Addressing Drift and Overfitting in Deep Visual Odometry

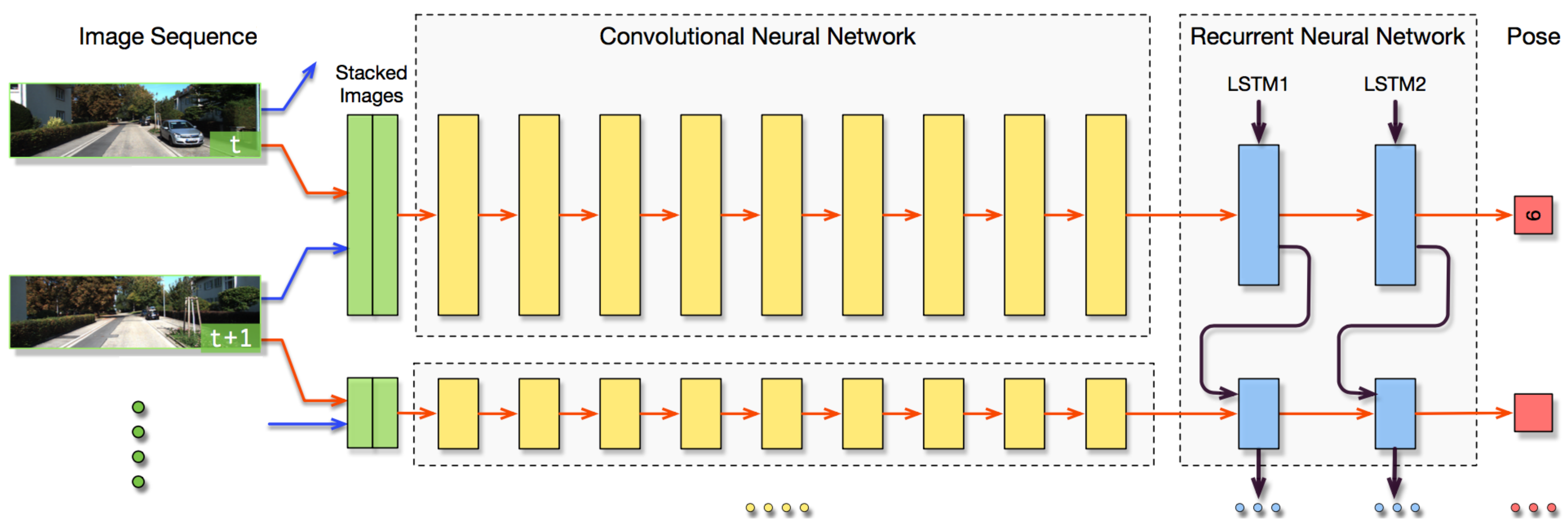
Fabian B. Fuchs, Alex Bewley, Ingmar Posner
Applied AI Lab, Oxford Robotics Institute, University of Oxford

Visual Odometry (VO)
means estimating an agent's ego-motion from an image sequence captured with a camera attached to the agent.

Deep Mono VO
is an end-to-end approach attacking this problem with deep learning for mono camera setups.

Our Contribution
First, we address drift in VO by introducing a global component to the loss function. Second, we develop the concept of stethoscopes to reduce overfitting.

