Collaborative Mapping with Street-level Images in the Wild

Yubin Kuang

Co-founder and Computer Vision Lead

Mapillary

Mapillary is a street-level imagery platform, powered by collaboration and computer vision.

Mapillary data → SfM/3D reconstruct → Sign recognition → Object recognition → Map features → Map updates

Computer Vision

Map updates → OEMs/Map Providers

OEMs/Map Providers → Data

Data → Images

Images → Mapillary data

Collaboration - Image Capture

Collaborative mapping - Capture

Any device combined with automation can scale infinitely

Collaborative mapping generates fresh, diverse and global map data for HD Maps

Collaborative mapping - Computer Vision

Localization and Mapping

- Structure from Motion (SfM)
- Simultaneous Localization and Mapping (SLAM)
- Positioning and scale estimation

Recognition

- Object Recognition
  - Stationary objects
  - Moving objects
- Semantic Scene Understanding
  - Semantic relations between the map objects

Sensors: Monocular Camera

Redundancy: LiDAR, Radar, Stereo Camera,

Sensors: Monocular Camera, GPS

Redundancy: LiDAR, Radar, Stereo Camera, Compass, IMU

Monocular Camera + GPS

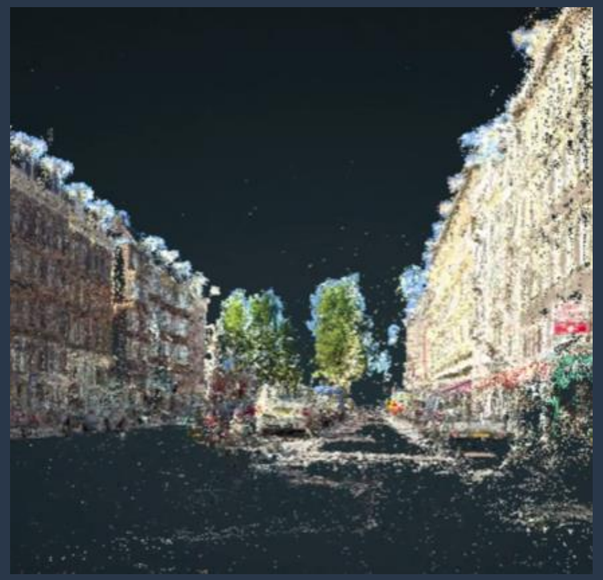
Key Components

Monocular Camera + GPS

Recognition

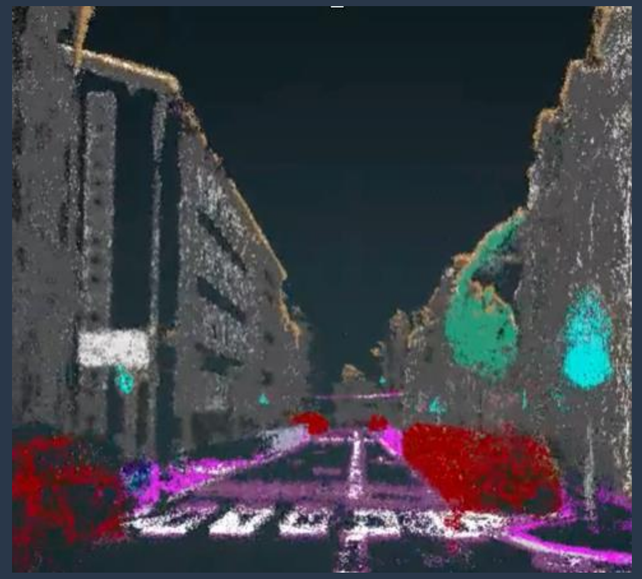
Object recognition

SfM



3D reconstruction

Map Data

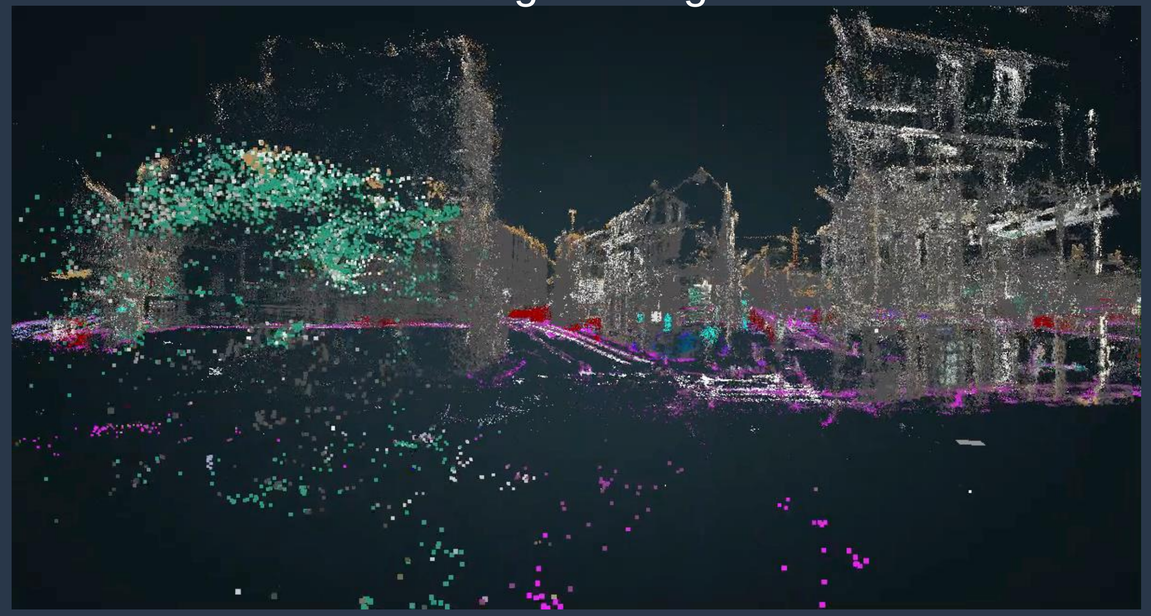


3D object extraction

Semantic Segmentation

Traffic Sign Recognition

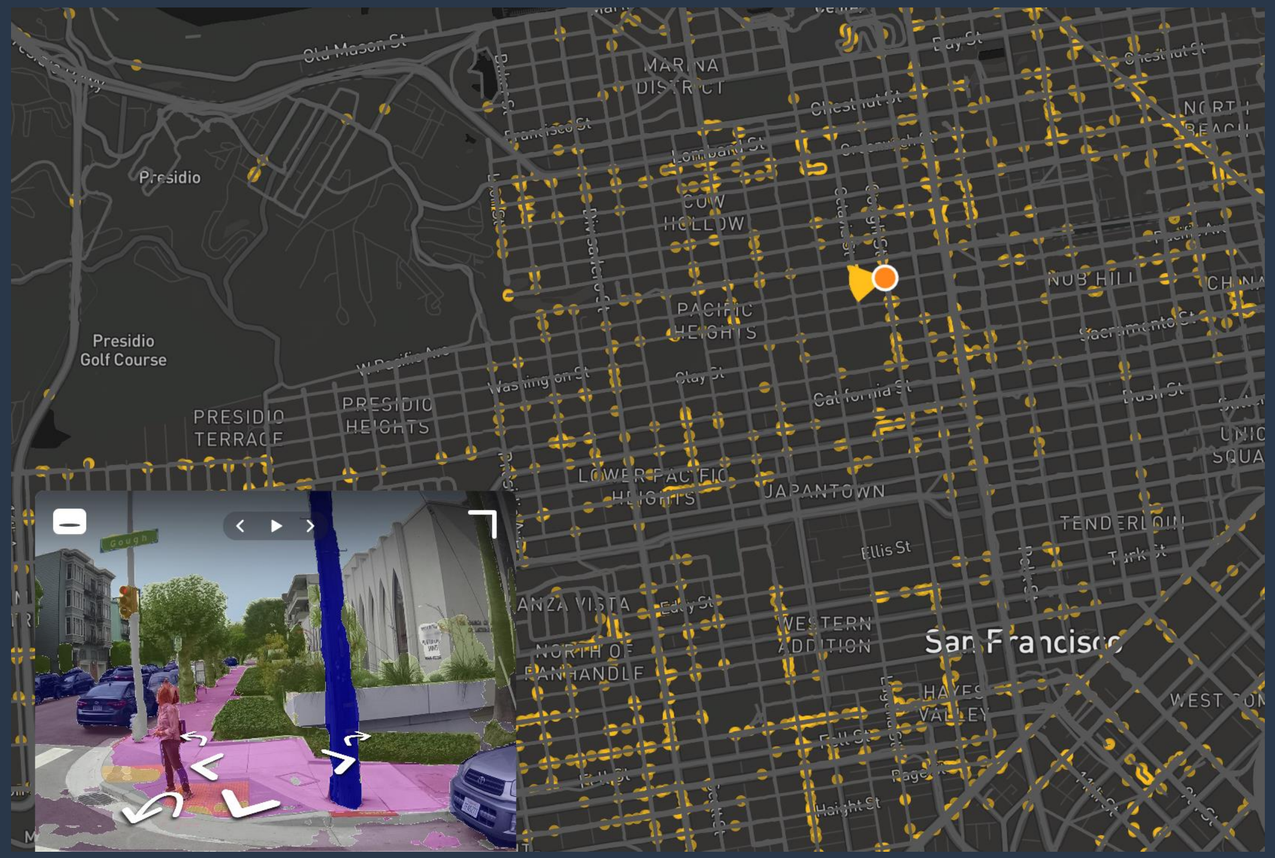
3D Point cloud



Semantic Point Cloud

Map Data - Visualization and API

Traffic Signs



Poles

Map data from 200M images accessible worldwide through API

Challenges and Solutions

Moving Objects

- Challenges:

