

From Visual Place Recognition to Navigation: Learning Sample-Efficient Control Policies across Diverse Real World Environments

Marvin Chancán^{1,2} and Michael Milford¹

Abstract— Visual navigation tasks in real world environments often require both self-motion and place recognition feedback. While deep reinforcement learning has shown success in solving these perception and decision-making problems in an end-to-end manner, these algorithms require large amounts of experience to learn navigation policies from high-dimensional inputs, which is generally impractical for real robots due to sample complexity. In this paper, we address these problems with two main contributions. We first leverage place recognition and deep learning techniques combined with goal destination feedback to generate compact, bimodal images representations that can then be used to effectively learn control policies at kilometer scale from a small amount of experience. Second, we present an interactive and realistic framework, called CityLearn, that enables for the first time the training of navigation algorithms across city-sized, real-world environments with extreme environmental changes. CityLearn features over 10 benchmark real-world datasets often used in place recognition research with more than 100 recorded traversals and across 60 cities around the world. We evaluate our approach in two CityLearn environments where our navigation policy is trained using a single traversal. Results show our method can be over 2 orders of magnitude faster than when using raw images and can also generalize across extreme visual changes including day to night and summer to winter transitions.

I. INTRODUCTION

The ability to sense their own location in time and space is a key in both robotic systems and living beings that enable them to navigate highly dynamic, real-world environments. For mobile robots, the way they can create a particular, internal world representation often depends on their perceptual limitations as well as how they interact and make decisions with the environment [1]. Visual feedback provides high-dimensional information that, when encoded properly, can be used to make sense of where they are and where they need to go. Similarly, self-motion feedback also provides information concerning current position within an environment. These two sensory input modalities are concurrent, time-aligned and often complementary during goal-driven navigation tasks.

Recent deep reinforcement learning approaches have successfully performed active navigation tasks on simulated environments using real-world street imagery [2] or synthetic scenarios [3], [4]. These algorithms, however, generally utilize additional/external feedback data, such as the agent-relative velocity or reward function values, that eventually

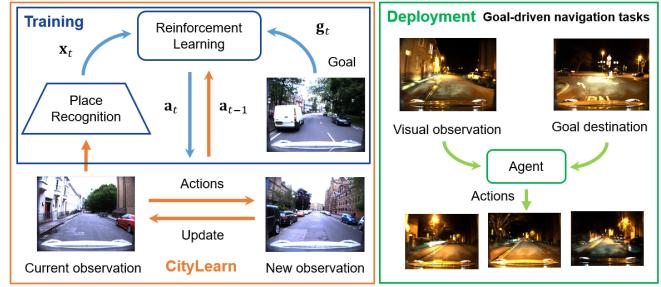


Fig. 1. Learning sample-efficient control policies from compact, bimodal representations for goal-driven navigation tasks at kilometer scale across extreme visual appearance changes in real-world environments.

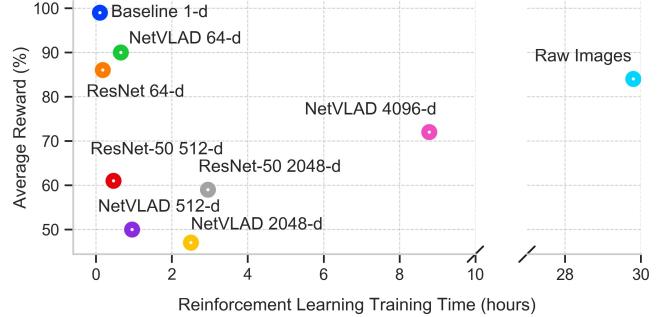


Fig. 2. Performance and compute characterization of reinforcement learning training using off-the-shelf place recognition (NetVLAD) and deep learning (ResNet-50) methods to encode raw images across a range of feature vector dimensions (64, 512, 2048, 4096).

increase their network policy architecture and sample complexity.

Visual place recognition methods, on the other hand, need to successfully match two or more image sequences of recorded traversals in real-world environments. While recent improvements using deep learning [5]–[7] and algorithmic techniques [8] have contributed to state-of-the-art results on city-sized datasets, the ability of those methods for performing navigation tasks on real robots is not well explored.

In this work, we leverage both visual place recognition and deep reinforcement learning techniques to efficiently learn control policies for navigation tasks. We demonstrate our proposed method, shown in Fig. 1, where the resulting control policy is able to perform goal-driven navigation tasks using only two sensory feedback modalities (goal destination and visual features). The results demonstrate that our policy is able to generalize over a range of extreme environmental changes on real-world datasets, while drastically reducing the amount of training experience, e.g. from 29h48m to 11m (see Fig. 2), achieving practical sample efficiency in an interactive and diverse environment that we call CityLearn.

¹ School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, QLD 4000, Australia

² School of Mechatronics Engineering, Universidad Nacional de Ingeniería, Lima, Rímac 15333, Peru. mchancanl@uni.pe

Our primary contributions are:

- 1) CityLearn: an interactive open framework with real-world environments for perception and decision-making to enable the evaluation of navigation algorithms across more than 10 robotic benchmark datasets (see Fig. 3).
- 2) A new approach to sample-efficient reinforcement learning training for goal-driven navigation tasks. We use place recognition and deep learning methods to encode our sensory input images, which combined with a goal destination, generate a compact, bimodal representation, from which a navigation policy can be learned to generalize across extreme visual changes such as day to night or summer to winter cycles.

