Robot Perception enables Complex Navigation Behavior via Self-Supervised Learning

Marvin Chancán and Michael Milford
QUT Centre for Robotics, School of Electrical Engineering and Robotics
Queensland University of Technology, Australia
Email: mchancan1@uni.pe

Abstract—Learning visuomotor control policies in robotic systems is a fundamental problem when aiming for long-term behavioral autonomy. Recent supervised-learning-based vision and motion perception systems, however, are often separately built with limited capabilities, while being restricted to few behavioral skills such as passive visual odometry (VO) or mobile robot visual localization. Here we propose an approach to unify those successful robot perception systems for active target-driven navigation tasks via reinforcement learning (RL). Our method temporally incorporates compact motion and visual perception data—directly obtained using self-supervision from a single image sequence—to enable complex goal-oriented navigation skills. We demonstrate our approach on two real-world driving dataset, KITTI and Oxford RobotCar, using the new interactive CityLearn framework. The results show that our method can accurately generalize to extreme environmental changes such as day to night cycles with up to an 80% success rate, compared to 30% for a vision-only navigation systems.

I. INTRODUCTION

Recent advances in self-supervised learning have show promising results in a range of visuomotor tasks including robotic manipulation [18, 20, 5, 11, 6, 16] using deep reinforcement learning (RL), both in simulation and on real hardware. For mobile robots, these self-supervised learning techniques are now being explored and have already show to achieve comparable results to classical robot perception pipelines for passive visual odometry (VO) [21, 19], visual localization or place recognition (VPR) [7], and also active outdoor navigation tasks [2] in real environments. Nevertheless, end-to-end learning of visuomotor policies for long-term, all-weather autonomous navigation tasks using self-supervision remains unexplored.

Large-scale outdoor navigation is a key component for enabling the deployment of mobile robots and autonomous vehicles in the real world. Recent RL-based navigation systems for real environments rely on GPS-based ground-truth data for labeling raw sensory images. They then reduce the problem of navigation to vision-only methods [15] or extend it with language-based sensory inputs. These approaches: 1) are generally hard to train—due to their weakly-related input sensor modalities, 2) rely on the precision of GPS data—which may not be reliable across month-spaced traversals of the same route, and 3) require a large amount of experience with the environment in terms of RL training episodes—which might be impractical for real robots. Moreover, their generalization capabilities to different environmental changes such as lighting or weather transitions are not explored.

In this paper, instead of relying on supervised learning methods for capturing motion and visual representations, we investigate how to leverage recent self-supervised learning approaches for enabling efficient and robust long-term robot navigation skills. Our key contributions are:

- An approach to temporally integrate motion states (classical VO self-optimized with optical flow and depth prediction) with visual observations (self-enhanced with image-to-region similarities) via RL for large-scale, all-weather navigation tasks (see Fig. 1), and
- Experimental trade-off between the RL navigation success rate and the motion estimation precision, providing key insights to decide which ego-motion sensor would be appropriate for a particular application.

We demonstrate the effectiveness and advantages of our method on two large, real driving datasets for goal-oriented navigation tasks, compared to motion-only and vision-only navigation systems. Furthermore, we report experimental results where our approach is capable of generalizing to extreme environmental transitions such as day to night cycles with high navigation success rate, where vision-only navigation systems typically fail.
II. Problem Formulation

We formulate the goal-driven navigation task as a Markov decision process $\mathcal{M}$: at any given discrete state $s_t \in S$ at time $t$, the robot executes a discrete action $a_t \in A$ following the policy $\pi_\theta : S \rightarrow \mathcal{P}(A)$, then transitions to a new state $s_{t+1}$ receiving a corresponding reward $r$. We train our policy to find an optimal $\pi^*$ that maximizes the objective function given by

$$E_{T \sim \pi_\theta(T)} \sum_{t=1}^{T} \gamma r(t),$$

with a transition operator $T : S \times A \rightarrow S$ and a $\gamma$-discounted reward function over a finite-horizon $T$.

In this work, following the main ideas proposed in [3] and [4], we investigate how to temporally incorporate in $S$ compact motion states, $m_t$, with equally compact visual observations, $x_t$, both obtained via self-supervised learning from a single monocular image sequence, $I_t$, using the state-of-the-art RL algorithm PPO [17] (see Fig. [1]).

