

Exploitation-Guided Exploration for Semantic Embodied Navigation

Justin Wasserman^{1,2} Girish Chowdhary¹ Abhinav Gupta² Unnat Jain^{2,3}

Abstract—In the recent progress in embodied navigation and sim-to-robot transfer, modular policies have emerged as a de facto framework. However, there is more to compositionality beyond the decomposition of the learning load into modular components. In this work, we investigate a principled way to syntactically combine these components. Particularly, we propose Exploitation-Guided Exploration (XGX) where separate modules for exploration and exploitation come together in a novel and intuitive manner. We configure the exploitation module to take over in the deterministic final steps of navigation *i.e.* when the goal becomes visible. Crucially, an exploitation module teacher-forces the exploration module and continues driving an overridden policy optimization. XGX, with effective decomposition and novel guidance, improves the state-of-the-art performance on the challenging object navigation task from 70% to 73%. Along with better accuracy, through targeted analysis, we show that XGX is also more efficient at goal-conditioned exploration. Finally, we show sim-to-real transfer to robot hardware and XGX performs over two-fold better than the best baseline from simulation benchmarking.
Project page: [gxvisnav.github.io](https://github.com/gxvisnav)

I. INTRODUCTION

The meaning of a whole is a function of the meanings of the parts and of **the way they are syntactically combined.**

Principle of Compositionality (Frege's Principle)

Consider the ‘planning’ problem you underwent when finalizing your last vacation. There is usually plenty of uncertainty in the decision-making – where to go, is the weather good, managing the budget, booking ground transport, *etc.* We tackle this uncertainty by breaking the problem down into sub-parts and then offload some of these sub-parts to specialists or expert websites. However, there is an ‘additional order’ that goes beyond offloading sub-parts to these reliable and modular solutions. The very fact that you know these modular solutions exist helps you make more ambitious plans, be more creative, and likely travel more often and better than without this information. Generally speaking, division of work or *modularity is only one aspect of a compositional approach.* The very awareness of our repertoire of skills or (associated modules) helps manage resources to plan uncertain aspects of planning better. This is true about compositional structures and software packages allowing us to think more creatively in tackling research as well. This intuition of our day-to-day intelligence guides our alternate take on visual robot navigation or *embodied navigation* which we discuss in this work.

While compositionality has been thoroughly investigated with classical [21] and neural lenses [5], [4] in natural

language processing, the utility of modular structure in learning-based navigation has only recently been caught attention, with the development of reproducible task definitions and platforms [39], [69], [7], [54], [6], [25], [59], [18]. Particularly, the decomposition of exploration and exploitation policies (frequently dubbed as global and local policies [10], [9]) is a common and intuitive one. Also, the exploitative final steps after exploration (also dubbed as last-mile navigation [76], [12], [67]) are shown to be better tackled by principled geometric vision or homography-based visuo-motor servoing (see [67] for a related study). As motivated above, we believe there is more to be tapped in compositionality beyond policy decomposition. To this end, we investigate the research question: “*Can modular navigation agents learn better exploration when guided by their exploitation abilities?*”

Given that we as a research community have made progress on these blocks, the investigation of this research question is rather intuitive and straightforward! For this study, we focus on the semantic visual navigation task of object-goal navigation [6] where the agent’s objective is to navigate to a goal category (akin to “find a chair”), which is a standardized [61]. Our policy *decomposition* employs a state-of-the-art neural policy [48] for the exploration module. For the exploitation module, we devise a simple-and-effective geometric policy for visuo-motor servoing the exploitation phase (steps after the goal is in view). Contrary to prior works, in XGX, the exploitation module provides *guidance* to the exploration module (implemented as teacher-forced [72] variation of the Proximal Policy Optimization). Note, prior works tackling vision-based navigation via modular policy *independently* optimize modules, with no guidance or interplay in loss objectives. Fig. 1 illustrates these two pillars of XGX— *decomposition*, and *guidance*.

We quantify the utility of XGX with benchmarking in simulation and sim-to-real transfer. Summary of our contributions: (1) We introduce XGX, a novel method for guiding exploration behavior through teacher-forcing of an exploitation module. (2) on the competitive object-goal navigation (HM3D) benchmark, across baselines from several learning paradigms and frontier-based methods, XGX leads to a 70% → 73% increase in success rate over the previous state-of-the-art ; (3) XGX needs lesser surveying while increasing success (captured via a new metric that balances task success with exploration efficiency), (4) rigorous real-robot experiments showing over two-fold improvement over best-performing baselines from simulation; (5) Detailed ablations, quantitative analysis of efficiency, and error modes.

¹ University of Illinois at Urbana-Champaign

² Carnegie Mellon University ³ FAIR at Meta

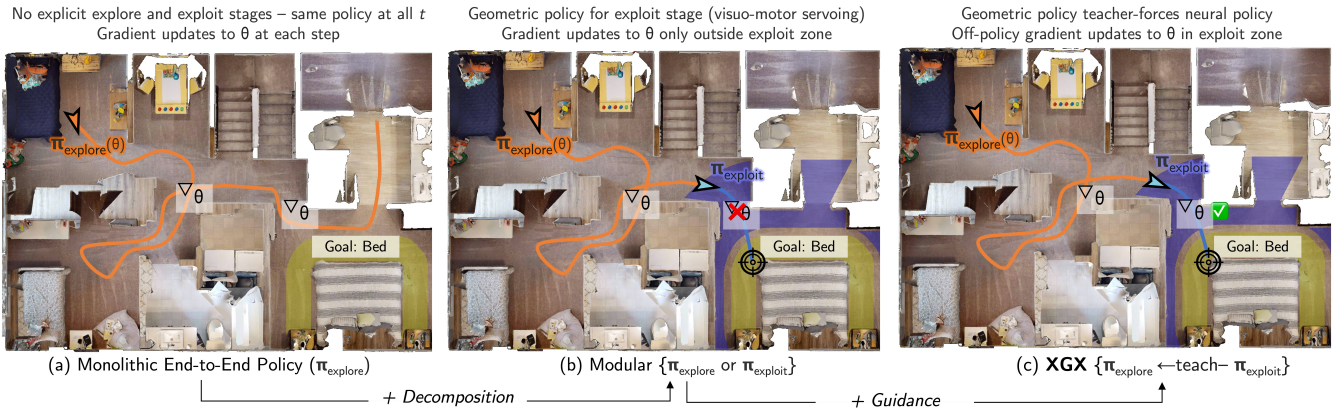


Fig. 1: **Overview of Exploitation-Guided Exploration (XGX)**. Within learning-based visual navigation (e.g. “navigate to category: *Bed*”), (a) most prior works adopt a neural policy, trained end-to-end. (b) Some works employ a modular policy, with a dedicated module to explore the environment and another for local navigation near the goal; we devise a simple and effective decomposition in XGX. (c) Unlike prior work, XGX enables the exploitation module to guide the exploration module via off-policy updates from teacher forcing. Both pillars of XGX (*policy decomposition* and *guidance*) are detailed in Sec. III.

