HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 2 (2024)

Architectural Space Recognition from Blueprints Using Machine Learning-Based Semantic Segmentation

Shiho Yu[1] and Su Jin[#]

[1]Korea International School Jeju Campus, Republic of Korea
[#]Advisor

ABSTRACT

Architectural blueprints have been providing fundamental processes for urban planning, construction, and interior design for professionals. It contains a visual representation of building plans and specifications in the construction and design industry. However, for the non-expert, it is difficult to understand these complex technical drawings. This is because of the specialized language, symbols, and technical knowledge required to understand architectural blueprints. This challenge often leads to misunderstandings which can extend the construction process and escalate costs. To bridge this knowledge gap, in this research paper, I propose a machine-based blueprint interpretation method using semantic segmentation. The proposed method takes blueprints as input and generates semantic segmentation maps. These maps categorize and isolate distinct architectural areas into predefined classes, including rooms, kitchens, and bathrooms. The proposed machine learning model is tested on a publicly available dataset of architectural blueprints. Through comprehensive quantitative and qualitative assessments, it is shown that the proposed method achieves state-of-the-art performance.

Introduction

The architecture and construction industry has relied heavily on analyzing blueprints for designing buildings. However, the inherent complexity of these technical drawings often creates a significant barrier for individuals lacking specialized knowledge in the field. The use of specialized language, symbols, and intricate details makes it challenging for non-experts to comprehend and interpret architectural blueprints accurately.

This knowledge gap can lead to misunderstandings which delays construction projects, and increased costs. Bridging this gap is crucial for fostering effective communication among stakeholders and ensuring the smooth progression of construction processes. The motivation behind this research stems from the need to empower non-experts with the ability to interpret and understand architectural blueprints more intuitively. Addressing this challenge holds the potential to streamline communication, enhance collaboration, and reduce errors in construction projects. By leveraging advancements in machine learning, particularly semantic segmentation, I aim to develop a method that facilitates the automatic recognition and categorization of distinct architectural spaces within blueprints.

To achieve this, I propose a machine-based blueprint interpretation method using semantic segmentation.

The proposed method takes blueprints as input and generates semantic segmentation maps. These maps categorize and isolate distinct architectural areas into predefined classes, including rooms, kitchens, and bathrooms. The proposed machine learning model is tested on a publicly available dataset of architectural blueprints. Through comprehensive quantitative and qualitative assessments, it is shown that the proposed method achieves state-of-the-art performance.

Related Work

Image Semantic Segmentation

Image semantic segmentation is a computer vision task that takes the digital image as input and outputs the segmentation map that divides an image into meaningful and semantically homogeneous segments or regions. Unlike image classification, where the goal is to assign a single label to an entire image, semantic segmentation aims to label each pixel in the image with a specific class or category.

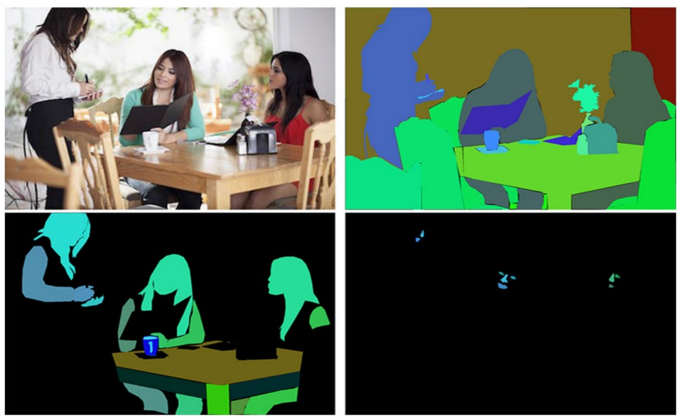

Figure 1. Example of image semantic segmentation (Liu et al. 2019)

Figure 1 illustrates an example of the image semantic segmentation process. The objects within the input image are segmented into distinct groups, represented by colored areas in the segmentation maps. Applications of image semantic segmentation are diverse and span various domains. In the context of architectural blueprints or floor plans, as mentioned in Chapter 1, semantic segmentation can be employed to automatically identify and categorize different architectural elements such as rooms, doors, windows, and other spatial features.