  Differentiate between the ego motion and distractor motions in the scene

- Solutions:
- **Motion segmentation**: Identify motion clusters in the scene and recover ego motion
- **Moving object removal**: Semantically ignore moving objects in SfM

A moving bus in front of the camera

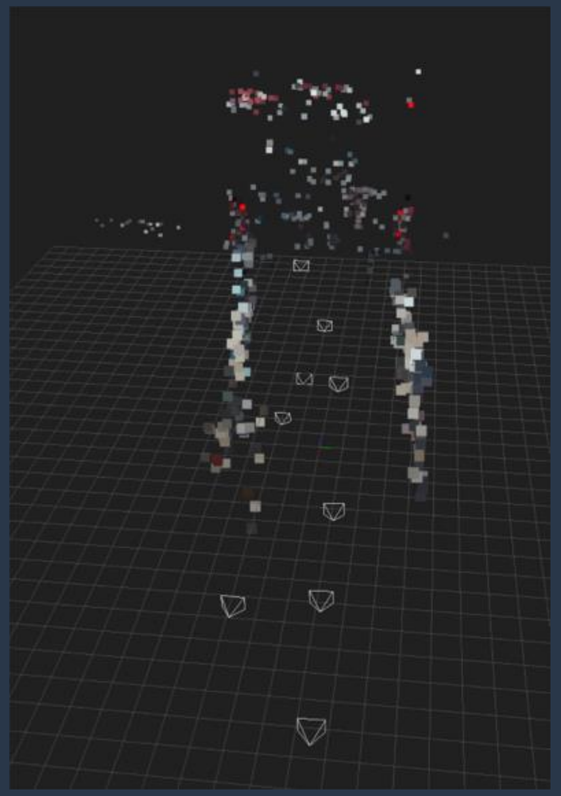
Moving Objects

Imag
e

Segmentation

Static vs.
Dynamic

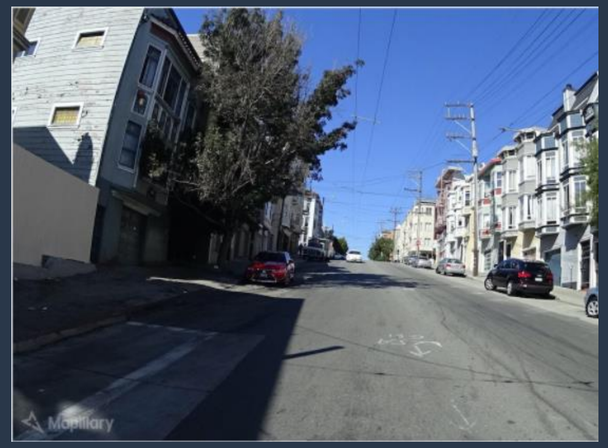
Removal of moving objects



Before

After

Camera Calibration



Action Cameras

Fisheye

Equirectangular (360)

Calibration:

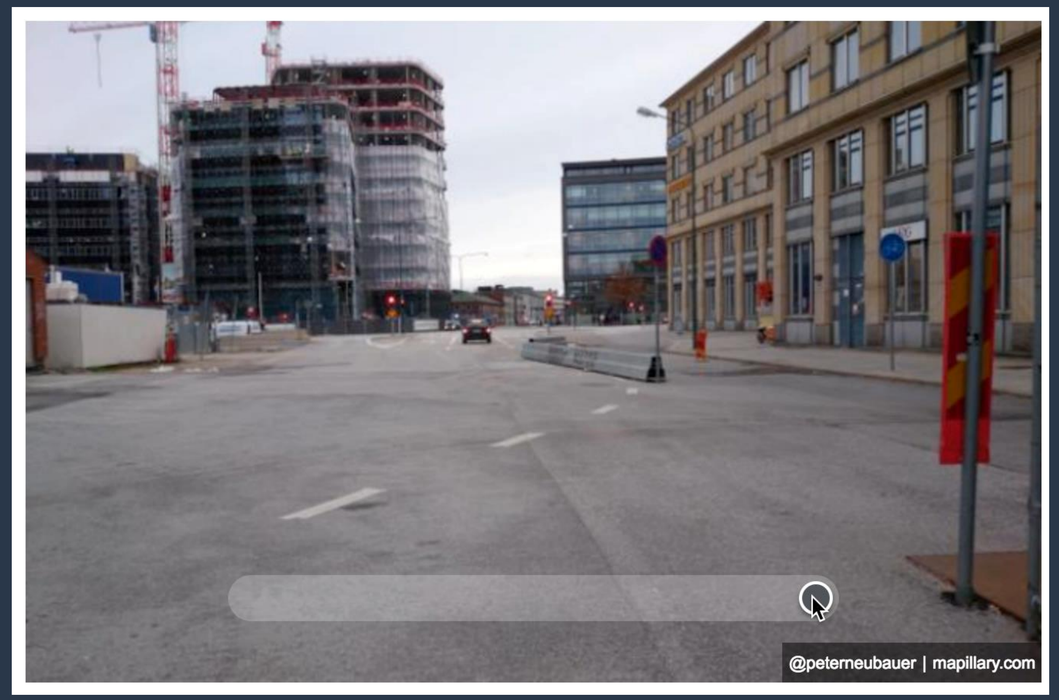
- Crowdsourced calibration
- Self-calibration with multiple images
- End-to-end self-calibration with CNN