II. RELATED WORK

Embodied Navigation. A popular paradigm for solving object-goal navigation has been to learn directly from simulation. This has been achieved through modular [25], [66] approaches, end-to-end [71] approaches, and by fine-tuning to the environment [48]. DDPPO [71] has been used as a baseline for many experiments for its success in solving point-goal navigation. They train a policy via a decentralized and distributed update rule, allowing them to train a policy in the high-fidelity simulator, AIHabitat [54], [58], [44], with over 2.5 billion frames. Habitat-Web [49] was able to achieve competitive performance on the 2022 Habitat Challenge by utilizing imitation learning over trajectories collected from a human. In contrast to directly using the simulator for fully training a policy from scratch, several methods have also attempted to use pretraining [75], [74], [35] to improve the navigation results. OVRl-v2 has achieved the current SOTA performance on the image-goal and object-goal by first pretraining a transformer head via masked autoencoding on HM3D [46] and Gibson [73]. Moving further away from simulation, a number of methods have been proposed to solve visual navigation tasks without any training in simulation. These zero-shot methods include frontier-based exploration [24] or heuristics [67] in conjunction with a module to solve the navigation task, as well as learning from real-world demonstrations [25], [8].

Modular Policies in Visual Navigation. Using modular policies is a popular paradigm for solving the visual navigation task. Previous methods such as NRNS [25] use a modular strategy to build a topological map and with a Graph Neural Network predicts where on the topological map the agent should explore to find the image-goal. They propose a neural exploitation policy to complete the task. In Chaplot *et al.* [9] the authors optimize for exploration by training a semantic-map conditioned policy to predict a long-term goal of where in the map the agent should explore. Once

their agent is near the goal, they utilize a deterministic local policy to arrive at the goal. In NTS [11] the authors follow a similar paradigm by training a topological map to explore the environment and then using a local policy to reach a relative waypoint from the agent. Finally, in PONI [45], the authors propose a potential function network that predicts “where to look” in the environment by predicting the most promising area along the boundaries of a map. The authors also use an analytical planner to predict an action to solve how to navigate to a given boundary. In all of these previous works, the authors keep the exploration and exploitation modules separated. In contrast to these previous works, XGX does not learn to explore separately from the exploitation module. Instead, we learn to explore the environment with feedback from our exploitation module included during training.

Goal-Conditioned Navigation on Physical Robots. Several recent works study robot navigation and locomotion, particularly in sim-to-real settings [64], [24], [1], [22]. This is an especially challenging task as real-world robots are susceptible to error modes not seen in simulation such as actuation and sensing noise, no access to ground truth data, and collecting trajectory examples is much slower [63], [13], [62]. Related works have studied image-goal and point-goal tasks and have shown transfer to a physical robot. ViKiNG [56], [57] was proposed to solve the image-goal task at a kilometer-scale by utilizing geographical hints through a top-down map. SLING [67] was able to improve over all previous baselines in simulation and real by utilizing a geometric-based solution for solving last-mile navigation. TGSM [37] utilizes a cross-graph mixer over a topological map that incorporates both image and object nodes to achieve SOTA performance on the image-goal task when a panoramic camera is utilized. However, in these tasks, an exact image or position of the goal is required. This does not allow for generalizing to our semantic visual navigation setting of navigating to a label such as “chair”.

III. EXPLOITATION-GUIDED EXPLORATION

As visualized in Fig. 1, the key components of the Exploitation-Guided Exploration methodology is its exploration module (π_{explore}), phase transition to exploitation module (π_{exploit}), and optimization of the exploration module via guidance from the exploitation module.

A. Exploration Module (π_{explore})

The neural architecture choice for the exploration module π_{explore} is orthogonal to our novelty of XGX (decomposition and guidance). To this end, we adopt the best-performing publicly available design – PIRLNav by Ramrakhya *et al.* [48]. The architecture includes a ResNet [27] *i.e.* convolutional blocks for visual encoding and gated-recurrent unit [17] for connecting observation across time. A multi-layer perceptron policy head on top outputs a categorical action distribution. For further details, kindly refer to [48]. Taking lessons from their extensive benchmarking, we too warm-start the neural exploration module by imitation learning over offline demonstrations (from [49]) and then fine-tune with reinforcement learning via XGX. The decomposed policy transitions from the exploration module to the exploitation module, if certain criterion is met, which we explain next.

B. Phase Transition

An ideal transition from exploration to exploitation phases would be characterized by handing off deterministic final stage of navigation to the exploitation module. One way to realize this is by training with an *updated* reward signal, $\mathbb{1}\{\text{Exploitation Module Succeeds}\}$ rather than direct success rewards akin to $\mathbb{1}\{\text{Within } 1m \text{ of the goal}\}$. However, calculating such an *updated* reward signal, at every step, is computationally intractable (requires a full trajectory rollout at each step). We implement a tractable approximation of this *updated* reward signal. Intuitively, we check if the agent can ‘see’ the goal and is close to it. If so, the phase transition from the exploration module to the exploitation module is triggered.

Specifically, if semantic segmentation (of the agent’s visual observation) contains the goal category and the agent is within a prescribed distance δ from the goal (as inferred from depth observation), the control transitions to the exploitation module. The set of states for which this transition will occur can be more mathematically described by the set:

$$\mathcal{S}_{\text{exploit}}^j \triangleq \{(s) \in \mathcal{S} : \text{dist}_j(s) \leq \delta, O_j \in f_{\text{semantic}}(s)\}, \quad (1)$$

where j is an index corresponding to the object-goal category labels $\{O_1, \dots, O_n\}$, and f_{semantic} an off-the-shelf semantic segmentation model. In this set, which we call the ‘exploitation states’, the agent employs simple and effective visuo-motor servoing *i.e.* our exploitation module, detailed next.