Architectural Blueprint Image

A blueprint image is a specialized visual representation of architectural or engineering plans, providing a detailed and technical overview of a structure's design. Blueprint images contain intricate technical details of a building or structure, including dimensions, floor plans, elevations, and other specifications. These details are essential for guiding construction processes and ensuring accuracy in implementation.

Figure 2. Example of blueprint image

Blueprint images are integral to the communication and realization of architectural and engineering projects. While they are invaluable tools for professionals in the field, they can pose challenges for individuals without specialized training due to the technical complexity of their content.

To address this problem, this research aims to develop a novel machine learning-based blueprint interpretation system using semantic segmentation. The detailed explanation of the proposed approach will be provided in Chapter 3.

Proposed Method

Architecture Overview

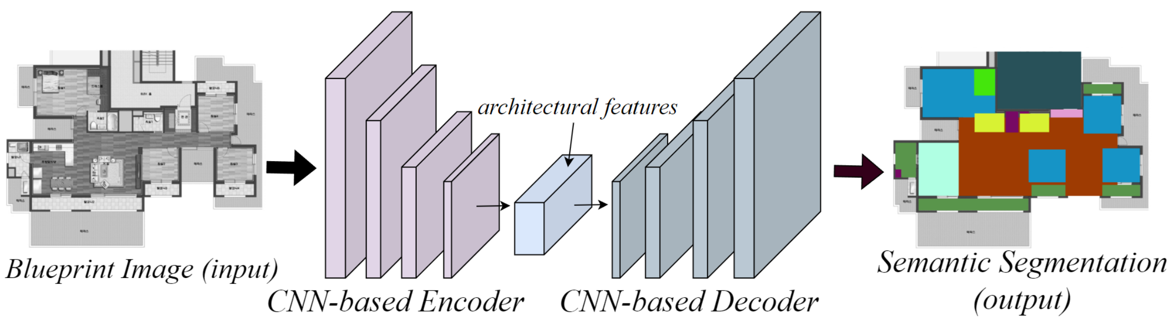

Figure 3. Diagram of the proposed blueprint segmentation system

Figure 3 illustrates the overall architecture of the proposed blueprint semantic segmentation method. The proposed method is developed upon the architecture of a convolutional neural network-based autoencoder, consisting of an encoder and a decoder. The encoder processes the blueprint image as input, generating latent features or architectural feature maps that encapsulate the visual patterns inherent in the provided blueprint image. The decoder utilizes these features as input and generates a semantic segmentation map which isolates each room's location and type through visually isolated bounding boxes, as shown in Figure 3.

The encoder and decoder components can be designed using various types of convolutional neural network architectures. I have experimented with five established architectures to determine which one yields the most accurate results. Further details on this investigation are discussed in Chapter 4.

To train the proposed system, I utilized the dice coefficient loss function as explained in Equation 1.

Equation 1: Dice coefficient loss function

$$L = 1 - D_{score}$$

Where, D_{score} indicates the dice score which is a metric commonly used to evaluate the performance of image segmentation algorithms. It measures the similarity between the predicted segmentation and the ground truth and provides a quantitative assessment of how well the trained model delineates object boundaries in an image. The dice score is explained in Equation 2.

Equation 2: Dice score

$$D_{score} = \frac{2 \times |GT \cap Pred|}{|GT| + |Pred|}$$

In Equation 2, *Pred* and *GT* denote the model's predicted segmentation map and its corresponding ground truth. Dice coefficient loss function quantifies the dissimilarity between predicted and ground truth segmentations. The value of dice coefficient loss function ranges from 0 to 1, with 1 indicating perfect overlap between the predicted and ground truth segmentations. The loss function is minimized during the training of the proposed blueprint segmentation network.

Implementation

In this chapter, I provide an overview of the implementation details employed in the development of the proposed method. I set the batch size to 512 and trained the proposed network for 120 epochs with an initial learning rate of 0.0001. The learning rate decayed at epoch 40 and 80 to fine-tune the model's parameters as it progresses through different stages of training. I maintained the exact same training parameter settings for five image segmentation methods to accurately compare their performance.

