aerospace

MDPI

Article

Simultaneous Incremental Map-Prediction-Driven UAV Trajectory Planning for Unknown Environment Exploration

Jianing Tang, Guoran Jiang, Jingkai Yang and Sida Zhou *

Yunnan Key Laboratory of Unmanned Autonomous Systems, School of Electrical and Information Engineering, Yunnan Minzu University, Kunming 650500, China; 041749@ymu.edu.cn (J.T.); noname2301@ymu.edu.cn (G.J.); jingkai_yang@ymu.edu.cn (J.Y.)
* Correspondence: zhousida@sina.com

Abstract

Efficient autonomous exploration in unknown environments is a core challenge for Unmanned Aerial Vehicle (UAV) applications in unstructured settings. The primary challenges are exploration speed, coverage efficiency, and the autonomous, efficient, and obstacle-/threat-avoiding global guidance of UAV under local observational information. This paper proposes an autonomous exploration method driven by simultaneous incremental map prediction and the fusion of global frontier information to enhance the exploration efficiency of UAVs in unknown unstructured environments. Based on generative deep learning, we introduce an incremental map prediction method for 3D unstructured mountainous terrain, enabling the simultaneous acquisition of map predictions and their uncertainty estimates. Map prediction and trajectory planning are conducted concurrently: by utilizing the simultaneously predicted 3D map and its confidence (i.e., the uncertainty estimates), an overlap analysis is conducted between the flyable areas in the predicted map and the high-confidence regions. Dynamic guidance subspaces are generated by extracting global frontier points, within which shortest-time optimization is adopted for trajectory planning to maximize information gain and coverage per step. Experimental results demonstrate that compared to classical methods, our proposed approach achieves significant performance improvements in key metrics, including map coverage rate, total exploration time, and average path length.

Keywords: UAV; trajectory planning; simultaneous incremental map prediction; unknown environment exploration

Check for updates

Academic Editor: Gokhan Inalhan

Received: 12 December 2025
Revised: 27 January 2026
Accepted: 29 January 2026
Published: 30 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

In diverse tasks, the Unmanned Aerial Vehicle (UAV) not only undertakes high-resolution image acquisition and target recognition but also needs to achieve autonomous trajectory planning and high-precision mapping in unknown environments, a core capability for long-term autonomous missions. Currently, the applied research and applications of UAVs in typical structured environments (e.g., urban streets and industrial parks) are relatively mature. However, unstructured environments (e.g., forests, mining areas, and post-disaster ruins), characterized by highly undulating terrain and complex obstacles, pose significant challenges to the perception and planning systems of UAVs [1]. Additionally, such environments typically lack prior mapping information and feature disordered obstacle distribution, which significantly elevates the difficulty of real-time data fusion and environmental understanding [2]. Furthermore, dynamic obstacles and perception

blind spots in such complex terrains further compromise sensor stability and positioning accuracy [3]. Traditional localized exploration or heuristic guidance approaches struggle to balance navigational safety with global coverage. Therefore, how to enhance the exploration efficiency of UAVs in complex unknown unstructured environments is a critical issue that urgently requires addressing.

To tackle the challenge of UAV autonomous exploration in unknown unstructured environments, researchers have proposed various path planning approaches. For scenarios with extreme environmental constraints (e.g., dense obstacles and uncharted terrain), Hong et al. [4] developed an enhanced sampling-based path planning method adapted for forest environments, which improves UAV safety by penalizing paths close to obstacles and boosts efficiency by minimizing overlaps with existing trajectories via a potential field function. Complementarily, Xie et al. [5] proposed an autonomous drone exploration system (ADES) integrated with a lightweight Lite-TA-MSNet saliency prediction model and visual-inertial odometry (VIO)-based exploration approach, enabling collision avoidance and anti-repetitive exploration in real unknown environments without pre-constructed maps. For general unstructured scenarios, Cieslewski et al. [6] developed a frontier-based selection strategy suitable for high-speed flight, which significantly enhances the exploration speed of multi-rotor UAVs. Schmid et al. [7] developed a sampling-based online path planning method enabling efficient information acquisition in unknown unstructured environments. To balance exploration efficiency and environmental reconstruction accuracy, Meng et al. [8] proposed a two-stage optimization framework that enables 3D structural reconstruction of unknown unstructured 3D environments through next-view planning. Additionally, Selin et al. [9] presented an efficient autonomous exploration strategy for large-scale unknown unstructured environments, further extending the operational capabilities of UAVs in complex unstructured environments. Despite certain achievements, most existing methods still suffer from low exploration efficiency and incomplete map reconstruction in large-scale unstructured scenarios, failing to meet the requirements of time-critical applications. The core issue stems from the fact that many exploration strategies rely on localized heuristic approaches for decision making. Among these approaches, the most prominent one is the frontier-based exploration approach, whose core lies in guiding UAVs to fly toward the boundary of unknown regions to realize rapid environmental coverage [10]. To enhance global decision quality, some studies formulate information gain as an optimizable objective function, obtaining globally suboptimal paths via gradient-based optimization methods [11]. Furthermore, integrating UAVs' physical constraints and sensor field-of-view (FOV) models into the planning framework has also been demonstrated to significantly enhance flight stability and exploration efficiency [12]. However, this short-term gain-driven greedy strategy overlooks global coverage efficiency, which often results in path redundancy and unbalanced coverage of exploration regions [13]. On the other hand, to ensure flight safety in unknown unstructured environments, existing methods typically produce overly conservative flight trajectories, which inherently restrict the dynamic performance of UAVs [14]. Moreover, some algorithms exhibit high computational complexity in perception processing and trajectory replanning, with inadequate real-time responsiveness, which hinders the achievement of high-frequency updates in rapidly changing unstructured environments. Therefore, exploration planning systems are in urgent need of a design framework featuring a global-oriented perspective, high dynamic performance adaptability, and real-time response capabilities, so as to substantially improve overall exploration efficiency.

The limitations of the aforementioned studies offer clear directions for improvement in the present study. First, to tackle the issues of path redundancy and uneven exploration caused by localized heuristic strategies [10,11], this paper integrates global frontier

search, aiming to maximize global information gain at each planning step. Second, to address the issue of overly conservative trajectory generation in unknown unstructured environments—a problem that restricts the dynamic performance of UAVs—we incorporate simultaneous incremental map prediction (SIMP), and leverage the predicted flyable regions and high-confidence regions as guidance to support the generation of more aggressive yet efficient flight trajectories. Finally, to tackle the challenge of high computational load in complex unstructured environments, this framework incorporates an SIMP-driven incremental frontier update and target waypoint selection mechanism, thereby ensuring the system's real-time response capability in highly dynamic unstructured environments. The main contributions of this paper are summarized as follows:

- A SIMP-integrated trajectory planning framework with global frontier search is proposed. This framework incorporates SIMP, uncertainty estimation, and global frontier point search. By dynamically analyzing the intersection of predicted flyable regions and high-confidence regions, it achieves globally optimal trajectory planning while ensuring flight safety, and effectively addresses the issues of path redundancy and uneven exploration caused by localized heuristic strategies.
- A deep learning-based SIMP and confidence estimation model is constructed. This model can incrementally predict the terrain structure of unknown unstructured areas based on local observational information and output corresponding uncertainty maps, thereby providing reliable prior information support for global exploration decision making.
- An efficient incremental frontier exploration mechanism is implemented. Built on the frontier-based approach, this mechanism introduces predicted information gain (derived from SIMP) as guidance. Through incremental frontier updates and dynamic target waypoint selection, it significantly enhances exploration efficiency and path quality in complex unstructured environments, while addressing the high computational load issue of traditional methods.

The rest of this paper is organized as follows. Section 2 categorizes and analyzes existing methods. Section 3 elaborates on the proposed incremental map prediction approach. Section 4 presents a frontier exploration framework guided by predicted maps. Section 5 provides a comprehensive quantitative and qualitative analysis of map prediction performance and UAV exploration results. Finally, Section 6 summarizes the main contributions and conclusions of this study and outlines potential directions for future research.

2. Related Work

Over the past few decades, autonomous exploration of UAVs in unknown unstructured environments has garnered extensive research attention, and numerous approaches have been proposed. These methods can be roughly classified into three categories: frontier-based methods, sampling-based methods, and hybrid methods.

Among the three categories, frontier-based exploration methods represent the most classical research paradigm, owing to their intuitive concept and efficient implementation. The foundational work in this field was proposed by Yamauchi [10], with its core principle being to define the boundary between known and unknown regions as the "frontier" as the exploration target, and employs a greedy strategy to select the nearest frontier point for rapid environmental coverage. However, as exploration extends from 2D to complex unstructured 3D environments, traditional frontier-based methods confront significant challenges in frontier extraction and decision-making effectiveness. To tackle this challenge, Shen et al. [15] proposed a stochastic differential equation model, thereby significantly improving the robustness and accuracy of frontier detection. At the decision-making level, owing to its localized perceptual range, the simplistic shortest-distance greedy strategy is

prone to leading to path redundancy and flight instability [10]. To address this limitation, Cieslewski et al. [6] proposed a frontier selector aiming to minimize velocity variations, thereby significantly enhancing flight smoothness and stability during high-speed flight. Deng et al. [11] further explicitly formalized information gain as a differentiable function, thereby obtaining near-globally optimal exploration paths via gradient-based optimization. In recent years, FOV constraints have been widely incorporated to enhance exploration efficiency. The Fast UAV Exploration using Incremental Frontier Structure and Hierarchical Planning (FUEL) framework proposed by Zhou et al. [12,16] is a prominent example, and achieves high-frequency global environmental coverage and local agile maneuvering by maintaining an incremental frontier structure and hierarchical planning. Building on the FUEL framework, Zhao et al. [17] proposed the Fast Autonomous Exploration Planner (FAEP), which aims to address the issue of back-and-forth maneuvers and enhance global decision awareness. This method integrates flight-level factors and frontier-level spatial features into the Asymmetric Traveling Salesman Problem (ATSP) to generate rational exploration sequences. Frontier-based methods excel at guiding UAVs in large-scale unknown environments to rapidly cover unknown spaces, but their decision-making quality is highly dependent on the accuracy of frontier extraction and the global awareness of the selection strategy.

