

Learning on the Go: A Meta-learning Object Navigation Model

Xiaorong Qin^{1,2}, Xinhang Song^{1,2}, Sixian Zhang^{1,2}, Xinyao Yu^{1,2}, Xinmiao Zhang², Shuqiang Jiang^{1,2,3}

¹Key Lab of Intelligent Information Processing Laboratory of the Chinese Academy of Sciences (CAS),

Institute of Computing Technology, Beijing, ²University of Chinese Academy of Sciences, Beijing

³Institute of Intelligent Computing Technology, Suzhou, CAS

{xiaorong.qin, xinhang.song, sixian.zhang, xinyao.yu}@vipl.ict.ac.cn, sqjiang@ict.ac.cn

Abstract

Object navigation tasks require an agent to locate a target object using visual observations in unseen environments, where unfamiliar layouts and novel object appearances can hinder navigation. Most existing methods lack the adaptability needed to handle these uncertainties, as their navigation models remain fixed during testing. In this paper, we address this challenge by examining object-conditioned trajectory distribution shifts in navigation caused by changes in environmental dynamics. We propose learning a central conditional distribution as a prior that approximates the specific distributions of diverse environments. To retain environment-specific information during navigation, we allow each environment-specific distribution to approximate this central distribution rather than relying on it directly. To implement this, we introduce a meta-learning mechanism that integrates with traditional navigation methods, offering tailored solutions for various types of navigation approaches. Our approach, Learning on the Go (LOG), enables agents to learn on the go, allowing for flexible, adaptive, real-time learning during navigation. Our theoretical analysis highlights the benefits of learning a central distribution for effective generalization across environments, and empirical results confirm the proposed method's effectiveness, demonstrating superior performance compared to existing approaches.

1. Introduction

The object navigation problem [11, 12, 61, 65] requires an agent to navigate towards an instance of an object category (e.g., a sink) using first-person visual observations. A fundamental challenge in navigation lies in the ability to generalize to unseen environments, which often introduce system uncertainties such as unfamiliar layouts and objects with novel appearances. These uncertainties make it difficult to navigate effectively to the target during test. However, most current navigation methods exhibit limited

flexibility and adaptability when faced with this challenge, as their navigation models tend to specialize in training environments and remain fixed at the test stage, resulting in a lack of degrees of freedom that can be relearned to adapt to changes under uncertainty [4, 24]. The limitation of poor adaptability is evident in both mainstream modular approaches [7, 27, 42, 62, 68, 73] and end-to-end techniques [12, 14, 22, 31, 38, 43, 44, 54, 65, 74] for navigation. Unlike these methods, human learners exhibit the ability to flexibly adapt to dynamic environments—a capability that is crucial for developing visually intelligent agents [31, 54]. Currently, several zero-shot approaches [6, 30, 64, 70, 71] based on the generalization capability of large language models (LLMs). However, these methods highly depend on pre-trained knowledge bases, have complex input conditions, and lack the flexibility of human adaptability. To better simulate humans' dynamic adaptability, some research in visual recognition and detection tasks has been investigated from the perspective of meta-learning [15, 20, 21]. However, there is still a lack of systematic research on navigation, which is a more complex and real-time task.

Towards better dynamic adaptability, we propose a plugable meta-learning mechanism that enhances the generalization capabilities of traditional navigation models by enabling efficient "learning on the go." Our approach focuses on analyzing object-conditioned trajectory distribution shifts caused by environmental changes, which is illustrated in Fig. 1. As the environments vary, the trajectories for locating the sink also appear different. To achieve effective generalization in new environments, a model must go beyond simply replicating training distributions and instead be capable of generalizing to related but distinct distributions. Ideally, prior learning provides a strong foundation for rapid adaptation, ensuring effective transfer to new environments. However, successful transfer of prior knowledge to unknown distributions requires certain assumptions to hold true [2, 3, 39]. In our scenario, the observation trajectories of agents across different environments may appear significantly distinct, yet they exhibit common patterns

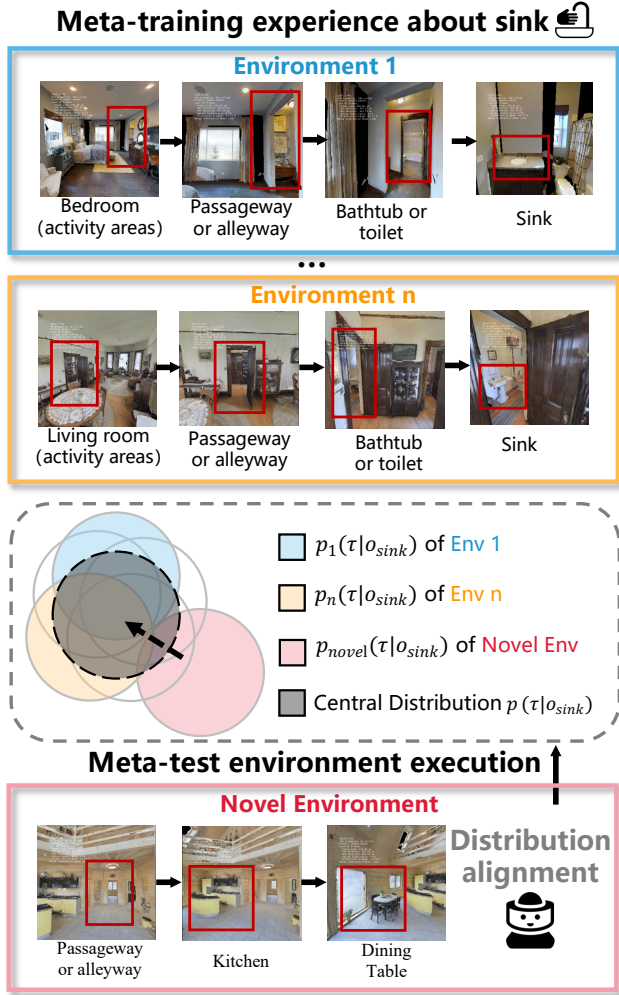


Figure 1. Learning a prior for trajectory distributions to generalize across environments via adaptation-time alignment.

in their overall navigation logic. By capturing these shared patterns, we aim to leverage the learned experience for effective adaptation to new environments. Consequently, we assume that the object-conditioned trajectory distributions across environments are sufficiently close to a central conditional distribution, which we learn through meta-learning.

