Robust Detection for Autonomous Elevator Boarding Using a Mobile Manipulator

Seungyoun Shin[1], Joon Hyung Lee[1], Junhyug Noh[2], and Sungjoon Choi[1(✉)]

[1] Korea University, 13 Jongam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea
{2022021568,dlwnsgud8823,sungjoon-choi}@korea.ac.kr
[2] Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Republic of Korea
junhyug@ewha.ac.kr

Abstract. Indoor robots are becoming increasingly prevalent across a range of sectors, but the challenge of navigating multi-level structures through elevators remains largely uncharted. For a robot to operate successfully, it's pivotal to have an accurate perception of elevator states. This paper presents a robust robotic system, tailored to interact adeptly with elevators by discerning their status, actuating buttons, and boarding seamlessly. Given the inherent issues of class imbalance and limited data, we utilize the YOLOv7 model and adopt specific strategies to counteract the potential decline in object detection performance. Our method effectively confronts the class imbalance and label dependency observed in real-world datasets, Our method effectively confronts the class imbalance and label dependency observed in real-world datasets, offering a promising approach to improve indoor robotic navigation systems.

Keywords: Object detection · Mobile manipulator · Class imbalance in detection methods

1 Introduction

Indoor robots have become a ubiquitous presence in diverse fields, ranging from hospitality and delivery services to cleaning and security. The development of localization techniques and the study of legged robot locomotion on stairs has been the focus of significant research efforts. Nevertheless, navigating multi-level buildings using elevators remains an underexplored topic in the field. A crucial skill that robots require to leverage elevators effectively is perception. This involves accurately determining parameters, such as the current floor and the elevator's location, by processing and interpreting sensory information. Advanced perception capabilities are thus essential for robots to navigate multi-level buildings with accuracy and efficiency. However, equipping robots with additional sensors can be prohibitively expensive and may not be a scalable solution. To address this challenge, we propose a novel method to recognize the state of an elevator using only an image sensor, thereby eliminating the need for additional equipment.

H. Lu et al. (Eds.): ACPR 2023, LNCS 14406, pp. 15–28, 2023.
https://doi.org/10.1007/978-3-031-47634-1_2

Fig. 1. GAEMI is a sophisticated mobile manipulator equipped with a 5DoF robotic arm and a ZED camera. Its non-holonomic base features a 2D LiDAR sensor for obstacle detection and localization within a mapped environment. Additionally, GAEMI has a forward-facing RGB camera.

When attempting to perform object detection under elevator conditions, it became evident that real-world datasets present significant challenges that deviate from pre-existing benchmark datasets, such as COCO [10]. Two prominent issues that arose were label dependency and small object detection.

Addressing the challenges of issues requires careful consideration and specialized techniques to mitigate their impact on object detection performance. To this end, we have developed a comprehensive system for indoor robots that focuses on these challenges and enables intelligent interaction with elevators.

In summary, we introduce a **robotic system adept at indoor navigation and intelligent elevator interaction**. It effectively addresses challenges like small object detection and label dependency to ensure accurate elevator state recognition and precise interaction. The primary contributions are:

1. Development of an autonomous robotic system that interacts with elevators using advanced SLAM, kinematics, and perception, ensuring dependable real-world navigation and interaction.
2. Tackling small object detection and label dependency in our dataset, enhancing the robot's perception and performance in real-world tasks.

In Fig. 1, we present GAEMI, the robot employed throughout our experiments, which demonstrates the capabilities of the autonomous system developed in this work.

2 Related Work

Our study tackles the challenge of autonomous elevator operation by emphasizing perception and its translation into robotic actions. Utilizing advances in object detection and elevator button recognition, we aim to establish a dependable system for indoor robots' inter-floor movement. This section provides a brief overview of pertinent literature on object detection and prior research on elevator-interacting robots.

2.1 Object Detection

Object detection has been a crucial area of research within the field of computer vision, with numerous techniques and models developed to advance its capabilities. In this section, we briefly discuss key approaches, including real-time object detection and anchor-based and anchor-free methods.

Real-time object detection focuses on achieving high-speed detection while maintaining acceptable levels of accuracy. This aspect is particularly important for applications where real-time response is vital, such as autonomous vehicles and robotic navigation. Several notable models, such as YOLO [14], SSD [11], and MobileNet [3], have been proposed to address the trade-off between speed and accuracy.

