

MetaNav: A Bio-Inspired Meta-Imitation Learning Framework for Robotic Visual Semantic Navigation

Carlos Gutiérrez-Álvarez¹[0000–0002–4727–1575], Rafael
Flor-Rodríguez¹[0009–0000–5120–0334], Sergio
Lafuente-Arroyo¹[0000–0002–4841–2501], Saturnino
Maldonado-Bascón¹[0000–0001–6472–5359], and Roberto J.
López-Sastre¹[0000–0002–2477–0152]

University of Alcalá, Department of Signal Theory and Communications, GRAM
Research Group, Alcalá de Henares, Spain
`{carlos.gutierrezalva}@uah.es`
<https://gram.web.uah.es/>

Abstract. In this work we address the robotic visual semantic navigation problem from a bio-inspired learning perspective, motivated by the ability of natural agents to rapidly adapt their behavior by exploiting prior experience. We explore meta-imitation learning as an artificial counterpart of biological learning-to-learn mechanisms, enabling fast adaptation to new tasks from limited demonstrations. Technically, we propose the model Meta Visual Semantic Navigation (MetaNav), a meta-imitation learning framework that combines imitation learning from human demonstrations with meta-learning to acquire adaptable navigation policies. Rather than optimizing for a single task, the model learns an adaptive prior that can be efficiently specialized to new object-goal navigation tasks using only a few examples. The approach is trained on a small set of environments from the HM3D dataset following a task-based formulation inspired by natural adaptation across related scenarios. Experimental results show that the learned policies exhibit fast adaptation to previously unseen tasks, highlighting the potential of bio-inspired meta-learning for robotic navigation, while also revealing current limitations and directions for future research.

Keywords: Imitation Learning · Meta-Learning · Visual Navigation.

1 Introduction

Autonomous navigation continues to be a core challenge in robotics, especially in unstructured and dynamic environments. In this context, Visual Semantic Navigation (VSN) has emerged as a promising paradigm, driven by recent advances in machine learning and the availability of large-scale datasets, with the aim of explicitly addressing these challenges. VSN models are predominantly learning-based navigation approaches that employ end-to-end architectures to

directly map visual sensory inputs to control actions governing robot motion (e.g. [13, 18]). Despite their effectiveness, current VSN approaches typically rely on large amounts of training data and extensive interaction with highly realistic simulation environments in order to achieve acceptable performance and generalization. Moreover, models trained under these conditions often exhibit limited adaptability when deployed in novel environments or confronted with unseen goals, requiring additional data collection or retraining. These constraints highlight a fundamental gap between the data efficiency of natural navigation systems and that of current learning-based VSN models, motivating the exploration of alternative learning paradigms that emphasize rapid adaptation and experience reuse.

In contrast, in this work, we explore a novel training paradigm for VSN models that pursues two main objectives: first, to achieve more efficient training pipelines; and second, to learn navigation models with improved generalization capabilities.

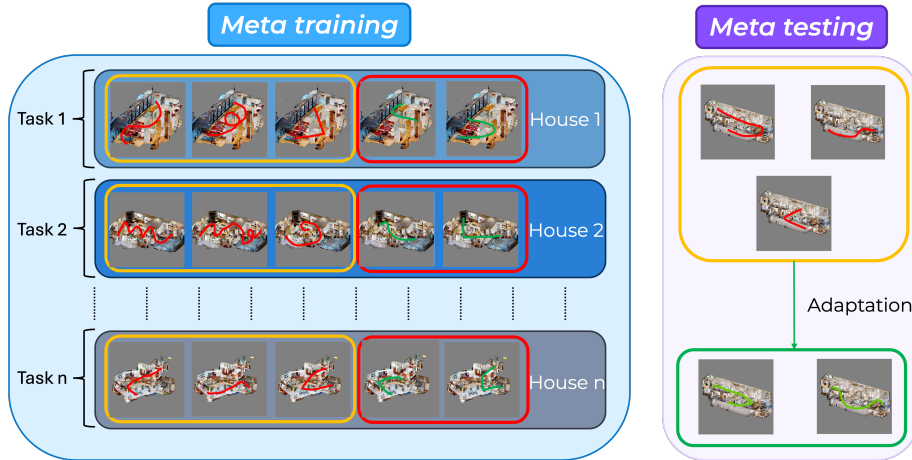


Fig. 1. MetaNav reformulates VSN as a meta-imitation learning problem. MetaNav learns a navigation policy prior across multiple training tasks (meta-training) and enables fast adaptation to unseen object-goal navigation tasks with few demonstrations (meta-testing), achieving both generalization and rapid specialization.