II. RELATED WORK

In robotics research, the use of probabilistic techniques played an important role to solve robotic problems such as how the robot's sensory information should be integrated to generate internal states to support the decision-making process [1]. In the mid-1990s, these methods allowed the deployment of navigation algorithms on a real robot by using conditional probability distributions, instead of deterministic functions at a fixed time interval (as in classical control), in order to compute more general control actions that govern the robot's states [9]–[11]. In the same decade, moreover, the field of reinforcement learning (RL) started to attract the interest of roboticists. RL agents learn specific behavior through interactions with dynamic environments only by reward and punishment signals [12]. Therefore, the use of RL to solve more complex robot-learning problems started to be extended with the incorporation of neural networks to obtain broader generalization capabilities. Ideas like hierarchical or curriculum learning were also proposed to reduce the learning time and solve these complex, physically realistic robotic problems in simulation environments [12].

A. Deep Reinforcement Learning based Navigation

The spread of *convolutional neural networks* (CNN) has yielded impressive state-of-the-art results in computer vision, natural language processing and many other related domains over the past eight years [13]. Similarly, recent works incorporating deep neural networks to more advanced RL algorithms for navigation tasks have shown promising results in simulated environments. [14], [21], [?], [39], [40] demonstrate deep RL agents performing goal-oriented navigation tasks using real-world street images and at city scale [14], [23], [25], [3], generalizing over different cities with minimal additional training and network architecture changes. However, these approaches often use additional feedback data such as reward function values or the agent-relative velocity that further increase the policy architecture and sample complexity. These factors also increase the number of interactions required with the environment, typically to the order of billions. Moreover, those systems are often evaluated on the same environment used for training, thus their generalization capabilities to different conditions are

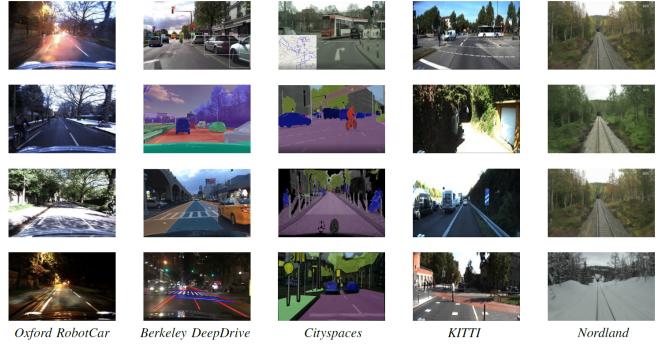


Fig. 3. Five of ten benchmark robotic datasets featured by CityLearn.

often unknown; alternatively it is necessary to increase the complexity of their architectures for them to successfully generalize.

Recent work using deep RL models has shown success on target-driven navigation tasks [37], but is only demonstrated on indoor environments [37], [54]. Only recently researchers have started to use realistic images to show how the use of pre-trained deep learning models can lead to efficiently learn navigation policies at kilometer scale in unstructured environments [14].

B. Visual Place Recognition

Visual place recognition methods for localization tasks typically perform a multi-frame matching procedure between two traversals (query and database) using stationary real-world datasets (see Fig. 3). Both query and database sequence of images often include challenging appearance and viewpoint changes between them, such as different weather or seasonal conditions, illumination changes due to time of day, and dynamic objects. A visual place recognition algorithm can be broadly split into two main steps [31]: (1) a feature extraction process utilizing either hand-crafted or deep-learning-based techniques to obtain image representations of both query and database traversals, that can then be (2) matched by using conventional similarity metrics (e.g. cosine or L_2 distance) or more elaborate multi-frame temporal filtering algorithms [15]. Though recent improvements using temporal filtering approaches have shown state-of-the-art results [8], we note that those methods need to have at least two traversals from the same environment at the time of performing their final matching procedure.

In this work, for the proposed goal-oriented navigation tasks, we need only one traversal to train our control policy network, which can then be evaluated on the other traversals. This is why we particularly choose place recognition and deep learning techniques that are focused on obtaining better feature representations from raw images such as NetVLAD [5], which performs well compared to related place recognition methods [32], or ResNet, [16], than algorithmic techniques that build on top of feature representations.

III. THE CITYLEARN ENVIRONMENT

Visual place recognition methods are often evaluated on variety-rich, real-world datasets (see Fig. 3) collected over

TABLE I
CITYLEARN: DETAILED COMPARISON WITH OTHER RECENT REAL-WORLD ENVIRONMENT

Environment	region/dataset	#trav.	#imgs.	av. step	#sensors	journey	city	country
StreetLearn [2]	Wall Street	1		9.8m		548.8km	New York	
	Union Square	1	56k	9.8m				
	Hudson Rive	1		9.9m	1×panoramic cam. (360°view)			USA
	CMU	1		9.9m		574.2km	Pittsburgh	
	Allergheny	1		9.8m				
CityLearn [Ours]	South Shore	1		9.9m				
	Oxford RobotCar	133	20M	0.2m	6×stereo cam. (360° view) 2×2D, 1×3D Lidar	1010.46km	Oxford	UK
	Berkeley DeepDrive	100k	120M	30fps	1×vid. cam. (front)	1100h	multiple	USA
	<i>Cityspaces</i>	50	25k	20m	1×vid. cam. (front)	100h	multiple (50)	Germany
	KITTI	22	8.4k	10fps	4×vid. cam. (front, rear)	6h	Karlsruhe	
	Nordland Railway	4	3.6M	0.05m	1×cam. (front)	2912km	Trondheim–Bodø	Norway
	Multi-Lane Road	4	-	-	1×cam. (front)	16km	Gold Coast (GC)	
	Gold Coast Drive	1	-	-	1×cam. (front)	87km	Brisbane–GC	
	UQ St. Lucia	2	-	-	1×cam. (front)	9.5km	Brisbane	Australia
	Alderley Day/Night	2	-	50fps	1×cam. (front)	16km	Brisbane	