Sampling-based exploration methods conduct exploration by randomly generating candidate viewpoints within the mapped space, and are closely associated with "Next-Best-View (NBV)" planning [18]. Bircher et al. [19] pioneered the application of sampling-based methods to UAV exploration, proposing a Recurrent NBV Planner that guides exploration by rapidly expanding a rapidly exploring random tree (RRT) and assessing the information gain of candidate nodes. The bottleneck of such methods lies in the challenge of traditional early sampling frameworks in balancing exploration completeness and computational efficiency in complex unstructured environments. To improve the utilization efficiency of the sampling tree, subsequent studies proposed constructing sampling-based roadmaps to reutilize historical sampling data [20,21]. Schmid et al. [7] adopted the optimal reconnection idea of Rapidly Exploring Random Tree Star (RRT*) to significantly improve the quality of online trajectories by continuously maintaining and optimizing a single sampling tree. Sampling-based methods exhibit high flexibility in complex cluttered environments, but their inherent randomness can easily lead to local optima, making it challenging to ensure complete coverage of large-scale unknown environments [22].

Hybrid methods are designed to integrate the global guidance of frontier methods with the local fine-grained maneuvering and search capability of sampling-based methods to balance coverage efficiency and trajectory quality [23]. A typical paradigm of this class of methods is to integrate the global frontier guidance of frontier-based methods into the Traveling Salesman Problem (TSP) model to plan an optimized viewpoint visiting sequence. For example, Huang et al. [24] approximated the information gain by counting the number of frontiers observable from sampled candidate viewpoints, thereby significantly reducing computational overhead, and further employed the 2-opt algorithm to solve the TSP to determine the optimal candidate viewpoint sequence. Tang et al. [25] utilized unoccluded spheres to evaluate viewpoint visibility and optimize candidate viewpoint selection, and formulated it as an asymmetric Traveling Salesman Problem for efficient solution of the optimal candidate viewpoint sequence. To address the computational scalability challenges of large-scale environments, some studies have employed a hierarchical planning framework: first performing coarse-grained global path planning; and then executing fine-grained candidate viewpoint sampling and trajectory optimization in local subregions to balance efficiency and accuracy [26]. Although hybrid methods have achieved significant performance improvements, their structures are often highly complex, and their performance

heavily relies on the efficiency of core components such as TSP solvers and hierarchical planners in large-scale unstructured environments.

To overcome the aforementioned limitations, we incorporate the global frontier point set into the trajectory optimization framework to generate the near-globally optimal exploration trajectory and mitigate the computational complexity of hybrid strategies, thereby maximizing environmental information coverage. We utilize the incrementally explored map region to dynamically update and maintain predictive maps as well as their confidence, which characterize the spatial distribution and uncertainty of unknown environments, thereby providing dynamic prior knowledge for subsequent exploration decisions in unknown spaces. The intersection of the predicted flyable region and the high-confidence region is adopted as a safety-critical flight constraint, which integrates the environmental predictability and reliability from the dual-map module, to generate efficient and collision-free exploration trajectories, thereby achieving high-performance exploration.

In the field of image inpainting, deep learning-based image inpainting methods have achieved remarkable advances in recent years, especially in handling large-scale missing regions compared with traditional methods. The Mask-Aware Transformer (MAT) framework, proposed in [27] by Li et al., greatly improves the semantic coherence and quality of inpainting for large-scale missing regions by introducing a mask-aware attention mechanism to capture long-range semantic dependencies while adapting to mask distribution. Liu et al. [28] proposed a novel paradigm of parameterized coordinate queries to model spatial correlations in high-resolution feature maps, thereby enabling efficient and high-fidelity restoration of high-resolution images. To address the inherent prediction uncertainty issue in the inpainting process, Ma et al. [29] proposed an uncertainty-aware adaptive feedback network architecture that progressively reduces inherent prediction uncertainty via iterative adaptive feedback optimization and provides a quantitative reliability assessment for the inpainting results. Furthermore, Zheng et al. [30] integrated global context modeling and a multi-scale collaborative learning framework to capture long-range semantic dependencies and refine local details, respectively, thus significantly improving the restoration quality and robustness for severely damaged images. The attention-based Transformer model, proposed in [31] by Dong et al., operates in a low-resolution sketch space to capture the global structure prior of images, thus effectively reconstructing the global structural coherence of images, especially for those with severe structural damage. The significant advancements of these deep learning-based image inpainting methods in enhancing the reliability of inpainting results, addressing large-scale missing regions, and enabling structural reconstruction lay a solid technical foundation for UAVs to realize incremental map prediction and uncertainty estimation in unknown environments based on sparse local sensor observations.

3. Incremental Map Prediction

In this section, we first detail the collection and preprocessing pipeline of the constructed UAV exploration map dataset. Subsequently, we provide a detailed description of the designed network architecture, including the computation methods of the predicted map and confidence map. Finally, we present the formal definitions and optimization motivations of the key loss functions adopted in our framework, thereby providing theoretical support for subsequent experimental evaluations [27].

3.1. Dataset Construction

A Digital Elevation Model (DEM) is a digital topographic map that encodes surface elevation information in raster data format to represent the three-dimensional topographic features of the Earth's surface. This core elevation information can be used for a wide

range of geographic and environmental science applications, such as generating topographic maps, performing Geographic Information System (GIS) analysis, constructing terrain models, and developing hydrological models. We randomly crop out multiple non-overlapping 512 × 512-pixel sub-blocks from the DEM by uniformly selecting the top-left pixel coordinates (i, j) of each sub-block.

$$(i, j) \sim Uniform([0, W - L] \times [0, H - L]) \tag{1}$$

where W and H represent the width and height of the DEM, respectively, and L = 512 pixels. To obtain sub-images with moderate terrain undulations, we constrain the standard deviation of the elevation values, denoted as $\sigma(i, j)$.

$$\mu_{i,j} = \frac{1}{L^2} \sum_{(x,y) \in P_{i,j}} d(x, y) \tag{2}$$

$$\sigma(i, j) = \sqrt{\frac{1}{L^2} \sum_{(x,y) \in P_{i,j}} (d(x, y) - \mu_{i,j})^2} \tag{3}$$

where $d(x, y)$ represents the elevation value of the DEM at pixel (x, y), and $\mu_{i,j}$ denotes the average elevation value of the current DEM sub-block. In addition to using the standard deviation of elevation to constrain terrain undulations, we also impose constraints such as the number of local peaks and the proportion of flat areas to filter out maps that lack sufficient terrain complexity. However, the elevation values of the filtered sub-maps exhibit non-uniformity across samples. Therefore, we normalize the elevation values of each selected sub-map to ensure consistency in the input data for subsequent model training.

$$\frac{d(x, y) - d_{\min}(x, y)}{d_{\max}(x, y) - d_{\min}(x, y)} \times Z_{\max} \tag{4}$$

$$S = \{S_{i_1,j_1}, S_{i_2,j_2}, \ldots, S_{i_N,j_N}\} \tag{5}$$

Using Equation (4), the elevation values of each sub-map are normalized to the range $[0, Z_{\max}]$, where $Z_{\max}$ denotes the preset maximum elevation value, yielding the final dataset denoted as S, where S_{i_N,j_N} represents each normalized sub-map in the dataset.

Through rigorous random sampling and terrain feature filtering, we ensure the input samples not only cover diverse undulating terrains but also exclude large-scale flat regions or noisy areas, as shown in Figure 1a. This yields high-quality training and testing sub-maps that capture various environmental features, providing reliable input for the incremental map prediction model. For all samples in the map dataset, a color-to-height mapping method is employed to encode different elevation information in the images: cooler colors indicate lower elevation values, whereas warmer colors indicate higher elevation values, as shown in Figure 1b.

In addition to constructing the map dataset, a mask dataset is also required. We feed samples from the map dataset into a UAV exploration algorithm; as the UAV explores continuously, the scope of its explored regions expands while the scope of unknown regions shrinks dynamically. Explored regions are represented by black pixels, while unknown regions are denoted by white pixels, as shown in Figure 1c.

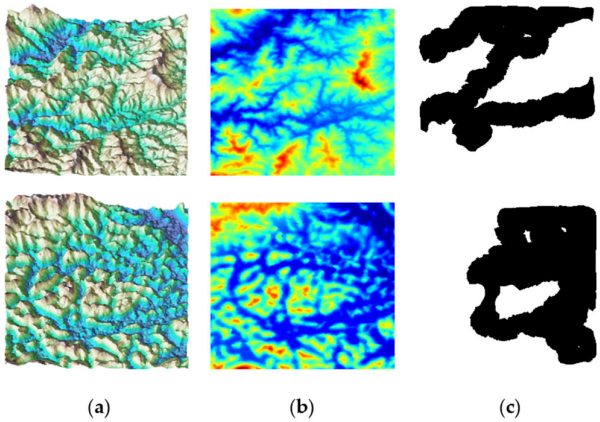


Figure 1. Examples of the constructed datasets. (**a**) Three-dimensional map dataset sample; (**b**) 2D map dataset sample; (**c**) mask dataset sample.

