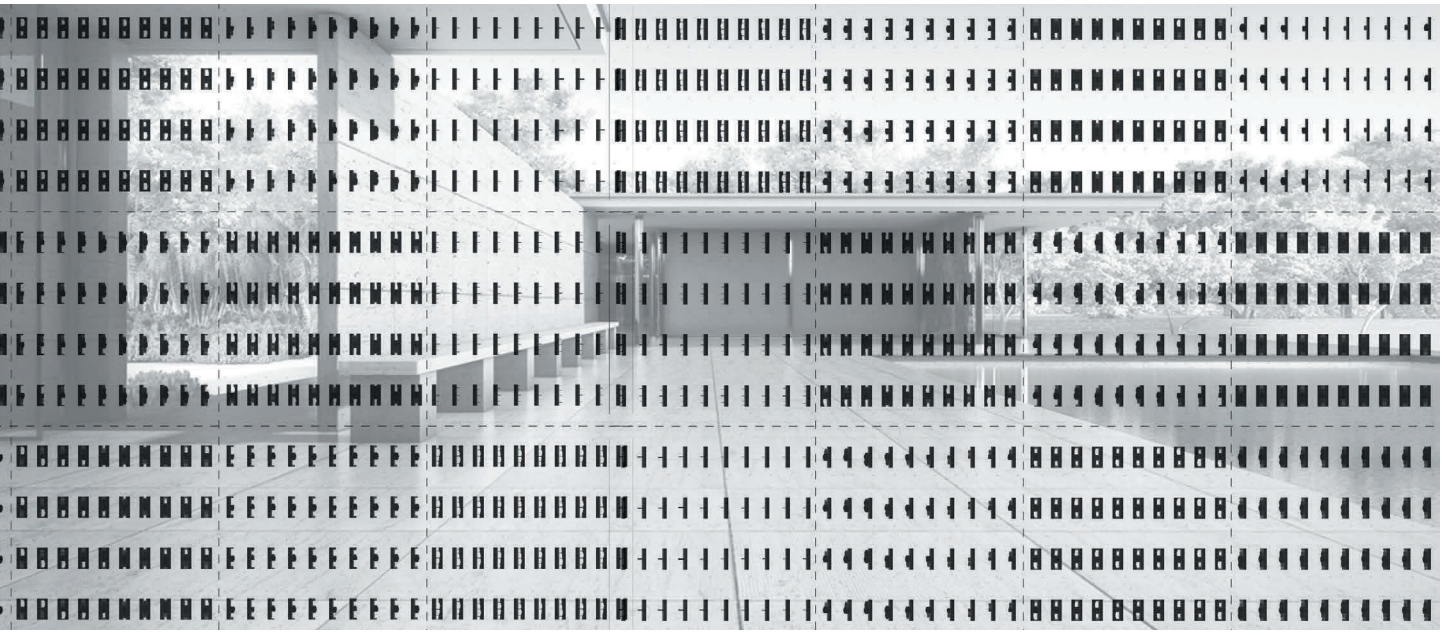
Machines' Perception of Space

Employing 3D Isovist Methods and a Convolutional Neural Network in Architectural Space Classification

Wenzhe Peng
Massachusetts Institute of Technology

Fan Zhang
The Chinese University of Hong Kong

Takehiko Nagakura
Massachusetts Institute of Technology



1

ABSTRACT

Simple and common architectural elements can be combined to create complex spaces. Different spatial compositions of elements define different spatial boundaries, and each produces a unique local spatial experience to observers inside the space. Therefore an architectural style brings about a distinct spatial experience.

While multiple representation methods are practiced in the field of architecture, there lacks a compelling way to capture and identify spatial experiences. Describing an observer's spatial experiences quantitatively and efficiently is a challenge. In this paper, we propose a method that employs 3D isovist methods and a convolutional neural network (CNN) to achieve recognition of local spatial compositions. The case studies conducted validate that this methodology works well in capturing and identifying local spatial conditions, illustrates the pattern and frequency of their appearance in designs, and indicates peculiar spatial experiences embedded in an architectural style. The case study used small designs by Mies van der Rohe and Aldo van Eyck.

