2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

A Novel Framework for Structure Descriptors-Guided Hand-drawn Floor Plan Reconstruction

Zhentong Zhang[1,3], Juan Liu[2], Xinde Li[1,3,4,5], ChuanFei Hu[3,4], Fir Dunkin[3,4] and Shaokun Zhang[4]

***Abstract*— In the absence of a pre-built indoor map, robot navigation suffers from the limitations of sensors and environments, resulting in decreased efficiency in performing ad-hoc tasks. Given that blueprints are difficult to obtain, an intuitive method is to provide robots with prior knowledge via hand-drawn floor plans. However, due to the inability of robots to directly comprehend hand-drawn styles, the applicability of this method is limited. In this paper, we present a novel framework for hand-drawn floor plan reconstruction that can recognize abstract hand-drawn elements and standardize the reconstruction of hand-drawn floor plans, thereby providing robots with valuable global map information. Specifically, we design a new series of structure descriptors as reconstruction components and employ a deep learning-based model for recognition. Then the standardized results are obtained through the proposed floor plan reconstruction algorithm. To verify the effectiveness of the framework, we conduct experiments on electronic and paper hand-drawn floor plans. Compared with other state-of-the-art methods, our proposed method achieves superior reconstruction results. This work expands the application scenarios for indoor robots, enabling them to quickly comprehend the semantics of complex scenes, thereby enhancing the competitiveness in downstream tasks.**

I. Introduction

In recent years, robot scene understanding has attracted significant attention from both the industrial and academic communities [1]–[5]. In exploring the unknown environment, indoor maps can help robots improve the efficiency of mapping and navigation [6]–[9]. As shown in Fig. 1, every second is crucial in emergency firefighting rescue missions, and the inefficiency of robots that rely only on random searches without pre-built maps has become increasingly prominent. Although blueprints are difficult to obtain in reality, the trapped individuals can employ available tools to provide hand-drawn floor plans. While humans can easily understand the information in hand-drawn drawings and determine the optimal rescue path, the uncertainties and non-standard structures in such drawings pose challenges for robotic comprehension. In this paper, we focus attention on the reconstruction of hand-drawn floor plans to facilitate

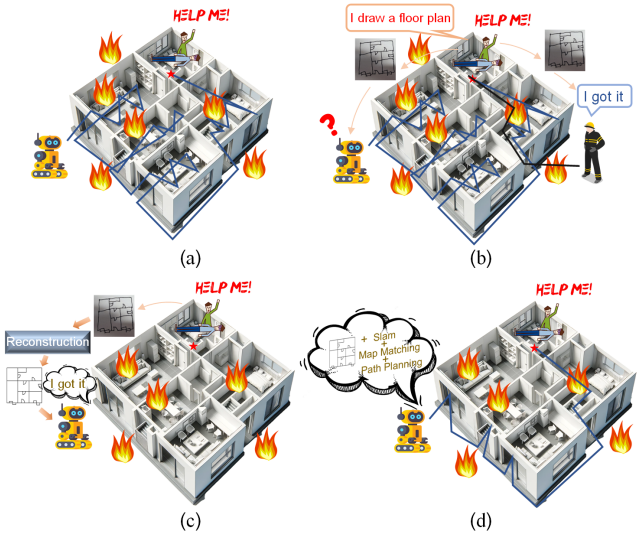

Fig. 1. (a) Robots rescue without maps. (b) Differences between robots and humans with hand drawings. (c) Reconstruction algorithm helps robots understand. (d) Combine downstream algorithms to perform rescue tasks.

robots to establish a basic understanding of uncertain indoor environments. By integrating other downstream navigation algorithms, robots can achieve higher efficiency in completing complex tasks in unknown environments. As far as we know, there is currently no method for reconstructing hand-drawn floor plans. To address this challenge, we propose a novel framework for structure descriptors-guided hand-drawn floor plan reconstruction.

Hand-drawn styles exhibit variations among individuals and suffer from distortions, inconsistent alignment, non-standard structures, complex backgrounds, and shadows. The primary challenges in reconstructing floor plans mainly focus on the recognition of noise elements and the understanding of the correct relationship between structures. These challenges hinder the application of traditional blueprint reconstruction algorithms on hand-drawn floor plans [10]–[12]. In the domain of sketch synthesis [13], [14], such algorithms typically prioritize overall stylistic similarity and cannot have precise control over lines and angles. Moreover, from a sampling perspective, the output of sketch synthesis is unstable.

In this paper, we propose a framework for reconstructing floor plans to help robots understand hand-drawn floor plans in uncertain indoor environments. The framework utilizes a deep learning algorithm to recognize the information of structure descriptors required for reconstruction. Then, the standardized reconstruction algorithm is employed, which

*This work was supported by the National Natural Science Foundation of China under Grant 62233003 and 62073072, the Key Projects of Key R&D Program of Jiangsu Province under Grant BE2020006 and Grant BE2020006-1, and the Shenzhen Science and Technology Program under Grant JCYJ20210324132202005 and JCYJ20220818101206014.

[1]School of Cyber Science and Engineering, Southeast University, Nanjing 211102, China
[2]Samsung Electronics(China)R&D, Nanjing 210018, China
[3]Nanjing Center for Applied Mathematics, Nanjing 211135, China
[4]School of Automation, Southeast University, Nanjing, 210012, China
[5]Southeast University Shenzhen Research Institute, Shenzhen, 518063, China
*Xinde Li is corresponding author: xindeli@seu.edu.cn

consists of three sub-algorithms: basic structural descriptor connection, coordinate adjustment, and door/window reconstruction. In consideration of practical usage scenarios, the reconstruction framework unifies the handling of both electronic and paper hand-drawn floor plans, directly obtaining the final reconstruction results through an end-to-end processing pipeline, without the necessity for explicitly distinguishing between them during the computation process. The main contributions of this paper can be summarized as follows.

