BRIDGE: Building plan Repository for Image Description Generation, and Evaluation

Shreya Goyal[1], Vishesh Mistry[1], Chiranjoy Chattopadhyay[1], Gaurav Bhatnagar[2]

[1] Department of Computer Science and Engineering

[2] Department of Mathematics

Indian Institute of Technology Jodhpur, India

International Conference on Document Analysis and Recognition, Sydney (2019)

ICDAR 2019

Outline

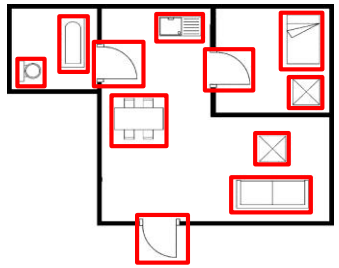
- Introduction
- Various approaches in floorplan research
- Existing datasets
- Requirement
- Construction of dataset
- Experiments
  - Décor symbol detection
  - Region wise captioning
  - Description generation
- Analysis

Introduction

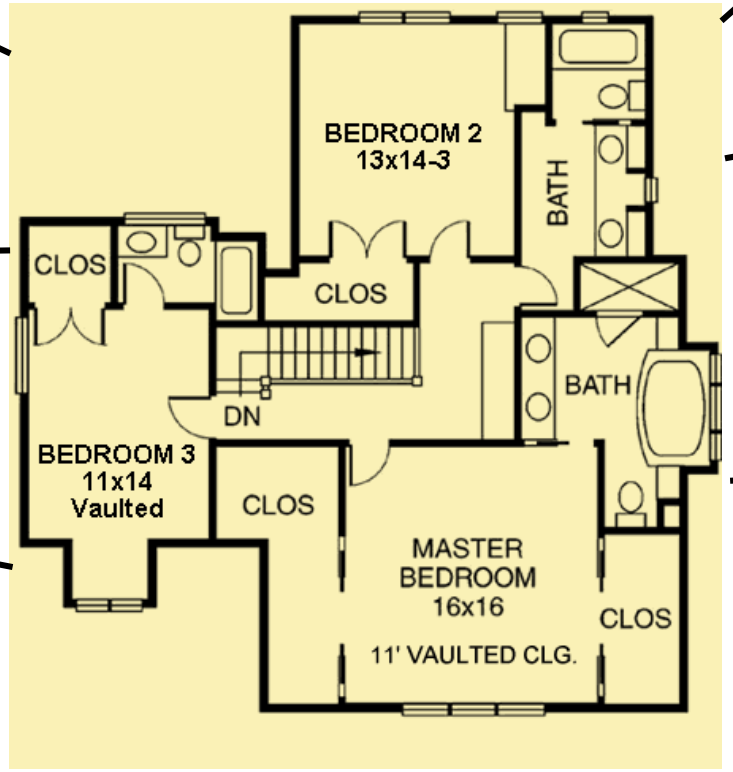
- Floor plan is an architectural drawing of a building.
- Tasks involved:
  - Semantic understanding
  - Décor symbol spotting and classification
  - Textual description generation
  - Image segmentation and retrieval
- Approaches:
  - Non learning based methods
  - Machine/ Deep learning based methods

Various approaches in floorplan research

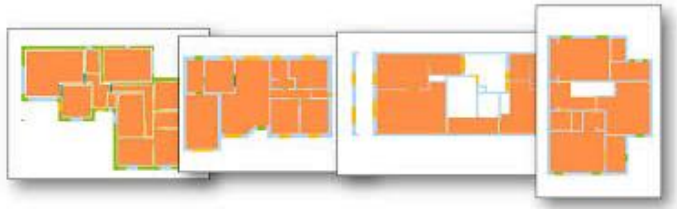
Symbol Spotting



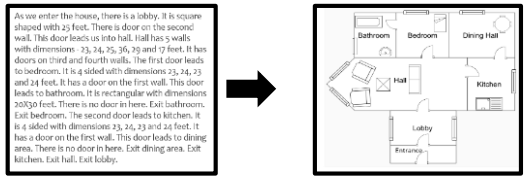
Floor plan retrieval



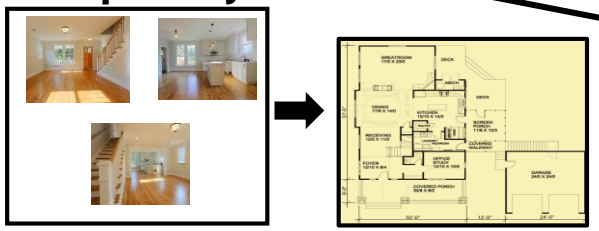
Segmentation



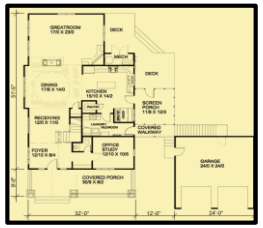
Floorplan Synthesis



Floorplan Synthesis



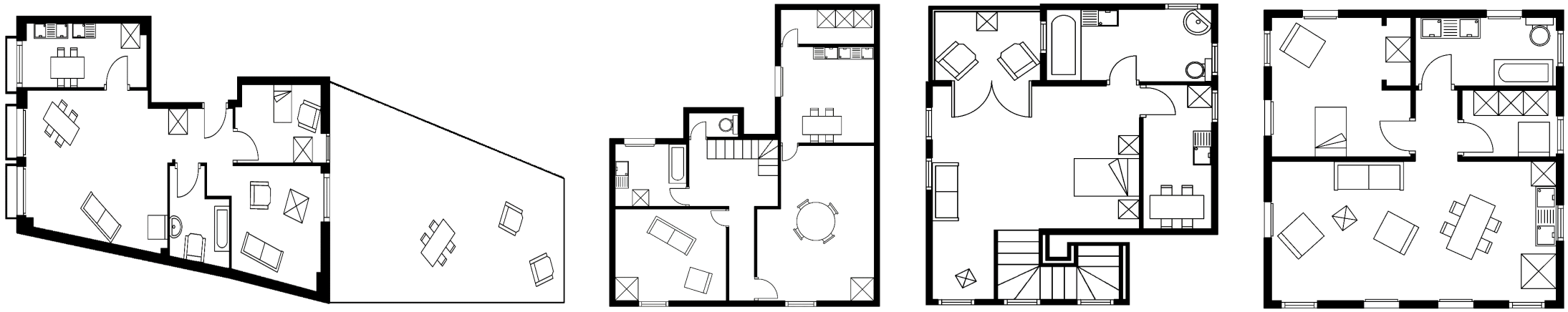
Description Generation



This floor plan has a kitchen, dining, living room, garage. Kitchen has cooking range and eating facility. Dining room is large in size and has sitting facility. Garage is connected with stairs...