Anchor-based object detection methods, such as Faster R-CNN [15] and RetinaNet [9], use predefined anchor boxes to generate region proposals for detecting objects. These methods benefit from improved localization accuracy but may suffer from increased computational complexity due to the need to evaluate multiple anchors per image.

Conversely, anchor-free object detection approaches, such as CornerNet [8] and CenterNet [20], eliminate the need for predefined anchor boxes by directly predicting object bounding boxes and class probabilities. These methods have the potential to simplify the detection pipeline and reduce computational overhead, making them attractive for real-time object detection tasks.

In the following sections, we will delve into the specifics of these object detection techniques and explore their relevance to addressing the challenges associated with real-world vision tasks, such as extreme class imbalance and label dependency.

2.2 Autonomous Elevator Boarding Using Mobile Manipulators

Recent research advancements in indoor robots have contributed significantly to understanding indoor environments and the development of robot navigation and interaction capabilities [1,4,13,16,17]. However, limited attention has been dedicated to the problem of robots autonomously moving between floors using elevators. Most existing studies have primarily focused on elevator button recognition, lacking a comprehensive pipeline for autonomous inter-floor movement [5–7]. This underscores the need for further investigation to address the challenges associated with autonomous elevator operation for indoor robots.

Conventional computer vision algorithms have been employed in some studies for elevator button recognition due to their low data requirements. However, these methods suffer from limited accuracy and necessitate specific postures or environments for the robot's operation. This indicates the need for more advanced techniques for autonomous elevator operation in real-world robotic services. To address the challenges associated with conventional methods in button recognition, Dong *et al.* [2] introduced a deep learning approach to improve elevator button recognition. Nevertheless, button location identification still relied on conventional methods.

Yang *et al.* [19] proposed an end-to-end method for button recognition using the YOLO [14], which enables real-time object detection. Zhu *et al.* [21] introduced a large-scale dataset specifically for button recognition and highlighted the presence of a high-class imbalance in this task.

A related work worth mentioning is an autonomous robotic system that utilizes an eye-in-hand configuration [22]. This system addresses button operation by incorporating a deep neural network for button detection and character recognition along with a button pose estimation algorithm. However, it remains essential to develop a comprehensive pipeline for autonomous inter-floor movement and advanced techniques for autonomous elevator operation in real-world scenarios.

3 Proposed Method

3.1 Perception System

Label Superset and Elevator Status Perception. The primary objective of robotic perception for elevator usage is to ascertain the elevator's status, such as whether the door is open or closed, the current floor of the robot, and the location of the elevator floor. To achieve this level of perception, essential for seamless navigation and interaction with the elevator system, we first design a label superset that clearly defines the problem and covers all possible scenarios to accommodate diverse sites, as shown in Table 1.

With our designed label superset, we can address various elevator statuses vital for robot planning. For a task like "Go to room 406," the robot needs to recognize its floor and the elevator's status. This helps decide whether to press up or down and prepare for boarding based on the elevator's position. The robot must also ascertain the elevator door's movement. Our label superset equips the robot to make decisions and navigate intricate settings effectively.

Object Detection Model. We utilized the YOLOv7 [18] for both object detection and instance segmentation tasks. YOLOv7 is a cutting-edge object detection model that outperforms other detectors in terms of speed and accuracy. We deemed YOLOv7 an appropriate choice for detecting elevator status. To optimize the YOLOv7 object detection and instance segmentation models for our hardware (NVIDIA Orin), we implemented float16 quantization. This optimization led to a single forward path inference time with an FPS ranging between 50 and 80, which is well within the acceptable range for real-time processing.

Table 1. Label superset. White rows represent labels processed by the indicator detection module, while the gray row indicates labels handled by the button detection module. We employ an instance segmentation model for the gray row.