Database:

- Build a database for camera intrinsics and projection models

Camera Calibration

Panorama to Perspective



Time Travel

Map Updates

- Challenges:
  - Traditional SfM pipeline is designed for static/batch processing
  - Map updates need to be scalable and consistent
- Solutions:
  - Stream processing architecture over batch processing
  - Robust local reconstruction alignments under varying imaging conditions
  - Distributed map updates given GPS (straightforward)
  - Handling boundary conditions

Annotations - Recognition

Cityscape Dataset

- 30 object classes
- 5K fine / 20K coarse annotations
- European cities
- Diverse weather/season
- Instance labels

Mapillary Vistas Dataset (MVD)

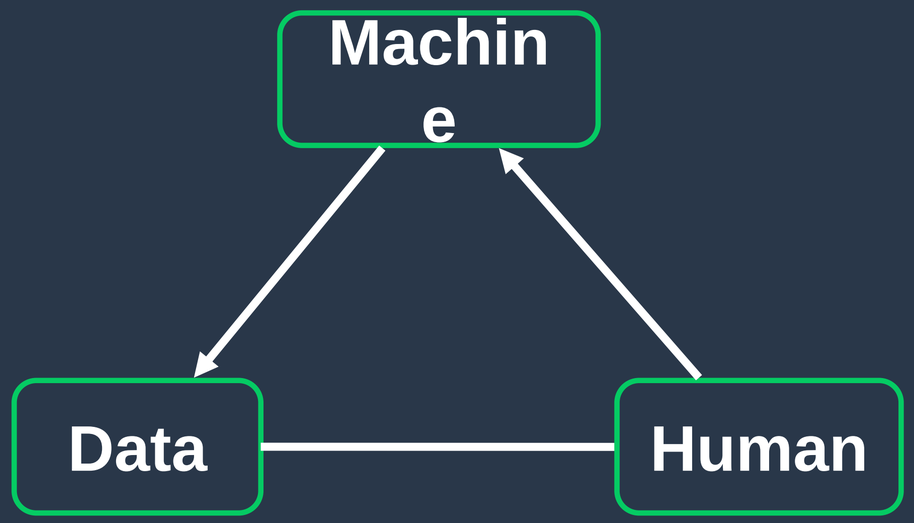
- 100 object classes
- 25K fine annotations
- 6 continents
- Diverse weather/season/cameras
- Instance labels

Neuhold et al. ICCV 2017 Mapillary

Annotations - Recognition

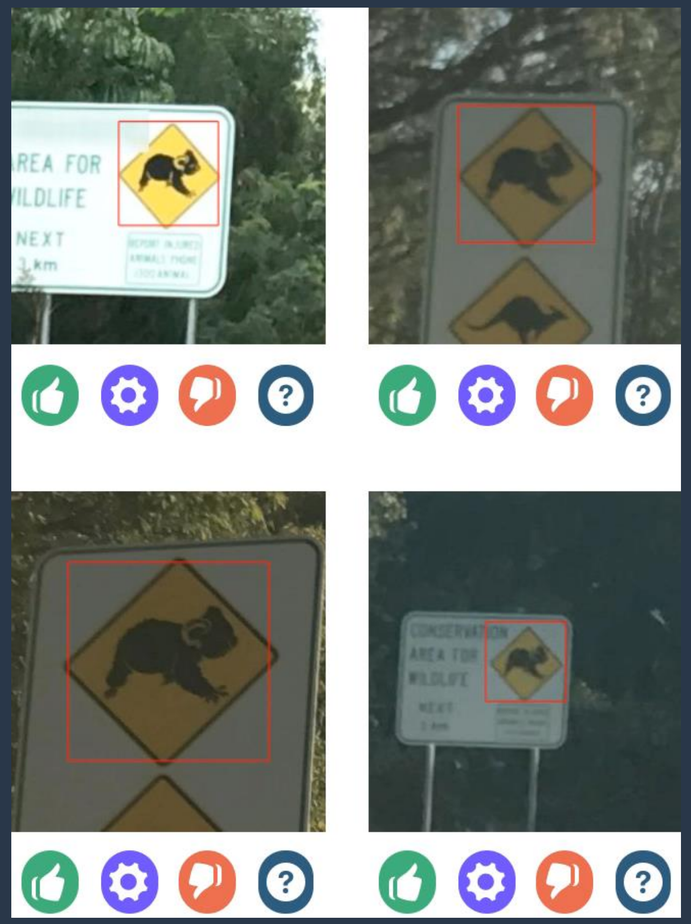
- Challenge:
  - Annotation is time-consuming in terms of **specification**, **annotations** and **QA**.
- Solutions:
  - Synthetic data
  - GAN for domain adaptation
  - Active learning
  - Semi-automatic annotation
  - Human in the loop

