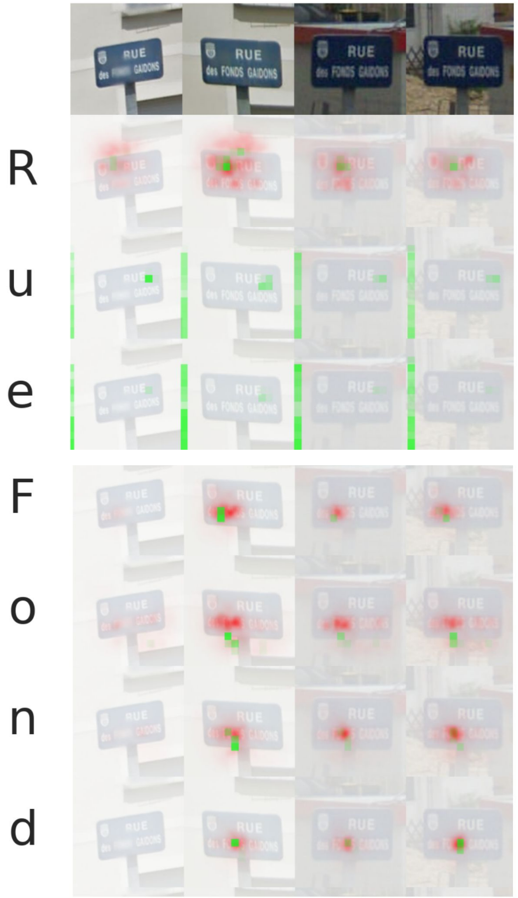

Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks
Ian J. Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, Vinay Shet

# Street Names


Attention-based Extraction of Structured Information from Street View Imagery
Zbigniew Wojna, Alex Gorban, Dar-Shyang Lee, Kevin Murphy, Qian Yu, Yeqing Li, Julian Ibarz

# Street Names



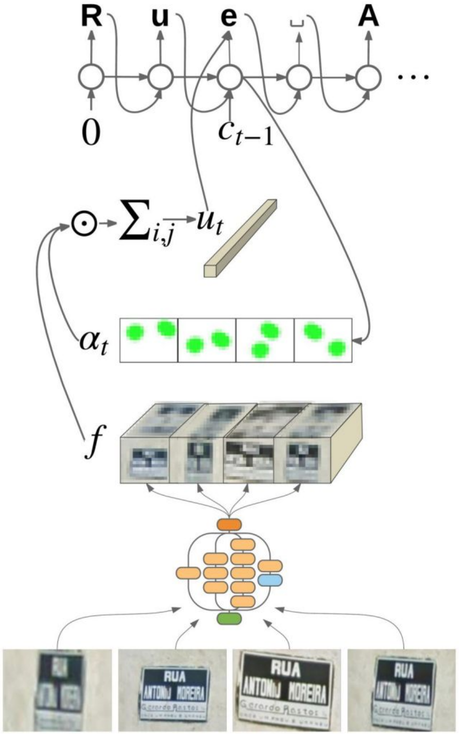

Attention-based Extraction of Structured Information from Street View Imagery
Zbigniew Wojna, Alex Gorban, Dar-Shyang Lee, Kevin Murphy, Qian Yu, Yeqing Li, Julian Ibarz

# Business Front Sign Detector


Large Scale Business Store Front Detection from Street Level Imagery
Qian Yu, Christian Szegedy, Martin C. Stumpe, Liron Yatziv, Vinay Shet, Julian Ibarz, Sacha Arnoud

