

Adaptive Path Planning for UAVs for Multi-Resolution Semantic Segmentation

Felix Stache* Jonas Westheider* Federico Magistri Cyrill Stachniss Marija Popović⁺

ARTICLE INFO

Keywords:

unmanned aerial vehicles
semantic segmentation
planning
terrain monitoring

ABSTRACT

Efficient data collection methods play a major role in helping us better understand the Earth and its ecosystems. In many applications, the usage of unmanned aerial vehicles (UAVs) for monitoring and remote sensing is rapidly gaining momentum due to their high mobility, low cost, and flexible deployment. A key challenge is planning missions to maximize the value of acquired data in large environments given flight time limitations. This is, for example, relevant for monitoring agricultural fields. This paper addresses the problem of adaptive path planning for accurate semantic segmentation of using UAVs. We propose an online planning algorithm which adapts the UAV paths to obtain high-resolution semantic segmentations necessary in areas with fine details as they are detected in incoming images. This enables us to perform close inspections at low altitudes only where required, without wasting energy on exhaustive mapping at maximum image resolution. A key feature of our approach is a new accuracy model for deep learning-based architectures that captures the relationship between UAV altitude and semantic segmentation accuracy. We evaluate our approach on different domains using real-world data, proving the efficacy and generability of our solution.

1. Introduction

Remote sensing and monitoring methods provide abundant data for ecology and environment research in a broad range of applications. However, the applicability of conventional remote sensing methodologies in large-scale environments is limited when both fast and high-quality data collection is required. Unmanned aerial vehicles (UAVs) are experiencing a rapid uptake in a variety of aerial monitoring applications, including search and rescue [16], wildlife conservation [15, 9, 11], industrial inspection [6, 33], and precision agriculture [24, 26, 38, 20]. Compared to traditional data acquisition methods, such as manual or static sampling procedures [15], they offer a more flexible and easier to execute way to monitor areas at high spatial and temporal resolutions [15, 33]. In recent years, the advent of deep learning (DL) has unlocked their potential for image-based remote sensing, enabling flexible, low-cost data collection and processing [6]. However, a key challenge is planning paths to efficiently gather the most useful data in large environments, while accounting for the constraints of physical platforms, e.g. on fuel/energy, as well as the characteristics of the on-board sensor and DL model used for data processing.

This paper examines the problem of DL-based semantic segmentation using UAVs. Specifically, we investigate how

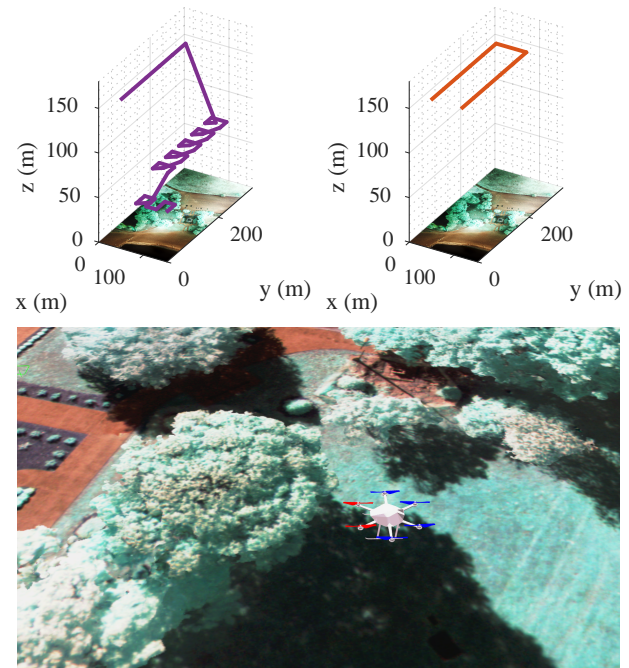


Figure 1: A comparison of our proposed adaptive path planning strategy (top-left) against lawnmower coverage planning (top-right) for UAV-based field segmentation, evaluated using real-world data from the RIT-18 dataset [13] (bottom). By allowing the paths to change online, our approach enables selecting high-resolution (low-altitude) imagery in areas with more semantic detail, enabling higher-accuracy, fine-grained segmentation in these regions.

semantic information can be exploited for intelligent path planning during a mission, i.e., online. Our problem setup considers a UAV flying above a 2D terrain and taking images of it using a downwards-facing camera, as depicted in Fig. 1. The goal is to adaptively select the next sensing locations for the UAV above the terrain to maximize the

* Authors with equal contribution.

⁺Corresponding: Marija Popović, Niebuhrstraße 1A, 53113 Bonn, Germany. mpopovic@uni-bonn.de.

All authors are with the University of Bonn, Germany. This work has partially been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy, EXC-2070 – 390732324 – PhenoRob and by the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support programme under funding no 28DK108B20 (RegisTer).

ORCID(s):

classification accuracy of objects or areas of interest seen in images, e.g. animals on grassland or crops on a field. This enables us to perform targeted high-resolution classification only where necessary and thus maximize the value of data gathered during a mission.

Aerial data acquisition campaigns rely often on coverage-based planning to generate UAV paths at a fixed flight altitude [4, 10]. Although they are easily implemented, the main drawback of such methods is that they assume an even distribution of features in the target environment; mapping the entire area at a constant image spatial resolution governed by the altitude. Recent work has explored *informative planning* for terrain mapping, whereby the aim is to optimize an information-theoretic mapping objective subject to platform constraints. By modifying plans online, such strategies enable adapting the flight path according to the mission aim to maximize the value of collected data. This can be used for improving the geometric estimation [22] or for semantic estimation tasks [35, 5]. Several previous studies in UAV-based informative planning either consider planning at a fixed altitude [37, 31], i.e. on a 2D plane above the terrain, or apply simple heuristic predictive sensor models [24, 26, 16, 7], which limit the quality of future plans. Two open challenges are therefore: (1) reliably characterizing how the accuracy of segmented images varies with the altitude and relative scales of the objects in registered images and (2) designing strategies to incorporate such models into the planning pipeline for improved targeted data collection efficiency.

The contribution of this paper is a new adaptive planning algorithm that directly tackles the altitude dependency of the DL semantic segmentation model using UAV-based imagery. First, our approach leverages prior labeled terrain data to empirically determine how classification accuracy varies with altitude; we train a deep neural network with images obtained at different altitudes that we use to initialize our planning strategy. Based on this analysis, we develop a *decision function* using Gaussian Process (GP) regression that governs the decision-making strategy. This function is first initialized on a training scenario and then updated online on a spatially separate testing scenario during a mission as new images are received.

