Inference of Relevant BIM Objects Using CNN for Visual-input Based Auto-Modeling

J. Kim[a], J. Song[a], and J. Lee[a]

[a] Department of Interior Architecture and Built Environment, Yonsei University, South Korea
E-mail: wlstjd1320@gmail.com, songjy92@gmail.com, leejinkook@yonsei.ac.kr

Abstract –

This paper aims to propose an approach to inferring relevant BIM objects using techniques for recognition of design attributes and calculation of visual similarity using a deep convolutional neural network. Main objective is a visual-input based modelling approach to the automated building and interior design, and this paper represents a preliminary yet critical part of the process. While designing a building, it is important to consider requirements such strong constraints (e.g. regulations, a request for proposal) as well as relatively weak and qualitative constraints such as preference style of clients or users, design trend and etc. Building Information Modeling (BIM) enables to execute to check and review a building design according to the constraints. Until now there is no research that has focused on relatively qualitative and "soft" constraints such as preference or design trend. As a part of research on design supporting system for such "soft" constraints, this paper focuses on training deep learning models that recognize design attributes of BIM object, calculate visual similarity with other objects, and for visual-input based auto-modeling on BIM using the models. A deep convolutional neural network is utilized to extract a visual feature of the 3D object. The input data type for extracting feature data is a 2D rendering image of an object with a specific view and option. The target object is a chair. The feature data is used as input to training models inferring design attributes such as design style, seating capacity and sub-type of a chair and also calculating the visual similarity between objects. This models plays an important role of visual-input based automated modelling system.

Keywords –

Automated Modeling, Building Information Modeling (BIM); Design Attributes; Visual data; Visual similarity;

1 Introduction

In general, design problems have been known as "wicked problem [1]" that means ill-defined problem including complex and various requirement [2]. In particular, while designing a building, requirements include strong constraints such building permit regulations, a request for proposal (RFP) as well as relatively weak and qualitative constraints such as preference style of clients or users, design trend and etc.

Building Information Modeling (BIM) enables a computer to understand a building design model and execute to check and review and simulate the digital model with consideration of the above constraints [3]. Until now the researches have focused on relatively quantitative constraints (e.g. energy simulation [4], quantity take-off [5] and etc.), or strong constraints (e.g. rule-checking [6], in particular, indoor circulation [7] and space program [8]). It is true that consideration of such constraints is important as increasing the size of buildings [9] and changing standards of building performance.

In a high level of detailed design phase such as interiors, however, the relatively qualitative and "soft" constraints such as preference or design trend need to be considered as well. Nevertheless, research focusing on "soft" constraints using BIM has not been studied so far. As a part of research on design supporting system for such "soft" constraints, the ultimate goal of this study is to propose a BIM-based design approach using a deep learning model that can understand the design. The approach targets the visual representation of building design references, among other media representing such constraints. Traditionally, visual representation has been the role of important sources to analysis and understand architectural design [10].

As a preliminary study for the implementation of this approach, this paper explores the ways to combine deep learning-based model understanding design reference image or picture with BIM-based design. The scope of this paper is on training deep learning models that recognize design attributes of BIM object, calculate

