# Towards Map-Agnostic Policies for Adaptive Informative Path Planning

Julius Rückin[1] David Morilla-Cabello[2] Cyrill Stachniss[1,4] Eduardo Montijano[2] Marija Popović[3]

***Abstract*—Robots are frequently tasked to gather relevant sensor data in unknown terrains. A key challenge for classical path planning algorithms used for autonomous information gathering is adaptively replanning paths online as the terrain is explored given limited onboard compute resources. Recently, learning-based approaches emerged that train planning policies offline and enable computationally efficient online replanning performing policy inference. These approaches are designed and trained for terrain monitoring missions assuming a single specific map representation, which limits their applicability to different terrains. To address this limitation, we propose a novel formulation of the adaptive informative path planning problem unified across different map representations, enabling training and deploying planning policies in a larger variety of monitoring missions. Experimental results validate that our novel formulation easily integrates with classical non-learning-based planning approaches while maintaining their performance. Our trained planning policy performs similarly to state-of-the-art map-specifically trained policies. We validate our learned policy on unseen real-world terrain datasets.**

***Index Terms*—Motion and Path Planning, Aerial Systems: Perception and Autonomy, Reinforcement Learning**

## I. Introduction

DECISION-MAKING under uncertainty in unknown terrains is a crucial skill for autonomous robots in many real-world scenarios, such as exploration [14], [15], environmental monitoring [16], [23], [29], precision agriculture [22], [30], and search and rescue [18], [26]. To complete their mission goals, robots gather relevant information about the terrain using onboard sensors. A key challenge is to adapt planned paths online based on newly incoming noisy measurements under limited onboard compute and mission budget as the robot's understanding of the terrain evolves. This problem is known in the literature as the adaptive informative path planning (IPP) problem [5], [8], [9], [16], [19], [21].

Specifically, this work examines the problem of mapping user-defined areas of interest using a budget-constrained robot with noisy onboard sensors [8], [9], [16], [22]. To this end, the robot adaptively replans paths online to maximise the information gathered about the initially unknown areas of interest based on the evolving understanding of the terrain. The gathered information is captured in a continuously updated terrain map using newly acquired sensor measurements.

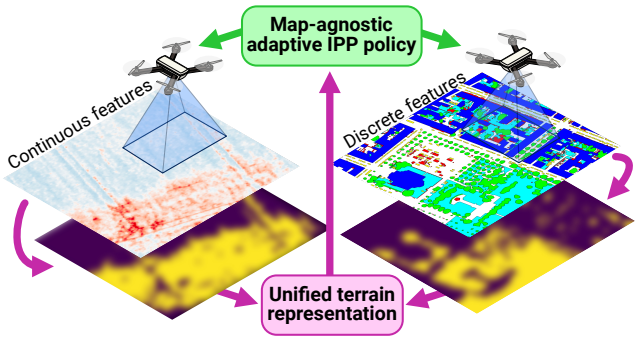


Fig. 1: Robots perform continuous- or discrete-valued terrain feature monitoring missions, e.g. mapping surface temperature or urban semantics. We transform mission-specific terrain map representations, e.g. Gaussian processes or occupancy grid maps, into a novel unified state representation for adaptive informative path planning (IPP). In this way, we design and train a single map-agnostic planning policy applicable to largely varying terrain monitoring missions.

Various adaptive IPP approaches emerged for different to-be-mapped information, i.e. terrain features, of interest during a mission. Mapping continuous-valued terrain features, e.g. bacteria levels [8] or signal strength [16], is commonly performed using Gaussian processes [8], [20], [30] or Kalman filter [22], [24] map representations. Mapping discrete-valued terrain features, e.g. crop-weed [32] or rural area semantic segmentation [22], is commonly performed using grid maps. Based on the current map, non-learning-based planning algorithms iteratively select candidate paths and evaluate their expected information value [8], [9], [16], [19]. These approaches can be adapted for different map representations. However, they tend to be too compute-intensive for frequent online replanning as they rely on costly evaluations of many potential future paths. To overcome these issues, learning-based approaches have been proposed. These methods train adaptive IPP policies offline in simulation and perform compute-efficient policy inference at deployment [1], [2], [4], [14], [18], [24], [31]. Although learning-based approaches show promising performance, they are specifically designed for and trained on a single terrain map representation. This prohibits their direct application to a larger variety of terrain monitoring missions.

Manuscript received: Oct, 21, 2024; Revised Jan, 24, 2025; Accepted Mar, 18, 2025. This paper was recommended for publication by Editor Giuseppe Loianno upon evaluation of the Associate Editor and Reviewers' comments.

[1]J. Rückin. and C. Stachniss are with the Center for Robotics, University of Bonn, Germany. [2]D. Morilla-Cabello and E. Montijano are with DIIS-I3A, Universidad de Zaragoza, Spain. [3]M. Popović is with the MAVLab, TU Delft, Netherlands. [4]C. Stachniss is also with the Lamarr Institute for Machine Learning and Artificial Intelligence, Germany.

This work has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy, EXC-2070 – 390732324 (PhenoRob), DGA project T45_23R, MCIN/AEI/ERDF/European Union NextGenerationEU/PRTR project PID2021-125514NB-I00, ONR grant N62909-24-1-2081 and grant FPU20-06563. Corresponding author: jrueckin@uni-bonn.de.

Digital Object Identifier (DOI): see top of this page.

We argue that the broad pool of existing adaptive IPP approaches should be viewed along two dimensions: the map-specific formulation modelling the adaptive IPP problem and the algorithm used to offline-train or online-search the planning policy. The formulation of the adaptive IPP problem is the most critical design decision to ensure the unified applicability of planning policies across various terrain monitoring missions. This motivates the need for a map-agnostic formulation of the adaptive IPP terrain monitoring problem that directly integrates with any (non)-learning-based policy search algorithm used for adaptive IPP. Particularly, this formulation ensures training and deploying learned policies in largely varying terrain monitoring missions.

The main contribution of this paper is such a novel map-agnostic formulation of the adaptive IPP problem for terrain monitoring illustrated in Fig. 1. Our formulation unifies continuous-valued, i.e. regression, and discrete-valued, i.e. classification, terrain feature monitoring for adaptive IPP policies. To achieve this, we unify state space representations across terrain map representations utilised for replanning online. Based on this unified state space and a new reward function, we train and deploy a single generally applicable planning policy on previously unmet variations of terrain monitoring missions using reinforcement learning (RL).

In sum, we make the following claims. First, our map-agnostic planning policy trained and deployed on vastly varying simulated terrain monitoring missions performs on par or better than state-of-the-art map-specifically trained policies and non-learning-based adaptive IPP approaches. Second, our map-agnostic policy performs similarly to state-of-the-art adaptive IPP approaches on various real-world terrain datasets. Third, in our experiments, we demonstrate that our map-agnostic adaptive IPP formulation easily integrates with previous non-learning-based state-of-the-art adaptive IPP algorithms while maintaining or improving their performance. We will open-source our code for usage by the community at: https://github.com/dmar-bonn/ipp-rl-gen.

## II. Related Work

This work addresses the problem of robotic information gathering in initially unknown terrains where certain areas are considered more interesting than others, e.g. temperature hotspots [4], [22] or search and rescue victims [18], [26]. This problem is known as the adaptive IPP problem [8], [9], [16], [19], [21], where the aim is to efficiently discover and precisely map the areas of interest using a resource-constrained robot, e.g. an unmanned aerial vehicle (UAV) with limited battery capacity [22], [29], [30]. Various adaptive IPP approaches have been proposed, which actively replan paths online during a mission based on the robot's state and previously collected measurements. In contrast, often less efficient non-adaptive approaches, e.g. coverage paths [7], [27], pre-compute static paths that cannot be modified during a mission.