(1) To address the limitations of traditional mapping and navigation methods in enhancing robots' comprehension of scene semantics, we propose a potential technology approach to augment scene semantics using hand-drawn floor plans.

(2) With the aim of addressing the significant variations in hand-drawn drawing styles among different individuals, which can hinder scene understanding, we propose a novel framework for structure descriptors-guided hand-drawn floor plan reconstruction.

(3) Compared to other state-of-the-art methods, the framework achieves superior reconstruction results, an enhancement that greatly boosts the competitiveness of indoor robots in downstream tasks.

II. Related works

A. Floor plan blueprint reconstruction

The technique of floor plan blueprint reconstruction is a relatively mature field and can be applied to subsequent robot tasks. Previous works mainly focused on designing rule-based algorithms according to the floor plan blueprint characteristics [15]–[18]. Since the floor plan blueprint is generally drawn according to rules, specific algorithms can be used to recognize and correct walls [19]. Ahmed et al. used expert algorithms to segment the blueprint into different parts and obtain standardized floor plans by replacing them with the original ones [20], [21]. To address the problem of uneven wall thickness in the floor plan blueprint, a deep learning model was designed to obtain a segmented image of the walls [22], [23]. However, blueprints are drawn according to specific standards, which simplifies the challenges of reconstruction. As shown in Fig. 2, in the hand-drawn drawing with background and shadows, the Hough transform, contour methods, and filtering methods commonly used in traditional blueprint reconstruction completely fail.

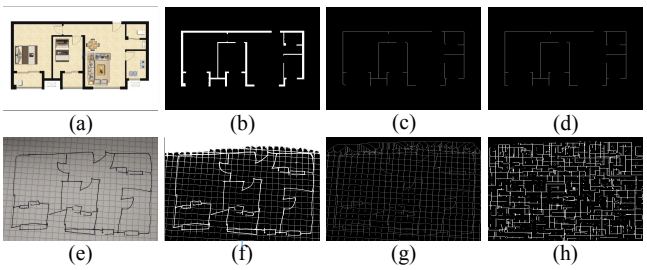

Fig. 2. Blueprint reconstruction algorithm applies to blueprint and hand-drawn. (a), (e) Blueprint and hand-drawn. (b), (f) Foreground extraction. (c), (g) Contour extraction. (d), (h) Standardized reconstruction.

B. Hand-drawn object recognition

Hand-drawn object recognition is a technique used to identify specific objects in hand-drawn sketches. Previous research on hand-drawn sketch recognition can be divided into online and offline methods [24]–[26]. Online methods rely on user input information, such as pen speed and curvature. Gennari et al. proposed a two-dimensional dynamic programming algorithm, where the object category is determined by evaluating strategies and connectivity [27]. Offline methods use texture and shape information to recognize hand-drawn components [28], [29]. Dey et al. used deep learning and boundary tracing to recognize hand-drawn circuit components and connections [30]. However, in the case of paper version hand-drawn floor plans, texture and shape information are ambiguous on account of lighting and photograph. Furthermore, these methods focus on improving component recognition accuracy without considering the relationship between recognition and reconstruction.

C. Hand-drawn sketch synthesis

Hand-drawn sketch synthesis translates hand-drawn style images into another style image while maintaining the original image content [31]–[33]. Our target could be considered as translating hand-drawn floor plan styles into standardized ones. The methods can be classified into supervised and unsupervised learning. Pix2pix [34] established a correspondence between hand-drawn sketches and translated images by adversarial networks. CycleGAN [35] divided images into content and style parts to achieve multi-to-multi mapping between image domains. Diffusion model [36] enabled noisy images to approach the forward-sampled domain during training, and the translated image was obtained through reverse diffusion processes. However, the methods based on hand-drawn image translation cannot process each line and corner accurately of hand-drawn floor plan standardization. The output results are limited in terms of controllability, editing ability, and interpretability.

III. Method

In this section, we will introduce the overall workflow of the reconstruction framework as shown in Fig. 3. First, define the structure descriptors and design a deep learning network to recognize them. Then, employ the descriptor recognition information to perform the standardized reconstruction algorithm. In the Fig. 3, potential applications of the reconstruction framework in downstream robot tasks are also displayed.

Our definition of reconstruction is the transformation of hand-drawn floor plans into standardized floor plan structures. This involves removing the background and shadows, correcting misaligned, distorted lines and angles, recognizing doors&windows and restoring their position. The standard for reconstruction is the vector of the original floor plan. This paper focuses on the restoration of straight lines, angles, and doors&windows. Although this paper does not discuss scaling due to the unknown original proportions in hand-drawn floor plans, it is noted that these proportions can