How can this prior assumed be exploited and implemented? With the correct prior, the central distribution, and based on the assumption that the specific conditional distribution of a new environment is very close to the prior, the effective search space during the adaptation process is significantly reduced. This allows for rapid adaptation with only a few updates and a small amount of data. In terms of implementation, exact inference of the prior $p(\tau|o)$ is intractable, so first we redefine the meta-learning objective by employing a variational approximation method to infer it through maximizing a variational lower bound associated

with $p(o|\tau)$ and $p(\tau)$. Here $p(o|\tau)$ represents the target distribution conditioned on the trajectory variable, while $p(\tau)$ denotes the prior distribution of τ . At this time, the prior learned actually consists of two components: $p(o|\tau)$ represented by a decoder meta-learned across all environments, and $p(\tau)$ evaluated from trajectories sampled by an initial model that also requires to be meta-learned. During inner-layer optimization, we train an environment-specific model to approximate the environment-specific distribution to the central distribution, providing rewards or losses based on integrated navigation method types. We apply tailored optimization objectives for the outer-layer stage. Through this optimization mechanism, we ultimately learn a central conditional distribution that facilitates adaptation to unknown environments.

Our contributions are summarized as follows: (1) We enable agents to learn on the go, facilitating generalization by maintaining behavioral consistency, thus extending the agent’s capability beyond traditional navigation limits. (2) Theoretically, we build on the generalization framework of learning invariant representations [3] and analyze the benefits of learning a central target-conditioned latent behavior variable distribution. (3) Empirically, our method demonstrates superior performance in our experimental setup, validating its effectiveness.

2. Related works

Object-goal navigation. Previous works addressing the object-goal navigation problem can be divided into two main categories. The first category, end-to-end methods [10, 12, 14, 31–34, 38, 59, 60, 65, 67, 69, 74], employs a single unified network that takes the current observation and object goal as inputs and outputs the action to be performed, following either the reinforcement learning (RL) or imitation learning (IL) paradigm. The second category is modular approaches [7, 8, 27, 42, 62, 68, 73], which construct a top-down semantic map and perform target position prediction and navigation action execution sequentially. Both categories rely on pre-trained and fixed models to perform navigation during inference. However, we argue that this approach lacks adaptability in novel environments. Furthermore, it does not leverage the agent’s ability to collect environment-relevant, unsupervised data—such as trajectories comprising sequences of state-action pairs—through interaction with the environment during navigation. This interaction data supports learning while navigating, which can enhance adaptability to new environments. In the context of object-goal navigation, SAVN [54] addresses this adaptability challenge by meta-learning a self-supervised loss function that aligns with the navigation loss. This method encourages gradient updates that improve adaptation by utilizing trajectory data and the loss function with SGD during test. We share the perspective that “learning to learn”

is a critical skill for visually intelligent agents, particularly given the natural advantage of accessible interaction data compared to third-party data. However, the current design lacks a deeper understanding of the relationships between environments, object goals, and trajectories.

Meta Learning. Meta-learning [19, 29, 50], or “learning to learn,” is a machine learning approach focused on enabling models to adapt quickly to new tasks by generalizing across various experiences. Unlike traditional methods optimized for specific tasks, meta-learning trains models to apply knowledge from one task to another, even with limited data. This involves two main stages: meta-training, where generalized patterns are learned across tasks, and meta-test, where models adapt quickly to novel tasks. Key approaches in meta-learning include metric-based methods [25, 48, 49, 51], which focuses on learning a similarity metric between data points, often using neural networks such as Siamese networks or prototypical networks. It allows models to classify new instances by comparing them to known examples, effectively enabling “few-shot learning.”; model-based methods [18, 36], which enhance a model’s capacity for rapid learning through mechanisms like memory-augmented neural networks [35] or Meta Networks [47]. These models leverage an external memory component to store representations and use them for fast adaptation, making them suitable for sequential and reinforcement learning tasks; and optimization-based methods [37, 40, 45, 46, 56, 72], which modify the optimization process to enable quick adaptation with minimal updates. Model-Agnostic Meta-Learning (MAML) [15], for example, finds an optimal initialization of model parameters that facilitates rapid adaptation across tasks, while Reptile [37] uses gradient-based steps to achieve similar flexibility with lower computational cost. Meta-learning applications are vast, spanning AutoML for automating model selection, few-shot learning for image recognition [41], recommendation systems [52], transfer learning [28], embodied AI [17, 36, 54, 66]. Ongoing research aims to enhance these models’ efficiency and flexibility, making meta-learning a promising path for adaptive AI applications. Our method is based on an optimization-based meta-learning approach to learn a central conditional distribution for adapting to unknown environments.

3. Proposed method

3.1. Task and Notation Definition

In the object navigation problem, an agent is assigned the task of navigating to an object given its category label (*e.g.*, bed) within an unfamiliar environment. At the beginning of each episode, the agent is positioned at a random, navigable point in the environment. At each time step t , the

agent receives observation consisting of RGB-D sensor input s_t , odometer readings, and the goal category o . Based on the observation, the agent selects an action a_t from MOVE_FORWARD, TURN_LEFT, TURN_RIGHT, and STOP. Within the episode, the agent can collect historical trajectories in the form $\tau_t^o = (s_0, \dots, s_{t-1}, s_t)$ at each t . To successfully complete the task, the agent must navigate to within 1.0 meter of the target object and execute the STOP action. The episode ends either when the agent performs the STOP action or when the time budget T of 500 steps is exceeded.

3.2. Meta-problem for object navigation

As previously mentioned, frozen navigation policies during the test phase struggle to generalize to unknown environments. In this section, we explore learning adaptive policies from historical trajectories. To address this, we propose meta-learning a prior, specifically, the central trajectory distribution, to guide policy adaptation in new environments. To this end, we introduce LOG, a meta-learning framework that enables the agent to learn dynamically while navigating, thereby enhancing its adaptability to novel scenarios.

Meta-problem formulation of LOG. Given a target space $\mathcal{O}$ and an environment space $\mathcal{E}$, where $o \in \mathcal{O}$ represents a navigation target and τ denotes a trajectory collected based on o and $\mathcal{E}$, our goal is to learn a robust prior hypothesis, specifically, a central target-conditioned trajectory distribution $p_\theta(\tau|o)$ parameterized by meta-parameters θ . This prior serves as a foundation for a class of environment-specific target-conditioned trajectory distributions $p_{\theta_i}(\tau|o)$, parameterized by environment-specific parameters θ , facilitating efficient and adaptive learning in new environments. When encountering a new environment, the environment-specific model is refined by aligning its distribution $p_{\theta_i}(\tau|o)$ with the prior $p_\theta(\tau|o)$ as Eq. (1).

$$\min_{p_{\theta_i}(\tau|o)} D_{KL}(p_{\theta_i}(\tau|o) \parallel p_\theta(\tau|o)). \quad (1)$$