long traversals across different seasons, time of day or weather conditions, including dynamic objects, such as cars, traffic and pedestrians, along with longer term changes such as construction or roadworks [15], [43]–[49]. The data obtained typically includes videos or sequences of images providing panoramic or 360° views using stereo cameras, scans of 2D/3D Lidar sensors, visual odometry, and GPS/inertial data that can then be used as ground truth labels.

We leverage those real-world datasets to create our interactive framework CityLearn¹, that enables for the first time the training of navigation algorithms on city-sized, realistic environments. Our fully-configurable environment runs on top of the Unity game engine and their ML-Agents framework. CityLearn is related to the recent StreetLearn work presented in [2] but has a range of useful differences. We propose a range of more diversified environments across 5 countries and additionally enable loading any other dataset including in-house recorded data, raw sensory inputs such as LIDAR, GPS/IMU, and many others (see Tables I and II for a detailed comparison).

TABLE II
STREETLEARN VS. CITYLEARN: SUPPORT AND FEATURES

Description	StreetLearn [2]	CityLearn [Ours]
Operating system	Ubuntu 18.04	Windows/Linux/Mac
Environment engine	StreetLearn	Unity/ML-Agents
Language/ML frameworks	C++, Python/TF	C#, Python/TF
Min. RAM per env.	12GB	2GB
Number of public datasets	1	+10
Number of cities	2	+60
Number of traversals	1	+100
Min. average agent step	9.8m	0.05m
Multi-environment training	✓	✓
Feature public datasets	✗	✓
Appearance changes	✗	✓
Viewpoint changes	✗	✓
Multiple times of day	✗	✓
Multiple weather/seasons	✗	✓

¹Code at <https://github.com/mchancan/citylearn>

In Table I, all the environments (region/dataset) also include GPS localization data. Related frameworks for city-scale navigation based on real-world images were not considered in Table I as they interact differently with the environment via natural language communication [33]–[35].

IV. PROBLEM STATEMENT AND METHODS

Our objective is to train a policy network to perform goal-driven navigation tasks. To enable sample-efficiency we can use either off-the-shelf place recognition or large deep learning techniques to encode our sensory input images and obtain multi-dimensional feature vectors. Then, using reinforcement learning we combine these features with the goal destination resulting in a compact, bimodal representation that we use to train our policy using a single traversal in our CityLearn environment.

We use a Markov Decision Process \mathcal{M} with discrete state $s_t \in \mathcal{S}$ and action $a_t \in \mathcal{A}$ spaces, and transition operator $\mathcal{T} : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$ to model our navigation tasks as a finite-horizon T problem. Our goal is to find θ^* that maximizes the objective function:

$$J(\theta) = \mathbb{E}_{\tau \sim \pi_\theta(\tau)} \left[\sum_{t=1}^T \gamma r(\tau) \right] \quad (1)$$

where $\pi_\theta : \mathcal{S} \rightarrow \mathcal{P}(\mathcal{A})$ is the stochastic policy we want to learn, and $r : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is the reward function with discount factor γ . To optimize our navigation policy π_θ we parameterize it with a neural network that learn θ as described in Sec. IV-B. \mathcal{S} is defined by our compact, bimodal space representation b_t generated by combining the agent's visual feature observation x_t and its goal destination g_t , as detailed in Secs. IV-A and IV-B. \mathcal{A} is defined over discrete action movements in the agent's action space a_t . We evaluate our approach on two challenging CityLearn environments (see Fig. 5) including extreme visual changes such as day to night for Oxford RobotCar [43], and summer to winter for the Nordland [18] dataset.

A. Visual Feature Observation

We encode our sensory input images using off-the-shelf visual place recognition (NetVLAD) and deep learning (ResNet-50) techniques. For NetVLAD [5], we use their best performing network, based on VGG-16 [17] with PCA plus whitening techniques, to encode our images into a range of image representations including 4096-d, 2048-d, 512-d and 64-d feature vectors. For ResNet-50 [16], we use the network trained on ImageNet [22] and extract feature vectors of 2048-d, which we then reduce to more compact representations such as 512-d and 64-d using the algorithm provided in NetVLAD for dimensionality reduction. The input image size we use for both techniques are either 1920×1080 RGB images for Nordland or 1280×960 RGB images (previously converted from stereo images) for Oxford RobotCar dataset.