Various Datasets in Floor plans

SESYD dataset

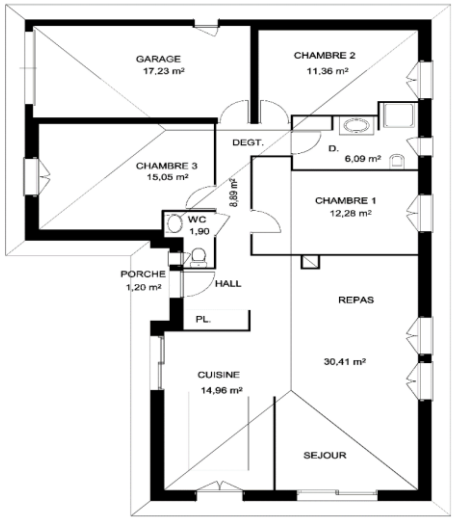


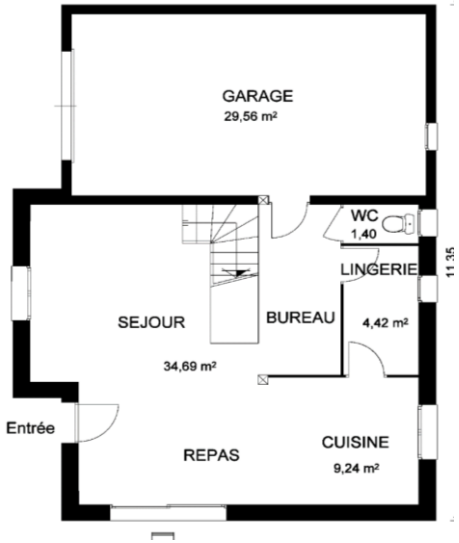
- 1000 floorplan sample in 10 categories
- Synthetically generated
- Designed for retrieval and symbol spotting tasks
- Categories differ in global layout of the floor plan

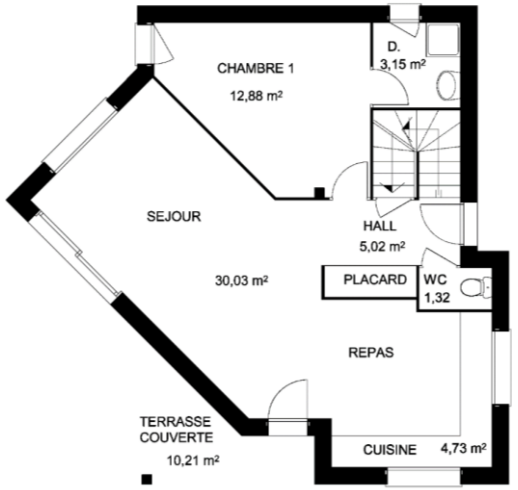
M. Delalandre, Generation of synthetic documents for performance evaluation of symbol recognition & spotting systems," IJDAR, 2010

Various Datasets in Floor plans

CVC-FP dataset





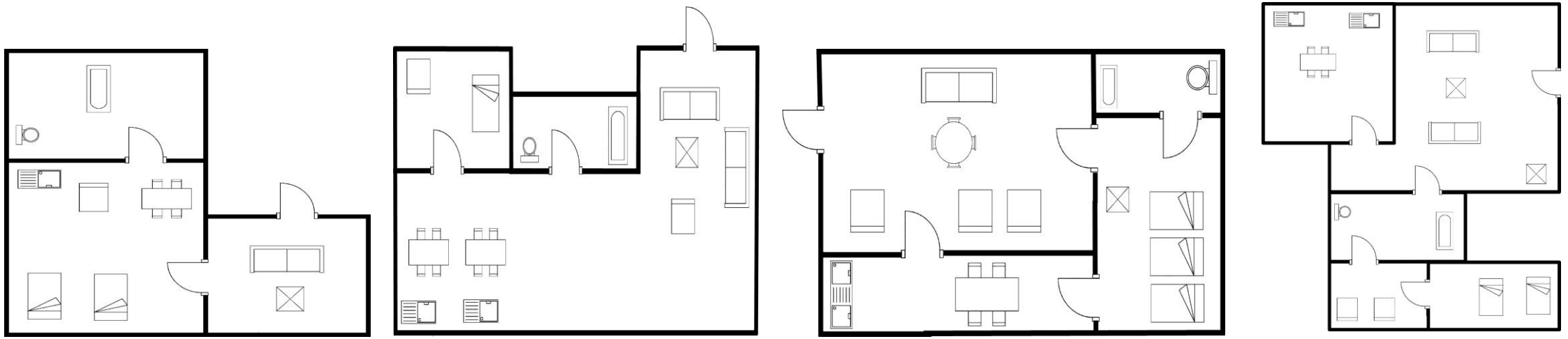


- 122 floorplan sample in 4 categories
- Scanned Documents
- Designed for segmentation of rooms and other components
- Categories differ in origin and style
- Contains ground truth for segmentation

L.P. de las Herras et al. , CVC-FP and SGT, a new database for structural floor plan analysis and its groundtruthing tool," IJDAR, 2015

Various Datasets in Floor plans

ROBIN dataset

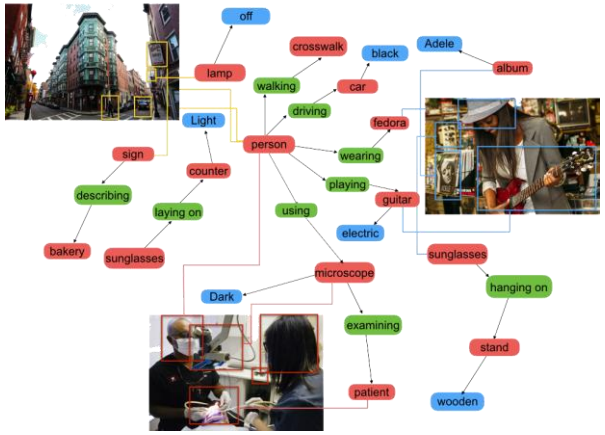


- 510 floorplans in 3 categories
- Hand crafted floorplans
- Designed for retrieval and symbol spotting purpose.
- Categories differ in number of rooms and global layout.

