2022 IEEE International Conference on Robotics and Automation (ICRA)
May 23-27, 2022. Philadelphia, PA, USA

Uncertainty-driven Planner for Exploration and Navigation

Georgios Georgakis[1], Bernadette Bucher[1], Anton Arapin[2], Karl Schmeckpeper[1],
Nikolai Matni[1], and Kostas Daniilidis[1]

***Abstract*— We consider the problems of exploration and point-goal navigation in previously unseen environments, where the spatial complexity of indoor scenes and partial observability constitute these tasks challenging. We argue that learning occupancy priors over indoor maps provides significant advantages towards addressing these problems. To this end, we present a novel planning framework that first learns to generate occupancy maps beyond the field-of-view of the agent, and second leverages the model uncertainty over the generated areas to formulate path selection policies for each task of interest. For point-goal navigation the policy chooses paths with an upper confidence bound policy for efficient and traversable paths, while for exploration the policy maximizes model uncertainty over candidate paths. We perform experiments in the visually realistic environments of Matterport3D using the Habitat simulator and demonstrate: 1) Improved results on exploration and map quality metrics over competitive methods, and 2) The effectiveness of our planning module when paired with the state-of-the-art DD-PPO method for the point-goal navigation task.**

I. Introduction

A major prerequisite towards true autonomy is the ability to navigate and explore novel environments. This problem is usually studied in the context of specific tasks such as reaching a specified point goal [1], finding a semantic target [2], or covering as much area as possible while building a map. Each of these tasks has its own idiosyncrasies, but all of them represent examples where one must often reason beyond what is currently observed and incorporate the uncertainty over the inferred information into the decision making process. For example, in point-goal navigation it is important to predict whether a certain path can lead to a dead-end. Likewise, in exploration strong confidence over a particular region's representation may prompt the agent to visit new areas of the map.

We investigate the tasks of point-goal navigation and exploration, and propose a planning module that leverages contextual occupancy priors. These priors are learned by a map predictor module that is trained to estimate occupancy values outside the field-of-view of the agent. Using the epistemic (model) uncertainty associated with these predictions we define objectives for path selection for each task of interest. Earlier work in this field focused mainly on learning how to actively control the agent for the purpose of reducing the uncertainty over the map [3] (Active SLAM), without considering navigation tasks in the process, while methods that did consider navigation often operated in relatively simple environments of artificially placed cylindrical obstacles [4], [5].

With the recent introduction of realistic and visually complex environments serving as navigation benchmarks [6], [7], the focus shifted on learning-based end-to-end approaches [8], [9], [10]. While end-to-end formulations that map pixels directly to actions are attractive in terms of their simplicity, they require very large quantities of training data. For instance, DD-PPO [10] needs 2.5 billion frames of experience to reach its state-of-the-art performance on Gibson [7]. On the other hand, modular approaches [11], [12], [13] are able to encode prior information into explicit map representations and are thus much more sample efficient. Our method falls into the latter category, but differs from other approaches by its use of the uncertainty over predictions outside the field-of-view of the agent during the planning stage. In contrast to [13], [12] this allows our method more flexibility when defining goal selection objectives, and does not require re-training between different tasks.

In this paper, we introduce *Uncertainty-driven Planner for Exploration and Navigation* (UPEN), in which we propose a planning algorithm that is informed by predictions over unobserved areas. Through this spatial prediction approach our model learns layout patterns that can guide a planner towards preferable paths in unknown environments. More specifically, we first train an ensemble of occupancy map predictor models by learning to hallucinate top-down occupancy regions from unobserved areas. Then, a Rapidly Exploring Random-Trees [14] (RRT) algorithm generates a set of candidate paths. We select paths from these candidates using epistemic (model) uncertainty associated with a path traversibility estimate as measured by the disagreement of ensemble models [15], [16], and we choose appropriate short-term goals based on the task of interest. Our contributions are as follows:

- We propose UPEN, a novel planning framework that leverages learned layout priors and formulates uncertainty-based objectives for path selection in exploration and navigation tasks.

Research was sponsored by the Army Research Office and was accomplished under Grant Number W911NF-20-1-0080. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. Further support was provided by the following grants: NSF IIS 1703319, NSF MRI 1626008, NSF TRIPODS 1934960, NSF CPS 2038873, ARL DCIST CRA W911NF-17-2-0181, ONR N00014-17-1-2093, the DARPA-SRC C-BRIC, CAREER award ECCS-2045834, and a Google Research Scholar award.

[1]GRASP Laboratory, Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104. ggeorgak@seas.upenn.edu

[2]Department of Computer Science, The University of Chicago, Chicago, IL, 60637. aarapin@uchicago.edu

2022 IEEE International Conference on Robotics and Automation (ICRA) | 978-1-7281-9681-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICRA46639.2022.9812423