3.2. Network Architecture

The map prediction network architecture comprises three encoders, two decoders, and a dual-branch output layer, as illustrated in Figure 2. The input image and its corresponding mask are denoted as **I** and **M**, respectively. To extract discriminative shallow features from the input image **I**, three convolutional layers are employed to generate the encoder output feature denoted as **F**.

$$\mathbf{F} = \varepsilon_3(\varepsilon_2(\varepsilon_1([\mathbf{I}, \mathbf{M}]))) \tag{6}$$

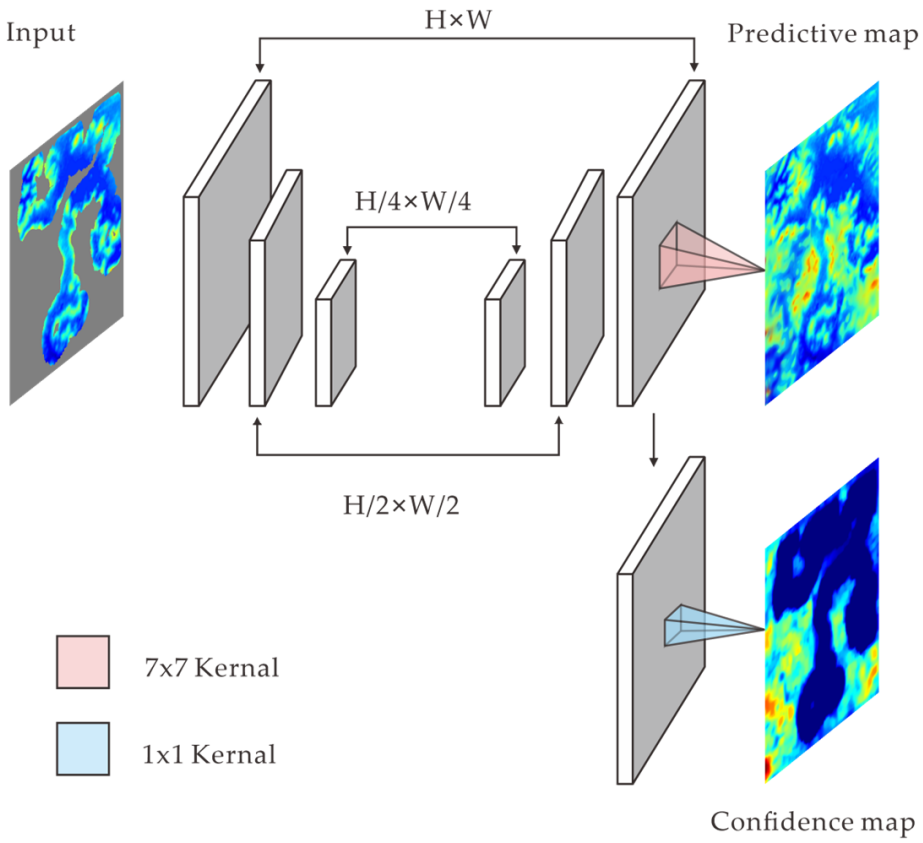


Figure 2. Architecture of the proposed incremental map prediction network.

The concatenated feature vector of the input image and its corresponding mask is denoted as [**I**, **M**] in Equation (6). This concatenated feature vector [**I**, **M**] is sequentially fed through three convolutional layers to extract multi-scale semantic features. The first

encoder layer ε_1 employs a convolution with a large receptive field to preserve spatial resolution, while the second and third encoder layers ε_2 and ε_3 adopt convolutions with smaller receptive fields combined with downsampling operations. This design has two core objectives: first, capturing local contextual information via large-kernel convolutions; second, progressively aggregating global contextual information through downsampling. This process yields a compact feature tensor that integrates local details and global semantics to support subsequent modules.

The decoder module comprises two transposed convolutional layers, which are responsible for generating the inpainting result denoted as $\hat{\mathbf{I}}^{(\mathbf{t})}$ and the corresponding uncertainty map denoted as $\hat{\mathbf{U}}^{(\mathbf{t})}$. Two-level shallow features, denoted as $\mathbf{f_1}$ and $\mathbf{f_2}$, are introduced in Equation (7), which correspond to the outputs of the first and second encoder layers from the preceding encoder module. In the map reconstruction process, the encoder output feature $\mathbf{F}$ is first upsampled (denoted as U) and then fused with the two-level shallow features $\mathbf{f_1}$ and $\mathbf{f_2}$. This enables the direct injection of shallow details into the final features tensor, thereby facilitating the enhanced recovery of high-frequency information and the preservation of spatial consistency.

$$\begin{cases} \mathbf{D}^{(\mathbf{t})} = \mathrm{Dec}_2(\mathrm{Dec}_1(U(\mathbf{F}), \mathbf{f_2}), \mathbf{f_1}), \\ \hat{\mathbf{I}}^{(\mathbf{t})} = \tanh(\mathrm{Conv}(\mathbf{D}^{(\mathbf{t})})), \\ \hat{\mathbf{U}}^{(\mathbf{t})} = \phi(\mathrm{Conv}(\mathbf{D}^{(\mathbf{t})})), \end{cases} \tag{7}$$

In the generation process of the inpainting result $\hat{\mathbf{I}}^{(\mathbf{t})}$, a larger convolutional kernel is employed to aggregate broader contextual information from the neighborhood regions. This facilitates the extraction of information (e.g., color textures and edge orientations) around the occluded regions in the original image, thereby achieving smooth and consistent reconstruction of the missing regions. Thus, the upper output branch in Figure 2 leverages wide-receptive-field convolutions and color-to-height mapping to learn how to reasonably inpaint the missing or occluded regions using texture and structural information from surrounding pixels, thereby fulfilling the image inpainting task. The generation of the uncertainty map $\hat{\mathbf{U}}^{(\mathbf{t})}$ involves performing a channel-wise weighted sum independently for each spatial location, which compresses a multi-dimensional feature vector into a single scalar value. This scalar value can be regarded as an aggregated score of the recovery quality or prediction confidence for the current spatial location. Specifically, this branch generates inpainting confidence for each spatial location by performing pointwise channel aggregation on the multi-channel features, followed by a non-negative activation function, thereby explicitly quantifying the reliability of the network's prediction results across different regions.

3.3. Loss Function

In the proposed incremental map prediction method, we adopt five common loss functions: adversarial loss $\mathcal{L}_{adv}$, style loss $\mathcal{L}_{style}$, perceptual loss $\mathcal{L}_{percep}$, total variation loss $\mathcal{L}_{tv}$, and mutual information loss $\mathcal{L}_{MI}$. These losses primarily focus on the appearance and statistical characteristics of the reconstructed map, making it challenging to intuitively quantify the model's confidence in unknown regions. To address this limitation, we design a probabilistic reconstruction loss $\mathcal{L}_{rec}$, which incorporates the network's prediction uncertainty for each spatial location as a learnable term into the loss formulation. This loss aims to simultaneously optimize reconstruction accuracy and uncertainty estimation during model training.

1. Probabilistic Reconstruction Loss: This loss is formulated based on the Gaussian negative log-likelihood function, enabling the network to simultaneously learn recon-

struction values and prediction uncertainty for each spatial location. In Equation (8), σ_i denotes the uncertainty estimation value output by the network at the i-th spatial location. Herein, $\hat{\mathbf{I}}_\mathbf{i}$ represents the predicted mean value of the i-th pixel in the inpainted result, $\mathbf{I_i}$ denotes the ground-truth value of the i-th pixel in the input image $\mathbf{I}$, and $\mathbf{M_i}$ stands for the ground-truth value of the i-th pixel in the mask image $\mathbf{M}$. The first term γ enforces an inverse proportionality between the reconstruction error and the prediction uncertainty. If the network outputs a large σ_i, it indicates that the model deems its prediction at this spatial location to be highly unreliable (e.g., due to high noise interference or inherently difficult-to-predict data). In such cases, we avoid excessive inflation of the overall loss caused by reconstruction errors within the noise range by scaling down the error using $\sigma_i^2 + \delta$. The second term $\ln(\sigma_i^2 + \delta)$ imposes a penalty on the network for assigning excessively large σ_i values. Meanwhile, the third term, $\ln \sigma_0^2$, serves as a constant offset to prevent the loss value from becoming excessively large. By decomposing the total loss into the loss for missing region $\mathcal{L}_{holc}$ and the loss for the valid region $\mathcal{L}_{valid}$—where λ_{holc} denotes the weight for the missing region loss and λ_{valid} represents the weight for the valid region loss—we ensure the model receives adequate supervision signals for both regions, thereby enhancing the reconstruction quality and the accuracy of confidence estimation.

$$\begin{cases} \gamma = \dfrac{\left(\hat{\mathbf{I}}_\mathbf{i} - \mathbf{I_i}\right)^2}{\sigma_i^2 + \delta} \\ E_i = \gamma + \ln(\sigma_i^2 + \delta) - \ln \sigma_0^2 \\ \mathcal{L}_{holc} = \lambda_{holc} \dfrac{1}{\sum\limits_i \mathbf{M_i}} \sum\limits_i \left(\dfrac{1}{2} E_i \mathbf{M_i} \right) \\ \mathcal{L}_{valid} = \lambda_{valid} \dfrac{1}{\sum\limits_i (1 - \mathbf{M_i})} \sum\limits_i \left[\dfrac{1}{2} E_i (1 - \mathbf{M_i}) \right] \\ \mathcal{L}_{rec} = L_{holc} + L_{valid} \end{cases} \tag{8}$$

2. Overall Loss: The overall loss of the proposed method is a weighted combination of the aforementioned loss functions.

$$\begin{aligned} \mathcal{L}_G \quad &= \lambda_{rec}\mathcal{L}_{rec} + \lambda_{adv}\mathcal{L}_{adv} + \lambda_{style}\mathcal{L}_{style} \\ &+ \lambda_{percep}\mathcal{L}_{percep} + \lambda_{tv}\mathcal{L}_{tv} + \lambda_{mi}\mathcal{L}_{MI} \end{aligned} \tag{9}$$

where λ_{rec}, λ_{adv}, λ_{style}, λ_{percep}, λ_{tv}, and λ_{mi} are the hyperparameters corresponding to the respective loss functions. Collaborative optimization of these losses imposes multi-level, multi-scale constraints on the generated results. It ultimately enables high-quality, reliably controllable UAV environment inpainting and map prediction.