# Business Front Sign Detector→Business Name Map

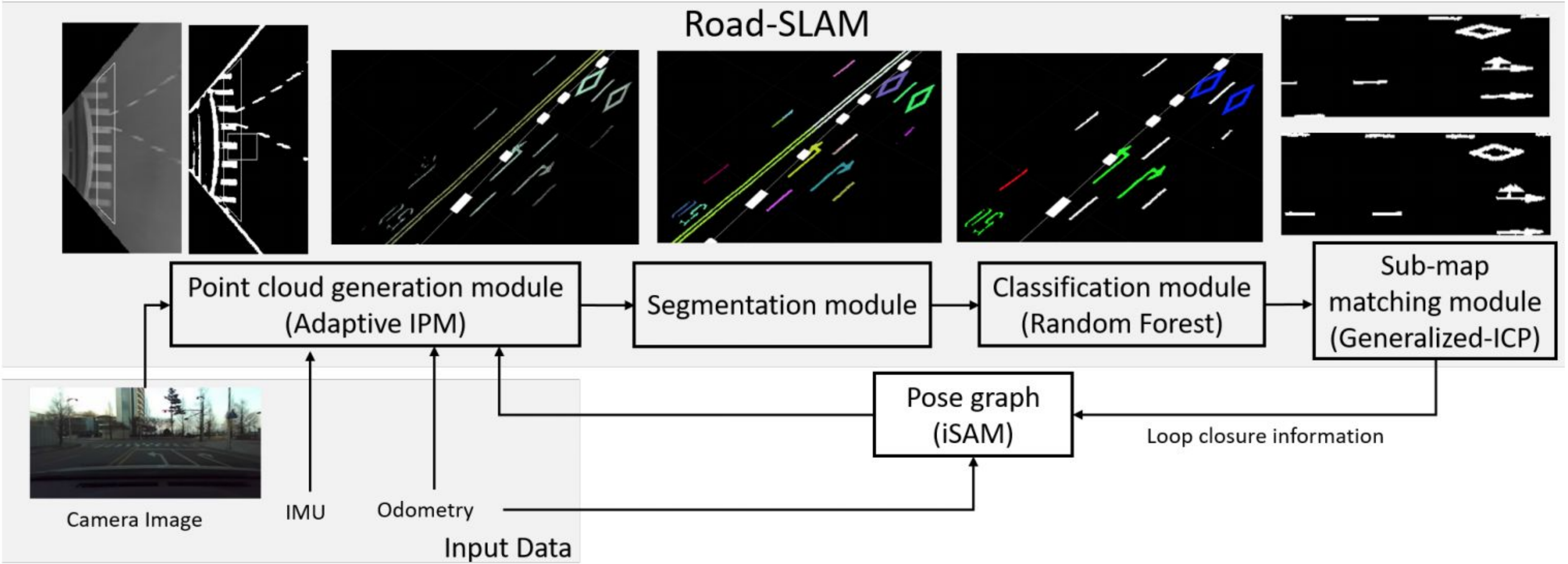

+




Large Scale Business Store Front Detection from Street Level Imagery

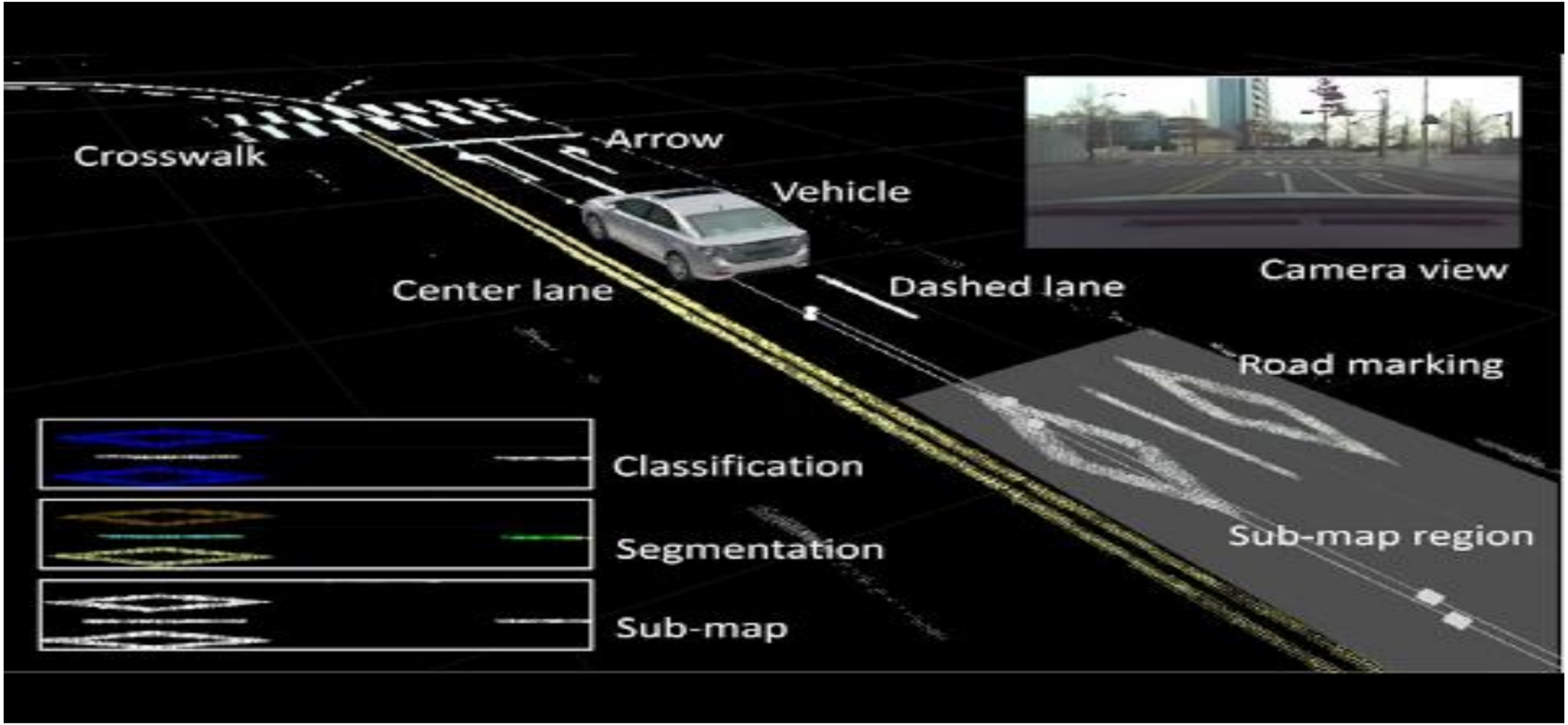
Large Scale Business Store Front Detection from Street Level Imagery
Qian Yu, Christian Szegedy, Martin C. Stumpe, Liron Yatziv, Vinay Shet, Julian Ibarz, Sacha Arnoud

# Road Marker Map




Road-SLAM : Road Marking based SLAM with Lane-level Accuracy

# Road Marker Map




Road-SLAM : Road Marking based SLAM with Lane-level Accuracy

# 3D Maps

https://blog.mapillary.com/tech/2017/07/26/improving-3d-reconstruction-with-semantic-understanding.html

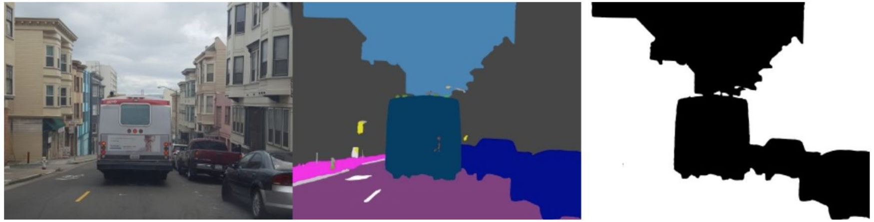
# 3D Maps - Remove moving objects

### An example

Let's take a look at an example. The sequence below shows nine images taken roughly ten meters apart along a straight road in San Francisco. The mapper was unfortunate to end up behind a bus that takes up a large part of the images.

