SenseExpo: Lightweight Neural Networks for Efficient Autonomous Exploration and Scene Prediction

Haohua Que[1], Haojia Gao[2], Mingkai Liu[3], Hoiian Au[2], Handong Yao[1,*], Fei Qiao[2,*]

***Abstract*— This work presents *SenseExpo*, a frontier-based exploration framework powered by a lightweight local map predictor that combines GAN training, a Transformer encoder, and Fast Fourier Convolution. Our smallest model (709k parameters) surpasses much larger baselines (U-Net 24.5M, LaMa 51M) on KTH dataset, achieving PSNR 9.026 and SSIM 0.718, and shows strong cross-domain robustness on HouseExpo (FID 161.55). Leveraging predicted free space for goal selection, *SenseExpo* accelerates exploration, reducing time by 67.9% on KTH dataset and 77.1% on MRPB 1.0 relative to MapEx, while sustaining high coverage and accuracy. Delivered as a plug-and-play ROS node, it is practical for resource-constrained robots and easy to integrate into existing navigation stacks.**

I. INTRODUCTION

Autonomous exploration systems play a pivotal role in numerous fields, such as planetary exploration [1] and environmental monitoring [2]. These systems are crucial for enabling robots to navigate and operate in unknown or partially known environments without human intervention. However, one of the significant challenges faced by robots in these settings is the ability to efficiently explore and map their surroundings in real-time.

For robots, the exploration of unknown environments remains a complex challenge due to the limited sensor field of view (FOV) and the computational cost of generating maps in real-time. *So how to make robots explore smarter with less cost?* Biological systems hold the key. Unlike robots, humans and animals navigate unknown spaces by building mental maps that extend far beyond their immediate view, allowing them to make better decisions and anticipate future states [3]. So we think robots also can obtain such 'memory' ability by pre-trained models to explore more efficiently.

Prediction-aided exploration has seen progress: MapEx [4] uses LaMa [5] for global map completion, and Katyal et al. [6] apply U-Net for prediction. Yet these approaches are computationally heavy and hard to deploy on resource-limited robots, and they often generalize poorly under domain shift. We propose *SenseExpo*, a frontier-based framework powered by a lightweight local predictor (as small as 709k parameters) operating on robot-centric observations. The predictor integrates GAN training [7], a Transformer encoder [8] for long-range structure, and Fast Fourier Convolution [9] for efficient global context. This design delivers accurate free-space predictions at low cost, enabling faster, more informed goal selection and practical deployment.

[1]University of Georgia, Athens, USA.
[2]Tsinghua University, Beijing, China.
[3]Peking University, Beijing, China.
*Corresponding authors: Handong Yao (email: yaohandong@tsinghua.edu.cn) and Fei Qiao (email: qiaofei@tsinghua.edu.cn).

II. PROBLEM DEFINITION

Consider an unknown environment modeled as a continuous 2D space $E \subseteq \mathbb{R}^2$. The ground truth of this environment is represented by an occupancy grid map, an unknown 2D discrete matrix $M \in \{0,1\}^{H \times W}$. This matrix is a discretization of the continuous environment E, where the space is divided into a grid of uniform (square) cells. Each cell in M corresponds to a specific area in E. A value of '1' in a cell signifies that the corresponding area is **occupied** by an obstacle, while a value of '0' signifies it is **free** space. In this environment, n mobile robots $\{R_i\}_{i=1}^n$ are deployed, and each robot satisfies the following conditions:

- The pose of each robot R_i at time t is denoted as $P_i^t \in E$, which exists in the continuous space E. At each time step Δt, it moves a fixed distance L along the planned path, such that the movement satisfies the constraint:

$$\|P_i^{t+\Delta t} - P_i^t\|_2 = L \tag{1}$$

- Each robot R_i is equipped with a laser scanner with an effective radius r, and at time t, it can obtain a local observation of the obstacle map $m_i^t = S(E, P_i^t, r, \delta)$. These observations are used to build a **global observation map** M_i^t, which contains only ground-truth information from the sensor. The side length of the square local observation map is $\sigma = \delta \cdot 2r$, where δ is a unitless scaling factor that determines the size of the observation window relative to the sensor's diameter.
- Each robot R_i is also equipped with a map prediction network $f_\rho : m_i^t \to \hat{m}_i^t$ where $\hat{m}_i^t$ is the predicted local obstacle map of robot R_i at time t, and the network has a parameter size of ρ.
- Each robot maintains a **global predicted map** $\hat{M}_i^t$, which is constructed by concatenating the local predicted maps $\{\hat{m}_i^\tau\}_{\tau=0}^t$ over time. This map layer consists solely of predictions generated by the model.