The contribution of this paper is threefold. First, it introduces a sampling method based on 3D Isovist that generates a 2D image that can be used to represent a 3D space from a specific observation point. Second, it employs a CNN model to extract features from the sampled images, then classifies their corresponding space. Third, it demonstrates a few case studies where this space classification method is applied to different architectural styles.

1 Decoding architectural space using machine-learning and computer vision techniques.

2 Isovist

INTRODUCTION

Background and Problem Description

Architecture generally consists of columns, walls, shades, windows, etc. Although these elements are simple and common components, they can be combined to create complex and specific compositions of space. Different spatial compositions define different spatial boundaries, and therefore produce different feelings to the observers inside the space. Against a one-sided wall, enclosed by an L-shaped wall, surrounded by columns, and with or without a shade; all these situations create unique local spatial experiences for the observers. Architects usually consider these local spatial experiences, and their sequence is critical to architectural design.

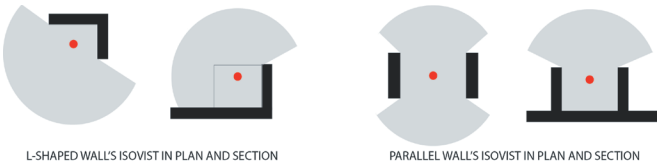
While multiple representation methods are practiced in the field of architecture, there is a lack of compelling ways to capture and identify local spatial experience. Therefore, it can be challenging for architects to describe spatial experience quantitatively and efficiently. In this paper, we propose a system that employs a 3D isovist spatial representation method and a convolutional neural network (CNN) machine-learning algorithm to achieve recognition of local spatial compositions. Case studies are then conducted using existing buildings designed by notable architects. The tests evaluate if the method reveals spatial experience typically found in an architectural style.

Related Works

In his book *Architecture: Form, Space, and Order*, Francis D.K. Ching stated the relationship between form and space:

> Space constantly encompasses our being. Through the volume of space, we move, see forms, hear sounds, feel breezes, smell the fragrances of a flower garden in bloom. It is a material substance like wood or stone. Yet it is an inherently formless vapor. Its visual form, its dimensions and scale, the quality of its light—all of these qualities depend on our perception of the spatial boundaries defined by elements of form. As space begins to be captured, enclosed, molded, and organized by the elements of mass, architecture comes into being." (Ching 2014)

An isovist is one spatial representation method to study space composition (Benedikt 1979). It is the volume visible from a given point in space, together with a specification of the location of that point. A 2D isovist captures the essential boundary of a 2D plan. Similarly, a 3D isovist, acquired by projecting rays spherically, captures the spatial boundary of a three-dimensional space composition from an observer's perspective. Figure 2 shows the isovists of two different spatial compositions, in plan and section. The isovist, in gray shades, represents the differences of spatial boundaries defined by different elements.



L-SHAPED WALL'S ISOVIST IN PLAN AND SECTION

PARALLEL WALL'S ISOVIST IN PLAN AND SECTION

2

Recent works in computer vision have proved that using CNNs has been an efficient approach to perform object detection and scene classification (Krizhevsky and Geoffrey 2012; Zeiler and Rob 2014). There is a new research trend on using CNN models to recognize places (Zhang et al. 2016). Compared with traditional methods of image classification, a machine-learning approach has a great advantage at learning features, which has proven more feasible and efficient than hand-crafted features. The importance of CNNs has been well recognized due to their impressive performance of visual recognition tasks.