Once we obtain our images feature observations \mathbf{x}_t , we combine them with the goal destination \mathbf{g}_t to produce our compact, bimodal representation \mathbf{b}_t that serves as input to our navigation policy (see Fig. 4-b). We encode the goal \mathbf{g}_t using a 1-d feature vector to preserve the compactness of our final bimodal feature, resulting in 65-d, 513-d, 2049-d, and 4097-d representations.

B. Policy Learning for Visual Navigation

Our final objective is to learn a policy for goal-oriented navigation tasks using our compact bimodal representation \mathbf{b}_t . While there has been some success using deep reinforcement learning in navigation tasks with raw images [2], [3], they require the addition of more feedback modalities (e.g. reward values or agent's velocity) that eventually increase the number of required interactions with the environment and consequently the training time. We aim to investigate the performance of using our bimodal representation \mathbf{b}_t obtained in Sec. IV-A to train our policy.

Task Setup : We design our navigation tasks where a successfully task requires reasoning using our encoded visual features observations and goal destination \mathbf{b}_t to find a required target \mathbf{g}_t over a single traversal in the City Learn environment (see Figs. 1 and 4).

Our Agent Architecture: We choose the proximal policy optimization (PPO) algorithm [19] to optimize our objective function in Eq. (1). PPO is a variation of the TRPO algorithm [20] that constrains/optimizes the policy update, while striking the balance between sample complexity and hyperparameter tuning to achieve state-of-the-art results on a range of benchmark RL problems. First, we use a single linear *multilayer perceptron* (MLP) of 512 units to encode our bimodal representation \mathbf{b}_t , see Fig. 4 (b). We then combine it with the agent's previous action \mathbf{a}_{t-1} using a single recurrent layer LSTM (long short-term memory) [55] of 256 units to estimate the required actions from the estimated policy π and the value function V to perform our navigation tasks. Details on the other two policy networks shown in Fig. 4 (a) and (c) are discussed in Sec. IV-D.

Reward Design and Curriculum Learning: We use 7 levels of curriculum learning [27], [28] to encourage the

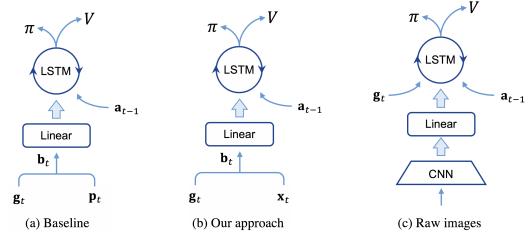


Fig. 4. **Navigation agent baselines.** Our approach uses the goal destination \mathbf{g}_t and our compact visual feature observations \mathbf{x}_t to generate bimodal representations \mathbf{b}_t which can then be combined with the agent's previous action \mathbf{a}_{t-1} to estimate our stochastic policy π and value function V . We also train a baseline agent using its current position \mathbf{p}_t instead of \mathbf{x}_t , and another agent using raw images from scratch.

agent to explore the environment gradually in order to find increasingly distant destinations [2], [27], [28]. Our sparse reward function gives the agent a reward of 1 only when it finds the target, potentially receiving a punishment of $-1/\text{ms}$ when it heads away from the required destination (ms being the maximum number of agent's steps per episode).

C. Evaluation Metrics

We evaluate the performance of our visual feature observations, obtained as described in Sec. IV-A, using *area under the curve* (AUC) metrics across both the place recognition and deep learning models over a range of feature vector dimensions. We start by training a classifier for each database traversal using a single-layer MLP that takes our feature observations as an input. We then use this classifier and evaluate it over the remaining query traversals. Once we have the scores for both query and database, we compute the precision-recall curves from where we can obtain the overall AUC performance.

For our navigation policy, we evaluate its performance over the traversals used for training as well as two different traversals including challenging appearance changes such as day to afternoon or night for the Oxford RobotCar and summer to winter or fall for the Nordland dataset. During this evaluation, we limit the maximum number of agent steps in an episode to the number of frames within the traversal, measuring in this way how well the agent can find a target destination with a moderate, environment-appropriate number of steps. We provide statistics of the number of navigation tasks that our policy can achieve by reporting the percentage of evaluation episodes in two categories: (1) *completed* tasks, when the agent gets the target using the minimum number of steps as defined above, or (2) *failed* tasks, otherwise.

D. Agent Baselines

In order to have different agent baselines to compare with, we present two additional agent architectures: baseline and raw images as shown in Fig. 4 (a) and (c) respectively. For baseline, we define a relatively trivial navigation agent that uses its current position \mathbf{p}_t instead of \mathbf{x}_t ; while this substantially simplifies the problem, it is a competitive baseline reference since it achieves 100% completed tasks on deployment. In contrast, the raw images



Fig. 5. The two diversified real-world benchmark datasets used in our experiments. **Nordland** (left to right-center): summer, fall and winter traversals. **Oxford RobotCar** (left-center to right): day, overcast and night traversals.