Category	Parameters
Elevator Door	Opened, Moving, Closed
Current Robot Floor	B6, B5, . . . , B1, 1, . . . , 63
Current Elevator Floor (Outside/Inside)	B6, B5, . . . , B1, 1, . . . , 63
Current Elevator Direction (Outside/Inside)	Up, Down, None
Elevator Button (Outside/Inside)	Up, Down, B6, B5, . . . , B1, 1, . . . , 63

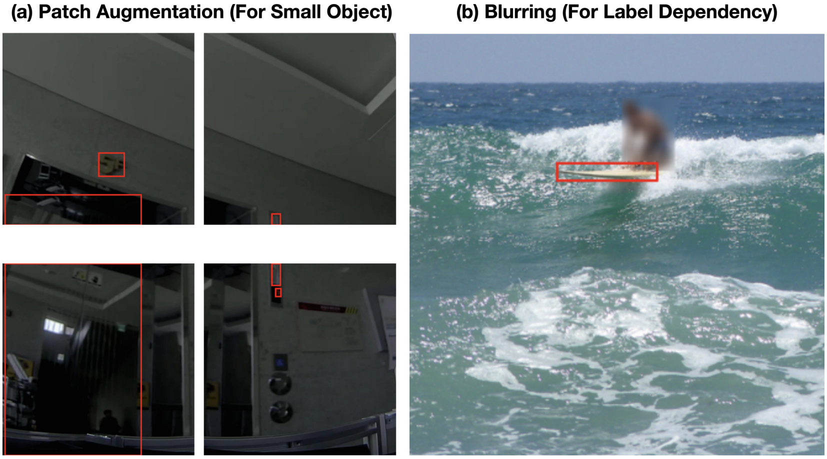

Fig. 2. Our two-fold strategy for addressing **label dependency** and **small object detection**. (a) The **patch-augmentation** technique involves cropping high-resolution images, which helps to maintain the resolution of smaller objects when they are resized for input to the model, ultimately enhancing recognition of fine details. (b) The **Gaussian blur** applied to the bounding box region effectively mitigates label dependency and class imbalance by altering the visual features and removing certain labels from the image.

3.2 Addressing Label Dependency and Small Object Detection

In our data collection phase for training elevator indicators, we faced two significant challenges: label dependency and small object detection. Label dependency is an issue where some labels, like those for elevator doors, appear regularly in the images, while others are seen less often. This unequal distribution creates a class imbalance, making it hard to compile a balanced dataset with a variety of labels. Our second challenge was related to the size of the objects in our

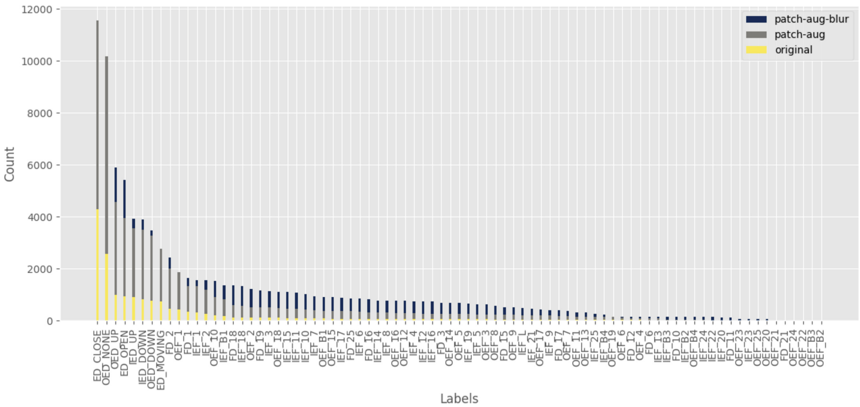

Fig. 3. Augmentation strategy. To tackle challenges like small indicator detection, label dependency, and class imbalance, we employ a two-step augmentation process. **(1) Original dataset**. **(2) Patch-aug**: We crop the original dataset for higher-resolution training images. **(3) Patch-aug-blur**: We blur frequent objects such as elevator doors in high-resolution images and remove their labels. This strategy increases the dataset size and effectively mitigates class imbalance and label dependency issues.

images. Most indicators, except for the elevator doors, are quite small, which can make accurate detection by object detection models more difficult. This size discrepancy posed an additional hurdle in our pursuit of precise detection and identification.

Patch Augmentation. In addressing the challenge of detecting small objects, we employed a technique called patch augmentation. We divided high-resolution images into cropped sections, thereby increasing the resolution of smaller objects and enriching the visual features in the dataset. The patch-augmentation process is illustrated in Fig. 2(a).