METHODOLOGY

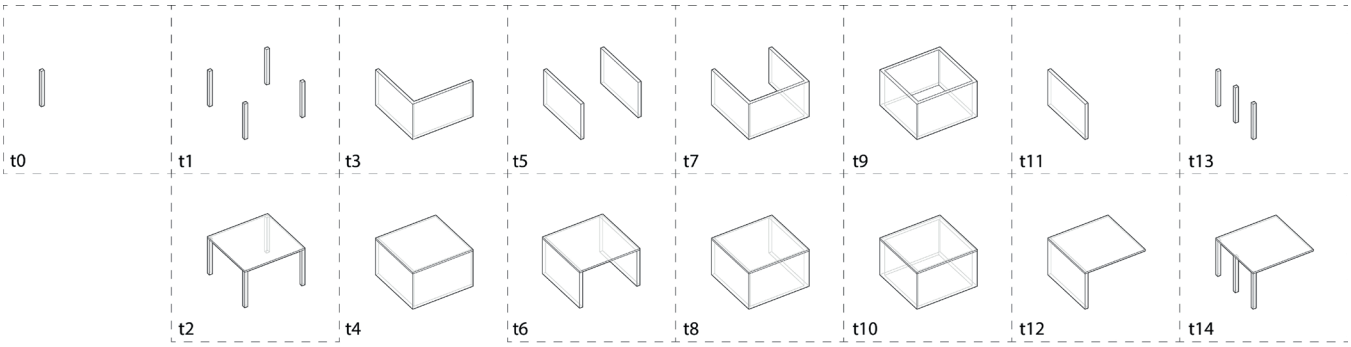
In this section, we introduce the method that we developed to achieve the recognition of local spatial compositions. The proposed method formulates this recognition problem as a discriminative classification task: with several predefined local spatial compositions, or what we call Seed-Spaces, we train a classifier to predict the type of a given space, and in that approach we can sample a building to get a statistical distribution of the predefined elements. The methodology can be applied to different settings of predefined compositions. The details of this methodology are introduced in the following subsections.

Spatial Composition Classification Workflow

Figure 3 shows the framework of the spatial composition classification system. It consists of two phases, training and service. In each phase, the system first performs spatial sampling, and then runs the result through the network, to either train the network or use the pre-trained network to acquire a presumption. In the training phase we use data sampled from pre-defined Seed-Spaces to train our network. The labels of these sampled datasets correspond to the originated Seed-Spaces. In the service phase, the network presumes the most similar Seed-Space for any input data sampled from a given space using the same methodology.

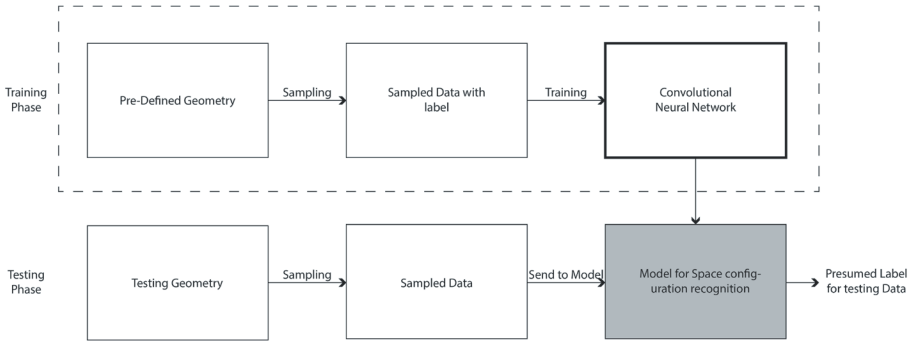
Collection of Local Spatial Compositions

As stated earlier by Ching, space can be composed of elements. Limited elements can create unlimited possibilities of space. Horizontal planes (including base plane, elevated base plane, depressed base plane, and overhead plane), vertical linear elements (like columns), and vertical planes (walls) are considered as the primary architectural space-defining elements.

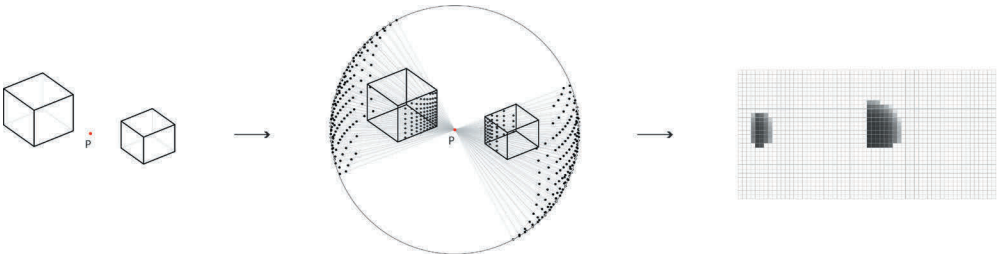


t0 - BY ONE COLUMN
t1 - SUROUNDED BY COLUMNS
t2 - SUROUNDED BY COLUMNS WITH SHADING
t3 - L-SHAPE WALL
t4 - L-SHAPE WALL WITH SHADING
t4 - L-SHAPE WALL WITH SHADING
t5 - PARALLEL WALLS
t6 - PARALLEL WALLS WITH SHADING
t7 - U-SHAPE WALL
t8 - U-SHAPE WALL WITH SHADING
t9 - 4-SHADED WALL
t10 - 4-SHADED WALL WITH SHADING
t11 - BY SINGLE WALL
t12 - BY SINGLE WALL WITH SHADING
t13 - BY ONE-SIDED COLUMNS
t14 - BY ONE-SIDED COLUMNS WITH SHADING