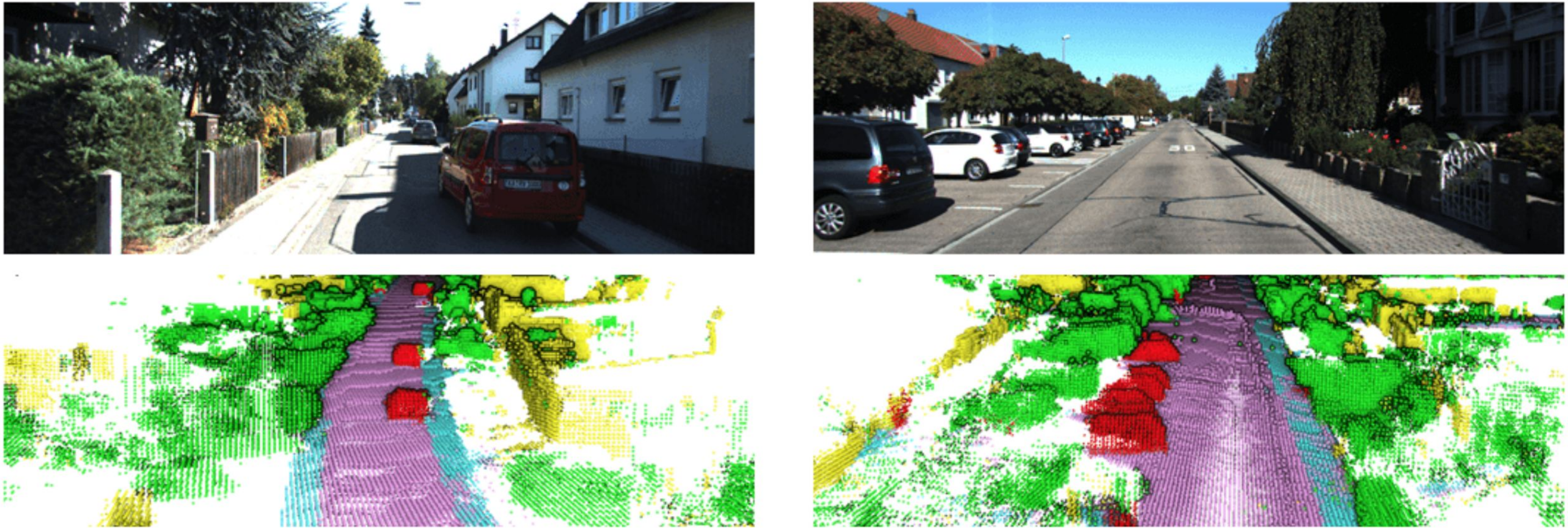
### The solution

**With the help of sematic understanding of the captured scene, it is possible to resolve the problem of most moving and temporary objects.** By making use of highly accurate machine-genereated semantic segmentations, we can determine which keypoints belong to potentially moving objects. We can then **do the reconstruction using only keypoints on static object classes** like buildings and roads. The result is a 3D reconstruction that depicts the important non-moving parts of the scene and the camera motion with respect to them.



*Creating a keypoint mask. Left: original sequence image. Middle: sematic segmentation for the image. Right: binary mask with white for included segmentation classes and black for excluded. Excluded classes are vehicle, pedestrian, and sky, among others.*

https://blog.mapillary.com/tech/2017/07/26/improving-3d-reconstruction-with-semantic-understanding.html