D.Sharma et al., Daniel: A deep learning architecture for automatic analysis and retrieval of building floorplans, ICDAR 2017.

Various datasets in natural images

- **Visual Genome**
  - over 1 million images.
  - object annotations, region descriptions, scene graph, region graphs.



- **MS-COCO**
  - 328k images
  - Object instance category labelling, spotting and segmentation.

- **MS-COCO captions**
  - Images from MS-COCO.
  - 5 captions per image.


The man at bat readies to swing at the pitch while the umpire looks on.


A large bus sitting next to a very tall building.

Requirement

- Bridging the two modalities- Image and Text.
- Document images lacks large scale datasets.
- To enable training the data hungry algorithms.
- Increase accuracy of the existing algorithms with more data.
- Existing floorplan datasets lack textual annotations.
- Also they lack large scale décor symbol annotations.
- BRIDGE dataset also caters description generation models.
- For evaluation of the existing description generation models.

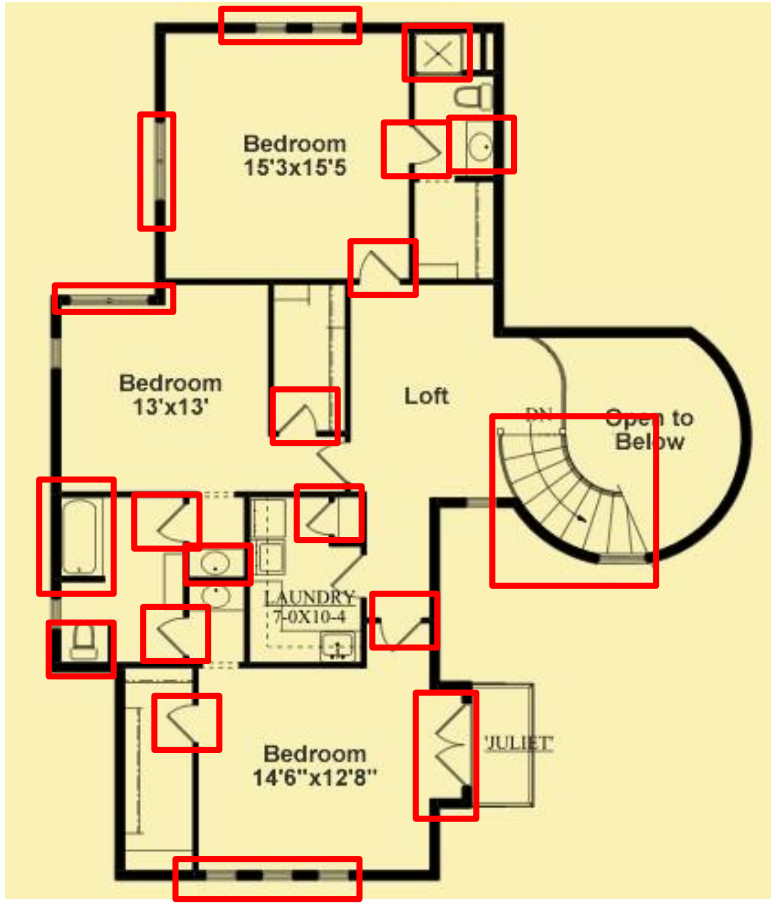
Construction of BRIDGE

- BRIDGE dataset contains ~13000 samples of floorplan images.
  - Images were collected from two websites
  - website: www.architecturalhouseplans.com , www.houseplans.com
- Décor symbol annotations
  - Human annotators
  - Stored in XML format
- Region wise captions
  - Human annotators
  - Stored in JSON format
- Paragraph descriptions.
  - Collected from websites.

Challenges in labelling dataset

- Number of image samples is large.
- Less variability in images.
- Requires precision unlike natural images.
- Difficult to understand by human annotators.
- Difficult to label in cluttered images.

Décor Symbol annotation

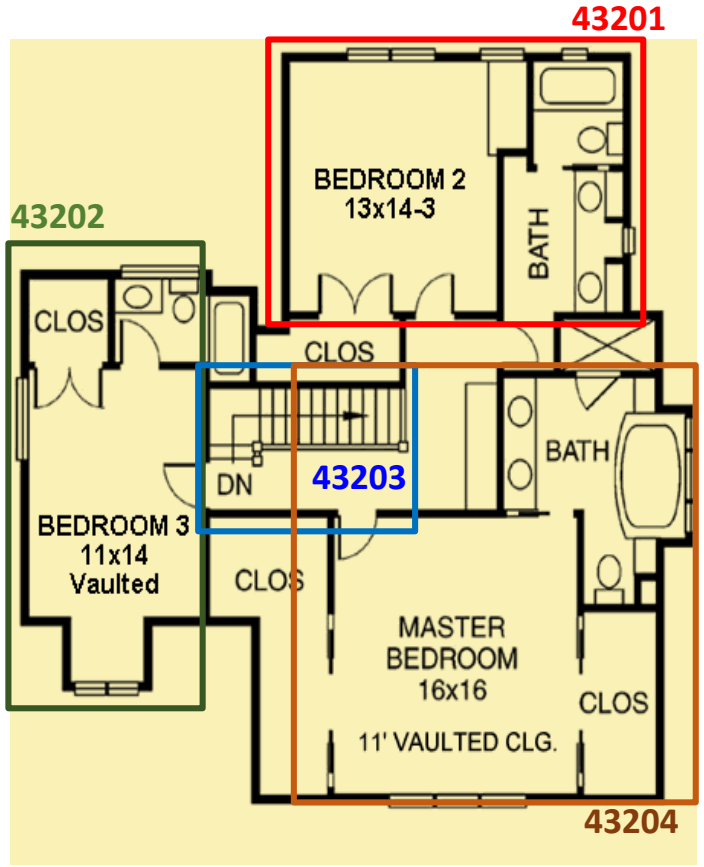


Décor Symbol annotations

```
<?xml version="1.0"?>
- <annotation>
    <folder>D1_new</folder>
    <filename>432.jpg</filename>
    <path>G:/data_new/D1_new/432.jpg</path>
  - <source>
      <database>Unknown</database>
    </source>
  - <size>
      <width>448</width>
      <height>668</height>
      <depth>3</depth>
    </size>
    <segmented>0</segmented>
  - <object>
      <name>door</name>
      <pose>Unspecified</pose>
      <truncated>0</truncated>
      <difficult>0</difficult>
    - <bndbox>
        <xmin>134</xmin>
        <ymin>130</ymin>
        <xmax>187</xmax>
        <ymax>161</ymax>
      </bndbox>
    </object>
  - <object>
```