C. Exploitation Module (π_{exploit})

Once the goal is in sight, *i.e.*, within the exploitation state, visuo-motor servoing is both simple and effective. To this end, our exploitation module π_{exploit} can triangulate

the goal, navigate to it, and execute the ‘stop’ action. The exploitation module, utilizing the same off-the-shelf semantic segmentation model as phase transition, transforms the RGB to a semantic mask of the goal object. Next, it lifts the 2D semantic mask to 3D by utilizing the depth mask. From this depth mask, and with knowledge of the camera intrinsics, a waypoint is calculated. To navigate to this waypoint, direct planning works well. For this, we utilize a local metric map (free from the depth sensor) and employ a fast-marching method [10], [25] for collision avoidance. Next, we include XGX loss function and guidance-based policy updates.

D. Exploitation Module Teacher-Forcing (Guidance)

The high-level objective of a policy-gradient approach like XGX with PPO is to learn a policy towards maximizing expected, discounted, cumulative reward. The clipped surrogate loss objective from PPO [55] is as follows:

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathcal{D}} \left[\min(p_t(\theta) \hat{A}_t(\theta), \text{clip}(p_t(\theta), 1 - \epsilon, 1 + \epsilon) \hat{A}_t(\theta)) \right], \quad (2)$$

where $\hat{A}_t(\theta)$ is the advantage function, $p_t(\theta)$ is ratio function to correct for stale policy used for rollouts, and clip is performed to keep ratio function within ϵ dispersion of 1 (prevents too large of a gradient). The replay buffer of rollouts is denoted by \mathcal{D} consisting of (s_t, a_t, r_t) tuples for state, action, and environment rewards. The modification that allows π_{exploit} to guide π_{explore} is rooted in the ratio function. In standard PPO, the ratio function is $p_t(\theta) = \frac{\pi(a_t|s_t;\theta)}{\pi(a_t|s_t;\theta^-)}$, where $a_t \sim \pi(a_t|s_t;\theta^-)$ and θ^- is the stale parametrization of policy used to collect rollouts in \mathcal{D} and θ is the current policy’s parameters. In XGX, actions in the replay buffer (a_t) are instead sampled from a teacher-forced policy. Concretely:

$$a_t = \begin{cases} a_t^{\text{exploit}} \sim \pi_{\text{exploit}}(a|s) & \text{if } s \in \mathcal{S}_{\text{exploit}} \\ a_t^{\text{explore}} \sim \pi_{\text{explore}}(a|s;\theta^-) & \text{otherwise} \end{cases}, \quad (3)$$

where π_{explore} , $\mathcal{S}_{\text{exploit}}$, and π_{exploit} were introduced in Sec. III-A, Sec. III-B, and Sec. III-C, respectively. The value function relies on behavior policy as well. Since we employ the above teacher-forced policy, XGX value (and thereby advantage) estimates are also different. Notably, with π_{exploit} employed in Eq. (3), the time horizon of the RL formulation shortens significantly. This allows more effective learning from sparse success cues to optimize the parameters θ . Exploitation module π_{exploit} is specialized for getting to the goal, it increases the reward the agent receives overall during training. Another advantage of the decomposition in Eq. (3), also allows for back-tracking, *i.e.*, if the agent leaves ‘exploitation states’ $\mathcal{S}_{\text{exploit}}$, the agent returns to following the exploration module.

We choose to base XGX on PPO [55] for two practical reasons. First, it is the de facto algorithm for most RL formulations, with a stable convergence. Second, this choice allows us to build off the excellent support the AIHabitat photorealistic simulator has for distributed and decentralized PPO [71].

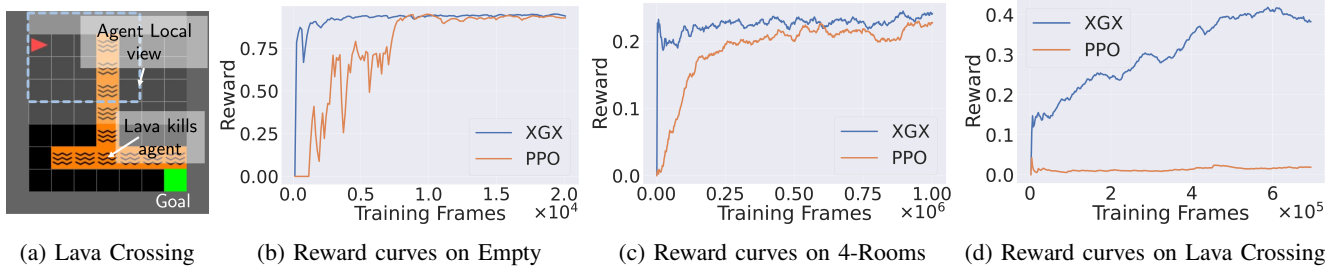


Fig. 2: (a) Visualization of one of the three navigation tasks XGX is tested on. (b,c,d) Across these tasks, XGX outperforms PPO and trains faster. As task complexity increases (b→c→d), the gains from XGX become increasingly prominent.

IV. EXPERIMENTS

In this section, we test our XGX methodology across (1) three 2D gridworld tasks, (2) large-scale photorealistic simulation across 1000 indoor environments, (3) physical robot runs across three diverse real-world scenes.

A. Diagnostic Task: Navigation in 2D MiniGrid

We investigate XGX in the fast MiniGrid environments [16], commonly used for proof of concept in prior works of visual navigation [68], [42], [30], [41]. Particularly, three tasks we benchmark on are: Empty, 4-Rooms, and Lava Crossing. These tasks require the navigation agent to explore the environment, avoid obstacles (particularly the detrimental lava) and stop at the goal represented as a green square. The standard reward structure used in these tasks is $1 - 0.9 * \frac{\text{step count}}{\text{max steps}}$ where max steps is the maximum number of steps the agent is allowed to take in a given episode. The agent observes a ‘egocentric’ (along the direction the agent is facing) patch of the 2D maze. An example Lava Crossing environment (mazes are procedurally generated and randomized) with helpful annotations is visualized in Fig. 2a. For XGX’s exploitation module π_{exploit} , the agent uses a local planner to create an action that will navigate it toward the goal (only when the goal is in the agent’s egocentric view). **Results.** As demonstrated in Fig. 2, XGX outperforms PPO across the three tasks. XGX demonstrates faster convergence over vanilla PPO. In the empty environment task, XGX achieves a reward of 0.9 after just 5% of training budget, while PPO needs 40%. In the Lava Crossing environment, XGX attains a reward of 0.41 while PPO is stuck at ~ 0.01 despite training it for over 500k steps. The immediate gain in XGX is attributed to π_{exploit} bootstrapping learning by reaching the goal immediately in some starting configurations.

