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: mchancanl@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 [9] 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.

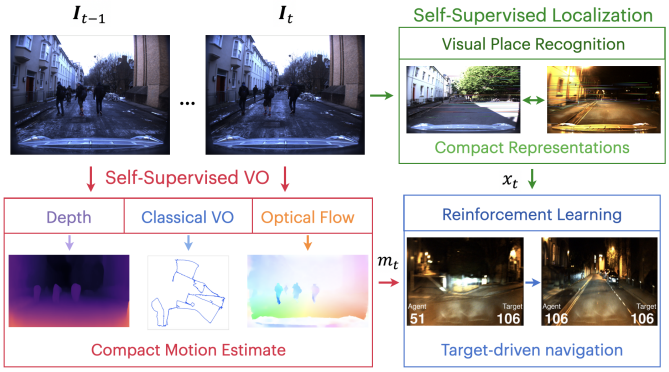

Fig. 1. Overview of our unified robot learning framework for navigation tasks. Given a single traversal of a car ride (I_t), we use self-supervised learning to obtain optimized VO data (m_t) and visual representations (x_t). We then temporally combine these compact visuomotor signals to learn control policies for goal-driven navigation skills via RL. Our method can accurately generalize to extreme environmental changes such as day to night transitions.

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 $\mathbf{s}_t \in \mathcal{S}$ at time t, the robot executes a discrete action $\mathbf{a}_t \in \mathcal{A}$ following the policy $\pi_\theta : \mathcal{S} \rightarrow \mathcal{P}(\mathcal{A})$, then transitions to a new state $\mathbf{s}_{t+1}$ receiving a corresponding reward r. We train our policy to find an optimal θ^* that maximizes the objective function given by $E_{\tau\sim\pi_\theta(\tau)} \sum_{t=1}^{T} \gamma r(\tau)$, with a transition operator $\mathcal{T} : \mathcal{S}\times\mathcal{A} \rightarrow \mathcal{S}$ and a γ-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 $\mathcal{S}$ compact motion states, $\mathbf{m}_t$, with equally compact visual observations, $\mathbf{x}_t$, both obtained via self-supervised learning from a single monocular image sequence, $\mathbf{I}_t$, 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-vision-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 [7]. 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]. Here we incorporate a self-supervised deep pose corrections method (SS-DPC-Net) [19], which combine depth estimation, optical flow, and classical VO in a hybrid manner, for robust VO into our RL-based system, providing compact and optimized ego-motion estimate data.

Fig. 2. **Deployment results** on the KITTI dataset. The agent navigates from left to right towards the goal destination.

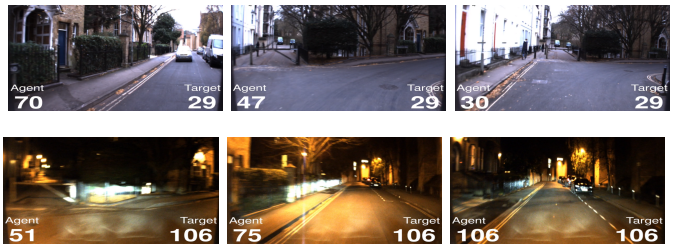

Fig. 3. **Deployment results** on the Oxford RobotCar dataset. The agent navigates from left to right towards the goal destination on the traversal it was trained (top), and generalize well at night (bottom).

C. Reinforcement learning-based navigation

Goal-driven navigation: We merge both motion states, $\mathbf{m}_t$, and visual observations, $\mathbf{x}_t$, obtained via self-supervision from raw image sequences, for learning to navigate through actions, $\mathbf{a}_t$, towards a required goal destination, $\mathbf{g}_t$, via RL [17].

Architecture: Our policy network is inspired by [15], which includes a single *linear* layer with 512 units to encode $\mathbf{m}_t$ and $\mathbf{x}_t$. Then, using a single recurrent layer long short-term memory (LSTM) with 256 units, current states and observations are combined with the agent's previous actions, $\mathbf{a}_{t-1}$. The updated agent's actions, $\mathbf{a}_t$, are then used to estimate both the new actions and the value function V from π_θ.

Reward design and curriculum learning: We use multiple levels of curriculum learning to gradually encourage our agent to explore the environment, and a sparse reward function that gives the agent a reward of $+1$ only when it finds the target.

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).