In this model, the robot moves within a continuous physical space, while its perception and representation of that space for planning and prediction are managed through the discrete grid map M. To validate the core contributions of our predictive model, the experimental evaluation in this paper is focused on the foundational single-robot ($n = 1$) case.

The optimization goal is to minimize the model parameter size ρ and the total exploration time T while ensuring that

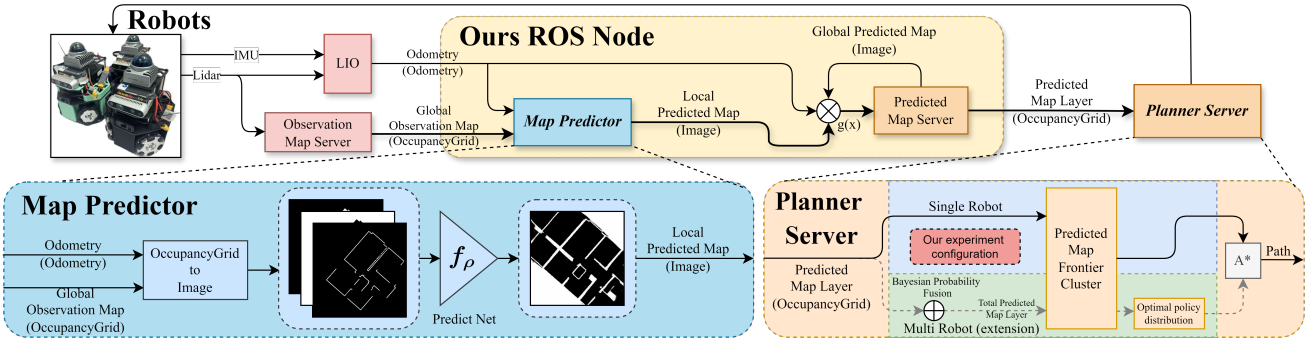

Fig. 1. **The Overview of *SenseExpo*.** The Map Predictor uses odometry and global observation map obtained from the robot to produce local predicted map. The Predicted Map Server then concatenates the local predicted map and the global predicted map to form a predicted map layer. For a potential multi-robot extension, the Planner Server would be responsible for fusing maps from each robot (if there are multi robots) into a total predicted map layer and outputs the path to the navigation point.

the total predicted map $\hat{M}^T_{\text{Total}}$ is as close as possible to the ground truth map M:

$$\min(\theta_1 \cdot \rho + \theta_2 \cdot T + \theta_3 \cdot \|M - \hat{M}^T_{\text{Total}}\|_1) \tag{2}$$

Where:

- When there is only one robot ($n = 1$), the total predicted map is $\hat{M}^T_{\text{Total}} = \hat{M}^t_0$.
- When there are multiple robots ($n \neq 1$), the total predicted map is $\hat{M}^T_{\text{Total}} = \mathscr{F}(\{\hat{M}^T_i\}_{i=1}^n)$, where $\mathscr{F}$ is the **map fusion operator**. For cells with conflicting information (e.g., seen differently by different robots), it employs a probabilistic update mechanism.

III. Methodology

We present the overall pipeline for efficient exploration in unknown environments (Fig. 1). First, we introduce our lightweight local map prediction model and its GAN-based training procedure. Next, we detail the frontier-based exploration strategy, focusing on the single-robot case.

A. Map Prediction

Most map prediction algorithms struggle with deployment due to large parameter sizes and poor generalization, especially in robots with limited training data or varying environments [10]. Common approaches like U-Net or LaMa [5] often fail under domain shifts. Our method addresses these limitations by introducing a lightweight local map predictor that enhances generalization while significantly reducing parameters. As shown in Fig. 2, we optimize the U-Net architecture by reducing channels, removing redundant layers, and integrating Transformer Encoder [8] and Fast Fourier Convolutions (FFC) [9] for large-scale perception. Dropout layers further improve robustness. These enhancements enable accurate map predictions with minimal computational cost, making the model suitable for resource-constrained robots. To represent the three distinct states of the local map, we encode it into three binary channels. The **'Free'** and **'Obstacle'** channels mark cells that are explicitly known from sensor data. As can be inferred from Fig. 2, the **'Uncertain'** channel is then determined as the logical complement, marking all cells that are neither known to be free nor occupied ($Uncertain = \neg(Free \vee Obstacle)$).