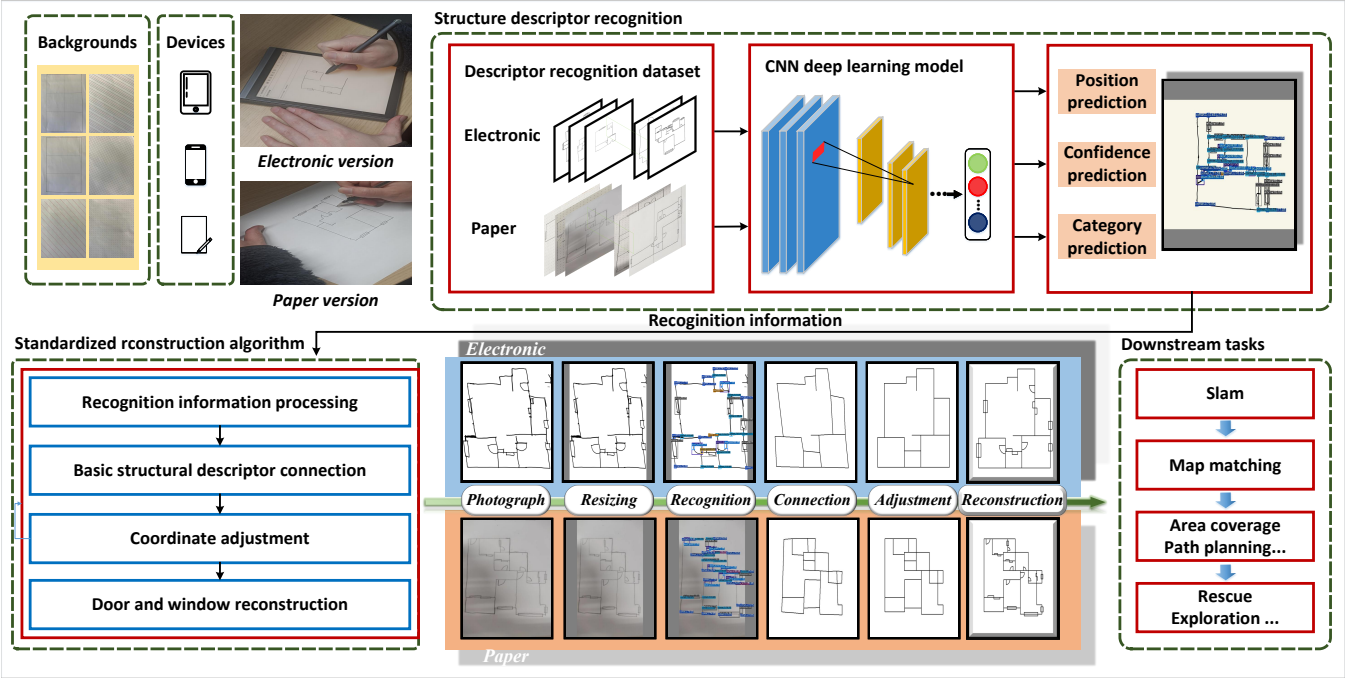

Fig. 3. Our framework for hand-drawn floor plan reconstruction, including structure descriptor recognition and standardized reconstruction algorithms, and potential applications in the field of robotics.

be corrected through downstream matching algorithms. A possible scenario involves using SLAM to obtain incomplete mappings, which are then matched with floor plans to achieve accurate proportion correction.

A. Structure Descriptor recognition algorithm

1) Structure Descriptor definition: We define a new series of structure descriptors for the reconstruction framework. As shown in Fig. 4, we define twelve sub-categories of descriptors within three major structures. The first major structure is the door and window descriptors used for reconstructing the door and window structure, including door, window, sliding door, and bay window. The second major structure is the basic structural point descriptors used for reconstructing the lines, including normal point, L point, T point, cross point, window-wall composite point, and sliding door-wall composite point. The third major structure is the direction descriptors used to determine the direction of the door and window, including the door direction point and bay window direction point.

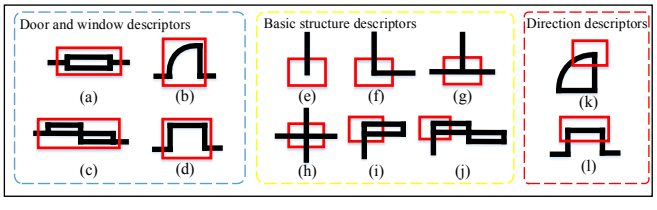

Fig. 4. Examples of descriptor categories. (a) window. (b) door. (c) sliding door. (d) bay window. (e) normal point. (f) L point. (g) T point. (h) cross point. (i) window-wall composite point. (j) sliding door-wall composite point. (k) door direction point. (l) bay window direction point.

2) Detector Design: The detector in this paper is based on YOLOX [37], as shown in Fig. 5. Our annotation definition is shown in Fig. 4. In the standardized reconstruction algorithm, we employ the center point of the bounding box as the location of the basic structure descriptions, and the aspect ratio of the bounding box as the scale for reconstructing doors and windows. We employ the overlap area to supervise the offset of bounding box coordinates $\{x_{\text{center}}, y_{\text{center}}, w, h\}$. Considering the above three geometric parameters, we employ the $CIoU$ loss function to be represented as follows:

$$\mathcal{L}_{CIoU} = 1 - IoU + \left(\frac{\rho^2\left(b, b^{gt}\right)}{c^2}\right) + \alpha v \qquad (1)$$

where IoU is the intersection over union between the predicted bounding box b and the ground truth bounding box b^{gt}, ρ is Euclidean distance, c is the smallest enclosing box covering both b and b^{gt}, α is balance parameter, $\boldsymbol{v} = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2$ calculates width and height consistency.

We utilize a confidence $conf > 0.25$ to select the initial information for the standardized reconstruction from the detection results of structure descriptors. The confidence loss can be represented as follows:

$$\mathcal{L}_{conf} = -\sum_{i=1}^{N}\left[y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)\right] \qquad (2)$$

where N is the total number of b, y_i denotes the ground truth label of b_i, $\hat{y}_i$ denotes the predicted confidence of b_i.