Experimental Results

Dataset

The dataset used in this research comprises a diverse collection of architectural blueprints representing different types of residential structures. The dataset is composed of a total of 41,556 samples. The samples are categorized into three groups: apartments, multiplex housing, and regular houses. Blueprint samples for apartments and multiplex housing exhibit uniformity and similarity in appearance. In contrast, the samples for regular houses display significant variability as their designs typically depend on individual customer preferences.

Inference Metric

To assess the performance of the proposed method, I measure the intersection over union (IoU) that is commonly used to evaluate the accuracy of object detection and segmentation algorithms. The IoU is calculated by determining the ratio of the area of overlap between the predicted and ground truth regions to the total area encompassed by their union.

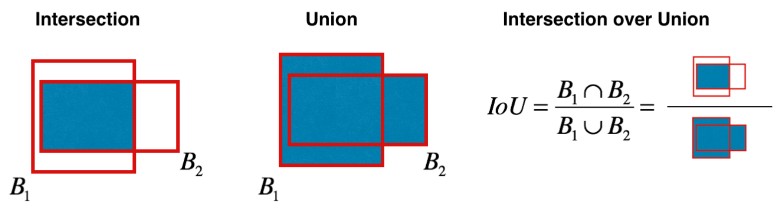

Figure 4. Explanation of intersection over union

Figure 4 illustrates how the intersection over union is calculated. This metric assesses the degree of spatial agreement between the predicted segmentation output generated by the proposed method and the actual ground truth delineated in the dataset. A higher IoU indicates better alignment and accuracy in segmenting the architectural areas of interest. Practically, an IoU value of 1 signifies perfect overlap which indicates that the predicted and ground truth regions perfectly match. Conversely, an IoU of 0 indicates no overlap between the predicted and ground truth regions.

Experimental Result

To develop the blueprint segmentation model, I employed five well-established image semantic segmentation methods such as PSPNet (Zhao et al. 2017), Multipath-RetineNet (Lin et al. 2017), Resnet-38-MS-COCO (Wu et al. 2019), IDW-CNN (Wang et al. 2017), and DeepLabV3 (Chen et al. 2017). To ensure a fair performance comparison, I trained each segmentation model using identical training hyperparameter settings. The IoU evaluation results for each method are summarized in Table 1 and Figure 5.

Table 1. Experimental result on blueprint dataset

Method	Multiplex Housing	Apartment	House	Total Samples
PSPNet (Zhao et al. 2017)	82.7	83.0	80.1	81.9
Multipath-RetineNet (Lin et al. 2017)	82.8	84.2	80.1	83.2
Resnet-38-MS-COCO (Wu et al. 2019)	83.7	85.2	81.5	84.0
IDW-CNN (Wang et al. 2017)	84.3	85.7	81.7	84.6
DeepLabV3 (Chen et al. 2017)	**86.4**	**87.8**	**83.7**	**86.7**