For replanning, the UAV path is chosen according to the decision function and segmented images to obtain higher classification accuracy in more semantically detailed or interesting areas. This allows us to gather more accurate data in targeted areas without relying on a heuristic sensor model for informative planning.

This article corresponds to an extension of the authors' preliminary conference work [35]. We consolidate our previous contribution in adaptive planning with additional explanations and experimental results. A major difference is that the journal version formulates our methods in a general way, rendering them directly applicable to *any* UAV-based semantic segmentation scenario, e.g. search and rescue [16], urban scene analysis, wetland assessment, in

addition to precision agriculture and lake monitoring, which are studied as motivating applications in our experiments.

2. Related Work

There is an emerging body of literature addressing online mission planning for UAV-based remote sensing. In this section, we briefly overview recent work related to semantic segmentation of aerial images and adaptive path planning approaches for efficient data acquisition using resource-constrained platforms.

2.1. Semantic Segmentation Using Aerial Imagery

The goal of semantic segmentation is to assign a pre-determined class label to each pixel of an image. State-of-the-art approaches are predominantly based on fully convolutional neural networks (CNNs) due to their rich feature representation and end-to-end training capabilities, which generally allow for superior performance compared against handcrafted vision pipelines [6]. In remote sensing, CNNs have been successfully applied to aerial image datasets in various scenarios, e.g. for crop/weed segmentation in precision agriculture [30, 31, 20], tracking and infrastructure inspection [19], urban scene analysis [14, 1], wildlife detection [9], among others.

In the past few years, technological advancements have enabled computationally efficient segmentation on board small UAVs with constraints on size, weight, and power. Nguyen et al. [19] introduced MAVNet, a light-weight network designed for real-time aerial surveillance and inspection. Sa et al. [30] and Deng et al. [8] proposed CNN architectures to segment vegetation for smart farming using similar platforms. Recently, Bultmann et al. [3] designed a UAV system for real-time semantic fusion using multiple sensor modalities. Our work examines a problem setup similar to these studies: we adopt the light-weight ERFNet architecture, introduced by Romera et al. [29] for semantic segmentation of terrain based on aerial images obtained from a downwards-facing camera.

A fundamental difference with respect to the works above is that, instead of flying predetermined paths for data collection, our focus is on adaptive planning. We propose to exploit accurate real-time segmentation capabilities to modify the flight plan online to achieve targeted data acquisition. Specifically, our goal is to localize areas of interest and finer detail (e.g. high vegetation cover in a field or victims in a disaster site) online and steer the robot for adaptive, high-accuracy mapping in these regions.

2.2. Multi-Resolution Monitoring

An important trade-off in aerial imaging arises from the fact that the same point on a field can be observed from different altitudes. As a result, image spatial resolution degrades with increasing ground area coverage, i.e. the closer a UAV flies to the ground plane, the greater the level of image detail, but the smaller the observed area and hence the higher the flight time required to completely cover a field of fixed size. Peña et al. [23] established that

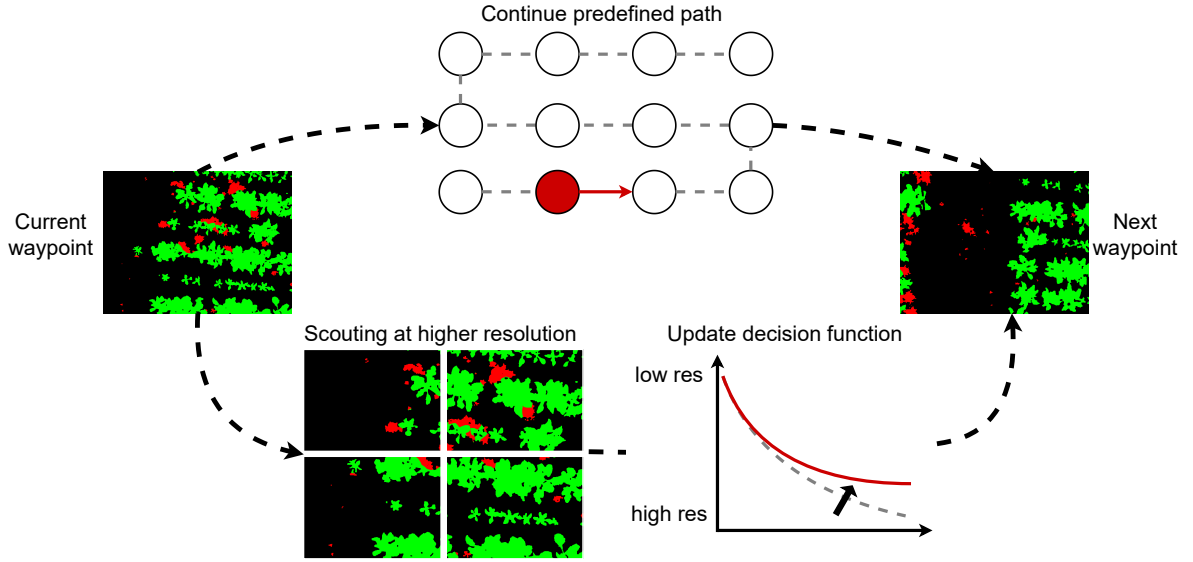


Figure 2: Overview of our adaptive planning approach. Each time one area of the field is segmented, we decide whether the UAV should follow its predefined path ('Next waypoint') or scout the same region at a lower altitude to obtain higher-resolution images here ('Scouting at higher resolution'). In the second case, we update the decision-making strategy ('Update decision function') by comparing the segmentation results of the re-observed regions at different altitudes.

there are optimal altitudes for monitoring plants based on their size. Duporge et al. [9] presented similar findings in wildlife monitoring applications. These studies motivate our approach for adaptively modifying the flight altitude during a mission based on the image content.

Relatively limited research has addressed the altitude-resolution trade-off in the contexts of semantic segmentation and robotic motion planning. Several works [1, 20] investigated network architectures with variable-size kernels that are robust to flight altitude changes. However, such methods have not yet been studied in adaptive decision-making scenarios where the physical robot constraints during data collection are taken into account.

Various methods have been proposed to address planning with multi-resolution sensors. For 3D mapping with image-based semantic information, Dang et al. [7] employed an interesting method for weighting distance measurements according to their resolution. This sensor model is used to guide exploration planning in unknown environments based on the current map state. Sadat et al. [32] proposed an adaptive coverage-based strategy that assumes sensor accuracy increases with altitude. Other studies [38, 31] only considered fixed-altitude mission planning in terrain monitoring problems; thereby neglecting the altitude dependency of the camera. We follow previous approaches that empirically assess the effects of multi-resolution observations for trained models [16, 25, 27, 11]. In contrast to these works, which derive the sensor model for planning offline, i.e. before a mission, our contribution is a decision function that supports *online* updates based on incoming images for more reliable predictive planning performance.

