HIGH SCHOOL EDITION
Journal of Student Research

Volume 13 Issue 1 (2024)

A Deep Learning Approach to Semantic Segmentation for Architectural Blueprint Interpretation and Geographic-Based Material Recommendation

Ashley Chon[1] and Peter Gabor[#]

[1]TJHSST, USA
[#]Advisor

ABSTRACT

The planning and building of houses is time-consuming and takes a financial toll, causing housing development to struggle to keep up with rapidly growing populations. In the rapid expansion of living spaces, the environmental implications of such developments are frequently disregarded. Though few recent scientists have attempted to leverage machine learning to expedite the house modeling process, these models do not take into consideration environmental factors such as geographic location, temperature, and landscaping. There is a pressing need for developing an automated and accurate system that offers comprehensive interpretation of architectural blueprints, providing essential guidance and insights for various aspects of the construction process. This study incorporates these factors by leveraging machine learning to develop software models for cost-effective and environmentally sustainable housing. To address the aforementioned problem, I propose a machine learning-based semantic segmentation network for architectural blueprint interpretation. The proposed method is designed to address the complex task of understanding architectural blueprints, discerning room types and their spatial arrangement. To achieve semantic segmentation, I employ convolutional neural networks widely recognized for their effectiveness in image analysis tasks. I also introduce a material recommendation method that aligns with the geographical context of the construction site. Through extensive experiments, it is shown that the proposed method outperforms the previous state-of-the-art method in accurately generating semantic segmentation maps for the inputted blueprint images. I expect that these superior segmentation results will significantly enhance the architectural planning process by providing architects and designers with a more detailed and informative representation of blueprint layouts, thus aiding in better decision-making and design refinement.

Introduction

Problem Definition

The rapid expansion of urban areas and the global population have given rise to an unprecedented demand for housing and infrastructure development. Due to this, the construction industry finds itself facing two challenges: the need for expeditious architectural planning and to adopt environmentally sustainable practices. Architects and construction professionals need to work efficiently to design and plan new structures. Any delays or inefficiencies in the architectural planning phase can result in increased costs. Nevertheless, a frequent challenge arises from the misinterpretation of the blueprints, which not only prolongs the process but also drives up costs significantly.

To solve this problem, there is a need to develop an automated blueprint interpretation system to reduce errors, enhance efficiency, and ultimately facilitate cost-effective and sustainable construction projects. Additionally, with the increasing availability of digital architectural blueprints and data, there is an opportunity to leverage technology, such as machine learning, to expedite the interpretation process.

Proposed Method

In this paper, I proposed a novel machine learning-based semantic segmentation network for architectural blueprint interpretation. This system aims to expedite the blueprint interpretation process through discerning room types and their spatial arrangement. To create this network, I leverage convolutional neural networks as they are effective in computer vision analysis. With geographical input from the user, the model also outputs material recommendations to the environmental benefit of the house's location.

The structural organization of this research paper is as follows. Chapter 2 provides a comprehensive introduction to the necessary background knowledge required for a clear understanding of the proposed method. Chapter 3 explains the details of the proposed method. In Chapter 4, we present an extensive analysis of experimental results. In conclusion, Chapter 5 provides a summary and conclusion of the paper.

