Imagination-augmented Navigation Based on 2D Laser Sensor Observations

Zhengcheng Shen¹, Linh Kästner¹, Magdalena Yordanova¹, and Jens Lambrecht¹

Abstract—Autonomous navigation of mobile robots is an essential task for various industries. Sensor data is crucial to ensure safe and reliable navigation. However, sensor observations are often limited by different factors. Imagination can assist to enhance the view and aid navigation in dangerous or unknown situations where only limited sensor observation is available. In this paper, we propose an imagination-enhanced navigation based on 2D semantic laser scan data. The system contains an imagination module, which can predict the entire occupied area of the object. The imagination module is trained in a supervised manner using a collected training dataset from a 2D simulator. Four different imagination models are trained, and the imagination results are evaluated. Subsequently, the imagination results are integrated into the local and global cost map to benefit the navigation procedure. The approach is validated on three different test maps, with seven different paths for each map. The quality and numeric results showed that the agent with the imagination module could generate more reliable paths without passing beneath the object, with the cost of a longer path and slower velocity.

I. INTRODUCTION

In today’s navigation systems, accurate and complete sensor information that can observe a wide area of the environment is crucial to ensure safe and reliable navigation [1], [2] and a variety of different sensors are deployed to ensure safety robustness and redundancy. However, this can not always be guaranteed. Since the laser can not go through solid objects, there will be occluded areas [3] which may not be seen by the robot. Moreover, the installation height of the sensor will also strongly limit the observation. For instance, for 2D laser sensors in a lower position, only the legs of the chair or table will be detected. Moreover, in hazardous environments, or when the connection is low, precious information could go missing. In these scenarios, imagination could substantially help to recount old memories to rebuild and/or augment the field of view and aid for navigation tasks [4]. Additionally, with the imagination enhanced observation, semantic information can be leveraged to achieve high-level path planning tasks [5], such as to avoid sensitive areas, e.g. the area under tables or between chairs in specific situations like in Fig. 1. The potential benefits of the imagination for navigation are further explored in the paper. In this paper, we propose an approach to train and deploy an imagination module to enhance navigational safety and efficiency based on 2D sensor data observations. We train four different neural network models to detect and imagine real-world objects from limited sensor observations. A dataset, which contains the observation and 4 paired ground truths for different models, is extracted from the arena-ros for the training of the imagination module. The input of the imagination unit will be a 2D local egocentric map. The resulting observation is directly merged into the occupancy grid maps used by global and local planners.

The main contributions of this work are the following:

• Proposal of an imagination module to semantically detect and imagine real-world objects based solely on 2D laser scan data. Various filters and preprocessing steps are introduced to convert an egocentric 2D laser scan into a semantic top-down map for training. Subsequently, four different network models are trained.
• Integration of the imagination module into the robot operating system (ROS). The imagination results will directly be merged into the occupancy grid maps used by global and local planners.
• Evaluation and validation of the imagination models on three different maps

The paper is structured as follows. Sec. II begins with related works. Subsequently, the methodology is presented in Sec. III. Sec. IV presents the results and discussion. Finally,
II. RELATED WORKS

Autonomous navigation of mobile robots has been subject to a vast amount of research works [6], [7], [8], [9]. The majority of the work assumes complete sensor information, which is not always guaranteed in real-world settings. Thus, leveraging semantic information to additionally improve navigation has become an important aspect and has been explored by a variety of different research works in recent years. Everett et al. [10] proposed a navigation system that learns to navigate based on specific context learned by analyzing terrain context from semantic grid maps. Narasimhan et al. [11] utilize semantic maps using amodal top-down maps and learn the architectural shape of rooms. The researchers demonstrate that the agent could learn to navigate to specific rooms in unknown homes without seeing the entire environment. Works by Chen et al. [12], [13], [14] or Yu et al. [15] or Krantz et al. [16] have explored semantic navigation for high-level tasks such as finding specific objects, navigating and finding rooms based on visual and audio semantics. Another work that focuses on high-level semantic navigation is proposed by Batra et al. [17]. The researchers present methods and benchmarks to assess and evaluate navigation based on high-level information of surrounding objects. They conclude and stress the importance of merging semantic information to contribute to intelligent navigation.