2.3. Adaptive Path Planning

Adaptive algorithms for active sensing allow an agent to replan online as measurements are collected during a mission to focus on application-specific interests. Several works have successfully incorporated adaptivity requirements within informative path planning problems. Here, the objective is to minimize uncertainty in target areas as quickly as possible, e.g. for exploration [36, 7], underwater surface inspection [12], target search [16, 32, 34, 11], and environmental sensing [34, 25]. These problem setups differ from ours in several ways. First, they consider a probabilistic map to represent the entire environment, using a sensor model to update the map with new uncertain measurements. In contrast, our approach directly exploits the accuracy in semantic segmentation to drive the next actions in adaptive planning. This circumvents the computational expenses of storing and updating a global map. Second, they consider a predefined, i.e. non-adaptive, sensor model, whereas ours is adapted online using a GP according to the behavior of the semantic segmentation model. The usage of GPs for path planning has already been exploited by Nardi et al. [18]; in contrast to their method, we additionally use online segmentation to adapt the GP online.

Very few works have considered planning based on semantic information. Oßwald et al. [21] proposed to explore a scene exploiting background information given by a user. Bartolomei et al. [2] introduced a perception-aware planner for UAV pose tracking. Although, like us, they exploit semantics to guide next UAV actions, their goal is to triangulate high-quality sparse landmarks whereas we aim to obtain accurate pixel-wise semantic segmentation in dense images. Dang et al. [7] and Meera et al. [16] studied informative planning for active target search using object detection networks with distance-based uncertainties.

Ghods et al. [11] explored a similar problem setup with multiple search agents. Most similar to our approach is that of Popović et al. [25], which adaptively plans the 3D path of a UAV for terrain monitoring based on an empirical performance analysis of a SegNet-based architecture at different altitudes [30]. A key difference is that our decision function, representing the network accuracy, is not static. Instead, we allow it to change online and thus adapt to new unseen environments. Moreover, for path planning, we present a general approach applicable for problems with different numbers of semantic labels.

3. Our Adaptive Path Planning

Our problem setup considers a UAV surveying a flat field of known size using a downwards-facing camera and subject to flight time constraints. The goal is to maximize the accuracy in the semantic segmentation of RGB images taken by the camera. We propose a data-driven approach that uses information from incoming images to adapt an initial predefined UAV flight path online. The main idea behind our approach is to guide the UAV to take high-resolution images for fine-grained segmentation at lower altitudes (higher spatial resolutions) only in areas where high semantic detail is desired.

Fig. 2 shows an overview of our planning strategy. We first divide the target field into non-overlapping regions and, for each, associate a waypoint in the 3D space above the field from which the camera footprint of the UAV camera covers the entire area. From these waypoints, we then define a lawnmower coverage path that we use to bootstrap the adaptive strategy. Our planning strategy consists of two main steps. First, at each waypoint along the lawnmower path, we use a deep CNN to assign a semantic label to each pixel in the region observed in the image. Second, based on the segmentation output, we decide whether the current region contains enough semantic value for more detailed re-observation at a higher image resolution, i.e., lower UAV altitude; otherwise, the UAV continues its pre-determined coverage path. The replanning procedure is repeated for each waypoint on the original lawnmower path.

A key aspect of our approach is a new data-driven decision function which enables the UAV to select a new altitude for higher-resolution images only if they are needed. The decision function captures the relative pixel-wise ratio of semantic labels of interest in an image, allowing us to judge whether an area contains high semantic value and thus needs closer inspection. This function is updated adaptively during the mission by comparing the segmentation results of the current region at the different altitudes. In this way, we can precisely capture the relationship between image resolution (altitude) and segmentation accuracy when planning new paths.

In the following sub-sections, we first describe the CNN for semantic segmentation (Sec. 3.1) before detailing our path planning strategy, which consists of offline planning (Sec. 3.2.1 and Sec. 3.2.2) and online path adaptation (Sec. 3.2.3).

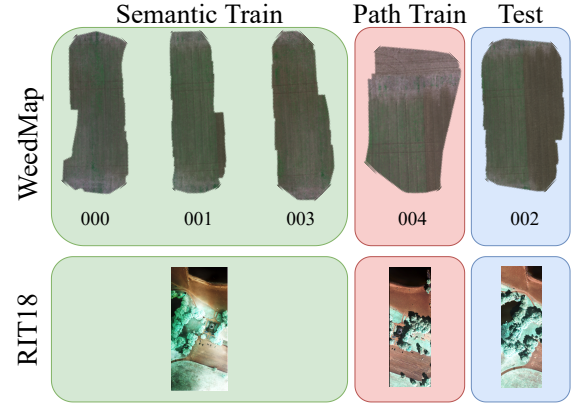


Figure 3: Our experimental setup using the WeedMap [31] and RIT-18 [13] datasets. Green, red, and blue indicate the fields used to train a CNN for semantic segmentation, initialize the planning strategy, and for evaluation. For an extensive evaluation of our approach, we swap the roles of the fields so that we test our algorithm on each field once.

3.1. Semantic Segmentation

In this work, we consider the semantic segmentation of RGB images not only as of the final mission goal but also as the tool within our planning algorithm used to define adaptive paths for re-observing given regions of the field. Each time the UAV reaches a waypoint, we perform pixel-wise semantic segmentation in the current view to assign a class label to each pixel from the set $C = \{l_1, l_2, \dots, l_C\}$, where C is the number of classes. Specifically, our proposed approach leverages the ERFNet [29] architecture provided by the Bonnetal framework [17] that allows for real-time inference. We train this CNN on RGB images collected at different altitudes to allow it to generalize across possible altitudes without the need for retraining. If the same region is observed by the camera from different altitudes, we preserve the results obtained with the highest resolution, assuming that higher-resolution images yield greater segmentation accuracy.

3.2. Path Planning

Given a trained CNN model, our path planning algorithm can be divided in three parts. First, we define a lawnmower strategy to cover the entire region of interest. Second, we initialize a decision function, based on the data obtained in previous flights on a spatially disjoint region, which serves as the starting point for our online planner. Third, while flying above the region of interest, we update the decision function as soon as new data is available.

3.2.1. Initial Strategy

In the first replanning step, the initial UAV flight path is calculated at a fixed altitude above the target field based on a standard zig-zag lawnmower strategy [10]. Such a path enables covering the field efficiently assuming no prior knowledge about it is available. In Sec. 3.2.3, we adapt this initial path according to the non-uniform distribution

of features of interest to improve semantic segmentation performance.