To learn the prior $p_\theta(\tau|o)$ for the distribution alignment described above, we consider a meta-navigation task distribution $\mathcal{T}$, where each sampled task $T_i \in \mathcal{T}$ corresponds to a navigation episode for a specific target and environment. Based on this, we formulate the problem as a meta-learning task

$$\begin{aligned} & \min_{p_{\theta_i}(\tau|o)} \mathbb{E}_{T_i \sim \mathcal{T}} [\mathcal{L}_{\text{nav}}(o, \tau)], \text{ where } \tau \sim p_{\theta_i}^*(\tau|o), \text{ and} \\ & p_{\theta_i}^*(\tau|o) = \arg \min_{p_{\theta_i}(\tau|o)} D_{KL}(p_{\theta_i}(\tau|o) \parallel p_\theta(\tau|o)), \end{aligned} \quad (2)$$

in which $\mathcal{L}_{\text{nav}}$ represents the empirical performance in the navigation task, with its precise definition varying across different navigation methods. A detailed explanation will be provided in the optimization solution section. Eq. (2)

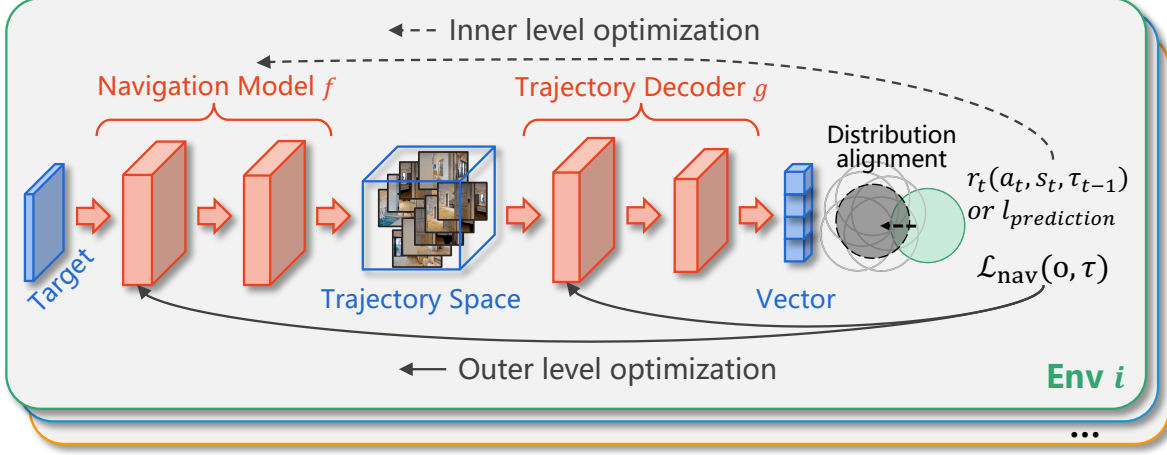


Figure 2. The overall framework of LOG: the inner-level optimization obtains the environment-specific parameters through Eq. (6) and Eq. (8), while the outer-level optimization learns the meta-parameters through Eq. (7) and Eq. (9).

encompasses two levels of learning. At the inner level (intra-task learning), it aims to find the current environment-specific conditional distribution $p_{\theta_i}(\tau|o)$ for a given task T_i , which is centered around the prior hypothesis $p_{\theta}(\tau|o)$. At the outer level (inter-task learning), the model leverages the performance of the biased optimal $p_{\theta_i}(\tau|o)$ to refine $p_{\theta}(\tau|o)$, ensuring that it maintains a small average distance to all true optimal $\{p_{\theta_i^*}(\tau|o)\}_i$. By sufficiently optimizing both the inner and outer objectives, the estimated prior, the estimated $p_{\theta}(\tau|o)$ is expected to be close to the optimal $p_{\theta_i}(\tau|o)$ for any task T_i sampled from $\mathcal{T}$ and thus can serve as a reliable prior hypothesis for a novel navigation task via Eq. (1).

Meta-Optimization of LOG. Exact inference of $p_{\theta}(\tau|o)$ is intractable. To simplify the problem, we first employ the Evidence Lower Bound [16, 23] to reformulate our distribution alignment objective Eq. (1) as follows (the proof is provided in appendix):

$$\max_{p_{\theta_i}(\tau|o)} \mathbb{E}_{p_{\theta_i}(\tau|o)} [\log p_{\theta_i}(o|\tau)] - D_{KL}(p_{\theta_i}(\tau|o) \parallel p_{\theta}(\tau)). \quad (3)$$

Notably, the prior hypothesis we aim to learn consists of the trajectory-conditioned distribution $p_{\theta}(o|\tau)$ and the prior distribution $p_{\theta}(\tau)$ of τ now. For practical implementation, we perform trajectory sampling from $p_{\theta_i}(\tau|o)$ using the following process: in the current environment of task T_i , the target object o is fed into a traditional navigation model f such as an end-to-end reinforcement learning (RL) approach or modular methods, to collect trajectories $\{\tau_t^o = (s_0, s_1, \dots, s_{t-1}, s_t)\}_t$. We model $p_{\theta}(o|\tau)$ using a probabilistic decoder, which is parametrized as a Gaussian centered around a deterministic encoding $g_{\theta}(\tau)$ with variance $\rho^2 I$, where g_{θ} takes τ as input and outputs a vector

representation of the target o . Additionally, we introduce a meta-initialization function f_{θ} and assume that the trajectories τ generated by f_{θ} follow the prior distribution $p_{\theta}(\tau)$.

At this stage, the first term of Eq. (3) approximately equals $-\log p_{\theta}(o|\tau)$ (as proved in the appendix). For the second term $D_{KL}(p_{\theta_i}(\tau|o) \parallel p_{\theta}(\tau))$, we approximate its implementation by initializing f_{θ_i} as f_{θ} and ensuring that f_{θ_i} remains close to f_{θ} by updating it through a few gradient steps. Then given N tasks sampled from $\mathcal{T}$, we derive our empirical meta-optimization objective as transformation of Eq. (2), formulated as follows.

$$\min_{f_{\theta}, g_{\theta}} \frac{1}{N} \sum_{m=1}^N \mathcal{L}_{\text{nav}}(o, \tau), \text{ where } \tau \sim f_{\theta_i}^*, \quad (4)$$

$$f_{\theta_i}^* = \arg \max_{f_{\theta_i}} \log p_{\theta}(o|\tau), \tau \sim f_{\theta_i},$$

where f_{θ} serves as the initialization of f_{θ_i} . To facilitate the practical optimization of f_{θ} during the agent's execution, we expand $\log p_{\theta}(o|\tau)$ using a telescoping series:

$$\log p_{\theta}(o|\tau) = \sum_{t=1}^T \log \frac{p_{\theta}(o|\tau_{t:t} = [s_t; \tau_{t-1}])}{p_{\theta}(o|\tau_{t-1})} \quad (5)$$

$$\propto \sum_{t=1}^T \|o - g_{\theta}(\tau_t)\|_2^2 - \|o - g_{\theta}(\tau_{t-1})\|_2^2,$$

where T is the number of time steps in the trajectory τ , and $\tau_{t-1} = (s_0, s_1, \dots, s_{t-1})$.

