8
Conclusions

Indoor space understanding is an essential field of study as it serves multiple requirements in the current time. These requirements can be reconstruction or renovation of an existing house or building, planning of interior design, indoor navigation, description of a property to be submitted to a real estate website, AR/VR applications, architectural solutions, etc. However, a system could completely understand an indoor space by giving its holistic interpretation in a textual narration or any other representation. The indoor space images consist of a lot of subjectivity and lack global information to describe the entire house/building. Hence, there was a requirement of a representation of the house, which could be understood and interpreted. Floor plans are the heart of constructing a building that represents an indoor space and captures the intricate details. Hence, understanding a floor plan and generating its interpretation is proposed in this thesis. A floor plan generation scheme is proposed, which produces floor plans to interpret the indoor space using its RGB images taken by a conventional mobile phone. In this chapter, the contributions made for each approach are summarized, and their strengths and limitations are discussed. Finally, several strategies to improve the proposed techniques are discussed in future work.

In this thesis, classical machine learning techniques and advanced deep learning methods have been proposed to understand floor plans. To experiment with classical machine learning methods, novel features BoD and LOFD are presented, representing the individual rooms in a floor plan based on the decor contents. Experiments are done with several classifiers and artificial neural networks, generating labels for each room. Techniques for decor localization and classification are proposed using hand-crafted features and features learned by a deep neural network. Subsequently, a description is generated from the information extracted using grammar-based and learning-based models.

A system to automatically understand the indoor space from its RGB images and generate a pictorial interpretation of a floor plan is also proposed. In the proposed methodology, the floor plan for the entire indoor space is developed from the RGB images taken from a conventional mobile phone's camera and built-in SLAM technology in the Google ARcore library. The proposed system gives a user the freedom to use a traditional mobile phone without relying upon the depth-sensing hardware or advanced camera images such as panoramic or spherical. Extensive experimental evaluations were performed by comparing the area and aspect ratios of the generated layouts with manually generated layouts using multiple devices and existing layout generation applications.

8.1 CONTRIBUTIONS

The key contributions of this thesis are:

- In Chapter 3, state-of-the-art datasets of floor plans are discussed. Annotation and augmentation done on existing dataset ROBIN for the decor classification and room label detection schemes proposed in this thesis are described in detail. A novel large scale dataset, BRIDGE, is presented, which contains over 13000 floor plan images and annotations for decor symbol detection, region-wise caption generation, and paragraph description for each image. Different experiments for decor detection, region-wise caption generation, and paragraph generation are shown with the annotations proposed to prove the proposed dataset's utility.

- In Chapter 4, a novel end to end framework for understanding and generating a textual interpretation of annotated floor plans is proposed. In this work, the annotated ROBIN, discussed in Chapter 3, is used for experimentation and validation. Floor plan images in ROBIN are annotated with text labels for each room, and the walls are annotated with textures for different construction materials. An ego-centric textual narration is generated after extracting information about each component of the floor plan.

- In Chapter 5, a novel framework for understanding and interpreting an un-annotated floor plan is proposed, which takes advantage of classical machine learning techniques. Two novel features, BoD and LOFD, are presented, which capture the entire floor plan's global information and represent each room in terms of its decor contents. In this chapter, an improvement is made for the existing decor characterization technique, and a comparison is presented with other existing state-of-the-art schemes. An algorithm for door-to-door obstacle avoidance indoor navigation is also proposed, identifying the path within a floor plan and narrating it with the general description generated. The proposed scheme takes an edge over existing methods as it improves the model of floor plan understanding by not requiring any textual annotations or prior knowledge.

- In Chapter 6, a knowledge-driven model for automatic understanding and interpreting floor plans is proposed. In this chapter, advanced deep learning models are proposed to classify rooms in the floor plan images, and characterize decor symbols. A hierarchical recurrent neural network-based model (DSIC) is proposed, which directly takes visual cues from image and map to it the respective textual description, generating a paragraph-based description for each floor plan. A knowledge-driven description generation model (TBDG) is also proposed, taking word cues as input and visual cues. The proposed model is more robust than previously submitted models since the floor plan is a technical drawing and contains less subjectivity than a natural image. Hence, specific word cues boost the description generation's accuracy and make the generated description closer to human written descriptions. Extensive experiments are done on several metrics such as ROUGE, METEOR, BLUE, with all the models proposed. Dataset BRIDGE is used for the experimentation and validation of the proposed methods.

- In Chapter 7, a novel framework, GRIHA for automatic understanding and interpreting the indoor space directly from RGB images in the pictorial form, i.e., floor plan, is proposed. The existing solution for floor plan generation largely depends upon depth-sensing hardware and advanced visuals such as videos, panorama, and spherical images. However, in the proposed method, these requirements are released. The floor plan is generated from RGB images of indoor space taken from a conventional mobile phone's camera and camera motion information extracted from built-in SLAM in the Google ARcore library. This method gives users the freedom to use conventional devices and automatically interpret the indoor space.

8.2 SCOPE OF FUTURE WORK

The work presented in this thesis opens the scope of a chain of tasks in many diverse applications for researchers to solve in the future. Some of them are given below.

- In the future description generation from floor plan images, more advanced deep learning models can be used, which could use more intricate details from floor plans. The generation methods can be made more generalized for diverse floor plan images including more heterogeneity. The description generation framework can be made real-time in the future, as a person could upload a few images and descriptions could be generated simultaneously.

- In the future, researchers may explore the possibility of the generated description from the floor plan images to be used for several other tasks like automatic tagging, multimodal retrieval. Additionally, the proposed models can be extended for multi-story buildings connecting information from multiple floor plans.

- Having a digitized floor plan always is not possible. There might be situations where a printed floor plan for old buildings has to be scanned. While scanning a printed document, there are chances of getting noise in the digital copy. This noise can also occur because of the old floor plan's existing noise, which might have a yellow tint or quality degradation. Hence, noise removal while understanding the floor plan as an end to end process can be explored as future work.

- In future the proposed work can be adapted to other domains in document analysis and recognition. Areas such as natural language reasoning over scientific plots, graphs, charts, mathematical equations, textbook data can be a few such examples. The proposed algorithms can be adapted for these domains as well. These domains can be used for various other multimodal tasks such as visual question answering, natural language inferencing, visual entailment, question generation from images etc.

- In future, the proposed work can be extended for machine based refinement of the description of the floor plans. A Human In The Loop (HITL) based machine learning model can be proposed where human experts can provide some rough description for the given floor plan and the machine participates to refine/ fine-tune the provided description by making corrections and elaborate it with machine understanding of the engineering drawings.

- The Manhattan assumption is always not valid in terms of an indoor space. The proposed framework of floor plan generation from RGB images can be improved by releasing the Manhattan world assumption. More generalized shapes for rooms can be captured, and more realistic floor plans can be obtained as future work.

- Also, enhancing the depth estimation method for occluded regions is another area of improvement and can be explored. This will help users improve accuracy in terms of aspect ratio for occluded indoor spaces such as offices and laboratories heavily equipped with experimental facilities.