For a desired region of interest, we define a lawnmower path based on a series of waypoints. A waypoint is defined as a position \mathbf{w}_i in the 3D UAV workspace above the field where: (i) the UAV camera footprint does not overlap the footprints of any other waypoint; (ii) the UAV performs the semantic segmentation of its current field of view; (iii) the UAV decides to revise its path or to execute the path as previously determined; and (iv) we impose zero velocity and zero acceleration.

The initial flight path is calculated in form of fixed waypoints at the highest altitude $\mathbf{W}^{h_{\max}} = \{\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_n\}$, which is empirically set. If necessary, we modify this coarse plan by inserting further waypoints based on the new camera imagery as it arrives. At each waypoint \mathbf{w}_i , the UAV decides either to follow the pre-computed coverage path, i.e. moving to \mathbf{w}_{i+1} , or to inspect the current region more closely at a lower altitude. In the second case, we define a second series of waypoints, $\mathbf{W}^{h'} = \{\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_n\}$, at the desired altitude, h' , that will be inserted before $\mathbf{w}_{i+1} \in \mathbf{W}^{h_{\max}}$ so that the resulting path, at the desired altitude, is a lawnmower strategy covering the camera footprint from $\mathbf{w}_i \in \mathbf{W}^{h_{\max}}$.

3.2.2. Decision Function Initialization

We develop a decision function that takes a given waypoint as input and outputs the next waypoint, either $\mathbf{w}_{i+1} \in \mathbf{W}^{h_{\max}}$ or $\mathbf{w}_0 \in \mathbf{W}^{h'}$, given the semantic segmentation result. In the second case, where there is an altitude change, our decision function also outputs the value of the desired altitude h' . To do this, we start by defining a subset of class labels $\mathcal{L} = \{l_1, l_2, \dots, l_L\}$, $\mathcal{L} \subseteq \mathcal{C}$ considered as being interesting for more detailed semantic analysis. For a segmented image, we compute the number of pixels belonging to these labels as a fraction of the total number of pixels:

$$\sigma = \frac{\sum_{l \in \mathcal{L}} P_l}{P_{\text{tot}}}, \quad (1)$$

where P_{tot} is the total number of pixels in the image and p_l is the total number of pixels classified as the labels in \mathcal{L} . The *semantic ratio* σ gives us a way to infer how valuable it is to spend time on the current region of the field. It captures the intuition that higher values of this ratio indicate more possible misclassifications among the class labels of interest. To quantify such a relationship, we let the UAV run on a separate field, where we have access to ground truth data, segmenting regions of the fields with different altitudes. Segmenting the same region of the field with different altitudes provides two pieces of information that we use to shape the decision function. On one hand, we have the difference between the altitudes from which we segment the field, $\Delta h = h_{\max} - h'$. On the other hand, we have the difference between the semantic ratio σ in the predicted segmentation, $\Delta \sigma = \sigma_{h_{\max}} - \sigma_{h'}$. At the same time, we can compare σ to the accuracy of the predicted segmentation by computing the mean intersection over union (mIoU). The

mIoU is defined as the average over the classes \mathcal{L} of the semantic ratio between the intersection of ground truth (gt) and predicted segmentation (pred) and the union of the same quantities:

$$\text{mIoU} = \frac{1}{|\mathcal{L}|} \sum_{l \in \mathcal{L}} \frac{\text{gt}_l \cap \text{pred}_l}{\text{gt}_l \cup \text{pred}_l}. \quad (2)$$

Again, we define the difference between mIoUs at different altitudes as $\Delta \text{mIoU} = \text{mIoU}_{h_{\max}} - \text{mIoU}_{h'}$.

Our method thus considers two sets of observations, representing the relationships between the semantic ratio σ and UAV altitude h (called \mathcal{O}) and between the ratio σ and mIoU (called \mathcal{I}) as follows:

$$\mathcal{O} = \begin{bmatrix} \Delta \sigma_0 & \Delta h_0 \\ \Delta \sigma_1 & \Delta h_1 \\ \vdots & \vdots \\ \Delta \sigma_n & \Delta h_n \end{bmatrix}, \quad \mathcal{I} = \begin{bmatrix} \Delta \sigma_0 & \Delta \text{mIoU}_0 \\ \Delta \sigma_1 & \Delta \text{mIoU}_1 \\ \vdots & \vdots \\ \Delta \sigma_n & \Delta \text{mIoU}_n \end{bmatrix}. \quad (3)$$

While both sets are initialized offline, we only update \mathcal{O} online given that \mathcal{I} requires access to ground truth data, which is clearly not available on testing fields. We fit both sets of observations using Gaussian Process (GP) regression, a nonparametric Bayesian regression approach [28]. A GP assumes a Gaussian process prior $f(x)$, which is fully defined by a mean function $m(x)$ and a covariance function $k(x_i, x_j)$:

$$f(x) \sim \mathcal{GP}(m(x), k(x_i, x_j)). \quad (4)$$

To capture environmental phenomena, a common choice is to set the mean function $m(x) = 0$ and to use the squared exponential covariance function:

$$k(x_i, x_j) = \zeta_f^2 \exp\left(-\frac{1}{2} \frac{|x_i - x_j|^2}{\ell^2}\right) + \zeta_n^2, \quad (5)$$

where $\theta = \{\ell, \zeta_f^2, \zeta_n^2\}$ are the model hyperparameters and represent respectively the length scale ℓ , the variance of the output ζ_f^2 and of the noise ζ_n^2 . Typically, the hyperparameters are learned from the training data by maximizing the log marginal likelihood. Given a set of observations y of f for the inputs \mathbf{X} (i.e. our sets \mathcal{O} , \mathcal{I}), GP regression allows for learning a predictive model of f at the query inputs \mathbf{X}_* by assuming a joint Gaussian distribution over the samples. The predictions at \mathbf{X}_* are represented by the predictive mean μ_* and variance σ_*^2 defined as:

$$\begin{aligned} \mu_* &= \mathbf{K}(\mathbf{X}_*, \mathbf{X}) \mathbf{K}_{\mathbf{X}\mathbf{X}}^{-1} y, \\ \sigma_*^2 &= \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) - \mathbf{K}(\mathbf{X}_*, \mathbf{X}) \mathbf{K}_{\mathbf{X}\mathbf{X}}^{-1} \mathbf{K}(\mathbf{X}, \mathbf{X}_*), \end{aligned} \quad (6)$$

where $\mathbf{K}_{\mathbf{X}\mathbf{X}} = \mathbf{K}(\mathbf{X}, \mathbf{X}) + \zeta_n^2 \mathbf{I}$, and $\mathbf{K}(\cdot, \cdot)$ are matrices constructed using the covariance function $k(\cdot, \cdot)$ evaluated at the training and test inputs, \mathbf{X} and \mathbf{X}_* . In the following, we will use the ground sampling distance (GSD) to identify

