Deep Probabilistic Regression of Elements of SO(3) using Quaternion Averaging and Uncertainty Injection

Valentin Peretroukhin, Brandon Wagstaff, and Jonathan Kelly
University of Toronto
v.peretroukhin@mail.utoronto.ca, brandon.wagstaff@mail.utoronto.ca, jkelly@utias.utoronto.ca

Abstract

Consistent estimates of rotation are crucial to vision-based motion estimation in augmented reality and robotics. In this work, we present a method to extract probabilistic estimates of rotation from deep regression models. First, we build on prior work and develop a multi-headed network structure we name HydraNet that can account for both aleatoric and epistemic uncertainty. Second, we extend HydraNet to targets that belong to the rotation group, SO(3), by regressing unit quaternions and using the tools of rotation averaging and uncertainty injection onto the manifold to produce three-dimensional covariances. Finally, we present results and analysis on a synthetic dataset, learn consistent orientation estimates on the 7-Scenes dataset, and show how we can use our learned covariances to fuse deep estimates of relative orientation with classical stereo visual odometry to improve localization on the KITTI dataset.

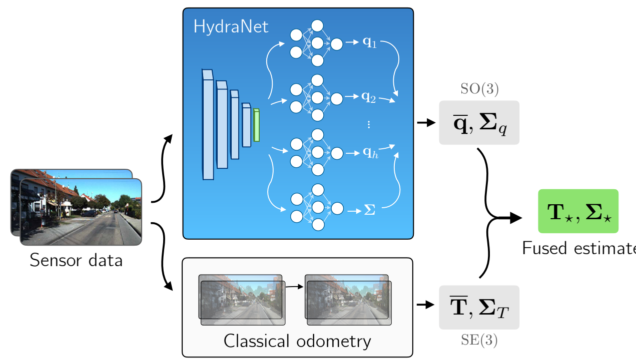

Figure 1: Our model outputs a probabilistic representation of rotation as a unit quaternion and a full covariance matrix. A consistent uncertainty estimate allows us to improve classical estimators using pose-graph optimization.

1. Introduction

Accounting for position and orientation, or pose, is at the heart of many computer vision algorithms like visual odometry, structure from motion, and SLAM. By tracking motion, these algorithms form the basis of visual localization pipelines in autonomous ground and aerial vehicles, visual mapping, and augmented reality.

Recent work [4, 15, 12] has attempted to transfer the success of deep neural networks in many areas of computer vision to the task of camera pose estimation. These approaches, however, can produce arbitrarily poor pose estimates if sensor data differs from what is observed during training (i.e., it is 'out of training distribution') and their monolithic nature makes them difficult to debug. Further, despite much research effort, classical motion estimation algorithms, like stereo visual odometry, still achieve state-of-the-art performance in nominal conditions[1]. Nevertheless, the representational power of deep regression algorithms makes them an attractive option to complement classical motion estimation when these latter methods perform poorly (e.g., under diverse lighting conditions or low scene texture). By endowing deep regression models with a useful notion of uncertainty, we can account for out-of-training-distribution errors (a critical step to ensure safe operation) and fuse these models with classical methods using probabilistic factor graphs. In this work, we choose to focus on rotation regression, since many motion algorithms are sensitive to rotation errors [18], and good rotation initializations can be critical to robust optimization [3]. Our novel contributions are

1. a deep network structure we call *HydraNet* that builds on prior work [14, 17, 16] to build covariance matrices that account for both *epistemic* and *aleatoric* uncertainty [13],

2. a formulation that extends HydraNet to means and covariances defined on the rotation group $\mathrm{SO}(3)$,

3. and open-source code for $\mathrm{SO}(3)$ regression[2].

[1] Based on the KITTI odometry leaderboard [6] at the time of writing.

[2] https://github.com/utiasSTARS/so3_learning

2. Approach

2.1. HydraNet

To endow deep networks with uncertainty, we extend prior work on multi-headed networks [17] and ensembles of networks trained through bootstrapped training sets [14] to a single network structure we call HydraNet (see Figure 1).