Corresponding XML file

Region wise captioning

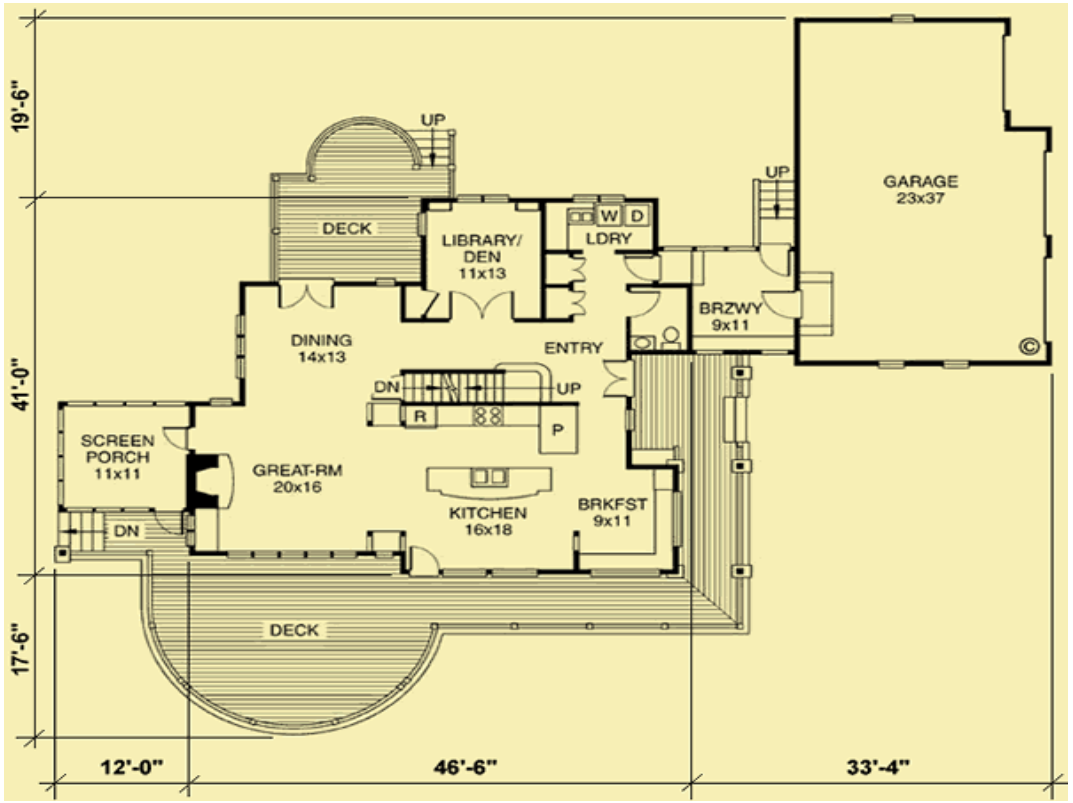


Region annotations

[
{“image_id”: 432,
“regions”:[{“region_id”: **43201**, “width”: 220,
“height”: 160, “image_id”:432, “x”: 5, “y”: 227,
“phrase”: “the bedroom is with a private bathroom with dual sink, tub and toilet space”},
{“region_id”: **43202**, “width”: 140, “height”: 270,
“image_id”: 432, “x”: 150, “y”: 5, “phrase”: “another bedroom with private bathroom having tub, sink and walk in closet”},
{“region_id”: **43203**, “width”: 250 “height”: 90,
“image_id”:432, “x”: 255, “y”: 170, “phrase”: “There are stairs to the other floors”},
{“region_id”: **43204**, “width”: 190 “height”:310,
“image_id”:432, “x”: 256, “y”: 240, “phrase”:
“master bedroom has a master bathroom with dual sink, tub, toilet and closet”}]
}
]

Region wise captions

Paragraph description



Floorplan image

The great room is anchored by a finely crafted stone fireplace, and it is open to both the kitchen and the dining room. It also accesses a screened porch that has unlimited views in three directions. The large wrap–around deck can be accessed from the screened porch, the kitchen, and the entryway. There's a sunlit breakfast nook next to the kitchen for casual dining, and the more formal dining area accesses a large deck for outdoor dining on warm evenings.