Background Knowledge

Blueprint Image

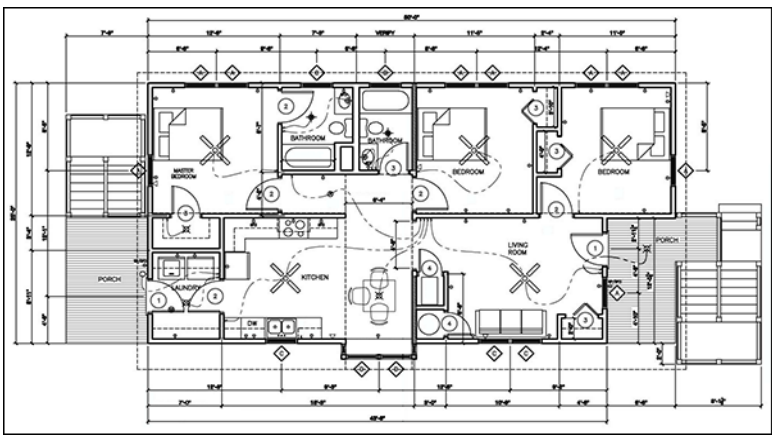


Figure 1. Example of architectural blueprint image

Architectural blueprints are used for planning and constructing buildings. They are a reproduction of a technical or engineering drawing. As displayed in Figure 1, blueprints are 2D drawings with building features such as doors, furniture, and windows indicated with symbols. Architectural blueprints are key to communicating the building plan between anyone working on the project and the building team.

Architectural blueprints use a standardized set of symbols, lines, and notations to represent various elements of a building, such as walls, doors, windows, and electrical systems. Deciphering these symbols and conventions can be daunting for those not trained in reading blueprints. Also, to understand and read the blueprint images, people need to be able to visualize how different elements fit together in space and understand the spatial relationships between rooms, components, and systems. This spatial visualization skill can be challenging for some people. To address this issue, in this paper, I present a novel approach employing image semantic segmentation. Detailed information regarding image segmentation is provided in Chapter 2.2.

Image Segmentation

Image segmentation is one of the most common computer vision techniques that divides an image into multiple meaningful and distinct regions or segments. These segments are typically based on certain criteria such as color, intensity, texture, or other visual properties. The goal of the image segmentation task is to simplify the representation of an image by partitioning it into smaller, more manageable and semantically meaningful regions as shown in Figure 2.

Figure 2. Example of image segmentation (Geiger et al. 2013), (left: input image, and right: output segmentation map)

Image segmentation has a wide range of applications across various machine learning fields such as medical image analysis, environmental monitoring, and object detection and recognition. Numerous research studies have been proposed to develop image segmentation architectures (Minaee et al. 2021).

In this research, I address architectural blueprint interpretation using image segmentation techniques. The proposed blueprint interpretation system produces segmented regions that can pinpoint specific areas like kitchens, rooms, and living rooms within architectural blueprint images. Comprehensive details, including the network design and training strategy, will be elaborated in Chapter 3.

Proposed Method

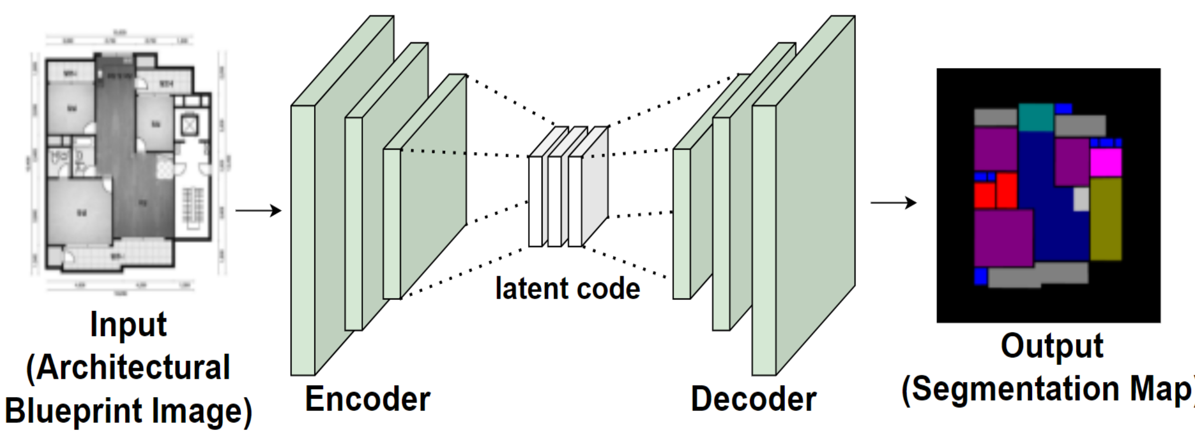


Figure 3. Architecture of the proposed architectural blueprint segmentation network

Figure 3 provides an overview of the architectural blueprint segmentation network proposed in this research paper. The network is constructed using the U-Net architecture (Ronneberger et al. 2015), featuring both a convolutional neural network-based encoder and decoder. The encoder processes an architectural blueprint image as its input, producing a latent code that encapsulates meaningful mathematical features extracted from the input blueprint image. The decoder utilizes this latent code to generate a segmentation map that segments the specific rooms within the house. In the following subheading chapters, I will explain the detailed steps of each component including their mathematical formulations and associated loss functions.

