3
Dataset of Floor plans

In the field of document image analysis, a floor plan is a document containing a drawing of the house, apartment, or any other building. These are documents that aid architects to show the interior of a building along with components. Hence, understanding every component of these documents is necessary to generate an interpretation in any form. Floor plan image understanding involves semantic segmentation, symbol spotting, and identifying a relationship between them. Tasks such as symbol spotting, thick and thin wall classification, door and window detection, room and sub-room detection are various aspects in floor plan image analysis. These are the set of graphical parts that are required to be identified accurately to generate a comprehensive interpretation. In Fig. 3.1 shows various research possibilities for floor plan understanding. By analyzing floor plan images, tasks such as symbol spotting, retrieval, segmentation, text generation from floor plans, floor plan generation from the text, and indoor scene images can be done.

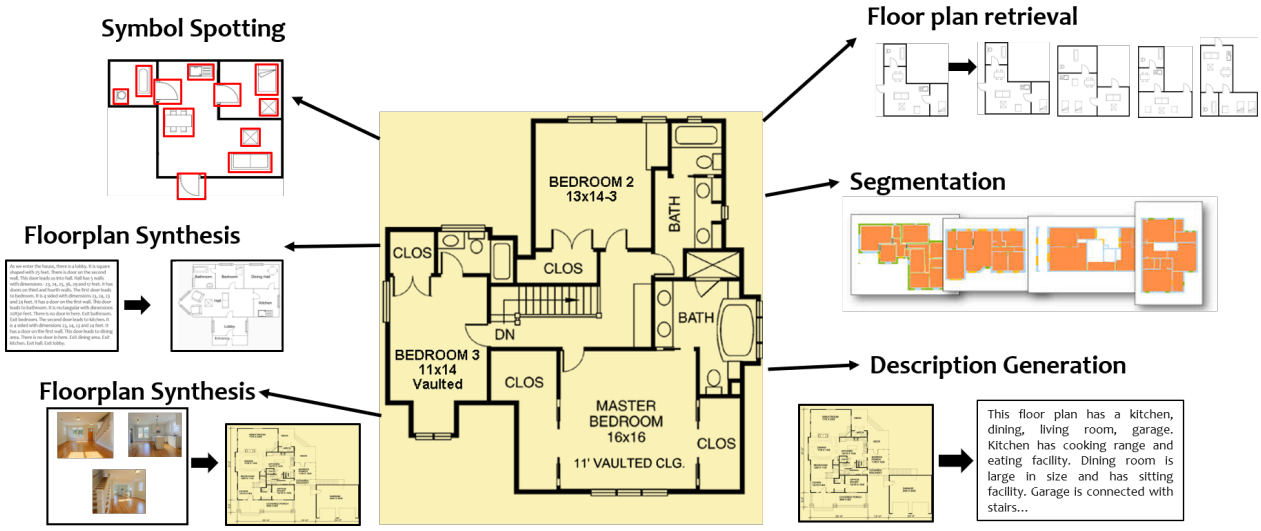


Figure 3.1 : Various tasks performed on floor plan images.

All the tasks highlighted in Fig. 3.1, require good datasets for their implementation and experiments. Datasets in any research play a pivotal role and the lack of which can act as a bottleneck situation. Hence, good quality datasets are much required to investigate these problems further and overcome the current shortcomings. These datasets should include a large number of images and respective annotations for the specified applications. With the advent of advanced deep learning and artificial intelligence models, there is a requirement for datasets with a large number of images and annotations. For training and experimenting with larger models to make them more generalized, it is required to have large-scale datasets that could cater to multiple tasks. Hence, in this chapter, we discuss a large-scale floor plan dataset BRIDGE, which contains $\sim$ 13000 floor plan images along with annotations for various tasks. This is a novel dataset that bridges the gap

between two modalities such as document images and natural language text. We also discuss other datasets available in the context of floor plan images and natural images.

The remaining of the chapter is divided in Sec. 3.1 giving a brief overview of the datasets in the literature and the requirement of BRIDGE dataset, Sec. 3.2 and Sec. 3.3 describes the annotation and augmentation of the existing dataset ROBIN for the floor plan understanding the schemes proposed in this thesis. Sec. 3.4 describe the construction of the BRIDGE dataset, Sec. 3.5 describes the various experiments performed over the proposed dataset and Sec. 3.6 summarize the chapter.

3.1 BRIEF OVERVIEW

In the literature, the number of floor plan datasets, which are publicly available for research is four. They are: ROBIN Sharma *et al.* [2017], CVC-FP de las Heras *et al.* [2015], SESYD Delalandre *et al.* [2010], and FPLAN-POLY Barducci and Marinai [2012]. Figure. 3.2 shows sample images from these datasets, where Fig. 3.2 (a), (b), (c), (d) are from SESYD, CVC-FP, ROBIN and BRIDGE respectively. However, the datasets SESYD, CVC-FP, ROBIN, FPLAN-POLY contain a very less number of sample images, various symbols for furniture, doors, and windows and do not contain annotations for objects, and their descriptions. These datasets are primarily constructed to find the solution for problems such as image segmentation, retrieval, and layout analysis. These datasets are not suitable for the purpose of caption generation and description synthesis. Nowadays, with the advent of deep neural networks, tasks such as symbol spotting, caption generation, retrieval, and semantic segmentation are getting more accurate and robust. However, there is a requirement for a large number of samples and corresponding annotations specific to each task for training these models. There are many large-scale datasets publicly available in the literature in the context of natural images. Examples include visual genome Krishna *et al.* [2017], MS COCO Lin *et al.* [2014], which has a large number of natural images along with their descriptions or captions, object annotations, region graphs, and other metadata. Since the publicly available floor plan datasets are not having a large number of image samples, the deep learning models perform poorly when trained and tested on them.