In our standardized reconstruction algorithm, structure descriptors are processed separately according to their de-

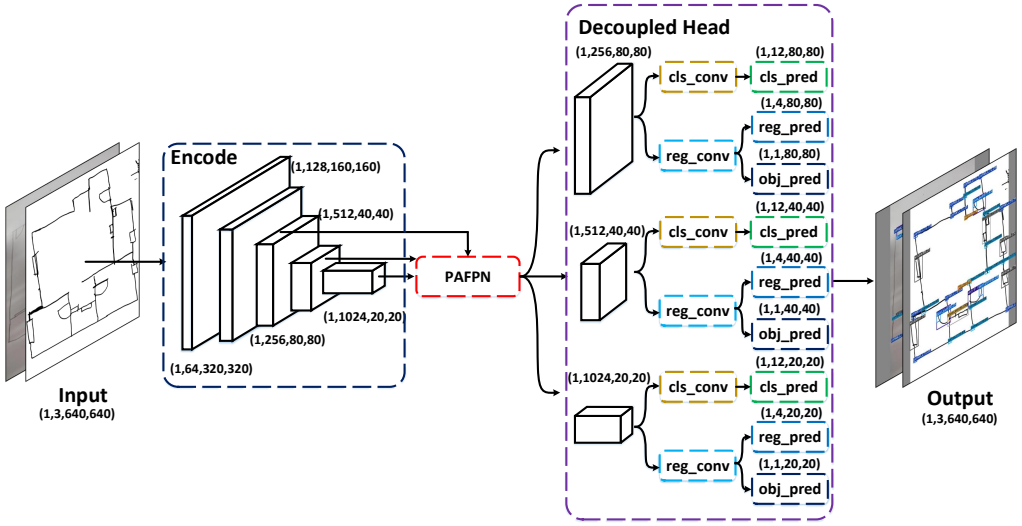

Fig. 5. Deep learning descriptor recognition model.

tection categories. The classification loss can be represented as follows:

$$\mathcal{L}_{class} = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_{i,c}) \tag{3}$$

where C is the total number of classes, $y_{i,c}$ and $\hat{y}_{i,c}$ denote the ground truth class and predicted class of b_i, respectively.

The total loss for network optimization is a combination of $CIoU$ loss, confidence loss, and classification loss as follows:

$$\mathcal{L}_{total} = \lambda_{CIoU}\cdot\mathcal{L}_{CIoU} + \lambda_{conf}\cdot\mathcal{L}_{conf} + \lambda_{class}\cdot\mathcal{L}_{class} \tag{4}$$

where λ_{CIoU}, λ_{conf}, and λ_{class} are the weights that balance the above three losses. In this paper, they are all set to 1.

B. Standardized reconstruction algorithm

After obtaining the descriptor information, we employ the standardized reconstruction algorithm, which includes three sub-algorithms as follows:

Basic structural descriptor connection. The bounding box of the descriptor is extracted from the floor plan based on the recognition information. Fig. 3 shows the representation of the basic structural descriptor connection algorithm, which closely corresponds to the transition from the recognition to the connection. We define the basic structural point descriptor as B_i, where i is the ith bounding box. The mth highest pixel density is located at the edges of B_i bounding box as $P_{i,m}$. The direction angle $\theta_{i,m}$ is the angle between $P_{i,m}$ and the center point in the bounding box. Given the influence of background, the number of the $P_{i,m}$ might be inconsistent with the number of the B_i categories. We propose the following connection rules:

- *Connected domain condition*. D is a connected domain function. $P_{i,m}$ and $P_{j,n}$ must have the same connected domain $D(P_{i,m}) = D(P_{j,n})$.
- *Angle condition*. We define to be the horizontal and vertical direction $\theta_{i,m} \in A$, and the other angles as the diagonal direction. The connection rule:

$$\min_{i\neq j}||\theta_{i,m}| - |\theta_{j,n}|| \leq \begin{cases} 15^\circ & \text{if } \theta_{i,m}\in A, \\ 30^\circ & \text{otherwise,} \end{cases}$$
$$\text{where } A = [-15^\circ, 15^\circ]\cup[75^\circ, 105^\circ] \cup[165^\circ, 195^\circ]\cup[255^\circ, 285^\circ] \tag{5}$$

- *Direction condition*. Two connected points have opposite angles:

$$\min_{i\neq j}|\theta_{i,m} - \theta_{j,n}| \geq \begin{cases} 150^\circ & \text{if } \theta_{i,m}\in A, \\ 120^\circ & \text{otherwise,} \end{cases} \tag{6}$$

- *Bidirectional condition*. When multiple points match the above three conditions. Let $\mathcal{P}_{i,m}$ and $\mathcal{P}_{j,n}$ be two match sets of $P_{i,m}$ and $P_{j,n}$. Two connection points are closest to each other and can be represented as:

$$(P_{i,m}, P_{j,n}) = \underset{P_{i,m}\in\mathcal{P}_{i,m}, P_{j,n}\in\mathcal{P}_{j,n}}{argmin} \|P_{i,m} - P_{j,n}\|_2 \tag{7}$$

Coordinate adjustment. Following the establishment of connections, the lines within the floor plan remain distorted and misaligned. We develop a coordinate adjustment algorithm employing a depth-first search (DFS) strategy to standardize the points. Fig. 3 shows the representation of the coordinate adjustment algorithm, which closely corresponds to the transition from the connection to the adjustment. The algorithm is as follows in Alg.1.

Door and window reconstruction. According to the coordinate information of the door and window descriptor, bind each door and window to the connection line containing its coordinates. After adjusting the coordinates, the center coordinates of the door and window descriptors are calculated based on the new coordinates of the binding

Algorithm 1 Adjust Points Coordinates

Input: $\mathcal{N} = \{B_1, B_2, \ldots, B_n\}$, set of all basic structural points with coordinate information $(B_n.x, B_n.y)$, $\mathcal{C} = \{L_1, L_2, \ldots, L_m\}$, set of point connections states, Alg.2.
Output: $\mathcal{R}$, set of points with adjusted coordinates