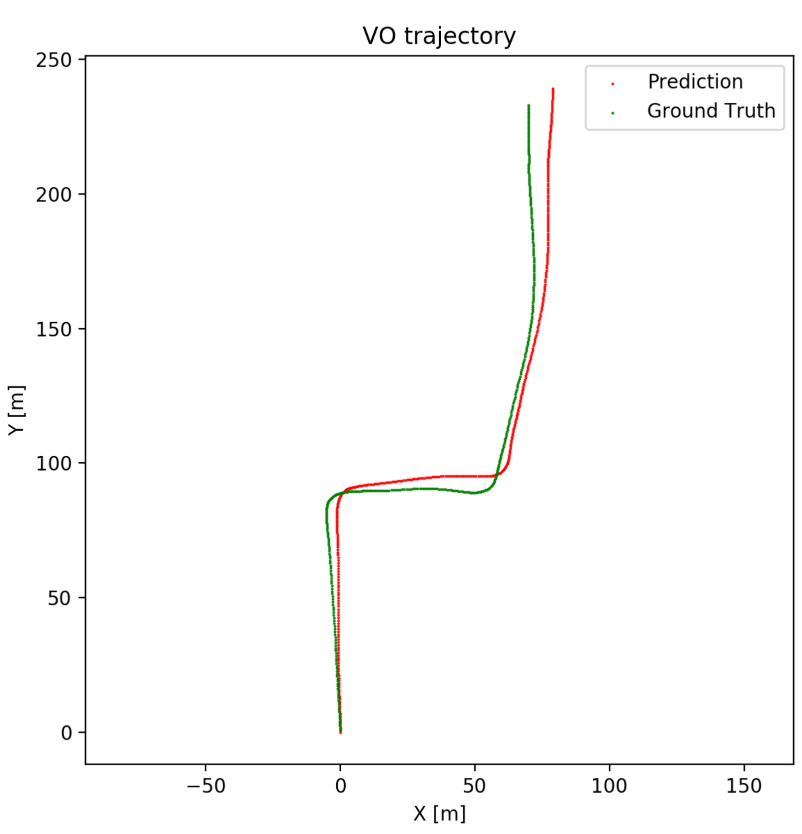
Architecture



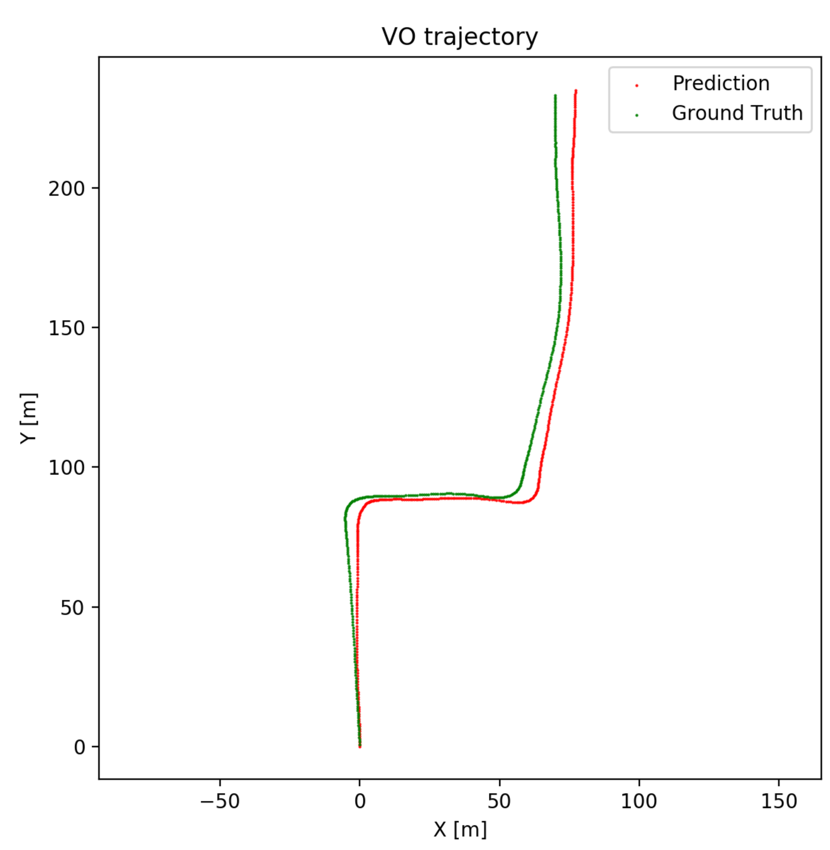
Following the architecture of Wang et al.[1], the network is made up of a CNN pre-trained for optical flow[2] followed by an RNN.

Global Loss Function

Only Local Loss:



With global Loss:



	only local loss	with global loss
Error (displacement over traveled distances)	3.8%	2.6%

Drift - the accumulation of local errors - is an inherent problem of visual odometry. We found that adding a ***global*** component to the ***loss*** function which penalises drift improves the accuracy.

- Local loss measures the accuracy of frame-to-frame pose changes:

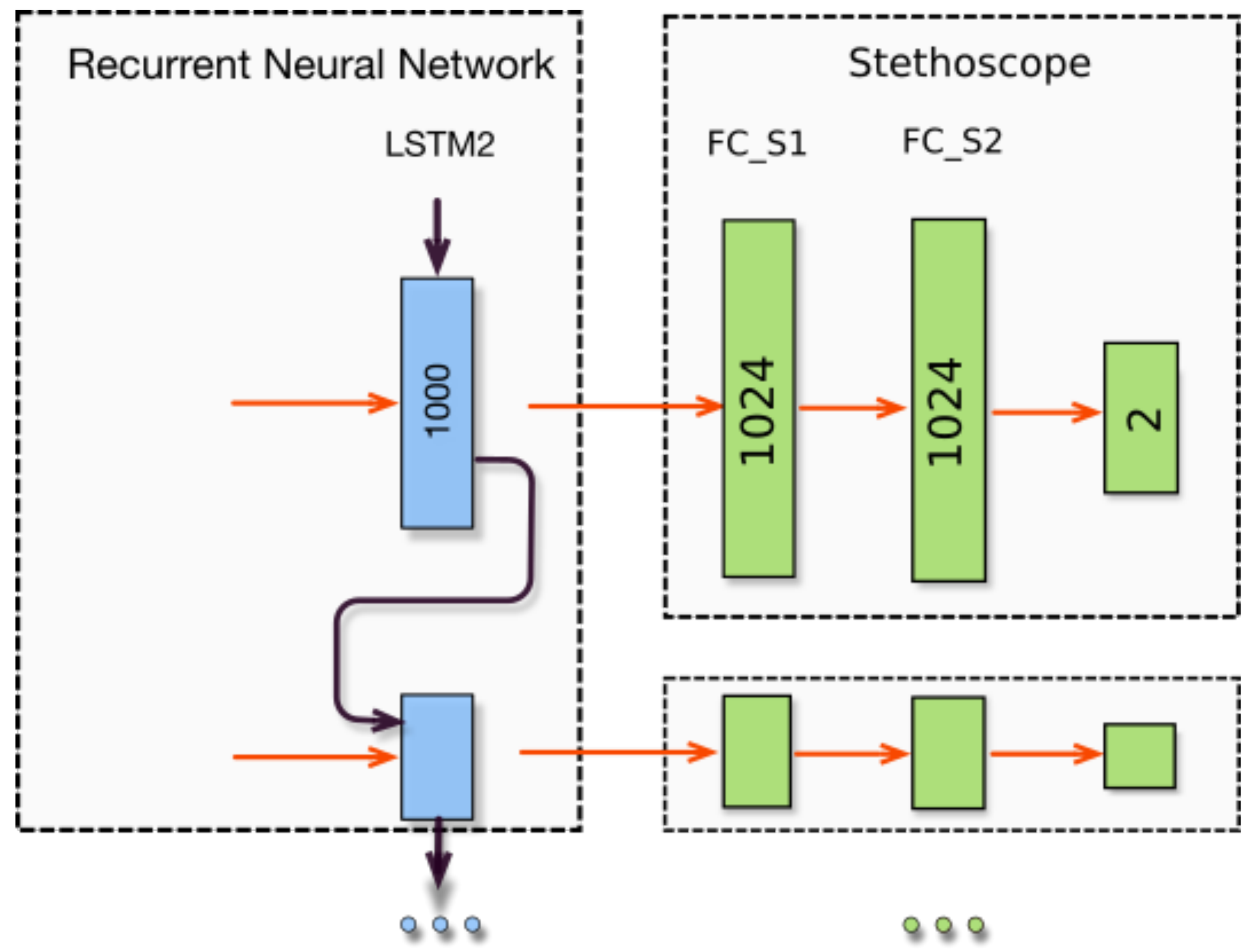
$$\mathcal{L}_{\text{VO-local}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{t}\left(\|\hat{\boldsymbol{p}}_k - \boldsymbol{p}_k\|_2^2 + \beta\|\hat{\boldsymbol{\varphi}}_k - \boldsymbol{\varphi}_k\|_2^2\right)$$

- Global loss measures the absolute displacement at each point:

$$\mathcal{L}_{\text{VO-global}} = \frac{1}{N}\sum_{i=1}^{N}\left(\|\hat{x}_{abs} - x_{abs}\|_2^2 + \|\hat{y}_{abs} - y_{abs}\|_2^2\right)$$

Where p and φ denote the relative translation and rotation between frames and (x,y) is the agent's absolute position.