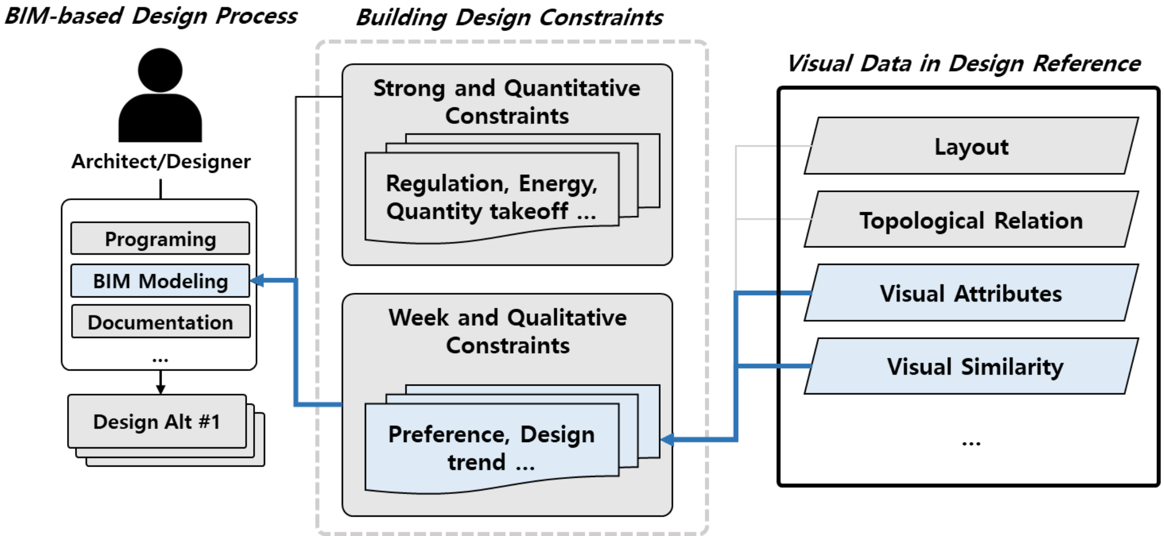


Figure 1. The scope of this paper for visual-input based modelling approach

visual similarity with other objects, and propose an approach to visual-input based auto-modeling on BIM using the models (Figure 1).

For training such intelligent models, this paper utilizes a deep convolutional neural network (CNN) model that showed the successful performance of understanding general object [11], context [12]and even design attributes of interior design object [13] on visual data. The pre-trained CNN model (VGGnet [14]) is used to extract visual feature map data from 2D rendering image of a design object. The feature map data is used as input data to train models recognizing design attributes of chair object (e.g. design style, seating capacity and sub-type). The results of recognition are used to query BIM object in BIM library as a semantic annotation. The feature map data is also used to calculate the visual similarity between objects for the similarity-based query.

2 Background

2.1 Reusing Design Objects in BIM-based Design Process

While solving design problems such as architectural, interior design and etc, the strategy based on generating possible design alternatives is generally adopted [15]. Since it is difficult to make a clear and explicit definition of the problem ("wicked problem") [16] architects and designers compare, analyze and test the alternatives to reach a satisfactory answer.

Building information modeling (BIM) provides the advantages of rapidly generating design alternatives, evaluating and analyzing them. BIM supports the model-oriented design process, enabling integrated and automatic design tasks about as drawing, simulation, various review processes, and visualization with BIM model and related applications. BIM is derived from the idea that a building is considered as a spatial composition of the spatial composition of a set of parts [17]. In other words, determining which design elements will be utilized and how to combine them is an important process for BIM-based architectural design.

The Search accuracy of traditional keyword-based retrieval models, such as Boolean model, vector space model, or probabilistic model, has been often problematic because of the semantic ambiguity of terminologies in BIM documents and queries. The semantic ambiguity in BIM documents can be alleviated by using a domain ontology. As compared to the traditional metadata retrieval of BIM models, Ge Gao et. al. [18] present a way to search into the content of BIM models, which contains more information. It provides a way to customize their requirement specifically. Both accurate numerical value search and text search is supported A survey of the state of the art is beyond the scope of this paper. Instead, this section briefly reviews the most related studies associated with our work.

2.2 Deep Convolutional Neural Network for Extracting Visual Feature

Deep convolutional neural networks (CNN) is known as the most powerful method to extract a visual feature from pixel data of image [11]. For the feature extraction, stacked convolution filters and pooling filters are used