HydraNet is composed of a large, main body, with multiple heads attached prior to the output. Each head is trained independently and from different weight initializations to provide a different estimate of the regressed quantity. This diversity allows the model to capture *epistemic* uncertainty (σ_e^2) by computing a sample variance over the heads. The epistemic uncertainty is designed to grow when inputs are 'far' from training data in some latent space. Further, an additional head is reserved for regressing an *aleatoric* estimate of uncertainty directly (σ_a^2). This aleatoric uncertainty accounts for effects like sensor noise that are present even if a test input is 'close' to training data. See [13] for a further discussion of aleatoric and epistemic uncertainty.

2.2. Deep Probabilistic $\mathrm{SO}(3)$ Regression

In order to extend the ideas of HydraNet to the matrix Lie group SO(3), we investigate different ways to regress and combine several estimates of rotations. Namely, given a network, $g(\cdot)$, and an input $\mathcal{I}$, we consider how to to process several outputs, $g_i(\mathcal{I})$, and combine them into an estimate of a 'mean' rotation, $\overline{\mathbf{R}}$, and an associated 3×3 covariance matrix, $\mathbf{\Sigma}$. To produce estimates of rotation for a given HydraNet head, we choose to normalize outputs, $g(\mathcal{I}) \in \mathbb{R}^4$, to produce a unit quaternions that reside on S^3, $\mathbf{q} = \frac{g(\mathcal{I})}{\|g(\mathcal{I})\|_2}$. The choice of unit quaternions (as opposed to using exponential coordinates, for example), is motivated by a simple analytic mean expression based on the *quaternionic* metric which we describe next.

2.2.1 Rotation Averaging

Given several estimates of a rotation, we define the mean as the rotation which minimizes some squared metric defined over the group [10],

$$\overline{\mathbf{R}} = \underset{\mathbf{R} \in \mathrm{SO}(3)}{\operatorname{argmin}} \sum_{i=1}^{n} d(\mathbf{R}_i, \mathbf{R})^2. \tag{1}$$

There are three common choices for a bijective metric [10, 3] on $\mathrm{SO}(3)$. The angular, chordal and quaternionic:

$$d_{\mathrm{ang}}(\mathbf{R}_a, \mathbf{R}_b) = \left\| \mathrm{Log}\left(\mathbf{R}_a \mathbf{R}_b^T\right) \right\|_2, \tag{2}$$

$$d_{\mathrm{chord}}(\mathbf{R}_a, \mathbf{R}_b) = \|\mathbf{R}_a - \mathbf{R}_b\|_{\mathrm{F}}, \tag{3}$$

$$d_{\mathrm{quat}}(\mathbf{q}_a, \mathbf{q}_b) = \min\left(\|\mathbf{q}_a - \mathbf{q}_b\|_2, \|\mathbf{q}_a + \mathbf{q}_b\|_2\right), \tag{4}$$

where $\mathrm{Log}(\cdot)$, represents the capitalized matrix logarithm [19], and $\|\cdot\|_F$ the Frobenius norm. In the context of Equation (1), using the angular metric leads to the *Karcher mean*, which requires an iterative solver and has no known analytic expression. Applying the chordal metric leads to an analytic expression for the average but requires the use of Singular Value Decomposition. Using the quaternionic metric, however, leads to a simple, analytic expression for the rotation average as the normalized arithmetic mean of a set of unit quaternions [10],

$$\overline{\mathbf{q}} = \underset{\mathbf{R}(\mathbf{q}) \in \mathrm{SO}(3)}{\operatorname{argmin}} \sum_{i=1}^{H} d_{\mathrm{quat}}(\mathbf{q}_i, \mathbf{q})^2 = \frac{\sum_{i=1}^{H} \mathbf{q}_i}{\left\| \sum_{i=1}^{H} \mathbf{q}_i \right\|}. \tag{5}$$

This expression is simple to evaluate numerically, and if necessary, can be easily differentiated with respect to its constituent parts. For these reasons, we opt to construct our $\mathrm{SO}(3)$ HydraNet using unit quaternion outputs, and evaluate the rotation average using the quaternionic metric.