Another aspect to be considered is that oftentimes, information is not perfectly given or the sensor does not work properly due to limited observations. Thus, semantic scene completion has also gained increasing interest in the research community. Song et al. [18] proposed an approach to generate semantic 360 panoramas from partial observations of less than 50 percent. The authors leveraged a convolutional neural network to learn priors from a large-scale dataset. Other researchers combined this semantic scene completion with navigation. Liang et al. [19] proposed SSCNav to complete scenes with unknown obstacles to improve a DRL agent. The authors compute the agent’s observation online by deploying a semantic scene completion module to enhance the agent’s observations. The researchers demonstrated efficient and more reliable trajectories in unknown scenes. Georgakis et al. [20] proposed an approach to actively learn to generate semantic maps of the environment outside the agent’s field of view. The authors project the depth and RGB image into an egocentric top-down view to train a model that learns to predict semantics of unobserved areas in a supervised manner. The model could learn to predict semantically plausible scenes such as chairs around tables. Most of the aforementioned papers deal with photo-realistic 3D environments and visual information as they contain the most information. However, it is not always possible to have a visual sensor, and processing this data is computationally intensive. On that account, our work focuses on the usage of more lightweight sensor data such as 2D laser scan data. we utilize 2D laser scan data and enhance it with semantic information to improve navigation.

III. METHODOLOGY

In this chapter, we present the methodology of our proposed framework. The main objective is to detect and imagine different object classes solely from 2D laser scan data and directly integrate this additional information for path planning. Our system design is illustrated in Fig. 2. We utilize a 2D simulation environment arena-rosnav, which simulates the robot and laser scan data. Using the simulator, training data is first generated and prepared for training. More specifically, the 2D egocentric laserscan information is transformed into a semantic top-down view map using the ground truth map. The network then learns to imagine the complete shape of the object from limited sensor observation and outputs an occupancy grid. This additional occupancy information is then merged with the existing global and local cost map of the ROS navigation stack, which is used for the global and local planner respectively. In the following, each module is described in more detail.

A. Simulation Platform

This work is utilizing arena-rosnav [21] an efficient 2D simulation environment for navigation integrated into ROS. It is based on the open-source package Flatland, which allows simulation of dynamic and static 2D objects and implementation of ROS-based functionalities like sensor data, navigation models, etc. In this work, we spawn different chairs and tables that are classified using a specific ID. This ID is later used to semantically generate the semantic laser scan information. The middleware is ROS and respective information is gathered through topics and services. Since it’s full integration into ROS, conventional planning methods can be used, which will be later on utilized to validate our approach. Furthermore, arena-rosnav also consists of tools to generate new worlds and scenarios, record data, and test navigation approaches. In this paper, a Turtlebot3 robot with a laser height of 0.2m is deployed. Several different maps consisting of static obstacles such as tables and chairs are spawned. Due to the low height of the laser scan sensor, the robot is only able to perceive the legs of objects like tables and chairs. Within this paper, our objective is to utilize the proposed imagination module to augment the limited observation space using imagination.

B. Data Acquisition and Preprocessing

The capability to get the semantic map from imagination is proved by [22]. In this paper, we focus on the navigation influence created by the imagination. To integrate the system with arena-rosnav, the 2D laser scan data will be generated according to the following steps. 2D laser scan data is acquired using ROS topics. The simulated robot will provide a stream of 2D sensor data together with its position on the map. Within the ground truth map, all obstacles are assigned
with a unique ID, which specifies their class. Using this information, each laser beam can be assigned an additional semantic id of the object it is perceiving. Thus, the egocentric laser scan observation of the robot can be semantically labeled to provide a top-down semantic map. Therefore, the ROS scan message type is extended by an ID specifying the respective class. This will then be used by the imagination unit. The procedure is illustrated in the pre-processing unit in Fig. 2. To consider noise and improve the semantic object detection capability as presented in [22], we cut the laser sensor into a 60x60 top view image, which corresponds to a 3-meter by 3-meter area. Using imagination, the unseen area can be anticipated. For the imagination beyond the scope, we have to extend the 60x60 sensor observation into a 100x100 sensor observation, as illustrated in Fig. 3. We fill the ID value of unknown environments for the expanded area. For this expanded observation space, a different model will be trained. The result of our imagination module is a probability occupancy map, which is used to enhance the existing global and local occupancy maps of the navigation stack. A threshold $\theta$ will be set to control the intensity of the imagination, and all the points with a value larger than $\theta$ are labeled as occupied. The threshold in this paper is set to 0.2. The final result will be used to calculate the cost map used by the path planners.