Methods for adaptive IPP can be categorised into non-learning-based and learning-based planning approaches. Non-learning-based approaches have been successfully applied to many different variants of the adaptive IPP problem, such as exploration [14], [15], search and rescue [5], [26], and terrain monitoring [8], [9], [30]. Sampling-based methods solve the adaptive IPP problem by iteratively (re-)sampling potential paths and evaluating their information value based on the robot's terrain understanding, building upon established sampling-based search, such as receding horizon planning [9] or Monte-Carlo tree search (MCTS) [5], [19]. Optimisation-based methods utilise derivative-free optimisation, such as evolutionary algorithms [8], [22] or Bayesian optimisation [16], [30], directly maximising the information acquired along the path. Although non-learning-based planning methods for adaptive IPP show promising results, they tend to be computationally inefficient [1], [22], [24] as they evaluate expensive-to-compute information criteria for many potential future paths, prohibiting fast online replanning or sacrificing path quality. Further, these approaches directly use mission-specific terrain map representations to design the planning state space, which requires adapting planning methods for deploying them in monitoring missions with different terrain maps.

Recently, learning-based methods were proposed to tackle the adaptive IPP problem, providing higher compute efficiency and achieving similar or better planning performance. This is done by shifting the computational burden to an offline training phase, simulating many terrain monitoring missions, and inferring the learned planning policy at deployment [1], [3], [4], [14], [24], [28], [29], [31], [32]. RL methods have been proposed for specific adaptive IPP applications, such as terrain exploration [2], [14], [18] and monitoring [1], [4], [24], [28], [31]. These works mainly differ in their reward function design influenced by the mission goal and terrain map representation, and the used policy networks trained with different RL algorithms. Methods for exploration design reward functions measuring coverage of the terrain [2], [14], while works considering efficiently finding and precisely mapping areas of interest commonly reward decreasing terrain map uncertainty in these areas [1], [24], [29], [32]. All these approaches maintain a mission-specific spatial terrain map representation to fuse onboard measurements, such as occupancy maps [2], [14], [18], [32], [33] or sub-sampled Gaussian processes [1], [4], [28], [31]. These approaches directly use mission-specific terrain map representations to design the planning algorithm's map-specific state space and train the adaptive IPP policy on this map-specific state space.

Overall, all previous adaptive IPP approaches propose terrain map-specific solutions, assuming either occupancy maps [2], [14], [18], [22], [32], [33], pre-trained Gaussian processes [1], [4], [8], [19], [28], [31] or Kalman filters [22], [24], as their planning state representation. Thus, these methods require adaptation and re-training as the map, and hence, planning state representations change. This prohibits the application of learned policies to various monitoring missions that require different map representations. In contrast, we present a novel map-agnostic state formulation for adaptive IPP unifying terrain monitoring missions across various map representations. Combining this map-agnostic state space with a new reward function, we train a single planning policy using RL that is applicable to continuous- and discrete-valued terrain feature monitoring missions.

## III. Problem Formulation

This work aims to formulate the adaptive IPP problem for terrain monitoring [9], [10], [16], [17], [22] in a terrain map-agnostic fashion to offline-learn or online-solve planning policies across different monitoring missions and map representations without adapting or re-training policies. We consider a robot with pose $\mathbf{p}_t \in \mathbb{R}^{D_r}$ at time $t$, moving in an *a priori* unknown terrain. The terrain $\xi \subset \mathbb{R}^{D_e}$ is characterised by its initially unknown and stationary feature field $F : \xi \to \mathcal{F}$, where $\mathcal{F}$ is the mission-specific continuous or discrete terrain feature space. The goal is to estimate and precisely map the terrain feature field $F$ in interesting areas,

$$\xi_I = \{\mathbf{x} \in \xi \mid F(\mathbf{x}) \in \mathcal{F}_I\} \subseteq \xi\,, \tag{1}$$

where $\mathcal{F}_I \subseteq \mathcal{F}$ is the user-defined subset of feature values qualifying the interest in a point $\mathbf{x} \in \xi$, e.g. a value range or semantic classes, with each $\mathbf{x} \in \xi_I$ being of equal interest.

To accomplish this objective, the robot is equipped with a sensor to collect measurements $z \in \mathcal{Z}$ from the terrain, e.g. semantically segmented RGB images, thermal images, or radiation levels. At each time step $t$, the measurements provide noisy information about $F$ according to $z_t \sim p(z \mid \mathbf{p}_t, F)$ and are used to model a stochastic process $\hat{F}_t$ over all possible terrain feature field functions $F$,

$$\hat{F}_t \sim p(F \mid z_{1:t}, \mathbf{p}_{1:t}, \theta_F)\,, \tag{2}$$

where $z_{1:t}$ is the set of all collected measurements at robot poses $\mathbf{p}_{1:t}$, and $\theta_F$ indicates the chosen map representation and its hyperparameters. Most works update the belief for continuous-valued feature spaces $\mathcal{F} \subseteq \mathbb{R}$ with pre-trained Gaussian processes or Kalman filters, and for discrete-valued feature spaces $\mathcal{F} \subseteq \mathbb{N}$ with occupancy grid mapping.

We aim to find an optimal action sequence $\psi^* = (\mathbf{a}_1, \dots, \mathbf{a}_N)$, where $\mathbf{a}_t \in \mathcal{A} \subseteq \mathbb{R}^{D_a}$ are relative pose changes. The action sequence $\psi^*$ maximises an information criterion $I : \mathcal{A}^N \times \xi_I \to \mathbb{R}$, where $\mathcal{A}^N$ encompasses all action sequences of variable length $N$, associating the sensor measurements collected while executing an action sequence $\psi$ with their information value about areas of interest $\xi_I$,

$$\psi^* = \underset{\psi \in \mathcal{A}^N}{\operatorname{argmax}}\ I(\psi, \xi_I),\ \text{s.t. } C(\psi) \leq B\,, \tag{3}$$

where $C : \mathcal{A}^N \to \mathbb{R}$ is the action sequence execution cost, e.g. battery capacity or travel time, and $B \geq 0$ is the robot's fixed maximum mission budget. As $F$ and thus $\xi_I$ are *a priori* unknown, Eq. (3) cannot be solved offline. The optimal action sequence $\psi^*$ in Eq. (3) changes as $\hat{F}_t$ is updated based on new measurements. Therefore, online replanning is required to find an optimal $\psi^*$ that *adaptively* focuses on areas of interest $\xi_I$ as they are discovered.

The concrete formulation of Eq. (3) depends on the specific terrain monitoring mission. Depending on the mission characteristics, the spatially mapped terrain feature space $\mathcal{F}$ might be discrete, such as semantic classes, or continuous, such as surface temperature. For a given mission, the user defines interesting features $\mathcal{F}_I \subseteq \mathcal{F}$ and chooses the map representation $\hat{F}_t$ with map hyperparameters $\theta_F$. We denote $\mathcal{H} = \{\mathcal{F}, \mathcal{F}_I, \theta_F\}$ as the set of mission hyperparameters defining the specific instantiation of Eq. (3).