4. Frontier Exploration Guided by Predicted Maps

The proposed framework operates on a voxel grid map. As illustrated in Figure 3, it comprises a frontier incremental update module and a predictive information gain module. Whenever the map is updated with sensor measurements, the system updates and maintains the predicted map along with its corresponding confidence map based on the current exploration state, while checking for affected frontier clusters. If such affected frontier clusters exist, they are removed, and new frontier clusters are extracted. Subsequently, the exploration planning process is triggered, which selects the final target point by evaluating the frontier information gain and predictive information gain of each candidate viewpoint.

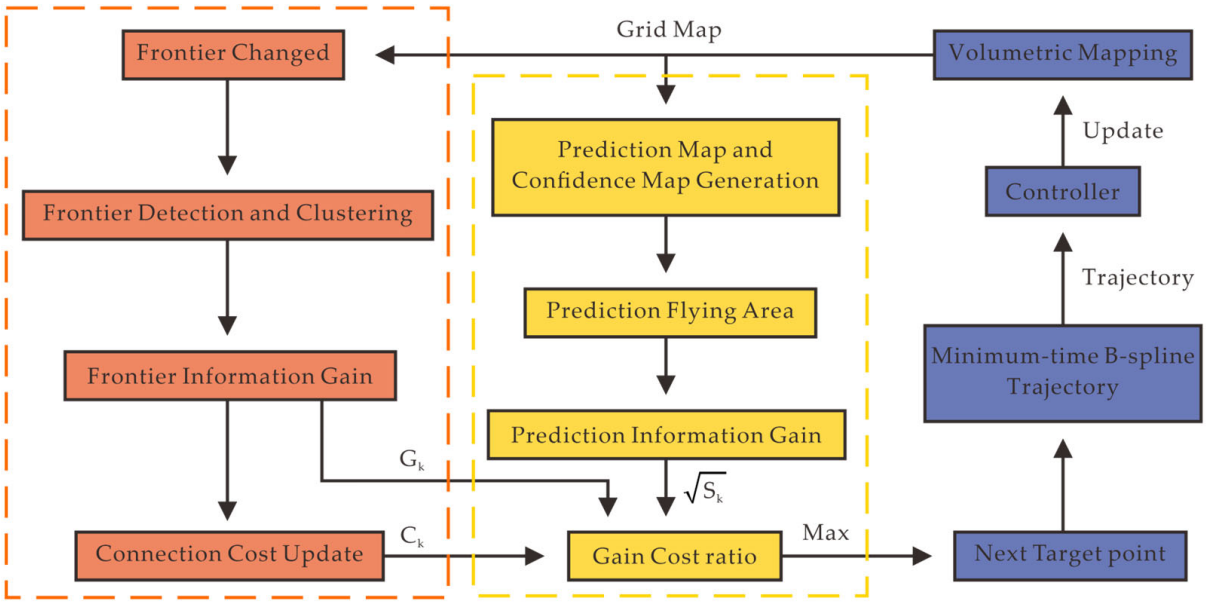


Figure 3. Overview of the predicted map-guided frontier exploration framework.

4.1. Incremental Frontier Update

In the proposed incremental map prediction-driven UAV exploration framework, the incremental frontier update module serves as a key component for the efficient operation of the entire system. Its primary objective is to maintain accurate representations and dynamic tracking of unknown space boundaries during UAV flight, while minimizing computational overhead and latency. Unlike traditional full-map scanning approaches, this module focuses solely on local voxel regions affected by the latest sensor observations, thus greatly reducing global data structure traversal frequency.

Specifically, when the UAV's sensor acquires new environmental information and writes it to the voxel grid map, the module first identifies pixels with changed states—typically those transitioning from an unknown state to an observed free state or an observed occupied state. These pixel state changes are not isolated; instead, they exert a cascading effect on neighboring pixels; if a voxel transitions from unknown to known, and it has at least one adjacent voxel remaining in an unknown state, this location is designated as a new "frontier point", denoted as x_i in Equation (10). Conversely, if the voxel corresponding to a frontier point is reclassified as occupied, or if no unknown pixels remain in its neighborhood, the frontier point is removed during the incremental update. Throughout this process, the system performs detection operations exclusively on affected voxel blocks, rather than conducting a point-by-point check of the entire grid map, thereby achieving a substantial reduction in overall computational load.

$$X = \{x_1, x_2, x_3, \ldots, x_i, \ldots\} \tag{10}$$

After completing the addition and removal of frontier points, the next step is to cluster the newly formed set of frontier points X to generate more representative frontier clusters. This step not only reduces the number of nodes for subsequent exploration planning but also preserves key information (e.g., spatial distribution and boundary geometry). We adopt the efficient tree-based data structure, Disjoint Set Union (DSU), to manage frontier clusters: for each newly generated frontier point, the system first queries the identifiers of existing clusters in its neighborhood. If the point is adjacent to multiple clusters, a cluster merging operation is triggered to unite these clusters under a single root node in the DSU. Following merging, the system immediately updates the cluster's properties, such as its

centroid position, total number of frontier points, and boundary polygon. If the removal of certain frontier points disrupts the connectivity of an original cluster, the system employs a connected-component segmentation algorithm based on Depth-First Search (DFS) to split the fragmented parts into several sub-clusters, while recalculating and storing the attribute data for each sub-cluster. For any new frontier points that are neither connected to an existing cluster nor assigned to one, the system creates a new cluster proactively and assigns it an initial identifier and attribute data.

For each frontier cluster, its centroid and boundary features are extracted. Candidate viewpoints are then generated diversely by offsetting outward along the boundary normals by a safe distance, based on the cluster's principal direction and key boundary directions. Subsequently, a rapid collision detection is performed on each candidate position to filter out unreachable viewpoints. The number of observable unknown pixels is estimated by simulating the sensor's frustum, and the corresponding information gain is calculated. Incorporating a balancing strategy between coverage and motion cost, the system selects several cost-optimal viewpoints for local path planning. Furthermore, as the UAV navigates and receives observational data feedback, these viewpoints are dynamically adjusted based on the updated frontier clusters. This ensures that the exploration process remains focused on the most valuable unknown space, thereby achieving efficient environmental information acquisition while guaranteeing flight safety.

4.2. Predicted Map-Guided Goal Point Selection

For continuous exploration, three complementary maps are maintained in our framework with distinct state information: the observed map records real-time free and occupied pixels acquired by the sensor; the predictive map leverages a deep learning model to inpaint unknown regions during the observation process; and the confidence map quantifies the reliability of prediction results by scoring the accuracy of different regions. Upon receiving each new frame of sensor data, the observed map immediately marks the newly added free and occupied regions. The predicted map then activates its prediction model, using the updated observed map as a basis to fill in unknown regions, while the confidence map synchronously updates the reliability values for the predicted regions.

In our approach, the inpainting of the masked regions is performed via a deep learning model, yielding the network's predicted output $\hat{\mathbf{I}}$ as defined in Equation (7). According to the binary mask M illustrated in Figure 1b (where black pixels denote observed regions and white pixels denote regions to be inpainted), a linear interpolation strategy is employed.

$$\mathbf{P} = \mathbf{M}\hat{\mathbf{I}} + (1 - \mathbf{M})\mathbf{I} \tag{11}$$

It seamlessly fuses the model's inpainting result with the original input I, maximizing the preservation of real observed pixels while directing the network to focus on the inpainting task and avoiding unnecessary error propagation. In practice, the input map and binary mask are first normalized to the range $[-1, 1]$ and concatenated into a four-channel tensor, which is then fed into the network. After the network generates the output, the output undergoes denormalization and the aforementioned element-wise fusion to ultimately produce the predictive map P, which is used for subsequent visualization and spatial analysis.

The network also outputs a raw uncertainty map $\hat{\mathbf{U}}$ which denotes the model's confidence level in the prediction for each voxel. To eliminate scale variations under different map environments and mask configurations, morphological opening and Gaussian smooth-

ing are first applied to $\hat{\mathbf{U}}$ for a smoothed map $\widetilde{\mathbf{U}}$. Subsequently, within the masked region Ω, the following operation is performed:

$$\widetilde{\mathbf{U}}_{\mathbf{norm}} = \frac{\widetilde{\mathbf{U}} - \min_{(p,q)\in\Omega}\widetilde{\mathbf{U}}(p,q)}{\max_{(p,q)\in\Omega}\widetilde{\mathbf{U}}(p,q) - \min_{(p,q)\in\Omega}\widetilde{\mathbf{U}}(p,q)} \tag{12}$$

which maps the values to the interval [0, 1] linearly. After normalization, $\widetilde{\mathbf{U}}_{\mathbf{norm}}$ directly reflects the relative confidence: a value closer to 0 indicates a more reliable prediction, while a value closer to 1 denotes higher uncertainty. This processing not only facilitates unified threshold-based judgments (e.g., filtering high-confidence regions) but also enables convenient hybrid computation with other normalized features on the same scale.

To quantify the coverage of the current voxel grid map, we introduce the exploration rate E defined as

$$E = \frac{1}{|\Omega|}\sum_{(i,j\in\Omega)} \mathrm{f}[M(i,j) = 0] \tag{13}$$

where the function $\mathrm{f}[M(i,j) = 0]$ quantifies the proportion of pixels observed by the point cloud (i.e., where the mask value is 0). The exploration rate $E \in [0,1]$ reflects the known-to-unknown region ratio in the environment. When E is small, the system is in the initial exploration phase and should prioritize frontier-based methods to expand its field of view. When E is large, we adjust the strategy to guide the UAV for efficient exploration via the predicted map. Based on the predicted map and confidence map, we select regions that satisfy three conditions simultaneously: being a predicted flyable region, having high confidence, and being unexplored by the UAV—as the next candidate regions. However, each candidate region has a different gain–cost ratio. As illustrated in Figure 4, for each target point k, we determine its position and orientation to identify the corresponding voxel region S_k of its predicted flyable region. We then calculate the gain–cost ratio for each target point individually:

$$B_k = \frac{G_k + \sqrt{S_k}}{C_k} \tag{14}$$

where G_k denotes the frontier information gain, defined as the number of visible frontier pixels when the UAV reaches the specific target point. C_k represents the motion cost, which includes the path length and yaw adjustment cost for the UAV to navigate from its current state to the specific target point. The final next target point for the UAV is determined by comparing the magnitude of B_k values.