B. Autonomous Exploration based on Map Prediction

Our exploration strategy leverages the classical frontier-based [11] search algorithm to ensure systematic coverage. The primary innovation, however, is not in redefining this paradigm, but in how our novel map predictor (Fig. 2) provides high-quality, long-range predictive information. This information transforms the goal selection process, allowing the robot to make more informed decisions and drastically improving exploration efficiency compared to using only currently observed data. The traditional Frontier-based algorithm searches for boundary points between the uncertain and free areas. However, the map predicted by our model does not explicitly include an uncertain region. If we define thresholds to partition the free, uncertain, and obstacle regions, the uncertain area will mostly be concentrated at the edges of the obstacles. This makes it unsuitable for use as a navigation point selection area.

To address the problem mentioned above, we designed a frontier-based search algorithm [11] based on local predicted maps. To accommodate potential multi-robot scenarios, after each robot updates its local predicted map $\hat{m}^t_i$, a fusion mechanism such as Bayesian Updating could be used to obtain the total predicted map $\hat{M}^t_{\text{Total}}$. For the single-robot case, which is the focus of our experiments, the total predicted map is simply the local predicted map $\hat{m}^t_i$. The boundary points between the free region (red region) in $\hat{M}^t_{\text{Total}}$ are extracted as shown in Fig. **??**(b). Thus, we can derive the quality for each frontier $\{\mu_j\}_{j=1}^k$. We compute the cost for each frontier's central point using the following equation (Eq. 3):

$$\text{cost}_{i,j} = \left| \|\mu_j - P^t_i\|_2 - \sigma \right| \tag{3}$$

where $i \in [1, n]$ and $j \in [1, k]$.

We associate each frontier with its corresponding area, allowing the robot to focus on exploring larger regions in the unexplored areas. The exploration process continues iteratively, with the robot moving to the highest-utility frontier at each step. The process is considered complete and terminates

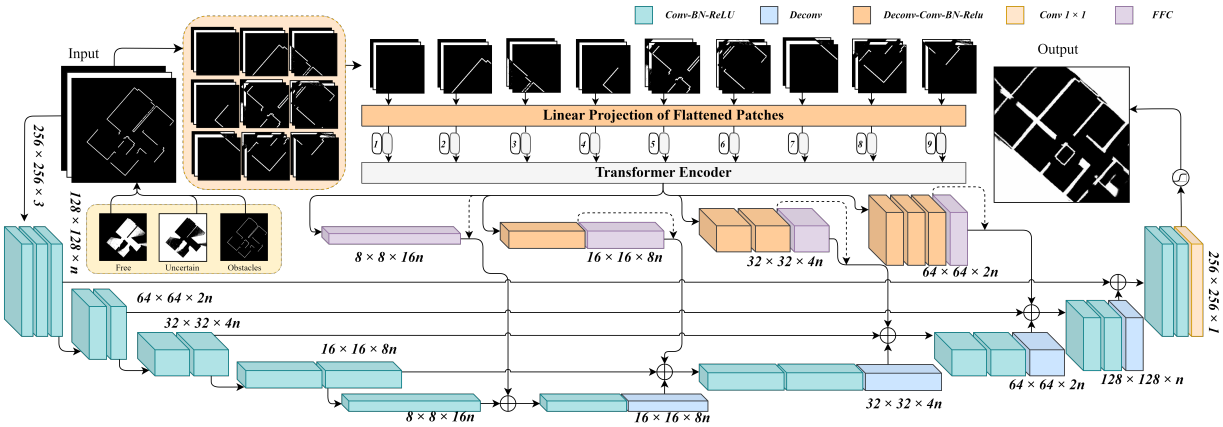

Fig. 2. **Complete Architecture of the Map Prediction Network.** The input, which has 3 channels (free, uncertain, obstacles), is processed through the network to produce a single-channel grayscale image output, where each pixel value represents the probability of occupancy. Simultaneously, input patches are passed through a Transformer Encoder and then Fast Fourier Convolution, with the outputs concatenated to the feature maps in the U-Net.

when no new frontiers can be identified from the map, which signifies that all reachable areas have been fully explored or predicted. This termination condition is explicitly handled in our goal selection algorithm .