Label Blurring. To tackle the label dependency issue, we adopted a method that involves duplicating portions of the dataset and applying a Gaussian blur to the bounding box region, effectively eliminating the visual features from the image. In our dataset, certain labels such as Elevator Door Closed (`ED_CLOSE`) and Current Elevator Direction Outside with None (`OED_NONE`) appeared more frequently. This was because elevator doors are present in every image, and most sites have a direction indicator. As a consequence, sampling scarce labels would result in an increase in `ED_CLOSE` and `OED_NONE` occurrences, leading to an imbalanced dataset. To address this issue, we selectively blurred these two classes. For instance, if an image displayed an elevator door in the closed state and the current floor was 4, we blurred the elevator door to remove the label from the image. This strategy efficiently mitigates the label dependency issue and generates a more compact, balanced dataset. Figure 3 illustrates the resolution

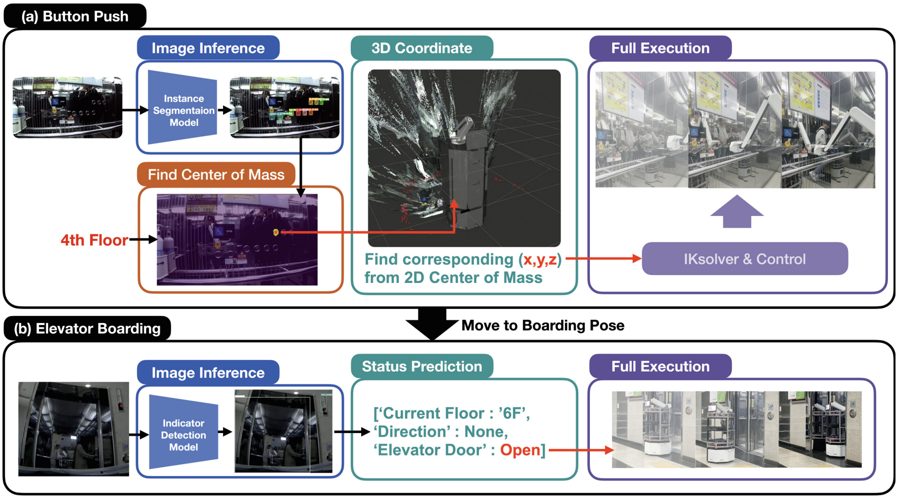

Fig. 4. An overview of the autonomous elevator boarding process. The procedure is divided into two main categories: (a) button-pushing operations and (b) elevator boarding, which encompass tasks such as path planning, object detection, and interaction.

of label dependency and class imbalance, as demonstrated by the dark blue bar. The implementation of blurring is shown in Fig. 2(b).

3.3 GAEMI: Robotic System

Robot Configuration. In this study, we employed a GAEMI robot, produced by ROBOTIS, as our indoor mobile manipulator for self-driving service applications. The GAEMI robot measures $50 \times 50 \times 117$ cm^3 and features a non-holonomic base along with a 5-degree-of-freedom (DoF) manipulator. The base of the robot is equipped with a 2D LiDAR sensor, and a ZED camera is mounted on the manipulator arm's end effector. A comprehensive illustration of the entire robot is available in Fig. 1.

Robot Operation. Our primary objective for the robot operation is to achieve autonomous elevator boarding without relying on additional equipment, such as network connections with the elevator or specialized sensors. We divided the mobile manipulator tasks into two main categories: (1) navigation and (2) interaction. Our complete procedure is depicted in Fig. 4.

Navigation. We employed Cartographer to generate a 2D map. Subsequently, we used the ROS2 Navigation2 package [12] to direct the robot toward the target pose. We utilized the Adaptive Monte Carlo Localization (AMCL) method to determine the robot's current position within the map and the Dynamic Window Approach (DWA) for path following.

Fig. 5. Dataset overview. (a) Indicator dataset: object detection dataset is tailored to capture the basic status of an elevator. (b) Button dataset: instance segmentation dataset designed to identify points of interaction between the robot and the elevator, facilitating precise and successful task execution.

Interaction. In the elevator boarding task, the primary physical interaction with the environment involves 'button pushing.' To achieve this, we first perform instance segmentation on RGB images to identify relevant objects within the scene. For example, if we intend to push the 'down' button, we locate the corresponding mask that represents the 'down' button in the image.

Next, we calculate the 2D center of mass from the obtained mask image, resulting in a 2D (x, y) point. To determine the corresponding 3D (x, y, z) coordinate, we use the camera's intrinsic parameters to transform the 2D point. This transformation allows us to obtain the relative pose of the button with respect to the camera's position.