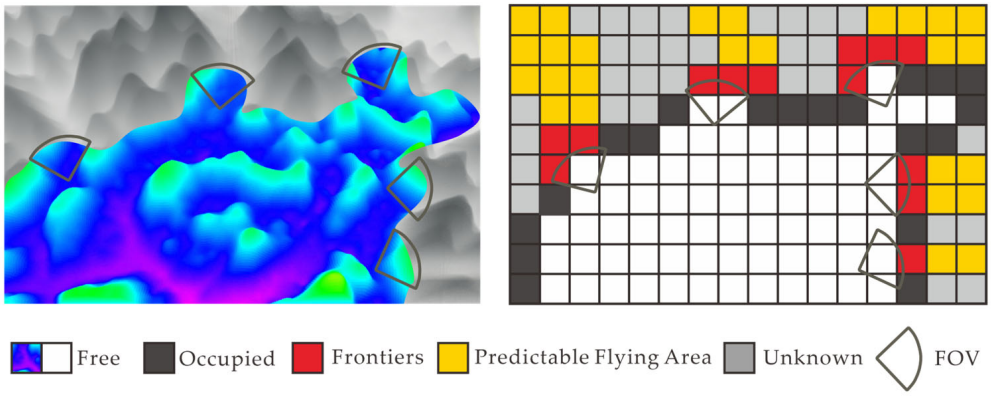


Figure 4. Frontier information gain structure and predicted flyable voxel region.

5. Experimental Design

5.1. Experimental Setup

1. Dataset: The DEM data utilized in this study were extracted from a geographical region spanning 97°31′–106°11′ E longitude and 21°8′–29°15′ N latitude. Following the procedure detailed in Section 3.1, the entire area was subjected to random cropping and terrain feature filtering, ultimately constructing a dataset of approximately 260,000 high-quality sub-map samples, each with a resolution of 512 × 512-pixels. To obtain the corresponding mask annotations, a subset of these sub-maps was randomly selected and input into an algorithm (used solely for generating coverage maps and unrelated to the algorithm proposed in this paper). By running this simulator multiple times under different exploration rate configurations, a series of binary masks were generated: explored regions were denoted by black pixels (value 0), while the remaining unknown regions were denoted by white pixels (value 1). In this manner, a mask set with coverage levels ranging from low to high was obtained, simulating the FOV distribution of a UAV during the early, middle, and late stages of exploration. This mask dataset is large scale and well distributed, capable of meeting the training and evaluation requirements of our network under various degrees of incompleteness. Additionally, it ensures the model's strong generalization ability for map prediction and uncertainty estimation across different exploration progress stages.
2. Implementation Details: The experiments in this work were conducted on a computer equipped with an Intel Core i7-12700F CPU with 32 GB of RAM, and an NVIDIA GeForce RTX 4060 Ti GPU, running the Ubuntu 18.04 operating system. The implementation uses Python 3.10 and the PyTorch 2.0.0 framework. In the simulation experiments, Equation (4) is used to normalize the map's elevation values to the range $[0, Z_{\max}]$. The high-quality sub-maps with a resolution of 512 × 512 pixels were scaled to three different actual physical sizes, namely 30 m × 30 m × 8 m, 51.2 m × 51.2 m × 10 m, and 80 m × 80 m × 12 m, and the scaled maps were employed as the exploration maps for the UAV.
3. Hyperparameter Sensitivity Analysis: In addition to the above experimental configurations, this study further carried out a sensitivity analysis on the hyperparameters of the loss functions in Equations (8) and (9). The values of each hyperparameter were set as follows in the experiments: the weight of the missing region loss λ_{holc} was 6, the weight of the valid region loss λ_{valid} was 1, the weight of the probabilistic reconstruction loss λ_{rec} was 1, the weight of the adversarial loss λ_{adv} was 0.1, the weight of the style loss λ_{style} was 120, the weight of the perceptual loss λ_{percep} was 0.05, the weight of the total variation loss λ_{tv} was 0.1, and the weight of the mutual information loss λ_{mi} was 0.0001. To evaluate the impact of different hyperparameters on model performance, we categorized the above hyperparameters into two types: core parameters and secondary parameters. The core parameters mainly included λ_{holc}, λ_{adv}, λ_{rec}, and λ_{style}, which played a key role in map prediction accuracy. The secondary parameters covered λ_{valid}, λ_{mi}, λ_{tv}, and λ_{percep}. For the two types of parameters, validation experiments were conducted under two mask scenarios (Small Mask and Large Mask), with the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as the evaluation metrics. The influence of each parameter on the map prediction effect under different values is shown in Table 1. Finally, the above optimal hyperparameter combination was determined to ensure the stability and accuracy of the model in the tasks of terrain reconstruction and uncertainty estimation.

Table 1. Effects of core parameters and secondary parameters on map prediction accuracy.

Evaluation Dimension	Core Metrics	PSNR		SSIM	
	Mask Ratio	Small Mask	Large Mask	Small Mask	Large Mask
Core Parameters	$\lambda_{holc} = 4$	23.25	20.75	0.9505	0.8432
	$\lambda_{holc} = 6$	27.55	20.44	0.9603	0.8644
	$\lambda_{holc} = 8$	26.22	20.08	0.9591	0.8536
	$\lambda_{adv} = 0.05$	24.84	20.28	0.9545	0.8601
	$\lambda_{adv} = 0.1$	27.55	20.44	0.9603	0.8644
	$\lambda_{adv} = 0.15$	24.16	20.11	0.9528	0.8542
	$\lambda_{rec} = 0.8$	26.49	20.14	0.9639	0.8559
	$\lambda_{rec} = 1.0$	27.55	20.44	0.9603	0.8644
	$\lambda_{rec} = 1.2$	25.32	20.04	0.9547	0.8519
	$\lambda_{style} = 100$	24.04	19.67	0.9532	0.8784
	$\lambda_{style} = 120$	27.55	20.44	0.9603	0.8644
	$\lambda_{style} = 140$	22.64	19.91	0.9500	0.8508
Secondary Parameters	$\lambda_{valid} = 1$	27.55	20.44	0.9603	0.8644
	$\lambda_{valid} = 2$	26.18	20.31	0.9634	0.8590
	$\lambda_{mi} = 0.0001$	27.55	20.44	0.9603	0.8644
	$\lambda_{mi} = 0.005$	27.12	20.03	0.9591	0.8630
	$\lambda_{tv} = 0.05$	26.96	19.618	0.9489	0.8509
	$\lambda_{tv} = 0.1$	27.55	20.44	0.9603	0.8644
	$\lambda_{percep} = 0.05$	27.55	20.44	0.9603	0.8644
	$\lambda_{percep} = 0.1$	26.51	20.32	0.9589	0.8619

5.2. Map Prediction and Confidence Map: Results and Analysis

Based on the optimal hyperparameter combination determined in Section 5.1, we conducted comprehensive experiments on diverse maps and masks to evaluate the performance of the proposed map prediction and confidence map generation model, and further compared it with the state-of-the-art MAT method [27] on the custom dataset.

Table 2 presents the experimental results, which demonstrate that the proposed method outperforms current state-of-the-art methods on our custom dataset in both PSNR and SSIM metrics. These results confirm that the proposed method not only effectively inpaints missing regions in the map but also largely preserves the original map's structural and topographic semantic information.

Table 2. Quantitative evaluation results of different algorithms on the custom dataset.

Mask Ratio		Small Mask	Large Mask
MAT	PSNR	23.66	18.04
Ours		27.55	20.44
MAT	SSIM	0.7899	0.6662
Ours		0.9603	0.8644

In this section, we further evaluate the coherence of our custom dataset at the visual and semantic levels through qualitative experiments. Specifically, we simulated the UAV to complete exploration with a coverage rate of 40–50% under multiple sets of conditions, including different terrain maps, initial takeoff positions, and headings. At this stage, we performed prediction on the unknown regions (as illustrated in Figure 5). Here, Figure 5c presents the predicted map generated at the current exploration progress, while Figure 5d shows the corresponding confidence map. By comparing the predicted results with the complete global map, it is evident that our method achieves semantically con-

sistent terrain inpainting while preserving rich topographic details. The confidence map intuitively reflects the reliability of prediction for each region through a cool–warm color scheme: cool-colored areas correspond to high confidence (i.e., high prediction accuracy), whereas warm-colored areas indicate higher uncertainty (i.e., low prediction accuracy). This visualization confirms that our method not only exhibits high-quality inpainting performance but also outputs a credibility assessment for each pixel, providing valuable reference for subsequent task planning and decision making.

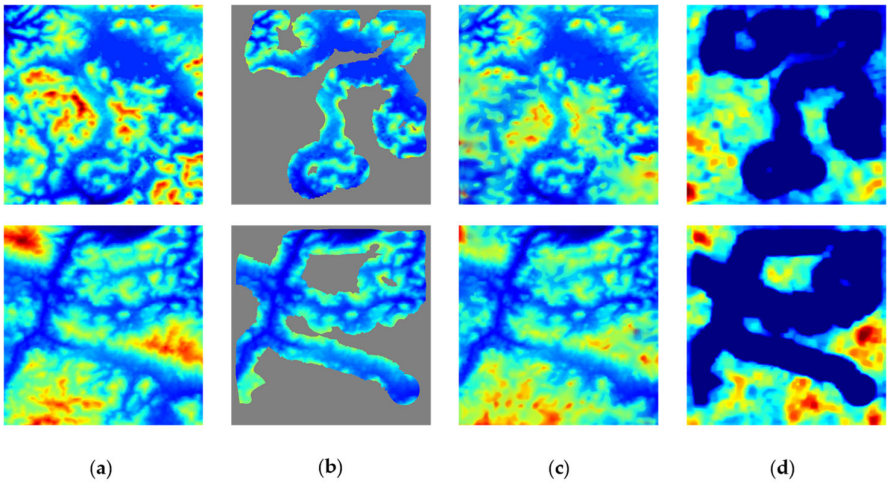


Figure 5. Map visualization results. (**a**) Global map; (**b**) exploration map; (**c**) predicted map; (**d**) confidence map.