4



3



5

3 Workflow of setting up the classification system

4 The 15 different space types, or Seed-Spaces, labeled as t0, t1, ... , t14.

5 3D isovist sampling of a given space from observation point P.

In this research, we select fifteen basic types of element compositions that make what we call Seed-Spaces to describe the local spatial conditions of one-story buildings. This collection of Seed-Spaces can be customized when dealing with different spatial conditions. These Seed-Spaces are considered to be the primary space types taken as elements to compose other spaces to be analyzed. Therefore, the selected Seed-Spaces are made up of columns, walls, and overhead planes in various ways, and they are listed in Figure 4, which depicts them using isometric drawings.

Space Sampling (Data Collection Using 3D Isovist)

In order to analyze a specific local spatial composition from an observer's perspective, we need a way to capture space that addresses both essentiality and conciseness. Essentiality means the representation should contain the basic configuration of the space from the selected observation point, capturing the corresponding spatial boundaries. Conciseness makes sure that the data captured should also be simplified and can extract certain characters of the original space composition that makes the later analysis feasible.

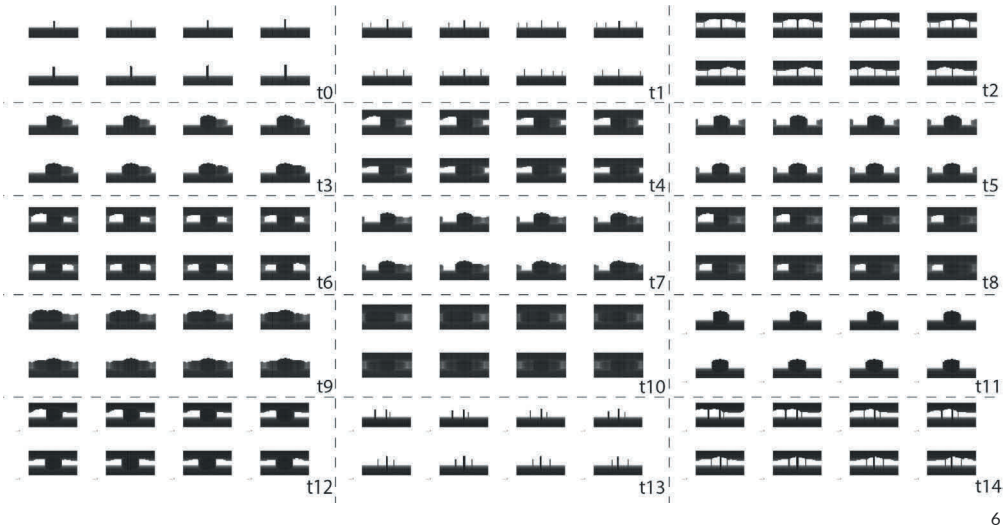
We propose this sampling method that suits both essentiality and conciseness based on the concept of a 3D isovist. By projecting the space information onto a sphere centered on the observation point, we then remap the spherical depth values to a 2D equirectangular "image" with a resolution of 60 x 30. In our experiment, for simplicity, only the value of the distance from the surface of the objects to the observation point is chosen as the criteria of the recognition, which makes up one channel of the image. If values other than distance (like the brightness of the environment/the RGB value of the surrounding space, etc.) are included, they can also be added as additional channels to this "image" to be utilized in the matrix calculation. The maximum sampling distance in our experiment is 10 meters, which is also the radius of the sampling sphere. After the sampling process, the original spatial composition (from the observation point) can be represented as a one-channel image that contains its necessary spatial boundary (distance). This sampling process is shown in Figure 5.

The orientation of the equirectangular representation might be one issue, as different sampling orientations creates horizontally shifted representations. However, the network we developed (introduced in the next section) can learn the orientation of the invariant features. Also, in this practice we shift the image horizontally by centralizing the average darkest column, making sure that the closest boundary is centralized so that the most influential features to the observation point are captured.