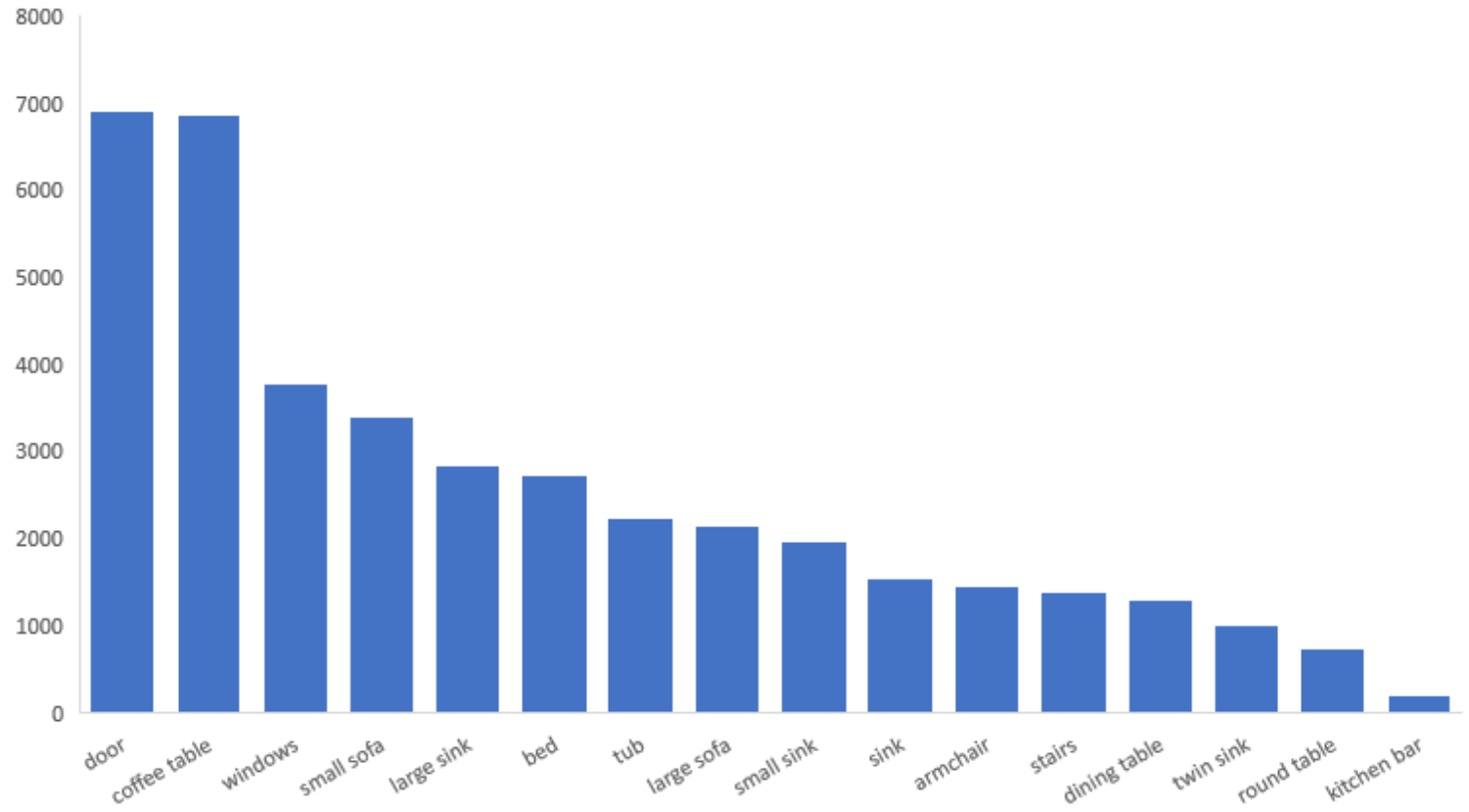
Paragraph annotation

Fine-tuning pre-trained networks

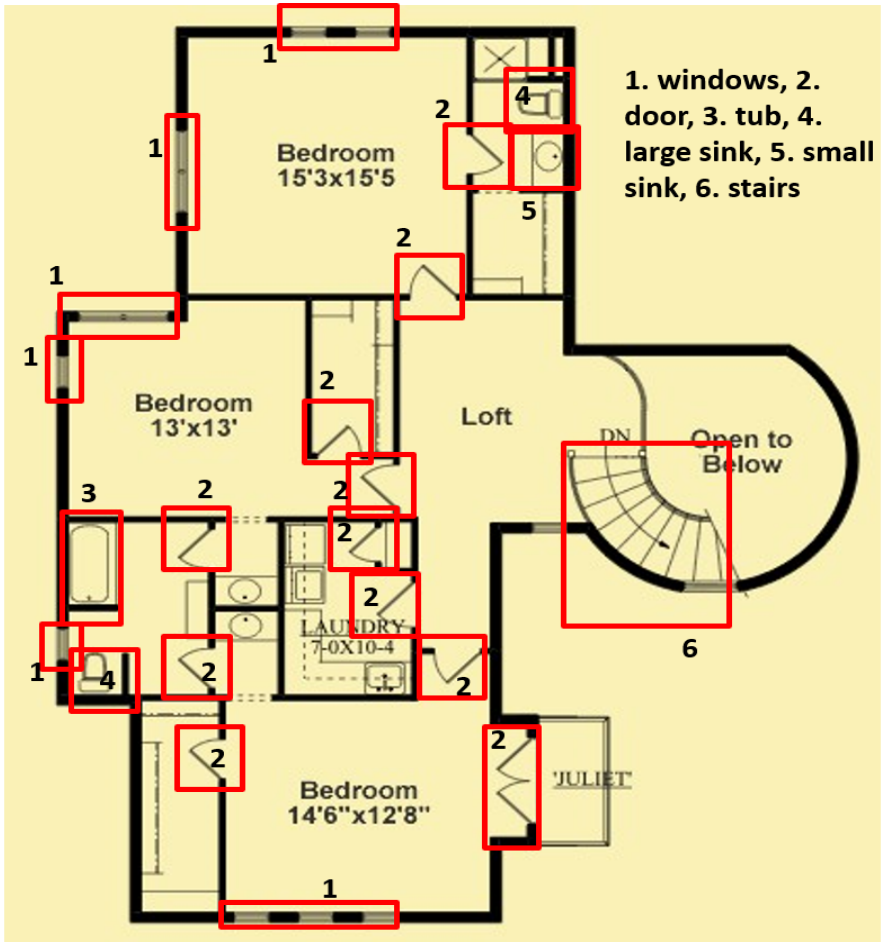
- Original network tiny-YOLO has 9 convolutional layer.
- Trained on 1000 class Image-net dataset.
- Fine-tuned original network with 16 classes of objects (BRIDGE)
- Final layer has 105 filters and linear activation function.
- Faster RCNN has two modules:
  - Region proposal network
  - Fast-RCNN detector
- Confidence score calculated using Intersection over union (IoU).