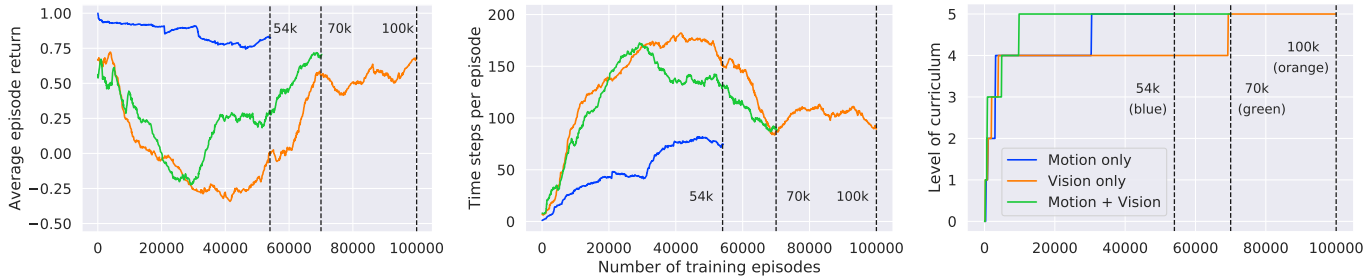

Fig. 4. **RL training curves on the KITTI dataset**. Our approach incorporates the desirable properties of motion- and vision-only methods for navigation tasks. Using images (alone) seems to increase complexity and reduce performance, but when combining it with motion data we compensate these shortcomings.

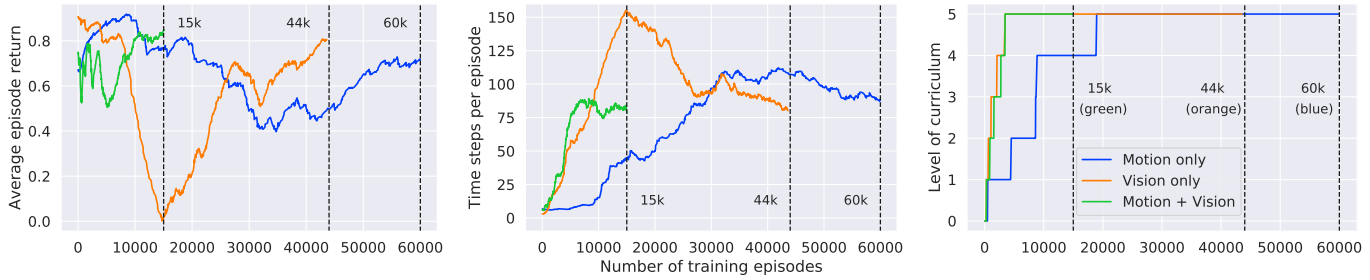

Fig. 5. **RL training curves on the Oxford RobotCar dataset**. In contrast to the results in Fig. 4, we found that using motion+visual data can actually boost the RL training. In this case, our full model required 15k training episodes, compared to 60k and 44k for the motion- and vision-only baselines.

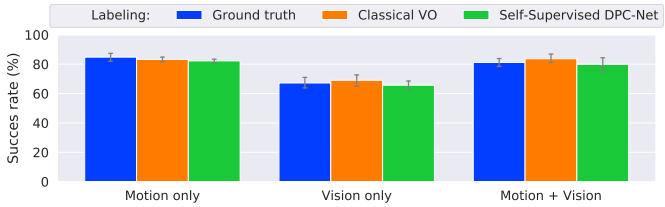

Fig. 6. **RL deployment statistics on the KITTI dataset**. We trained and tested on sequence 05 of the raw data.

Fig. 7. **RL deployment statistics on the Oxford RobotCar dataset** on the traversal it was trained (top) and generalizing at night (bottom).

These two baselines use the same setup as the full model, except that they only use either motion estimate data or visual observations, respectively, as shown in Fig. 1.

We also report deployment statistics in Figs. 6 and 7. For the KITTI dataset, our full model can solve navigation tasks with 80% success rate, compared to 65% for the vision-only system. 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 traversal 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

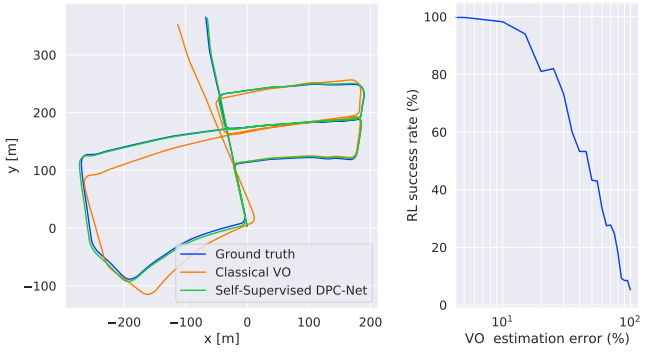

Fig. 8. **Influence of motion estimation precision**. Ground truth and VO-based data of the KITTI dataset (left). Trade-off between the RL navigation success rate and the ego-motion estimation precision (right).