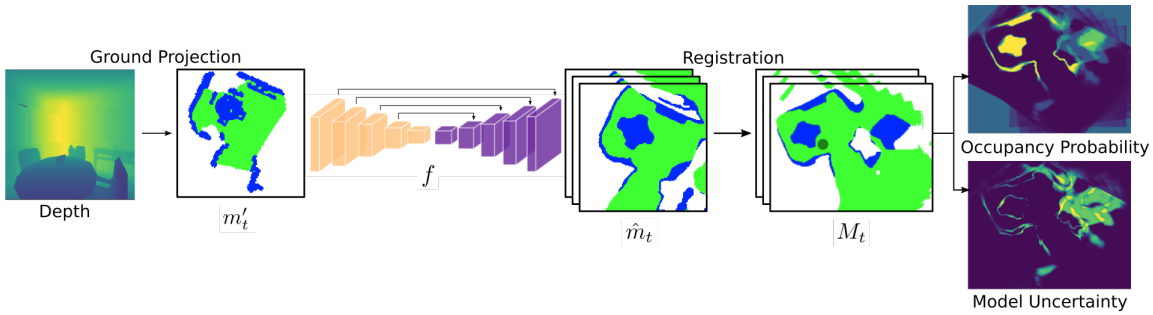

Fig. 1: Occupancy map prediction (blue-occupied, green-free) and uncertainty estimation for a time-step t. The egocentric depth observation is first ground-projected and passed through an ensemble f of encoder-decoder models that each infers information in unobserved areas ($\hat{m}_t$). Each $\hat{m}_t$ is then registered to a separate global map M_t. The final occupancy probabilities and model uncertainty are given by the mean and variance over the set of global maps.

- We show improved exploration results over competitive methods on the Matterport3D [17] dataset.
- We demonstrate the effectiveness of our planner when used to complement existing end-to-end methods on the point-goal navigation task.

II. Related Work

a) Navigation approaches: Traditional approaches to visual navigation focus on building a 3D metric map of the environment [18], [3] before using that representation for any downstream navigation tasks, which does not lend itself favourably for task-driven learnable representations that can capture contextual cues. The recent introduction of large-scale indoor environments and simulators [7], [17], [6] has fuelled a slew of learning based methods for indoor navigation tasks [1] such as point-goal [10], [19], [20], [21], [22], object-goal [23], [24], [25], [26], [27], and image-goal [8], [28], [29]. Modular approaches which incorporate explicit or learned map representations [11], [23], [25] have shown to outperform end-to-end methods on tasks such as object-goal, however, this is not currently the case for the point-goal [10], [20] task. In our work, we demonstrate how an uncertainty-driven planning module can favourably complement DD-PPO [10], a competitive method on point-goal navigation, and show increased performance in challenging episodes.

b) Exploration methods for navigation: A considerable amount of work was also devoted to planning efficient paths during map building, generally referred to as Active SLAM [30], [31], [32], [33], [34], [35]. For example, [32], [35] define information gain objectives based on the estimated uncertainty over the map in order to decide future actions, while [33] investigates different uncertainty measures. Recent methods focus on learning policies for efficient exploration either through coverage [9], [13], [36], [37] or map accuracy [12] reward functions. Furthermore, several works have gone beyond traditional mapping, and sought to predict maps for unseen regions [12], [38], [24], [27], [39] which further increased robustness in the decision making process. Our approach leverages the uncertainty over predicted occupancy maps for unobserved areas and shows its effectiveness on exploring a novel environment.

c) Uncertainty estimation: To navigate in partially observed maps, uncertainty has been estimated across nodes in a path [4], [40], via the marginal probability of landmarks [5], and with the variance of model predictions across predicted maps [24], [41]. Furthermore, uncertainty-aware mapping has been shown to be effective in unknown and highly risky environments [42], [43]. In our work, we use uncertainty differently for exploration and point goal navigation. In exploration, we estimate uncertainty over a predicted occupancy map via the variance between models in an ensemble. This variance across the ensemble specifically estimates model (epistemic) uncertainty [44], [45]. We select paths by maximizing epistemic uncertainty as a proxy for maximizing information gain following prior work in exploration [16], [24]. In point goal navigation, we compute traversability scores for candidate paths using an ensemble of map predictors and compute uncertainty with respect to these traversability scores using the variance over the scores given by each model in the ensemble. We use this uncertainty regarding path traversability to construct an upper confidence bound policy for path selection to balance exploration and exploitation in point goal navigation [46], [47], [48], [24].

III. Approach

We present an uncertainty-driven planning module for exploration and point-goal navigation tasks, which benefits from a learned occupancy map predictor module. Our approach takes as input the agent's egocentric depth observation and learns to predict regions of the occupancy map that are outside of the agent's field-of-view. Then it uses the uncertainty over those predictions to decide on a set of candidate paths generated by RRT. We define a separate policy to select a short-term goal along a path for each task of interest. In exploration we maximize uncertainty over the candidate paths, while for point-goal navigation we choose paths with an upper confidence bound policy for efficient and traversable paths. Finally, a local policy (DD-PPO [10]) predicts navigation actions to reach the short-term goal.