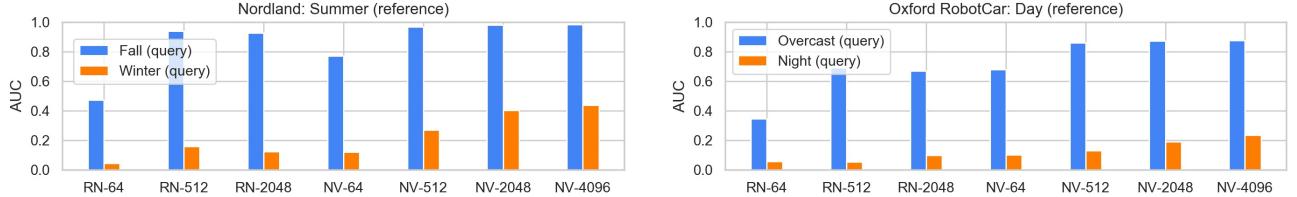


Fig. 6. AUC place recognition performance on **Nordland** (left) and **Oxford Robotcar** (right) datasets over moderate and extreme visual changes using off-the-shelf place recognition (NetVLAD: NV) and deep learning (ResNet-50: RN) methods with a range of feature dimensions (64, 512, 2048, 4096).

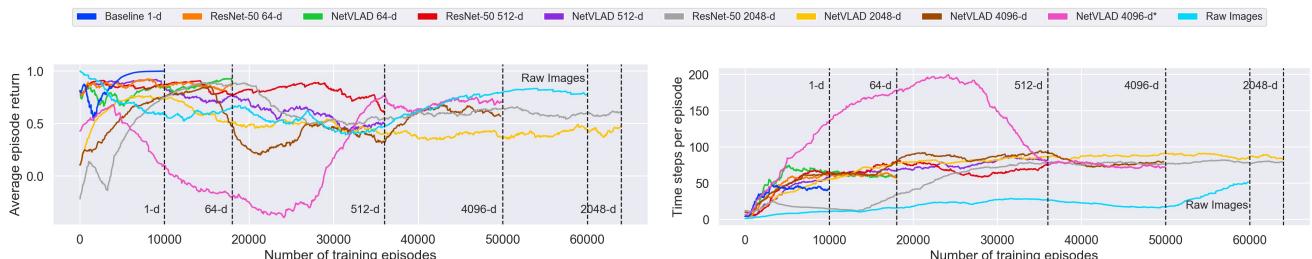


Fig. 7. Reinforcement learning training curves using compact, bimodal representations: goal destination and image features observations feedback modalities. Visual place recognition (NetVLAD) and deep learning (ResNet-50) techniques encoded our input images into 64-d, 512-d, 2048-d, 4096-d.

agent uses a CNN visual module of 2 *convolutional* layers, as in previous works [3], [30], with inputs consisting of RGB 84×84 images. The implementation code of these two agent baselines including our approach, shown in Fig. 4, are made publicly available alongside with the CityLearn environment.

V. EXPERIMENTS: RESULTS

We first conduct conventional visual place recognition experiments by using our visual feature representations across two stationary real-world datasets including challenging environmental changes between the provided traversals (see Fig. 5). We then use these compact place representations to train our policy network for efficiently learning goal-driven navigation tasks using our CityLearn environment, generalizing across the extreme appearance changes of the two datasets.

A. Place Recognition Experiments

To investigate the trade-off of using compact visual feature representation dimensions for place recognition tasks, we present the results of our single-frame matching experiments, as described in Secs. IV-A and IV-C, in Fig. 6 across our two datasets. From that Fig. we observe how the AUC performance decreases as we decrease the feature dimension from 4096-d all the way to 64-d in both NetVLAD and ResNet-50 models.

From Fig. 6 we can observe how well these networks generalize when facing small appearance variations such as summer to fall for Nordland (see Fig. 6 left). Day/sunny to overcast for Oxford RobotCar provides moderate viewpoint changes (see Fig. 6 right), so the global performance is

lower than for Nordland as it does not include viewpoint changes. In contrast, for extreme appearance changes, such as summer to winter or day to night respectively, we can observe that the global AUC performance is compromised by reducing it to less than a half for Nordland or even to less than a quarter for Oxford RobotCar when compared to small appearance changes. It is worth noting we are performing only a single-frame matching procedure here; the results may not be as good as expected for these state-of-the-art techniques, since multi-frame algorithmic techniques are typically incorporated on top of those single-frame results, as previously described in Sec. II-B.

B. Sample-Efficient Navigation Policy Training

We illustrate the reinforcement learning training curves in Fig. 7; complete-related visualization as a function of the required training time is presented in Fig. 2. In Fig. 7-left we observe that our approach using 64-d compact representations achieves comparable average reward performance to the baseline agent, being 92% for NetVLAD, 80% for ResNet-50 and 99% for the baseline agent. This small difference between these three agents is reflected in Fig. 7-right, where the number of agent steps stabilize slightly below 50 at 10,000 steps for the baseline agent, while for the remaining two agents (NetVLAD and ResNet-50 with 64-d) this occurs slightly above 50 steps at 18,000 episodes.

This behavior is consequently observed again in Fig. 7, as we increase the visual feature dimensions from 512-d to 4096-d. The final average number of steps for these agents is around 75 and the required number of training episodes increases as we increase the feature dimension, except for

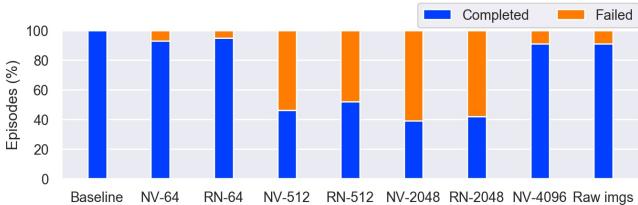


Fig. 8. Navigation policy evaluation statistics on the traversal it was trained. **Nordland** (left): summer, **Oxford RoborCar** (right): day.