To perform the above-mentioned tasks efficiently with deep neural networks, a large-scale dataset of floor plan images along with their task-specific annotations is the need of the hour. In this work, we propose the BRIDGE (Building plan Repository for Image Description Generation, Evaluation, and other purposes) dataset. The rationale behind the nomenclature is that the proposed dataset bridges two modalities, i.e., image and text for floor plan images. In the domain of floor plan datasets, BRIDGE contains a collection of 13000+ floor plan images. Annotations for symbols, region descriptions, and paragraph descriptions are part of the dataset. The construction of BRIDGE is inspired by the visual genome dataset, which includes annotations for similar tasks in the context of natural images. In this chapter, along with the features of the dataset, we also present experimental results on tasks such as symbol spotting, region-wise caption generation, and paragraph generation, using deep learning models. Along with the construction of the BRIDGE dataset, the existing dataset is also annotated and augmented for performing/ evaluating various sets of experiments with hand-crafted features and classical machine learning techniques. In the proposed thesis, floor plan understanding is done for annotated floor plans, un-annotated floor plans using machine learning methods, and generalized floor plans using deep learning algorithms. In the coming sections, Annotated-ROBIN, A-ROBIN, and BRIDGE datasets are explained in detail to perform the aforementioned tasks.

3.2 ANNOTATED-ROBIN

To understand the annotated floor plan dataset, ROBIN was annotated with wall material, textures and text data. Since there was no dataset available with textual description or textual labels of floor plans, the existing floor plans in ROBIN were annotated using three wall material symbols, concrete, glass, and wood. Figure. 3.3 shows the sample annotated floor plans. They also had the textual annotation as labels for each room name. In chapter 4, floor plan understanding technique using annotated floor plans is proposed and experiments are performed on Annotated-ROBIN.

3.3 A-ROBIN

To understand the floor plan images better, and produce an automated textual description narrating them, we propose a dataset *Augmented* ROBIN (A-ROBIN). In A-ROBIN there are four

Figure 3.2 : Sample images of floor plans in some publicly available datasets.

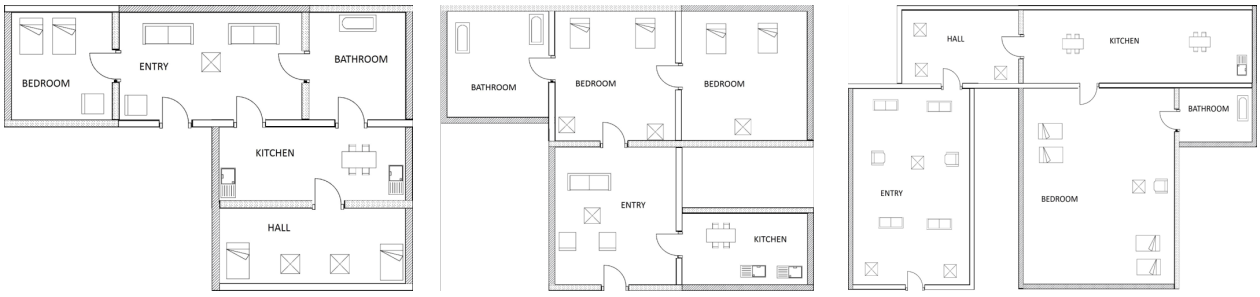
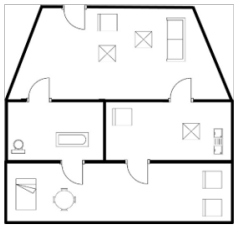


Figure 3.3 : Sample images of floor plans in annotated ROBIN dataset.

human written descriptions for each image in ROBIN dataset. In the literature, there are a few image datasets, for example, Flicker8k Hodosh *et al.* [2013], COCO dataset Chen *et al.* [2015b], which has associated descriptions of the image samples. However, and these datasets describe the natural images. In the context of document images, a dataset consisting the description of floor plan images was lacking. Hence, we require a novel dataset which contains annotations and descriptions for document images. The descriptions for a floor plan image were collected from the volunteers in the following manner. Each volunteer was supplied with a set of 10 images in a



Description 1

The floor plan has 4 rooms. Rooms are entry, bathroom, bedroom and kitchen. Entry is in the north and has two tables in center, one small sofa in north and one large sofa in east corner. Bathroom is in east direction and has a tub and a sink in south west side. Kitchen is in west direction and has a sink, table in west side and sofa in east side. Bedroom is in south direction and has a bed, a table in west corner and two small sofa in the east corner. Entry is the largest room in floor plan, while bedroom being the second largest. Bathroom and kitchen are medium sized rooms.

Description 2

The entry door in the floor plan leads to the drawing room which is the largest room in the house. It contains a big sofa adjacent to the east wall and two square tables and a small sofa around it. This room leads to two rooms, out of which one is a bathroom containing a tub and a sink and the other is a kitchen containing a big sink to the right, a table and sofa to the left of the sink. The bathroom and kitchen give way to one more room that is the bedroom containing a bed and a round table adjacent to the west wall and two sofas near to the right wall. The shape of all the rooms is rectangular except the first room which resembles a parallelogram with two sides not parallel to each other.

Description 3

The house is has 4 rooms. Entrance is a big hall having 4 decors items. There is a big sofa to east wall, small sofa, and two tables in center. Neighboring rooms are bathroom and kitchen. Kitchen has a sink to east wall, a table in center and a small sofa adjacent to west wall. Bathroom has a tub and a large sink which are adjacent to south wall. Bedroom has a bed aligned to west wall and a round table in its side. While there are two small sofa aligned to east wall.

Description 4

In the given plan there is a big size hall located at west direction. It has main entry to the house. The hall contains 2 coffee table located in the middle, 1 small sofa beside the main entry door and one large sofa in the east corner wall. There are two other doors in the hall which are connected to the bathroom and one living room. Both bathroom and living room are besides of each other. Size of the bathroom is small and it is situated in the north direction. The bathroom contains one tub and one large sink situated in the opposite corners of the bathroom. The living room is situated on the east side beside the bathroom. The living room contains one small sofa and one coffee table and one twin sink. Both bathroom and living room have one more entry door to a big size bedroom. This bedroom is situated in the south direction. The bedroom contains one bed and one roundtable on the north corner wall and 2 small sofas on the east corner wall.

Figure 3.4 : Example of a sample floor plan image from ROBIN dataset and the annotation collected for the same to synthesize A-ROBIN.