A. Occupancy Map Prediction

The first component in our planning module aims to capture layout priors in indoor environments. Such information

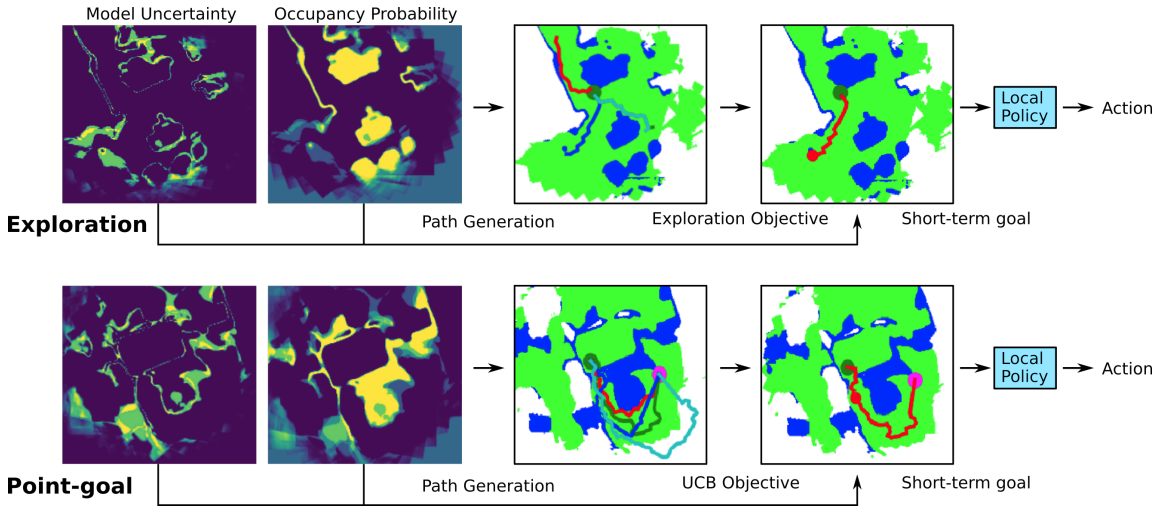

Fig. 2: Examples of path selections for exploration (top row) and point-goal navigation (bottom-row) tasks. Given the model uncertainty and occupancy probabilities we first generate a set of paths which are evaluated either with an exploration objective (section III-B) or an upper confidence bound objective (section III-C). The agent position is denoted as a dark green dot, the goal is shown as magenta, and red dots signify short-term goals.

can lead to a more intelligent decision making process for a downstream navigation task. Following the recent success of [12], [24] we formulate the occupancy map prediction as a semantic segmentation problem. Our model takes as input a depth image D_t at time-step t which is ground projected to an egocentric grid $m'_t \in \mathbb{R}^{|C|\times h\times w}$, where C is the set of classes containing $unknown$, $occupied$, and $free$, and h, w are the dimensions of the local grid. The ground projection is carried out by first using the camera intrinsic parameters to unproject D_t to a 3D point cloud and then map each 3D point to the $h \times w$ grid coordinates: $x' = \lfloor \frac{x}{r} \rfloor + \frac{w-1}{2}$, $z' = \lfloor \frac{z}{r} \rfloor + \frac{h-1}{2}$, where x', z' are the grid coordinates, x, z are the 3D points, and r is the grid cell size. Since the agent has a limited field of view, m'_t represents a local incomplete top-down occupancy grid of the area surrounding the agent. Our objective is to predict the missing values and produce the complete local occupancy map $\hat{m}_t \in \mathbb{R}^{|C|\times h\times w}$. To do so, we pass m'_t through an encoder-decoder UNet [49] model f that outputs a prediction for each grid location over the set of classes C. The model f is trained with a pixel-wise cross-entropy loss:

$$L = -\frac{1}{K}\sum_{k}^{K}\sum_{c}^{C} m_{k,c} \log \hat{m}_{k,c} \tag{1}$$

where $K = h \times w$ corresponds to the number of cells in the local grid and $m_{k,c}$ is the ground-truth label for pixel k. The ground-truth occupancy is generated by ground-projecting the available semantic information of the 3D scenes. To ensure diversity in the training examples, we sample training pairs across shortest paths between two randomly selected locations in a scene, where m'_t can contain a variable number of ground-projected depth images. Unlike [12] we do not use the RGB images during training, as we have found that the aforementioned sampling strategy is sufficient for the model to converge. This enables us to define a smaller and less memory intensive model f.

During a navigation episode, we maintain a global map $M_t \in \mathbb{R}^{|C|\times H\times W}$. Since f predicts a probability distribution over the classes for each grid location, we register $\hat{m}_t$ by updating M_t using Bayes Theorem. The global map M_t is initialized with a uniform prior probability distribution across all classes.

B. Exploration Policy

The main goal of exploration task is to maximize map coverage which requires navigating to new map regions around obstacles. To this end, we propose selecting paths using uncertainty of our map predictions as an objective in our planning algorithm. We are explicitly minimizing map uncertainty by collecting observations to improve the predicted global map M_t. Implicitly map coverage is maximized by minimizing map uncertainty because high coverage is required for predicting an accurate map with low uncertainty.

We use the epistemic (model) uncertainty as an objective for exploration [45], [44], [16], [24]. In order to estimate epistemic uncertainty, we construct f as an ensemble of N occupancy prediction models defined over the parameters $\{\theta_1, ..., \theta_N\}$. Variance between models in the ensemble comes from different random weight initializations in each network [16]. Our model estimates the true probability distribution $P(m_t|m'_t)$ by averaging over sampled model weights, $P(m_t|m'_t) \approx \mathbb{E}_{\theta\sim q(\theta)} f(m'_t;\theta) \approx \frac{1}{N}\sum_{i=1}^{N} f(m'_t;\theta_i)$ where the parameters θ are random variables sampled from the distribution $q(\theta)$[50], [51]. Then, following prior work [15], [16], [24], the epistemic uncertainty can be approximated from the variance between the outputs of the models in the ensemble, $\mathrm{Var}\, f(m'_t;\theta)$.