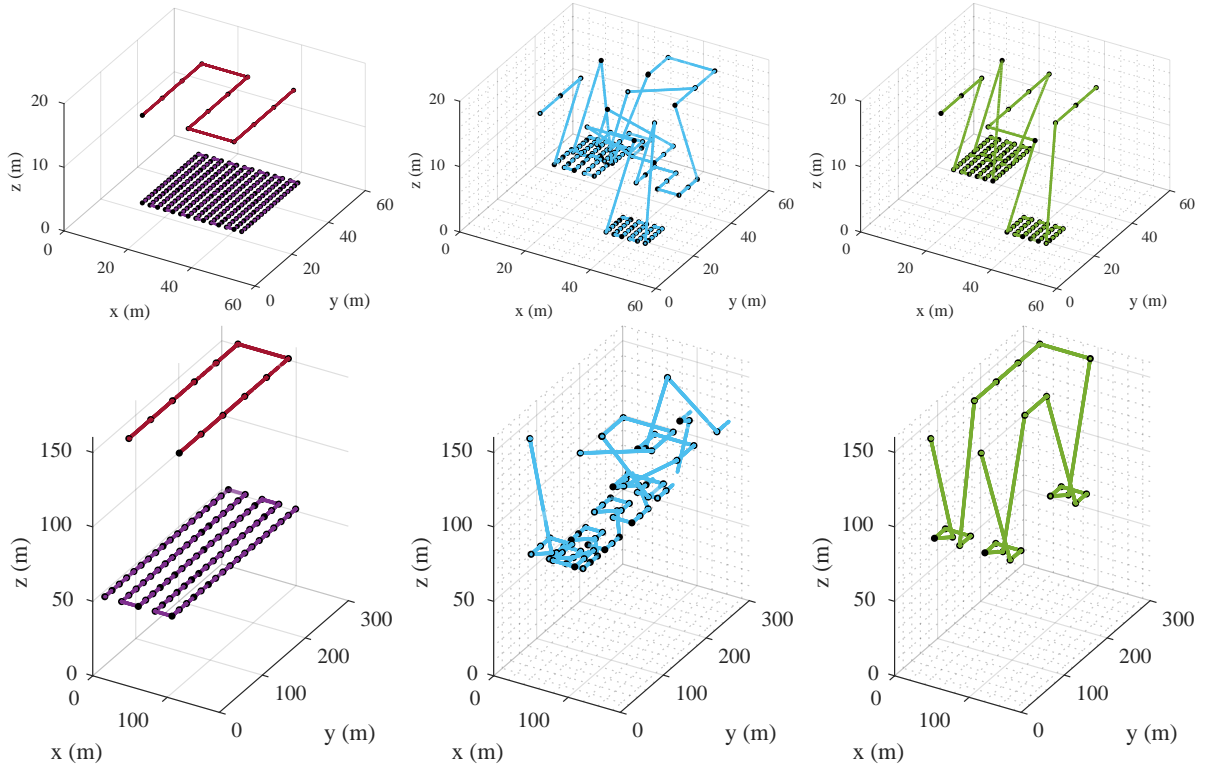


Figure 4: Visual comparison of trajectories traveled by the UAV over a field using different planning strategies. Top WeedMap, bottom RIT-18. The coverage paths (left) are restricted to fixed heights and cannot map targeted areas of interest. The linear decision function (middle) enables adaptive planning, but it is continuous with respect to altitude and leads to sudden jumps. Our adaptive approach overcomes this issue, leaving the path less often and more purposefully at selected heights for more efficient mapping. The black spheres indicate measurement points.

the image resolution (thus the UAV altitude) from which semantic segmentation is performed. The GSD is defined as: $GSD = \frac{hS_w}{fI_w}$, where h is the UAV altitude in meters, S_w is the camera sensor width in centimeters, f the focal length of the camera in centimeters and I_w is the image width in pixels.

3.2.3. Online Adaptation

To adapt the UAV behavior online to fit the differences between the testing and training fields, we update the GP defined by the set \mathcal{O} in the following way. In the testing field, each time the UAV decides to change altitude to a lower one, we compute a new pair $\Delta\sigma'$, $\Delta h'$ and re-compute the GP output as defined in Eq. (4). This procedure is repeated for each waypoint on the original lawnmower path.

4. Experimental Results

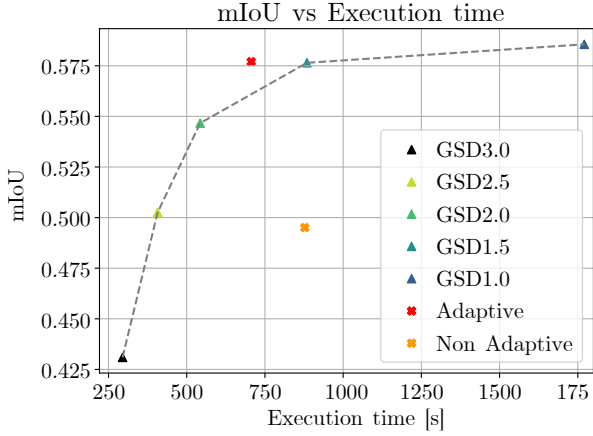
We validate our proposed algorithm for online adaptive path planning on the application of UAV-based semantic segmentation. The goal of our experiments is to demonstrate the benefits of using our adaptive strategy to maximize segmentation accuracy in missions while keeping a low execution time. Specifically, we show results to support two key claims: our online adaptive algorithm can (i) map high-interest regions with higher accuracy and (ii) improve

segmentation accuracy while keeping a low execution time with respect to the baselines described in Sec. 4.2.