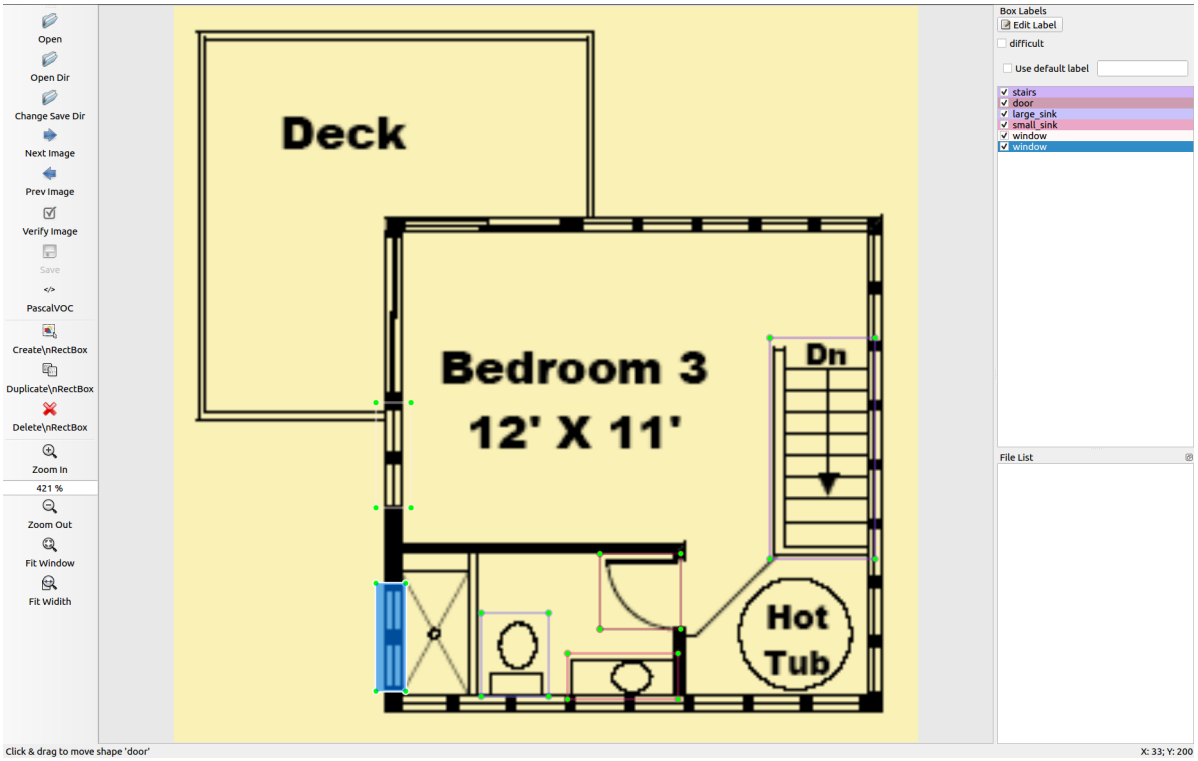


Figure 3.5 : Graphical User Interface of LabelImg tool for annotation proposed by Tzutalin [2015].

Google form and asked to describe them in their own words based on a set of instructions. Each form was given to 4 volunteers, so that each image has 4 set of a descriptions. These descriptions focus on the rooms and their decor content in the floor plan images. Also they focus on their relative positioning in the respective rooms and the relative position of each room within the floor plan image. These descriptions vary in the sequence of information given, the details of the information provided, the sentence conjunctions, and the vocabulary used for the components in floor plan image. Figure 3.4 shows one of the floor plan and its corresponding descriptions. It can be observed that different descriptions of the same image vary in the amount of information provided, the sequence of describing each room, and the names for each decor item could be different for different users. Moreover, a room may have variations in its name for different user. The length of the descriptions provided for each image is also varied. After the descriptions were collected, each set of 4 description were tagged to their respective images using image identifiers.

The volunteers were directed to annotate the: (1) entire plan (number of rooms, name of the rooms), (2) individual rooms- name, decor contained, relative size (small, medium, large, etc), (3) relative positioning of decors (north, east, west, south, aligned with wall, adjacent with something, etc), (4) relative position of each rooms (north, east, west, south, adjacent with bathroom, north of kitchen, etc). Apart from these information a volunteer may also provide additional information as per his/her choice. For 510 images in ROBIN datasets, there were 2040 descriptions collected in total. The dataset is tokenized and preprocessed for further processing. In our experiments, we have compared machine generated description with these descriptions and an analysis regarding the closeness of machine translation with human written descriptions is given. Next, we describe the steps in our framework in detail.

3.4 CONSTRUCTION OF BRIDGE

To construct the BRIDGE dataset we have followed two approaches. First, we have collected floor plan images from the publicly available datasets (i.e., ROBIN, SESYD). In the second approach, we have collected the remaining floor plan images from the internet. In total, we have over 13000 floor plan images in this dataset. Along with the images BRIDGE also has object annotations, region descriptions, and paragraph description for the floor plans. Till date, this is the largest annotated floor plan dataset created for the document analysis and research (DAR) community. For creating annotations, we asked volunteers for marking bounding boxes around each decor item. We used LableImg graphical annotation tool proposed by Tzutalin [2015] for marking the bounding boxes in the images. Figure. 3.5 shows the screenshot of the annotation tool LabelImg, with annotation of a floor plan image. For generating region descriptions also we used the same tool and later converted them to the JSON format.

3.4.1 Floor Plan images

Along with the images obtained from available public datasets, images were collected from two websites `www.architecturalhouseplans.com` and `www.houseplans.com`. These websites contain multiple floor plan images for a single house design for both single storied and multi-storied buildings. The similarity between the images taken from both websites is that the floor plans belong to real homes, available for a customer to use and they are not generated for any specific task, for example, retrieval or segmentation. They are similar to the symbols of the objects used in them. There is a diversity between the layouts and complexity of the plans. Each image in the dataset has a unique ID (image name) which depicts the source where it has been obtained from along with the floor plan images of the same house for the different floors. Figure 3.2 (d) shows some of the sample floor plan images available in BRIDGE. There is a lot of variability in the formation of the images in the dataset. Floor plan images from SESYD dataset are synthetically generated, while images from ROBIN are handcrafted. Floor plan images collected from the web are standard plans designed by the architects. The images taken did not have any ground truth data or annotations available which are required of any machine learning algorithm to train. Hence, we further proceeded to create an annotation for various task and evaluation purposes.