For path planning during exploration, our proposed objective can be used with any planner which generates a set S

of candidate paths. Each path $s \in S$ can be expressed as a subset of grid locations in our map. Each of these grid locations k has an associated uncertainty estimate given by the variance between model predictions in our ensemble. We specify this uncertainty map as $u_k := \text{Var}\, f(m'_t; \theta) \in \mathbb{R}^{1\times h\times w}$. We use this map to score each path s and construct the objective

$$\arg\max_{s\in S} \frac{1}{|s|} \sum_{k\in s} u_k \tag{2}$$

which selects the path with the maximum average epistemic uncertainty on the traversed grid.

In this work, we incorporate our uncertainty-based objective in RRT to plan to explore. We expand RRT for a set number of iterations, which generates candidate paths in random directions. We select between these paths using our objective from equation 2. In practice, equation 2 is evaluated over the accumulated global map M_t. Figure 1 shows the occupancy map prediction and the uncertainty estimation process using the ensemble f, while Figure 2 (top row) shows an example of path selection using the exploration objective.

C. Point-goal Policy

In the problem of point-goal navigation, the objective is to efficiently navigate past obstacles to a given goal location from a starting position. We again use RRT as a planner which generates a set of paths S between the agent's current location and the goal location. Thus, the primary objective when we select a path from these candidates to traverse is for the path to be unobstructed. Given a predicted occupancy map from model i in our ensemble and a candidate path $s \in S$ generated by our planner, we evaluate whether or not the path is obstructed by taking the maximum probability of occupancy in any grid cell k along each path. Specifically,

$$p_{i,s} = \max_{k\in s} \left(\hat{m}^i_{k,occ}|_{k\in s}\right) \tag{3}$$

where $\hat{m}^i_{k,occ}|_{k\in s}$ is the map of occupancy probabilities defined on the subset of grid cells $k \in s$ predicted by model i in the ensemble f. Choosing the path $s \in S$ by minimizing $p_{i,s}$ chooses the path we think most likely to be unobstructed. We can minimize this likelihood by selecting $\arg\min_{s\in S} \mu_s$ where $\mu_s := \frac{1}{N}\sum_{i=1}^{N} p_{i,s}$. However, we note that there may be multiple unobstructed candidate paths generated by our planner. We differentiate between these in our selection by adding a term d_s to our objective to incentivize selecting shorter paths. Furthermore, as an agent navigates to a goal, it makes map predictions using its accumulated observations along the way. Therefore, to improve navigation performance we can incorporate an exploration component in our navigation objective to incentivize choosing paths where it can gain the most information regarding efficient traversability.

We estimate uncertainty associated with efficient traversability of a particular path s for our exploration objective. Since there is zero uncertainty associated with path lengths d_s, we design our exploration objective to maximize information gain for path traversability. We denote $P_{s_{NT}}(m_t|m'_t)$ as the probability the path s is not traversable (NT) estimated by μ_s. We recall that μ_s is computed by averaging traversability scores over an ensemble of models. We compute the variance of these scores $\text{Var}_{i\in N}\, p_{i,s}$ to estimate uncertainty of our model approximating $P_{s_{NT}}(m_t|m'_t)$.

We combine exploration and exploitation in our full objective using an upper confidence bound policy [47], [46], [48], [24]. Our objective for efficient traversable paths is specified as

$$\arg\min_{s\in S} P_{s_{NT}}(m_t|m'_t) + d_s \tag{4}$$

and can be reconstructed as a maximization problem $\arg\max_{s\in S} -P_{s_{NT}}(m_t|m'_t) - d_s$. We denote $\sigma_s := \sqrt{\text{Var}_{i\in N}\, p_{i,s}}$ and observe the upper bound

$$-P_{s_{NT}}(m_t|m'_t) - d_s \leq -\mu_s + \alpha_1\sigma_s - d_s \tag{5}$$

holds with some fixed but unknown probability where α_1 is a constant hyperparameter. Using our upper bound to estimate $-P_{s_{NT}}(m_t|D_t)$, our full objective function as a minimization problem is

$$\arg\min_{s} \mu_s - \alpha_1\sigma_s + \alpha_2 d_s \tag{6}$$

where α_2 is a hyperparameter weighting the contribution of path length. Similarly to our exploration policy, in practice, equation 6 is evaluated over the accumulated global map M_t. Figure 2 (bottom row) illustrates path selection using our objective during a point-goal episode.

IV. Experiments

Our experiments are conducted on the Matterport3D (MP3D) [17] dataset using the Habitat [6] simulator. We follow the standard train/val/test environments split of MP3D which contains overall 90 reconstructions of realistic indoor scenes. The splits are disjoint, therefore all evaluations are conducted in novel scenes where the occupancy map predictor model has not seen during training. Our observation space consists of 256×256 depth images, while the action space contains four actions: `MOVE_FORWARD` by $25cm$, `TURN_LEFT` and `TURN_RIGHT` by $10°$ and `STOP`.

We perform two key experiments. First, we compare to other state-of-the-art methods on the task of exploration using both coverage and map accuracy metrics (sec. IV-B). Second we evaluate on the point-goal navigation task and demonstrate increased performance when DD-PPO [10] is complemented with our planning strategy (sec. IV-C).