III. Approach

Our objective is to train an RL agent to perform goal-driven navigation tasks across a range of real-world environmental conditions, especially where noise or poor GPS data typically limit the capabilities of supervised learning approaches. We therefore developed a combined motion-and-visual-based perception method that can be trained using self-supervision. Our approach operates by temporally associating local motion states, obtained from VO-based techniques, with visual observations to efficiently train our navigation policy network, Fig. [1]. This enables our policy to learn from both motion and visual information in a self-supervised manner, while training using an RL framework, to being robust to environmental visual changes and also poor GPS data.

A. Self-Supervised Single-Frame Visual Localization

In image-based localization, weak GPS- or geo-tagged labels can be problematic when training visual place recognition (VPR) systems using supervised learning. To overcome these challenges, successful VPR systems such as NetVLAD [1] have achieved state-of-the-art results via weakly-supervised learning, with a range of recent developments [10] [13]. More recently, however, a self-supervised fine-grained region similarities (SFRS) system, especially designed for dealing with noisy pairwise image-label, has outperformed these VPR pipelines [2]. In this work, we attempt to merge the desirable properties of SFRS into our RL-based navigation system for leveraging image-to-region similarities when GPS labels are poor or not available for large-scale image perception.

B. Self-Supervised Monocular Visual Odometry

In robot navigation research, visual odometry (VO) and SLAM techniques are also typically used for performing visual-based localization; providing key complementary information of the environment along with GPS, IMU or LiDAR sensors. While SLAM extends VO, along with loop closing and global map optimization, for building a geometrically consistent map of the environment, VO continues to be a fundamental component for proving ego-motion estimate data for mobile robots. With the rapid progress of deep learning techniques in computer vision, roboticists have been attracted to incorporate these learning capabilities for VO over the past 4 years [22] [12]. Only recently, however, the use of more advanced self-supervised learning techniques have enabled to outperform those purely geometry-based or deep-learning-based VO systems [21] [19].

IV. Experiments

We evaluate our model on two real driving datasets, Oxford RobotCar [14] and KITTI [8], using the CityLearn environment [2], see Figs. 2 and 3. We additionally conduct experiments to obtain the trade-off between the RL success rate and the motion estimation precision.

Figs. 4 and 5 provide the corresponding quantitative results averaged over 6 runs with different random seeds. We compare our full model (green) with two baselines which correspond to pure motion-only RL (blue), and vision-only RL (orange).
Our approach incorporates the desirable properties of motion- and vision-only methods for navigation tasks. Using images (alone) seems to increase complexity and reduce performance, but when combining it with motion data we compensate these shortcomings.

In contrast, the agent using motion states seems to compete with our full model, however, its main limitation is that it does not incorporate visual information for distinguishing between environmental changes. For the Oxford RobotCar dataset, where we test generalization from day to night, our full model is capable of consistently obtaining around an 80% success rate, compared to 30% for the vision-only system.

To further analyze the influence of motion estimation precision, in all our experiments we compare the ego-motion data obtained using classical VO and SS-DPC-Net against ground truth data provided by each dataset, see Fig. 8 (left) for the KITTI dataset. Interestingly, the difference between these ego-motion results does not seem to impact our three baselines on the KITTI dataset, as all these models are deployed on the same traversals used for training. Conversely, on the Oxford RobotCar dataset, as we also deploy under drastic visual changes (day to night), we note that our full model retains good navigation performance, compared to vision-only systems. We also provide insights on the influence of the VO precision to our full model in Fig. 8 (right).
V. Conclusion

We have shown that combining self-supervised learning for visuomotor perception and RL for decision-making considerably improves the ability to deploy robotic systems capable of solving complex navigation tasks from raw image sequences only. We proposed a method, including a new neural network architecture, that temporally integrates two fundamental sensor modalities such as motion and vision for large-scale target-driven navigation tasks using real data via RL. Our approach was demonstrated to be robust to drastic visual changing conditions, where typical vision-only navigation pipelines fail. This suggests that odometry-based data can be used to improve the overall performance and robustness of conventional vision-based systems for learning complex navigation tasks. In future work, we seek to extend this approach by using unsupervised learning for both decision-making and perception.

Acknowledgments

This research received funding from the Australian Government, via grant AU5MURIB000001 associated with ONR MURI grant N00014-19-1-2571, and was partially supported by funding from ARC grants FT140101229, CE140100016 and the QUT Centre for Robotics.

References