The specific implementation across different approaches. Next, we present the specific inner and outer level optimization solutions integrating meta-objective Eq. (4) with traditional navigation approaches, including both end-to-end RL methods and modular methods.

For **end-to-end RL methods**, such as DAT [12], f represents a RL model. At the inner level, we train f_{θ_i} of Eq. (4) using the reward function $r_t(a_t, s_t, \tau_{t-1})$ (see appendix) defined in Eq. (6) from the series Eq. (5). Intuitively, this reward reflects the idea that, given the cumulative trajectory up to the previous step, taking an action based on the current observation should make the cumulative trajectory at the current step more relevant to the target to be sought.

$$r_t(a_t, s_t, \tau_{t-1}) \propto \|o - g_\theta(\tau_{:t})\|_2^2 - \|o - g_\theta(\tau_{:t-1})\|_2^2 - \lambda_1, \quad (6)$$

where λ_1 is a tunable parameter.

For the outer-level optimization, the meta-model f_θ is updated using $\mathcal{L}_{\text{nav}}(o, \tau)$ as formulated in Eq. (7) (proved in the appendix), under the assumption of approximate optimality at the inner level. In this formulation, $l_{\text{a3c}}(o, \tau)$ represents the navigation loss, which aims to maximize the rewards provided by the simulator in the A3C algorithm, conditioned on $f_{\theta_i}^*$ learned in the inner stage. Additionally, we optimize g_θ using the environment-specific target-conditioned trajectory distribution $p_{\theta_i^*}(o|\tau)$, which is considered as another component of $\mathcal{L}_{\text{nav}}(o, \tau)$ as shown below:

$$\min_{f_\theta, g_\theta} \frac{1}{N} \sum_{i=1}^N \mathcal{L}_{\text{nav}}(o, \tau), \text{ where } \tau \sim f_{\theta_i}^*, \text{ and} \quad (7)$$

$$\mathcal{L}_{\text{nav}}(o, \tau) = l_{\text{a3c}}(o, \tau) + \mathbb{I}[\tau] \cdot \alpha_1 \|o - g_\theta(\tau)\|_2^2,$$

where α_1 is a tunable parameter, and $\mathbb{I}[\tau]$ is a 1/-1 indicator function, which takes a value of 1 if τ is a trajectory where the target is successfully found, and -1 otherwise.

For **modular methods**, such as PEANUT [62], f serves as a prediction network designed to estimate the locations of targets from incomplete semantic maps. In the inner stage, adaptively adjusting the prediction network based on Eq. (5) enables more precise long-term goals updates on the go. Every M times the network updates the long-term goal by outputting probability map z_i , we update the prediction network. The network is penalized according to Eq. (8) based on the historical trajectory since the last network update.

$$l_{\text{prediction}} = \lambda_2 \sum_{i=1}^M \sum_{j=1}^k C \|z_j - z_i\|_2^2, \quad (8)$$

where λ_2 is a tunable parameter, k is the step interval for long-term goal updates, $C = \|o - g_\theta(\tau_{:i*k})\|_2^2 - \|o - g_\theta(\tau_{:(i-1)*k})\|_2^2$, and z_k represents the prediction output based on the k -step historical semantic maps since the i -th long-term goal update after the previous network update, while z_i denotes the corresponding i -th prediction output. The loss penalty here encourages the network to explore alternative locations with similarly high probabilities.

At the outer level, we optimize the meta-model via Eq. (9). Since the prediction network of PEANUT independently learned the optimal solution through supervised training, we initialize it with pre-trained parameters following fine-tuning it, and apply meta-learning specifically to the decoder g_θ (details are provided in the appendix).

$$\max_{f_\theta, g_\theta} \frac{1}{N} \sum_{i=1}^N \mathcal{L}_{\text{nav}}(o, \tau), \text{ where } \tau \sim f_{\theta_i}^*, \text{ and}$$

$$\mathcal{L}_{\text{nav}}(o, \tau) = \mathbb{I}[\tau] \cdot [\alpha_2 \sum_{i=t}^T \|z_i - z_T\|_2^2 + \alpha_3 \|o - g_\theta(\tau)\|_2^2], \quad (9)$$

where α_2 and α_3 are tunable parameters, z_t is generated by the semantic maps after the inner-loop update of the network, and z_T represents the last long-term goal prediction map in each episode. Additionally, $\mathbb{I}[\tau]$ has the same meaning as in Equation Eq. (7).

Algorithm 1: The meta-training process on LOG

Input: Task distribution $\mathcal{T}$, the initialization of f_θ, g_θ , inner-level update step numbers J , trajectory step interval K , $i, j = 0$

Output: the meta-trained f_θ, g_θ

while $i < N$ **do**

 Sample N tasks randomly $\{T_i | i = 0, \dots, N-1\}$ from $\mathcal{T}$

 Let $f_{\theta_i} = f_\theta$

 // The inner-level optimization Eq. (3)

while $j < J$ **do**

 Roll out each initial model f_{θ_i} using the target o of T_i as input, generating trajectories

$\{\tau_t^o = (s_0, s_1, \dots, s_t) | t = 1, \dots, K\}$

if The "done" is true **then**

 | break

end

 Compute the inner-level reward or loss using Eq. (6) or Eq. (8) to update each θ_i

$j \leftarrow j + 1$

end

if The "done" is false **then**

 | Roll out each updated model $f_{\theta_i}^*$

end

 Update f_θ, g_θ through Eq. (7) and Eq. (9).

$i \leftarrow i + 1$

end

Overall, the overall framework of our method, LOG, is illustrated in Fig. 2, and the **meta-training** process is detailed in Algorithm 1. During **meta-test**, environment-specific models are learned by aligning via aligning $p_{\theta_i}(\tau|o)$ with the prior hypothesis $p_\theta(\tau|o)$, following a pro-

cess similar to the inner-level optimization in Algorithm 1.

3.3. Benefits of learning the central distribution across diverse environments

We borrow the idea of analyzing the generalization bounds for domain generalization problems of supervised learning, and analyze our method to make the generalization upper bound for unknown environments tighter.

For the same navigation target, trajectory data from different environments do not conform to the i.i.d. assumption because of the sequence correlation and diversity of environment dynamics, so we assume that there exists a central conditional distribution that is close to diverse environment-specific conditional distribution. We therefore learn f_θ and g_θ in the meta-learning paradigm to align distributions of data from different dynamics in the final view. So we proceed to illustrate the benefits or motivation of learning a identical conditional distribution for navigation tasks in an unknown environment by using the distribution-adaptive generalization theoretical framework on domain alignment proposed by [3]. Now we give the generalization bound for unknown environments.