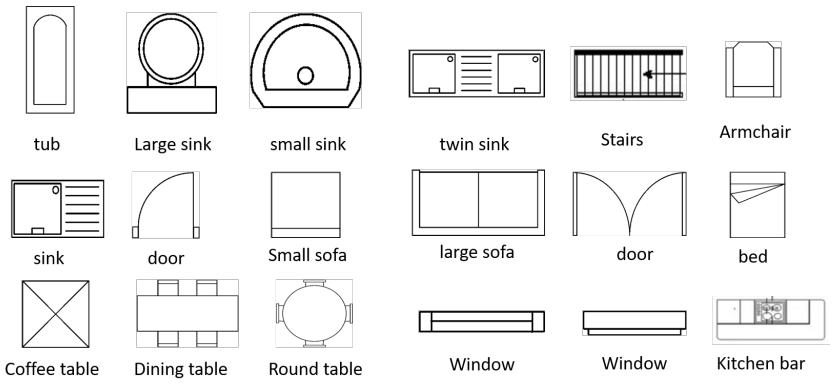


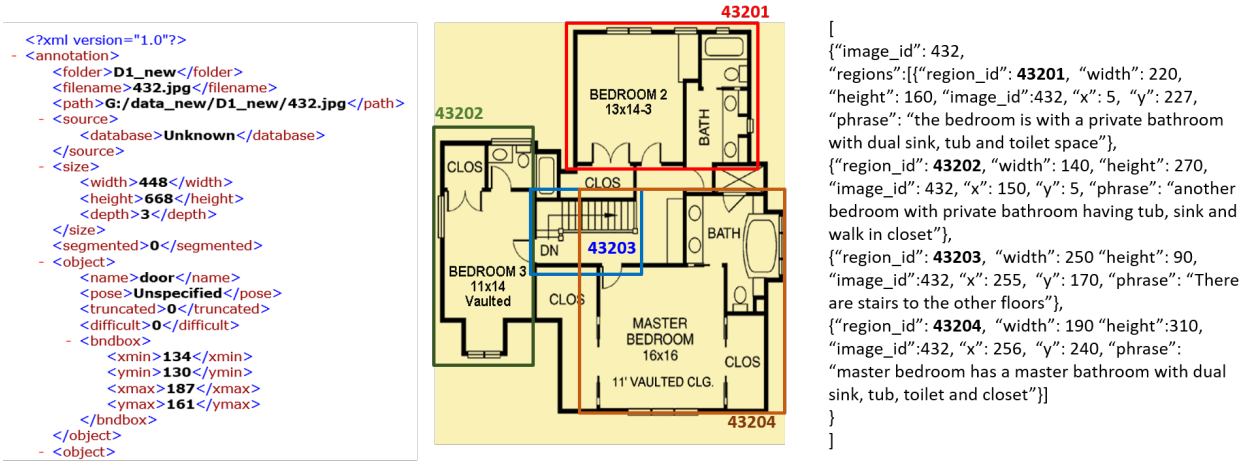
Figure 3.6 : Various classes of decor items available in the dataset.

3.4.2 Symbol Annotations

Detection of several decor items is an important step when for parsing a floor plan image and information extraction. Object detection schemes have been used in the context of objects in natural images. In the line of architectural drawings, technique involving handcrafted features is used multiple times in the literature. However, techniques using deep neural networks are still needed to be explored in the context of document images. For the task of symbol detection and localization, symbol annotations were required. Figure 3.6 shows various decor symbols for 16 classes of objects in the proposed dataset. Each annotation file is in XML format with information such as image name, path, and the bounding box for each decor symbol with their names. Figure 3.7 (a) shows the sample annotation XML for objects. All of the symbol annotations were generated using LabelImg tool Tzutalin [2015].

3.4.3 Caption Annotations

In the literature, there are image datasets with image captions (MS-COCO) and region-wise captions (visual genome). For a floor plan, region wise caption generation is an important step. In the work proposed by Goyal *et al.* [2018a,b], the authors have used handcrafted features for identifying decor symbol, room information and generating region wise caption generation. Deep neural networks, like CNN, RNN, and LSTM have shown superior performance for natural images for the same task. However, the same can not be translated for floor plans due to the lack of data and pre-trained models. To compensate this issue we annotated image regions by describing them in the form of a dictionary. The region descriptions are inspired by the region descriptions provided in the visual genome dataset. The regions in the floor plan images are taken as different rooms, for example, bedroom, kitchen, living room, etc. For each image, region description contains a region ID which is unique over the entire dataset. Additionally, coordinates of the bounding box for each region as *x*,*y* coordinate of the top left point, height and width along with a describing caption with the field "phrase" is also given. Figure 3.7(c) shows the sample region descriptions created for one of the floor plan images in Fig. 3.7(b)



(a) Symbol annotation (b) Regions in floor plan (c) Region Description

Figure 3.7 : Illustrations of the several steps of annotation process.

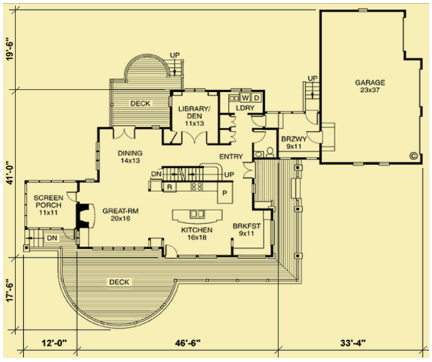


Figure 3.8 : An example of paragraph annotation for a floor plan image.