B. Object-Goal Navigation Task

We experiment with Exploitation-Guided Exploration on the semantic visual navigation task of Object-Goal Navigation or ObjectNav [6], [3]. This choice is based on several factors – reproducible, photorealistic simulation, suite of published baselines, and large-scale suitable for learning-based robotics and sim-to-real transfer. Next, we include a brief task description, information on the associated sensors, and details of the standard metrics.

Task Definition. We adopt the protocol laid out for AIHabitat [6] which has been standardized as a public benchmark

| # | Method | Type | SR \uparrow | SPL \uparrow |
|----|---------------------------------|------------|---------------|----------------|
| 1 | DD-PPO [71] | RL | 27.9 | 14.2 |
| 2 | OVRL [75] | SSL→RL | 62.0 | 26.8 |
| 3 | OVRL-V2 [74] | SSL→RL | 64.7 | 28.1 |
| 4 | RRR [52] | Modular TL | 30.0 | 14.0 |
| 5 | Frontier Based Expl. [77], [24] | Classical | 26.0 | 15.2 |
| 6 | Habitat-Web [49] (paper) | IL | 57.6 | 23.8 |
| 7 | Habitat-Web [49] (our impl.) | IL | 64.1 | 27.1 |
| 8 | PIRLNav [48] (paper) | IL→RL | 61.9 | 27.9 |
| 9 | PIRLNav [48] (our impl.) | IL→RL | 70.4 | 34.1 |
| 10 | XGX (ours) | IL→RL | 72.9 | 35.7 |

RL: reinforcement, IL: imitation, SSL: self-supervised, TL: transfer

TABLE I: **Quantitative results for semantic navigation on ObjectNav.** Exploitation-Guided Exploration (XGX) significantly outperforms prior works based on IL, RL, and self-supervised representation learning. An equivalent jump in SPL is much harder than the same jump in success rate, indicating significantly more efficient planning by XGX. \uparrow denotes higher is better.

as well [61]. At the start of an ObjectNav episode, an embodied agent is initialized at a random location and is tasked to navigate to a goal category. At every time step the agent can choose an action from the action space $\mathcal{A} := \{\text{move forward, turn right, turn left, stop, look down, look up}\}$. The episode is considered a success if the agent can navigate within 1 meter of an instance of the goal category and execute the *stop* action. Consistent with prior work, we too adopt a sparse reward structure that returns a 1 on success and a 0 otherwise, at every time step. The goal definition, *i.e.*, the category of object to navigate to is sampled from a set of 6 categories [6] that are visually and physically well-defined. Consistent with several works on ObjectNav [6], [40], [75], [50], [2], we adopt the Habitat-Matterport3D (HM3D) scenes dataset. HM3D consists of high-quality photorealistic scans of 1000 real-world indoor scans. HM3D is an ideal choice for our study as it includes several scene types – homes, workspaces, eateries, and retail shops. Scenes span many geographical locations, floor areas, and multiple floored scenes are also included. This diversity allows for a better chance at generalizing to real-robot runs. For evaluation, we adopt the standard and publicly available 2000 episodes HM3D-val split.

Sensors. We adopt the standard sensor suite for AIHabitat ObjectNav [60], [61]. Particularly, the agent has access to three sensor observations: (1) egocentric RGB image of 640×480 , (2) corresponding depth mask, (3) relative

localization obtained from a ‘GPS+Compass sensor’ [54], [6]. For off-the-shelf $f_{\text{semantic}}(\cdot)$, we utilize a RedNet [34] trained on the HM3D-training split to predict egocentric semantic information in simulation. In the real-world, we utilize DETIC [78] to predict semantic segmentation.

Metrics. Following the literature in embodied navigation [54], [71], [32], [31], [14], [6], [53], we adopt metrics of success rate (SR) and success weighted by path length (SPL). SPL captures the policy’s efficiency in path planning. For targeted evaluation of exploration modules, we also report a metric from learning-based exploration [15], [10], [43], [11], [47], particularly, the percentage of the environment seen (% Cov). A point in the environment is considered ‘covered’ if it is within 3.2m of the agent and its field of view. Note, less coverage is better only if the success rate is the same. To this end, symmetric to the SPL metric introduced in [3], we devise a metric (% SCov) to balance task success with efficient exploration, defined as:

$$\frac{1}{N} \sum_{i=1}^N S_i \frac{\% \text{Cov}_i^{\text{oracle}}}{\max(\% \text{Cov}_i, \% \text{Cov}_i^{\text{oracle}})} \quad (4)$$

Where N is the number of total episodes, S_i specifies the success of episode i , and $\% \text{Cov}_i^{\text{oracle}}$ is the % Cov of a given episode when using an shortest-path policy.

C. Methods

We benchmark utilizing a diverse set of baselines spanning purely reinforcement learning, imitation learning, self-supervised representation learning, transfer learning, classical baselines, and combinations of the former:

- DDPO [71]: Purely RL baseline that employs proximal policy optimization [55] in distributed and decentralized manner. DDPO is a widely adopted deep RL baseline in prior works, across task definitions [76], [25], [2], [75], [35], and perfectly solved point-goal navigation task [54].
- OVRL [75]: Powered by self-supervised representation learning, OVRL pretrains a modified ResNet50 [27] head on the Omnidata Starter Dataset [20]. Using this initialization, Yadav *et al.* finetune the encoder and the policy on top using 500M frames of experience.
- OVRL-V2 [74]: Similar to OVRL, but the choice of the encoder is changed to vision transformers (ViT) [19] with a compression layer through a masked auto-encoding [26]. 500M steps in the simulator are used for finetuning both the ViT and the policy head.
- Habitat-Web [49]: Purely IL on $\sim 80k$ human-collected episodes using a simulation-web interface. Utilizes an inflection-weight [70] to avoid learning trivial policies.
- PIRLNav [48]: Going a step further from Habitat-Web baseline, this adopts a two-stage training regime – imitation learning for 500M steps followed by PPO updates for 300M frames. This is shown to improve performance in simulation.
- RRR [52], [51]: Modular policy with a zero-shot transfer learning from an HM3D point-goal module to object-goal.
- FBE [77], [24]: Frontier-Based exploration while incorporating semantics to find the goal; adopted by Gervet *et*

| # | Method | SR \uparrow | SPL \uparrow |
|---|--|---------------|----------------|
| 1 | Init. with IL on Human Demos + Fine-Tune [48] | 70.4 | 34.1 |
| 2 | 1 + Policy Decomposition | 71.5 | 34.3 |
| 3 | 1 + Policy Decomposition + Guidance (<i>i.e.</i> XGX) | 72.9 | 35.7 |
| 4 | XGX + GT Semantics (<i>upper bound</i>) | 73.5 | 39.8 |