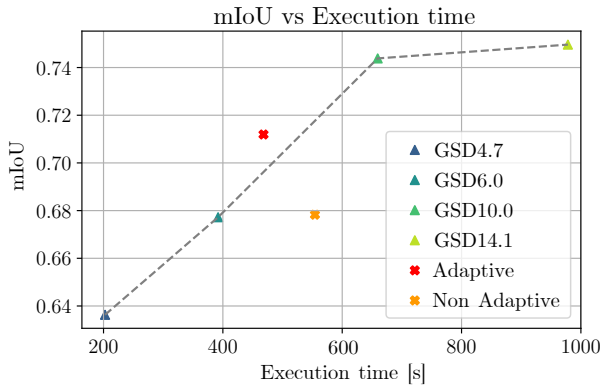
4.1. Datasets

Our approach is evaluated using two real-world datasets, WeedMap [31] and RIT-18 [13]. We consider aerial data captured from different domains to demonstrate the general applicability of our method. WeedMap consists of 8 different fields collected with two different having different channels. It also provides pixel-wise semantic segmentation labels for each of the 8 fields; the class labels present in this dataset are *soil*, *crop* and *weed*. In this study, we focus only on the 5 fields having RGB information. We split the 5 fields into training and testing sets (Fig. 3). One of the training fields is used to initialize the decision function that shapes altitude selection in the adaptive strategy, as described in Sec. 3.2.3. RIT-18 consists of labelled high-resolution multi-spectral orthomosaics obtained from remote sensing imagery, in this dataset we consider the labels for *asphalt*, *beach*, *vegetation*, *water*, and *building*. Note that, in order to reduce the complexity of semantic segmentation task for our experimental purposes, we define such labels by grouping together similar classes in the original dataset.

Fig. 3 specifies the the dataset splits studied in our experimental setup. For each experiment in the following subsections, we test our approach and the baselines described



(a) WeedMap.



(b) RIT-18.

Figure 5: Averaged results for the testing fields. The red cross lies to the left of all performances with a linear decision function, indicating performance improvement.

in Sec. 4.2 on each field once, and then report the average values for each run.

4.2. Baselines

To evaluate our proposed approach, we compare it against two main baselines.

The first one is the standard lawnmower strategy where a UAV covers the entire field at the same altitude, for this strategy we use consider five different altitudes resulting in $GSD \in \{1.0, 1.5, 2.0, 2.5, 3.0\}$ cm/px. The lawnmower strategy with a fixed GSD of 3.0 cm/px corresponds to the initial plan for our strategy described in Sec. 3.2. The second baseline is defined by *only* initializing the UAV behavior as described in Sec. 3.2.1 and without adapting the strategy online using the decision function as new segmentations arrive. We refer to this strategy as “Non Adaptive”. This benchmark allows us to study the benefit of adaptivity obtained by using our proposed approach (“Adaptive”).

4.3. Metrics

Our evaluation considers two main criteria: segmentation accuracy and mission execution time. For execution

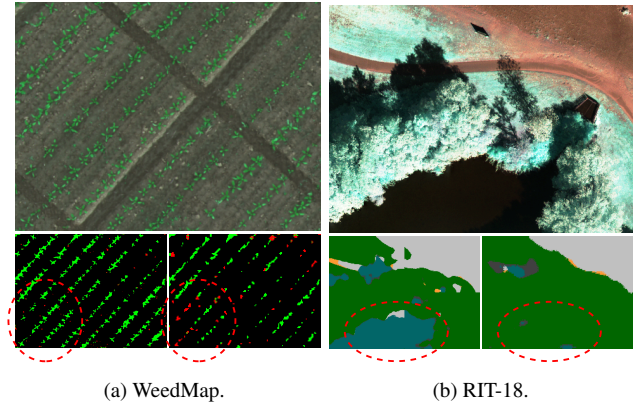


Figure 6: Qualitative field segmentation results using the proposed adaptive strategy using our decision function (bottom left) and lawnmower strategy (bottom right) for path planning. The circled details demonstrate that our adaptive planning approach enables targeted high-resolution segmentation to capture finer details at higher accuracy.

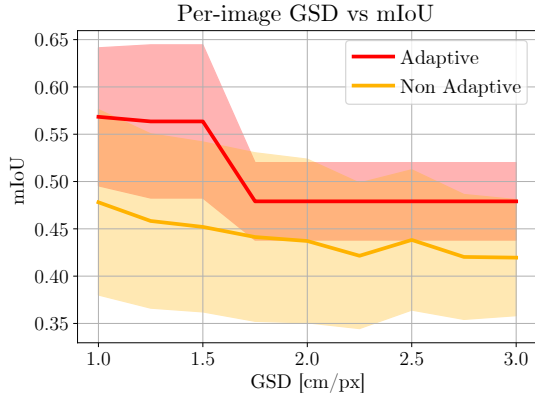
time, we compute the total time taken by the UAV to survey the whole field, including the time needed to move between waypoints, segment a new image, and plan the next path. To assess the quality of the semantic segmentation we compute the mIoU metric according to Eq. (2).

4.4. Field Segmentation Accuracy vs. Execution Time

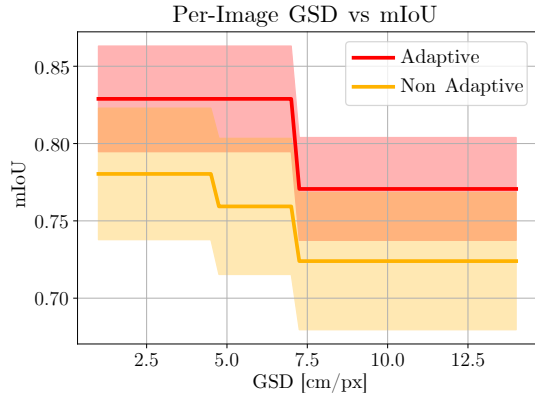
The first experiment is designed to show that our proposed strategy obtains higher accuracy when compared against the baseline methods while keeping low execution time. We show such results in Fig. 5a for WeedMap and Fig. 5b for RIT-18. For each strategy, we compute the mIoU (over the entire field) and the execution time needed by the UAV to complete its path. The adaptive strategy crosses the line defined by the lawnmower strategies at different altitudes, meaning that it can achieve better segmentation accuracy while keeping a lower execution time. The non-adaptive strategy instead lies under the curve, failing to overtake the lawnmower strategy. We plot exemplary paths results from the different strategies in Fig. 4; in the top row on the left, we show the lawnmower strategy with altitudes corresponding to GSDs of 1.0 cm/px and 3.0 cm/px, in the bottom row the lawnmower strategy correspond to GSDs of 4.7 cm/px and 14.1 cm/px. In both cases, the middle and right plots show the paths resulting from non-adaptive and adaptive strategy, respectively.

For the experiment on WeedMap, our set of targeted labels of interest comprises the two vegetation classes, $\mathcal{L} = \{crop, weed\}$. This is representative of a precision agriculture task where mission objective is to closely inspect plants. Using this subset of labels, the improvement of the mIoU is mainly given by the *crop* class while the accuracy values for the remaining classes remain stable.

For the experiment on RIT-18, we target one class at a time and plot the average results over each run, resulting in a total of 10 runs considering the 5 classes each with



(a) WeedMap.



(b) RIT-18.