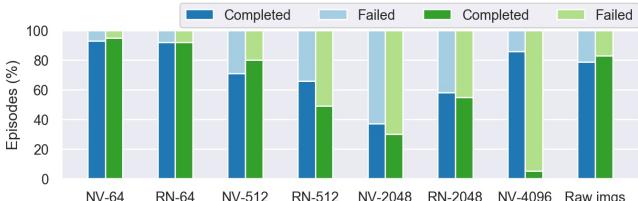


Fig. 9. **Generalization of our navigation policy.** Evaluation statistics over moderated (blue) and extreme (green) appearance environmental changes. For **Nordland** (left): fall (blue) and winter (green) traversals. For **Oxford RoborCar** (right): overcast (blue) and night (green).

4096-d that stabilizes at 50,000 episodes which is lower than 2048-d, which requires 60,000 episodes. This is why we have two training results for 4096-d where we show again this behavior in the curve 4096-d*. It is worth noting the training curves in Fig. 7 were obtained by averaging 5 trials each using different seed numbers, and then applying curve smoothing with weight 0.9 to enable cleaner visualization of our results. In all our experiments, the number of concurrent agents we used for training in our CityLearn environment was 32, all of them interacting with the same policy network presented in Fig. 4, only changing the dimension of the input as our bimodal representations b_t changes its dimensions according to the visual feature x_t .

C. Deployment and Generalization Evaluations

We report evaluation statistics of our trained navigation policy on both the database traversal used for training (see Fig. 8) and query traversals used to test their generalization performance (see Fig. 9) across our two datasets alongside with the CityLearn environment: Nordland (left) and Oxford RobotCar (right). We evaluated our trained stochastic policy every 100 episodes and calculated the number of *completed* and *failed* navigation tasks. From Figs. 8 and 9, it can be observed that when using compact representations (64-d) we can achieve better generalization results even under extreme environmental changes, such as summer to winter (see Fig. 9-left for Nordland) or day to night (see Fig. 9-right for Oxford RobotCar). While increasing the feature dimension in visual place recognition tasks reflect in better results in terms of AUC performance, as demonstrated in Fig. 6, for our reinforcement learning navigation tasks the opposite seems to occur; smaller representations are better for both final average performance and sample efficiency in terms of training time and number of episodes, as well as in generalization capabilities across extreme environmental changes, at least in an RL context.

Fig. 10 shows deployment comparisons between our approach and an agent trained using raw images in a route of the Oxford RobotCar dataset. Both agents starting at the same location with a common goal.

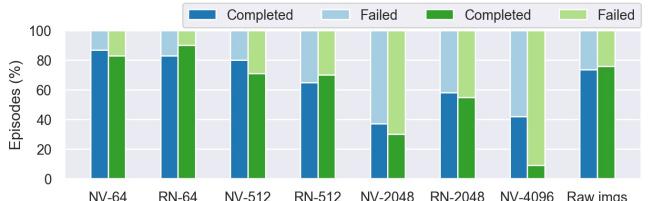
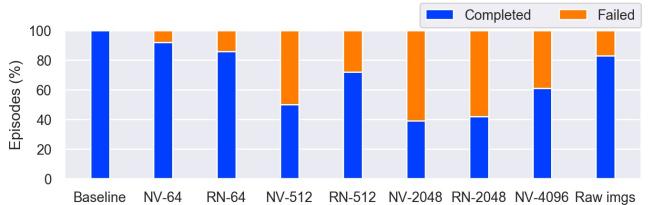
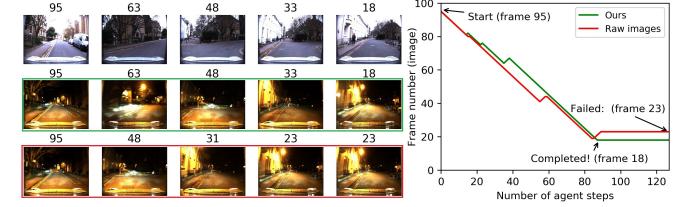


Fig. 10. **Deployment comparison.** The policies were trained on the Oxford Robotcar dataset (day: top-left) and evaluated under extreme visual changes (night). Our approach completed the task using NetVLAD 64-d (center-left), while the Raw images agent failed (bottom-left).

VI. CONCLUSIONS

We conducted comprehensive experiments applying place recognition and reinforcement learning techniques to examine the value of using visual and self-motion (in terms of the agent previous actions) sensory feedback to learn navigation policies at kilometer scale on two benchmark robotic datasets with challenging appearance changes. To enable efficient reinforcement learning training, we use place recognition techniques to encode our real-world sensory inputs that, combined with the goal destination, generates a compact bimodal representation. Once trained, we showed that smaller representations such as 64-d generalized better than larger representations over a range of environmental changes including day to night or summer to winter, while being around 2 orders of magnitude faster and requiring a small fraction of the amount of experience in terms of training time and number of iterations.

Our presented interactive environment, CityLearn, can also be used to load any other benchmark dataset (or even custom in-house recorded data), such as those from drones or under-water robots, to evaluate many different types of navigation algorithms as well as to further build and investigate the performance of advanced RL algorithms using real-world images. Future research could include other RL algorithms such as [30], modular architectures for transfer learning to new cities [36], and adding more functionalities to the environment such as creating a 2D map.