TABLE II: **Head-on ablations for XGX.** We observe steady gains in performance by including the exploitation module π_{exploit} in our policy decomposition and adding *guidance* to neural exploration module coming from teacher-forced policy optimization.

al. [24]. We report results on full HM3D validation split (not just environments with a single floor [24]).

D. Results and Analysis (Photorealistic Simulation)

Metrics from empirical runs in AIHabitat are reported in Tab. I, Tab. II, and Tab. III. We discuss these results and analysis in the following text. As detailed in Sec. IV-B, all results are reported on the HM3D validation scenes (never seen during training).

XGX improves over all prior IL, RL, and SSL baselines (Tab. I). Compared with diverse baselines (Sec. IV-C), XGX (in row 10) demonstrates significant gains, improving 70.4 \rightarrow 72.9 in success rate (relative 4% \uparrow) over the previous best IL+RL method – PIRLNav [48] (in row 9). As we intuitively guessed, communicative exploration-exploitation via XGX has more efficient path planning, by reducing the ‘learning load’ on the exploration module. This helps improve the challenging metric of SPL by a relative 5% (34.1 \rightarrow 35.7). The best SSL method, OVRL-V2 [74] (in row 3), is fairly competitive in success rate (64.7 *vs.* 72.9) but significantly lags behind on SPL (28.1 *vs.* 35.7). Notably, XGX also outperforms classical frontier-based methods (row 5), improving from 26.0% to our 72.9% success rate.

Both the exploitation module and guidance are helpful (Tab. II). Two key aspects of our proposed approach are (1) exploitation module: based on basic visuo-motor servoing and geometric computer vision and (2) guidance from this module to exploration. Here we undertake head-on ablations to quantitatively evaluate the utility of both in Tab. II. We observe improvements of 70.4 \rightarrow 71.5 \rightarrow 72.9% (row 1 \rightarrow 2 \rightarrow 3) in success rate by successively adding the policy decomposition and guidance to the initialization. These trends are even more pronounced for the efficiency metric of SPL. This is intuitive as geometric vision-based servoing is almost perfect in path planning when fired correctly. We observed the same trend when adding XGX to other models like Habitat-Web [49] (seconding our results with PIRLNav [48] in Tab. II. If accurate semantics are available, XGX can scale with them to improve performance to 73.5% (row 4). This shows XGX can successfully tap into improvements in vision towards better navigation systems.

Efficiency Analysis – XGX can explore faster (Tab. III). Here we undertake a focused investigation of learned-based exploration *i.e.* quantifying exploration and not navigation success to the goal. We find that XGX explores an average of

| Method | % SCov \uparrow | Cov |
|--------------|-------------------|------|
| PIRLNav [48] | 36.3 | 63.5 |
| XGX (ours) | 39.1 | 59.8 |

TABLE III: **Analysis for exploration efficiency.** We find that XGX outperforms PIRLNav (row 9 and row 10, Tab. I) and it does so with more efficient, goal-conditioned exploration.

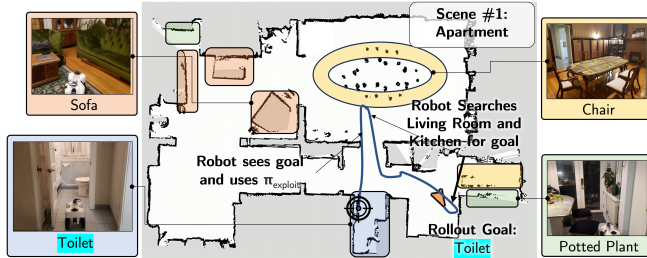


Fig. 3: **Physical robot navigation setup and rollout visualization in the Apartment environment.** The agent navigates to one of 4 goal categories. Here, we visualize an actual rollout for the ‘toilet’ category and mark the phase transition to the exploit module. Note: The agent has not seen the scene or accessed the top-down map (the map is for visualization).

59.8% per scene (compared to PIRLNav’s 63.5%), while still performing better on other navigation success metrics. This is because XGX’s exploration module was *a priori* optimized to offload servoing to the goal of the exploitation module. XGX taps into compositionality, making exploration significantly more efficient while also improving performance. This intuition is also empirically supported by XGX outperforming PIRLNav on % SCov at 36.3% to 39.1%.

Error Analysis. The failure modes of XGX are visualized in Fig. 4. The largest source of failure is missing annotations, where f_{semantic} of RedNet would correctly classify a goal object, only for the environment’s semantic mesh to not be correctly labeled for the given object. Major failure modes for the exploration module are due to not searching the environment well by either just looping over the same space (‘Exploration in Loops’) or not trying to go up/down stairs when it should (‘Missed Staircases’). Furthermore, the exploration module will occasionally call the ‘stop’ action immediately at the start of an episode (‘Stop Right Away’). Issues with the switching mechanism includes failing to correctly recognize the goal (‘Recognition Errors’) and failing to switch between the exploration and exploitation modules (‘Switching Error’). Other simulator errors include the floor geometry not being connected between spaces (‘Broken Floor Geometry’) and incomplete meshes where if the agent tries to go up/down stairs it can not (‘Stuck on Stairs’).