2.2.2 SO(3) Uncertainty

We parametrize uncertainty over $\mathrm{SO}(3)$ by injecting uncertainty onto the manifold [5, 2, 1] from a local tangent space about some mean element, $\overline{\mathbf{q}}$,

$$\mathbf{q} = \mathrm{Exp}(\boldsymbol{\epsilon}) \otimes \overline{\mathbf{q}}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \tag{6}$$

where $\otimes$ represents quaternion multiplication. In this formulation, $\mathbf{\Sigma}$ provides a 3×3 covariance matrix that can express uncertainty in different directions. Given a mean rotation, $\overline{\mathbf{q}}$, and samples, $\mathbf{q}_i$, we use the logarithmic map to compute a sample covariance matrix that expressed epistemic uncertainty as,

$$\mathbf{\Sigma}_e = \frac{1}{H-1} \sum_{i=1}^{H} \boldsymbol{\phi}_i \boldsymbol{\phi}_i^T, \quad \boldsymbol{\phi}_i = \mathrm{Log}\left(\mathbf{q}_i \otimes \overline{\mathbf{q}}^{-1}\right). \tag{7}$$

2.3. Loss Function

To capture aleatoric uncertainty, we train a direct regression of covariance through a parametrization of positive semi-definite matrices using a Cholesky decomposition [11, 9]). Given the network output of a unit quaternion $\mathbf{q}$, and a positive semi-definite matrix $\mathbf{\Sigma}_a$, we define a loss function as the negative log likelihood of a given rotation under Equation (6) (see [5]) for a given target rotation, $\mathbf{q}_t$, as

$$\mathcal{L}_{\mathrm{NLL}}(\mathbf{q}, \mathbf{q}_t, \mathbf{\Sigma}_a) = \frac{1}{2} \boldsymbol{\phi}^T \mathbf{\Sigma}_a^{-1} \boldsymbol{\phi} + \frac{1}{2} \log\det(\mathbf{\Sigma}_a), \tag{8}$$

where $\boldsymbol{\phi} = \mathrm{Log}\left(\mathbf{q} \otimes \mathbf{q}_t^{-1}\right)$. At test time, we combine both covariance matrices using a simple matrix additon, $\mathbf{\Sigma} = \mathbf{\Sigma}_e + \mathbf{\Sigma}_a$.

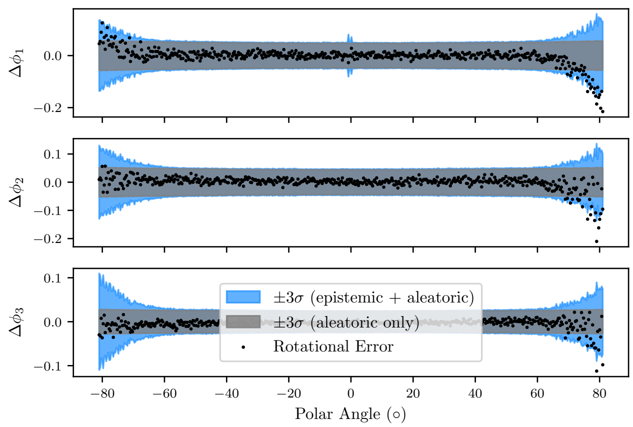

Figure 2: Rotation estimation errors and reported network covariance elements for HydraNet trained on synthetic data (noisy pixel locations). Epistemic uncertainty grows for out-of-training-distribution test samples (inputs captured at polar angles outside of $\pm 60°$).