REFERENCES

- [1] S. Thrun, W. Burgard and D. Fox. Probabilistic Robotics. MIT Press, 2005.
- [2] Mirowski, P.W., Grimes, M.K., Malinowski, M., Hermann, K.M., Anderson, K., Teplyashin, D., Simonyan, K., Kavukcuoglu, K., Zisserman, A., & Hadsell, R. (2018). Learning to Navigate in Cities Without a Map. NeurIPS.
- [3] Piotr Mirowski, Razvan Pascanu, Fabio Viola, Hubert Soyer, et al. Learning to navigate in complex environments. ICLR, 2017.
- [4] Beattie, C., Leibo, J.Z., Teplyashin, D., Ward, T., Wainwright, M., Ktler, H., Lefrancq, A., Green, S., Valds, V., Sadik, A., Schriftwieser, J., Anderson, K., York, S., Cant, M., Cain, A., Bolton, A., Gaffney, S., King, H., Hassabis, D., Legg, S., & Petersen, S. (2016). “DeepMind Lab.” ArXiv, abs/1612.03801.
- [5] R. Arandjelović, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, “NetVLAD: CNN Architecture for Weakly Supervised Place Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5297–5307.
- [6] Z. Chen, A. Jacobson, N. Sünderhauf, B. Upcroft, L. Liu, C. Shen, I. D. Reid, and M. Milford, “Deep learning features at scale for visual place recognition,” *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3223–3230, 2017.
- [7] S. Garg, N. Sünderhauf, and M. Milford, “LoST? Appearance-Invariant Place Recognition for Opposite Viewpoints using Visual Semantics,” *Proceedings of Robotics: Science and Systems XIV*, 2018.
- [8] S. Hausler, A. Jacobson and M. Milford, “Multi-Process Fusion: Visual Place Recognition Using Multiple Image Processing Methods,” in *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1924–1931, April 2019.
- [9] Durrant-Whyte, H. F. (1996). An Autonomous Guided Vehicle for Cargo Handling Applications. *The International Journal of Robotics Research*, 15(5), 407440.
- [10] R. Nourbakhsh, Illah & Bobenage, Judith & Grange, Sebastien & Lutz, Ron & Meyer, Roland & Soto, Alvaro. (1999). An affective mobile robot educator with a full-time job. *Artificial Intelligence*. 114. 95-124.
- [11] F. Dellaert, D. Fox, W. Burgard, and S. Thrun. Monte carlo localization for mobile robots. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, Leuven, Belgium, 1998.
- [12] Lin, L.-J. (1993). Reinforcement Learning for Robots Using Neural Networks. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA.
- [13] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep Learning,” *Nature*, vol. 521, pp. 436–444, Mar. 2015.
- [14] Bruce, J., Sünderhauf, N., Mirowski, P.W., Hadsell, R., & Milford, M. (2018). Learning Deployable Navigation Policies at Kilometer Scale from a Single Traversal. CoRL.
- [15] M. J. Milford and G. F. Wyeth, “SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights,” in *2012 IEEE International Conference on Robotics and Automation (ICRA)*, 2012.
- [16] K. He, X. Zhang, S. Ren and J. Sun, “Deep Residual Learning for Image Recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 770–778.
- [17] K. Simonyan, and A. Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” *CoRR* abs/1409.1556, 2015.
- [18] N. Sünderhauf, P. Neubert, and P. Protzel, “Are we there yet? Challenging SeqSLAM on a 3000 km journey across all four seasons,” in *IEEE International Conference on Robotics and Automation (ICRA)*, Karlsruhe, Germany, 2013.
- [19] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O, “Proximal Policy Optimization Algorithms.” ArXiv, abs/1707.06347. 2017.
- [20] Schulman, J., Levine, S., Abbeel, P., Jordan, M.I., & Moritz, P. “Trust Region Policy Optimization.” ICML, 2015.
- [21] S. Brahmbhatt and J. Hays, “DeepNav: Learning to Navigate Large Cities,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 3087-3096. doi: 10.1109/CVPR.2017.329
- [22] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, 2009, pp. 248–255.
- [23] Milford, Michael. (2013). Vision-based place recognition: How low can you go?. *The International Journal of Robotics Research*. 32. 766–789. 10.1177/0278364913490323.
- [24] Yu, F., Xian, W., Chen, Y., Liu, F., Liao, M., Madhavan, V., & Darrell, T. (2018). BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling. ArXiv, abs/1805.04687.
- [25] Xu, H., Gao, Y., Yu, F., & Darrell, T. (2016). End-to-End Learning of Driving Models from Large-Scale Video Datasets. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3530–3538.
- [26] Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Lioung, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., & Beijbom, O. (2019). nuScenes: A multimodal dataset for autonomous driving. ArXiv, abs/1903.11027.
- [27] Bengio, Y., Louradour, J., Collobert, R., Weston, J. Curriculum learning. ICML, 2009.
- [28] Wojciech Zaremba, Ilya Sutskever. “Learning to execute,” arXiv preprint arXiv:1410.4615, 2014.
- [29] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *International Conference on Machine Learning*, pages 1928–1937, 2016.
- [30] Lasse Espeholt, Hubert Soyer, Remi Munos, et al. Impala: Scalable distributed deep-rl with importance weighted actor-learner architectures. ICML, 2018.
- [31] S. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, “Visual Place Recognition: A Survey,” in *IEEE Transactions on Robotics*, vol. 32, no. 1, pp. 1–19, Feb. 2016.
- [32] M. Zaffar, A. Khalil, S. Ehsan, M. Milford, & K.D. McDonald-Maier, “Levelling the Playing Field: A Comprehensive Comparison of Visual Place Recognition Approaches under Changing Conditions,” ArXiv, abs/1903.09107.
- [33] Vries, H.D., Shuster, K., Batra, D., Parikh, D., Weston, J., & Kiela, D. (2018). Talk the Walk: Navigating New York City through Grounded Dialogue. ArXiv, abs/1807.03367.
- [34] Volkan Cirik, Yuan Zhang, and Jason Baldridge. Following formulaic map instructions in a street simulation environment. Visually Grounded Interaction and Language (ViGIL) Workshop, NeurIPS, 2018.
- [35] Chen, H., Suhr, A., Misra, D.K., Snavely, N., & Artzi, Y. (2018). Touchdown: Natural Language Navigation and Spatial Reasoning in Visual Street Environments. ArXiv, abs/1811.12354.
- [36] Devin, C., Gupta, A., Darrell, T., Abbeel, P., & Levine, S. (2016). Learning modular neural network policies for multi-task and multi-robot transfer. *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2169–2176.
- [37] Zhu, Y., Mottaghi, R., Kolve, E., Lim, J.J., Gupta, A., Fei-Fei, L., & Farhadi, A. (2016). Target-driven visual navigation in indoor scenes using deep reinforcement learning. *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 3357–3364.
- [38] Weyand, T., Kostrikov, I., & Philbin, J. (2016). PlaNet - Photo Geolocation with Convolutional Neural Networks. ECCV.
- [39] A. Khosla, B. An, J. J. Lim and A. Torralba, “Looking Beyond the Visible Scene,” 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, 2014, pp. 3710–3717.
- [40] R. F. Berriel et al., “Heading Direction Estimation Using Deep Learning with Automatic Large-scale Data Acquisition,” 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, 2018, pp. 1–8.
- [41] M. J. Milford, G. F. Wyeth, and D. Prasser, “RatSLAM: a hippocampal model for simultaneous localization and mapping,” in *IEEE International Conference on Robotics and Automation (ICRA)*, vol. 1, pp. 403–408, 2004.
- [42] M. J. Milford and G. F. Wyeth, “Mapping a Suburb With a Single Camera Using a Biologically Inspired SLAM System,” in *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1038–1053, Oct. 2008.
- [43] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, “1 Year, 1000km: The Oxford RobotCar Dataset,” *The International Journal of Robotics Research (IJRR)*, vol. 36, no. 1, pp. 3–15, 2017.
- [44] Z. Chen, F. Maffra, I. Sa, and M. Chli, “Only look once, mining distinctive landmarks from convnet for visual place recognition,” in *IROS*, pp. 916, IEEE, 2017.
- [45] N. Sünderhauf, S. Shirazi, A. Jacobson, F. Dayoub, E. Pepperell, B. Upcroft, and M. Milford, “Place Recognition with ConvNet Landmarks: Viewpoint-Robust, Condition-Robust, Training-Free,” *Proceedings of Robotics: Science and Systems XIV*, 2015.
- [46] A. J. Glover, W. P. Maddern, M. J. Milford, and G. F. Wyeth, “FAB-