Select an initial point B_{init} from $\mathcal{N}$, and fix $B_{\text{init}}.x$ and $B_{\text{init}}.y$
Remove B_{init} from $\mathcal{N}$ and add it to $\mathcal{R}$
for each direction dir in {horizontal, vertical} **do**
    FIX_EXTEND($\mathcal{N}, \mathcal{C}, B_{\text{init}}, dir, \mathcal{R}$)
end for
while there exists a B in $\mathcal{N}$ with either $B.x$ or $B.y$ not fixed **do**
    for $i = 1$ to $|\mathcal{N}|$ **do**
        $B = \mathcal{N}[i]$
        if $B.x$ or $B.y$ has only one coordinate fixed **then**
            Fix the B to its current coordinate
            FIX_EXTEND($\mathcal{N}, \mathcal{C}, B, dir, \mathcal{R}$)
        end if
        if both $B.x$ and $B.y$ have fixed **then**
            Remove B from $\mathcal{N}$ and add it to $\mathcal{R}$
        end if
    end for
    if all B in $\mathcal{N}$ have both $B.x$ and $B.y$ coordinates fixed **then**
        break
    else
        Select a new B_{new} from $\mathcal{N}$, repeat from step 2
    end if
end while

Algorithm 2 Fix_Extend Function

for each L in $\mathcal{C}$ **do**
    if L connects B and the other end B_{other} is not fully fixed **then**
        if dir is horizontal **then**
            $B_{\text{other}}.y = B.y$
        else if dir is vertical **then**
            $B_{\text{other}}.x = B.x$
        end if
        Remove B_{other} from $\mathcal{N}$ and add it to $\mathcal{R}$
        FIX_EXTEND($\mathcal{N}, \mathcal{C}, B_{\text{other}}, dir, \mathcal{R}$)
    end if
end for

line information. Then, the door and window descriptors are replaced with standardized door and window icons. The direction of doors and windows descriptors is determined by the direction descriptors.

IV. Experiments

Our experiments verify the performance of our framework from three perspectives: descriptor recognition, reconstruction results, and robot downstream applications.

A. Hand-drawn floor plan databset

The hand-drawn floor plans dataset includes paper and electronic versions, comprising a total of 1000 images, with sizes between 200-4500 pixels. For the paper version, the images were drawn with a pen and subsequently photographed using a camera. For the electronic versions, the images were drawn on electronic devices through touch. We have conducted a questionnaire and the following definitions conform to the drawing habits of the majority of individuals without training (Fig. 6):

- Each person applies their drawing style. Multiple drawing styles are allowed on one image.
- Dataset includes common hand-drawn situations such as smudging, non-standard drawing, protrusions, and erasing.
- For the paper version, uncertainties include camera parameters, camera angles, backgrounds and shadows. For the electronic version, uncertainties include the devices, software and backgrounds.

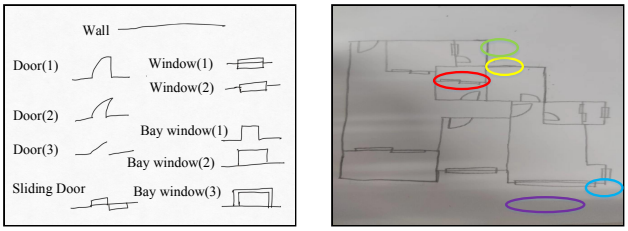

Fig. 6. The definition of hand-drawn floor plans. (a) The styles of hand-drawn floor plans includes wall, window, door, sliding door, bay window. (b) Examples of hand-drawn situations include: non-standard in red, smudging in yellow, erasing in green, protrusions in blue, and shadow in purple.

B. Structure descriptor recognition

The algorithm is implemented in PyTorch 1.9.0, with an input image size of 640×640. The model is trained for 400 epochs, with data augmentation enabled for the first 380 epochs. The base learning rate is set to 0.01/64, weight decay of 5×10^{-4}, and momentum of 0.9. The algorithms run on a computer with an Intel Core i7-11700F CPU and an RTX 3060Ti GPU.

1) Confusion matrix analysis: The confusion matrix reflects the relationship between the predicted results and the ground truth of descriptors, as shown in Fig. 7. From the predicted results, background misclassification dominates, with few classification errors occurring among the descriptors. From the ground truth, the prediction accuracy for each category descriptor exceeds 0.9, which satisfies the requirements of the reconstruction algorithm.

2) Comparison with recognition models: We compare our model with other advanced deep-learning models as shown in Tab. I. We measure the accuracy of the bounding boxes using AP, where AP50 focuses on detection results with $IoU \geq 0.5$. AP@50:5:95 calculates the average AP starting from $IoU \geq 0.5$, with steps of 0.05. It can be observed that the

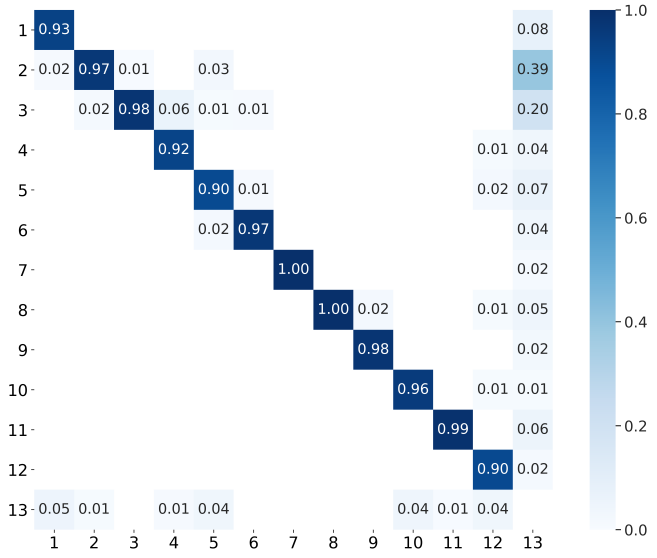

Fig. 7. The confusion matrix of our algorithm. 1: Normal point, 2: L point, 3: T point, 4: Cross point, 5: Window-wall, 6: Sliding door-wall, 7: Door, 8: Window, 9: Sliding door, 10: Bay window, 11: Door direction, 12: Window direction, 13: Background. The X-axis represents true labels, and the Y-axis represents predicted results.