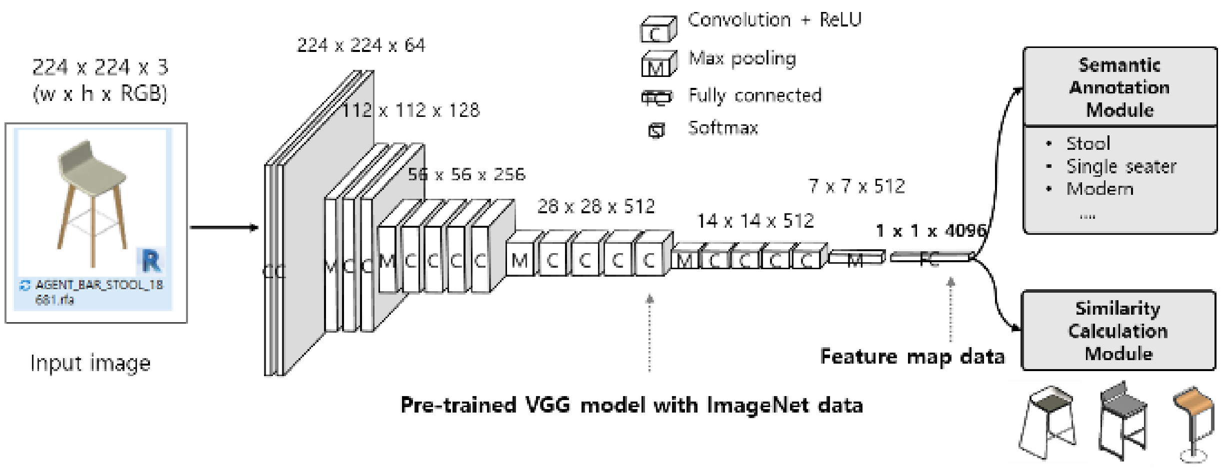


Figure 2. Extraction and utilization of feature map data to get semantic annotation and calculate similarity using VGG model [14]

(Figure 2.). These filters have a specific (n x n) size and are used to screen key pixel data as you scroll through the image to extract a simpler abstracted 'feature map'. The input images are converted to multi-dimensional vector values representing complex elements, such as points, lines, faces, shapes, color, and sizes. These feature map data can be used to train multi-class inference models with softmax algorithm [19] or support vector machine [20] and to find a visual similar one with the Euclidean distance or cosine angle distance [21]. A good model can extract meaningful visual feature. Such good models (e.g. VGG pre-trained with ImageNet [14]) and frameworks (e.g. Tensorflow [22]) are open to the public. The pre-trained model can be directly used to extract features of other images. In this paper, we utilize the pre-trained VGG model to extract visual feature data of BIM object.

3 Methodology

3.1 Approach

This paper proposes an approach to recommend a relevant collection of design objects using visual information from BIM objects. A collection of design objects means a set of objects that a user needs in the context of a particular design process. Metadata, geometry data, and topological relational data can be used when sorting a set of objects in the BIM library, but in this study, semantic annotation information and visual similarity information are only treated. The information of annotation is semantic string information that can be obtained from the shape. For example, when an image of a chair is given, string information such as category, sub-type, functional feature (e.g. armrests, backrest, wheel and etc.), the seating capacity, and even design styles can be inferred. The visual similarity means a degree of visual similarity between the shape of a target object and a shape of the others.

In this paper, the feature map data derived from 2D rendering image of BIM object is used to get semantic annotation and visual similarity information. To get 2D rendering images of BIM object, we develop the image generating module using rendering engine and API of BIM Tool. These images are input into the deep convolutional neural networks model [11] and the feature map data is generated. We utilize VGGnet [14] model that previously trained the ImageNet dataset as feature map extractor. It has been proven that the pre-trained model very well extracts important feature map of the image, and showed the high accuracy (93.2%) of image classification task with at ILSVRC challenge [14].

The feature map data are used as an input in two modules: 1) Semantic annotation module, 2) Similarity between objects calculation module. 2D images of objects are collected through the API of the BIM tool. The inference of semantic annotation information is executed using the softmax algorithm with, and the calculation of visual similarity is done using the Cosine angle distance. Through these two modules, the collection of related and visually similar objects can be recommended.

3.2 Extraction of Visual Feature from Object using CNN

In this study, the feature map data extracted from the 2D image is utilized as visual feature data of the 3D object. CNN model can extract the feature map data that means the abstraction of pixel data that is considered important in the image among the 3 channel (RGB color) pixel data of the input2D image. We extract the featured 4096-dimensional vector value, using the VGG model pre-trained with ImageNet dataset. Images of 3D objects for input into the model are generated by the rendering engine of the BIM software. We collect the BIM object from public BIM object library such as BIMObject [23], Revitcity [24] and NBS National BIM Library [25]. We developed plug-ins using API to automatically generate such images from consistent and accurate view angles with specific rendering option. In this paper, two different views are used: 1) front view and 2) view of horizontal 45 degrees with vertical 60 degrees. There are no shadows but only shading and realistic color with line stroke. The results of extracted feature maps can be saved as an ASCII file, so the feature map data is stored in an external database or in property field of the BIM object. The feature is used for inferring the semantic information of the object and calculating visual similarity. Sections 3.3 and 3.4 describe methodology and calculatation results using feature map data, respectively.