| Table I: Features From Unet Encoder |
|-------------------------------------|
| Features | Channels | Number of Features | Size |
| $x_1$    | 32       | 32                 | 60x60 |
| $x_2$    | 32       | 64                 | 30x30 |
| $x_3$    | 32       | 128                | 15x15 |
| $x_4$    | 32       | 256                | 7x7   |
| $x_5$    | 32       | 256                | 3x3   |

2) Post-process and Filters: Another important part of the imagination Unit is the post-processing module depicted in Fig. 4. Like humans, the imagination is usually accompanied by noise from relevant neuron activation. Sometimes it will suggest that there are objects even in free space. The noise from irrelevant neurons can be significantly reduced by deploying filters as shown in our previous works [22]. The filter is created based on the sensor data and no ground truth data is involved. For each local laser scan map $M_{scan}$, the following procedure is executed:

$$\text{filter} = \text{conv}(M_{scan}, \text{kernel}^{\text{gauss}})$$

where $\text{kernel}^{\text{gauss}}$ is a 31x31 gaussian kernel where:

$$\text{kernel}^{\text{gauss}}_{x,y} = \frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}$$

The result of our imagination module is a probability occupancy map, which is used to enhance the existing global and local occupancy maps of the navigation stack. A threshold $\theta$ will be set to control the intensity of the imagination, and all the points with a value larger than $\theta$ are labeled as occupied. The threshold in this paper is set to 0.2. The final result will be used to calculate the cost map used by the path planners.
Fig. 4: Detailed structure of the imagination unit. The imagination unit consists of an Unet encoder, an Unet decoder, the imagination mask generator, and the post-processing unit. The input data, 2D semantic laser data will first go through the Unet. The raw imagination output, combined with an input-based imagination mask will be then forwarded to the post-processing module for the final result.

D. Training Procedure

The training procedure will be taken offline to ensure efficiency. First, the training dataset will be created in the simulation environment. We have created three training maps within the simulation environment and collected training data from them. The training data contain 2 different categories, namely chair, and table. Totally 10 variants of the objects are included. The points from which training data is sampled are always in near proximity to the obstacles and shown in Fig. 5. The semantic laser scan input $M_{\text{scan}}$, will be generated by the semantic labeling procedure described in section II-B. For each $M_{\text{scan}}$, a paired ground truth for the local occupancy maps will be recorded. In this paper, we evaluated two different sizes for the local costmap, namely 60x60 and 100x100, while the robot is always in the center of the local map. For the training, we also have another set of data called extended ground truth. The object in the extended ground truth will be dilated by 5 pixels. An example is shown in Fig. 6.

Fig. 5: Data sampling points for training data acquisition. The blue points indicate the positions at which training data was sampled. The sampling points are close to the objects to gather a sufficient dataset.

To mitigate the issue of an imbalanced dataset due to more free space than occupied space in the dataset, the loss function and the weight matrix $W_{\alpha}$ described in [22] is used. Once the dataset is processed, it is given as input to the model. As we said in section II-B, our objective is to extend the imagination beyond the unseen area. The model is trained using different sizes of the input laser scan map.

E. Integration into the ROS navigation stack

To enhance observations and improve navigation, the imagination unit is integrated into the ROS navigation stack consisting of a global and local planner. Both planners work with ROS costmaps which are occupancy grid maps and represent areas that are not accessible by the robot. Especially the local costmap is dependent on the laser scan observations and is generated locally to react for dynamic changes and unknown areas not covered or considered by the global costmap. Both costmaps are consumed by a global and local planner respectively to plan and execute a path for the robot.
Fig. 7: Imagination results for three different scenarios from four different models. The area covered by the blue mask is the imagination part. The results from four different models are trained on the different ground truth, namely 60x60, 60x60 Extended, 100x100, 100x100 Extended. The imaginations are based on the observation and close to the real objects. The models trained with extended ground truth (denoted with ‘Ext’) predict the object more aggressively.

| Imagination | 60x60 | 60x60 Ext | 100x100 | 100x100 Ext |
|-------------|-------|-----------|---------|-------------|
| Imagination pixels | 32570 | 36529 | 32391 | 37754 |
| In object imagination pixels | 23616 | 24322 | 21737 | 26110 |
| Out object imagination pixels | 9954 | 12207 | 10654 | 11644 |

Table II: The number of imagination pixels from different imagination models

Fig. 8: Imagination results together with ground truth image, laser scan, semantic laser scan, preprocessed mask, and imagination result for different example local areas.