A. Implementation Details

The Unet [49] model used for the occupancy map prediction has four encoder and four decoder convolutional blocks with skip connections and it is combined with a ResNet18 [53] for feature extraction. We use Pytorch [54] and train using the Adam optimizer with a learning rate of 0.0002. The grid dimensions are $h = w = 160$ for local, and $H = W = 768$ for global, while each cell in the grid is $5cm \times 5cm$. For the path generation process, we run the RRT every 30 navigation steps for exploration and

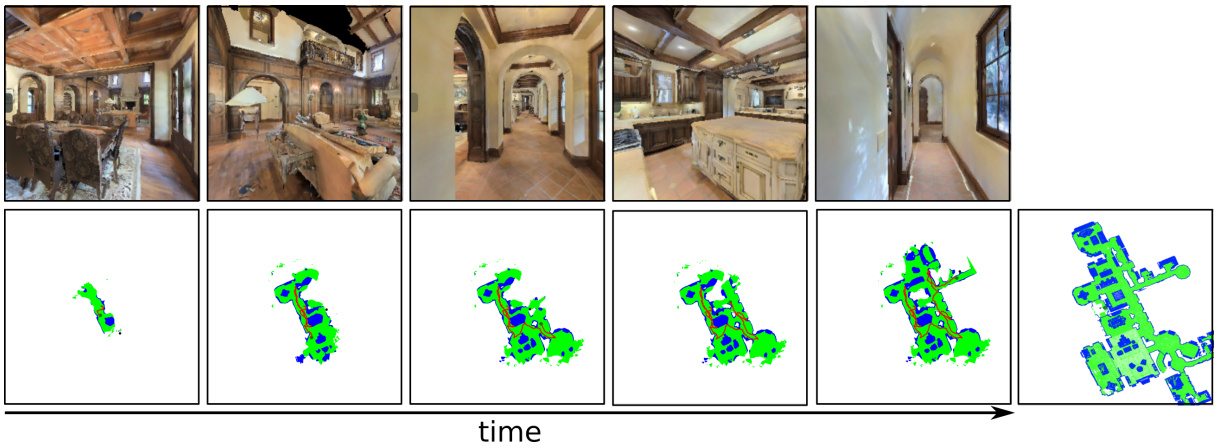

Fig. 3: Exploration example with T=1000 showing the trajectory followed by our agent (red line). The top row shows RGB images observed by the agent. The ground-truth map is visualized in the bottom right corner.

	Noisy		Noise-free	
Method	Map Acc (m^2)	IoU (%)	Map Acc (m^2)	IoU (%)
ANS(depth) [12]	72.5	26.0	85.9	34.0
OccAnt(depth) w/o AR [12]	92.7	29.0	104.7	38.0
OccAnt(depth) [12]	94.1	**33.0**	96.5	35.0
FBE [52] + DD-PPO [10]	100.9	28.7	120.2	44.7
UPEN + DD-PPO [10]	**110.3**	25.8	**141.6**	**45.6**

TABLE I: Exploration results on MP3D test scenes evaluating map quality at T=500. The "w/o AR" refers to the baseline that is trained without the anticipation reward in [12].

	Cov (m^2)	Cov (%)
ANS(rgb) [13]	73.28	52.1
FBE [52] + DD-PPO [10]	85.3	53.0
UPEN + DD-PPO [10]	**113.0**	**67.9**

TABLE II: Exploration results on MP3D test scenes evaluating area coverage at T=1000.

20 for point-goal. The RRT is set to generate a maximum of 10 paths every run, with a goal sampling rate of 20%. Finally, the RRT expands new nodes with a distance of 5 pixels at a time. A single step in a navigation episode requires 0.37s on average that includes map prediction and registration, planning using RRT, and DD-PPO. The timing was performed on a laptop using i7 CPU @ 2.20GHz and a GTX1060 GPU. All experiments are with ensemble size of 4. We provide code and trained models: `https://github.com/ggeorgak11/UPEN`.

B. Exploration

The setup from [12] is followed for this experiment, where the objective is to cover as much area as possible given a limited time budget $T = 1000$. Unless stated otherwise, the evaluation is conducted with simulated noise following the noise models from [13], [12]. We use the following metrics: 1) *Map Accuracy (m^2)*: as defined in [12] the area in the predicted occupancy map that matches the ground-truth map. 2) *IoU (%)*: the intersection over union of the predicted map and the ground-truth. 3) *Cov (m^2)*: the actual area covered by the agent. 4) *Cov (%)*: ratio of covered area to max scene coverage. We note that the two coverage metrics are computed on a map containing only ground-projections of depth observations. Our method is validated against the competitive approaches of Occupancy Anticipation [12] (*OccAnt*) and Active Neural SLAM [13] (*ANS*), which are modular approaches with mapper components. Both use reinforcement learning to train goal selection policies optimized over map accuracy and coverage respectively. Furthermore, we compare against the classical method of Frontier-based Exploration [52] (*FBE*). Since both UPEN and FBE are combined with DD-PPO and use the same predicted maps, this comparison directly validates our exploration objective.