Technically, we propose a bio-inspired approach that combines techniques from Imitation Learning (IL) and meta-learning to learn a VSN model, which we denote as MetaNav, see figure 1. Humans and animals rapidly acquire navigation skills by observing demonstrations and adapt them to new environments after only limited exposure, a capability commonly described as learning to learn. Meta-learning [5] and imitation learning [17] offer a computational abstraction of these bio-inspired mechanisms in artificial systems. Imitation learning enables robotic agents to acquire navigation behaviors from expert demonstrations, while meta-learning allows such behaviors to be organized into adaptive priors that fa-

cilitate fast specialization to new scenes and goals. Thus, meta-imitation learning is a combination of both realms, allowing agents to learn from a small number of demonstrations and adapt quickly to new tasks, in our case navigation tasks. Our hypothesis is that in the context of robotic navigation, this combination naturally supports efficient generalization and rapid adaptation across environments, reflecting key principles of biological navigation.

MetaNav (figure 1) reformulates VSN as a meta-imitation learning problem. During the meta-training phase, the model learns a navigation policy prior across multiple small object-goal navigation tasks defined in different training environments. In the meta-testing phase, the learned prior enables fast adaptation to new navigation tasks in previously unseen environments using only a small amount of demonstration data. The objective is to obtain a policy that is broadly effective across tasks while retaining the ability to specialize rapidly in novel navigation scenarios. Due to this adaptability, the policy can be trained on simulation using a set of task demonstrations and then deployed in unseen environments to perform new tasks.

The main contributions of this work are summarized as follows:

1. We implement a meta-imitation learning algorithm inspired by prior work on one-shot visual imitation learning and adapt it to the specific challenges of robotic visual navigation.
2. The proposed approach is trained and evaluated on a subset of large-scale 3D environments from the HM3D dataset, analyzing its adaptation capabilities across different experimental setups.
3. Experimental results show that the learned policies are capable of fast adaptation to previously unseen navigation tasks with limited examples.
4. To the best of our knowledge, this is the first work that applies meta-imitation learning to visual semantic navigation in large-scale environments, directly predicting navigation actions from raw sensory observations.

2 Related Work

Meta-imitation learning extends imitation learning into the meta-learning regime. Depending on the method used, two main approaches can be distinguished: gradient-based and contextual and latent variable meta-imitation learning.

Gradient-based Meta Imitation Learning. Many meta-imitation learning approaches have adopted gradient-based adaptation for fast imitation. In these methods, a policy (or policy initialization) is meta-learned such that a small number of gradient steps on a new task’s demo data leads to effective behavior cloning on that task. For example, Model Agnostic Meta-Learning (MAML) [5] learns a policy initialization that can be quickly adapted to new tasks via gradient descent. Later works have extended MAML to the context of imitation learning, such as learning from video demonstrations of robotic tasks [6, 7] or learning from human demonstrations in complex environments [16, 14].

Contextual and Latent Variable Meta-Imitation. An alternative line of work avoids explicit gradient updates at test time by conditioning the policy on

task-specific context, typically derived from demonstrations. Typically, one-shot imitation learning models process a demonstration trajectory (or a few demos) through an encoder and feed this context alongside the current state to the policy network [4]. Another instantiation of this idea is to use latent variables [11, 15, 8, 2, 3] to represent tasks: during meta-training, a latent context z is inferred for each demonstration set, and the policy or value function is conditioned on z rather than relying on one global initialization.

Our approach, described in section 3, extends the concepts of gradient-based meta-imitation learning by implementing a modified version of MAML [5] to train navigation policies using few demonstrations.

3 Meta-Imitation Learning for Robotic Visual Semantic Navigation

3.1 Problem setting

We address the problem of ObjectGoal Navigation [1], in which an embodied agent must navigate toward a specified target object by selecting actions from a discrete control set. The available actions include: `TURN_LEFT`, `TURN_RIGHT`, `MOVE_FORWARD`, `LOOK_UP`, `LOOK_DOWN`, and `STOP`. An episode is considered successful when the agent issues the `STOP` action within one meter of the target object and before reaching a maximum of 500 steps.

The agent operates exclusively on egocentric sensory inputs, as is standard in Visual Semantic Navigation (VSN). Specifically, it receives RGB visual observations together with relative displacement and orientation measurements, providing positional information with respect to the starting point. This perception-action loop mirrors the behavior of biological agents, which rely on first-person visual input and self-motion cues to incrementally construct spatial understanding and guide goal-directed navigation.