As shown in previous works [1], [2], [14], [18], [24], [31], [33], the adaptive IPP problem in Eq. (3) can be transformed into an RL problem for many terrain monitoring mission variants by

$$\begin{aligned}\pi^* &= \underset{\pi \in \Pi}{\operatorname{argmax}}\, I\left((\pi(s_1), \dots, \pi(s_N)), \xi_I\right) \\ &= \underset{\pi \in \Pi}{\operatorname{argmax}} \sum_{t=1}^{N} \gamma^{t-1} R\left(s_t, \pi(s_t), s_{t+1}, \xi_I\right),\end{aligned} \tag{4}$$

where $\pi : \mathcal{S} \to \mathcal{A}$ is a planning policy mapping state $s_t \in \mathcal{S}$ at a time step $t$ to an action $\mathbf{a}_t = \pi(s_t)$, and $\Pi$ is the function space of all possible policies. Thus, the action sequence $\psi$ is given by $\psi = (\pi(s_1), \dots, \pi(s_N))$. A mission- and map-specific reward function $R(s_t, \pi(s_t), s_{t+1}, \xi_I) \in \mathbb{R}$ resembles the information criterion $I$, rewarding taking actions $\mathbf{a}_t$ in state $s_t$ that lead to a next state $s_{t+1}$ with increased information about interesting areas $\xi_I$, and $\gamma \in [0, 1]$ is a discount factor. As interesting areas $\xi_I$ are unknown during a mission, prior adaptive IPP methods [1], [8], [22], [24], [32] approximate unknown areas of interest $\xi_I$ using map-specifically computed confidence intervals based on hand-tuned confidence thresholds, rewarding uncertainty reduction over map belief $\hat{F}_t$ in these approximated interesting areas.

Different from existing adaptive IPP approaches that consider terrain map-specific planning state formulations $s_t$ with approximated areas of interest, we formulate the problem in Eq. (4) in a fully probabilistic map-agnostic fashion. To this end, we propose a planning state $s_t$ that unifies the adaptive IPP problem across different map representations $\hat{F}_t$, allowing us to apply a single learned policy $\pi^*$ to varying terrain monitoring missions. Based on this planning state, we introduce a new reward function for Eq. (4) enabling training or online-solving policy $\pi^*$ for different monitoring missions.

## IV. Our Approach

Our approach is conceptually depicted in Fig. 1. We unify the adaptive IPP problem formulation introduced in Sec. III across different map representations required to spatially capture various continuous- and discrete-valued terrain features. To this end, we view any terrain monitoring mission as a binary classification task, probabilistically splitting the terrain into unknown interesting areas $\xi_I$ (Eq. (1)) and uninteresting areas $\xi - \xi_I$. Based on this belief over interesting areas $\xi_I$, we propose a map-agnostic planning state space (Sec. IV-A) and introduce a reward function to online-solve or offline-train a planning policy across terrain monitoring missions with different map representations (Sec. IV-B). Last, we show how we use our state formulation and reward function to offline-train adaptive IPP policies on varying terrain monitoring missions in simulation (Sec. IV-C).

### A. Unified Planning State Space for Adaptive IPP

Our formulation of planning states $s_t \in \mathcal{S}$ encodes all information required to solve the adaptive IPP problem in Eq. (4), i.e. the robot's state estimation, its current understanding of the

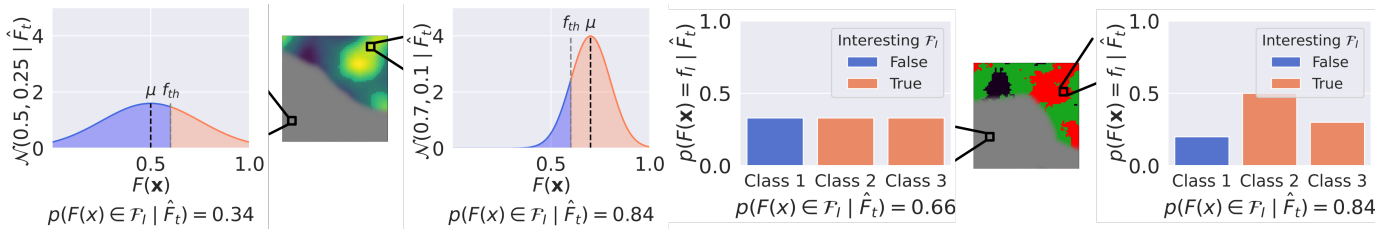


Fig. 2: Our unified belief $p\big(F(\mathbf{x}) \in \mathcal{F}_I \mid \hat{F}_t\big)$ over interesting areas $\mathbf{x} \in \xi_I$ for continuous- (left) and discrete-valued (right) terrain features. Grey areas are unknown with large map uncertainty. (Left) Posterior normal distributions inferred from a Gaussian process or Kalman filter map representation with an interesting value threshold $f_{th} = 0.6$. The unified belief is computed by the orange area under the curve, which is larger for known interesting areas than for unknown uncertain areas. (Right) The unified belief is given by the sum of posterior probability masses over interesting classes (orange) extracted from an occupancy map representation.

terrain, and mission hyperparameters. We propose a unified belief over interesting terrain areas reusable as input to the planning policy $\pi(s_t)$ for feature fields $F$ with continuous- and discrete-valued terrain features $\mathcal{F}$ that might require different map representations $\hat{F}_t$. Assume that $\mathcal{X}_t$ is a set of points $\mathbf{x}_t \in \xi$ sampled from the terrain $\xi$ at time step $t$ at which we aim to infer the state $s_t(\mathbf{x}_t)$. Then, for each $\mathbf{x}_t \in \mathcal{X}_t$, $s_t(\mathbf{x}_t)$ is defined as

$$s_t(\mathbf{x}_t) = \Big(p\big(F(\mathbf{x}_t) \in \mathcal{F}_I \mid \hat{F}_t\big), H\big(\hat{F}_t(\mathbf{x}_t)\big), \mathbf{p}_t, B_t, \mathcal{H}\Big), \quad (5)$$

where $p\big(F(\mathbf{x}_t) \in \mathcal{F}_I \mid \hat{F}_t\big)$ is the probability of $\mathbf{x}_t$ being part of an interesting area $\xi_I$, $H\big(\hat{F}_t(\mathbf{x}_t)\big)$ is the uncertainty of the mission-specific map belief $\hat{F}_t$ at $\mathbf{x}_t$, $\mathbf{p}_t$ is the robot's current position, $B_t \leq B$ is the robot's remaining budget, and $\mathcal{H}$ are the mission hyperparameters specifying Eq. (3). For occupancy maps, $H\big(\hat{F}_t(\mathbf{x}_t)\big)$ is the Shannon entropy at $\mathbf{x}_t$. For Gaussian processes or Kalman filters, $H\big(\hat{F}_t(\mathbf{x}_t)\big)$ is the variance at $\mathbf{x}_t$. Our state can be integrated with any representation $\mathcal{X}_t$ of terrain $\xi$ to compute the state representation $s_t$ over $\mathcal{X}_t$. For example, it supports equidistant grids [19], [22], [24], [32] or randomly sampled graphs [1], [2], [28], [31].

In contrast to previous works relying on map-specific formulations of $s_t$ with binary approximations of interesting areas, our planning state formulation in Eq. (5) introduces a fully probabilistic map-agnostic belief over interesting areas. Next, we show how to compute this map-agnostic belief $\hat{F}_{I,t} \sim p\big(F(\mathbf{x}_t) \in \mathcal{F}_I \mid \hat{F}_t\big)$ for continuous- and discrete-valued terrain feature mapping missions with different fully probabilistic map representations $\hat{F}_t$ as illustrated in Fig. 2.