3.4.4 Paragraph based description annotations

Going further in the line of describing images with captions as well as in the form of paragraphs, there was a requirement of the paragraph based textual descriptions which described images in free flow with variability. Previous works proposed in Goyal *et al.* [2018a] Goyal *et al.* [2018b] have generated template-based paragraph description which was generated by concatenating region based captions. To have more variability in description, there is a requirement of using RNN based deep learning models which require a bulk amount of data for training. Hence, we annotated floor plan images with paragraph based annotations. Along with the images, we have also collected the descriptions given on the web pages for each floor plan image. The images which did not have descriptions on them were written by the volunteers. The paragraph descriptions were collected in the form of raw text and cleaning was done by removing extra white spaces, alpha-numeric characters, new line characters, and non-ASCII characters by using NLTK tool. The paragraph annotations are converted into a JSON file for free flow description synthesis of floor plan images. Figure 3.8 shows paragraph annotation for the image shown. Table 3.1 shows the statistics of each paragraph annotation available in BRIDGE dataset. The average length of each paragraph in the word count is 116 with average length of each sentence being 5. It also shows the count of diversity, which is a measure of the richness of words used in sentences, nouns, verbs, adjectives, and proper nouns available in the dataset. Figure 3.9 shows the frequency distribution of the most frequent words in the paragraph annotations. It can be seen that the words chosen for annotating the category of rooms and decors are highly available in the dataset. In the next section, we demonstrate experimental results for various state-of-the-art techniques to prove the usability of our proposed dataset.

3.5 EXPERIMENTS

The experiments on the proposed dataset were performed to demonstrate its utility for the targeted problems on floor plan images. All the experiments were performed on a system with NVIDIA GPU Quadro *P*6000, with 24 GB GPU memory, 256 GB RAM.

Table 3.1 : Statistical details of the paragraph annotations

Statistics	Paragraphs
Average Length	116
Average no of sentence	5
Diversity	121
Nouns	134942
Verbs	5027
Adjectives	46379
Proper nouns	5476

3.5.1 Symbol Spotting

The symbol spotting algorithms are needed regarding identifying the decor and other symbols in the floor plan images. In the context of floor plan images, symbol spotting techniques using handcrafted features have been used widely in the literature. With the application of deep neural networks, the authors in Ziran and Marinai [2018] have explored Faster-RCNN techniques for symbol spotting on a very small scale dataset. However, the aforementioned task with deep neural networks, trained on a large-scale dataset, is hardly explored. We experimented with YOLO and Faster RCNN methods by fine-tuning the pretrained network on the proposed dataset. Fine tuning was done using 2500 annotations and testing was done with another 500 annotation. Figure 3.10 shows the distribution of various symbols over the training dataset. Results of the symbol spotting on BRIDGE are described next.

YOLO

YOLO is a single convolutional network, which simultaneously predicts multiple bounding boxes and class probabilities (confidence value) of those boxes. It defines confidence score as $Prob(object) * IoU$, where IoU is the intersection of union between the predicted bounding box and the ground truth bounding box (Eq. 3.1) and $Prob(object)$ is the probability of detecting the object in that bounding box. We fine-tuned the pre-trained tiny YOLO network using the BRIDGE dataset for 16 classes of objects. The original network has 9 convolutional layers, having max-pooling layers in between, where the final layer had 105 filters (for our dataset) and a linear

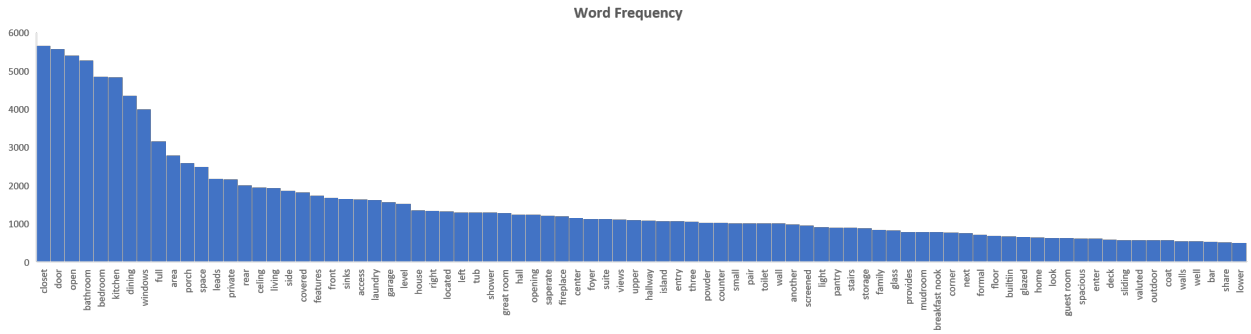


Figure 3.9 : Frequency distribution of the most frequent words occurring in the paragraph corpus.

activation function. Figure 3.11 (a) shows the average precision observed for each object after testing. Figure 3.12 shows the detected and localized objects using YOLO with their respective confidence score.

$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Union} \tag{3.1}$$

Faster RCNN

There are two modules in Faster RCNN, (i) A deep fully convolutional region proposal network, which proposes regions, (ii) a Fast-RCNN detector. The proposed regions are used for their classification. Figure. 3.11(b) shows the average precision observed for each object using faster-RCNN technique. Figure 3.13 shows the detected and localized objects using faster-RCNN and respective confidence score.

3.5.2 Caption generation

Captioning an entire image is a task which has been explored widely on natural images. A caption is a single line sentence consisting of information of the entire image. However, in case of floor plan images, generating a single line caption for the entire plan is insufficient for accurate representation. Hence it is required that captions to be generated region wise. Dense captioning is one such task which generates a caption for regions over the images. In this context, symbol spotting is a special case of dense captioning where the target labels generate one word. Along with fully convolutional network, the network has a fully convolutional localization layer which proposes a region of interest and respective confidence score. Recognition network and RNN language model succeed the previous networks where regions are refined by the earlier one and captions are generated by the later one. Figure 3.14(a) depicts the requirement of a trained dense captioning model on document images. The available data and pre-trained networks do not work on floor plan images because they are trained on natural images. Hence, there was a need of a dataset having floor plan