IV. Experiment and Discussion

A. The evaluation of Map Prediction

TABLE I
Comparison of Model Performance and Robustness

Exp.	Size/Method	Network	Param	PSNR ↑	SSIM ↑	LPIPS (VGG) ↓	LPIPS (Alex) ↓	FID ↓
Perf. Comp.	Small	U-Net	1.1M	7.707	0.670	0.280	0.326	56.904
		Ours	**709K**	**9.026**	**0.718**	**0.246**	**0.283**	**42.353**
	Medium	LaMa	27.0M	7.428	0.654	0.309	0.405	169.803
		U-Net	6.1M	8.267	0.695	0.263	0.305	47.097
		Ours	**5M**	**9.189**	**0.724**	**0.246**	**0.282**	**38.047**
	Large	Big LaMa	51.0M	6.508	0.608	0.334	0.408	117.863
		U-Net	24.5M	8.539	0.711	0.256	0.295	41.172
		Ours	**20.6M**	**9.209**	**0.731**	**0.248**	**0.280**	**33.160**
Pred. Rob.	Global	LaMa	27.0M	3.597	0.428	0.457	0.751	409.836
		Big LaMa	51.0M	3.824	0.444	0.461	0.697	397.363
	Local	LaMa	27.0M	4.296	0.506	0.321	0.452	261.365
		Big LaMa	51.0M	5.171	0.570	0.322	0.422	188.121
		Ours	**709K**	**5.673**	**0.605**	**0.294**	**0.377**	**161.548**

1) Experimental Results: Based on the above experimental environment and evaluation indices, we conducted several groups of comparison experiments, the experimental results are shown in Fig. 3 and Tab. I. We evaluated our model against baselines in terms of map prediction quality, structural similarity, and feature distribution distance. Generally, models with more parameters performed better. For instance, U-Net's PSNR improved from 7.707 (small) to 8.538 (large), while our model achieved 9.209 PSNR with 20.6M parameters and 9.026 with just 709K parameters.

Notably, our lightweight model (709K) outperformed the Big LaMa-Fourier model (51.0M), achieving a 38.7% higher PSNR and 18.1% higher SSIM, with significantly lower FID. It also performed comparably to the large U-Net (24.5M), showing better PSNR and SSIM despite slightly higher FID. This strong performance stems from our hybrid architecture, combining a Transformer Encoder for spatial reasoning and FFC for global context, integrated via a U-Net backbone. This design enables *SenseExpo* to deliver accurate and coherent maps with far fewer parameters.

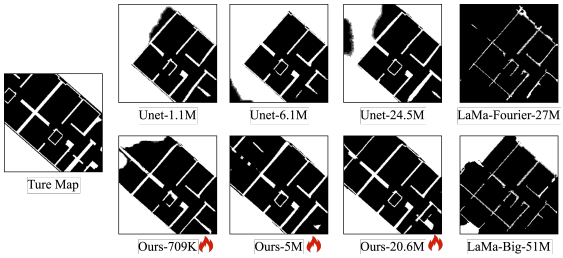

Fig. 3. **The example of map prediction.** Comparison of ground truth map and predicted maps. Our models outperform U-Net, which tends to cover a more limited area, as well as the LaMa model, whose predictions show structural inaccuracies and lower clarity.

B. The evaluation of Prediction Robustness

To evaluate prediction robustness under domain shift, experiments were conducted on the HouseExpo dataset [12], an unseen environment, using local and global map prediction tasks. Models trained on KTH [13] were tested for cross-domain generalization. As shown in Tab. I, local prediction consistently outperformed global prediction, with Big LaMa-Fourier reducing FID by 32.6% (from 397.36 to 188.12). This indicates localized feature extraction better handles domain discrepancies. Our compact model (709K parameters) achieved superior robustness, with an FID of 161.55 and LPIPS (Alex) of 0.377, outperforming larger LaMa variants (27M–51M parameters) by 14.1% and 10.7%, respectively. These results highlight the efficiency of lightweight models in cross-domain scenarios.

C. The comparison of Exploration Efficiency

We compared the exploration efficiency of the classic Frontier-based method, *SenseExpo* with our 709k model, and

MapEx [4] using 3 big LaMa models (51M) on KTH and MRPB 1.0 datasets. All models were trained solely on KTH, with *SenseExpo* predicting free space and MapEx predicting obstacles. Results are shown in Fig. 4 and Fig. 5. In KTH, *SenseExpo* achieved higher coverage than both methods, reducing exploration time by 92.5% compared to Frontier-based and 67.9% compared to MapEx for similar coverage. Accuracy remained above 90%, improving by 2.5% over MapEx. On MRPB 1.0, *SenseExpo* reduced exploration time by 77.1% while achieving higher coverage and an 8% accuracy gain. These improvements stem from its ability to predict expansive free space, enabling efficient long-horizon goal selection and faster map coverage.