3.3 Inferring Semantics Information of Object

We train the model inferring semantics information of BIM object using the softmax algorithm [26]. The softmax algorithm is the most popular function for inferring labeling of data. Softmax function takes an N-dimensional vector of real numbers and transforms it into a vector of a real number in the range (0,1). The outputs are marginal probabilities and therefore can be used for multiple-class classification.

The target object is a chair and the scope of semantic information includes a sub-type, seating capacity, a design style. There are other many design related attributes, but we adopt some attributes proposed by Kim et al [13]. The sub-type means conceptual and qualitative classification criteria according to function or shape. This paper covers a normal chair, stool, sofa and office chair. Seating capacity means how many seaters a chair has and this is seen as a quantitative property. The design style is a very qualitative feature that can be perceived differently according to culture, period, region and people. Therefore, this paper cover causal, classic and modern that can be seen as the simple and rough classification criteria of design style.

We use BIM seating object images that are generated from the BIM tool as well as real chair product images that are collected from the catalog. Finally, 250 images per class were collected, 20% of which are utilized for validating and 80% for training. Training model for the design style was conducted with only three detailed classes, so 750 images were utilized. While the others have four detailed classes, so 1000 images were utilized. Training is executed on the Tensorflow framework [22] that is the most useful to train machine learning model. Table 1 shows the results training of models. The accuracy of models training design style and sub-type is closed to 85%. On the other hand, the accuracy of model training seating capacity is relatively lower than others (closed to 75%).

Table 1. The results of training models to recognize design attributes of a chair object

Model	Class	Images	Training Accuracy
Seating capacity	Single 2-seaters 3-seaters 4-seaters	1000	75.2%
Design style	Casual Classic Modern	750	86.2%
Sub-type	Chair Sofa Stool Office chair	1000	85.1%

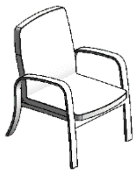
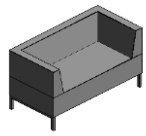
3.4 Calculating visual feature similarity with Cosine Angle Distance

The similarity is calculated through the cosine distance with vector feature maps. The cosine distance is the cosine value of the angle between two vectors, which the value is bound on the interval (-1~1). The correlation of distribution between random vectors may tend to get closer to zero as the dimensions of space increases. Therefore, the even small value of similarity become significant with growing dimension. Given two vectors of visual feature map, VF_a and VF_b on the n-dimensional space, and cosine similarity, $\cos(\theta)$, is represented as;

$$\cos(\theta) = \frac{\sum_{i=1}^{n} VF_{a,i} \times VF_{b,i}}{\sqrt{\sum_{i=1}^{n} {VF_{a,i}}^2} \times \sqrt{\sum_{i=1}^{n} {VF_{b,i}}^2}} \quad (1)$$

The resulting similarity ranges from −1 meaning exactly opposite, to 1 meaning exactly the same or a very similar one and with 0 indicating orthogonality (decorrelation). The feature map data is the positive number, so practical similarity value ranges from 0 to 1. Table 2 shows the examples of calculation of similarity between chair objects and the inferred semantic annotations. The target object is a famous stool and the

Table 2. The examples of inferring semantic annotation and calculating similarities between target and inputs objects using feature map data and cosine distance

	Target	Input #1	Input #2	Input #3	Input #4	Input #5
Image						
Similarity	-	88.9%	81.0%	66.6%	58.1%	51.6%
Seating capacity	Single	Single	Single	Single	Single	2-seaters
Sub-type	Stool	Stool	Stool	Chair	Stool	Sofa
Design style	Modern	Modern	Casual	Casual	Casual	Modern

input #1 is the same one, but slightly different sizes and width. The similarity between the above two is calculated as 88.9%. The input #2 is another stool with different color and detailed design (punched hole). The similarity between target and input #2 is calculated as 81.0% lower than the previous value. Input #3, 5 is not stool, but normal chair and sofa, respectively, and the similarity values are lower than 70%. The input #4 is a round stool, but the similarity value is relatively lower that the values of other stool.