Since an imitation learning setup is proposed, a previously collected dataset of navigation experience is needed. The dataset that was chosen is collected in [10]. It consists of 77k episodes of human navigation trajectories using the HM3D [9] dataset. This dataset consists of a set of photorealistic 3D indoor environments, each containing a variety of objects and different scenes.

A navigation task is defined as a tuple of the form $\mathcal{T} = (G, S)$, where G is the target object and S the scene in which the navigation is taking place. Each task can contain multiple demonstrations τ_i , each of them recorded by a human and always on the same scene and seeking the same target object. The variability of the demonstrations comes from the random starting points in each episode and the different paths that the human can take to reach any of the several instances of the target object. Thus, the goal of MetaNav is to learn a policy π_θ that can quickly adapt to combinations of new target objects G and scenes S .

3.2 Meta-imitation learning formulation

According to the problem formulation, the proposed MetaNav learning algorithm is described in this section. We propose a meta-imitation learning approach leveraging prior work on one-shot visual imitation learning with meta-learning [6].

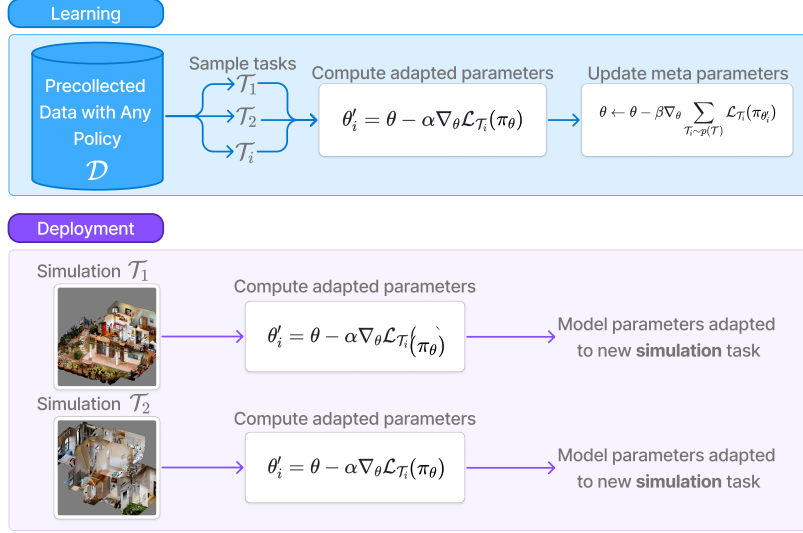


Fig. 2. MetaNav reformulates visual semantic navigation as a MAML-based meta-imitation learning process. During meta-training, task-specific parameters are obtained from demonstration data and used to update shared meta-parameters across navigation tasks. In deployment, the learned policy prior is quickly adapted to unseen environments with minimal additional experience, enabling efficient generalization and rapid task specialization.

Technically, we develop MetaNav as a VSN model that is trained following a MAML-based [5] meta-imitation learning process. This way, our vision-based navigation policy is trained to adapt to new navigation tasks with a few examples, in two stages, as it is shown in figure 2. First, during the meta-training stage, a set of tasks $\mathcal{T}_i = (G, S)$ drawn from a distribution $p(\mathcal{T})$ are considered. Each task $\mathcal{T}_i$ consists of a collection of observations o and actions a generated by an expert policy π_i^* :

$$\tau = \{o_1, a_1, \dots, o_T, a_T\} \sim \pi_i^*, \quad (1)$$

where T is the length of the demonstration or trajectory τ . Our objective is to learn a policy π_{θ} that maps observations o to predicted actions $\hat{a}$ from different demonstrations sampled from the task distribution. Since an imitation learning context with discrete actions is used, behavior cloning loss is employed to train

the policy from the demonstrations:

$$\mathcal{L}_{\mathcal{T}_i}^*(\pi_\theta) = \sum_{\tau_j \sim \mathcal{T}_i} \sum_{(o_t, a_t) \in \tau^i}^T -\log(\pi_\theta(a_t|o_T)). \quad (2)$$

This loss function is known as the inner loss, used during the meta-training stage, and its function is to adapt the policy to a specific task. By applying gradient descent to the inner loss, the adapted parameters $\theta'_i = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}^*(\pi_\theta)$ of the policy can be computed. Inflection weighting [12] is applied to the inner loss function to emphasize time steps corresponding to action changes.