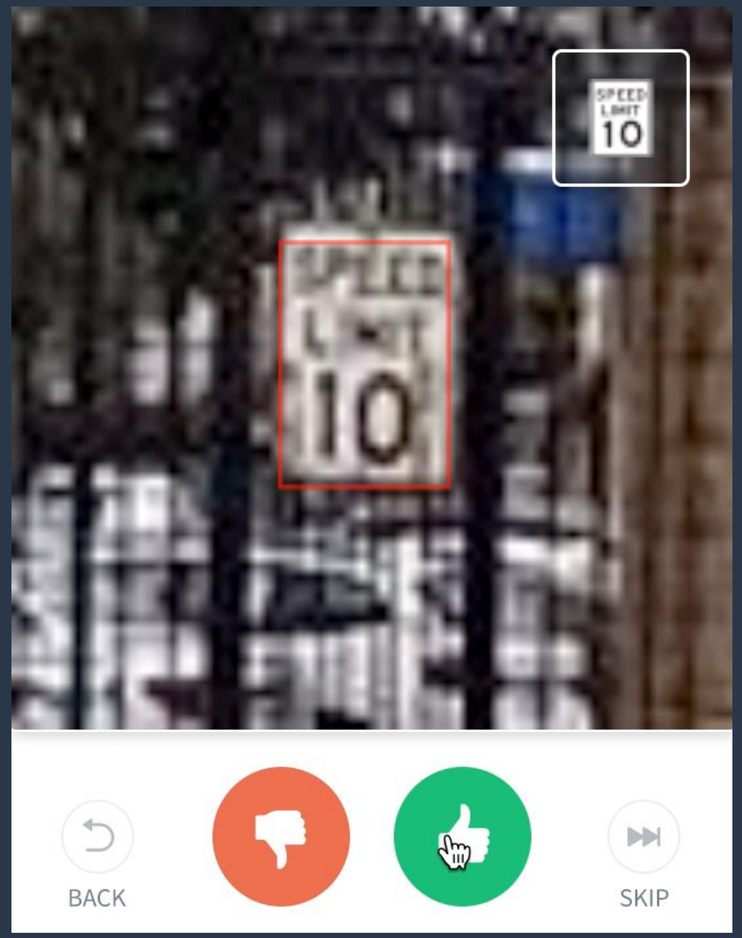
Annotations - Human in the loop



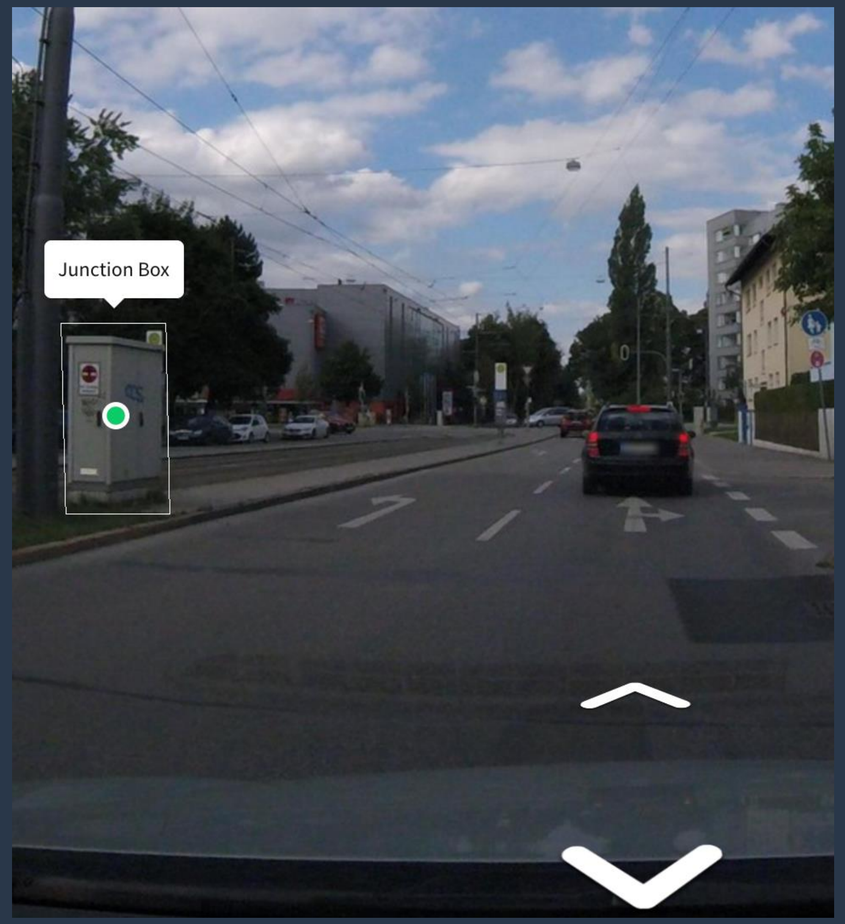
- Challenges:
  - Turnaround time from annotations to improvement of algorithms
  - Quality control is generally difficult with a large crowd of people
- Solutions:
  - Fully connected backend with automatic re-training
  - Work with the mapping community that understands and cares the quality of map data

