JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

psarlin.com/orienternet

OrienterNet

Visual Localization in 2D Public Maps with Neural Matching

Paul-Edouard Sarlin[1] Daniel DeTone[2] Tsun-Yi Yang[2] Armen Avetisyan[2] Julian Straub[2]
Tomasz Malisiewicz[2] Samuel Rota Bulo[2] Richard Newcombe[2] Peter Kontschieder[2] Vasileios Balntas[2]

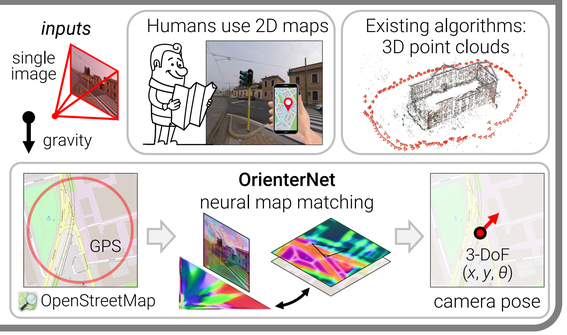
[1] ETH zürich [2] Meta Reality Labs

Overview

Task: camera pose estimation from a single image

Our approach: leverage 2D semantic maps like OpenStreetMap

Assumption: known gravity direction



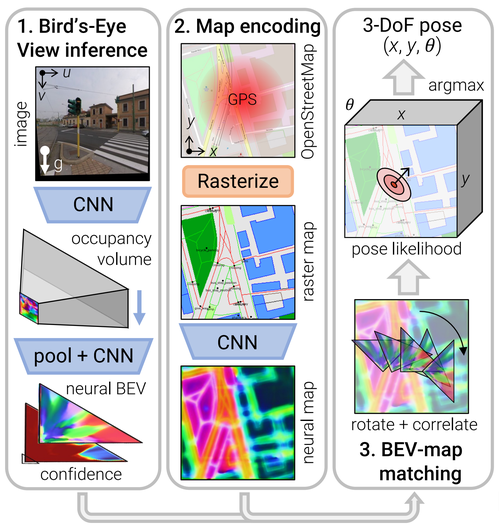
What are 2D semantic maps?

- Objects or surfaces with location and shape: **point, line, or polygon**
- Available for **free, globally**
- Rich semantic classes

Existing approaches

Map type	Storage (for 1 km²)	Privacy risk	Update	Accuracy
SfM /SLAM	high (+40 GB)	visual appearance	appearance changes	high ~0.1m
aerial /satellite	medium (75 MB)	overhead appearance	appearance changes	low >2m
OSM (ours)	compact (4.8 MB)	no PII (public info)	stable semantics, crowd-sourced	<1m

OrienterNet: end-to-end pose estimation



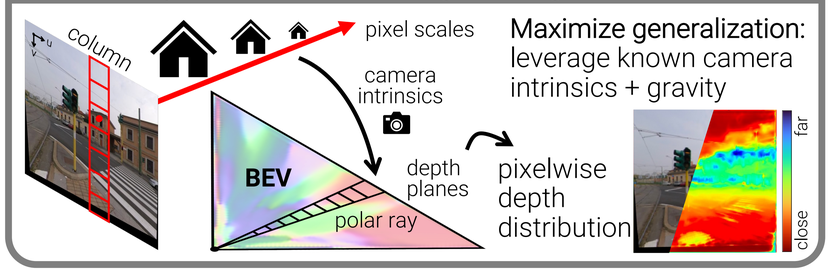
3-DoF pose estimation
- known gravity (IMU)
- motion is mostly 2D → height from SLAM

1. Infer a **mental map** (BEV) from the gravity-aligned image
2. Turn the vector map (polygon, line, points) into a **neural map**
3. **Exhaustive matching** for every x,y,θ on a grid → capture uncertainty (multi-modal pdf)

Loss: NLL given GT pose

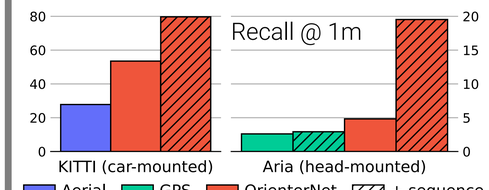
Training: 760k posed images from 12 cities in Europe+US, crowd-sourced from Mapillary.com: bikes, cars, hand-held

Bird's-Eye View from a single image



Maximize generalization: leverage known camera intrinsics + gravity

A single model for AR and robotics



Generalization to multiple datasets (only need OSM!)
- Semantic maps > aerial/satellite maps
- OrienterNet > GPS
- sequence > single image