Encoder and Decoder

The input image, the architectural blueprint image, is symbolized as I. This input image, is defined as $I \in R^{HW}$ where H and W denote the height and width of I, respectively. This image is placed into an encoder, $E: I \rightarrow Z$, where $Z \in R^{L}$ represents the latent code and L denotes the dimension of the latent code. The latent code consists of high-dimensional vectors, with each dimension representing a specific feature extracted from the input blueprint image, such as the location or the type of a particular room area. The latent code passes the decoder which then outputs the semantic segmentation map of the blueprint, recognizing and classifying each of the areas within the blueprint image. This is then run through the decoder, $D: Z \rightarrow S$, where S represents the final image, the segmentation map, where $S \in R^{HW}$ with H representing the height and W representing the width of the map. While both the encoder and decoder implement convolutional neural networks, the encoder has a down sampling layer while the decoder has an up sampling layer.

To construct the proposed encoder and decoder, I utilized six convolutional neural network architectures that have shown comparable performance in general image semantic segmentation. A detailed result of the comprehensive experiments can be found in Chapter 4.

Loss Function and Implementation Details

For training the proposed blueprint semantic segmentation system, I employed the Dice coefficient loss function, as depicted in Equation 1. The loss function quantifies how inaccurate a model is at predicting or classifying a data set. During the training process, a gradient descent algorithm is applied to minimize this loss, which, in turn, helps the model make better predictions.

Equation 1: Dice coefficient loss function

$$L_{dice} = 1 - \frac{2 \cdot ||GT \cap PRED||}{||GT|| + ||PRED||}$$

Here, $PRED$ and GT denote the prediction and ground truth respectively. As formulated in Equation 1, the loss function is determined by two times the intersection of GT and $PRED$ divided by the sum of GT and $PRED$ all subtracted from 1. With this formula, the loss function should return 0 when the function is completely accurate while it should return 1 if it is not accurate at all.

Material Recommendation Module

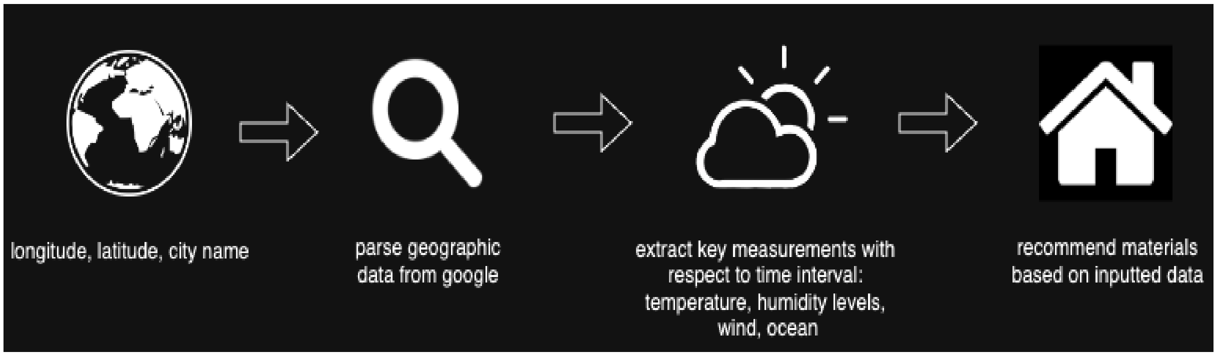
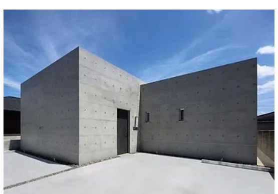


Figure 4. Flow chart of the proposed material recommendation module

Figure 4 describes the method in which the module ultimately proposes the materials recommendation. The user must input longitude, latitude, and the city name in which the residence will be constructed. The module uses this information by parsing through geographic data collected from google. From here key measurements including temperature, humidity levels, and wind from the area are collected to be analyzed and provide a final recommended materials list.