Annotations - Human in the loop





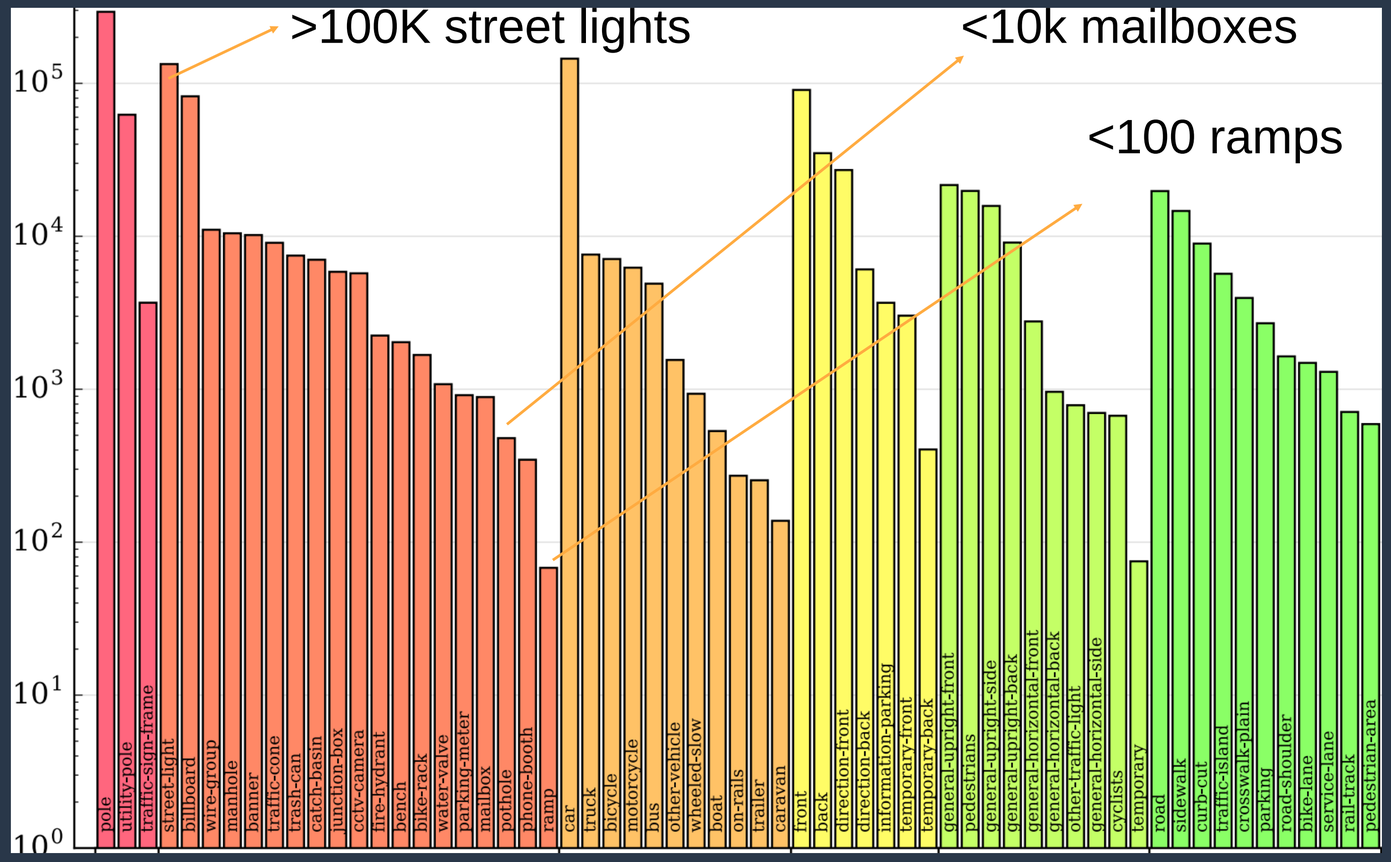
Machine detection to human verification



Tagging to machine detection

Rare Objects

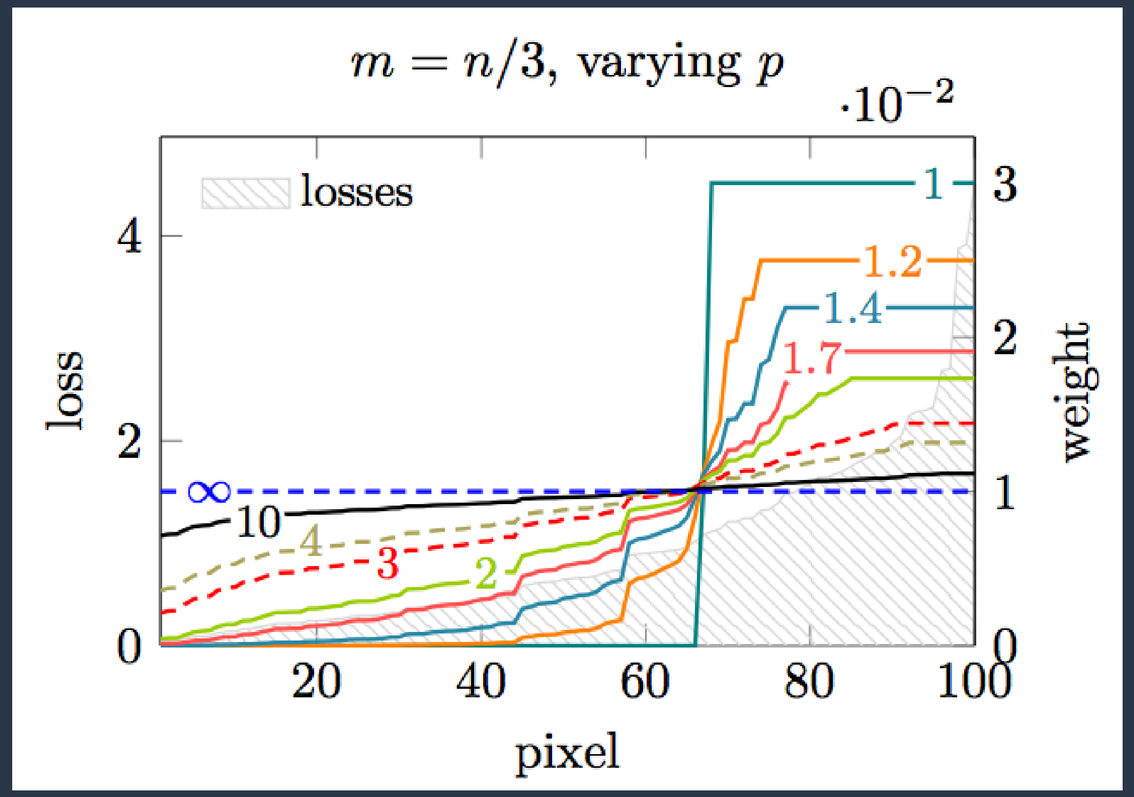
- Detecting rare objects (under-represented annotations) is key to the safety and map updates
- Long tail distribution for general objects on the road e.g. a koala on the road