4 Conclusion

As a preliminary study for the implementation of this approach, this paper explores the ways to combine deep learning-based model understanding design reference image or picture with BIM-based design. The scope of this paper is on training deep learning models that recognize design attributes of BIM object, calculate visual similarity with other objects, and propose an approach to visual-input based auto-modeling on BIM using the models. For training such intelligent models, this paper utilizes a deep convolutional neural network (CNN) model that showed the successful performance of understanding general object [11], context [12]and even design attributes of interior design object [13] on visual data. The pre-trained CNN model (VGGnet [14]) is used to extract visual feature map data from 2D rendering image of a design object. The feature map data is used as input data to train models recognizing design attributes of chair object (e.g. design style, seating capacity and sub-type). The results of recognition are used to query BIM object in BIM library as a semantic annotation. The feature map data is also used to calculate the visual similarity between objects for similarity-based query. This models plays an important role of visual-input based automated modelling system.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2019R1A2C1007920).

References

[1] Rittel H.W. and Webber M.M., Wicked problems, *Man-made Futures*, 26(1):272-280, 1974

[2] Simon H.A., The structure of ill structured problems, *Artificial intelligence*, 4(3-4):181-201, 1973

[3] Eastman C., et al., *Book BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors*, John Wiley & Sons, 2011

[4] Schlueter A. and Thesseling F., Building information model based energy/exergy performance assessment in early design stages, *Automation in construction*, 18(2):153-163, 2009

[5] Monteiro A. and Martins J.P., A survey on modeling guidelines for quantity takeoff-oriented BIM-based design, *Automation in Construction*, 35(238-253, 2013

[6] Eastman C., et al., Automatic rule-based checking of building designs, *Automation in construction*, 18(8):1011-1033, 2009

[7] Lee J.-k., et al., Computing walking distances within buildings using the universal circulation network, *Environment and Planning B: Planning*

and Design, 37(4):628-645, 2010
[8] Lee J.-K., et al., Development of space database for automated building design review systems, *Automation in Construction*, 24(203-212, 2012
[9] Choi J., et al., Development of BIM-based evacuation regulation checking system for high-rise and complex buildings, *Automation in Construction*, 46(38-49, 2014
[10] Koutamanis A. and Mitossi V., Computer vision in architectural design, *Design Studies*, 14(1):40-57, 1993
[11] LeCun Y., et al., Deep learning, *nature*, 521(7553):436, 2015
[12] Karpathy A. and Fei-Fei L., Deep visual-semantic alignments for generating image descriptions, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3128-3137, 2015
[13] Kim J.s., et al., Approach to the Extraction of Design Features of Interior Design Elements using Image Recognition Technique, *Proceedings of the 23rd International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2018*, pages 287-296, Beijing, China, 2018
[14] Simonyan K. and Zisserman A., Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*, 2014
[15] Cross N., *Book Designerly ways of knowing*, Springer, 2006
[16] Buchanan R., Wicked problems in design thinking, *Design issues*, 8(2):5-21, 1992
[17] Eastman C., An Outline of the Building Description System. Research Report No. 50, Institute of Physical Planning, Carnegie Mellon Univ., Pittsburgh, 1974
[18] Gao G., et al., A query expansion method for retrieving online BIM resources based on Industry Foundation Classes, *Automation in construction*, 56(14-25, 2015
[19] Szegedy C., et al., Rethinking the inception architecture for computer vision, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818-2826, 2016
[20] Niu X.-X. and Suen C.Y., A novel hybrid CNN–SVM classifier for recognizing handwritten digits, *Pattern Recognition*, 45(4):1318-1325, 2012
[21] Qian G., et al., Similarity between Euclidean and cosine angle distance for nearest neighbor queries, *Proceedings of the 2004 ACM symposium on Applied computing*, pages 1232-1237, 2004
[22] Abadi M., et al., Tensorflow: Large-scale machine learning on heterogeneous distributed systems, *arXiv preprint arXiv:1603.04467*, 2016
[23] BIMObject: BIM Libarary. Online: https://www.bimobject.com/, Ascessed: 15/01/2019
[24] Revitcity:Public BIM library. Online: https://www.revitcity.com, Ascessed: 15/01/2019
[25] NBS National Bim Library. Online: https://www.nationalbimlibrary.com, Ascessed: 15/01/2019
[26] Dunne R.A. and Campbell N.A., On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function, *Proc. 8th Aust. Conf. on the Neural Networks, Melbourne*, pages 185, 1997