Once the relative pose is established, we can solve the inverse kinematics (IK) problem to determine the target joint angles for the robotic manipulator. Finally, we implement a manipulator control system to execute the desired button-pushing operation. This approach enables us to achieve effective and precise interactions with the elevator's button panel.

4 Experiments

4.1 Mitigating Class Imbalance and Label Dependency

We partitioned the perception dataset into two distinct parts to address different aspects of elevator perception. The first part is the Indicator dataset, which focuses on identifying the overall status of the elevator without requiring manipulator interaction. This dataset encompasses categories such as elevator door status, current robot floor, and elevator direction. The second part of the dataset addresses perception skills that necessitate precise interaction points, such as the elevator button. For these cases, instance segmentation is employed to detect flexible and reliable actions once the button is perceived by the robot. This division enables a comprehensive understanding of the elevator environment, ultimately facilitating seamless robot operation. Figure 5 shows some examples of the two datasets, and further details are provided below:

- **Indicator dataset:** In this study, we compiled a dataset consisting of 5,000 images sourced from seven distinct origins. The images were captured using two types of devices: a robot-mounted camera, providing varied viewpoints, and a smartphone camera. The inclusion of images from multiple perspectives aimed to enhance the dataset's robustness. The dataset consists of pairs of (image, bounding boxes).
- **Button dataset:** Additionally, we collected button data to identify interaction points. This dataset differs from the indicator dataset in that it includes pairs of (image, bounding boxes, instance segmentation masks). We hypothesize that incorporating instance segmentation masks can improve interaction capabilities, resulting in more successful robot operations.

Before training the robot perception datasets (Indicator and Button datasets), we first evaluated the effectiveness of our proposed method (Blur elimination of class) on a general dataset. We aimed to assess how different variations in the dataset influenced the performance of our model. The evaluation results are summarized in Table 2, which presents the mean Average Precision (mAP) scores at different Intersection-over-Union (IoU) thresholds (0.5 and 0.95) for various dataset variations. The evaluation was conducted using the COCO-test dataset. We established a baseline using the COCO-mini dataset, which consists of 1,000 images randomly sampled from the COCO-train dataset. The baseline model achieved a mAP@0.5 score of 0.014 and a mAP@0.95 score of 0.007. To create the COCO-blur dataset, as described in Sect. 3, we applied our proposed Gaussian blur elimination technique to the COCO-mini images and replicated them ten times, resulting in a dataset of 10,000 images. The COCO-blur dataset exhibited improved performance, with a mAP@0.5 score of 0.018 and a mAP@0.95 score of 0.009. In contrast, the COCO-cutout dataset, which replaces the blurring process with zero-value regions, demonstrated lower mAP scores compared to both the baseline and COCO-blur datasets. These results highlight the effectiveness of our proposed Blur Elimination Technique in addressing two common challenges in machine learning: class imbalance and label dependency. By employing this technique, we successfully enhanced the performance of our model on more generalized datasets, specifically the COCO-mini dataset.

Our analysis provides strong evidence supporting the effectiveness of the Blur Elimination Technique in improving model performance on a general dataset. Based on these encouraging results, we extended the application of this method to our Indicator dataset.

Table 2. Comparison of mAP scores on COCO-mini variations.

Dataset	mAP@0.5	mAP@0.95
COCO-mini (base)	0.014	0.007
COCO-blur	0.018	0.009
COCO-cutout	0.012	0.006

Our experiments demonstrate that diverse augmentations boost YOLOv7's performance. Without patch and blur augmentations, YOLOv7's accuracy drops. Using patch augmentation alone increased mAP@0.5 by +0.054, due to improved visual features aiding model accuracy.

However, adding blur after patch augmentation showed a trade-off between localization and status accuracy. Even though exact object localization slightly suffered, status accuracy improved. This is expected, as blurring affects localization but adds noise beneficial against class imbalance.

For our mobile robot's needs, status accuracy was prioritized over exact localization since it directly affects the robot's actions. Thus, even with slight localization losses, the model's effectiveness hinges on improved status accuracy, shown in Table 3.