E. Results and Analysis (Physical Robot)

For comprehensive details regarding implementation, analysis, and results, please see our project page xgxvisnav.github.io. We sim-to-real transfer PIRLNav and our XGX on a wheeled navigation robot employed by navigation works [28], [23], [33], [67]. We navigate to five goal categories {couch, TV, chair, toilet, potted plant} across 14 tra-

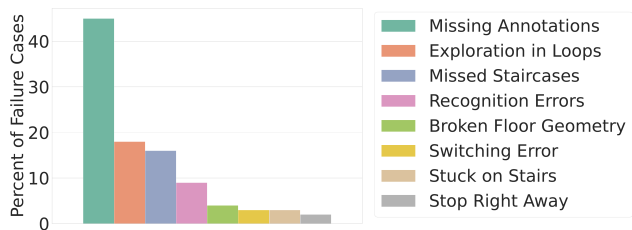


Fig. 4: **Error Analysis of XGX (simulation).** Beyond missing annotations in the dataset, the top three error modes of XGX are (1) π_{explore} manifesting a looping behavior, (2) missing staircases, (3) and errors in semantic segmentation.

jectories per method. Following success in robot navigation and locomotion research [65], [36], [29], we realize the low-level actuation using Model Predictive Control. We test our robotics setup in three diverse scenes (1) Apartment - as visualized in Fig. 3 (2) Office - environment with long hallways and many connecting rooms and (3) Food Court - containing many chairs and furniture that needs to be avoided. In order to acquire pose information, we employ localization using the robot’s LiDAR sensor and SLAM [38].

Accurate embodiment reduces the sim-to-real gap (Tab. IV, rows 1 & 2). The best-performing prior method, PIRLNav, does not demonstrate intuitive behaviors on the robot (see row 1, Tab. IV). This is despite us trying several lighting, goal categories, and distance to goal choices. We adapt PIRLNav (see row 2) by retraining it with a robot-specific configuration (height and field-of-view) as the default simulator configuration is different. This improved navigation behavior and a successful trajectory rollout. This is consistent with a prior study on end-to-end methods [24]. **XGX improves the performance of the robot as well (Tab. IV, row 3).** In Tab. IV, we find that XGX empirically has the best performance in physical robot experiments as well. Comparing rows 2 and 3, we observe an improvement from 8.4% \rightarrow 23.3% in SPL.

| Method | SR \uparrow | SPL \uparrow |
|-----------------------------|---------------|----------------|
| 1 PIRLNav (released ckpt) | 0.0 | 0.0 |
| 2 PIRLNav w/ our adaptation | 14.2 | 8.4 |
| 3 XGX (ours) | 35.7 | 23.3 |

TABLE IV: **Real-world object-goal navigation results.** In real-world robotics experiments, XGX performs the best.

V. CONCLUSION

In this work, we ask a fundamental question: “Can decomposed navigation agents learn better exploration when guided by their exploitation abilities?” To this end, we introduce XGX which utilizes teacher forcing from the exploitation to the exploration module, implemented via off-policy updates from teacher-forcing. We deploy XGX with standard choices of parametric exploration and simple geometric exploitation modules, leading to state-of-the-art performance in simulation, quantifiable efficiency gains in exploration, and an over 2-fold improvement in physical robotics experiments.

