Driven to Distraction: Self-Supervised Distractor Learning for Robust Monocular Visual Odometry in Urban Environments

Oxford Robotics Institute, University of Oxford, UK

Dan Barnes, Will Maddern, Geoffrey Pascoe, Ingmar Posner

UNIVERSITY OF OXFORD

Challenge

Robust urban visual odometry (VO) with only a monocular camera.

Limitations with Baseline VO :
- can **fail** with large moving distractor (**ephemeral**) objects
- often requires **expensive stereo** cameras

Addressed with Robust Mono VO :
- predict **ephemerality masks** to ignore distractors
- predict **disparity** to give scale with only a monocular camera

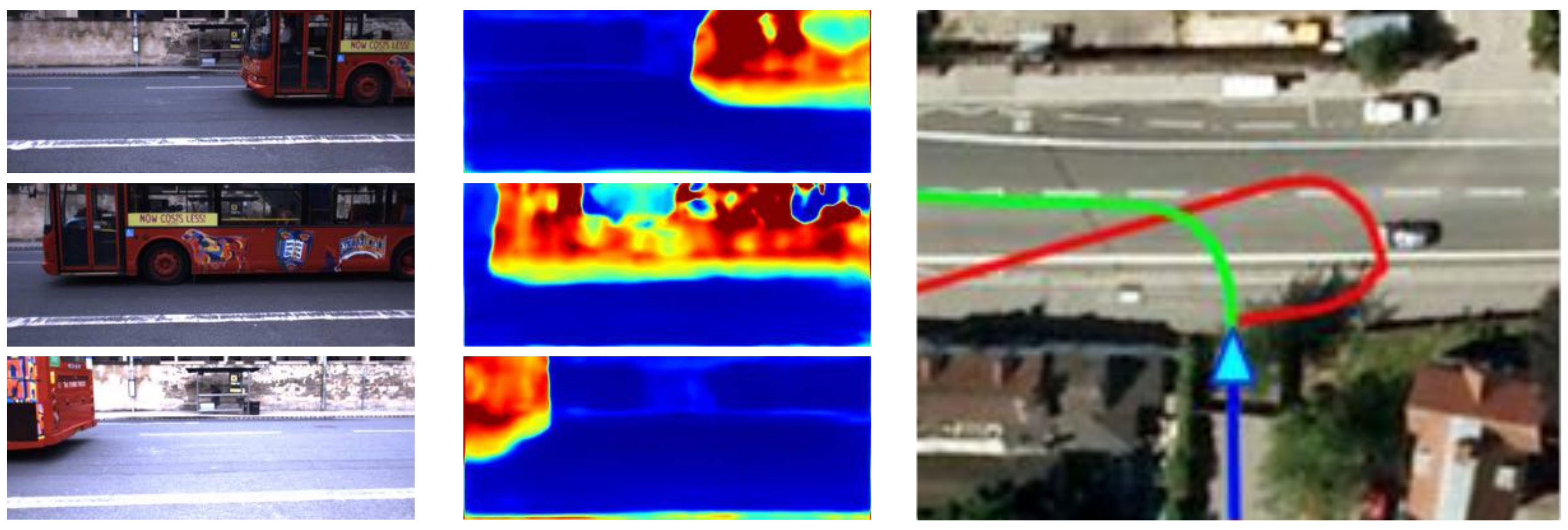
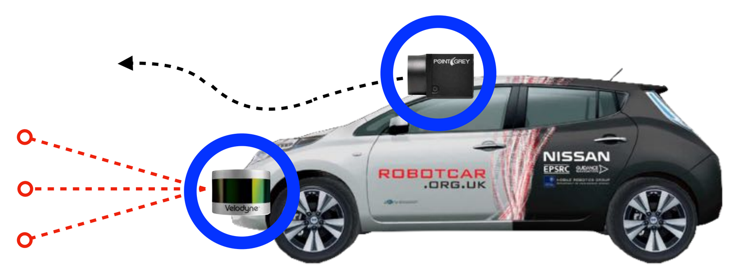