In a second stage, during the meta-testing stage, once the adapted parameters θ'_i are obtained, they can be used to compute the outer loss, which is used to update the policy parameters θ . The outer loss is also known as the meta-loss or meta-objective, and it is defined as:

$$\min_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(\pi_{\theta'_i}) = \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(\pi_{\theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}^*(\pi_\theta)}). \quad (3)$$

Intuitively, this meta-loss acts as a regularization term over a “fine-tuning” process over the tasks. This encourages the policy to find an average over θ meta-parameters with respect to the task distribution. Then, the meta-parameters can adapt to new tasks with few gradient updates, since the distance between the adapted parameters θ'_i and the original parameters θ is being minimized across a set of tasks. Note that for this goal, the meta-loss implies the use of double derivatives, which can be computationally expensive. In practice, a first-order approximation of the meta-loss [5] is used, which allows computing the outer loss without the need for second derivatives.

From a bio-inspired perspective, this formulation reflects the way natural agents acquire reusable navigation priors through experience and subsequently refine their behavior when exposed to new environments. The meta-parameters θ can be interpreted as a learned adaptive prior, analogous to long-term knowledge accumulated across tasks, while the task-specific updates model rapid short-term adaptation. Thus, MetaNav embodies a computational abstraction of learning-to-learn mechanisms observed in biological navigation systems, emphasizing experience reuse and efficient behavioral specialization.

4 Experiments

4.1 Experimental setup

Dataset. Since an imitation learning approach is followed, a previously collected dataset of navigation experience is needed. The dataset that was chosen for our experimental evaluation is a subset of the data collected in [10]. It consists of 77k episodes of human navigation trajectories using the HM3D [9] dataset. For our experiments, we design five progressively challenging setups that define how training and evaluation scenes are organized within the dataset, as illustrated

in Figure 3. The goal is to gradually increase the generalization difficulty from Setup 1 through Setup 5.

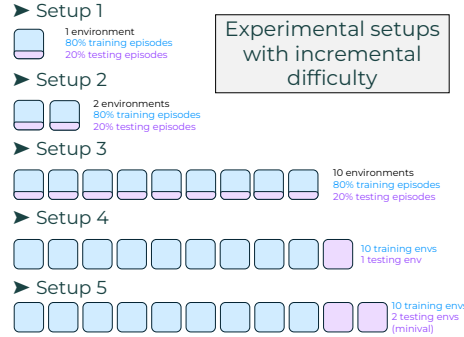


Fig. 3. Five experimental setups designed with an incremental difficulty.

Implementation details. For MetaNav, the policy architecture follows a CNN+RNN pipeline similar to the model described in [10], with the modification that a ResNet18 backbone is used for the visual encoders to improve feature extraction efficiency. The overall training framework builds on the decentralized distributed paradigm of DDPPPO [13], enabling large-scale parallel rollout collection and stable on-policy updates.

To adapt DDPPPO to the meta-imitation learning setting, we modify the inner optimization step to incorporate the behavior cloning loss defined in Eq. 2, and introduce the meta-objective (outer loss) described in Eq. 3. This integration allows the policy to learn from expert demonstrations while simultaneously optimizing for fast adaptation across tasks.

4.2 Evaluation strategies

Since MetaNav is trained under a meta-learning paradigm, the agent is exposed to a set of previously unseen tasks. For each task, the agent first must perform an adaptation phase, where it updates its policy using demonstrations from that task, yielding the adapted parameters θ'_i . The adapted policy is then evaluated by executing actions sampled from it and measuring downstream navigation performance.

We evaluate MetaNav using two complementary protocols: *continuous evaluation* and *per-episode evaluation*. In the continuous evaluation setting, the agent is given a fixed number of adaptation steps for each task, after which it is immediately evaluated on the same task for an additional fixed number of steps. This process is repeated continuously across the episode until it is finished. In the per-episode evaluation protocol, the agent again receives a fixed adaptation budget, but evaluation proceeds by running the adapted policy until each episode nat-

Table 1. Evaluation of MetaNav on the OBJECTNAV task using continuous and per-episode evaluation strategies.

Setup	Continuous Evaluation			Per-Episode Evaluation		
	SR (↑)	SPL (↑)	Dist. (↓)	SR (↑)	SPL (↑)	Dist. (↓)
1	89.18%	40.04%	0.29	83.33%	40.03%	0.29
2	76.10%	33.92%	0.97	60.78%	26.58%	1.74
3	64.19%	33.11%	1.99	55.19%	26.21%	2.54
4	23.07%	11.87%	12.23	16.67%	4.84%	12.72
5	21.74%	9.38%	7.99	25.00%	9.31%	8.19

usually terminates, either due to success or upon reaching the maximum allowed number of steps.