Proposition 1 *Assume that Γ is a navigation trajectory space for target o and $\mathcal{H}$ is a class of hypotheses of (f_θ, g_θ) with respect to τ over this space. Let the set $\{\mathcal{S}_i\}_{i=1}^k$ be trajectory distributions over Γ from k known environments, and $\mathcal{T}$ be trajectory distribution over Γ from unknown environment. Let $\mathcal{O} = \cap_i \mathcal{B}_\rho(\mathcal{S}_i)$, where $\mathcal{B}_\rho(\mathcal{S}) = \{\mathcal{P} | d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S}, \mathcal{P})\}$ and $\rho = \max_{\mathcal{S}_u, \mathcal{S}_v} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S}_u, \mathcal{S}_v)$. For any $h \in \mathcal{H}$ (see appendix for proof):*

$$\begin{aligned} \mathcal{E}_{\mathcal{T}}(h) &\leq \lambda_\phi + \sum_i \alpha_i \mathcal{E}_{\mathcal{S}}(h) + \frac{1}{2} \min_{\mathcal{Q} \in \mathcal{O}} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{Q}, \mathcal{T}) \\ &\quad + \frac{1}{2} \min_{\mathcal{S}_i, \mathcal{S}_j} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S}_i, \mathcal{S}_j). \end{aligned} \quad (10)$$

We proceed to show our method reduces the last two items to some extent by learning a central target-conditioned trajectory distribution. The first term is a convex combination of ideal-joint errors between each training environment and the unknown environment. The term is small, and does not depend on particular h . The second term $\sum_i \alpha_i \mathcal{E}_{\mathcal{S}}(h)$ is a convex combination of generalization errors the training environments. Empirical risk minimization of Q_m in our method is used to control this term. In the third and final terms, $d_{\mathcal{H}\Delta\mathcal{H}}$ is a discrepancy distance between two distributions defined by [3, 9]:

$$\begin{aligned} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S}_u, \mathcal{S}_v) &= 2 \sup_{h_1, h_2 \in \mathcal{H}} |P_{\tau \sim \mathcal{S}_u}[h_1(\tau) \neq h_2(\tau)] - P_{\tau \sim \mathcal{S}_v}[h_1(\tau) \neq h_2(\tau)]| \\ &\leq 2 \sup_{h \in \mathcal{H}} |P_{\tau \sim \mathcal{S}_u}[h(\tau) = e] - P_{\tau \sim \mathcal{S}_v}[h(\tau) = e]|, \end{aligned} \quad (11)$$

where e is a shared embedding as a constant. Our meta-objective Eq. (2) can reduce the divergence of two training environments by aligning the distributions related to τ indirectly. The training of (f_θ, g_θ) changes the representation space to reduce the third term $\frac{1}{2} \min_{\mathcal{Q} \in \mathcal{O}} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{Q}, \mathcal{T})$ which demonstrates the divergences of the unknown environment and diverse training environments, and changes dynamically throughout the training process by $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S}_u, \mathcal{S}_v)$, and also changes the final term $\frac{1}{2} \min_{\mathcal{S}_i, \mathcal{S}_j} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S}_i, \mathcal{S}_j)$, finally leading to the better approximation of $\mathcal{E}_{\mathcal{T}}(h)$.

From this, we can conclude that this method indirectly aligns trajectory-related distributions using the meta-learning objective, reducing various error terms in the generalization bound by adjusting the representation space. Ultimately, it achieves a better approximation of the generalization error $\mathcal{E}_{\mathcal{T}}(h)$ in unknown environments, demonstrating that learning and aligning a central distribution can effectively improve the model's generalization ability across diverse environments.

4. Experiments

4.1. Datasets and baselines

The evaluation of our meta-learning mechanism integrated with end-to-end RL methods is executed on iTHOR [26] within the AI2-THOR platform. For modular methods, we evaluate the meta-learning mechanism on the HM3D [58], MP3D [5] and Gibson [55] datasets.

The previous methods DAT [12] and PEANUT [62] have each demonstrated superior performance on their respective simulators. We consider them as the methods we intend to integrate with our meta-learning mechanism.

4.2. Evaluation setup and metrics

Our experimental evaluation includes cross-environment generalization evaluation following the traditional evaluation setup of previous works [11, 12, 27, 65] for object navigation task, cross-simulator generalization evaluation and adaptivity evaluation for one specific environment, which we describe in detail in the appendix.

We use the following metrics for evaluating our models in the object navigation experiments. (1) **Succ.**: the success rate. (2) **SPL**: the success weighted by path length [1].

4.3. Empirical analysis of performance gains

Comparison on cross-environment evaluation: Tab. 1 displays the results of our method in unseen iTHOR environments. The last two rows highlight the performance of combining the baseline DAT approach with different meta-learning mechanisms. As mentioned previously in Sec. 2, SAVN is a meta-learning technique for object navigation, achieved through a self-supervised loss. Our method, LOG,



Figure 3. Example navigation episodes from HM3D (val) using PEANUT_INITIAL (Initial policy of LOG) and PEANUT+LOG (Adaptive policy of LOG). The agent’s selected long-term goal is marked by an orange dot on the semantic maps. In PEANUT+LOG, the prediction model is updated every 20 steps. Notably, at $t=40$, PEANUT+LOG selects an alternative high-probability location based on the preceding few-step trajectories.

Table 1. Cross-environment evaluation on iTHOR.

Method	ALL(%)		L ≥ 5(%)	
	Succ.↑	SPL↑	Succ.↑	SPL↑
HOZ [65]	68.53	37.50	60.27	36.61
SSCNAV [27]	77.14	35.09	71.73	34.33
PONI [42]	78.58	37.27	72.92	36.40
DAT [12]	82.02	46.54	75.67	45.37
DAT+SAVN[54]	79.67	45.66	74.32	46.12
DAT_INITIAL	78.54	44.12	73.62	45.90
DAT_SUPERVISED	79.32	45.17	74.60	46.27
DAT+LOG(OURS)	84.02	46.95	77.52	47.68

surpasses the SAVN-based mechanism by addressing environmental generalization from a more insightful trajectory-level perspective. Overall, our approach exhibits superior performance. Likewise, Tab. 2 presents our method’s results on the HM3D and MP3D datasets, which demonstrates that

PEANUT+LOG, achieves the highest performance across both datasets, indicating the effectiveness of LOG.

Table 2. Cross-environment evaluation on HM3D (test-standard) and MP3D (val).