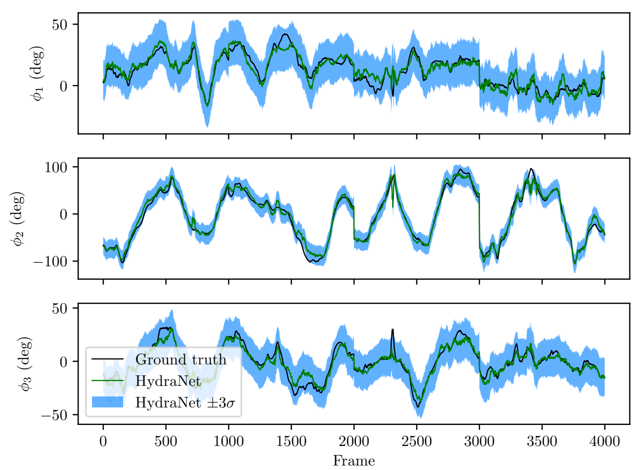

Figure 3: Orientation regression results for the 7scenes *chess* test set. Our HydraNet structure paired with a `resnet-34` results mean errors of 8.6 degrees, with consistent uncertainty.

3. Results & Future Work

To demonstrate our approach, we use three different datasets: synthetic (wherein we simulate a camera observing a static scene with point landmarks), 7-Scenes [8] RGB data, and KITTI monocular grayscale images [6]. Figure 2 shows the importance of epistemic uncertainty when training a model to compute rotations on the synthetic dataset. Figure 3 shows the consistent orientation regression resulting from training a HydraNet (with a ResNet body) to regress probabilistic estimates of orientation on the 7-Scenes dataset. Finally, Figure 4 shows the result of training a custom HydraNet to regress frame-to-frame rotations

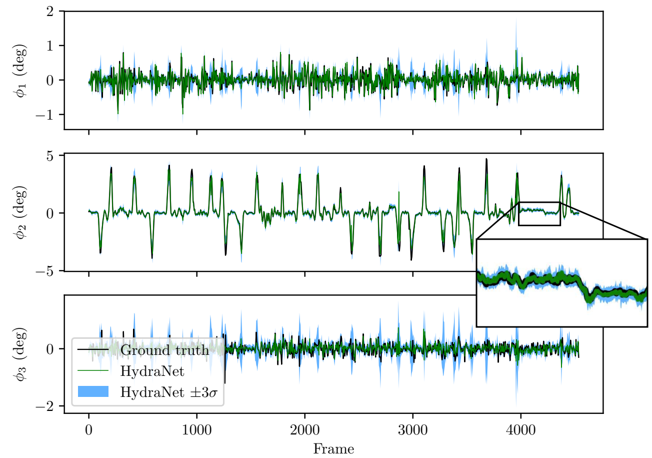

Figure 4: Frame-to-frame rotation regression for KITTI odometry dataset sequence `00`. Note how the uncertainty increases when the car turns (ϕ_2 represents the yaw angle).

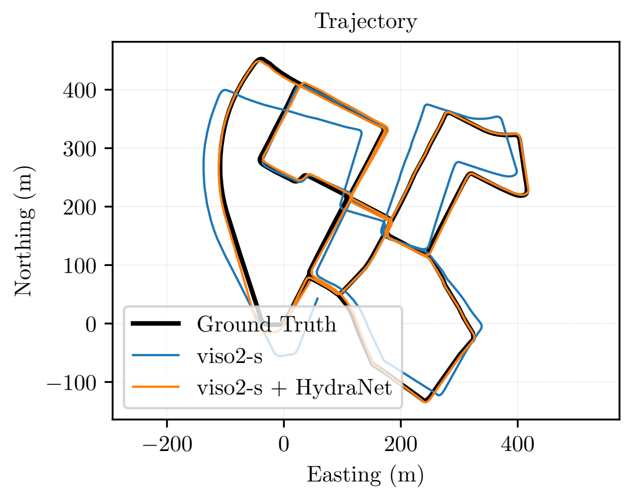

Figure 5: Top-down trajectories of KITTI odometry dataset sequence `00` with and without HydraNet fusion.

on the KITTI self-driving car dataset. By using the rotation estimates and their associated covariances, we fuse the outputs of HydraNet with a classical visual odometry pipeline (based on sparse features [7]) using probabilistic factor graphs and visualize the resulting increase in accuracy in Figure 5.