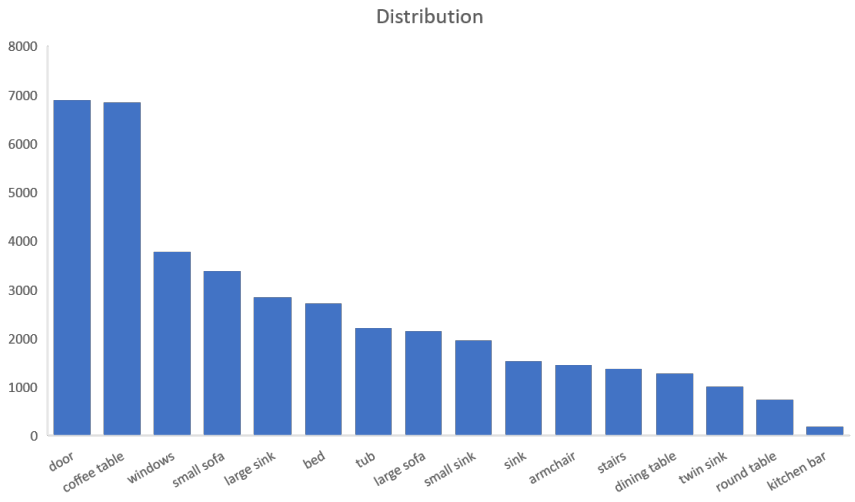
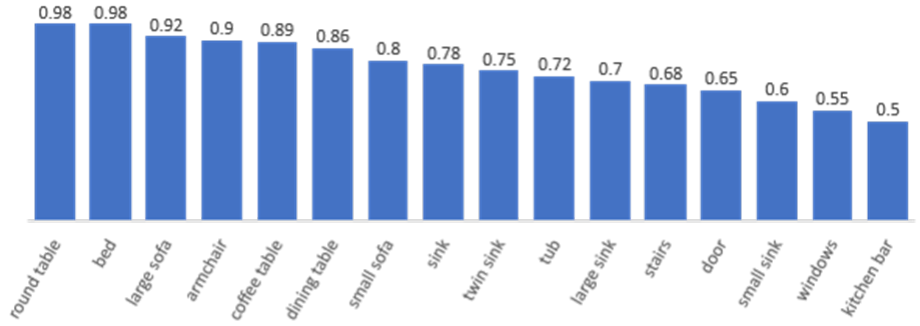
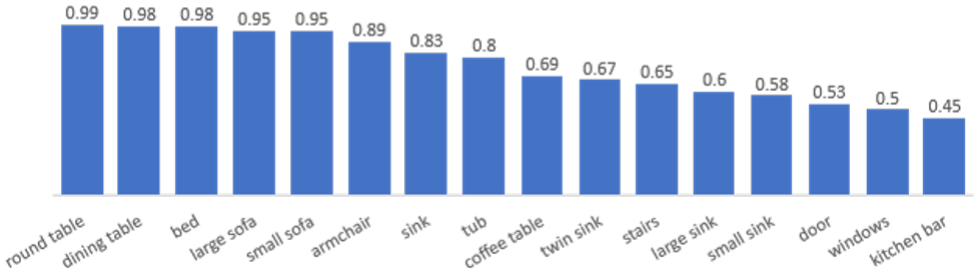


Figure 3.10 : Distribution of various objects in the training dataset.



(a) mAP=76.66 for YOLO



(b) mAP=75.25 for Faster-RCNN

Figure 3.11 : Quantitative analysis of the symbol spotting task.

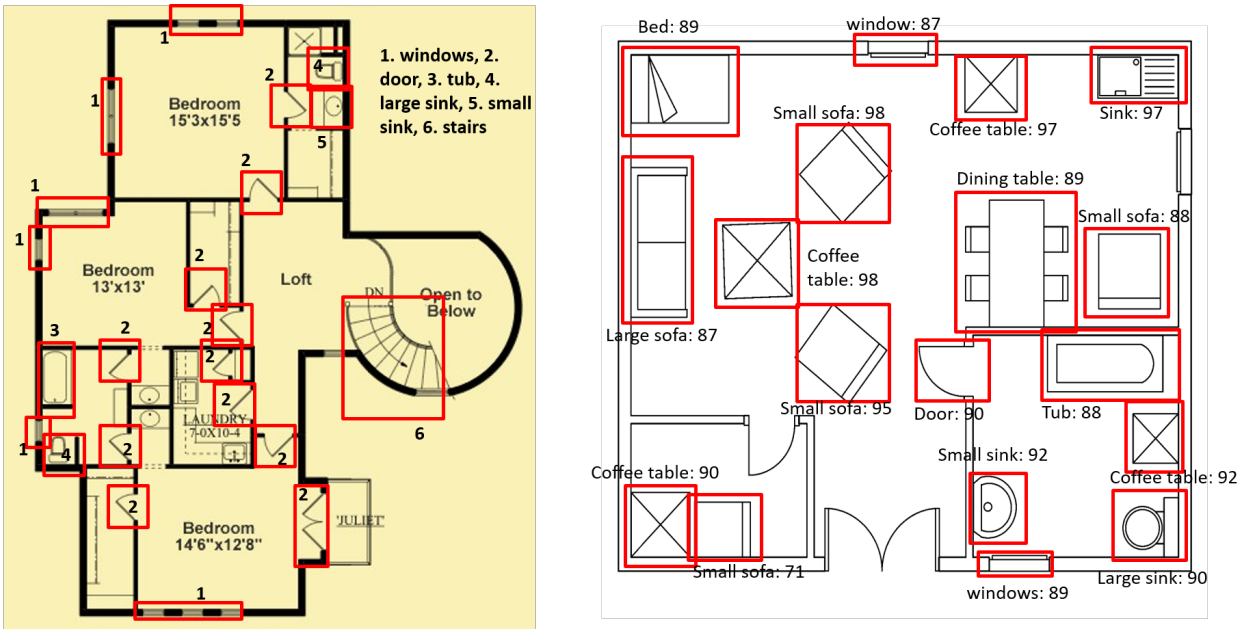


Figure 3.12 : Qualitative results of symbol spotting using the YOLO model.