We report two key results. First, in Table I our method outperforms all baselines in the noise-free case in both *Map Accuracy* and *IoU*. In fact, we show $21.4m^2$ and $36.9m^2$ improvement over *FBE* and *OccAnt* respectively on the *Map Accuracy* metric. In the noisy case even though we still surpass all baselines on *Map Accuracy*, our performance drops significantly in both metrics. In addition, the *Map Accuracy* increasing while *IoU* drops is attributed to increased map coverage with reduced accuracy. This is not surprising since unlike *OccAnt* and *Neural SLAM* we are not using a pose estimator. Second, in Table II we demonstrate superior performance on coverage metrics with a margin of $27.7m^2$ from *FBE* and $39.7m^2$ from *ANS*. This suggests that our method is more efficient when exploring a novel scene, thus validating our uncertainty-based exploration policy. Figure 3 shows an exploration episode.

C. Point-goal Navigation

We evaluate the performance of our uncertainty-driven planner when used to augment DD-PPO [10] against its vanilla version. DD-PPO is currently one of the best performing methods on point-goal navigation, achieving 97% SPL on the Gibson [7] validation set as shown in [10]. We follow the point-goal task setup from [1] where given a target coordinate the agent needs to navigate to that target and stop within a $0.2m$ radius. The agent is given a time-budget of

Dataset	MP3D Val		MP3D Test		MP3D Val-Hard	
Method	Success (%)	SPL (%)	Success (%)	SPL (%)	Success (%)	SPL (%)
DD-PPO [10]	47.8	**38.7**	37.3	30.2	38.0	28.1
UPEN-Occ + DD-PPO [10]	43.8	30.2	36.3	25.3	42.3	26.9
UPEN-Greedy + DD-PPO [10]	48.9	36.0	37.5	28.1	43.0	28.8
UPEN + DD-PPO [10]	**49.8**	36.9	**40.8**	**30.7**	**45.7**	**31.6**

TABLE III: Point-goal navigation results of our method against the vanilla DD-PPO[10]. "Occ" signifies a policy that uses only occupancy predictions, while "Greedy" refers to a policy taking into consideration path length without uncertainty.

	Avg GD (m)	Avg GEDR	Min GEDR
Gibson Val	5.88	1.37	1.00
MP3D Val	11.14	1.40	1.00
MP3D Test	13.23	1.42	1.00
MP3D Val-Hard	8.28	3.19	2.50

TABLE IV: Geodesic distance (GD) and geodesic to Euclidean distance ratio (GEDR) between different evaluation sets for point-goal navigation.

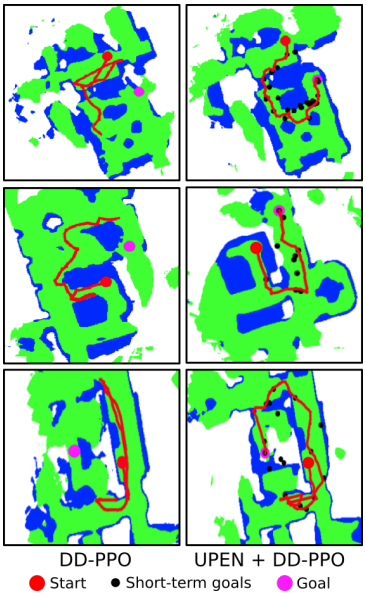

Fig. 4: Point-goal navigation examples from the MP3D Val-Hard set where the vanilla DD-PPO [10] fails to reach the target while our method is successful.

$T = 500$ steps to reach the target. For evaluation we use the standard metrics [1]: *Success:* percentage of successful episodes, and *SPL:* success rate normalized by path length. For this experiment we assume noise-free poses are provided by the simulator. To combine DD-PPO with our planner, we set the current short-term goal estimated by our approach as the target that DD-PPO needs to reach. For the vanilla DD-PPO we use the final target location in each test episode.

DD-PPO essentially solves Gibson point-goal navigation task so we turn our attention to MP3D where DD-PPO has lower performance due to the episodes having larger average geodesic distance (GD) to goal. However, we noticed that the average geodesic to euclidean distance ratio (GEDR) in MP3D is still low (a GEDR of 1 means there is a straight line path between the starting position and the goal). In order to demonstrate the effectiveness of our proposed method, we generated a new evaluation set (*MP3D Val-Hard*) with minimum GEDR=2.5. This created episodes which frequently involve sharp u-turns and multiple obstacles along the shortest path. Table IV illustrates episode statistics between different evaluation sets[1]. In addition to *MP3D Val-Hard*, we also test our method on the publicly available sets of *MP3D Val* and *MP3D Test*. We note that *MP3D Val-Hard* was generated using the same random procedure as its publicly available counterparts.

We define two variations of our method in order to demonstrate the usefulness of our uncertainty estimation by choosing different values for the α_1 and α_2 parameters of Eq. 6 from section III-C. First, *UPEN-Occ + DD-PPO* ($\alpha_1 = 0, \alpha_2 = 0$) considers only the occupancy probabilities when estimating the traversability difficulty of a path, while *UPEN-Greedy + DD-PPO* ($\alpha_1 = 0, \alpha = 0.5$) considers the path length and not the uncertainty. Our default method *UPEN + DD-PPO* uses $\alpha_1 = 0.1$ and $\alpha_2 = 0.5$.

The results are illustrated in Table III. We outperform all baselines in all evaluation sets with regards to *Success*. The largest gap in performance is observed in the *MP3D Val-Hard* set which contains episodes with much higher average GEDR that the other sets. This suggests that our method is able to follow more complicated paths by choosing short-term goals, in contrast to the vanilla DD-PPO which has to negotiate narrow passages and sharp turns only from egocentric observations. Regarding *SPL*, our performance gains are not as pronounced as in *Success*, since our policy frequently prefers paths with lower traversability difficulty in favor of shortest paths, to ensure higher success probability.