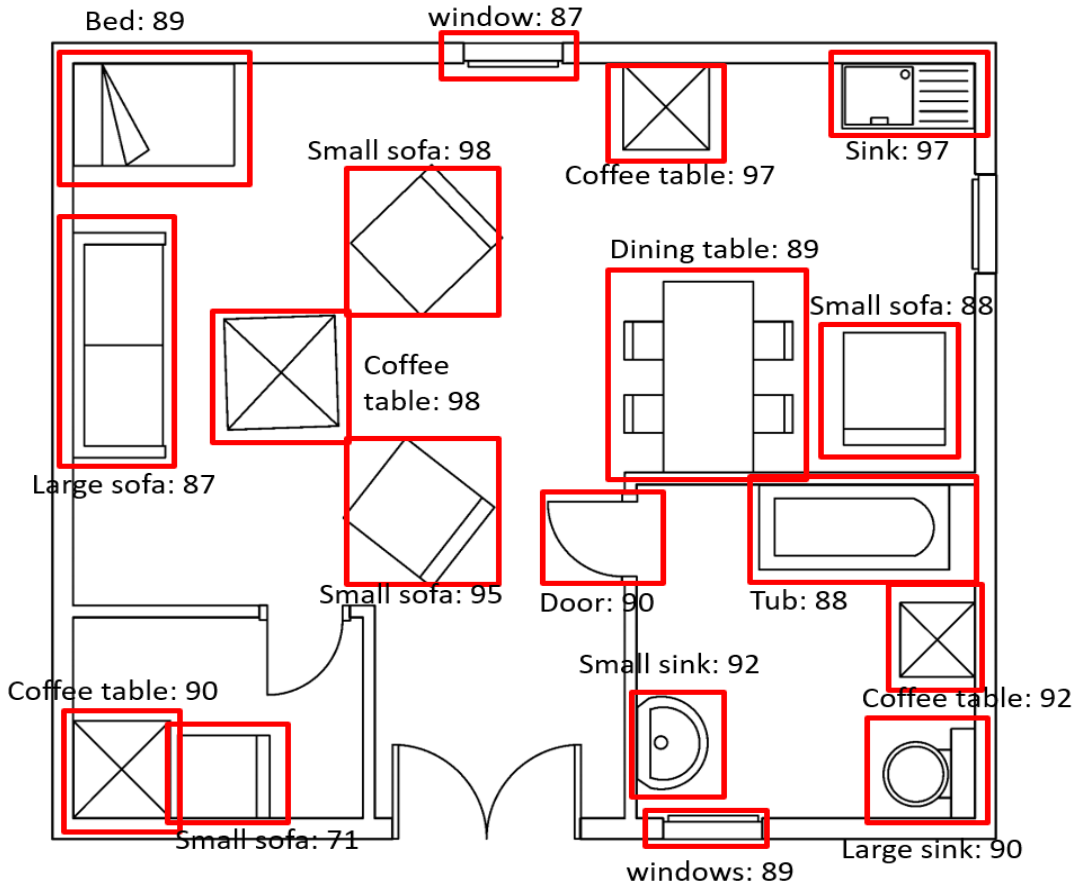
$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Union}$$

Distribution of décor symbols

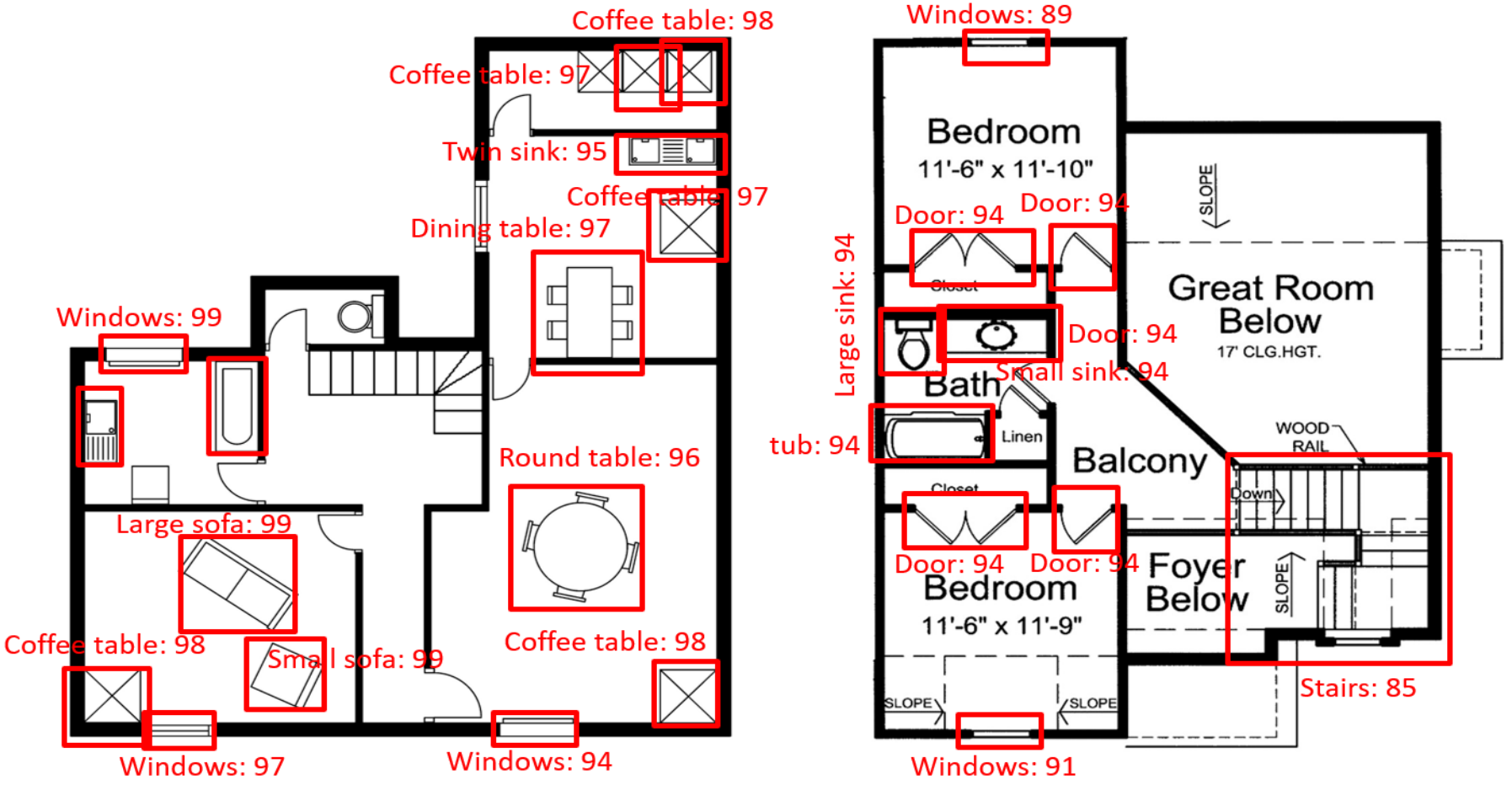


Results-Décor symbol spotting (YOLO)





Results-Décor symbol spotting (Faster-RCNN)



Dense-captioning

- Caption captures information over entire image.
- Insufficient to capture entire information.
- Dense-captioning task generates region wise captions.
- Existing network trained on natural images.
- VGG 16 architecture with 13 convolutional and 5 max pool layers.
- Region proposal network followed by RNN language model.
- Fine-tuned with region wise captions in BRIDGE dataset.