REFERENCES

- [1] Ananye Agarwal, Ashish Kumar, Jitendra Malik, and Deepak Pathak. Legged locomotion in challenging terrains using egocentric vision. In *CoRL*, 2022.
- [2] Ziad Al-Halah, Santhosh Kumar Ramakrishnan, and Kristen Grauman. Zero experience required: Plug & play modular transfer learning for semantic visual navigation. *CVPR*, 2022.
- [3] 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.
- [4] J. Andreas, M. Rohrbach, T. Darrell, and D. Klein. Deep compositional question answering with neural module networks. In *CVPR*, 2016.
- [5] Jacob Andreas, Marcus Rohrbach, Trevor Darrell, and Dan Klein. Neural module networks. In *CVPR*, 2016.
- [6] Dhruv Batra, Aaron Gokaslan, Aniruddha Kembhavi, Oleksandr Maksymets, Roozbeh Mottaghi, Manolis Savva, Alexander Toshev, and Erik Wijmans. Objectnav revisited: On evaluation of embodied agents navigating to objects. *arXiv preprint arXiv:2006.13171*, 2020.
- [7] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3d: Learning from rgb-d data in indoor environments. *3DV*, 2017.
- [8] Matthew Chang, Arjun Gupta, and Saurabh Gupta. Semantic visual navigation by watching youtube videos. In *NeurIPS*, 2020.
- [9] Devendra Singh Chaplot, Dhiraj Prakashchand Gandhi, Abhinav Gupta, and Russ R Salakhutdinov. Object goal navigation using goal-oriented semantic exploration. *NeurIPS*, 2020.
- [10] Devendra Singh Chaplot, Saurabh Gupta, Abhinav Gupta, and Ruslan Salakhutdinov. Learning to explore using active neural mapping. *ICLR*, 2020.
- [11] Devendra Singh Chaplot, Ruslan Salakhutdinov, Abhinav Gupta, and Saurabh Gupta. Neural topological slam for visual navigation. *CVPR*, 2020.
- [12] Prithvijit Chattopadhyay, Judy Hoffman, Roozbeh Mottaghi, and Aniruddha Kembhavi. Robustnav: Towards benchmarking robustness in embodied navigation. *ICCV*, 2021.
- [13] Yevgen Chebotar, Ankur Handa, Viktor Makoviychuk, Miles Macklin, Jan Issac, Nathan Ratliff, and Dieter Fox. Closing the sim-to-real loop: Adapting simulation randomization with real world experience. In *ICRA*, 2019.
- [14] Changan Chen, Unnat Jain, Carl Schissler, Sebastia Vicenc Amengual Gari, Ziad Al-Halah, Vamsi Krishna Ithapu, Philip Robinson, and Kristen Grauman. Soundspaces: Audio-visual navigation in 3d environments. *ECCV*, 2020.
- [15] Tao Chen, Saurabh Gupta, and Abhinav Gupta. Learning exploration policies for navigation. *ICLR*, 2019.
- [16] Maxime Chevalier-Boisvert, Bolun Dai, Mark Towers, Rodrigo de Lazcano, Lucas Willems, Salem Lahlou, Suman Pal, Pablo Samuel Castro, and Jordan Terry. MiniGrid & miniworld: Modular & customizable reinforcement learning environments for goal-oriented tasks. *CoRR*, abs/2306.13831, 2023.
- [17] K. Cho, A. C. Courville, and Y. Bengio. Describing multimedia content using attention-based encoder-decoder networks. *IEEE Transactions on Multimedia*, 2015.
- [18] Matt Deitke, Dhruv Batra, Yonatan Bisk, Tommaso Campari, Angel X Chang, Devendra Singh Chaplot, Changan Chen, Claudia Pérez D'Arpino, Kiana Ehsani, Ali Farhadi, et al. Retrospectives on the embodied ai workshop. *arXiv preprint arXiv:2210.06849*, 2022.
- [19] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [20] Ainaz Eftekhari, Alexander Sax, Jitendra Malik, and Amir Zamir. Omnidata: A scalable pipeline for making multi-task mid-level vision datasets from 3d scans. In *ICCV*, 2021.
- [21] Jerry A Fodor and Zenon W Pylyshyn. Connectionism and cognitive architecture: A critical analysis. *Cognition*, 1988.
- [22] Zipeng Fu, Ashish Kumar, Ananye Agarwal, Haozhi Qi, Jitendra Malik, and Deepak Pathak. Coupling vision and proprioception for navigation of legged robots. In *CVPR*, 2022.
- [23] Mateus V Gasparino, Arun N Sivakumar, Yixiao Liu, Andres EB Velasquez, Vitor AH Higuity, John Rogers, Huy Tran, and Girish Chowdhary. Wayfast: Navigation with predictive traversability in the field. *RAL*, 2022.
- [24] Theophile Gervet, Soumith Chintala, Dhruv Batra, Jitendra Malik, and Devendra Singh Chaplot. Navigating to objects in the real world. *arXiv preprint arXiv:2212.00922*, 2022.
- [25] Meera Hahn, Devendra Singh Chaplot, Shubham Tulsiani, Mustafa Mukadam, James M Rehg, and Abhinav Gupta. No rl, no simulation: Learning to navigate without navigating. *NeurIPS*, 2021.
- [26] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *CVPR*, 2022.
- [27] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CVPR*, 2016.
- [28] Vitor AH Higuity, Andres EB Velasquez, Daniel Varela Magalhaes, Marcelo Becker, and Girish Chowdhary. Under canopy light detection and ranging-based autonomous navigation. *JFR*, 2019.
- [29] Noriaki Hirose, Fei Xia, Roberto Martín-Martín, Amir Sadeghian, and Silvio Savarese. Deep visual mpc-policy learning for navigation. *RAL*, 2019.
- [30] Unnat Jain, Iou-Jen Liu, Svetlana Lazebnik, Aniruddha Kembhavi, Luca Weihs, and Alexander G Schwing. Gridtopix: Training embodied agents with minimal supervision. In *ICCV*, 2021.
- [31] Unnat Jain, Luca Weihs, Eric Kolve, Ali Farhadi, Svetlana Lazebnik, Aniruddha Kembhavi, and Alexander G. Schwing. A cordial sync: Going beyond marginal policies for multi-agent embodied tasks. In *ECCV*, 2020.
- [32] Unnat Jain, Luca Weihs, Eric Kolve, Mohammad Rastegari, Svetlana Lazebnik, Ali Farhadi, Alexander G. Schwing, and Aniruddha Kembhavi. Two body problem: Collaborative visual task completion. In *CVPR*, 2019.
- [33] Tianchen Ji, Arun Narenthiran Sivakumar, Girish Chowdhary, and Katherine Driggs-Campbell. Proactive anomaly detection for robot navigation with multi-sensor fusion. *IEEE Robotics and Automation Letters*, 2022.
- [34] Jindong Jiang, Lunan Zheng, Fei Luo, and Zhijun Zhang. Rednet: Residual encoder-decoder network for indoor rgb-d semantic segmentation. *arXiv preprint arXiv:1806.01054*, 2018.
- [35] Apoorv Khandelwal, Luca Weihs, Roozbeh Mottaghi, and Aniruddha Kembhavi. Simple but effective: Clip embeddings for embodied ai. In *CVPR*, 2022.
- [36] Donghyun Kim, Jared Di Carlo, Benjamin Katz, Gerardo Bledd, and Sangbae Kim. Highly dynamic quadruped locomotion via whole-body impulse control and model predictive control. *arXiv preprint arXiv:1909.06586*, 2019.
- [37] Nuri Kim, Obin Kwon, Hwiyeon Yoo, Yunho Choi, Jeongho Park, and Songhwai Oh. Topological semantic graph memory for image-goal navigation. In *CoRL*, 2023.
- [38] S. Kohlbrecher, J. Meyer, O. von Stryk, and U. Klingauf. A flexible and scalable slam system with full 3d motion estimation. In *Proc. IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR)*. IEEE, November 2011.
- [39] Eric Kolve, Roozbeh Mottaghi, Winson Han, Eli VanderBilt, Luca Weihs, Alvaro Herrasti, Daniel Gordon, Yuke Zhu, Abhinav Gupta, and Ali Farhadi. AI2-THOR: an interactive 3d environment for visual AI. *arXiv preprint arXiv:1712.05474*, 2019.
- [40] Arjun Majumdar, Gunjan Aggarwal, Bhavika Suresh Devnani, Judy Hoffman, and Dhruv Batra. Zson: Zero-shot object-goal navigation using multimodal goal embeddings. In *NeurIPS*, 2022.
- [41] Simone Parisi, Victoria Dean, Deepak Pathak, and Abhinav Gupta. Interesting object, curious agent: Learning task-agnostic exploration. *NeurIPS*, 2021.
- [42] Simone Parisi, Aravind Rajeswaran, Senthil Purushwalkam, and Abhinav Gupta. The unsurprising effectiveness of pre-trained vision models for control. In *ICML*, 2022.
- [43] Shivansh Patel, Saim Wani, Unnat Jain, Alexander Schwing, Svetlana Lazebnik, Manolis Savva, and Angel Chang. Interpretation of emergent communication in heterogeneous collaborative embodied agents. *ICCV*, 2021.
- [44] Xavier Puig, Eric Undersander, Andrew Szot, Mikael Dallaire Cote, Tsung-Yen Yang, Ruslan Partsey, Ruta Desai, Alexander William Clegg, Michal Hlavac, So Yeon Min, et al. Habitat 3.0: A co-habitat for humans, avatars and robots. *arXiv preprint arXiv:2310.13724*, 2023.
- [45] 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 *CVPR*, 2022.
- [46] Santhosh K Ramakrishnan, Aaron Gokaslan, Erik Wijmans, Oleksandr Maksymets, Alex Clegg, John Turner, Eric Undersander, Wojciech Galuba, Andrew Westbury, Angel X Chang, et al. Habitat-matterport 3d dataset (hm3d): 1000 large-scale 3d environments for embodied ai.