V. Conclusion

We introduced a novel uncertainty-driven planner for exploration and navigation tasks in previously unseen environments. The planner leverages an occupancy map predictor that hallucinates map regions outside the field of view of the agent and uses its predictions to formulate uncertainty based objectives. Our experiments on exploration suggests that our method is more efficient in covering unknown areas. In terms of point-goal navigation, we showed how DD-PPO [10] augmented with our method outperforms its vanilla version. This suggests that end-to-end navigation methods can benefit from employing an uncertainty-driven planner, especially in difficult episodes.

[1]The Gibson val, MP3D val, and MP3D test sets were downloaded from https://github.com/facebookresearch/habitat-lab before 09/09/2021.

REFERENCES

[1] P. Anderson, A. Chang, D. S. Chaplot, A. Dosovitskiy, S. Gupta, V. Koltun, J. Kosecka, J. Malik, R. Mottaghi, M. Savva, *et al.*, "On evaluation of embodied navigation agents," *arXiv preprint arXiv:1807.06757*, 2018.

[2] D. Batra, A. Gokaslan, A. Kembhavi, O. Maksymets, R. Mottaghi, M. Savva, A. Toshev, and E. Wijmans, "Objectnav revisited: On evaluation of embodied agents navigating to objects," *arXiv preprint arXiv:2006.13171*, 2020.

[3] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[4] N. A. Melchior and R. Simmons, "Particle rrt for path planning with uncertainty," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 1617–1624.

[5] K. Ok, S. Ansari, B. Gallagher, W. Sica, F. Dellaert, and M. Stilman, "Path planning with uncertainty: Voronoi uncertainty fields," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 4596–4601.

[6] M. Savva, A. Kadian, O. Maksymets, Y. Zhao, E. Wijmans, B. Jain, J. Straub, J. Liu, V. Koltun, J. Malik, *et al.*, "Habitat: A platform for embodied ai research," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 9339–9347.

[7] F. Xia, A. R. Zamir, Z. He, A. Sax, J. Malik, and S. Savarese, "Gibson env: Real-world perception for embodied agents," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9068–9079.

[8] Y. Zhu, R. Mottaghi, E. Kolve, J. J. Lim, A. Gupta, L. Fei-Fei, and A. Farhadi, "Target-driven visual navigation in indoor scenes using deep reinforcement learning," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 3357–3364.

[9] T. Chen, S. Gupta, and A. Gupta, "Learning exploration policies for navigation," *7th International Conference on Learning Representations, ICLR 2019*, 2019.

[10] E. Wijmans, A. Kadian, A. Morcos, S. Lee, I. Essa, D. Parikh, M. Savva, and D. Batra, "Dd-ppo: Learning near-perfect pointgoal navigators from 2.5 billion frames," *arXiv*, pp. arXiv–1911, 2019.

[11] S. Gupta, J. Davidson, S. Levine, R. Sukthankar, and J. Malik, "Cognitive mapping and planning for visual navigation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2616–2625.

[12] S. K. Ramakrishnan, Z. Al-Halah, and K. Grauman, "Occupancy anticipation for efficient exploration and navigation," *European Conference on Computer Vision*, pp. 400–418, 2020.

[13] D. S. Chaplot, D. Gandhi, S. Gupta, A. Gupta, and R. Salakhutdinov, "Learning to explore using active neural slam," *International Conference on Learning Representations*, 2020.

[14] S. M. LaValle *et al.*, "Rapidly-exploring random trees: A new tool for path planning," 1998.

[15] H. S. Seung, M. Opper, and H. Sompolinsky, "Query by committee," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 287–294.

[16] D. Pathak, D. Gandhi, and A. Gupta, "Self-Supervised Exploration via Disagreement," *ICML*, 2019.

[17] A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, "Matterport3d: Learning from rgb-d data in indoor environments," *2017 International Conference on 3D Vision (3DV). IEEE*, 2017.

[18] J. Fuentes-Pacheco, J. Ruiz-Ascencio, and J. M. Rendón-Mancha, "Visual simultaneous localization and mapping: a survey," *Artificial intelligence review*, vol. 43, no. 1, pp. 55–81, 2015.

[19] M. Savva, A. X. Chang, A. Dosovitskiy, T. Funkhouser, and V. Koltun, "Minos: Multimodal indoor simulator for navigation in complex environments," *arXiv preprint arXiv:1712.03931*, 2017.

[20] X. Zhao, H. Agrawal, D. Batra, and A. Schwing, "The surprising effectiveness of visual odometry techniques for embodied pointgoal navigation," *arXiv preprint arXiv:2108.11550*, 2021.

[21] P. Karkus, S. Cai, and D. Hsu, "Differentiable slam-net: Learning particle slam for visual navigation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2815–2825.

[22] D. Mishkin, A. Dosovitskiy, and V. Koltun, "Benchmarking classic and learned navigation in complex 3d environments," *arXiv preprint arXiv:1901.10915*, 2019.

[23] D. S. Chaplot, D. Gandhi, A. Gupta, and R. Salakhutdinov, "Object goal navigation using goal-oriented semantic exploration," *Advances in Neural Information Processing Systems 33*, 2020.