Navigation performance is assessed using the official ObjectGoal Navigation metrics [1]: Success Rate (SR) and Shortest Path Length (SPL). We additionally report the Distance to Goal, defined as the agent’s final Euclidean distance to the target at the end of each episode.

4.3 Results

We explicitly compare the two evaluation protocols introduced earlier—continuous evaluation and per-episode evaluation—in order to understand their impact on adaptation dynamics and final navigation performance. All metrics corresponding to this analysis are reported in Table 1.

The first trend we observe is that task difficulty indeed increases progressively from Setup 1 to Setup 5. Under both evaluation protocols, MetaNav starts with a strong SR, exceeding 83% in Setup 1. However, performance drops as the scenarios become more challenging: in Setup 5, the SR decreases to 21.74% under continuous evaluation and to 25% under the per-episode protocol. This behavior confirms that the proposed sequence of setups effectively induces a controlled increase in generalization difficulty.

These experiments also confirm that our meta-imitation learning framework is capable of effectively training a VSN model such as MetaNav. For instance, in Setup 3, MetaNav successfully navigates to the goal in up to 10 different environments with a success rate of 64%, demonstrating strong generalization within a moderately diverse set of tasks.

What are the main differences between the two evaluation strategies proposed? Table 1 reveals systematic differences between continuous and per-episode evaluation. Although performance varies across setups, continuous evaluation consistently achieves higher SR and SPL values than per-episode evaluation in all configurations except Setup 5. This behavior can be explained by the training protocol: MetaNav is trained in a continuous meta-learning fashion, which naturally favors evaluation scenarios where adaptation and execution are interleaved over extended horizons. Consequently, the evaluation setting better matches the training dynamics, yielding improved performance.

Another key factor lies in the effective adaptation budget. In the per-episode protocol, the agent is restricted to a single block of 64 steps for adaptation before executing the task until termination. In contrast, continuous evaluation alternates blocks of 64 adaptation steps and 64 evaluation steps, and if the episode is not completed, this process continues iteratively until termination. As a result, the model is exposed to substantially more experience per episode, enabling incremental refinement of the adapted parameters.

Finally, the results from Setup 5 are also encouraging. Even when evaluated on entirely unseen environments, MetaNav achieves a Success Rate of 25%, indicating that the model is able to leverage the meta-learned prior to perform meaningful adaptation beyond the training distribution. Interestingly, we can compare the performance of our MetaNav approach, with a the state-of-the-art model PirlNav [10] in Setup 5. While our meta-imitation learning reports a SR of 25.00%, PirlNav casts a 26.09% of SR. This reveals that our meta-imitation learning approach is able to report state-of-the-art results. Moreover, it is important to note that MetaNav is trained under a meta-learning paradigm, which allows us to adapt to new tasks with limited examples, while PirlNav is trained in a standard behavior cloning setup without any adaptation mechanism. While PirlNav achieves slightly better performance, MetaNav’s ability to adapt to new tasks with limited examples represents a significant ability that is not present in PirlNav, and it may lead to better performance in other setups or scenarios where adaptability is crucial.

5 Conclusions

We presented MetaNav, a meta-imitation learning framework for VSN that formulates object-goal navigation as a learning-to-learn problem. By acquiring a task-agnostic prior across environments, the policy can rapidly specialize to new navigation tasks from few demonstrations. To the best of our knowledge, this is the first application of meta-imitation learning to large-scale VSN with policies predicting actions directly from raw sensory inputs.

Experiments on HM3D show that MetaNav achieves state-of-the-art performance and consistent gains under continuous evaluation, highlighting strong adaptation when incremental experience is available.

Beyond empirical improvements, MetaNav offers a bio-inspired perspective by decoupling prior acquisition from rapid environment-specific adaptation, resembling experience-driven learning and behavioral adjustment in natural agents. Future work will investigate more scalable meta-optimization strategies, improved visual representations, and extensions toward real-world and continual learning scenarios.

Acknowledgments. This work was supported by projects: GUIDANCE-4D, with reference CM/DEMG/2024-028, of CAM-UAH; NAVIGATOR-D, with reference PID2023-148310OB-I00 from the Ministry of Science and Innovation of Spain; and MOEVE Chair at the University of Alcala.