A. Imagination Results

Fig. 8 illustrates the imagination masks, the raw imaginations (pre-processed), the imagination results (post-processed) together with the respective 2d laser scan observations, and the ground truth of three local areas of the test map. It is observed that the raw data created by the Unet decoder can already provide an accurate result. The mask is generated based on the raw laser data, which makes the prediction around the laser scan more reliable. The results demonstrate that the network can successfully predict the shape and the angle of the object based on the semantically labeled laser data.

Fig. 7 illustrates the imagination results of different scenarios and different models. We can observe that the imagination shows the correct position of the objects and enrich the sensor observation. To understand the difference of the imagination behavior between the different models, we calculate the total number of pixels of the imagination for three different evaluation scenarios and split them into two different classes, namely the in-object imagination pixels and the out-object imagination pixels. The results are listed in Table II. The model trained with extended ground truth predicts more pixels outside the object, while the model trained with 100x100-Ext ground truth data predicts significantly more pixels than other models. This is due to the fact that the model performs the imagination within a 100x100m area, although the observation is only limited in the size of 60x60m.

IV. Results and Evaluation

We evaluated different agents with and without different imagination models to assess the effect of our different imagination models. A simulation test containing 16 paths in three different test scenarios is performed.

to avoid occupied areas. Hence, to improve these planners, the costmaps are enhanced with information gathered from our imagination unit. To validate our approach, we utilize a A* global planner [24] and a TEB local planner [25].
extended ground truth also encourages the model to classify the points close to the object also as occupied.

B. Navigation Results

To qualitatively assess the planners’ behavior using the additional information of the imagination unit, we plotted the trajectories of the planners with and without imagination. Fig. 9 depicts the paths created by the different agents. It is clearly observed that the agent without imagination module will go through the object for the shortest path. At the same time, all the agents with an imagination module choose the path without driving beneath the objects. The imagination can significantly influence the path chosen by the planner. Based on the local cost map and global cost map with the imagination, it has the ability to observe a wider occupancy from limited sensor observations, which can be essential for high-level semantic tasks such as avoiding sensitive or hazardous areas. Although the sensor data suggests that the path is safe to go, the imagination module will stop the agent from navigating through the object. Although the imagination areas from the different models have differences, the paths from these models have no significant changes.

To compare the trajectories and kinematics of all models, we also perform a quantitative evaluation for the paths. Table III lists the results. It shows that the length of the path has increased with the imagination module, since it will not go through the object. On the other hand, the average velocity of the planner is significantly lower. The possible reason is that the imagination interference with the path planner so it has to find a new sub-goal more frequently.

V. Conclusion

In this paper, we proposed an imagination method to enhance the robot’s view based on 2D laser scan data. Therefore, we proposed data pre-and post-processing methods to semantically label 2D laser scan data, map them into a semantic top-down map and filter out imagination noise. Subsequently, we trained an Unet-based neural network with different input sizes and demonstrated that the model could learn to detect and imagine objects based solely on 2D laser scan data. The resulting occupancy information is directly merged with the cost map of the ROS navigation stack to improve the planner’s safety. For evaluation, we integrated the imagination module with conventional planning algorithms and tested the systems on three different maps. Results show that the imagination module could influence the planner behavior by providing a wider field of view and avoiding the specified areas. This could be essential for robots with a limited observation space to maintain safe navigation or for high-level tasks such as avoiding pre-defined objects. Future works will focus on the integration of more obstacles for detection and imagination, as well as more extensive evaluations of different planners and maps. Moreover, we aspire to validate our approach on real robots.

APPENDIX

The code is publicly available at https://github.com/ignc-research/2d-imagination

REFERENCES

[1] S. Robla-Gómez, V. M. Becerra, J. R. Llata, E. Gonzalez-Sarabia, C. Torre-Ferrero, and J. Perez-Oria, “Working together: A review on safe human-robot collaboration in industrial environments,” IEEE Access, vol. 5, pp. 26754–26773, 2017.