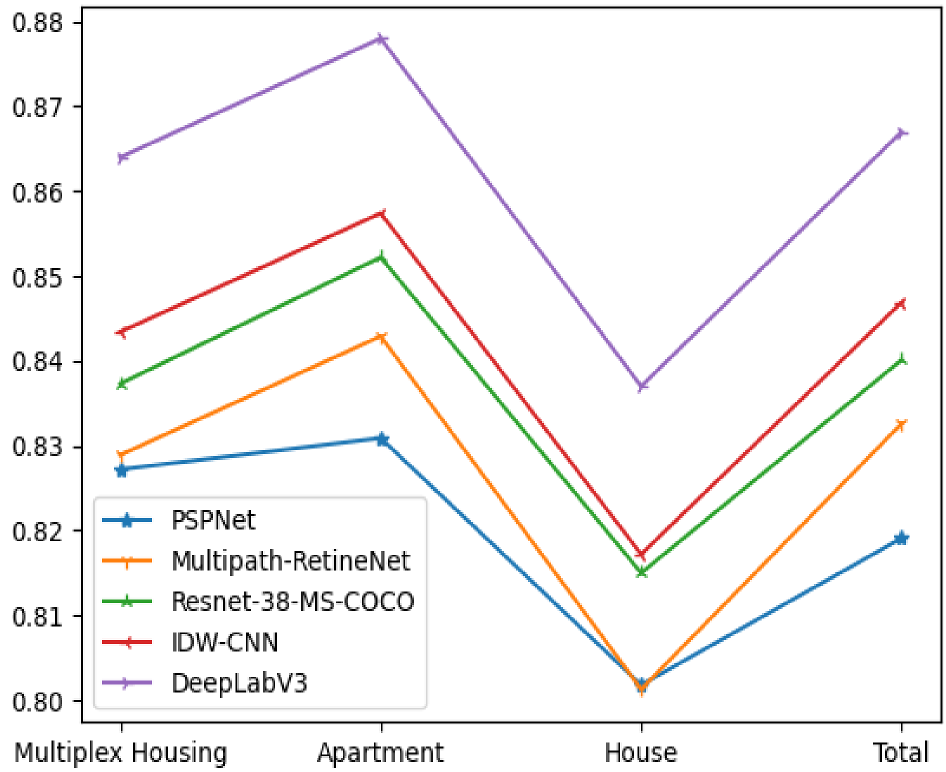

Figure 5. Performance (IoU) comparison graph

As shown in Table 1, PSPNet and Multipath-RetinaNet exhibit relatively inaccurate outcomes attributed to their shallow network depth. In contrast, Resnet-38-MS-COCO and IDW-CNN deliver slightly more accurate results, owing to their deeper network architectures. Notably, DeepLabV3 achieves the highest level of accuracy among the evaluated methods.

Notably, all methods encounter challenges when interpreting house samples, exhibiting the lowest accuracy among the three categories. This difficulty arises due to the diverse and varied visual patterns and characteristics present in regular house samples, as compared to the more standardized designs found in apartment and multiplex housing samples. It is expected that this limitation can be mitigated by augmenting the dataset with additional regular house samples. This will be explored in future research discussions.

(a) (b) (c)

Figure 6. Visual experimental result. (a): Multiplex housing, (b): Apartment, and (c): House

Finally, I present some qualitative experimental results. Figure 6 displays the visual outcomes of experiments conducted with DeepLabV3. Each colored bounding box represents a specific area of the room segmented from the original blueprint image. These visually interpreted images facilitate easy comprehension for both customers and non-experts which makes communication more accessible.

This finding holds potential for extending its application to automatic 3D modeling of houses. Given the traditionally time-consuming and labor-intensive nature of the 3D modeling process, I expect that this research could serve as a viable alternative which offers a more economically efficient approach.

Conclusion

In this research, I proposed a novel machine-based blueprint interpretation method using semantic segmentation. The proposed approach takes architectural blueprints as input and generates semantic segmentation maps. These maps categorize and isolate distinct architectural areas into predefined classes, including rooms, kitchens, and bathrooms. The evaluation of the proposed method on a publicly available dataset demonstrated its efficacy and state-of-the-art performance. Through comprehensive quantitative assessments, I have demonstrated promising results through the analysis of blueprint images using image semantic segmentation networks. In the future, I plan to broaden my analysis to encompass various types of constructions, including manufacturing factory layouts.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587. https://doi.org/10.48550/arXiv.1706.05587

Lin, G., Milan, A., Shen, C., & Reid, I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1925-1934). https://doi.org/10.48550/arXiv.1611.06612

Liu, X., Deng, Z., & Yang, Y. (2019). Recent progress in semantic image segmentation. Artificial Intelligence Review, 52, 1089-1106.

Wang, G., Luo, P., Lin, L., & Wang, X. (2017). Learning object interactions and descriptions for semantic image segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5859-5867). https://doi.org/10.1109/CVPR.2017.556

Wu, Z., Shen, C., & Van Den Hengel, A. (2019). Wider or deeper: Revisiting the resnet model for visual recognition. Pattern Recognition, 90, 119-133. https://doi.org/10.48550/arXiv.1611.10080

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2881-2890). https://doi.org/10.48550/arXiv.1612.01105