Figure 6 shows a part of the samplings of the 15 Seed-Spaces. Only the area enclosed by the space elements are sampled, as we believe that those are the areas that best represent the local spatial condition defined by the surrounding elements. Samples of the Seed-Spaces are labeled accordingly to be constructed as the training set of the network in the training phase of the system. Meanwhile, in the service phase, the same sampling method is applied to incoming spaces, and the sampled image is used to obtain predictions from the pretrained network.

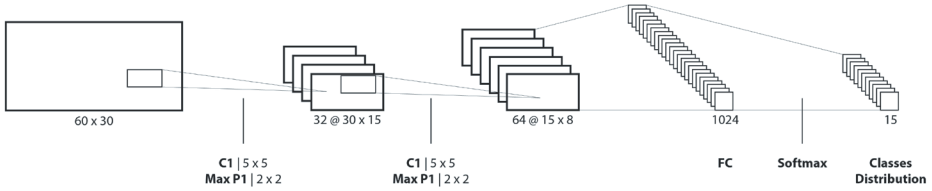
CNN Model (Classification Based on Feature Extraction)

Learning data representation is the fundamental part of pattern recognition and classification tasks. In this case, learning features for a local space-type classification task is more challenging due to the inherent complexity of space and human perceptions of it. To address this issue, we employ a CNN. This method is proved to have a similar performance to human vision in certain contexts (Kheradpisheh et al. 2016), and as a result is expected to simulate humans' perceptions to particular spatial compositions in this study.



6

6 A part of the space samples of the 15 different Seed-Spaces, in equirectangular format. t0, t1, ... , t14 represent each individual Seed-Space defined in the subsection "Collection of Local Spatial Compositions."



7

7 CNN architecture.

Considering the situation of this task, we design the CNN architecture based on the configuration of the input—the equirectangular dataset—as Figure 7 shows. The input image can be considered as a 60 x 30 matrix. After conducting two convolutional layers and two max poolings, a fully connected layer is extracted as the feature vector, and this vector is then used to classify the input image with a fifteen-label softmax classifier with dense connections. The network is set up with TensorFlow (Abadi et al. 2016).

We designed this neural network so that it takes a sampled image as input, and outputs a presumption, or class distribution, among the fifteen predefined Seed-Spaces. Once the CNN is properly trained, utilizing the training set built upon space samplings of the Seed-Spaces, it can be used to make a judgment upon a given space sample, i.e., classify the given space. Our network is trained with a training set of over 5000 images for the fifteen predefined Seed Spaces. It achieved higher than 99% accuracy on validation sets, in other words, it led to a top 1 presumption.

CASE STUDIES AND EXPERIMENTS

We sample and run network through several existing architecture projects using their digital models. With their samplings and the network trained ahead of time on predefined Seed-Spaces, we get spatial-type presumptions of the sampled models, which consists of a statistic report as long as a spatial distribution of Seed-Spaces.

The following three subsections show the results of the system being applied to different building models: Barcelona Pavilion, Exhibition House Berlin 1931, and Paviljoen van Aldo van Eyck. These results are followed by an analysis subsection. We sampled the three models with a resolution of 1 x 1 m from the height of 1.6 m from the floor, and only in areas close enough to the buildings (either inside of the building or within a distance of 4 m) and human accessible (pools are not included). These sampling points are drawn in Figures 8–10 with black or red dots depending on if the recognition result matches the space composition type legend below. Glass and windows and are removed from the model, as we only estimate visual boundaries.

Barcelona Pavilion, Mies van der Rohe

The Barcelona Pavilion was designed by Mies van der Rohe. It was originally designed as the German National Pavilion for the Barcelona International Exhibition, but it soon became an important building in the history of modern architecture, known for its form and use of materials. In our case we focus merely on the form and its style of the building. The result of our samplings and presumptions are shown in Figure 8 and Table 1.

Exhibition House, Berlin, 1931, Mies van der Rohe

The Exhibition House, Berlin is another spatial experiment conducted by Mies in 1931, after the Barcelona Pavilion. The results of our sampling and predictions are shown in Figure 9 and Table 2.