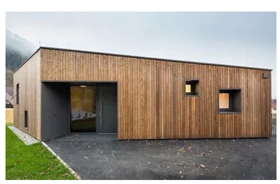

(a)


(b)


(c)


(d)


(e)


(f)


(g)

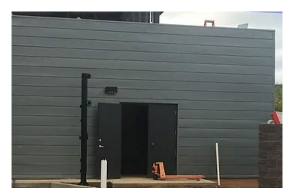

(h)

Figure 5. Example of building exterior finishing materials
(a): exposed concrete, (b): stucco, (c): siding, (d): masonry, (e): reclaimed bricks, (f): western red cedar, (g): zinc, and (h): metal sandwich panel

Figure 5 illustrates various examples of building exterior finishing materials. The selection of these materials is a critical consideration to ensure the durability and maintenance of a building, especially in response to weather, temperature fluctuations, and other factors.

Exposed concrete is frequently used when aiming for a sophisticated design, but it is susceptible to heat loss, and achieving complete external insulation is not easy, and it can be quite costly. Therefore, it is recommended for use primarily in areas with cooler weather. Starco wall finishing has the advantage of providing both insulation and finishing simultaneously. It is relatively cost-effective, but it is vulnerable to fire, so it is not recommended for use in dry regions. In the case of siding, ensuring effective prevention of water leakage from external rainfall can help reduce problems and incidents. When it comes to zinc finishing, minimal joints between the panels and well-finished flashing details alleviate concerns about leakage. Due to its elegant appearance, it is being widely used in recent buildings.

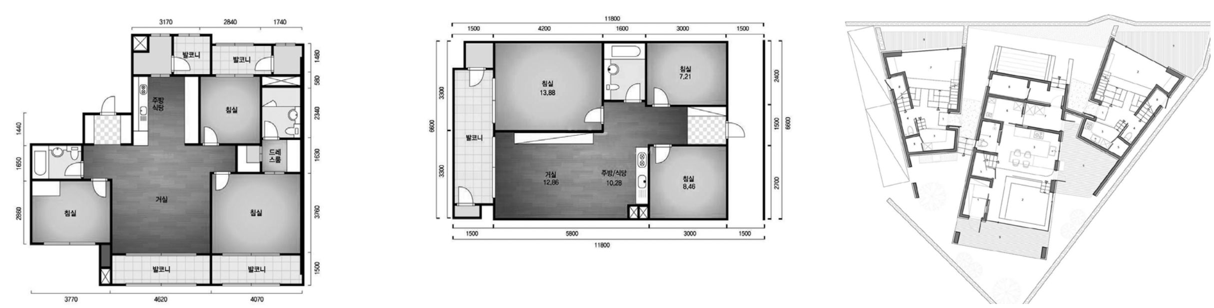
Results

Dataset

As shown in the Table 1, the dataset consists of three categories of residence: apartments, multiplex housing, and houses. Each datapoint consisted of an input, the blueprint, and the ground truth, the residence category.

Table 1. Dataset sample distribution

Residence Type	Amount of Data
Apartment	33,998 (81.8%)
Multiplex Housing	3,871 (9.32%)
House	3,687 (8.88%)
Total	41,556



(a) (b) (c)

Figure 6. Example of blueprint images
(a): Apartment, (b): multiplex housing, and (c): house

Figure 6 displays an example of blueprints for each type of residence. In comparison to (a) and (b), the blueprints of (c) are much more diverse. The model was trained through this dataset, searching for patterns within each category of residence to ultimately be able to accurately identify which type of residence a blueprint is representative of.

Evaluation Metrics

In this study, the mean intersection over union (I_oU) is used to evaluate the accuracy of a model. As mathematically represented in Equation2, the intersection over union takes the intersection between the data input and the model's prediction and divides by the union of the data input and the model's prediction. In this way, the most inaccurate prediction will result in an I_oU of 0 while the most accurate prediction will result in an I_oU of 1.