Image Sequence | Ephemerality Mask | Baseline VO | Robust Mono VO

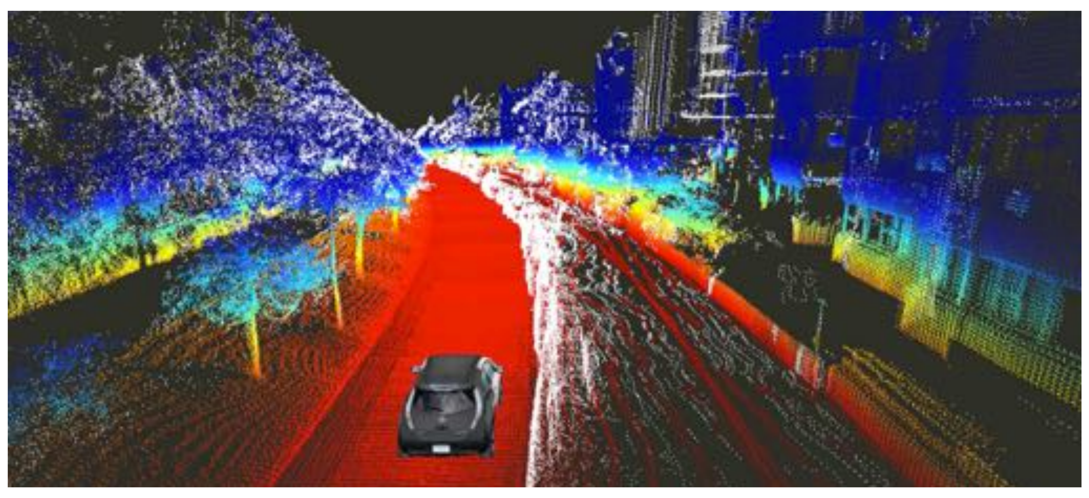
Training

1. Prior Mapping

Collect **numerous overlapping traversals** with a **stereo camera** and **LIDAR scanner.**



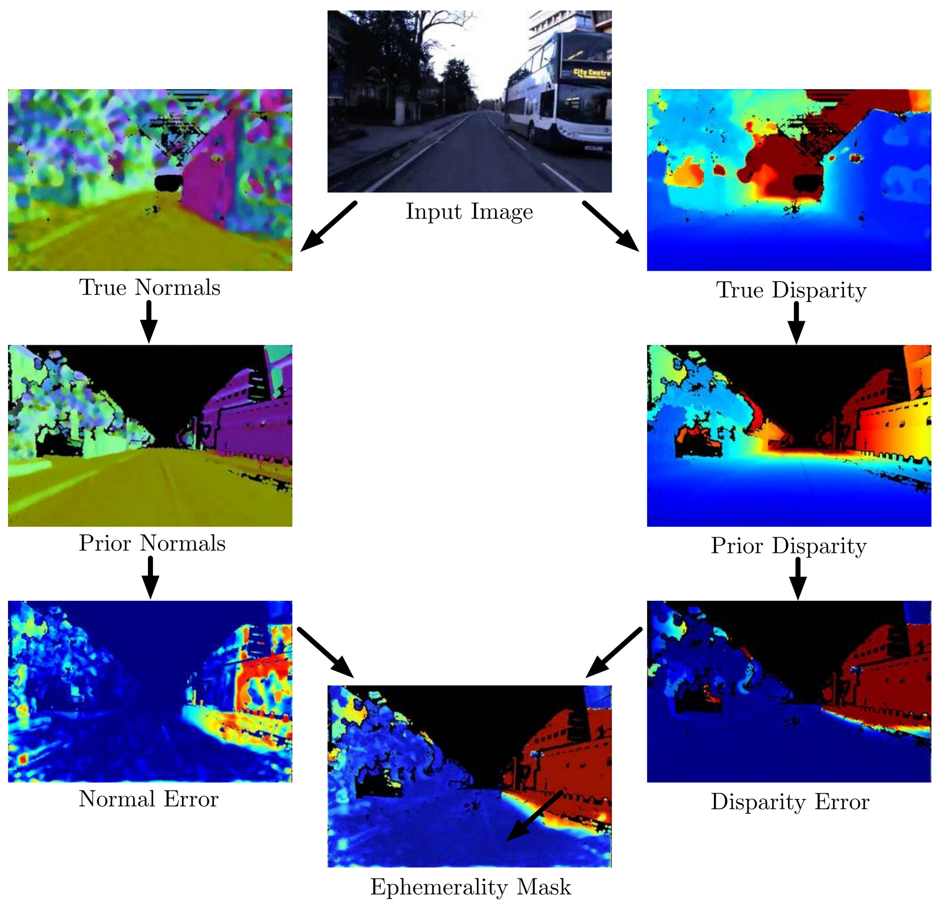
Remove **ephemeral points** (white) with an entropy-based, multi-session mapping approach.