Figure 7: Means and standard deviations of the per-image statistics for semantic segmentation. Our adaptive strategy leads to better performance when scouting the field at low altitudes (high GSDs).

the testing and validation sets reversed. Results from these experiments show that our approach yields better segmentation performance and lower execution time when the target class is not dominating the scene. Intuitively, the more spatially localized the class is, the more closely it can be inspected in a targeted way and hence the greater the benefit of using our adaptive multi-resolution strategy to reduce altitude only in this area. This is the case, for example, when targeting classes such as *ashpalt* or *water* in RIT-18. As qualitative examples, Fig. 6 compares the semantic masks obtained using the adaptive approach and a lawnmower strategy. The circled details illustrate situations where our proposed adaptive method produces visually more correct segmentations without loss of detail by allowing the UAV to reobserve a target area at higher resolutions.

4.5. Per-Image Segmentation Accuracy vs. Altitude

The second experiment shows the ability of our approach to achieve targeted semantic segmentation when compared to the non-adaptive strategy. At this stage, we compute mIoU for each image that contributes to the final

segmentation of the whole field. This gives us a way of evaluating the efficiency of our adaptation strategy. We then visualize the means and standard deviations. As can be seen in Fig. 7, our adaptive strategy provides higher per-image accuracies when the UAV is scouting the field at low altitudes. This entails that, with our strategy, the limited flight time resources are spent in a more efficient manner in terms of monitoring performance.

5. Conclusion

This paper presents a new approach for adaptive path planning for UAVs in general multi-resolution semantic segmentation applications. A key contribution of this paper is a new adaptive planning algorithm that directly tackles the altitude dependency of the deep learning semantic segmentation model using UAV-based imagery. Our strategy exploits the prior knowledge of a field and the new incoming segmentations to enable adaptive decision-making for mapping targeted areas of interest at higher image resolutions. Experimental results using real-world data from different domains validate that our strategy leads to high segmentation accuracy while minimizing flight time needed to cover the field. Our approach opens a direction for efficient UAV mapping, especially in applications such as precision agriculture, where certain areas of a field need to be closely inspected.

References

- [1] Avola, D., Pannone, D., 2021. MAGI: Multistream Aerial Segmentation of Ground Images with Small-Scale Drones. *Drones* 5. doi:10.3390/drones5040111.
- [2] Bartolomei, L., Teixeira, L., Chli, M., 2020. Perception-aware Path Planning for UAVs using Semantic Segmentation, in: *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*. doi:10.1109/IROS45743.2020.9341347.
- [3] Bultmann, S., Quenzel, J., Behnke, S., 2021. Real-Time Multi-Modal Semantic Fusion on Unmanned Aerial Vehicles, in: *Proc. of the Europ. Conf. on Mobile Robotics (ECMR)*.
- [4] Cabreira, T., Brisolar, L., Ferreira Jr., P.R., 2019. Survey on Coverage Path Planning with Unmanned Aerial Vehicles. *Drones* 3. doi:10.3390/drones3010004.
- [5] Carbone, C., Albani, D., Magistri, F., Ognibene, D., Stachniss, C., Kootstra, G., Nardi, D., Trianni, V., 2021. Monitoring and Mapping of Crop Fields with UAV Swarms Based on Information Gain, in: *Distributed Autonomous Robotic Systems*. doi:10.1007/978-3-030-92790-5_24.
- [6] Carrio, A., Sampedro Pérez, C., Rodríguez Ramos, A., Campoy, P., 2017. A Review of Deep Learning Methods and Applications for Unmanned Aerial Vehicles. *Journal of Sensors* 2017, 1–13. doi:10.1155/2017/3296874.
- [7] Dang, T., Papachristos, C., Alexis, K., 2018. Autonomous exploration and simultaneous object search using aerial robots, in: *Proc. of the IEEE Aerospace Conference*. doi:10.1109/AERO.2018.8396632.
- [8] Deng, J., Zhong, Z., Huang, H., Lan, Y., Han, Y., Zhang, Y., 2020. Lightweight Semantic Segmentation Network for Real-Time Weed Mapping Using Unmanned Aerial Vehicles. *Applied Sciences* 10. doi:10.3390/app10207132.