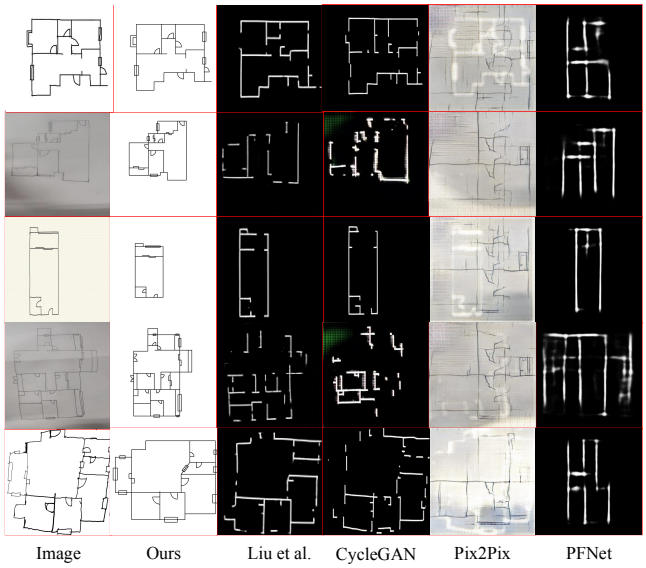

Fig. 8. Comparison of our algorithm with the results of Liu et al. [41], CycleGAN [35], Pix2Pix [34], and PFNet [42].

performances of our model are superior in most categories with mAP (0.893) and AP50@50:5:95 (0.498). Notably, for the basic structural point, our model is superior to other models. For the hand-drawn floor plan reconstruction, the results of the basic structural point descriptor are more important. In contrast, for the direction descriptors, it is only necessary to recognize the category correctly.

C. Standardized reconstruction algorithm

1) Deep learning algorithms design: Considering that the blueprint reconstruction algorithm fails in hand-drawn floor plans (Fig. 2), we design deep learning reconstruction comparison algorithms based on sketch synthesis. We view hand-drawn floor plans as input, directly synthesizing the reconstructed results using generative models.

2) Qualitative comparison: The reconstruction results of different methods are shown in Fig. 8. Deep learning models could restore most of the wall structures in the simple floor plan. However, it can be observed that in the complex hand-drawn floor plans, the structures might be lost. The model could not align lines horizontally or vertically and, worse yet, could not maintain the same direction as the original image. By comparison, our framework offers better reconstruction results and allows for more flexible editing.

3) Quantitative comparison: We quantitatively compare the standardized reconstruction results of advanced algorithms as shown in Tab. II. Our method with a reconstruction line accuracy of 0.932 and an angle reconstruction accuracy of 0.922, outperforming the deep learning algorithm. Comparing the door&window reconstruction accuracy, it is difficult for deep learning algorithms to predict the correct window/door position.

D. Efficiency analysis

After considering the requirements of real-world deployment scenarios, we conduct an efficiency evaluation of the proposed reconstruction framework. Specifically, we compute and comparatively analyze the efficiency of the structure descriptor recognition algorithm and the standardized reconstruction algorithm within the framework. As shown in Tab. III, the framework achieves an average of 0.545s for the electronic version and 0.444s for the paper version. Across both scenarios, the average computation time for the entire reconstruction framework is just 0.474s. It is particularly noteworthy that despite the standardized reconstruction algorithm's significantly longer average time, which is attributed to its integration of multiple complex sub-algorithms, we successfully manage to keep the total average time within 0.5s. This achievement meets the requirements for efficient practical applications, showcasing the potential to enhance robot downstream tasks.

E. Robot downstream application

We demonstrate the potential application of hand-drawn floor plan reconstruction for robot navigation by combining it with SLAM in two scenarios. In order to apply the reconstructed floor plan, we utilize the NBB [43] to match it with real-time LiDAR images. Subsequently, we meticulously adjust the rotation, flipping, and scaling of the floor plan to ensure alignment with the actual scenarios. As shown in Fig. 9 (a)-(d), we simulate a fire rescue scenario. In the absence of pre-built maps, a full map search strategy is applied as shown in (c), while applying our framework combined with map matching could reduce search paths (d). On the other hand, we simulate scenarios where dynamic objects obstruct and doors close as the robot creates a map for the first time (e)(f). This results in the robot failing to

TABLE I

COMPARISON WITH DIFFERENT RECOGNITION MODEL

AP50	ATSS-ResNet50 [38]	ATSS-ResNet101	FCOS [39]	YOLOv3 [40]	YOLOX [37]	Ours
window	0.942	0.959	0.967	0.959	0.901	0.904
door	0.987	0.998	0.979	0.998	0.909	0.999
sliding door	0.948	0.970	0.946	0.970	0.909	0.908
bay window	0.975	0.976	0.967	0.976	0.909	0.909
normal point	0.612	0.641	0.616	0.641	0.801	0.780
L point	0.839	0.826	0.871	0.826	0.861	0.863
T point	0.820	0.801	0.853	0.801	0.878	0.866
cross point	0.831	0.677	0.882	0.677	0.889	0.897
window-wall	0.796	0.820	0.882	0.820	0.797	0.892
sliding door-wall	0.828	0.868	0.946	0.868	0.908	0.909
window direction	0.898	0.927	0.934	0.927	0.909	0.909
door direction	0.714	0.881	0.896	0.881	0.894	0.885
mAP	0.852	0.862	0.876	0.862	0.878	0.893
AP50@50:5:95	0.439	0.390	0.443	0.390	0.494	0.498