2. Ephemerality Labelling

Automatically, with **no manual labelling**, compute reference **ephemerality masks**, $\mathcal{E}$, as a weighted difference between the static and true disparity, d, and normals, $\mathbf{n}$:

$$\mathcal{E}_i = \gamma \left\| d_i^S - d_i \right\|_1 + \delta \cos^{-1}\left(\mathbf{n}_i^S \cdot \mathbf{n}_i \right)$$



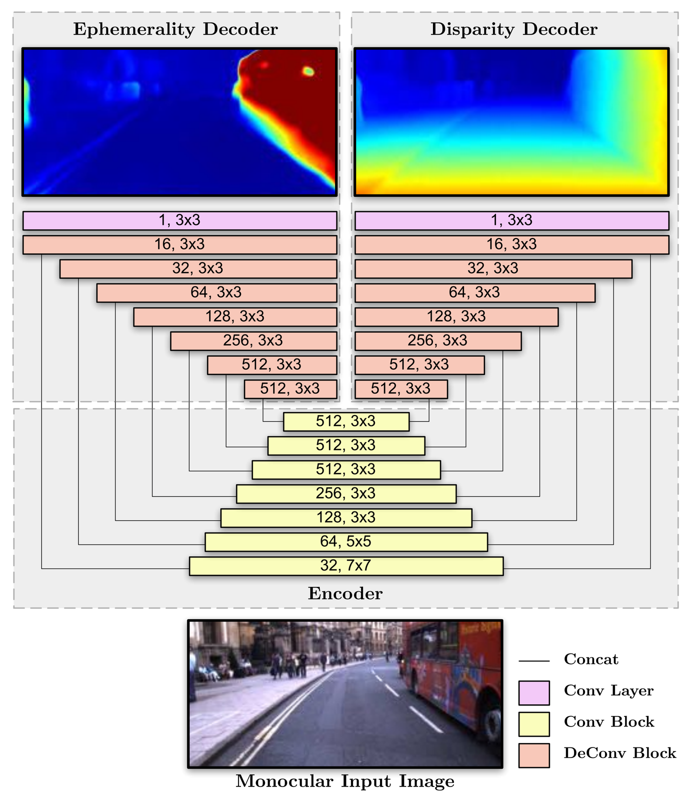
3. Network Training

Train a convolutional neural network to predict **ephemerality masks** and **disparity** from only monocular input images.

$$L = L_{ephemerality} + L_{disparity}$$

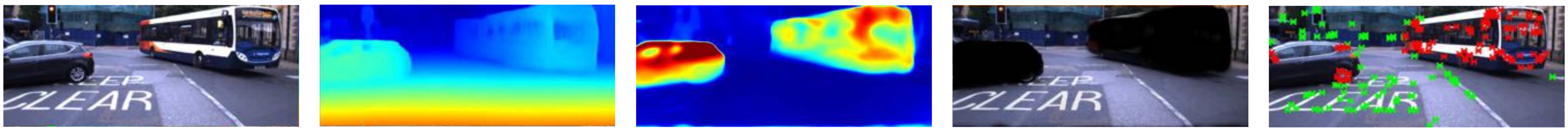
$$L_{ephemerality} = \|\hat{\varepsilon} - \varepsilon\|_1$$

$$L_{disparity} = \sum \alpha L_{recon} + \beta L_{smooth} + \gamma L_{SSIM}$$



Deployment

At run-time:
- **only requires a monocular camera**
- **operates in real-time** (CNN at 50Hz)
- **ephemerality** disables features in Sparse VO
- **ephemerality** weights photometric residuals in Dense VO
- **model-free distractor segmentation**
- substrate for **dynamic obstacle detection**



Input Image | Disparity | Ephemerality Mask | Robust Dense VO | Robust Sparse VO

Results

Over 400km of the Oxford RobotCar Dataset, we demonstrate:

- **reduced odometry drift**
- **significantly improved egomotion with large distractors**

Ephemerality mask predictions in challenging urban environments

Conclusion

Introduced **ephemerality masks**, the likelihood that a pixel corresponds to distractor objects.

Automatic self-supervised approach with **no manual labelling.**

Only requires a **single monocular camera** to produce **real-time, reliable ephemerality-aware visual odometry to metric scale.**

More Information

A²I APPLIED ARTIFICIAL INTELLIGENCE LAB
OXFORD ROBOTICS INSTITUTE