Consider *discrete feature spaces* $\mathcal{F} = \{1, \dots, K\}$ with $K \in \mathbb{N}$ semantic classes. Interesting areas $\xi_I$ are given by a user-defined set of interesting features $\mathcal{F}_I \subseteq \mathcal{F}$ with $|\mathcal{F}_I| \leq K$. As the map belief $\hat{F}_t \sim p(F \mid z_{1:t}, \mathbf{p}_{1:t})$ is represented using occupancy grid maps, the unified belief $p\big(F(\mathbf{x}) \in \mathcal{F}_I \mid \hat{F}_t\big)$ over interesting areas is defined as

$$p\big(F(\mathbf{x}) \in \mathcal{F}_I \mid \hat{F}_t\big) = \sum_{f_I \in \mathcal{F}_\mathcal{I}} p\big(F(\mathbf{x}) = f_I \mid z_{1:t}, \mathbf{p}_{1:t}\big), \quad (6)$$

where $f_I$ is a single class in the set of interesting classes $\mathcal{F}_I$ and $p\big(F(\mathbf{x}) = f_I \mid z_{1:t}, \mathbf{p}_{1:t}\big)$ is given by the $f_I$-th layer of the occupancy map at the grid cell corresponding to $\mathbf{x} \in \xi$, defining the categorical distribution over all $K$ classes.

Next, consider *continuous feature spaces* $\mathcal{F} = [f_a, f_b]$ with $f_a \leq f_b$. Interesting areas are given by user-defined thresholds $f_{th}$ with $f_a \leq f_{th} \leq f_b$, such that $\mathcal{F}_I = [f_{th}, f_b]$. As the map belief $\hat{F}_t$ is represented by Gaussian processes or Kalman filters, the probability density over feature values is given by $F(\mathbf{x}) \sim \mathcal{N}(\mu(\mathbf{x}), \sigma(\mathbf{x})^2 \mid \hat{F}_t)$ with mean $\mu(\mathbf{x})$ and variance $\sigma(\mathbf{x})^2$ of $\hat{F}_t$ at point $\mathbf{x}$. The unified belief $p\big(F(\mathbf{x}) \in \mathcal{F}_I \mid \hat{F}_t\big)$ over interesting areas is defined as

$$\begin{aligned} &p\big(F(\mathbf{x}) \in \mathcal{F}_I \mid \hat{F}_t\big) \\ &= \frac{1}{\sqrt{2\pi\sigma(\mathbf{x})^2}} \int_{f_{th}} \exp\left(-\frac{(f - \mu(\mathbf{x}))^2}{2\sigma(\mathbf{x})^2}\right) df \\ &= 1 - \Phi\left(\frac{f_{th} - \mu(\mathbf{x})}{\sqrt{\sigma(\mathbf{x})^2}}\right), \end{aligned} \quad (7)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution measuring $p\big(F(\mathbf{x}) \leq f_{th} \mid \hat{F}_t\big)$.

The mission-specific hyperparameters $\mathcal{H} = \{\mathcal{F}, \mathcal{F}_I, \theta_F\}$ directly influence the computation of our unified belief over interesting areas $\hat{F}_{I,t}$ in Eq. (6) or Eq. (7), making the effect of the chosen mission hyperparameters accessible to the planning policy, thus improving adaptivity to the concrete instance of Eq. (4) a planning method aims to solve. For learning-based planning methods aiming to train a policy $\pi^*$ offline, we additionally condition the planning policy on the mission-specific hyperparameters as it allows us to train a single policy that can solve Eq. (4) for various terrain monitoring variants $\mathcal{H}$ without retraining.

### B. Adaptive IPP Reward Function

We introduce a new reward function for the general adaptive IPP terrain monitoring problem in Eq. (4) based on our unified planning state space formulation $s_t$ presented in Sec. IV-A. The unified planning state space and reward function could be integrated into any non-learning-based planning method searching for the optimal policy $\pi^*$ online or learning-based planning method for training $\pi^*$ offline.

In adaptive IPP problems, we aim to quickly find initially unknown areas of interest $\xi_I$ (Eq. (1)) and precisely estimate the terrain feature field $F$ in these areas. To this end, we aim to maximise information about the map belief $\hat{F}_t \sim p(F \mid z_{1:t}, \mathbf{p}_{1:t})$ in unknown areas of interest $\xi_I$ (Eq. (2)). To adapt paths online towards areas likely of interest, we reward uncertainty reduction of map belief $\hat{F}_t$ proportionally to our unified belief over interesting areas

$\hat{F}_{I,t}(\mathbf{x}) \sim p\big(F(\mathbf{x}) \in \mathcal{F}_I \mid \hat{F}_t\big)$ (Eq. (6), Eq. (7)). Assume $\mathcal{X}$ is a finite subset of points $\mathbf{x} \in \xi$ sampled from an equidistant grid over the terrain $\xi$. The reward in Eq. (4) is defined as

$$R(s_t, \mathbf{a}_t, s_{t+1}) = \sum_{\mathbf{x} \in \mathcal{X}} \frac{H(\hat{F}_t(\mathbf{x})) - H(\hat{F}_{t+1}(\mathbf{x}))}{H(\hat{F}_t(\mathbf{x}))} p\big(F(\mathbf{x}) \in \mathcal{F}_I \mid \hat{F}_t\big), \tag{8}$$

where $H(\hat{F}_t(\mathbf{x}))$ is the uncertainty of the mission-specific map belief $\hat{F}_t$ at a point $\mathbf{x} \in \mathcal{X}$ and $\hat{F}_{t+1}$ is the updated map belief after executing action $\mathbf{a}_t$ and collecting a new observation $z_{t+1} \sim p(z \mid \mathbf{p}_{t+1}, F)$ from a next pose $\mathbf{p}_{t+1}$. For occupancy maps, $H\big(\hat{F}_t(\mathbf{x}_t)\big)$ is the exponential Shannon entropy at $\mathbf{x}$. For Gaussian processes and Kalman filters, $H\big(\hat{F}_t(\mathbf{x}_t)\big)$ is the variance trace at $\mathbf{x}$.

Assume two points $\mathbf{x}, \mathbf{x}' \in \xi$. If both points have the same probability of belonging to areas of interest $\xi_I$, the reward favours points $\mathbf{x}$ with higher expected map uncertainty reduction to foster exploration. If both points' expected map uncertainty reduction is the same, the reward favours point $\mathbf{x}$ with a higher probability of belonging to areas of interest $\xi_I$ to focus on these areas as they are discovered. By definition of Eq. (1), our reward also contains pure terrain exploration scenarios with areas of interest $\xi_I = \xi$ covering the whole terrain as a special case if all feature values $\mathcal{F} = \mathcal{F}_I$ are of interest. In these cases, $p(F(\mathbf{x}) \in \mathcal{F}_I \mid \hat{F}_t) = 1$ for all $\mathbf{x}, \mathbf{x}' \in \xi$ by definition of Eq. (6) and Eq. (7). Thus, points $\mathbf{x}$ with higher expected map uncertainty reduction are favoured, fostering the exploration of the whole terrain.