TABLE II

COMPARISON OF THE ACCURACY OF DIFFERENT ALGORITHM RECONSTRUCTION RESULTS

Accuracy	Line	Angle	Door&Window
CycleGAN	0.415	0.537	/
Liu et al.	0.683	0.738	/
Ours	0.932	0.922	0.936

TABLE III

EFFICIENCY ANALYSIS RESULTS

Average time(s)	Recognition	Stanardized	Total
Electronic	0.067	0.478	0.545
Paper	0.037	0.407	0.444
Electronic and paper	0.046	0.428	0.474

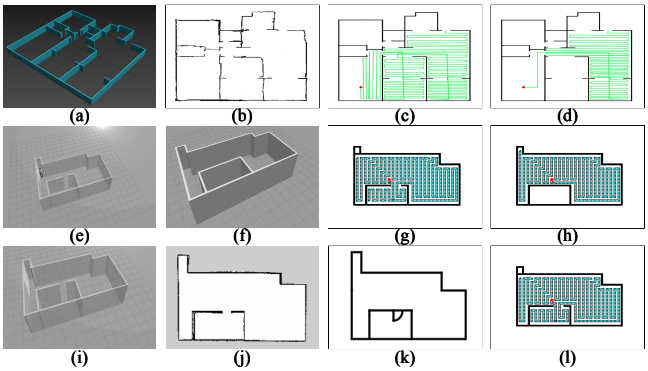

Fig. 9. Application results. (a)(e)(f) Fire, dynamic person, and closed door scenarios. (b)(j) Lidar map. (c) Traditional search. (d) Our framework applies to search. (g)(h) Traditional area coverage. (k) The reconstruction floor plan prompt. (l) Our framework applies to coverage.

cover these dynamic spaces in subsequent coverage tasks, even if objects are removed and doors are opened, (g)(h). However, hand-drawn floor plans provide information to prompt potential coverage spaces (k). The robot understands the scenarios and attempts to enter, thereby achieving more comprehensive coverage (l).

V. CONCLUSION

With the motivation of enabling indoor robots to quickly and comprehensively grasp the semantics of diverse and complex indoor scenes, this paper presents a hand-drawn reconstruction framework via structure descriptors guidance. Meanwhile, we demonstrate the framework's reconstruction results through ample experiments and apply the reconstructed maps to downstream tasks without pre-built maps, showcasing its potential to enhance robot scene understanding. We, therefore, anticipate that the outstanding performance will enhance the adaptability of robot application algorithms in unknown environments, e.g. rescue and exploration, particularly in complex scenarios. Nevertheless, this framework is currently limited to common indoor floor plans. In our future work, we will explore the applicability in more open scenarios, including unconventional floor plan structures and outdoor scene graphs.

REFERENCES

[1] J. R. Sánchez-Ibáñez, C. J. Pérez-del Pulgar, and A. García-Cerezo, "Path planning for autonomous mobile robots: A review," *Sensors*, vol. 21, no. 23, p. 7898, 2021.

[2] X. Li, J. Dezert, and X. Huang, "Selection of sources as a prerequesite for information fusion with application to slam," in *2006 9th International Conference on Information Fusion*. IEEE, 2006, pp. 1–8.

[3] C. Luo, S. X. Yang, X. Li, and M. Q.-H. Meng, "Neural-dynamics-driven complete area coverage navigation through cooperation of multiple mobile robots," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 1, pp. 750–760, 2016.

[4] X. Li, X. Li, M. O. Khyam, C. Luo, and Y. Tan, "Visual navigation method for indoor mobile robot based on extended bow model," *CAAI Transactions on Intelligence Technology*, vol. 2, no. 4, pp. 142–147, 2017.

[5] C. Luo, S. X. Yang, H. Mo, and X. Li, "Safety aware robot coverage motion planning with virtual-obstacle-based navigation," in *2015 IEEE International Conference on Information and Automation*. IEEE, 2015, pp. 2110–2115.

[6] X. Huang, X. Li, M. Wang, and J. Dezert, "A fusion machine based on dsmt and pcr5 for robot's map reconstruction," *International Journal of Information Acquisition*, vol. 3, no. 03, pp. 201–211, 2006.

[7] X. Li, X. Zhang, B. Zhu, and X. Dai, "A visual navigation method of mobile robot using a sketched semantic map," *International journal of advanced robotic systems*, vol. 9, no. 4, p. 138, 2012.

[8] C. Luo, J. Gao, X. Li, H. Mo, and Q. Jiang, "Sensor-based autonomous robot navigation under unknown environments with grid map representation," in *2014 IEEE Symposium on Swarm Intelligence*. IEEE, 2014, pp. 1–7.

[9] X. Li, X. Huang, and M. Wang, "Robot map building from sonar sensors and dsmt," *Information & Security Journal, Bulg. Acad. of Sci., Sofia*, vol. 20, 2006.

[10] V. Egiazarian, O. Voynov, A. Artemov, D. Volkhonskiy, A. Safin, M. Taktasheva, D. Zorin, and E. Burnaev, "Deep vectorization of technical drawings," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*. Springer, 2020, pp. 582–598.

[11] M. Liu, X. Li, J. Dezert, and C. Luo, "Generic object recognition based on the fusion of 2d and 3d sift descriptors," in *2015 18th International Conference on Information Fusion (Fusion)*. IEEE, 2015, pp. 1085–1092.

[12] P. N. Pizarro, N. Hitschfeld, I. Sipiran, and J. M. Saavedra, "Automatic floor plan analysis and recognition," *Automation in Construction*, vol. 140, p. 104348, 2022.