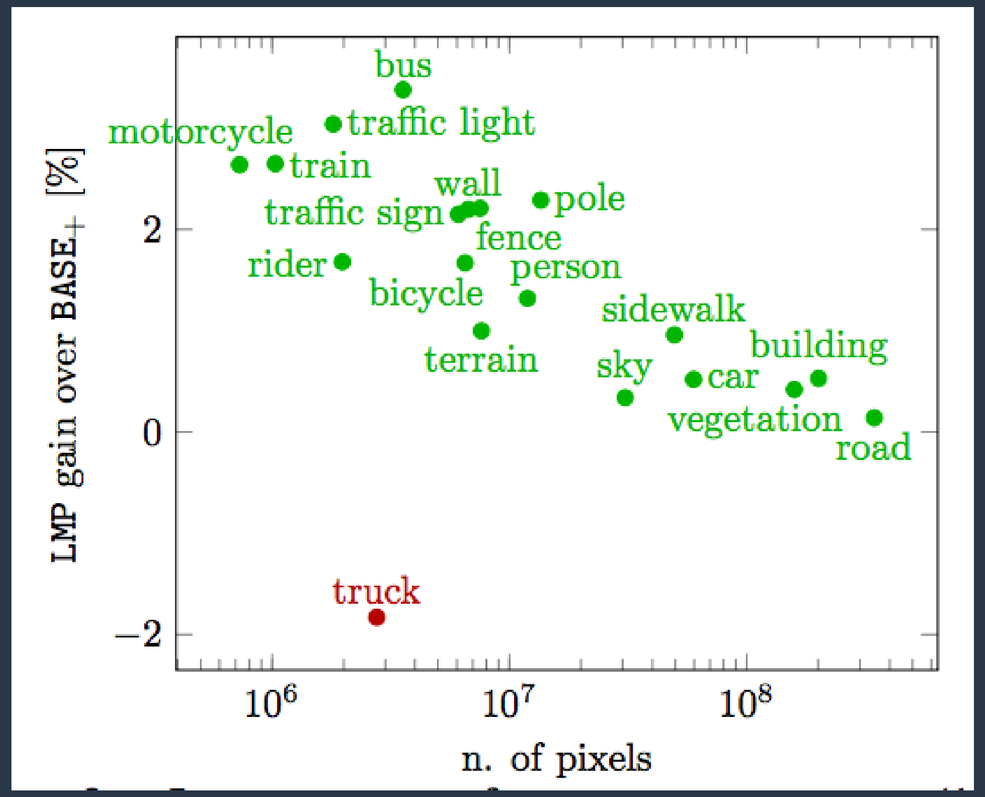


Number of instances for each object class in Mapillary Vistas Dataset

Rare Objects

- Use adaptive weighting in loss functions to boost performance for rare objects





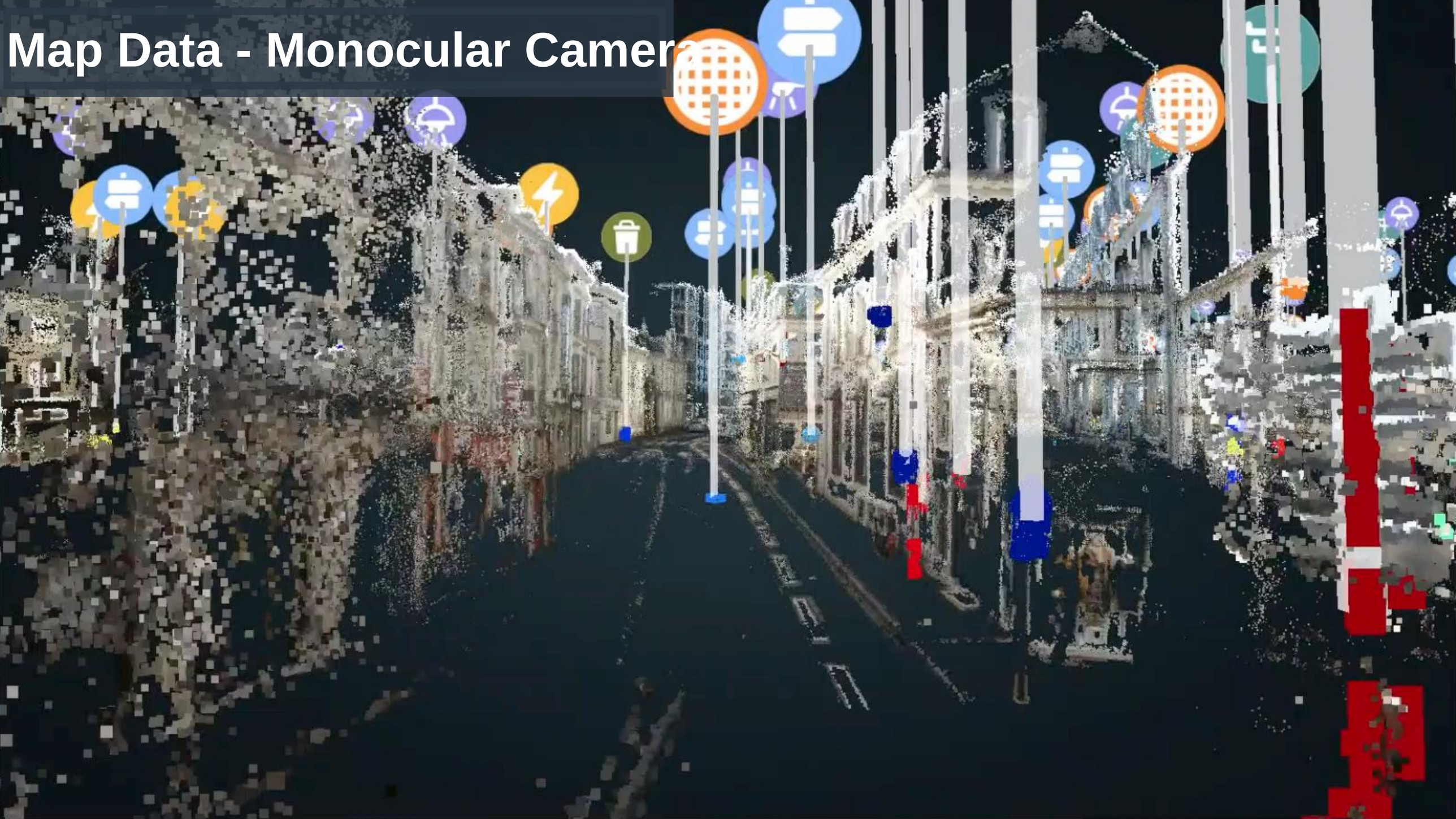
Loss Max-Pooling for Semantic Image Segmentation. Rota Bulò, Neuhold and Kontschieder CVPR 2017, **Mapillary**

Scaling

200 million Images
3.4 million km
15.6 billion objects
190 countries

- Challenges:
  - Constant and parallel updates
  - Serve billions of map features via API
  - Low latency and cost-effective processing
  - Time-consuming training
- Solutions:
  - Streaming processing over batch processing
  - Geo-Index and full-text search for map features
  - Optimized GPU processing in AWS ~$5K/100M images
  - In-house Titan-XP cluster significantly reduces training time

Map Data - Monocular Camera



Let’s map the world together!

To Date

200 million Images
3.4 million km mapped
15.6 billion objects
190 countries