Method	HM3D		MP3D	
	Succ. ↑	SPL ↑	Succ. ↑	SPL ↑
DD-PxPO [53]	26.0	12.0	8.0	1.8
PROCTHOR [13]	54.0	32.0	-	-
HABITAT-WEB [43]	55.0	22.0	35.4	10.2
OVRL [57]	60.0	27.0	28.6	7.4
3D-AWARE [63]	-	-	34.0	14.6
PEANUT [62]	62.6	32.5	39.4	15.0
SGM [68]	-	-	37.7	14.7
PEANUT_INITIAL	60.2	31.7	37.5	14.8
PEANUT_SUPERVISED	61.3	32.1	38.2	14.5
PEANUT+LOG(OURS)	64.3	34.2	40.3	15.4

Comparison on cross-simulator evaluation: As shown in Tab. 3, we evaluated our method in a cross-simulator

setting, using MP3D as the training simulator and Gibson as the test simulator. with PEANUT as the baseline method. The results demonstrate that incorporating our meta-learning mechanism, LOG, into the navigation model training significantly improves performance, even in environments with greater variation. This enhancement in generalization across diverse simulators further validates the effectiveness of our approach.

Table 3. Cross-simulator evaluation on MP3D and Gibson.

Method	Simulator		ALL(%)	
	Training	Testing	Succ.↑	SPL↑
PEANUT [62]	MP3D	Gibson	48.7	30.2
PEANUT+LOG(OURS)	MP3D	Gibson	63.9	37.1

Comparison on adaptivity evaluation for one specific environment: Tab. 4 shows the adaptivity evaluation of two models within one specific environment on the iTHOR platform. This experiment setup reflects a realistic scenario in which an agent is required to perform multiple navigation tasks within the same unfamiliar environment, assessing the adaptability of the model once it has been environment-specifically adapted. DAT+LOG consistently outperforms the baseline DAT method in an environment-specific setting. These findings validate the effectiveness of the proposed method in improving adaptability in a realistic, single-environment scenario.

Table 4. Adaptivity evaluation for one specific environment on iTHOR.

Method	ALL(%)		$L \geq 5$ (%)	
	Succ.↑	SPL↑	Succ.↑	SPL↑
DAT [12]	83.59	47.94	78.20	48.71
DAT+LOG(OURS)	86.43	49.72	82.46	50.60

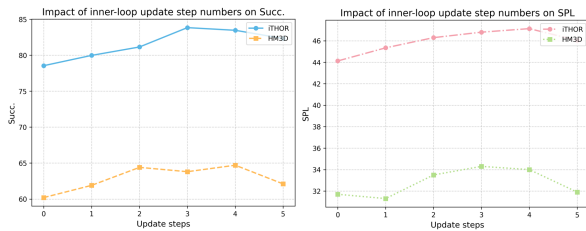


Figure 4. Ablation study on the number of inner-loop updates.

4.4. Ablations studies.

We analyze three major determinants of our model: (a) the performance difference before and after adaptation to the central distribution, (b) the effect of our meta-learning

mechanism, and (c) the impact of the number of gradient update steps in the inner level.

For (a), we introduce two baseline models: DAT_INITIAL and PEANUT_INITIAL, which represent the initialization of f_θ obtained through our meta-training process Algorithm 1. We compare the average performance of these initial models with their adapted counterparts, DAT+LOG and PEANUT+LOG, on cross-environment evaluation. The results, presented in Tab. 1 and Tab. 2, demonstrate that the adaptive models navigate more effectively in specific environments. For (b), we evaluate the impact of the meta-learning framework by constructing DAT_SUPERVISED and PEANUT_SUPERVISED, where f_θ and g_θ are trained via supervised learning, and then f_θ is finetuned to get f_{θ_i} using the same inner level optimization as our model LOG during meta-test. The corresponding results in Tab. 1 and Tab. 2 highlight the effectiveness of our meta-learning mechanism. For (c), we examine the influence of the number of gradient update steps in inner-level optimization, as shown in Fig. 4. The results indicate that within certain limits, increasing the number of gradient updates improves performance. However, excessive update steps lead to a decline in both accuracy and efficiency, emphasizing the importance of determining an optimal number of interaction gradient updates.

4.5. Qualitative results

As shown in the Fig. 3, we visualize the results of PEANUT_INITIAL and PEANUT+LOG, and observe that when searching for a bed, PEANUT+LOG has made adaptive adjustments upon encountering bathroom-related objects in the trajectory. This allows the agent to navigate to the bed more efficiently and accurately.

5. Conclusions

This paper addresses object navigation in unfamiliar environments by proposing LOG, a meta-learning framework that manages object-conditioned trajectory shifts via a learned central conditional distribution. LOG enhances generalization while preserving environment-specific details and integrates with traditional navigation methods for adaptive, real-time learning. Our test framework confirms LOG’s superior performance, supported by theoretical analysis highlighting the benefits of a central distribution for improved generalization in object navigation.

6. Acknowledgement

This work was supported by National Natural Science Foundation of China under Grant 62125207, 62272443, 62032022, U23B2012, and 62495084, in part by Beijing Natural Science Foundation under Grant JQ22012 and L242020.