### C. Planning Policy Training Details

We use RL to train a single unified planning policy $\pi^*$ on simulated terrain monitoring deployments with previously unmet mission variations. We detail our terrain monitoring mission simulations, encoding of mission hyperparameters $\mathcal{H}$ and the used RL algorithm and policy network representing $\pi^*$. In practice, any policy learning method, e.g. imitation learning, policy network architecture, and hyperparameter encoding could be used to train the policy with our new adaptive IPP formulation introduced in Sec. IV-A and Sec. IV-B.

**Mission simulations.** We sample mission hyperparameters $\mathcal{H} = \{\mathcal{F}, \mathcal{F}_I, \theta_F\}$, defining the monitoring mission. We randomly choose continuous- or discrete-valued terrain features $\mathcal{F}$. Note that we train our policy on both classes of terrain feature monitoring missions to minimise the training to deployment gap [13] and maximise planning performance. In case of continuous-valued features, we use a Gaussian process with sampled kernel parameters $\theta_F$ to represent the map belief $\hat{F}_t$. In case of discrete-valued features, we use an occupancy map to represent $\hat{F}_t$. For given features, we simulate randomised ground truth feature fields $F$ with spatial correlations of different extents as depicted in Fig. 3.

**Hyperparameter encoding.** We explicitly input mission hyperparameters $l_{GP}$ and $f_{th}$ into state $s_t$. The map hyperparameter $l_{GP} \geq 0 \in \theta_F$ is the lengthscale of a Gaussian process Matern kernel used to represent the map belief $\hat{F}_t$. This is important as different lengthscales result in different map updates along paths, potentially affecting decision-making. Map beliefs $\hat{F}_t$ assuming spatially independent measurements $z$, e.g. occupancy grid maps, are naturally encoded by $l_{GP} = 0$ as Matern kernels with $l_{GP} \to 0$ assume spatially independent measurements. The user-defined value threshold $f_{th} \in \mathcal{F}$ represents the interesting features $\mathcal{F}_I$.

**Policy training.** We train our policy $\pi^*$ using the proximal policy optimisation algorithm [25] and compute our state space in Eq. (5) over an equidistant grid $\mathcal{X}_t$. We use the IMPALA encoder [6] to process the interesting area belief $p\big(F(\mathbf{x}) \in \mathcal{F}_I \mid \hat{F}_t\big)$ and map belief uncertainty $H(\hat{F}_t(\mathbf{x}))$ for each $\mathbf{x} \in \mathcal{X}_t$. We use a multilayer perceptron (MLP) to process the current robot's pose $\mathbf{p}_t$, remaining budget $B_t$ and mission hyperparameters $\mathcal{H}$, and a MLP head to predict the stochastic policy $\pi(s_t)$ and value function $V_\pi(s_t)$.

## V. Experimental Results

The experiments are designed to support our claims. In Sec. V-B, we show that training our map-agnostic policy on various monitoring missions yields competitive performance with state-of-the-art online non-learning-based policy search methods and offline-learned policies adapted and re-trained for each class of monitoring missions with specific terrain map representation. In Sec. V-C, we verify that our map-agnostic policy trained in simulation performs similarly to these state-of-the-art adaptive IPP methods on unseen real-world datasets. In Sec. V-D, we show that our map-agnostic adaptive IPP formulation unifies existing adaptive IPP methods while maintaining or improving their performance.

### A. Experimental Setup

**Mission setup.** The general procedure for simulating monitoring missions used to train and evaluate planning policies is described in Sec. IV-C. For discrete-valued terrain features, we assume three semantic classes $\mathcal{F}$ with interesting classes $\mathcal{F}_I$ of varying spatial extent. We equip a simulated UAV with a sensor delivering image-like semantic measurements $z_t$ spanning a downwards-projected field of view. We use occupancy grid maps $\hat{F}_t$ for terrain mapping and confusion matrix-based sensor noise as in [22], [32]. For continuous-valued terrain features, we assume features $\mathcal{F} = [0, 1]$ with interesting thresholds $f_{th}$, such that $\mathcal{F}_I = [f_{th}, 1]$. Simulated UAVs are equipped with sensors delivering point measurements $z_t$ with Gaussian noise, mapped using Gaussian processes as in [1], [8], [19], [22]. We distinguish between the classical evaluation protocol of fixed mission hyperparameters $\mathcal{H} = \{f_{th}, l_{GP}\} = \{0.4, 0.35\}$ as in [1], [8], [22], [24], denoted as *Static*, and our more challenging scenario of randomly sampled $\mathcal{H}$ with $f_{th} \in [0.0, 0.8]$ and $l_{GP} \in [0.15, 0.55]$ denoted below as *Varying*. This resembles the static mission hyperparameters in expectation. The initial mission budget is set to $B = 100$s, and initial robot positions $\mathbf{p}_0$ are sampled at random. We assume actions $\mathbf{a}_t \in \mathcal{A}$ representing relative 2D robot position changes on an equidistant grid as in [19], [22], [24], [32]. To train and benchmark our approach, we simulate ground truth feature fields $F$ with varying spatial

TABLE I: Comparison of state-of-the-art map-specifically designed and trained methods to our map-agnostic planning policy (*RL-Ours*) on simulated continuous- and discrete-valued terrain feature monitoring missions. Best average performances are marked in bold, second-best average performances are underlined if standard deviations in brackets overlap. Our map-agnostic policy performs best in case of *Varying* user-defined mission hyperparameters and similar to state-of-the-art adaptive IPP methods in case of *Static* mission hyperparameters.

| | Approach | Static $\mathcal{H}$ II↑ | Unc.↓ | MLL↓ | RMSE↓ | Varying $\mathcal{H}$ II↑ | Unc.↓ | MLL↓ | RMSE↓ | Replanning time [s]↓ |
|---|---|---|---|---|---|---|---|---|---|---|
| Continuous | RL-Ours | 25.8 (0.17) | 60.6 (0.22) | -64.6 (0.12) | 3.83 (0.08) | **26.2** (0.65) | **60.4** (0.64) | **-60.5** (0.59) | 3.81 (0.07) | **0.004** |
| | RL-Base-C | **26.1** (0.25) | **59.6** (0.28) | **-66.2** (0.39) | **3.67** (0.02) | 24.0 (0.94) | 64.2 (1.07) | -48.4 (4.64) | 5.51 (0.97) | **0.004** |
| | MCTS | 25.6 (0.09) | 60.7 (0.08) | -64.9 (0.52) | 3.83 (0.16) | 25.3 (0.41) | 61.1 (0.24) | -59.5 (0.70) | 4.27 (0.22) | 2.86 |
| | CMA-ES | 23.1 (1.27) | 63.0 (3.41) | -60.1 (6.38) | 5.45 (2.67) | 21.5 (1.44) | 64.3 (2.32) | -55.4 (5.52) | **2.59** (0.85) | 6.05 |
| | Greedy | 24.5 (0.14) | 62.0 (0.12) | -62.0 (0.26) | 4.11 (0.11) | 25.2 (0.66) | 61.7 (0.29) | -58.0 (1.55) | 4.35 (0.24) | 0.05 |
| | Coverage | 15.3 (0.16) | 75.5 (0.39) | -28.0 (1.44) | 10.8 (0.12) | 13.4 (0.28) | 77.1 (0.25) | -30.6 (0.87) | 9.93 (0.20) | - |
| | | II↑ | Unc.↓ | mIoU↑ | F1↑ | II↑ | Unc.↓ | mIoU↑ | F1↑ | |
| Discrete | RL-Ours | **31.5** (0.12) | **38.3** (0.76) | **20.8** (0.26) | **25.6** (0.19) | 30.5 (0.12) | **39.2** (0.42) | **20.4** (0.17) | **25.3** (0.12) | **0.004** |
| | RL-Base-D | 31.1 (0.09) | 38.6 (0.43) | 20.7 (0.14) | 25.5 (0.09) | 30.4 (0.57) | 40.2 (0.95) | 20.1 (0.29) | 25.0 (0.26) | **0.004** |
| | MCTS | 30.7 (0.25) | 41.9 (0.31) | 19.5 (0.08) | 24.5 (0.12) | **30.8** (0.21) | 41.2 (0.85) | 19.8 (0.31) | 24.7 (0.24) | 1.95 |
| | CMA-ES | 29.6 (1.87) | 43.2 (2.38) | 19.2 (0.85) | 24.2 (0.76) | 30.0 (1.45) | 42.4 (0.61) | 19.5 (0.42) | 24.4 (0.51) | 3.75 |
| | Greedy | 29.9 (0.24) | 44.4 (0.83) | 18.7 (0.26) | 23.8 (0.22) | 29.4 (0.17) | 45.6 (0.59) | 18.2 (0.22) | 23.2 (0.22) | 0.03 |
| | Coverage | 29.7 (0.46) | 44.3 (0.37) | 18.7 (0.12) | 23.8 (0.12) | 27.9 (0.24) | 45.3 (0.37) | 18.4 (0.12) | 23.3 (0.08) | - |