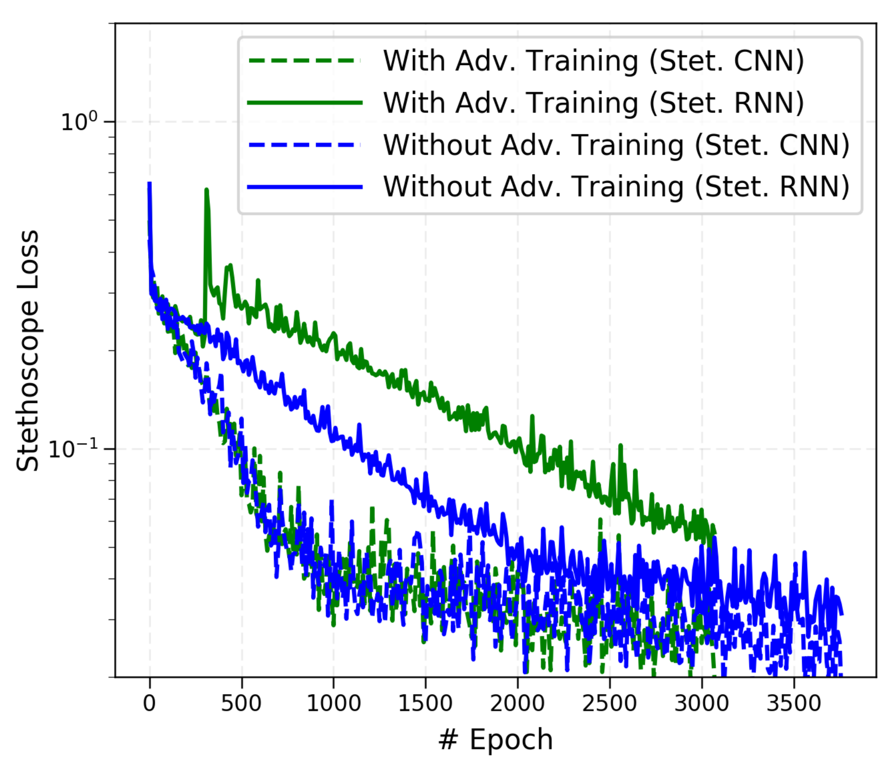
Stethoscopes

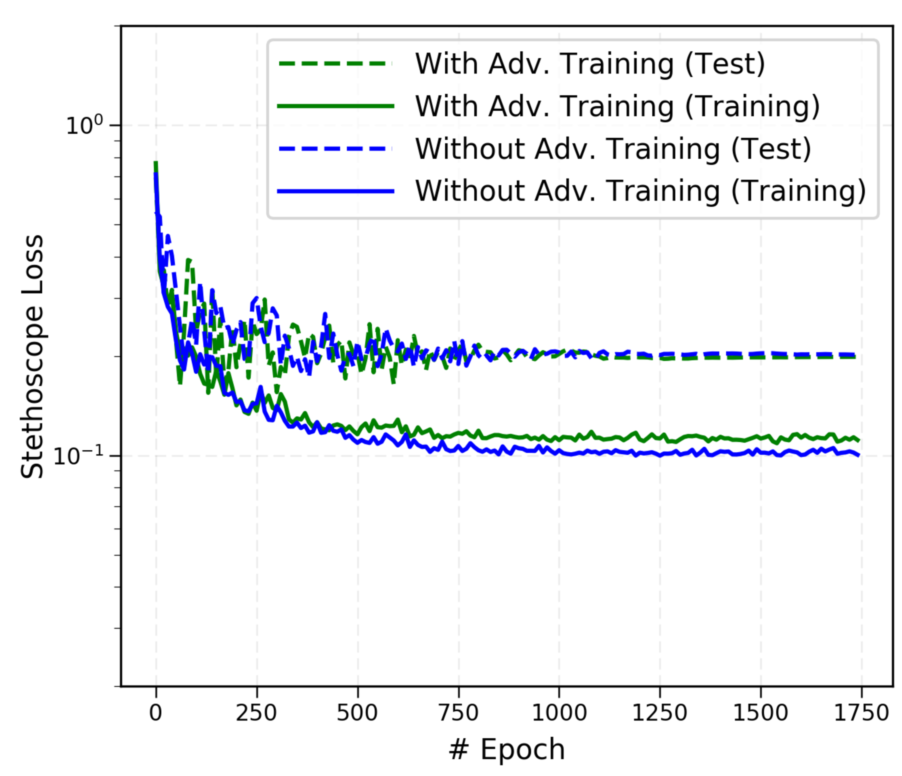


Our Hypothesis: VO algorithms ***overfit*** by extracting location-specific information from the input. E.g.: "Passing the blue house, the agent has a speed of 30mph."

With ***Stethoscopes***, we introduce a mechanism for assessing the amount of location-specific information at any part of the network. They can be attached to any layer of the network and are trained on the task of inferring the absolute position. The inspiration for this technique was taken from[3].

Adversarial Training





- We treat the ***stethoscope*** attached to the RNN ***as an adversary*** and add a confusion loss to the main network.
- This ***penalises*** the network for extracting ***location-specific*** information.
- The results show ***less overfitting*** (reduced gap between training and test performance) and a minor increase in test performance.

Acknowledgements

We would like to thank Ronnie Clark, recent graduate from the Sensor Networks Group, for sharing his valuable insights and his generous help to implement the DeepVO algorithm from[1].

[1] S. Wang, R. Clark, N. Trigoni, *DeepVO: Towards End-to-End Visual Odometry with Deep Recurrent Convolutional Neural Networks*. ICRA, 2017.
[2] P. Fischer, V. Golkov et al., *FlowNet: Learning Optical Flow with Convolutional Networks*. ICCV, 2015.
[3] P. Mirowski, R. Hadsell et al., *Learning to Navigate in Complex Environments*. CoRR, 2016.