Fig. 4. **Comparison of exploration efficiency in KTH and MRPB 1.0 Dataset.** When T reaches 500, our model has completed the prediction of the entire map, with no remaining uncertain areas, which are represented in gray.

V. Conclusion

SenseExpo couples a lightweight local map predictor with frontier-based planning, providing accurate free-space forecasts at very low compute cost. Across datasets it outperforms heavy baselines and is delivered as a plug-and-play ROS node for resource-constrained robots. Future work will target dynamic scenes, richer inputs (e.g., RGB), and broader layouts with multi-robot fusion.

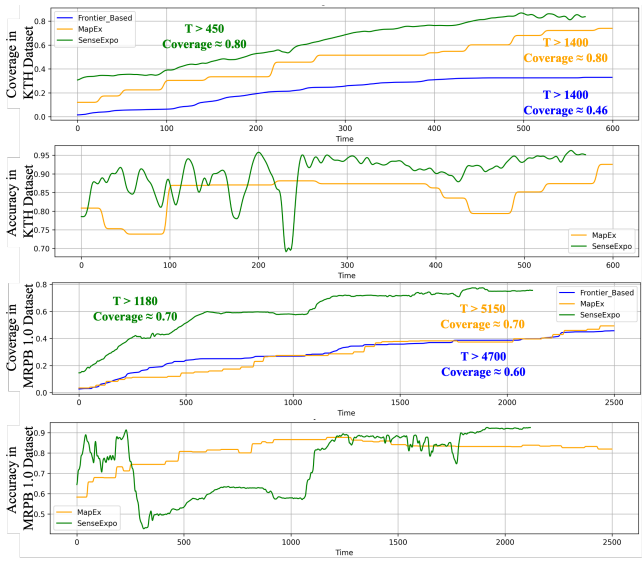

Fig. 5. **Comparison of models in terms of coverage and accuracy.** Our model demonstrated higher coverage and accuracy in both KTH and MRPB 1.0 datasets.

References

[1] T. Vögele, R. Sonsalla, A. Dettmann, F. Cordes, M. Maurus, R. Dominguez, and F. Kirchner, "Robotics concepts for future planetary exploration missions," in *Space Robotics: The State of the Art and Future Trends*. Springer, 2024, pp. 483–514.

[2] M. Dunbabin and L. Marques, "Robots for environmental monitoring: Significant advancements and applications," *IEEE Robotics & Automation Magazine*, vol. 19, no. 1, pp. 24–39, 2012.

[3] E. C. Tolman, "Cognitive maps in rats and men." *Psychological review*, vol. 55, no. 4, p. 189, 1948.

[4] C. Ho, S. Kim, B. Moon, A. Parandekar, N. Harutyunyan, C. Wang, K. Sycara, G. Best, and S. Scherer, "Mapex: Indoor structure exploration with probabilistic information gain from global map predictions," *arXiv preprint arXiv:2409.15590*, 2024.

[5] R. Suvorov, E. Logacheva, A. Mashikhin, A. Remizova, A. Ashukha, A. Silvestrov, N. Kong, H. Goka, K. Park, and V. Lempitsky, "Resolution-robust large mask inpainting with fourier convolutions," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2022, pp. 2149–2159.

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

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[9] L. Chi, B. Jiang, and Y. Mu, "Fast fourier convolution," *Advances in Neural Information Processing Systems*, vol. 33, pp. 4479–4488, 2020.

[10] E. Zwecher, E. Iceland, S. Y. Hayoun, A. Revivo, S. R. Levy, and A. Barel, "Deep-learning-aided path planning and map construction for expediting indoor mapping," *arXiv preprint arXiv:2011.02043*, 2020.

[11] B. Yamauchi, "A frontier-based approach for autonomous exploration," in *Proceedings 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA'97.'Towards New Computational Principles for Robotics and Automation'*. IEEE, 1997, pp. 146–151.

[12] T. Li, D. Ho, C. Li, D. Zhu, C. Wang, and M. Q.-H. Meng, "Houseexpo: A large-scale 2d indoor layout dataset for learning-based algorithms on mobile robots," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5839–5846.

[13] A. Aydemir, P. Jensfelt, and J. Folkesson, "What can we learn from 38,000 rooms? reasoning about unexplored space in indoor environments," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 4675–4682.