Region wise captioning-Densecap

Living room has kitchen and dining room with kitchen bar and access to laundry

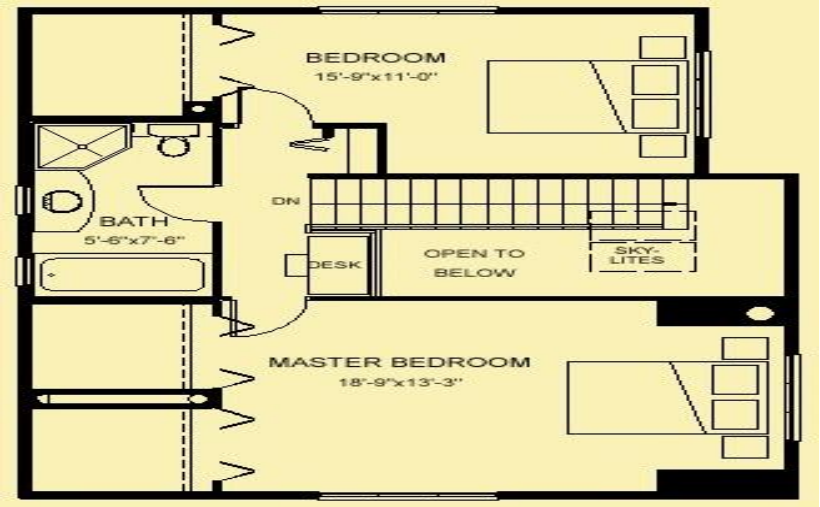
Bedroom has a private bathroom

Master bedroom has a private bathroom with tub space sink and walk in closet

Master bathroom with tub space sink and walk in closet

There are stairs to other floor outside

Paragraph generation



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.