References

- [1] Peter Anderson, Angel Chang, Devendra Singh Chaplot, Alexey Dosovitskiy, Saurabh Gupta, Vladlen Koltun, Jana Kosecka, Jitendra Malik, Roozbeh Mottaghi, Manolis Savva, et al. On evaluation of embodied navigation agents. *arXiv preprint arXiv:1807.06757*, 2018. 6
- [2] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. *Advances in neural information processing systems*, 19, 2006. 1
- [3] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79:151–175, 2010. 1, 2, 6
- [4] Yoshua Bengio, Tristan Deleu, Nasim Rahaman, Rosemary Ke, Sébastien Lachapelle, Olexa Bilaniuk, Anirudh Goyal, and Christopher Pal. A meta-transfer objective for learning to disentangle causal mechanisms. *arXiv preprint arXiv:1901.10912*, 2019. 1
- [5] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. Sparks of artificial general intelligence: Early experiments with gpt-4, 2023. 6
- [6] Yihan Cao, Jiazhao Zhang, Zhinan Yu, Shuzhen Liu, Zheng Qin, Qin Zou, Bo Du, and Kai Xu. Cognav: Cognitive process modeling for object goal navigation with llms. *arXiv preprint arXiv:2412.10439*, 2024. 1
- [7] Devendra Singh Chaplot, Dhiraj Prakashchand Gandhi, Abhinav Gupta, and Russ R Salakhutdinov. Object goal navigation using goal-oriented semantic exploration. *Advances in Neural Information Processing Systems*, 33:4247–4258, 2020. 1, 2
- [8] Devendra Singh Chaplot, Ruslan Salakhutdinov, Abhinav Gupta, and Saurabh Gupta. Neural topological slam for visual navigation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12875–12884, 2020. 2
- [9] Corinna Cortes and Mehryar Mohri. Domain adaptation in regression. In *International Conference on Algorithmic Learning Theory*, pages 308–323. Springer, 2011. 6
- [10] Ronghao Dang, Zhuofan Shi, Liuyi Wang, Zongtao He, Chengju Liu, and Qijun Chen. Unbiased directed object attention graph for object navigation. In *MM '22: The 30th ACM International Conference on Multimedia, Lisboa, Portugal, October 10 - 14, 2022*, pages 3617–3627. ACM, 2022. 2
- [11] Ronghao Dang, Zhuofan Shi, Liuyi Wang, Zongtao He, Chengju Liu, and Qijun Chen. Unbiased directed object attention graph for object navigation. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 3617–3627, 2022. 1, 6
- [12] Ronghao Dang, Liuyi Wang, Zongtao He, Shuai Su, Jiagui Tang, Chengju Liu, and Qijun Chen. Search for or navigate to? dual adaptive thinking for object navigation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8250–8259, 2023. 1, 2, 5, 6, 7, 8
- [13] Matt Deitke, Eli VanderBilt, Alvaro Herrasti, Luca Weihs, Kiana Ehsani, Jordi Salvador, Winson Han, Eric Kolve, Aniruddha Kembhavi, and Roozbeh Mottaghi. Proctor: Large-scale embodied ai using procedural generation. *Advances in Neural Information Processing Systems*, 35:5982–5994, 2022. 7
- [14] Heming Du, Lincheng Li, Zi Huang, and Xin Yu. Object-goal visual navigation via effective exploration of relations among historical navigation states. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2563–2573, 2023. 1, 2
- [15] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pages 1126–1135. PMLR, 2017. 1, 3
- [16] Charles W Fox and Stephen J Roberts. A tutorial on variational bayesian inference. *Artificial intelligence review*, 38: 85–95, 2012. 4
- [17] Kevin Frans, Jonathan Ho, Xi Chen, Pieter Abbeel, and John Schulman. Meta learning shared hierarchies. *arXiv preprint arXiv:1710.09767*, 2017. 3
- [18] James Harrison, Apoorva Sharma, Chelsea Finn, and Marco Pavone. Continuous meta-learning without tasks. *Advances in neural information processing systems*, 33:17571–17581, 2020. 3
- [19] Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(9):5149–5169, 2021. 3
- [20] Ting-I Hsieh, Yi-Chen Lo, Hwann-Tzong Chen, and Tyng-Luh Liu. One-shot object detection with co-attention and co-excitation. *Advances in neural information processing systems*, 32, 2019. 1
- [21] Bingyi Kang, Zhuang Liu, Xin Wang, Fisher Yu, Jiashi Feng, and Trevor Darrell. Few-shot object detection via feature reweighting. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8420–8429, 2019. 1
- [22] Apoorv Khandelwal, Luca Weihs, Roozbeh Mottaghi, and Aniruddha Kembhavi. Simple but effective: Clip embeddings for embodied ai. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14829–14838, 2022. 1
- [23] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014. 4
- [24] Robert Kirk, Amy Zhang, Edward Grefenstette, and Tim Rocktäschel. A survey of zero-shot generalisation in deep reinforcement learning. *Journal of Artificial Intelligence Research*, 76:201–264, 2023. 1
- [25] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, pages 1–30. Lille, 2015. 3
- [26] Eric Kolve, Roozbeh Mottaghi, Winson Han, Eli VanderBilt, Luca Weihs, Alvaro Herrasti, Matt Deitke, Kiana Ehsani, Daniel Gordon, Yuke Zhu, et al. Ai2-thor: An interactive 3d environment for visual ai. *arXiv preprint arXiv:1712.05474*, 2017. 6

- [27] Yiqing Liang, Boyuan Chen, and Shuran Song. Sscnav: Confidence-aware semantic scene completion for visual semantic navigation. In *2021 IEEE international conference on robotics and automation (ICRA)*, pages 13194–13200. IEEE, 2021. 1, 2, 6, 7
- [28] Q Sun Y Liu, TS Chua, and B Schiele. Meta-transfer learning for few-shot learning. In *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 3
- [29] Risheng Liu, Jiabin Gao, Jin Zhang, Deyu Meng, and Zhouchen Lin. Investigating bi-level optimization for learning and vision from a unified perspective: A survey and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):10045–10067, 2021. 3
- [30] Joel Loo, Zhanxin Wu, and David Hsu. Open scene graphs for open world object-goal navigation. *arXiv preprint arXiv:2407.02473*, 2024. 1
- [31] Oleksandr Maksymets, Vincent Cartillier, Aaron Gokaslan, Erik Wijmans, Wojciech Galuba, Stefan Lee, and Dhruv Batra. Thda: Treasure hunt data augmentation for semantic navigation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15374–15383, 2021. 1, 2
- [32] Bar Mayo, Tamir Hazan, and Ayellet Tal. Visual navigation with spatial attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16898–16907, 2021.
- [33] Mahdi Kazemi Moghaddam, Qi Wu, Ehsan Abbasnejad, and Javen Shi. Optimistic agent: Accurate graph-based value estimation for more successful visual navigation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 3733–3742, 2021.
- [34] Arsalan Mousavian, Alexander Toshev, Marek Fišer, Jana Košecká, Ayzaan Wahid, and James Davidson. Visual representations for semantic target driven navigation. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8846–8852. IEEE, 2019. 2
- [35] Tsendsuren Munkhdalai and Hong Yu. Meta networks. In *International conference on machine learning*, pages 2554–2563. PMLR, 2017. 3
- [36] Anusha Nagabandi, Ignasi Clavera, Simin Liu, Ronald S Fearing, Pieter Abbeel, Sergey Levine, and Chelsea Finn. Learning to adapt in dynamic, real-world environments through meta-reinforcement learning. *arXiv preprint arXiv:1803.11347*, 2018. 3
- [37] A Nichol. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018. 3
- [38] Anwesha Pal, Yiding Qiu, and Henrik Christensen. Learning hierarchical relationships for object-goal navigation. In *Conference on Robot Learning*, pages 517–528. PMLR, 2021. 1, 2
- [39] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009. 1
- [40] Eunbyung Park and Junier B Oliva. Meta-curvature. *Advances in neural information processing systems*, 32, 2019. 3
- [41] Xiaorong Qin, Xinhang Song, and Shuqiang Jiang. Bi-level meta-learning for few-shot domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15900–15910, 2023. 3
- [42] Santhosh Kumar Ramakrishnan, Devendra Singh Chaplot, Ziad Al-Halah, Jitendra Malik, and Kristen Grauman. Poni: Potential functions for objectgoal navigation with interaction-free learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18890–18900, 2022. 1, 2, 7
- [43] Ram Ramrakhya, Eric Undersander, Dhruv Batra, and Abhishek Das. Habitat-web: Learning embodied object-search strategies from human demonstrations at scale. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5173–5183, 2022. 1, 7
- [44] Ram Ramrakhya, Dhruv Batra, Erik Wijmans, and Abhishek Das. Pirlnav: Pretraining with imitation and rl finetuning for objectnav. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17896–17906, 2023. 1
- [45] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *International conference on learning representations*, 2017. 3
- [46] Andrei A Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell. Meta-learning with latent embedding optimization. *arXiv preprint arXiv:1807.05960*, 2018. 3
- [47] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. Meta-learning with memory-augmented neural networks. In *International conference on machine learning*, pages 1842–1850. PMLR, 2016. 3
- [48] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017. 3
- [49] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1199–1208, 2018. 3
- [50] Ricardo Vilalta and Youssef Drissi. A perspective view and survey of meta-learning. *Artificial intelligence review*, 18: 77–95, 2002. 3
- [51] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. *Advances in neural information processing systems*, 29, 2016. 3
- [52] Jianling Wang, Kaize Ding, and James Caverlee. Sequential recommendation for cold-start users with meta transitional learning. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1783–1787, 2021. 3
- [53] Erik Wijmans, Abhishek Kadian, Ari Morcos, Stefan Lee, Irfan Essa, Devi Parikh, Manolis Savva, and Dhruv Batra. Dd-ppo: Learning near-perfect pointgoal navigators from 2.5 billion frames. *arXiv preprint arXiv:1911.00357*, 2019. 7
- [54] Mitchell Wortsman, Kiana Ehsani, Mohammad Rastegari, Ali Farhadi, and Roozbeh Mottaghi. Learning to learn how to