[13] X. Xing, C. Wang, H. Zhou, J. Zhang, Q. Yu, and D. Xu, "Diffsketcher: Text guided vector sketch synthesis through latent diffusion models," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[14] S.-I. Cheng, Y.-J. Chen, W.-C. Chiu, H.-Y. Tseng, and H.-Y. Lee, "Adaptively-realistic image generation from stroke and sketch with diffusion model," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 4054–4062.

[15] L.-P. De Las Heras, S. Ahmed, M. Liwicki, E. Valveny, and G. Sánchez, "Statistical segmentation and structural recognition for floor plan interpretation: Notation invariant structural element recognition," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 17, no. 3, pp. 221–237, 2014.

[16] Z. Zeng, X. Li, Y. K. Yu, and C.-W. Fu, "Deep floor plan recognition using a multi-task network with room-boundary-guided attention," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9096–9104.

[17] S. Kim, S. Park, H. Kim, and K. Yu, "Deep floor plan analysis for complicated drawings based on style transfer," *Journal of Computing in Civil Engineering*, vol. 35, no. 2, p. 04020066, 2021.

[18] Y. Tang, H. PAN, Y. Zhu, and X. Li, "A survey of image super-resolution reconstruction," *Acta Electronica Sinica*, vol. 48, no. 7, pp. 1407–1420, 2020.

[19] S.-h. Or, K.-H. Wong, Y.-k. Yu, M. M.-y. Chang, and H. Kong, "Highly automatic approach to architectural floorplan image understanding & model generation," *Pattern Recognition*, pp. 25–32, 2005.

[20] S. Ahmed, M. Weber, M. Liwicki, and A. Dengel, "Text/graphics segmentation in architectural floor plans," in *2011 International Conference on Document Analysis and Recognition*. IEEE, 2011, pp. 734–738.

[21] S. Ahmed, M. Liwicki, M. Weber, and A. Dengel, "Automatic room detection and room labeling from architectural floor plans," in *2012 10th IAPR international workshop on document analysis systems*. IEEE, 2012, pp. 339–343.

[22] H. Jang, K. Yu, and J. Yang, "Indoor reconstruction from floorplan images with a deep learning approach," *ISPRS International Journal of Geo-Information*, vol. 9, no. 2, p. 65, 2020.

[23] Z. Cai and N. Vasconcelos, "Cascade r-cnn: high quality object detection and instance segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 5, pp. 1483–1498, 2019.

[24] R. Dinesh *et al.*, "Handwritten electronic components recognition: An approach based on hog+ svm." *Journal of Theoretical & Applied Information Technology*, vol. 96, no. 13, 2018.

[25] G. Feng, C. Viard-Gaudin, and Z. Sun, "On-line hand-drawn electric circuit diagram recognition using 2d dynamic programming," *Pattern Recognition*, vol. 42, no. 12, pp. 3215–3223, 2009.

[26] X. Li, X. Li, Z. Li, X. Xiong, M. O. Khyam, and C. Sun, "Robust vehicle detection in high-resolution aerial images with imbalanced data," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 3, pp. 238–250, 2021.

[27] L. Gennari, L. B. Kara, T. F. Stahovich, and K. Shimada, "Combining geometry and domain knowledge to interpret hand-drawn diagrams," *Computers & Graphics*, vol. 29, no. 4, pp. 547–562, 2005.

[28] S. Roy, A. Bhattacharya, N. Sarkar, S. Malakar, and R. Sarkar, "Offline hand-drawn circuit component recognition using texture and shape-based features," *Multimedia Tools and Applications*, vol. 79, pp. 31 353–31 373, 2020.

[29] R. Lakshman Naika, R. Dinesh, and S. Prabhanjan, "Handwritten electric circuit diagram recognition: an approach based on finite state machine," *Int J Mach Learn Comput*, vol. 9, pp. 374–380, 2019.

[30] M. Dey, S. M. Mia, N. Sarkar, A. Bhattacharya, S. Roy, S. Malakar, and R. Sarkar, "A two-stage cnn-based hand-drawn electrical and electronic circuit component recognition system," *Neural Computing and Applications*, vol. 33, pp. 13 367–13 390, 2021.

[31] S. Lu, Y. Liu, and A. W.-K. Kong, "Tf-icon: Diffusion-based training-free cross-domain image composition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 2294–2305.

[32] Y. Peng, C. Zhao, H. Xie, T. Fukusato, and K. Miyata, "Difffacesketch: High-fidelity face image synthesis with sketch-guided latent diffusion model," *arXiv preprint arXiv:2302.06908*, 2023.

[33] C. H. Wu and F. De la Torre, "Making text-to-image diffusion models zero-shot image-to-image editors by inferring" random seeds"," in *NeurIPS 2022 Workshop on Score-Based Methods*.

[34] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[35] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[36] C. Meng, Y. He, Y. Song, J. Song, J. Wu, J.-Y. Zhu, and S. Ermon, "Sdedit: Guided image synthesis and editing with stochastic differential equations," in *International Conference on Learning Representations*, 2021.

[37] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," *arXiv preprint arXiv:2107.08430*, 2021.

[38] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S. Z. Li, "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9759–9768.

[39] Z. Tian, C. Shen, H. Chen, and T. He, "Fcos: Fully convolutional one-stage object detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 9627–9636.

[40] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[41] R. Liu, Q. Yu, and S. X. Yu, "Unsupervised sketch to photo synthesis," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. Springer, 2020, pp. 36–52.

[42] H. Mei, G.-P. Ji, Z. Wei, X. Yang, X. Wei, and D.-P. Fan, "Camouflaged object segmentation with distraction mining," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8772–8781.

[43] K. Aberman, J. Liao, M. Shi, D. Lischinski, B. Chen, and D. Cohen-Or, "Neural best-buddies: Sparse cross-domain correspondence," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, pp. 1–14, 2018.