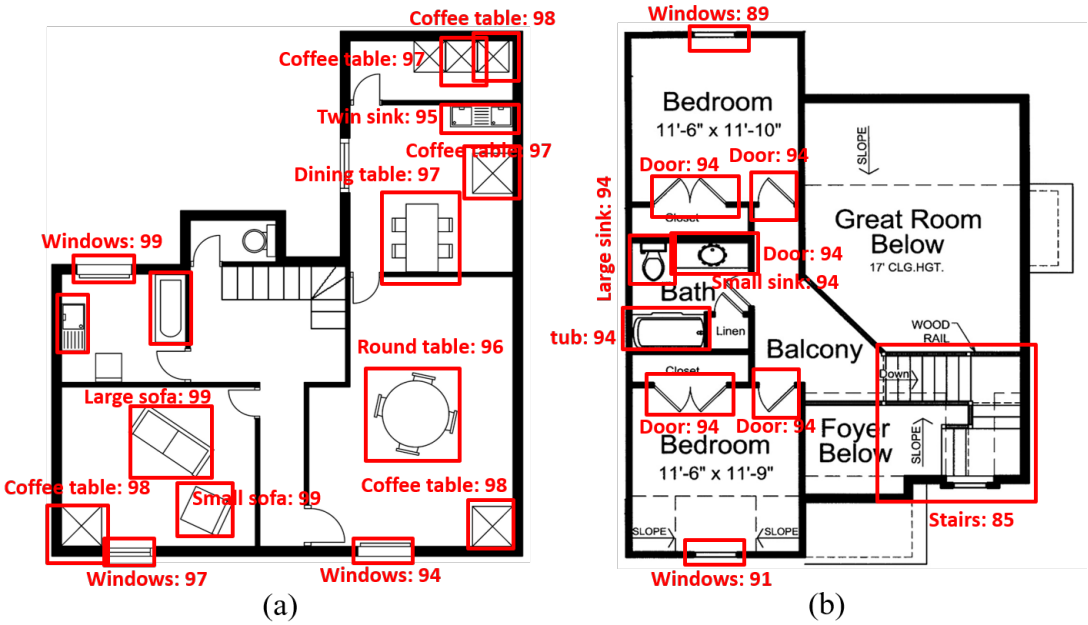


Figure 3.13 : Qualitative results of symbol spotting using the faster-RCNN model.

images along with their region descriptions as shown in Fig. 3.7(b), (c). Figure 3.14(b), (c) shows the results of dense captioning task on the floor plan images from BRIDGE. The generated caption for a specific region is depicted in the same color as the bounding box for the region.

In the publicly available models for natural images, maximum number of tokens/captions is 15. For floor plans, we released the limit since these images require more words for a caption to make sense. The word which appears less than 15 times is given a special token ⟨UNK⟩. Figure 3.14(d) depicts a failure case, where the caption is incorrect, because the detected region is not widespread in the training dataset. Most of the failure cases are due to less variability in the text data for the floor plan images. The limitation of annotating a floor plan image with region-wise description is less variability in the design of floor plans compared to natural images. Hence, it affects the efficiency of the model.

3.5.3 Description synthesis

It is insufficient to describe an image by a single caption. Hence we need a system which could generate paragraph based descriptions and has variability. There are many state-of-the-art techniques which have generated paragraphs in the context of natural images. However, these models need annotated images to be used to train the deep neural networks and further test and evaluate the models. There is no such publicly available floor plan dataset to support this task. We have generated paragraphs by using two techniques and evaluated them with the annotations in BRIDGE.