5.3. UAV Exploration Results and Analysis

We validated our proposed framework through simulation experiments and further evaluated it in real-world scenarios using the aforementioned custom dataset. We compared our framework against three baseline methods: FUEL, FAEP, and Classic method. It should be noted that no open-source code was available for the Classic method; thus, we implemented it independently for the comparative experiments.

For all experiments, the dynamic constraints were uniformly set to the same values for all methods: a maximum linear speed of 2 m/s and a maximum angular velocity of 0.9 rad/s. The sensor's FOV was configured as 80° horizontally and 60° vertically (as illustrated in Figure 6), with a maximum detection range of 4.5 m.

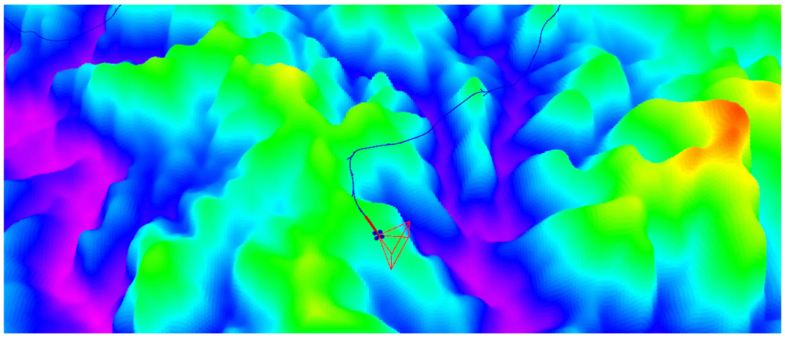


Figure 6. UAV local exploration in a digital elevation map (DEM) and sensor field of view (FOV) visualization. Blue lines represent the UAV's flight path.

We selected three representative maps from our custom map dataset, as illustrated in Figure 7; each method was executed ten times under the identical initial configuration. To verify that the performance improvements of the proposed method over the

Classic, FUEL, and FAEP methods were statistically significant—rather than mere artifacts of random experimental variation—we adopted the Wilcoxon signed-rank test for pairwise comparative analyses. The test data were extracted from the core performance metrics in Tables 3–6, which cover the three test maps and experimental subscenarios with different flight altitudes. We set the significance level to $\alpha = 0.05$, a widely recognized benchmark in engineering and robotics research. For each performance metric, we conducted pairwise comparisons between the proposed method and the Classic, FUEL, and FAEP methods. Furthermore, we added significance labels to the corresponding cells in Tables 3–6 (where * indicates $p < 0.05$ and ** indicates $p < 0.01$) to intuitively reflect the statistical significance across different experimental subscenarios. As shown in Tables 3–6, for exploration time and flight distance, most comparisons with Classic and FUEL reached the highly significant level of $p < 0.01$, while those with FAEP also met the significance level of $p < 0.05$. This indicates that the performance advantages of the proposed method in these metrics were not attributable to random fluctuations in repeated experiments, but reliable improvements were demonstrated consistently across all experimental scenarios.

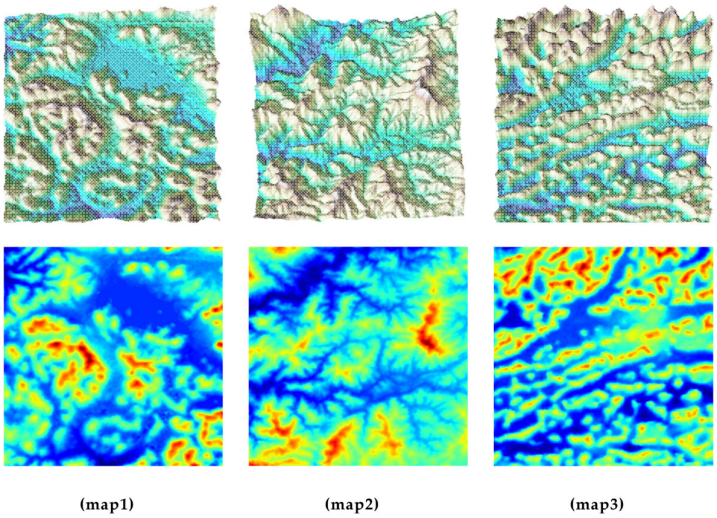


Figure 7. Different test scenarios for UAV exploration.

Table 3. Exploration statistics in the digital elevation map scenario (UAV maximum flight altitude 5.5 m, map size: 51.2 m × 51.2 m × 10 m).

Map	Method	Exploration Time (s)				Flight Distance (m)			
		Avg	Std	Max	Min	Avg	Std	Max	Min
Map1	Classic [10]	1601 **	309	2955	1155	2063 **	283	2503	1354
	FUEL [16]	823 **	36	914	767	1282 **	73	1446	1164
	FAEP [17]	888 **	48	948	749	1201 **	70	1258	1009
	Ours	745	28	776	683	1145	66	1217	995
Map2	Classic	2520 **	296	2955	2024	2099 **	389	2660	1409
	FUEL	1078 **	178	2110	903	1601 *	71	1745	1352
	FAEP	966 **	67	1206	860	1570 *	110	1840	1441
	Ours	922	59	1078	820	1546	68	1645	1377
Map3	Classic	1511 *	243	1956	1010	2043 *	367	2677	1192
	FUEL	719 *	36	819	639	1050 *	76	1225	865
	FAEP	730 *	34	766	681	997 *	60	1089	897
	Ours	671	32	761	577	967	58	1121	807

Table 4. Exploration statistics in the digital elevation map scenario (UAV maximum flight altitude 6.5 m, map size: 51.2 m × 51.2 m × 10 m).

Map	Method	Exploration Time (s)				Flight Distance (m)			
		Avg	Std	Max	Min	Avg	Std	Max	Min
Map1	Classic [10]	2557 **	228	2875	2089	2453 **	475	3057	1696
	FUEL [16]	1344 **	87	1631	1161	1923 **	65	2036	1786
	FAEP [17]	1265 **	87	1440	1102	1829 *	63	1985	1742
	Ours	1173	69	1361	1023	1805	56	1893	1676
Map2	Classic	2526 **	572	3990	1593	2210 **	611	3306	1428
	FUEL	2057 **	233	2582	1586	2156 *	69	2335	2007
	FAEP	2129 **	173	2582	1945	2190 *	46	2255	2108
	Ours	1594	120	1832	1322	2190	60	2285	2077
Map3	Classic	2662 **	462	3373	2007	2373 **	469	3158	1836
	FUEL	1324 **	86	1507	1148	1979 *	111	2202	1641
	FAEP	1294 **	73	1389	1122	1884 *	62	1959	1792
	Ours	1187	65	1307	1035	1851	101	2035	1623

Table 5. Exploration statistics in the digital elevation map scenario (UAV maximum flight altitude 5.5 m, map size: 30 m × 30 m × 8 m).

Map	Method	Exploration Time (s)				Flight Distance (m)			
		Avg	Std	Max	Min	Avg	Std	Max	Min
Map1	Classic [10]	513 **	31	634	475	785 **	48	919	674
	FUEL [16]	360 **	18	403	321	523 *	34	617	457
	FAEP [17]	355 **	17	385	319	507 *	43	586	435
	Ours	328	17	365	255	477	27	533	381
Map2	Classic	579 **	48	752	407	818 **	64	945	636
	FUEL	406 **	18	453	362	503 **	31	570	432
	FAEP	366 **	20	418	322	481 **	28	551	416
	Ours	325	17	366	291	472	25	540	431
Map3	Classic	568 **	36	654	515	679 **	80	775	526
	FUEL	370 **	19	395	334	528 **	35	572	454
	FAEP	351 *	25	412	297	464 *	30	522	387
	Ours	337	13	358	308	426	17	452	392

Table 6. Exploration statistics in the digital elevation map scenario (UAV maximum flight altitude 8 m, map size: 80 m × 80 m × 12 m).

Map	Method	Exploration Time (s)				Flight Distance (m)			
		Avg	Std	Max	Min	Avg	Std	Max	Min
Map1	Classic [10]	12,758 **	2000	15,724	10,446	4431 **	1300	6997	2806
	FUEL [16]	2123 **	74	2317	2012	3226 **	60	3365	3115
	FAEP [17]	2025 *	58	2109	1921	3069 **	59	3165	2956
	Ours	1916	48	2014	1785	2943	53	3043	2846
Map2	Classic	14,064 **	2279	18,391	10,096	5308 **	1507	6976	2795
	FUEL	2164 **	116	2519	1916	3218 **	183	3553	2750
	FAEP	1961 *	106	2120	1741	2963 *	129	3157	2641
	Ours	1831	91	2161	1686	2869	163	3247	2504
Map3	Classic	13,493 **	1322	15,542	11,220	4873 **	762	6238	3982
	FUEL	1528 **	89	1803	1328	2292 **	179	2770	1908
	FAEP	1728 **	76	1883	1655	2093 *	105	2279	1987
	Ours	1361	64	1432	1202	1996	82	2121	1838

The paired experimental data across all experimental subscenarios were integrated into a unified sample set, and a comprehensive statistical analysis was conducted to generate overall statistical results. The detailed results of the Wilcoxon signed-rank test are summarized in Table 7, which includes all comparison pairs, test statistics, *p*-values, and corresponding significance labels. The data in this table represent the aggregated results obtained by merging the experimental data across all flight altitude scenarios of the three test maps. As shown in Table 7, for both exploration time and flight distance, the proposed method significantly outperformed all baseline methods, with all pairwise comparisons attaining statistical significance at the $p < 0.001$ level (***). Specifically, the test statistics ranged from 0 to 182 for exploration time and from 11 to 628 for flight distance, demonstrating statistically highly significant performance differences between the proposed method and Classic, FUEL, and FAEP, respectively. These results confirm that the performance advantages of the proposed method in exploration efficiency are robust and not attributable to random experimental fluctuations.