- learn: Self-adaptive visual navigation using meta-learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6750–6759, 2019. 1, 2, 3, 7
- [55] Fei Xia, Amir R Zamir, Zhiyang He, Alexander Sax, Jitendra Malik, and Silvio Savarese. Gibson env: Real-world perception for embodied agents. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9068–9079, 2018. 6
- [56] Zhongwen Xu, Hado P van Hasselt, and David Silver. Meta-gradient reinforcement learning. *Advances in neural information processing systems*, 31, 2018. 3
- [57] Karmesh Yadav, Ram Ramrakhya, Arjun Majumdar, Vincent-Pierre Berges, Sachit Kuhar, Dhruv Batra, Alexei Baevski, and Oleksandr Maksymets. Offline visual representation learning for embodied navigation. In *Workshop on Reincarnating Reinforcement Learning at ICLR 2023*, 2023. 7
- [58] Karmesh Yadav, Ram Ramrakhya, Santhosh Kumar Ramakrishnan, Theo Gervet, John Turner, Aaron Gokaslan, Noah Maestre, Angel Xuan Chang, Dhruv Batra, Manolis Savva, et al. Habitat-matterport 3d semantics dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4927–4936, 2023. 6
- [59] Wei Yang, Xiaolong Wang, Ali Farhadi, Abhinav Gupta, and Roozbeh Mottaghi. Visual semantic navigation using scene priors. *arXiv preprint arXiv:1810.06543*, 2018. 2
- [60] Joel Ye, Dhruv Batra, Abhishek Das, and Erik Wijmans. Auxiliary tasks and exploration enable objectgoal navigation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 16117–16126, 2021. 2
- [61] Xinyao Yu, Sixian Zhang, Xinhang Song, Xiaorong Qin, and Shuqiang Jiang. Trajectory diffusion for objectgoal navigation. *Advances in Neural Information Processing Systems*, 37:110388–110411, 2024. 1
- [62] Albert J Zhai and Shenlong Wang. Peanut: Predicting and navigating to unseen targets. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10926–10935, 2023. 1, 2, 5, 6, 7, 8
- [63] Jiazhaoh Zhang, Liu Dai, Fanpeng Meng, Qingnan Fan, Xuelin Chen, Kai Xu, and He Wang. 3d-aware object goal navigation via simultaneous exploration and identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6672–6682, 2023. 7
- [64] Lingfeng Zhang, Qiang Zhang, Hao Wang, Erjia Xiao, Zixuan Jiang, Honglei Chen, and Renjing Xu. Trihelper: Zero-shot object navigation with dynamic assistance. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10035–10042. IEEE, 2024. 1
- [65] Sixian Zhang, Xinhang Song, Yubing Bai, Weijie Li, Yakui Chu, and Shuqiang Jiang. Hierarchical object-to-zone graph for object navigation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 15130–15140, 2021. 1, 2, 6, 7
- [66] Sixian Zhang, Weijie Li, Xinhang Song, Yubing Bai, and Shuqiang Jiang. Generative meta-adversarial network for unseen object navigation. In *European Conference on Computer Vision*, pages 301–320. Springer, 2022. 3
- [67] Sixian Zhang, Xinhang Song, Weijie Li, Yubing Bai, Xinyao Yu, and Shuqiang Jiang. Layout-based causal inference for object navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10792–10802, 2023. 2
- [68] Sixian Zhang, Xinyao Yu, Xinhang Song, Xiaohan Wang, and Shuqiang Jiang. Imagine before go: Self-supervised generative map for object goal navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16414–16425, 2024. 1, 2, 7
- [69] Sixian Zhang, Xinhang Song, Xinyao Yu, Yubing Bai, Xinlong Guo, Weijie Li, and Shuqiang Jiang. Hoz++: Versatile hierarchical object-to-zone graph for object navigation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 2
- [70] Gengze Zhou, Yicong Hong, Zun Wang, Xin Eric Wang, and Qi Wu. Navgpt-2: Unleashing navigational reasoning capability for large vision-language models. In *European Conference on Computer Vision*, pages 260–278. Springer, 2024. 1
- [71] Gengze Zhou, Yicong Hong, and Qi Wu. Navgpt: Explicit reasoning in vision-and-language navigation with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7641–7649, 2024. 1
- [72] Pan Zhou, Xiaotong Yuan, Huan Xu, Shuicheng Yan, and Jiashi Feng. Efficient meta learning via minibatch proximal update. *Advances in Neural Information Processing Systems*, 32, 2019. 3
- [73] Minzhao Zhu, Binglei Zhao, and Tao Kong. Navigating to objects in unseen environments by distance prediction. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10571–10578. IEEE, 2022. 1, 2
- [74] Yuke Zhu, Roozbeh Mottaghi, Eric Kolve, Joseph J Lim, Abhinav Gupta, Li Fei-Fei, and Ali Farhadi. Target-driven visual navigation in indoor scenes using deep reinforcement learning. In *2017 IEEE international conference on robotics and automation (ICRA)*, pages 3357–3364. IEEE, 2017. 1, 2