Table 3. Experimental results on Indicator dataset. Different variations of the YOLOv7 model are evaluated based on mAP@0.5 and Status Accuracy. Rows in white represent standard and patched YOLOv7 models, while the gray row denotes the performance of the YOLOv7 model augmented with both a patch and the proposed blur elimination technique.

Method	mAP@0.5	Status Accuracy
YOLOv7	0.730	0.813
YOLOv7 + patch	0.784	0.878
YOLOv7 + patch + blur	0.779	0.879

4.2 Real-World Robot Operation

To evaluate the performance of our proposed method in real-world scenarios, we conducted various tasks in the Woojung Hall of Informatics at Korea University. For this purpose, we constructed an occupancy map of the 6th floor of the building, as shown in Fig. 6. We focused on three essential tasks to assess the effectiveness of our method. The tasks were as follows: (1) navigating to the button position (`GOTO_BUTTON_POSE`), (2) pressing elevator buttons (`BUTTON_PUSHING`), and (3) boarding the elevator (`ELEVATOR_BOARDING`). The success rates for each

Fig. 6. Occupancy map of Woojung Hall of Informatics. This figure illustrates the constructed occupancy map of the 6th floor of Woojung Hall of Informatics at Korea University, which serves as the operational landscape for all our real-world robot experiments.

Table 4. Real-world experiment results. This table reports the success rates of three distinct tasks executed by the robot. Each success rate corresponds to the proportion of successful trials out of a total of ten attempts.

Task	Success Rate
`GOTO_BUTTON_POSE`	10/10
`BUTTON_PUSHING`	9/10
`ELEVATOR_BOARDING`	3/10

task are presented in Table 4, demonstrating the performance of our method in real-world applications.

In the `GOTO_BUTTON_POSE` task, the robot successfully achieved optimal positioning for button actuation (the button click pose, as shown in Fig. 6) in all ten trials, resulting in a success rate of 100%. The task was considered successful if the positioning error was within 15cm of the target button click pose. In the `BUTTON_PUSHING` task, the robot effectively pressed the correct elevator buttons in nine out of ten trials, leading to a success rate of 90%. The evaluation cri-

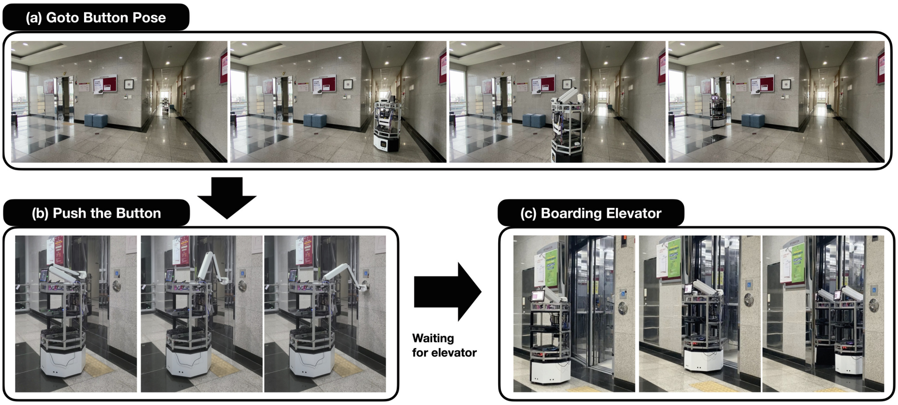

Fig. 7. Demonstration of our integrated robotic system. A comprehensive illustration of the robot successfully performing tasks within a real-world indoor environment.

terion for this task was the successful actuation of the targeted button by the robot. However, in the `ELEVATOR_BOARDING` task, the success rate was only 30% over ten trials. The suboptimal performance in this task can be attributed to environmental limitations, such as potential obstructions at the elevator door.

The experimental results presented in Table 4 validate the efficacy of our proposed method and the robustness of our integrated robotic system in real-world indoor environments. The high success rates in the `GOTO_BUTTON_POSE` and `BUTTON_PUSHING` tasks demonstrate the ability of our method to overcome label dependency and detect small objects effectively. These achievements significantly contribute to enhancing robotic navigation and interaction capabilities in multi-floor buildings.

Figure 7 provides a comprehensive demonstration of the autonomous elevator boarding process.