Table 7. Overall statistical results of the Wilcoxon signed-rank test.

Performance Metrics	Comparison Pairs	Test Statistic	*p*-Value	Significance Labels
Exploration time (s)	The proposed method and Classic	0	<0.001	***
	The proposed method and FUEL	23	<0.001	***
	The proposed method and FAEP	182	<0.001	***
Flight distance (m)	The proposed method and Classic	11	<0.001	***
	The proposed method and FUEL	147	<0.001	***
	The proposed method and FAEP	628	<0.001	***

The proposed method and FAEP

In subsequent experiments, we will continue to adopt the map size of 51.2 m × 51.2 m × 10 m as the default (unless otherwise specified). This is because this size can not only well preserve the geographic information features in real-world environments, but also greatly save the time cost and memory cost required for experiments. To further verify the real-time application potential of the proposed algorithm, we supplementally compared the computational efficiency of different methods, with the results presented in Table 8. Similar to advanced methods such as FAEP and FUEL, our framework avoids the global traversal and redundant computation inherent in traditional methods via the incremental map prediction and frontier update mechanism, thus achieving lower computational overhead.

Table 8. Average computation time in different scenes, map size: 51.2 m × 51.2 m × 10 m.

Method	Experimental Results in Three Scenes (ms)		
	Map1	Map2	Map3
Classic	246	142	124
FUEL	49	44	48
FAEP	37	35	36
Ours	32	27	29

To provide a more intuitive understanding of exploration performance, we also plotted the curves of exploration rate versus time, as illustrated in Figure 8. In the initial exploration stage, all methods exhibit comparable performance. However, as exploration proceeds, the predicted map generated by our method becomes increasingly accurate—this

accuracy enables the UAV to be further guided toward regions with higher information gain. Consequently, the performance advantage of our method becomes simultaneously more pronounced.

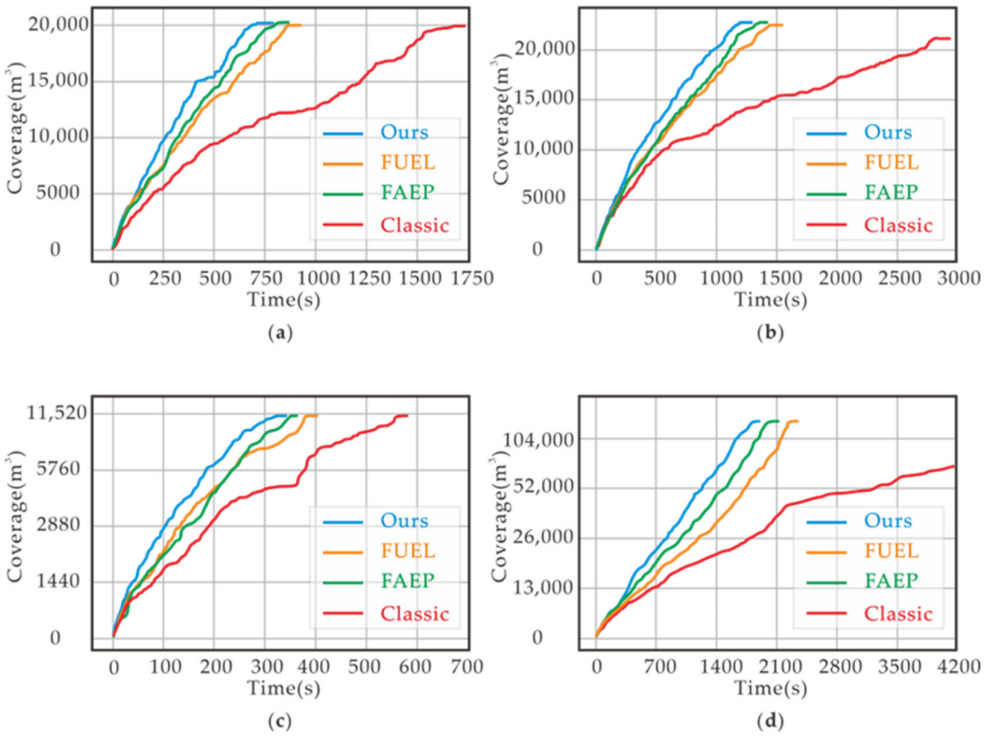


Figure 8. Exploration rate versus time curves. (**a**) UAV maximum flight altitude: 5.5 m, map size: 51.2 m × 51.2 m × 10 m; (**b**) UAV maximum flight altitude: 6.5 m, map size: 51.2 m × 51.2 m × 10 m; (**c**) UAV maximum flight altitude: 5.5 m, map size: 30 m × 30 m × 8 m; (**d**) UAV maximum flight altitude: 6.5 m, map size: 80 m × 80 m × 12 m.

In contrast, our method uses the predicted map to perform a priori estimation of future visible regions and updates the prediction map and confidence map in real time before each exploration step, thereby greatly reducing biases in gain evaluation. As the flight progresses, prediction accuracy continuously improves (Figure 9), enabling the model to consistently select the next target point with the maximum information gain. This avoids indecisive back-and-forth movement in confined spaces, ensures each decision advances toward the global optimum, and effectively enhances the exploration rate and path coverage quality.

Furthermore, we evaluated the accuracy of the predicted map and confidence map, as well as their effectiveness in guiding flight path planning, under various coverage conditions. Experiments show that as the UAV's exploration progresses, the cool-colored regions in the generated confidence map—representing high prediction reliability—gradually expand. This demonstrates that the model's estimation accuracy for unknown terrain is continuously improving. To utilize these predictions for path decision making, we compute the intersection of three region types: first, a flyable space is selected from the predicted map; this space is then intersected with high-confidence areas and currently unexplored regions to obtain several "high-value" sub-regions (marked with white pixels). The spatial size of these white sub-regions can be used as a metric for potential information gain, where a larger size indicates more new information to be acquired in that direction. Based on this assessment, the UAV can dynamically lock onto the white sub-region with the highest information gain, prioritizing it for flight and exploration. This ensures maximum information collection efficiency in each planning cycle, significantly improving the overall coverage speed and map reconstruction quality.

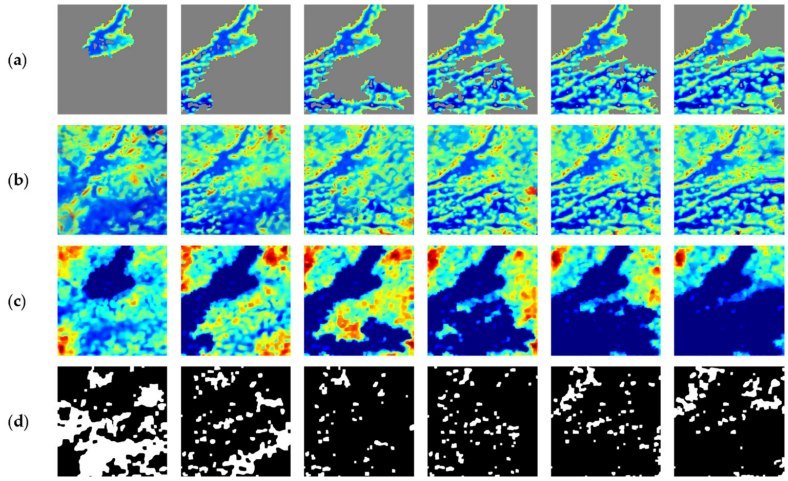


Figure 9. Maps at different exploration rates. (**a**) Exploration map; (**b**) predicted map; (**c**) confidence map; (**d**) flyable map.

Under the same Digital Elevation Map conditions, all baseline methods exhibited long task execution times to achieve full coverage of the explorable region, due to the complex terrain structure and numerous occluded regions. The comparison results are illustrated in Figure 10, where it can be clearly observed that the flight paths generated by the methods differ significantly upon completion of the exploration task. Compared to traditional methods, the strategy proposed in this paper demonstrates higher rationality in path planning decisions, avoiding repeated exploration of already visited regions. This effectively shortens the coverage path length and yields an approximately linear exploration rate (Figure 8). This performance advantage is primarily attributed to the guidance of the predicted map: with its assistance, the UAV can prioritize exploring regions with greater potential information content. In contrast, without a predicted map, baseline methods rely solely on frontier-based information gain for flight decisions. However, the scale of the unknown region behind a single frontier is typically small, which to some extent limits the global exploration rate and reduces path planning efficiency.

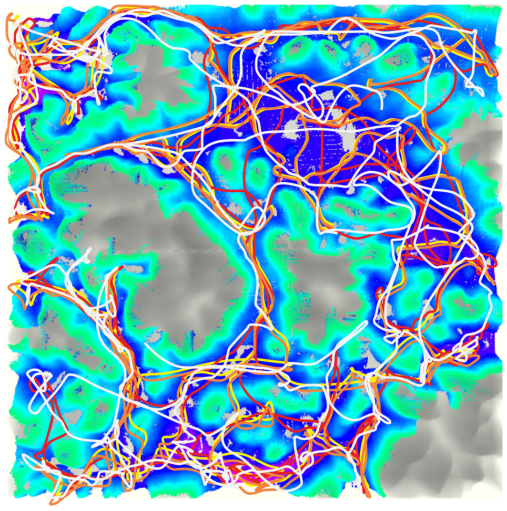


Figure 10. Flight paths obtained by different methods. Proposed method (Yellow); FUEL method (Red); classic method (Orange); FAEP method (White).