- MAP+RatSLAM: Appearance-based SLAM for multiple times of day, in *ICRA*, pp. 35073512, IEEE, 2010.
- [47] M. Milford and et al., “Sequence searching with deep-learnt depth for condition-and viewpoint-invariant route-based place recognition,” in *CVPR Workshops*, pp. 1825, 2015.
- [48] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The KITTI dataset,” *IJRR*, 2013.
- [49] T. Naseer, W. Burgard, and C. Stachniss, “Robust visual localization across seasons,” *IEEE T-RO*, vol. 34, no. 2, pp. 289302, 2018.
- [50] M. Warren, D. McKinnon, H. He, and B. Upcroft, “Unaided stereo vision based pose estimation,” in *ACRA*, 2010.
- [51] M. Kempka, M. Wydmuch, G. Runc, J. Toczek and W. Jakowski, “ViZDoom: A Doom-based AI research platform for visual reinforcement learning,” 2016 IEEE Conference on Computational Intelligence and Games (CIG), Santorini, 2016, pp. 1-8.
- [52] Lample, G., & Chaplot, D.S. (2016). “Playing FPS Games with Deep Reinforcement Learning.” ArXiv, abs/1609.05521.
- [53] Khan, A., Zhang, C., Atanasov, N., Karydis, K., Kumar, V.S., & Lee, D.D. (2017). “Memory Augmented Control Networks.” ICLR ArXiv, abs/1709.05706.
- [54] Chaplot, D.S., Parisotto, E., & Salakhutdinov, R. (2018). Active Neural Localization. ICLR ArXiv, abs/1801.08214.
- [55] S. Hochreiter and J. Schmidhuber. “Long short-term memory,” *Neural computation*, 9(8):17351780, 1997.