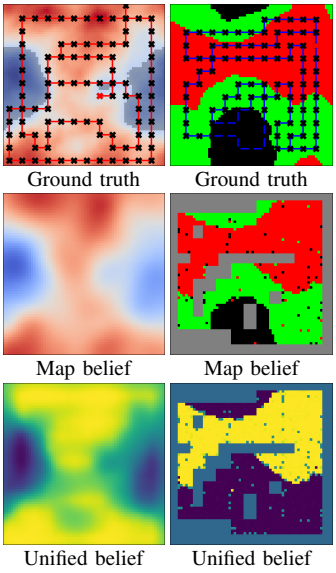


Fig. 3: Continuous (left) and discrete terrain feature fields (right).

correlations as shown in Fig. 3-top. Additionally, we evaluate the performance on real-world orthomosaic fields $F$.

**Baselines.** We consider state-of-the-art adaptive IPP methods performing online planning or offline-trained policy inference. In contrast to our *RL-Ours* method, all baseline policies rely on map-specific planning state spaces. All methods consider the current robot position and remaining budget in their state. Continuous-valued terrain features are modelled by directly using posterior mean and variance of the Gaussian process as in [1], [8], [19], [22], [24]. Discrete-valued terrain features are modelled by directly using the posterior occupancy map and its entropy as in [22], [32]. All baselines reward map uncertainty reduction in approximated areas of interest relying on hand-tuned confidence intervals as in [1], [8], [22]. Based on these states and rewards, we implement finite-horizon rollout-based *MCTS* [5], [19], finite-length path optimisation using the *covariance matrix adaptation evolution strategy (CMA-ES)* [8], [22], and *Greedy* planning [22] as online policy search methods. To offline-train *RL-Base* planning policies, we use RL assuming *Static* hyperparameters and perform policy inference online [1], [18], [24], [32]. Further, we pre-compute lawnmower-like *Coverage* paths [7] commonly used in real-world monitoring deployments.

**Evaluation metrics.** All adaptive replanning performance metrics are computed over areas of interest $\xi_I$ (Eq. (1)) after a mission is terminated. For continuous-valued mapping missions, we compute the final covariance log-trace of map $\hat{F}_t$ normalised by the prior covariance log-trace of $\hat{F}_0$ (Unc.) and root mean squared error (RMSE) as in [1], [19], [22], and mean log loss (MLL) of $\hat{F}_t$ w.r.t. the ground truth feature field $F$ as computed by Marchant and Ramos [16], Eq. (23). For discrete-valued mapping missions, we compute the final Shannon entropy of map $\hat{F}_t$ normalised by the prior Shannon entropy of $\hat{F}_0$ (Unc.), and mean Intersection-over-Union (mIoU) and F1-score of $\hat{F}_t$ w.r.t. the ground truth feature field $F$ as in [18], [22], [32]. Further, we compute an information integral (II) as one minus the area under the normalised map uncertainty (Unc.) over budget curve. The II captures the uncertainty reduction speed over the depleted budget in a single metric. All metrics are averaged over 100 missions, repeated with three different random seeds. We report mean and standard deviations over the three seeds.

### B. Simulation Results

The first set of experiments shows that our single map-agnostic adaptive IPP policy yields competitive performance with state-of-the-art online policy search methods while substantially reducing replanning runtime. Further, our map-agnostic policy outperforms state-of-the-art map-specifically designed and offline-trained policies on various terrain monitoring missions. We evaluate all methods in simulated continuous- and discrete-valued terrain feature monitoring scenarios as described in Sec. V-A. We consider the classical *Static* mission hyperparameter and our *Varying* mission hyperparameter evaluation protocol to benchmark adaptive IPP approaches on challenging inter-mission variations.

Tab. I summarises the results. In line with previous RL-based adaptive IPP works, map-specifically designed *RL-Base-C* and *RL-Base-D* policies outperform state-of-the-art online policy search methods in their respective continuous- and discrete-valued terrain feature monitoring missions with *Static* hyperparameters they were trained on. Our single map-agnostic *RL-Ours* policy shows competitive performance on continuous- and discrete-valued monitoring missions with *Static* hyperparameters compared to online policy search methods and the *RL-Base-C/D* policies. Noticeably, our map-agnostic policy outperforms the map-specific *RL-Base-C/D* policies on *Varying* hyperparameters, causing larger inter-mission variations. This verifies the advantage of our unified policy being trained and conditioned on larger mission variations, while the *RL-Base-C* trained on *Static* hyperparameters does not match the performance of online policy search. Further, our map-agnostic policy outperforms the strongest

TABLE II: Comparison of state-of-the-art map-specifically designed and trained methods to our map-agnostic planning policy (*RL-Ours*) on real-world continuous-valued surface temperature (*Temperature-1/2*) and discrete-valued urban (*Potsdam*) and rural (*RIT-18*) semantic terrain datasets. Best average performances are marked in bold, second-best average performances are underlined if standard deviations in brackets overlap. Our map-agnostic policy performs similarly to state-of-the-art adaptive IPP methods.

| **Approach** | **Temperature-1** | | **Temperature-2** | | **Potsdam** [11] | | **RIT-18** [12] | |
|---|---|---|---|---|---|---|---|---|
| | II↑ | Unc.↓ | II↑ | Unc.↓ | II↑ | Unc.↓ | II↑ | Unc.↓ |
| RL-Ours | **25.4** (0.31) | 62.2 (0.09) | 27.6 (0.26) | 58.6 (0.22) | **32.9** (1.92) | **35.8** (3.10) | 31.7 (1.25) | 39.9 (0.64) |
| RL-Base | 24.6 (0.62) | 63.1 (0.46) | 26.7 (0.14) | 60.6 (0.85) | 31.9 (1.82) | 36.4 (2.24) | **32.2** (0.31) | **39.6** (0.59) |
| MCTS | 25.3 (0.22) | **61.9** (0.21) | **27.7** (0.49) | **58.4** (0.49) | 31.7 (0.45) | 40.8 (0.93) | 30.7 (0.50) | 42.2 (0.17) |
| Greedy | 25.2 (0.46) | 62.4 (0.45) | 27.1 (0.43) | 58.9 (0.29) | 29.8 (0.29) | 44.0 (0.66) | 29.7 (0.45) | 46.0 (0.54) |
| Coverage | 14.7 (0.34) | 75.9 (0.08) | 15.8 (0.84) | 73.4 (0.98) | 29.3 (0.60) | 44.6 (0.17) | 29.7 (0.62) | 45.4 (0.91) |

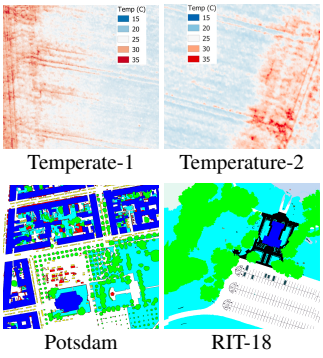


Fig. 4: Real-world datasets.