[24] G. Georgakis, B. Bucher, K. Schmeckpeper, S. Singh, and K. Daniilidis, "Learning to map for active semantic goal navigation," *arXiv preprint arXiv:2106.15648*, 2021.

[25] G. Georgakis, Y. Li, and J. Kosecka, "Simultaneous mapping and target driven navigation," *arXiv preprint arXiv:1911.07980*, 2019.

[26] A. Mousavian, A. Toshev, M. Fišer, J. Košecká, A. Wahid, and J. Davidson, "Visual representations for semantic target driven navigation," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 8846–8852.

[27] Y. Liang, B. Chen, and S. Song, "SSCNav: Confidence-aware semantic scene completion for visual semantic navigation," *International Conference on Robotics and Automation (ICRA)*, 2021.

[28] D. S. Chaplot, R. Salakhutdinov, A. Gupta, and S. Gupta, "Neural topological slam for visual navigation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 875–12 884.

[29] O. Kwon, N. Kim, Y. Choi, H. Yoo, J. Park, and S. Oh, "Visual graph memory with unsupervised representation for visual navigation."

[30] H. J. S. Feder, J. J. Leonard, and C. M. Smith, "Adaptive mobile robot navigation and mapping," *The International Journal of Robotics Research*, vol. 18, no. 7, pp. 650–668, 1999.

[31] T. Kollar and N. Roy, "Trajectory optimization using reinforcement learning for map exploration," *The International Journal of Robotics Research*, vol. 27, no. 2, pp. 175–196, 2008.

[32] L. Carlone, J. Du, M. K. Ng, B. Bona, and M. Indri, "Active slam and exploration with particle filters using kullback-leibler divergence," *Journal of Intelligent & Robotic Systems*, vol. 75, no. 2, pp. 291–311, 2014.

[33] H. Carrillo, I. Reid, and J. A. Castellanos, "On the comparison of uncertainty criteria for active slam," in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 2080–2087.

[34] J.-L. Blanco, J.-A. Fernandez-Madrigal, and J. González, "A novel measure of uncertainty for mobile robot slam with rao—blackwellized particle filters," *The International Journal of Robotics Research*, vol. 27, no. 1, pp. 73–89, 2008.

[35] C. Stachniss, G. Grisetti, and W. Burgard, "Information gain-based exploration using rao-blackwellized particle filters." in *Robotics: Science and systems*, vol. 2, 2005, pp. 65–72.

[36] K. Fang, A. Toshev, L. Fei-Fei, and S. Savarese, "Scene memory transformer for embodied agents in long-horizon tasks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 538–547.

[37] J. Zhang, L. Tai, M. Liu, J. Boedecker, and W. Burgard, "Neural slam: Learning to explore with external memory," *arXiv preprint arXiv:1706.09520*, 2017.

[38] M. Narasimhan, E. Wijmans, X. Chen, T. Darrell, D. Batra, D. Parikh, and A. Singh, "Seeing the un-scene: Learning amodal semantic maps for room navigation," *European Conference on Computer Vision. Springer, Cham*, 2020.

[39] Y. Katsumata, A. Taniguchi, L. El Hafi, Y. Hagiwara, and T. Taniguchi, "Spcomapgan: Spatial concept formation-based semantic mapping with generative adversarial networks," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 7927–7934.

[40] E. Beeching, J. Dibangoye, O. Simonin, and C. Wolf, "Learning to plan with uncertain topological maps," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. Springer, 2020, pp. 473–490.

[41] K. Katyal, K. Popek, C. Paxton, P. Burlina, and G. D. Hager, "Uncertainty-aware occupancy map prediction using generative networks for robot navigation," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5453–5459.

[42] D. D. Fan, K. Otsu, Y. Kubo, A. Dixit, J. Burdick, and A.-A. Agha-Mohammadi, "Step: Stochastic traversability evaluation and planning for risk-aware off-road navigation," in *Robotics: Science and Systems*. RSS Foundation, 2021, pp. 1–21.

[43] È. Pairet, J. D. Hernández, M. Carreras, Y. Petillot, and M. Lahijanian, "Online mapping and motion planning under uncertainty for safe navigation in unknown environments," *IEEE Transactions on Automation Science and Engineering*, 2021.

[44] Y. Gal, "Uncertainty in deep learning," Ph.D. dissertation, University of Cambridge, 2016.
[45] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" in *Advances in neural information processing systems*, 2017, pp. 5574–5584.
[46] M. G. Azar, I. Osband, and R. Munos, "Minimax regret bounds for reinforcement learning," in *International Conference on Machine Learning*. PMLR, 2017, pp. 263–272.
[47] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2, pp. 235–256, 2002.
[48] R. Y. Chen, S. Sidor, P. Abbeel, and J. Schulman, "UCB exploration via q-ensembles," *arXiv preprint arXiv:1706.01502*, 2017.
[49] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
[50] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," *Advances in Neural Information Processing Systems 30*, 2017.
[51] Y. Gal, R. Islam, and Z. Ghahramani, "Deep bayesian active learning with image data," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1183–1192.
[52] B. Yamauchi, "A frontier-based approach for autonomous exploration," in *Proceedings 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA'97.'Towards New Computational Principles for Robotics and Automation'*. IEEE, 1997, pp. 146–151.
[53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[54] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.