References

1. Batra, D., Gokaslan, A., Kembhavi, A., Maksymets, O., Mottaghi, R., Savva, M., Toshev, A., Wijmans, E.: ObjectNav Revisited: On Evaluation of Embodied Agents Navigating to Objects. In: arXiv (2020)
2. Bhoopchand, A., Brownfield, B., Collister, A., Lago, A.D., Edwards, A., Everett, R., Frechette, A., Oliveira, Y.G., Hughes, E., Mathewson, K.W., Mendolicchio, P., Pawar, J., Pislár, M., Platonov, A., Senter, E., Singh, S., Zacherl, A., Zhang, L.M.: Learning few-shot imitation as cultural transmission. *Nature Communications* (2023)
3. Cho, S., Kim, D., Lee, J., Hong, S.: Meta-controller: Few-shot imitation of unseen embodiments and tasks in continuous control. In: NeurIPS (2024)
4. Duan, Y., Andrychowicz, M., Stadie, B.C., Ho, J., Schneider, J., Sutskever, I., Abbeel, P., Zaremba, W.: One-shot imitation learning. In: NeurIPS (2017)
5. Finn, C., Abbeel, P., Levine, S.: Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In: ICML (2017)
6. Finn, C., Yu, T., Zhang, T., Abbeel, P., Levine, S.: One-Shot Visual Imitation Learning via Meta-Learning. In: *Proceedings of the 1st Annual Conference on Robot Learning*. pp. 357–368. PMLR (Oct 2017)
7. Gao, C., Jiang, Y., Chen, F.: Transferring hierarchical structures with dual meta imitation learning. In: 6th Annual Conference on Robot Learning (2022)
8. Matsushima, T., Kondo, N., Iwasawa, Y., Nasuno, K., Matsuo, Y.: Modeling task uncertainty for safe meta-imitation learning. *Frontiers in Robotics and AI* **Volume 7 - 2020** (2020). <https://doi.org/10.3389/frobt.2020.606361>
9. Ramakrishnan, S.K., Gokaslan, A., Wijmans, E., Maksymets, O., Clegg, A., Turner, J., Undersander, E., Galuba, W., Westbury, A., Chang, A.X., Savva, M., Zhao, Y., Batra, D.: Habitat-Matterport 3D Dataset (HM3D): 1000 large-scale 3D environments for embodied AI. In: NeurIPS (2021)
10. Ramrakhya, R., Batra, D., Wijmans, E., Das, A.: PIRLNav: Pretraining with Imitation and RL Finetuning for ObjectNav. In: CVPR (2023)
11. Sontakke, S.A., Zhang, J., Arnold, S., Pertsch, K., Biyik, E., Sadigh, D., Finn, C., Itti, L.: RoboCLIP: One demonstration is enough to learn robot policies. In: NeurIPS (2023)
12. Wijmans, E., Datta, S., Maksymets, O., Das, A., Gkioxari, G., Lee, S., Essa, I., Parikh, D., Batra, D.: Embodied question answering in photorealistic environments with point cloud perception. In: CVPR (2019)
13. Wijmans, E., Kadian, A., Morcos, A., Lee, S., Essa, I., Parikh, D., Savva, M., Batra, D.: DD-PPO: Learning Near-Perfect PointGoal Navigators from 2.5 Billion Frames. In: ICLR (2020)
14. Wu, S., Wang, Y., Huang, Y.: One-shot robust imitation learning for long-horizon visuomotor tasks from unsegmented demonstrations. arXiv (2024)
15. Yu, L., Yu, T., Finn, C., Ermon, S.: Meta-inverse reinforcement learning with probabilistic context variables. In: NeurIPS (2019)
16. Yu, T., Finn, C., Xie, A., Dasari, S., Zhang, T., Abbeel, P., Levine, S.: One-shot imitation from observing humans via domain-adaptive meta-learning. arXiv (2018)
17. Zare, M., Kebria, P.M., Khosravi, A., Nahavandi, S.: A survey of imitation learning: Algorithms, recent developments, and challenges. *IEEE Transactions on Cybernetics* **54**(12), 7173–7186 (2024). <https://doi.org/10.1109/TCYB.2024.3395626>
18. Zhu, Y., Mottaghi, R., Kolve, E., Lim, J.J., Gupta, A., Fei-Fei, L., Farhadi, A.: Target-driven Visual Navigation in Indoor Scenes using Deep Reinforcement Learning. In: ICLR (2017)