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 suggest 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 AUSMURIB000001 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

[1] Relja Arandjelovic, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic. NetVLAD: CNN Architecture for Weakly Supervised Place Recognition. In *CVPR*, 2016.

[2] Marvin Chancán and Michael Milford. From visual place recognition to navigation: Learning sample-efficient control policies across diverse real world environments. *arXiv preprint arXiv:1910.04335*, 2019.

[3] Marvin Chancán and Michael Milford. MVP: Unified Motion and Visual Self-Supervised Learning for Large-Scale Robotic Navigation. *arXiv preprint arXiv:2003.00667*, 2020.

[4] Marvin Chancán, Luis Hernandez-Nunez, Ajay Narendra, Andrew B. Barron, and Michael Milford. A hybrid compact neural architecture for visual place recognition. *IEEE Robotics and Automation Letters*, 5(2):993–1000, April 2020.

[5] Xinke Deng, Yu Xiang, Arsalan Mousavian, Clemens Eppner, Timothy Bretl, and Dieter Fox. Self-supervised 6D Object Pose Estimation for Robot Manipulation. *arXiv preprint arXiv:1909.10159*, 2019.

[6] Frederik Ebert, Sudeep Dasari, Alex X. Lee, Sergey Levine, and Chelsea Finn. Robustness via retrying: Closed-loop robotic manipulation with self-supervised learning. In *Proceedings of The 2nd Conference on Robot Learning*, volume 87 of *Proceedings of Machine Learning Research*, pages 983–993. PMLR, 2018.

[7] Yixiao Ge, Haibo Wang, Feng Zhu, Rui Zhao, and Hongsheng Li. Self-supervising fine-grained region similarities for large-scale image localization. *arXiv preprint arXiv:2006.03926*, 2020.

[8] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets Robotics: The KITTI Dataset. *The International Journal of Robotics Research*, 2013.

[9] Gregory Kahn, Pieter Abbeel, and Sergey Levine. BADGR: An autonomous self-supervised learning-based navigation system. *arXiv preprint arXiv:2002.05700*, 2020.

[10] H. J. Kim, E. Dunn, and J. Frahm. Learned contextual feature reweighting for image geo-localization. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3251–3260, 2017.

[11] Michelle Lee *et al.* Making sense of vision and touch: Learning multimodal representations for contact-rich tasks. *IEEE Transactions on Robotics*, 2020.

[12] Ruihao Li, Sen Wang, Zhiqiang Long, and Dongbing Gu. UnDeepVO: Monocular Visual Odometry Through Unsupervised Deep Learning. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 7286–7291, 2018.

[13] Liu Liu, Hongdong Li, and Yuchao Dai. Stochastic attraction-repulsion embedding for large scale image localization. In *ICCV*, October 2019.

[14] Will Maddern, Geoffrey Pascoe, Chris Linegar, and Paul Newman. 1 year, 1000 km: The Oxford RobotCar dataset. *The International Journal of Robotics Research*, 36(1):3–15, 2017.

[15] Piotr Mirowski *et al.* Learning to navigate in cities without a map. In *Advances in Neural Information Processing Systems 31*, pages 2419–2430. 2018.

[16] Suraj Nair and Chelsea Finn. Hierarchical foresight: Self-supervised learning of long-horizon tasks via visual subgoal generation. In *ICLR*, 2020.

[17] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[18] Pierre Sermanet, Corey Lynch, Jasmine Hsu, and Sergey Levine. Time-contrastive networks: Self-supervised learning from multi-view observation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 486–487, 2017.

[19] Brandon Wagstaff, Valentin Peretroukhin, and Jonathan Kelly. Self-supervised deep pose corrections for robust visual odometry. *arXiv preprint arXiv:2002.12339*, 2020.

[20] Andy Zeng, Shuran Song, Stefan Welker, Johnny Lee, Alberto Rodriguez, and Thomas Funkhouser. Learning synergies between pushing and grasping with self-supervised deep reinforcement learning. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4238–4245, 2018.

[21] Huangying Zhan, Chamara Saroj Weerasekera, Jiawang Bian, and Ian Reid. Visual odometry revisited: What should be learnt? *arXiv preprint arXiv:1909.09803*, 2019.

[22] Tinghui Zhou, Matthew Brown, Noah Snavely, and David G. Lowe. Unsupervised learning of depth and ego-motion from video. In *CVPR*, 2017.