Equation 2: Intersection over union

$$I_oU = \frac{GT \cap Pred}{GT \cup Pred}$$

Here, *GT* ad *Pred* denotes the ground truth and model's prediction, respectively.

Performance Comparison

The evaluation metrics of this study were measured and compared to other models through three different channels: Multiplex Housing, Apartments, and Houses. The following chart represents the results from these evaluation metrics:

Table 2. Performance comparison

Method	Multiplex Housing	Apartment	House	Total
PSPNet (Zhao et al. 2017)	82.4	82.8	81.1	81.8
Multipath-RetineNet (Lin et al. 2017)	82.9	84.1	81.0	83.1
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
CASIA_IVA_SDN (Fu et al. 2019)	85.4	86.8	82.3	85.6
DeepLabV3 (Chen et al. 2017)	**87.2**	**88.5**	**83.9**	**87.9**

As shown in Table 2, the number in the cell represents the accuracy of the model, with greater numbers indicating greater accuracy. The last row in this table displays the metrics of the DeepLabV3, which proves to have higher accuracy than the other models. These evaluation metrics provide a method of comparison, proving the greater accuracy of this study's model.

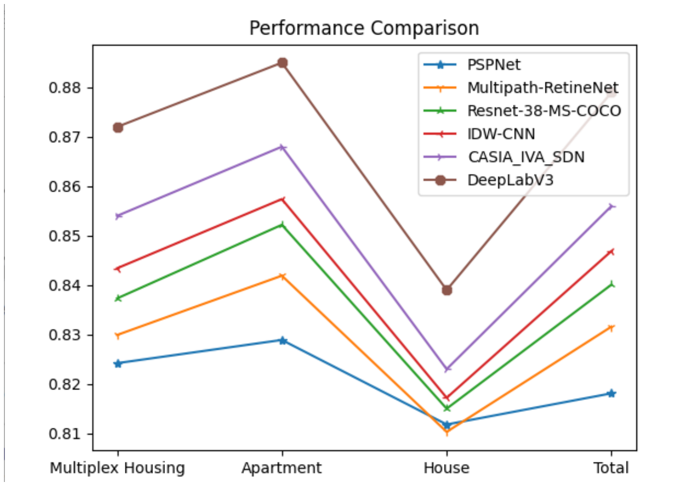
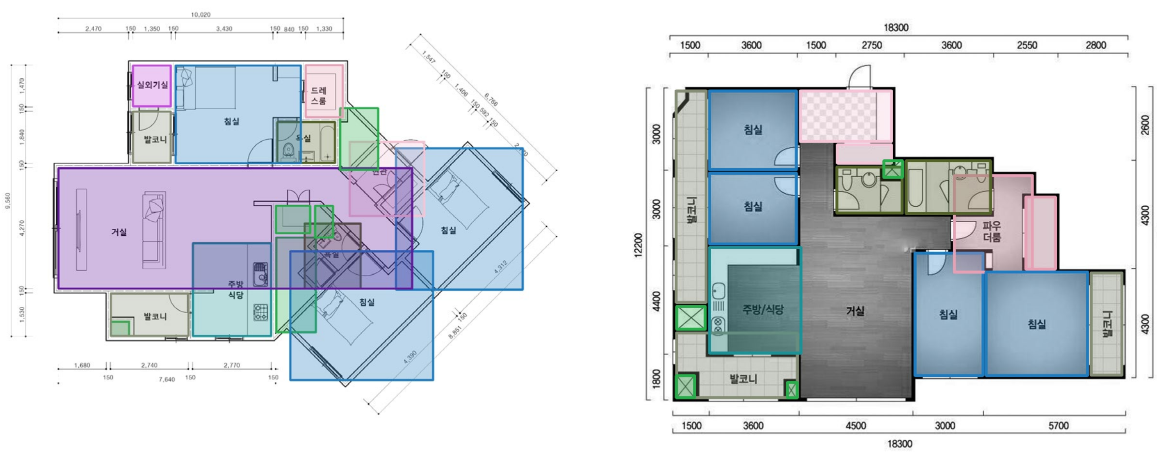