- arXiv preprint arXiv:2109.08238*, 2021.
- [47] Santhosh K Ramakrishnan, Dinesh Jayaraman, and Kristen Grauman. An exploration of embodied visual exploration. *arXiv preprint arXiv:2001.02192*, 2020.
- [48] Ram Ramrakhya, Dhruv Batra, Erik Wijmans, and Abhishek Das. Pirlnav: Pretraining with imitation and rl finetuning for objectnav. *arXiv preprint arXiv:2301.07302*, 2023.
- [49] Ram Ramrakhya, Eric Undersander, Dhruv Batra, and Abhishek Das. Habitat-web: Learning embodied object-search strategies from human demonstrations at scale. In *CVPR*, 2022.
- [50] Ram Ramrakhya, Eric Undersander, Dhruv Batra, and Abhishek Das. Habitat-web: Learning embodied object-search strategies from human demonstrations at scale. In *CVPR*, 2022.
- [51] Sonia Raychaudhuri, Tommaso Campari, Unnat Jain, Manolis Savva, and Angel X. Chang. Mopa: Modular object navigation with pointgoal agents, 2023.
- [52] Sonia Raychaudhuri, Tommaso Campari, Unnat Jain, Manolis Savva, and Angel X Chang. Reduce, reuse, recycle: Modular multi-object navigation. *arXiv preprint arXiv:2304.03696*, 2023.
- [53] Sonia Raychaudhuri, Saim Wani, Shivansh Patel, Unnat Jain, and Angel Chang. Language-aligned waypoint (law) supervision for vision-and-language navigation in continuous environments. In *EMNLP*, 2021.
- [54] Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, Devi Parikh, and Dhruv Batra. Habitat: A Platform for Embodied AI Research. *ICCV*, 2019.
- [55] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.
- [56] Dhruv Shah and Sergey Levine. ViKiNG: Vision-Based Kilometer-Scale Navigation with Geographic Hints. *RSS*, 2022.
- [57] Dhruv Shah, Ajay Sridhar, Nitish Dashora, Kyle Stachowicz, Kevin Black, Noriaki Hirose, and Sergey Levine. ViNT: A foundation model for visual navigation. In *CoRL*, 2023.
- [58] Andrew Szot, Alex Clegg, Eric Undersander, Erik Wijmans, Yili Zhao, John Turner, Noah Maestre, Mustafa Mukadam, Devendra Chaplot, Oleksandr Maksymets, et al. Habitat 2.0: Training home assistants to rearrange their habitat. *arXiv preprint arXiv:2106.14405*, 2021.
- [59] Andrew Szot, Unnat Jain, Dhruv Batra, Zsolt Kira, Ruta Desai, and Akshara Rai. Adaptive coordination in social embodied rearrangement. In *ICML*, 2023.
- [60] Habitat Team. Habitat CVPR challenge, 2020.
- [61] Habitat Team. Habitat CVPR challenge, 2021.
- [62] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *IROS*, 2017.
- [63] Joanne Truong, Sonia Chernova, and Dhruv Batra. Bi-directional domain adaptation for sim2real transfer of embodied navigation agents. *IEEE Robotics and Automation Letters*, 2021.
- [64] Joanne Truong, Max Rudolph, Naoki Harrison Yokoyama, Sonia Chernova, Dhruv Batra, and Akshara Rai. Rethinking sim2real: Lower fidelity simulation leads to higher sim2real transfer in navigation. In *CoRL*, 2023.
- [65] Joanne Truong, Denis Yarats, Tianyu Li, Franziska Meier, Sonia Chernova, Dhruv Batra, and Akshara Rai. Learning navigation skills for legged robots with learned robot embeddings. In *IROS*, 2021.
- [66] Saim Wani, Shivansh Patel, Unnat Jain, Angel Chang, and Manolis Savva. Multion: Benchmarking semantic map memory using multi-object navigation. *NeurIPS*, 2020.
- [67] Justin Wasserman, Karmesh Yadav, Girish Chowdhary, Abhinav Gupta, and Unnat Jain. Last-mile embodied visual navigation. In *CoRL*, 2022.
- [68] Luca Weihs, Unnat Jain, Iou-Jen Liu, Jordi Salvador, Svetlana Lazebnik, Aniruddha Kembhavi, and Alexander Schwing. Bridging the imitation gap by adaptive insubordination. In *NeurIPS*, 2021.
- [69] Luca Weihs, Jordi Salvador, Klemen Kotar, Unnat Jain, Kuo-Hao Zeng, Roozbeh Mottaghi, and Aniruddha Kembhavi. Allenact: A framework for embodied ai research. *arXiv preprint arXiv:2008.12760*, 2020.
- [70] Erik Wijmans, Samyak Datta, Oleksandr Maksymets, Abhishek Das, Georgia Gkioxari, Stefan Lee, Irfan Essa, Devi Parikh, and Dhruv Batra. Embodied question answering in photorealistic environments with point cloud perception. In *CVPR*, 2019.
- [71] 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. *ICLR*, 2019.
- [72] Ronald J Williams and David Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural computation*, 1989.
- [73] Fei Xia, Amir R Zamir, Zhiyang He, Alexander Sax, Jitendra Malik, and Silvio Savarese. Gibson env: Real-world perception for embodied agents. *CVPR*, 2018.
- [74] Karmesh Yadav, Arjun Majumdar, Ram Ramrakhya, Naoki Yokoyama, Alexei Baevski, Zsolt Kira, Oleksandr Maksymets, and Dhruv Batra. Ovrl-v2: A simple state-of-art baseline for imagenav and objectnav. *arXiv preprint arXiv:2303.07798*, 2023.
- [75] 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. *arXiv preprint arXiv:2204.13226*, 2022.
- [76] Joel Ye, Dhruv Batra, Abhishek Das, and Erik Wijmans. Auxiliary tasks and exploration enable objectgoal navigation. *ICCV*, 2021.
- [77] Sriram Yenamandra, Arun Ramachandran, Karmesh Yadav, Austin Wang, Mukul Khanna, Theophile Gervet, Tsung-Yen Yang, Vidhi Jain, Alexander William Clegg, John Turner, et al. Homerobot: Open-vocabulary mobile manipulation. *arXiv preprint arXiv:2306.11565*, 2023.
- [78] Xingyi Zhou, Rohit Girdhar, Armand Joulin, Philipp Krähenbühl, and Ishan Misra. Detecting twenty-thousand classes using image-level supervision. In *ECCV*, 2022.