- [9] Duporge, I., Spiegel, M.P., Thomson, E.R., Chapman, T., Lamberth, C., Pond, C., Macdonald, D.W., Wang, T., Klinck, H., 2021. Determination of optimal flight altitude to minimise acoustic drone disturbance to wildlife using species audiograms. *Methods in Ecology and Evolution* 12, 2196–2207. doi:10.1111/2041-210X.13691.
- [10] Galceran, E., Carreras, M., 2013. A survey on coverage path planning for robotics. *Robotics and Autonomous Systems* 61, 1258–1276. doi:10.1016/j.robot.2013.09.004.
- [11] Ghods, R., Durkin, W.J., Schneider, J.G., 2021. Multi-Agent Active Search using Realistic Depth-Aware Noise Model, in: *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, pp. 9101–9108. doi:10.1109/ICRA48506.2021.9561598.
- [12] Hollinger, G.A., Englot, B., Hover, F.S., Mitra, U., Sukhatme, G.S., 2013. Active planning for underwater inspection and the benefit of adaptivity. *Intl. Journal of Robotics Research (IJRR)* 32, 3–18. doi:10.1177/0278364912467485.
- [13] Kemker, R., Salvaggio, C., Kanan, C., 2018. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing* URL: <http://www.sciencedirect.com/science/article/pii/S0924271618301229>, doi:https://doi.org/10.1016/j.isprsjprs.2018.04.014.
- [14] Lyu, Y., Vosselman, G., Xia, G.S., Yilmaz, A., Yang, M.Y., 2020. UAVid: A semantic segmentation dataset for UAV imagery. *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)* 165, 108 – 119. doi:10.1016/j.isprsjprs.2020.05.009.
- [15] Manfreda, S., McCabe, M.F., Miller, P.E., Lucas, R., Pajuelo Madrigal, V., Mallinis, G., Ben Dor, E., Helman, D., Estes, L., Ciraolo, G., Müllerová, J., Tauro, F., De Lima, M.I., De Lima, J.L.M.P., Maltese, A., Frances, F., Caylor, K., Kohv, M., Perks, M., Ruiz-Pérez, G., Su, Z., Vico, G., Toth, B., 2018. On the Use of Unmanned Aerial Systems for Environmental Monitoring. *Remote Sensing* 10. doi:10.3390/rs10040641.
- [16] Meera, A.A., Popović, M., Millane, A., Siegwart, R., 2019. Obstacle-aware Adaptive Informative Path Planning for UAV-based Target Search, in: *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, pp. 718–724. doi:10.1109/ICRA.2019.8794345.
- [17] Milioto, A., Mandtler, L., Stachniss, C., 2019. Fast Instance and Semantic Segmentation Exploiting Local Connectivity, Metric Learning, and One-Shot Detection for Robotics, in: *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*. doi:10.1109/ICRA.2019.8793593.
- [18] Nardi, L., Stachniss, C., 2019. Actively Improving Robot Navigation On Different Terrains Using Gaussian Process Mixture Models, in: *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*. doi:10.1109/ICRA.2019.8794079.
- [19] Nguyen, T., Shivakumar, S.S., Miller, I.D., Keller, J., Lee, E.S., Zhou, A., Özasan, T., Loianno, G., Harwood, J.H., Wozencraft, J., Taylor, C.J., Kumar, V., 2019. MAVNet: An Effective Semantic Segmentation Micro-Network for MAV-Based Tasks. *IEEE Robotics and Automation Letters (RA-L)* 4, 3908–3915. doi:10.1109/LRA.2019.2928734.
- [20] Ocer, N.E., Kaplan, G., Erdem, F., Matci, D.K., Avdan, U., 2020. Tree extraction from multi-scale UAV images using Mask R-CNN with FPN. *Remote Sensing Letters* 11, 847–856. doi:10.1080/2150704X.2020.1784491.
- [21] Osswald, S., Bennewitz, M., Burgard, W., Stachniss, C., 2016. Speeding-Up Robot Exploration by Exploiting Background Information. *IEEE Robotics and Automation Letters (RA-L)* URL: <http://www.ipb.uni-bonn.de/wp-content/papercite-data/pdf/osswald16ral.pdf>, doi:10.1109/LRA.2016.2520560.
- [22] Palazzolo, E., Stachniss, C., 2018. Effective Exploration for MAVs Based on the Expected Information Gain. *Drones* 2. doi:10.3390/drones2010009.
- [23] Peña, J.M., Torres-Sánchez, J., Serrano-Pérez, A., De Castro, A.I., López-Granados, F., 2015. Quantifying Efficacy and Limits of Unmanned Aerial Vehicle (UAV) Technology for Weed Seedling Detection as Affected by Sensor Resolution. *Sensors* 15. doi:10.3390/s150305609.
- [24] Popović, M., Hitz, G., Nieto, J., Sa, I., Siegwart, R., Galceran, E., 2017a. Online Informative Path Planning for Active Classification Using UAVs, in: *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*. doi:10.1109/ICRA.2017.7989676.
- [25] Popović, M., Vidal-Calleja, T., Hitz, G., Chung, J.J., Sa, I., Siegwart, R., Nieto, J., 2020. An informative path planning framework for UAV-based terrain monitoring. *Autonomous Robots* 44, 889–911. doi:10.1007/s10514-020-09903-2.
- [26] Popović, M., Vidal-Calleja, T., Hitz, G., Sa, I., Siegwart, R., Nieto, J., 2017b. Multiresolution Mapping and Informative Path Planning for UAV-based Terrain Monitoring, in: *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*. doi:10.1109/IROS.2017.8202317.
- [27] Qingqing, L., Taipalmaa, J., Queralta, J.P., Gia, T.N., Gabbouj, M., Tenhunen, H., Raitoharju, J., Westerlund, T., 2020. Towards Active Vision with UAVs in Marine Search and Rescue: Analyzing Human Detection at Variable Altitudes, in: *Proc. of the IEEE International Symposium on Safety, Security, and Rescue Robotics*, pp. 65–70. doi:10.1109/SSRR50563.2020.9292596.
- [28] Rasmussen, C.E., Williams, C.K.I., 2006. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA. doi:10.1007/978-3-540-28650-9_4.
- [29] Romera, E., Alvarez, J.M., Bergasa, L.M., Arroyo, R., 2017. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems* 19, 263–272. doi:10.1109/TITS.2017.2750080.
- [30] Sa, I., Chen, Z., Popović, M., Khanna, R., Liebisch, F., Nieto, J., Siegwart, R., 2018a. weedNet: Dense Semantic Weed Classification Using Multispectral Images and MAV for Smart Farming. *IEEE Robotics and Automation Letters (RA-L)* 3, 588–595. doi:10.1109/LRA.2017.2774979.
- [31] Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebisch, F., Nieto, J., Stachniss, C., Walter, A., Siegwart, R., 2018b. Weedmap: A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. *Remote Sensing* 10. doi:10.3390/rs10091423.
- [32] Sadat, S.A., Wawerla, J., Vaughan, R., 2015. Fractal trajectories for online non-uniform aerial coverage, in: *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, pp. 2971–2976. doi:10.1109/ICRA.2015.7139606.
- [33] Shakhathreh, H., Sawalmeh, A.H., Al-Fuqaha, A., Dou, Z., Almaita, E., Khalil, I., Othman, N.S., Khreishah, A., Guizani, M., 2019. Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges. *IEEE Access* 7, 48572–48634. doi:10.1109/ACCESS.2019.2909530.
- [34] Singh, A., Krause, A., Kaiser, W.J., 2009. Nonmyopic Adaptive Informative Path Planning for Multiple Robots, in: *Proc. of the Intl. Conf. on Artificial Intelligence (IJCAI)*, p. 1843–1850. doi:10.5555/1661445.1661741.
- [35] Stache, F., Westheider, J., Magistri, F., Popović, M., Stachniss, C., 2021. Adaptive Path Planning for UAV-based Multi-Resolution Semantic Segmentation, in: *Proc. of the Europ. Conf. on Mobile Robotics (ECMR)*. doi:10.1109/ECMR50962.2021.9568788.
- [36] Stachniss, C., Grisetti, G., Burgard, W., 2005. Information Gain-based Exploration Using Rao-Blackwellized Particle Filters, in: *Proc. of Robotics: Science and Systems (RSS)*, Cambridge, MA, USA. pp. 65–72. URL: <http://www.informatik.uni-freiburg.de/~stachnis/pdf/stachniss05rss.pdf>.
- [37] Vivaldini, K., Guizilini, V., Oliveira, M., Martinelli, T., Ramos, F., Wolf, D., 2016. Route Planning for Active Classification with UAVs, in: *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*. doi:10.1109/ICRA.2016.7487412.

- [38] Vivaldini, K.C., Martinelli, T.H., Guizilini, V.C., Souza, J.R., Oliveira, M.D., Ramos, F.T., Wolf, D.F., 2019. UAV Route Planning for Active Disease Classification. *Autonomous Robots* 43, 1137–1153. doi:10.1007/s10514-018-9790-x.