In future work, we are looking for ways to obviate the need for supervised training (e.g., by embedding the HydraNet structure within a Bayesian filter like that in [9]), and compare our notion of epistemic uncertainty to a separate classification model that predicts whether a sample is in or out of training distribution. We also see great promise in using deep probabilistic rotation regression to aid in initialization within large scale pose graph optimization that relies on vision-based sensors.

References

[1] Timothy D Barfoot. *State Estimation for Robotics*. Cambridge University Press, July 2017.

[2] T D Barfoot and P T Furgale. Associating uncertainty with Three-Dimensional poses for use in estimation problems. *IEEE Trans. Rob.*, 30(3):679–693, June 2014.

[3] L Carlone, R Tron, K Daniilidis, and F Dellaert. Initialization techniques for 3D SLAM: A survey on rotation estimation and its use in pose graph optimization. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4597–4604, May 2015.

[4] Ronald Clark, Sen Wang, Hongkai Wen, Andrew Markham, and Niki Trigoni. VINet: Visual-inertial odometry as a sequence-to-sequence learning problem. In *AAAI Conf on Artificial Intelligence*, 2017.

[5] Christian Forster, Luca Carlone, Frank Dellaert, and Davide Scaramuzza. IMU preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation. 2015.

[6] A Geiger, P Lenz, C Stiller, and R Urtasun. Vision meets robotics: The KITTI dataset. *Int. J. Rob. Res.*, 32(11):1231–1237, 1 Sept. 2013.

[7] A Geiger, J Ziegler, and C Stiller. StereoScan: Dense 3D reconstruction in real-time. In *Proc. Intelligent Vehicles Symp. (IV)*, pages 963–968. IEEE, June 2011.

[8] B. Glocker, S. Izadi, J. Shotton, and A. Criminisi. Real-time rgb-d camera relocalization. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 173–179, Oct 2013.

[9] Tuomas Haarnoja, Anurag Ajay, Sergey Levine, and Pieter Abbeel. Backprop KF: Learning discriminative deterministic state estimators. In *Proceedings of Neural Information Processing Systems (NIPS)*, 2016.

[10] Richard Hartley, Jochen Trumpf, Yuchao Dai, and Hongdong Li. Rotation averaging. *Int. J. Comput. Vis.*, 103(3):267–305, July 2013.

[11] H Hu and G Kantor. Parametric covariance prediction for heteroscedastic noise. In *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Syst. (IROS)*, pages 3052–3057, 2015.

[12] Alex Kendall, Kendall Alex, Grimes Matthew, and Cipolla Roberto. PoseNet: A convolutional network for Real-Time 6-DOF camera relocalization. In *Proc. of IEEE Int. Conf. on Computer Vision (ICCV)*, 2015.

[13] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in neural information processing systems*, pages 5574–5584, 2017.

[14] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 6402–6413. Curran Associates, Inc., 2017.

[15] Iaroslav Melekhov, Juha Ylioinas, Juho Kannala, and Esa Rahtu. Relative camera pose estimation using convolutional neural networks. In *Proc. Int. Conf. on Advanced Concepts for Intel. Vision Syst.*, pages 675–687. Springer, 2017.

[16] R. Minetto, M. P. Segundo, and S. Sarkar. Hydra: An ensemble of convolutional neural networks for geospatial land classification. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–12, 2019.

[17] Ian Osband, Charles Blundell, Alexander Pritzel, and Benjamin Van Roy. Deep exploration via bootstrapped DQN. *CoRR*, abs/1602.04621, 2016.

[18] Valentin Peretroukhin, Lee Clement, and Jonathan Kelly. Inferring sun direction to improve visual odometry: A deep learning approach. *The International Journal of Robotics Research*, 37(9):996–1016, 2018.

[19] Joan Solà, Jeremie Deray, and Dinesh Atchuthan. A micro lie theory for state estimation in robotics. Dec. 2018.