*MCTS* adaptive IPP method on missions with *Varying* hyperparameters while substantially reducing replanning runtimes at deployment. This shows we can successfully train a single adaptive IPP policy applicable and well-performing in monitoring scenarios with larger inter-mission variations in user-defined hyperparameters and terrain map representations. Fig. 3 shows simulated ground truth feature fields $F$ with paths planned based on our unified belief $\hat{F}_{I,t}$ over initially unknown non-shaded and red areas of interest $\xi_I$, derived from mission-specific map beliefs $\hat{F}_t$ with yellow indicating a high probability of interesting areas according to $\hat{F}_{I,t}$.

### C. Results on Real-World Datasets

The experiments on real-world orthomosaics are designed to show that our single unified policy trained in simulation performs similarly to state-of-the-art adaptive IPP methods on previously unseen real-world terrain datasets. We compare our map-agnostic policy (*RL-Ours*) to map-specifically designed online non-learning-based planning methods and map-specifically designed and trained policies (*RL-Base-C/D*). We consider two continuous-valued surface temperature orthomosaics of crop fields near Bonn, Germany, mapped using Gaussian processes, where high surface temperatures above 25°C are interesting. Further, we execute discrete-valued semantic monitoring of an urban area in Potsdam, Germany [11] and a rural area [12] (RIT-18), mapped using occupancy maps, where vegetation features are of interest. Orthomosaic datasets are illustrated in Fig. 4. All RL-based policies are trained in simulation as in Sec. V-B and deployed on the real-world datasets without adaptation.

Tab. II summarises the results. In line with state-of-the-art methods, our map-agnostic policy consistently outperforms traditionally used non-adaptive *Coverage* paths, showcasing the advantages of adaptive online replanning. Notably, in most scenarios, our map-agnostic policy outperforms *Greedy* planning and performs similarly to *MCTS* planning while substantially reducing replanning runtimes. Furthermore, our map-agnostic policy performs comparably to map-specifically designed and trained learning-based *RL-Base-C/D* policies. Generally, we observe an expected small performance degradation of RL-based policies compared to their performance in simulated missions due to simulation to real-world dataset gaps [13]. While our single map-agnostic policy is applied to all real-world dataset missions, each baseline requires two map-specific versions before deployment. Overall, these results highlight the advantages of our map-agnostic policy, validating its performance on unseen real-world terrain data while facilitating deployment.

TABLE III: Integration of our map-agnostic adaptive IPP formulation (*ours*) into state-of-the-art online policy search methods. Best average performances are marked in bold, second-best average performances are underlined if standard deviations in brackets overlap. Our map-agnostic formulation unifies existing adaptive IPP methods while consistently maintaining or improving performance over previous map-specific formulations for continuous- (*prev-C*) and discrete-valued (*prev-D*) terrain feature monitoring missions.

| | **Policy** | **IPP** | **Varying $\mathcal{H}$** II↑ | Unc.↓ | MLL↓ | RMSE↓ |
|---|---|---|---|---|---|---|
| Continuous | Greedy | prev-C | 25.2 (0.66) | 61.7 (0.29) | -58.0 (1.55) | 4.35 (0.24) |
| | | ours | **25.3** (0.41) | **61.3** (0.42) | **-59.0** (0.73) | **4.15** (0.16) |
| | MCTS | prev-C | 25.3 (0.24) | 61.1 (0.24) | -59.5 (0.70) | 4.27 (0.22) |
| | | ours | **27.0** (0.42) | **59.6** (0.26) | **-63.8** (0.34) | **4.00** (0.24) |
| | CMA-ES | prev-C | 21.5 (1.44) | **64.3** (2.32) | **-55.4** (5.52) | **2.59** (0.85) |
| | | ours | **21.8** (2.25) | 64.5 (1.97) | -54.1 (5.67) | 2.73 (1.26) |
| | | | II↑ | Unc.↓ | mIoU↑ | F1↑ |
| Discrete | Greedy | prev-D | 29.4 (0.17) | 45.6 (0.59) | 18.2 (0.22) | 23.2 (0.22) |
| | | ours | **30.5** (0.33) | **44.5** (0.19) | **18.6** (0.00) | **23.6** (0.05) |
| | MCTS | prev-D | 30.8 (0.21) | 41.2 (0.85) | **19.8** (0.31) | **24.7** (0.24) |
| | | ours | **31.4** (0.78) | **41.0** (0.96) | **19.8** (0.31) | **24.7** (0.31) |
| | CMA-ES | prev-D | **30.0** (1.45) | 42.4 (0.61) | 19.5 (0.42) | 24.4 (0.51) |
| | | ours | 29.6 (1.33) | **41.6** (0.37) | **19.7** (0.21) | **24.6** (0.42) |

### D. Map-Agnostic Online Adaptive IPP Policy Search

The next set of experiments aims to answer if we can easily integrate our map-agnostic adaptive IPP formulation into state-of-the-art online non-learning-based policy search methods without planning performance loss. We show that our map-agnostic adaptive IPP formulation unifies existing adaptive IPP methods while maintaining or improving performance in various terrain monitoring missions.

To showcase the general applicability of our approach, we integrate our map-agnostic adaptive IPP formulation (*ours*) with the greedy, MCTS and CMA-ES algorithms described in Sec. V-A using our state formulation in Eq. (5) and reward function in Eq. (8) for policy search. We compare it to previously used map-specific adaptive IPP formulations

for continuous- (*prev-C*) and discrete-valued terrain feature monitoring (*prev-D*) resembling our baselines in Sec. V-A. Tab. III summarises the planning performances .Our map-agnostic adaptive IPP formulation consistently performs on par with adaptive IPP formulations specifically designed for continuous- and discrete-valued monitoring missions, irrespective of the policy search method. Notably, in some scenarios, our map-agnostic formulation even improves the average planning performance of policy search algorithms. These results verify that our method successfully integrates with state-of-the-art adaptive IPP methods without requiring map-specific adaptation. This way, our approach contributes to unifying the broad family of adaptive IPP approaches.

## VI. Conclusion

We proposed a novel map-agnostic formulation of the adaptive informative path planning (IPP) problem for terrain monitoring. Our adaptive IPP formulation is generally applicable to various continuous- or discrete-valued terrain feature monitoring missions. Our main contribution is a planning state space unifying different map representations. Based on our formulation and a newly introduced reward function, we show how to train a single adaptive IPP policy for terrain monitoring missions with varying map representations and user-defined areas of interest. Our experimental results show that our single learned policy performs similarly to state-of-the-art map-specifically designed and trained non-learning- and learning-based adaptive IPP methods on simulated and real-world terrain datasets. Our map-agnostic formulation easily integrates with state-of-the-art online policy search methods for adaptive IPP while maintaining performance.