Figure 7. Performance Comparison Between Various Models

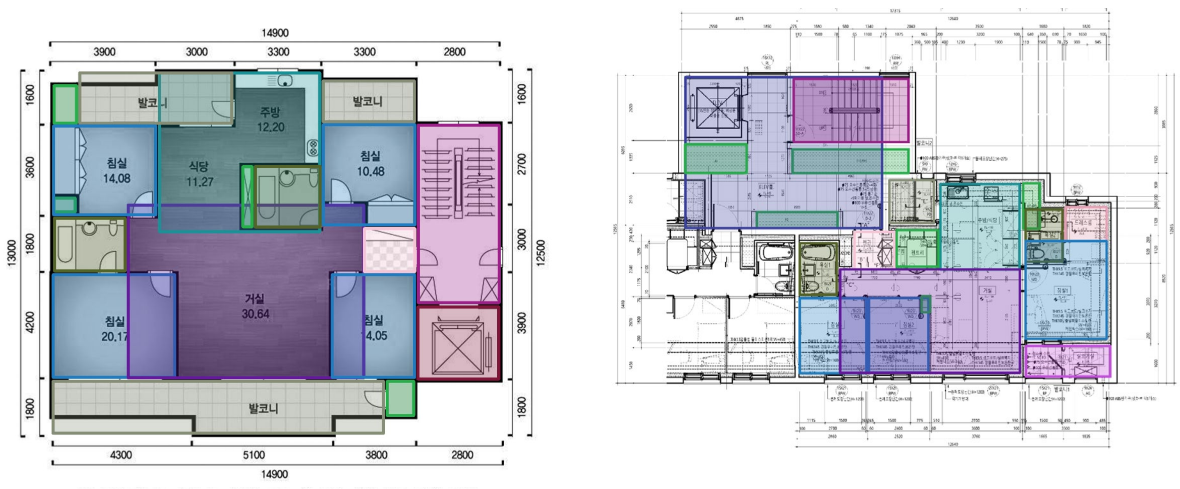
Figure 7 models the results of the calculated mI_oU values for each channel in a line graph. The figure shows a clear distinction between the DeepLab V3model in comparison to the other models. DeepLabV3 is significantly more accurate than all the other tested models in each of the channels. The second most accurate model was the CASIA_IVA_SDN, which also performed better than the other models in all channels.

Qualitative Experiment



(a) (b)

HIGH SCHOOL EDITION
Journal of Student Research



(c) (d)

Figure 8. Example samples of qualitative experiment
(a): house, (b): multiplex housing, (c): apartment, and (d): house

Figure 8 presents samples from a visual qualitative experiment. Each colored box indicates the location and type of specific rooms. For instance, purple and blue boxes represent the living room and bedroom, while pink and green boxes denote the dressing room and storage.

The proposed system tends to produce poor results when dealing with non-rectangular areas commonly found in house blueprints. This issue stems from the training sample annotations, which are exclusively in rectangle shapes. I anticipate that this problem can be easily addressed by employing a more detailed polygon annotation format.

Conclusion

In this paper, I proposed the use of a machine learning-based semantic segmentation network for architectural blueprint interpretation to eliminate the inefficiencies in the blueprint interpretation and residence construction process. I then use convolutional neural networks to introduce a material recommendation method to promote environmentally aware construction decisions. Overall, the experimental results prove that this new model is more efficient and accurate in comparison to other models that attempt to achieve similar goals.
For the future, I intend to further the accuracy and efficiency of this model as well as add more features and variables to improve recommendation outputs.

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

Fu, J., Liu, J., Wang, Y., Zhou, J., Wang, C., & Lu, H. (2019). Stacked deconvolutional network for semantic segmentation. IEEE Transactions on Image Processing. https://doi.org/10.1109/TIP.2019.2895460

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

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2021). Image segmentation using deep learning: A survey. IEEE transactions on pattern analysis and machine intelligence, 44(7), 3523-3542.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing. https://doi.org/10.48550/arXiv.1505.04597

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