5 Conclusion

In this work, we introduced a comprehensive robotic system capable of intelligent interaction with elevators in multi-floor environments. We developed a unique solution that successfully addresses the challenges of class imbalance and label dependency in object detection, leading to an enhanced perception system. Our system integrates cutting-edge SLAM, kinematics, and perception technologies, enabling the robot to reliably navigate within its environment and interact effectively with elevator buttons and doors. In real-world scenarios, our system demonstrated high accuracy and reliability, achieving commendable success rates in targeted tasks. The developed approach significantly improves the robot's functionality and effectiveness, paving the way for broader applications of robotics in complex indoor environments.

Acknowledgement. This work was supported by "Research of Elevator Indicator Recognition Technology for Indoor Autonomous Navigation" project funded by ROBOTIS Co. Ltd. and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2019-0-00079, No. 2022-0-00871, No. 2022-0-00612, No. 2022-0-00480).

References

1. Chaplot, D.S., Gandhi, D.P., Gupta, A., Salakhutdinov, R.R.: Object goal navigation using goal-oriented semantic exploration. Adv. Neural. Inf. Process. Syst. **33**, 4247–4258 (2020)
2. Dong, Z., Zhu, D., Meng, M.Q.H.: An autonomous elevator button recognition system based on convolutional neural networks. In: 2017 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 2533–2539. IEEE (2017)
3. Howard, A.G., et al.: MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
4. Huang, C., Mees, O., Zeng, A., Burgard, W.: Visual language maps for robot navigation. arXiv preprint arXiv:2210.05714 (2022)
5. Kang, J.G., An, S.Y., Choi, W.S., Oh, S.Y.: Recognition and path planning strategy for autonomous navigation in the elevator environment. Int. J. Control Autom. Syst. **8**, 808–821 (2010)
6. Kim, H.H., Kim, D.J., Park, K.H.: Robust elevator button recognition in the presence of partial occlusion and clutter by specular reflections. IEEE Trans. Industr. Electron. **59**(3), 1597–1611 (2011)
7. Klingbeil, E., Carpenter, B., Russakovsky, O., Ng, A.Y.: Autonomous operation of novel elevators for robot navigation. In: 2010 IEEE International Conference on Robotics and Automation, pp. 751–758. IEEE (2010)
8. Law, H., Deng, J.: CornerNet: detecting objects as paired keypoints. Int. J. Comput. Vision **128**(3), 642–656 (2019). https://doi.org/10.1007/s11263-019-01204-1
9. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
10. Lin, T.-Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
11. Liu, W., et al.: SSD: single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
12. Macenski, S., Martín, F., White, R., Clavero, J.G.: The marathon 2: a navigation system. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2718–2725. IEEE (2020)
13. Menini, D., Kumar, S., Oswald, M.R., Sandström, E., Sminchisescu, C., Van Gool, L.: A real-time online learning framework for joint 3D reconstruction and semantic segmentation of indoor scenes. IEEE Robot. Autom. Lett. **7**(2), 1332–1339 (2021)
14. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
15. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems 28 (2015)

16. Song, J., Patel, M., Ghaffari, M.: Fusing convolutional neural network and geometric constraint for image-based indoor localization. IEEE Robot. Autom. Lett. **7**(2), 1674–1681 (2022)
17. Vidanapathirana, K., Ramezani, M., Moghadam, P., Sridharan, S., Fookes, C.: LoGG3D-Net: locally guided global descriptor learning for 3d place recognition. In: 2022 International Conference on Robotics and Automation (ICRA), pp. 2215–2221. IEEE (2022)
18. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696 (2022)
19. Yang, P.Y., Chang, T.H., Chang, Y.H., Wu, B.F.: Intelligent mobile robot controller design for hotel room service with deep learning arm-based elevator manipulator. In: 2018 International Conference on System Science and Engineering (ICSSE), pp. 1–6. IEEE (2018)
20. Zhou, X., Wang, D., Krähenbühl, P.: Objects as points. arXiv preprint arXiv:1904.07850 (2019)
21. Zhu, D., Li, T., Ho, D., Zhou, T., Meng, M.Q.: A novel OCR-RCNN for elevator button recognition. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 3626–3631. IEEE (2018)
22. Zhu, D., Min, Z., Zhou, T., Li, T., Meng, M.Q.H.: An autonomous eye-in-hand robotic system for elevator button operation based on deep recognition network. IEEE Trans. Instrum. Meas. **70**, 1–13 (2020)