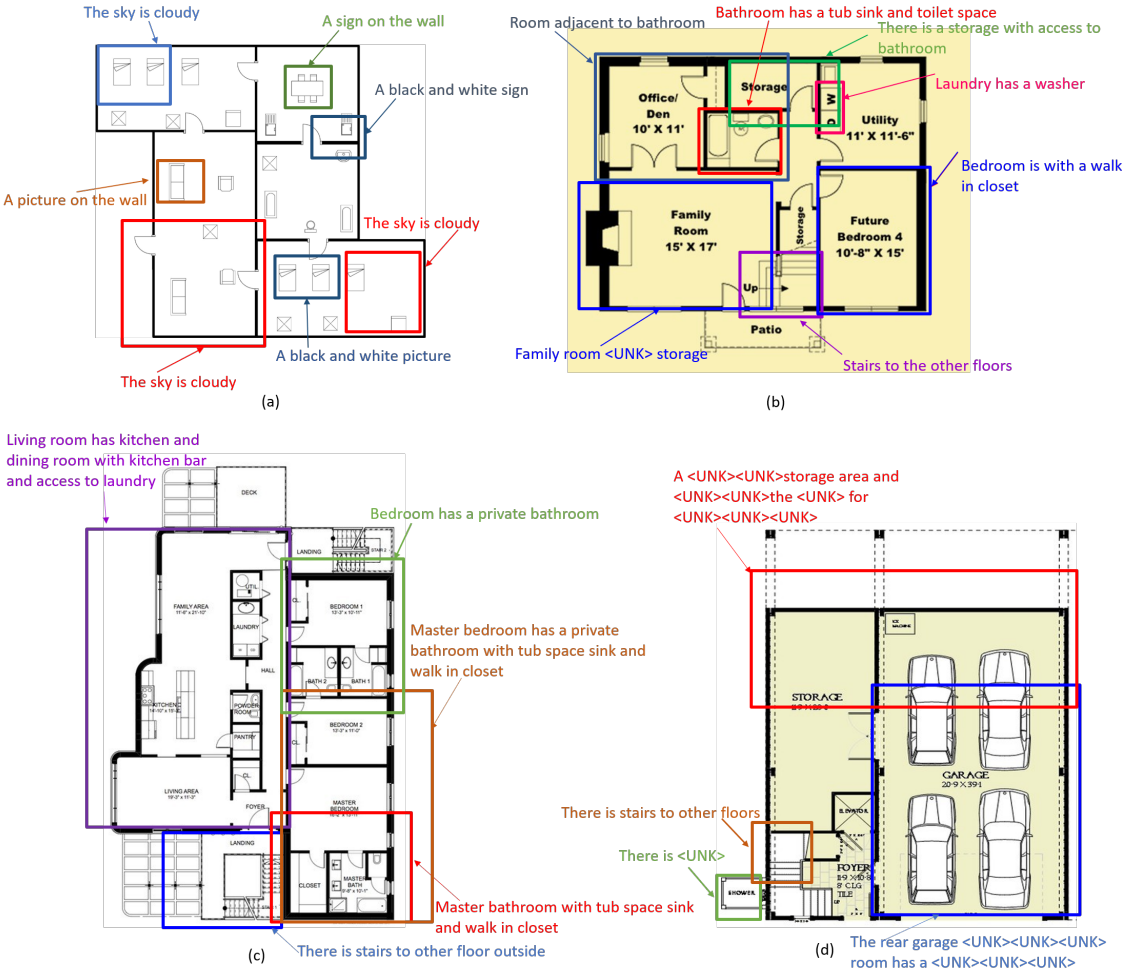


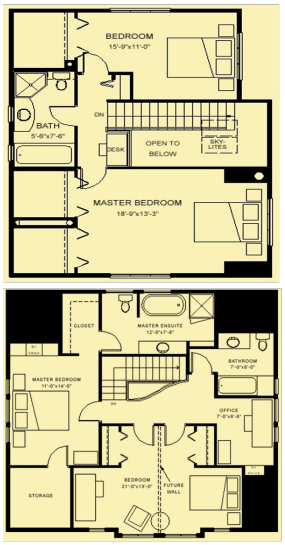
Figure 3.14 : Qualitative results of dense captioning with the proposed BRIDGE dataset.

Template based

Paragraph based descriptions are generated by using the technique proposed in Goyal *et al.* [2018a]. In this method rooms are detected by semantic segmentation of the floor plan and symbols are detected by symbol detection techniques using handcrafted features. Furthermore, a feature "Bag of Decor" is formed and the classifier is trained using that feature for each image. Paragraph based description is generated by using a manually specified template where words for room and decor symbols are proposed by above-mentioned methods. Figure 3.15 depicts comparison between the descriptions generated using template based, and Denscap model (described next), along with the ground truth (collected from web).

Densecap-concat

In this method, the paragraphs were synthesized by concatenating top 5 captions generated by Densecap. This method yields the bounding boxes of the regions along with the respective captions. Difference between Densecap-concat and template-based methods is that the former predicts the whole sentence without having a predefined structure within. Table 3.2 shows the



Densecap Concatenation

Bedroom is with bed. Bedroom is with bathroom which has tub, sink and toilet space. Master bedroom is near stairs and has a bathroom with tub shower and toilet space. Bathroom has a separate shower and sink space. There are stairs to the other floors.

The bathroom has space for tub sink toilet and separate shower. Bathroom has tub and sink. There are stairs to upper floors. Master bedroom has a bed and closet. Bedroom has a bed.

Template based

In this architectural floor plan there are 3 rooms. There is one bedroom. Bedroom has a bed in the east side of the room. There is one bathroom. Bathroom has 1 tub in south side of the room, 1 large sink in the north side of the room, 1 small sink in the west side of the room. There is one bedroom. Bedroom has a bed in the east side of the room

In this architectural floor plan there are 4 rooms. There is a bathroom. Bathroom has 1 tub in the north side of the room, 1 large sink in the north west of the room, 1 small sink in the south side of the room. There is a bathroom. Bathroom has 1 tub in the south side of the room, 1 large sink in the north side of the room, 1 small sink in the north west side of the room. There is one bedroom. Bedroom has 1 bed in the west side of the room. There is one bedroom. Bedroom has 1 bed in the east side of the room.

Collected Description

On the second floor, the balcony is open to the entrance foyer below, and has a nook for a desk. There is a master bedroom, a third bedroom, and a full bath to share. The bedrooms have a 9' flat ceiling that slope with the roof at the end walls. Closet spaces are tucked under the sloping roofs.

The upper level has a flexible floor plan which works well for growing families. The master bedroom offers a walk–in closet, spacious master bath, and a bonus storage room. The second bedroom is extra large, and can be easily divided by adding a center wall to create two smaller bedrooms. There is also an extra nook which can be used as an office or play area. The curved handrail in the hallway creates a small opening to view those on the ground floor.

Figure 3.15 : Paragraphs generated using Densecap-concat, Template based method and evaluation.

Table 3.2 : Evaluation of generated paragraphs with various metrices.

Method	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE_L
Densecap-concat	0.117	0.054	0.0195	0.003	0.189	0.122
Template	0.199	0.044	0.025	0.002	0.133	0.126

evaluation of the generated paragraphs with the paragraph annotations available in BRIDGE using metrics, METEOR proposed by Banerjee and Lavie [2005], ROUGE, proposed by Lin [2004], and BLEU-{1,2,3,4}, proposed by Papineni *et al.* [2002]. Results show that the used schemes for paragraph generation are not performing well on BLEU-3 and BLEU-4 because generated descriptions do not contain flexibility in terms of positioning of words. Hence, they contain similarity in terms of words chosen to describe (BLEU-1, BLEU-2) but not in the structuring of sentences. Hence a robust paragraph generation method for floor plan images is required.

3.6 SUMMARY

In this chapter, we discuss various potential research possibilities in floor plan images and publicly available datasets for corresponding tasks. The current available datasets of floor plans are targeted for a specific purpose, but they lack diversity and are very small in number. Also they are not suitable for multimodal understanding of floor plans and lack text modality in them. With the advent of advanced artificial intelligent models, there was a requirement of large-scale dataset which could be used in training these models efficiently. Hence, we present, for the first time, a novel large-scale (13000+ images) floor plan dataset BRIDGE, which has images and other metadata.The dataset includes annotations for decor symbol detection and classification, region-wise captions and paragraph based annotations. The dataset also includes images from ROBIN and SESYD datasets and annotations for them. This dataset can be used for various tasks including symbol spotting, room and wall segmentation, textual description generation, etc.

In this chapter, we also present experiments for symbol spotting using state of the art models such as, YOLO and Faster-RCNN. Furthermore, we present experiments for textual description generation using region wise caption generation model. The paragraph annotations can be used for evaluation purpose. The experiments depict the wide utility of the dataset BRIDGE and it opens up new research avenues for the graphic analysis community. The subsequent chapters describe the work done on indoor space understanding on publicly available datasets.