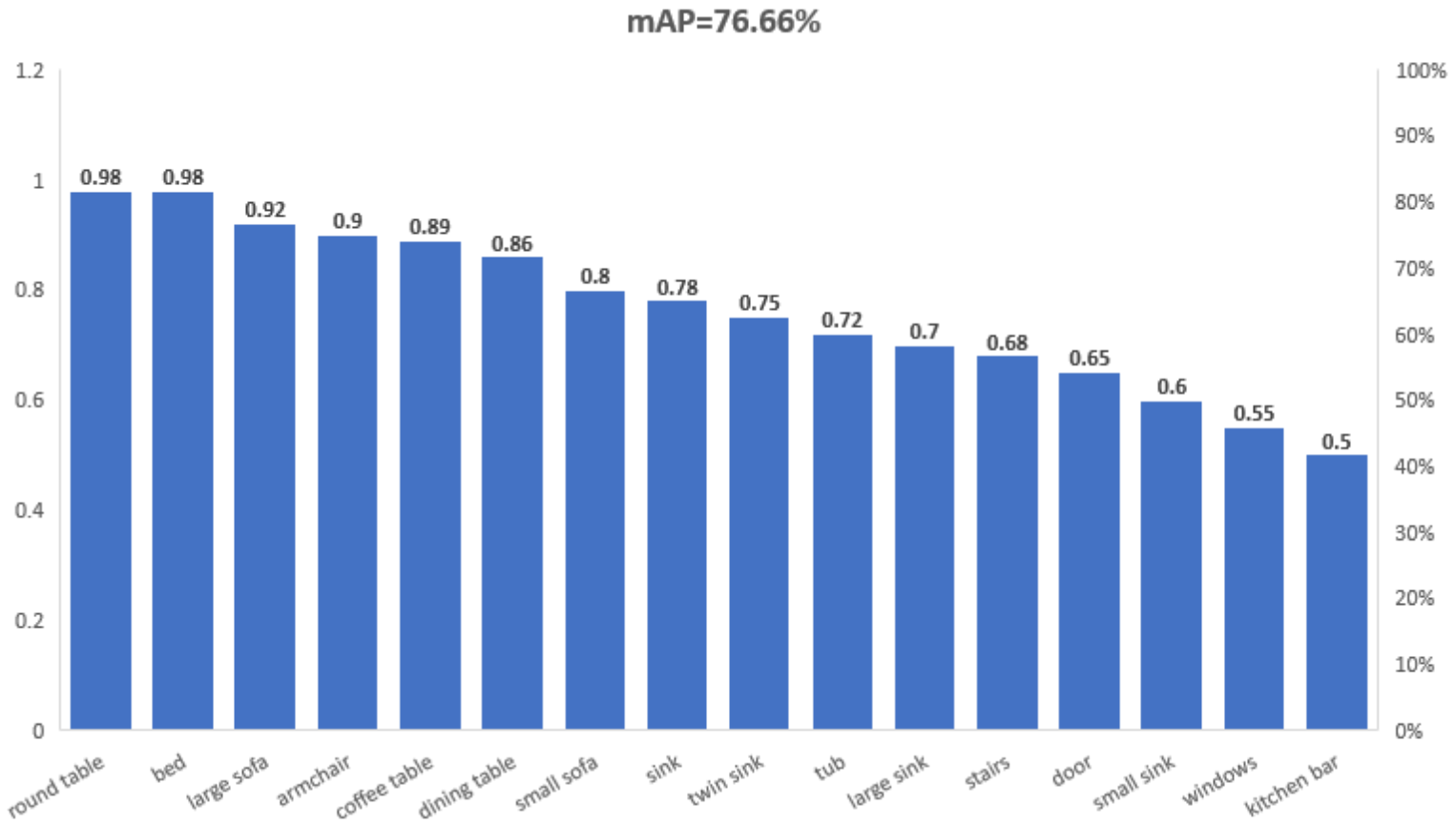
Template based

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

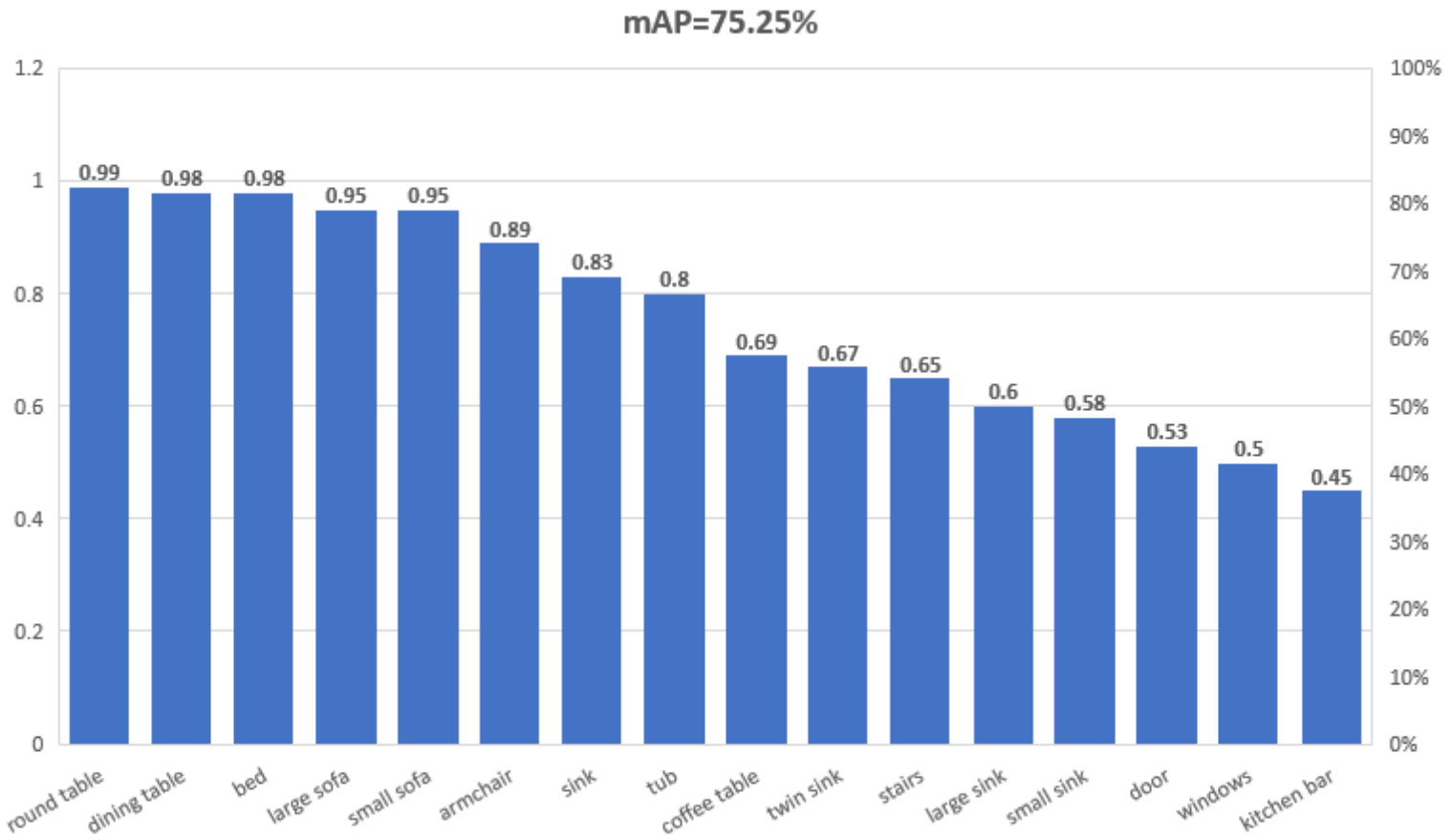
Collected Description

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.

Experimental Analysis (symbol spotting)



Experimental Analysis (symbol spotting)



Experimental Analysis (Paragraph Generation)

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-Based	0.199	0.044	0.025	0.002	0.133	0.126

BLEU- Bilingual Evaluation Understudy

METEOR- Metric for Evaluation of Translation with Explicit ORdering

ROUGE- Recall Oriented Understudy for Gisting Evaluation

Conclusion

- A new large scale dataset BRIDGE, of floor plan images is proposed.
- Contains ~13000 images, with annotations.
- Décor symbol, region wise captions and paragraphs.
- First dataset of its kind in floorplan images.
- Required for bridging the gap of document images and text.
- Useful for data hungry algorithms.
- Good start for text generation models for document images in future.

References

[1] D. Sharma, N. Gupta, C. Chattopadhyay, and S. Mehta, "Daniel: A deep architecture for automatic analysis and retrieval of building floor plans," in ICDAR 2017.

[2] L-P. de las Heras, O. R. Terrades, S. Robles, and G. Sanchez, "Cvc-fp and sgt: a new database for structural floor plan analysis and its groundtruthing tool," IJDAR, 2015.

[3] M. Delalandre, E. Valveny, T. Pridmore, and D. Karatzas, "Generation of synthetic documents for performance evaluation of symbol recognition & spotting systems," IJDAR, 2010.

[4] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, "Microsoft coco: Common objects in context," in ECCV, 2014.

[5] X. Chen, H. Fang, T.-Y. Lin, R. Vedantam, S. Gupta, P. Dollar, and C. L. Zitnick, "Microsoft coco captions: Data collection and evaluation server," arXiv preprint arXiv:1504.00325, 2015.

[6] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D. A. Shamma et al., "Visual genome: Connecting language and vision using crowdsourced dense image annotations, IJCV 2017.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances in neural information processing systems , 2015

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in CVPR , 2016, pp. 779–788.

[9] J. Johnson, A. Karpathy, and L. Fei-Fei, "Densecap: Fully convolutional localization networks for dense captioning," in CVPR , 2016.

Thank You!