6. Conclusions

In this study, we utilized a self-built dataset of elevation maps and corresponding masks to conduct comprehensive training and validation of our prediction network.

The experimental results demonstrate that the proposed model not only recovers terrain details with high fidelity but also generates a confidence assessment map for each voxel—providing a quantitative basis for subsequent decision making. Leveraging this capability, we introduced a simultaneous incremental map prediction-driven framework for UAV autonomous exploration. During execution, the system organically combines traditional global frontier information gain with the area gain of flyable regions derived from the predicted map, enabling the UAV to continuously select the next target point with the maximum potential information gain between consecutive steps. This strategy significantly shortens the total path length and reduces exploration time, while maintaining a near-linear coverage rate across various complex 3D Digital Elevation Map scenarios. This fully validates the guiding value of the predicted map in global path planning. Future work will focus on deploying this framework on a real UAV platform, and further investigating online map prediction and trajectory replanning strategies in the presence of dynamic obstacles—with the goal of enhancing the system's robustness and real-time performance.

Author Contributions: Conceptualization, J.T. and G.J.; methodology, G.J.; software, G.J.; validation, G.J.; formal analysis, S.Z.; data curation, J.Y.; writing—original draft preparation, G.J.; writing—review and editing, J.T. and J.Y.; supervision, S.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 62063035) and the Yunnan Provincial Department of Education Scientific Research Fund Project (Grant No. 2025Y0666).

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, N.; Li, X.; Zhang, K.; Wang, J.; Xie, D. A survey on path planning for autonomous ground vehicles in unstructured environments. *Machines* **2024**, *12*, 31. [CrossRef]
2. Ginerica, C.; Zaha, M.; Floroian, L.; Cojocaru, D.; Grigorescu, S. A vision dynamics learning approach to robotic navigation in unstructured environments. *Robotics* **2024**, *13*, 15. [CrossRef]
3. Tadić, S.; Krstić, M.; Veljović, M.; Čokorilo, O.; Milovanović, M. Risk analysis of the use of drones in city logistics. *Mathematics* **2024**, *12*, 1250. [CrossRef]
4. Hong, Y.; Kim, S.; Kwon, Y.; Choi, S.; Cha, J. Safe and Efficient Exploration Path Planning for Unmanned Aerial Vehicle in Forest Environments. *Aerospace* **2024**, *11*, 598. [CrossRef]
5. Xie, M.R.; Jung, S.Y.; Chen, K.W. Three-Dimensional Drone Exploration with Saliency Prediction in Real Unknown Environments. *Aerospace* **2023**, *10*, 488. [CrossRef]
6. Cieslewski, T.; Kaufmann, E.; Scaramuzza, D. Rapid exploration with multi-rotors: A frontier selection method for high speed flight. In *Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017*; IEEE: Piscataway, NJ, USA; pp. 2135–2142. [CrossRef]
7. Schmid, L.; Pantic, M.; Khanna, R.; Ott, L.; Siegwart, R.; Nieto, J. An efficient sampling-based method for online informative path planning in unknown environments. *IEEE Robot. Autom. Lett.* **2020**, *5*, 1500–1507. [CrossRef]
8. Meng, Z.; Qin, H.; Chen, Z.; Chen, X.; Sun, H.; Lin, F.; Ang, M.H. A two-stage optimized next-view planning framework for 3-d unknown environment exploration, and structural reconstruction. *IEEE Robot. Autom. Lett.* **2017**, *2*, 1680–1687. [CrossRef]
9. Selin, M.; Tiger, M.; Duberg, D.; Heintz, F.; Jensfelt, P. Efficient autonomous exploration planning of large-scale 3-d environments. *IEEE Robot. Autom. Lett.* **2019**, *4*, 1699–1706. [CrossRef]
10. Yamauchi, B. A frontier-based approach for autonomous exploration. In *Proceedings of the 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA'97. 'Towards New Computational Principles for Robotics and Automation', Monterey, CA, USA, 10–11 July 1997*; IEEE: Piscataway, NJ, USA; pp. 146–151. [CrossRef]
11. Deng, D.; Duan, R.; Liu, J.; Sheng, K.; Shimada, K. Robotic exploration of unknown 2d environment using a frontier-based automatic-differentiable information gain measure. In *Proceedings of the 2020 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Boston, MA, USA, 6–9 July 2020*; IEEE: Piscataway, NJ, USA; pp. 1497–1503. [CrossRef]

12. Zhou, B.; Xu, H.; Shen, S. Racer: Rapid collaborative exploration with a decentralized multi-uav system. *IEEE Trans. Robot.* **2023**, *39*, 1816–1835. [CrossRef]
13. Zhang, X.; Wang, J.; Wang, S.; Wang, M.; Wang, T.; Feng, Z.; Zhu, S.; Zheng, E. FAEM: Fast Autonomous Exploration for UAV in Large-Scale Unknown Environments Using LiDAR-Based Mapping. *Drones* **2025**, *9*, 423. [CrossRef]
14. Wang, J.; Yu, Z.; Zhou, D.; Shi, J.; Deng, R. Vision-Based Deep Reinforcement Learning of Unmanned Aerial Vehicle (UAV) Autonomous Navigation Using Privileged Information. *Drones* **2024**, *8*, 782. [CrossRef]
15. Shen, S.; Michael, N.; Kumar, V. Stochastic differential equation-based exploration algorithm for autonomous indoor 3D exploration with a micro-aerial vehicle. *Int. J. Robot. Res.* **2012**, *31*, 1431–1444. [CrossRef]
16. Zhou, B.; Zhang, Y.; Chen, X.; Shen, S. Fuel: Fast uav exploration using incremental frontier structure and hierarchical planning. *IEEE Robot. Autom. Lett.* **2021**, *6*, 779–786. [CrossRef]
17. Zhao, Y.; Yan, L.; Xie, H.; Dai, J.; Wei, P. Autonomous exploration method for fast unknown environment mapping by using UAV equipped with limited FOV sensor. *IEEE Trans. Ind. Electron.* **2023**, *71*, 4933–4943. [CrossRef]
18. Connolly, C. The determination of next best views. In *Proceedings of the 1985 IEEE International Conference on Robotics and Automation, St. Louis, MO, USA, 25–28 March 1985*; IEEE: Piscataway, NJ, USA; pp. 432–435. [CrossRef]
19. Bircher, A.; Kamel, M.; Alexis, K.; Oleynikova, H.; Siegwart, R. Receding horizon "next-best-view" planner for 3d exploration. In *Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016*; IEEE: Piscataway, NJ, USA; pp. 1462–1468. [CrossRef]
20. Witting, C.; Fehr, M.; Bähnemann, R.; Oleynikova, H.; Siegwart, R. History-aware autonomous exploration in confined environments using mavs. In *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018*; IEEE: Piscataway, NJ, USA; pp. 1–9. [CrossRef]
21. Wang, C.; Zhu, D.; Li, T.; Meng, M.Q.H.; De Silva, C.W. Efficient autonomous robotic exploration with semantic road map in indoor environments. *IEEE Robot. Autom. Lett.* **2019**, *4*, 2989–2996. [CrossRef]
22. Zhou, X.; Yi, Z.; Liu, Y.; Huang, K.; Huang, H. Survey on path and view planning for UAVs. *Virtual Real. Intell. Hardw.* **2020**, *2*, 56–69. [CrossRef]
23. Respall, V.M.; Devitt, D.; Fedorenko, R.; Klimchik, A. Fast sampling-based next-best-view exploration algorithm for a MAV. In *Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021*; IEEE: Piscataway, NJ, USA; pp. 89–95. [CrossRef]
24. Huang, J.; Zhou, B.; Fan, Z.; Zhu, Y.; Jie, Y.; Li, L.; Cheng, H. FAEL: Fast autonomous exploration for large-scale environments with a mobile robot. *IEEE Robot. Autom. Lett.* **2023**, *8*, 1667–1674. [CrossRef]
25. Tang, B.; Ren, Y.; Zhu, F.; He, R.; Liang, S.; Kong, F.; Zhang, F. Bubble Explorer: Fast UAV Exploration in Large-Scale and Cluttered 3D-Environments using Occlusion-Free Spheres. In *Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Detroit, MI, USA, 1–5 October 2023*; IEEE: Piscataway, NJ, USA. [CrossRef]
26. Hui, Y.; Zhang, X.; Shen, H.; Lu, H.; Tian, B. Dppm: Decentralized exploration planning for multi-uav systems using lightweight information structure. *IEEE Trans. Intell. Veh.* **2023**, *9*, 613–625. [CrossRef]
27. Li, W.; Lin, Z.; Zhou, K.; Qi, L.; Wang, Y.; Jia, J. Mat: Mask-aware transformer for large hole image inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022*; IEEE: Piscataway, NJ, USA; pp. 10758–10768.
28. Liu, W.; Cun, X.; Pun, C.M.; Xia, M.; Zhang, Y.; Wang, J. Coordfill: Efficient high-resolution image inpainting via parameterized coordinate querying. In *Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 7–14 February 2023*; AAAI Press: Washington, DC, USA; pp. 1746–1754. [CrossRef]
29. Ma, X.; Zhou, X.; Huang, H.; Jia, G.; Wang, Y.; Chen, X.; Chen, C. Uncertainty-aware image inpainting with adaptive feedback network. *Expert Syst. Appl.* **2024**, *235*, 121148. [CrossRef]
30. Zheng, H.; Zhang, Z.; Wang, Y.; Zhang, Z.; Xu, M.; Yang, Y.; Wang, M. GCM-Net: Towards effective global context modeling for image inpainting. In *Proceedings of the 29th ACM International Conference on Multimedia, Chengdu, China, 20–24 October 2021*; Association for Computing Machinery (ACM): New York, NY, USA; pp. 2586–2594. [CrossRef]
31. Dong, Q.; Cao, C.; Fu, Y. Incremental transformer structure enhanced image inpainting with masking positional encoding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022*; IEEE: Piscataway, NJ, USA; pp. 11358–11368.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.