## References


[1] Y. Cao, Y. Wang, A. Vashisth, H. Fan, and G.A. Sartoretti. CAtNIPP: Context-Aware Attention-based Network for Informative Path Planning. In *Proc. of the Conf. on Robot Learning (CoRL)*, 2023.

[2] Y. Cao, T. Hou, Y. Wang, X. Yi, and G. Sartoretti. ARiADNE: A Reinforcement Learning Approach using Attention-based Deep Networks for Exploration. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2023.

[3] F. Chen, P. Szenher, Y. Huang, J. Wang, T. Shan, S. Bai, and B. Englot. Zero-Shot Reinforcement Learning on Graphs for Autonomous Exploration Under Uncertainty. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2021.

[4] T. Choi and G. Cielniak. Adaptive Selection of Informative Path Planning Strategies via Reinforcement Learning. In *Proc. of the Europ. Conf. on Mobile Robotics (ECMR)*, 2021.

[5] S. Choudhury, N. Gruver, and M.J. Kochenderfer. Adaptive Informative Path Planning with Multimodal Sensing. In *Intl. Conf. on Automated Planning and Scheduling (ICAPS)*, 2020.

[6] L. Espeholt, H. Soyer, R. Munos, K. Simonyan, V. Mnih, T. Ward, Y. Doron, V. Firoiu, T. Harley, I. Dunning, et al. Impala: Scalable distributed deep-rl with importance weighted actor-learner architectures. In *Proc. of the Intl. Conf. on Machine Learning (ICML)*, 2018.

[7] E. Galceran and M. Carreras. A Survey on Coverage Path Planning for Robotics. *Journal on Robotics and Autonomous Systems (RAS)*, 61(12):1258–1276, 2013.

[8] G. Hitz, E. Galceran, M.E. Garneau, F. Pomerleau, and R. Siegwart. Adaptive continuous-space informative path planning for online environmental monitoring. *Journal of Field Robotics (JFR)*, 34(8):1427–1449, 2017.

[9] G.A. Hollinger and G.S. Sukhatme. Sampling-based robotic information gathering algorithms. *Intl. Journal of Robotics Research (IJRR)*, 33(9):1271–1287, 2014.

[10] G.A. Hollinger, B. Englot, F.S. Hover, U. Mitra, and G.S. Sukhatme. Active planning for underwater inspection and the benefit of adaptivity. *Intl. Journal of Robotics Research (IJRR)*, 32(1):3–18, 2013.

[11] ISPRS. 2D Semantic Labeling Contest, 2018.

[12] R. Kemker, C. Salvaggio, and C. Kanan. Algorithms for Semantic Segmentation of Multispectral Remote Sensing Imagery Using Deep Learning. *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)*, 145:60–77, 2018.

[13] R. Kirk, A. Zhang, E. Grefenstette, and T. Rocktäschel. A survey of zero-shot generalisation in deep reinforcement learning. *Journal of Artificial Intelligence Research (JAIR)*, 76:201–264, 2023.

[14] Z. Liu, M. Deshpande, X. Qi, D. Zhao, R. Madhivanan, and A. Sen. Learning to Explore (L2E): Deep Reinforcement Learning-based Autonomous Exploration for Household Robot. In *Proc. of the RSS Workshop on Robot Representations for Scene Understanding, Reasoning, and Planning*, 2023.

[15] I. Lluvia, E. Lazkano, and A. Ansuategi. Active Mapping and Robot Exploration: A Survey. *Sensors*, 21(7), 2021.

[16] R. Marchant and F. Ramos. Bayesian Optimisation For Informative Continuous Path Planning. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2014.

[17] A. Meliou, A. Krause, C. Guestrin, and J.M. Hellerstein. Nonmyopic informative path planning in spatio-temporal models. In *Proc. of the Conf. on Advancements of Artificial Intelligence (AAAI)*, 2007.

[18] F. Niroui, K. Zhang, Z. Kashino, and G. Nejat. Deep Reinforcement Learning Robot for Search and Rescue Applications: Exploration in Unknown Cluttered Environments. *IEEE Robotics and Automation Letters (RA-L)*, 4(2):610–617, 2019.

[19] J. Ott, E. Balaban, and M.J. Kochenderfer. Sequential Bayesian optimization for adaptive informative path planning with multimodal sensing. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2023.

[20] J. Ott, M.J. Kochenderfer, and S. Boyd. Approximate Sequential Optimization for Informative Path Planning. *Journal on Robotics and Autonomous Systems (RAS)*, 182:104814, 2024.

[21] M. Popović, J. Ott, J. Rückin, and M.J. Kochenderfer. Learning-based methods for adaptive informative path planning. *Journal on Robotics and Autonomous Systems (RAS)*, 179:104727, 2024.

[22] M. Popović, T. Vidal-Calleja, G. Hitz, J.J. Chung, I. Sa, R. Siegwart, and J. Nieto. An informative path planning framework for UAV-based terrain monitoring. *Autonomous Robots*, 44:889–911, 2020.

[23] I.M. Rayas Fernández, C.E. Denniston, D.A. Caron, and G.S. Sukhatme. Informative Path Planning to Estimate Quantiles for Environmental Analysis. *IEEE Robotics and Automation Letters (RA-L)*, 7(4):10280–10287, 2022.

[24] J. Rückin, L. Jin, and M. Popović. Adaptive Informative Path Planning Using Deep Reinforcement Learning for UAV-based Active Sensing. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2022.

[25] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[26] A. Singh, A. Krause, and W.J. Kaiser. Nonmyopic Adaptive Informative Path Planning for Multiple Robots. In *Proc. of the Intl. Conf. on Artificial Intelligence (IJCAI)*, 2009.

[27] C.S. Tan, R. Mohd-Mokhtar, and M.R. Arshad. A Comprehensive Review of Coverage Path Planning in Robotics Using Classical and Heuristic Algorithms. *IEEE Access*, 9:119310–119342, 2021.

[28] A. Vashisth, J. Rückin, F. Magistri, C. Stachniss, and M. Popović. Deep Reinforcement Learning with Dynamic Graphs for Adaptive Informative Path Planning. *IEEE Robotics and Automation Letters (RA-L)*, 9(9):7747–7754, 2024.

[29] A. Viseras, M. Meissner, and J. Marchal. Wildfire Front Monitoring with Multiple UAVs using Deep Q-Learning. *IEEE Access*, pages 1–1, 2021.

[30] K.C.T. Vivaldini, T.H. Martinelli, V.C. Guizilini, J.R. Souza, M.D. Oliviera, F.T. Ramos, and D.F. Wolf. UAV route planning for active disease classification. *Autonomous Robots*, 43:1137–1153, 2019.

[31] Y. Wei and R. Zheng. Informative path planning for mobile sensing with reinforcement learning. In *Proc. of the IEEE Conf. on Computer Communications*, 2020.

[32] J. Westheider, J. Rückin, and M. Popović. Multi-UAV Adaptive Path Planning Using Deep Reinforcement Learning. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2023.

[33] P. Yang, Y. Liu, S. Koga, A. Ashgharivaskasi, and N. Atanasov. Learning Continuous Control Policies for Information-Theoretic Active Perception. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2023.