https://arxiv.org/pdf/1802.10271.pdf

# TODO: GPS + Dashboard Camera (Association)

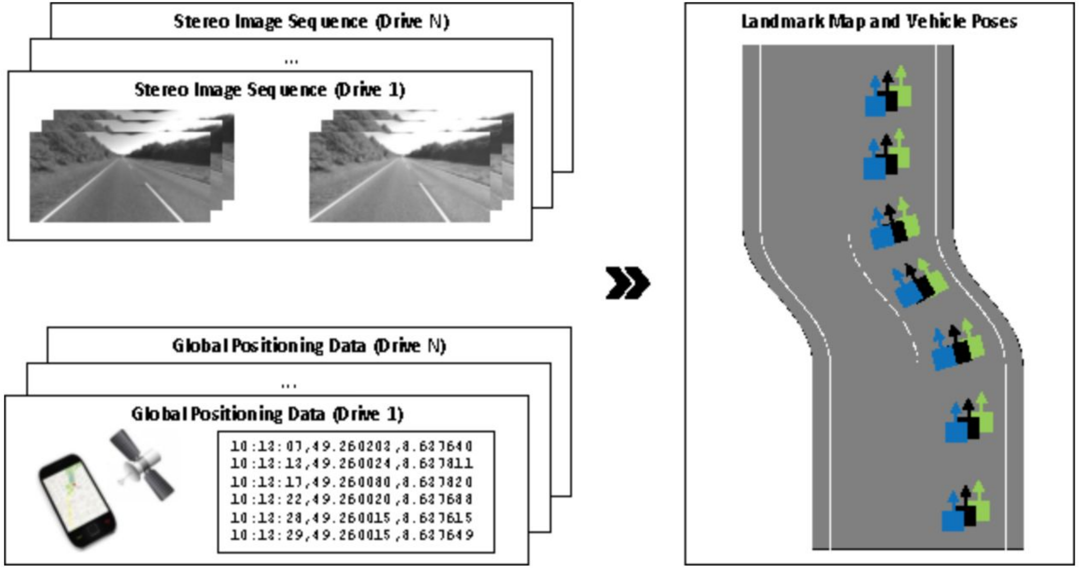
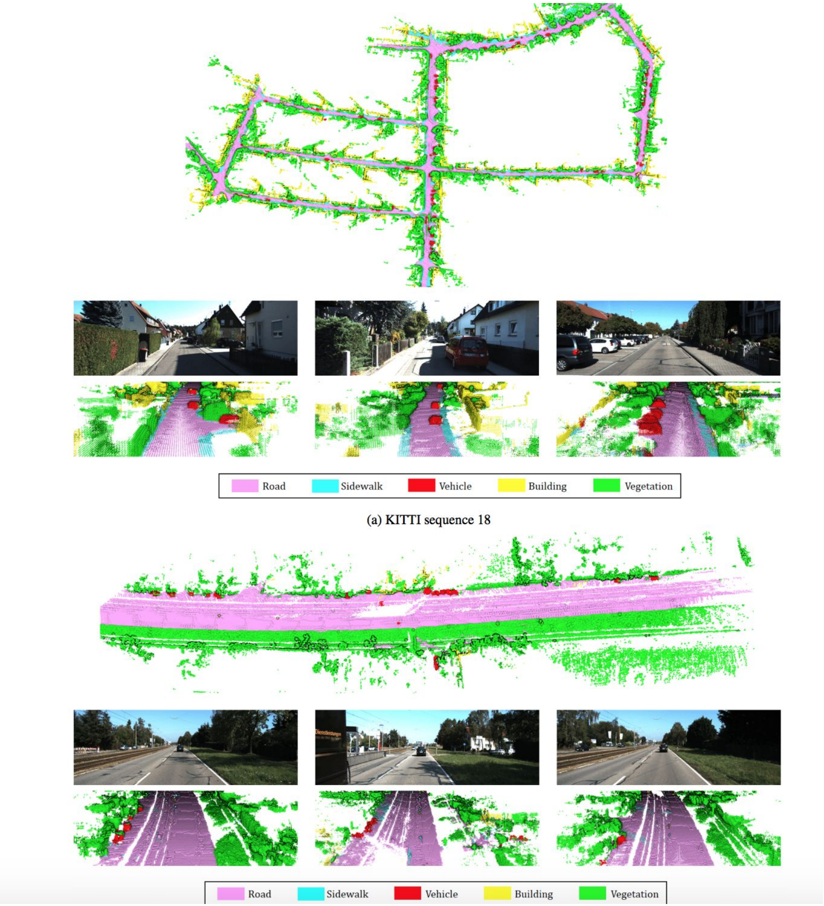



**Fig. 1. Overview of Input Modalities and Generated Output. For each drive, stereo imagery and GPS data is recorded.**

Schreiber, Markus et al. “Multi-drive feature association for automated map generation using low-cost sensor data.” 2015 IEEE Intelligent Vehicles Symposium (IV) (2015): 1140-1147.

# TODO: Camera + Lidar

https://arxiv.org/pdf/1802.10271.pdf




https://arxiv.org/pdf/1802.10271.pdf

# TODO: Other Papers

Monocular Visual Odometry and Dense 3D Reconstruction for On-Road Vehicles
https://pdfs.semanticscholar.org/52e9/4935e03936e2e9f1770fe5a960010f8576fe.pdf

Automatic lane-level map generation for advanced driver assistance systems using low-cost sensors

todo