[2] G. Fragapane, D. Ivanov, M. Peron, F. Sgarbossa, and J. O. Strandhagen, “Increasing flexibility and productivity in industry 4.0 production networks with autonomous mobile robots and smart intralogistics,” Annals of operations research, pp. 1–19, 2020.
[3] S. Song, F. Yu, A. Zeng, A. X. Chang, M. Savva, and T. Funkhouser, “Semantic scene completion from a single depth image,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1746–1754.

[4] S. K. Ramakrishnan, Z. Al-Halah, and K. Grauman, “Occupancy anticipation for efficient exploration and navigation,” in European Conference on Computer Vision. Springer, 2020, pp. 400–418.

[5] D. S. Chaplot, D. P. Gandhi, A. Gupta, and R. R. Salakhutdinov, “Object goal navigation using goal-oriented semantic exploration,” Advances in Neural Information Processing Systems, vol. 33, pp. 4247–4258, 2020.

[6] H.-T. L. Chiang, A. Faust, M. Fiser, and A. Francis, “Learning navigation behaviors end-to-end with autors,” IEEE Robotics and Automation Letters, vol. 4, no. 2, pp. 2007–2014, 2019.

[7] A. Faust, K. Oslund, O. Ramirez, A. Francis, L. Tapia, M. Fiser, and J. Davidson, “Prm-rl: Long-range robotic navigation tasks by combining reinforcement learning and sampling-based planning,” in 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018, pp. 5113–5120.

[8] A. Francis, A. Faust, H.-T. Chiang, J. Hsu, J. C. Kew, M. Fiser, and T.-W. E. Lee, “Long-range indoor navigation with prm-rl,” IEEE Transactions on Robotics, 2020.

[9] P. Long, W. Liu, and J. Pan, “Deep-learned collision avoidance policy for distributed multiagent navigation,” IEEE Robotics and Automation Letters, vol. 2, no. 2, pp. 656–663, 2017.

[10] M. Everett, J. Miller, and J. P. How, “Planning beyond the sensing horizon using a learned context,” in 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2019, pp. 1064–1071.

[11] M. Narasimhan, E. Wijmans, X. Chen, T. Darrell, D. Batra, D. Parikh, and A. Singh, “Seeing the un-scene: Learning amodal semantic maps for room navigation,” in European Conference on Computer Vision. Springer, 2020, pp. 513–529.

[12] C. Chen, U. Jain, C. Schissler, S. V. A. Gari, Z. Al-Halah, V. K. Ithapu, P. Robinson, and K. Grauman, “Soundspaces: Audio-visual navigation in 3D environments,” in European Conference on Computer Vision. Springer, 2020, pp. 17–36.

[13] C. Chen, Z. Al-Halah, and K. Grauman, “Semantic audio-visual navigation,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 15 516–15 525.

[14] C. Chen, S. Majumdar, Z. Al-Halah, R. Gao, S. K. Ramakrishnan, and K. Grauman, “Learning to set waypoints for audio-visual navigation,” arXiv preprint arXiv:2008.09622, 2020.

[15] Y. Yu, W. Huang, F. Sun, C. Chen, Y. Wang, and X. Liu, “Sound adversarial audio-visual navigation,” arXiv preprint arXiv:2202.10910, 2022.

[16] J. Krantz, E. Wijmans, A. Majumdar, D. Batra, and S. Lee, “Beyond the nav-graph: Vision-and-language navigation in continuous environments,” in European Conference on Computer Vision. Springer, 2020, pp. 104–120.

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

[18] S. Song, A. Zeng, A. X. Chang, M. Savva, S. Savarese, and T. Funkhouser, “Im2pano3d: Extrapolating 360 structure and semantics beyond the field of view,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3847–3856.

[19] Y. Liang, B. Chen, and S. Song, “Sscnav: Confidence-aware semantic scene completion for visual semantic navigation,” in 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2021, pp. 12004–12004.

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

[21] L. Küstner, T. Buiyan, L. Jiao, T. A. Le, X. Zhao, Z. Shen, and J. Lambrecht, “Arena-rosnav: Towards deployment of deep-reinforcement-learning-based obstacle avoidance into conventional autonomous navigation systems,” in 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021, pp. 6456–6463.