Paviljoen van Aldo van Eyck

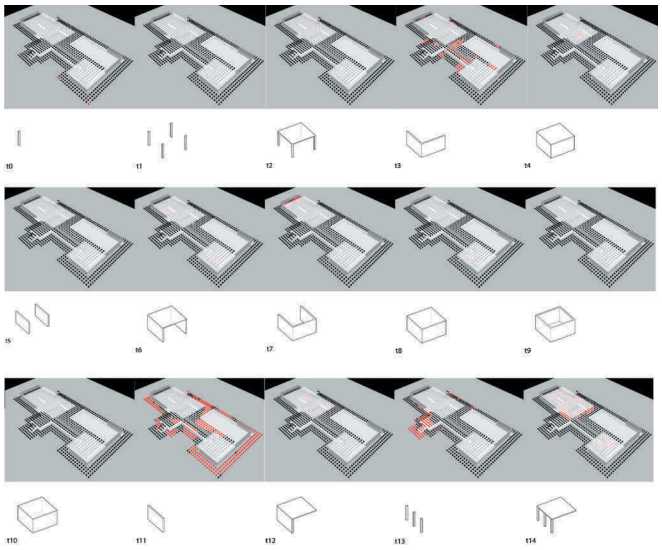
Paviljoen van Aldo van Eyck is the third example we chose to examine. It was designed in 1965-1966 for the 5th International Sculpture Exhibition and heavily featured circles and curves as key elements of Van Eyck's "humane architecture." The results of our sampling and predictions are shown in Figure 10 and Table 3.

Result Analysis

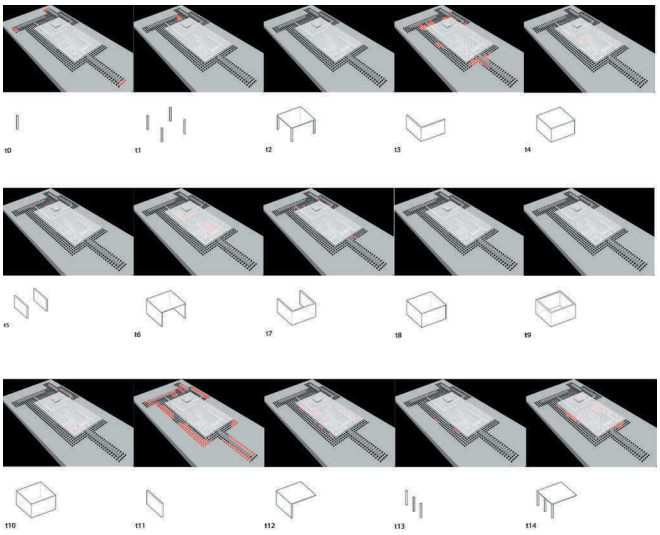
From the results shown in the previous three sections, we see that the system is able to give a reasonable space-type presumption for new sampled spaces. Also, we get statistical results for the whole building by running through a sampling of it. The results can be utilized to distinguish the building or its style. Figure 11 and Table 4 show the accumulated results acquired from the tests of the three buildings.

For case of the Paviljoen van Aldo van Eyck (PVAVE), it is interesting to notice that although the only compositional elements of the building are walls, it's still presumed to have 25.9% column-like space. By checking the distribution mapping of space in Figure 10, we can see the one-sided columns (t13) are primarily situated against the ends of walls, where the local experience is more similar to columns. Similarly, in the PVAVE case, pocket-walls (t7) are shown in many of the samplings, where the observer is surrounded by curved walls on one side. This fact that we can compose one type of space using other types of elements is totally reasonable, though it may not be obvious when only considering the type of the original elements. Spatial experience is too obscure to be described merely using drawings or models, but our methodology suggests a possible solution. Additionally, our system can be improved by adding these new space types, "against walls" or "enclosed by curved walls," to the Seed-Spaces. The system will be able to identify these new spatial compositions with an adjusted network, so that the result will describe the PVAVE's design more precisely.

Comparing the Barcelona Pavilion case and the Exhibition House Berlin case, the Mies designs both have walls and columns in the models. Mies van der Rohe is famous for designing free-flowing spaces, and we can see that the less open space types, such as t7, t8, t9, and t10, rarely appear in these two buildings. The difference of the two also can be identified from the sampling results. The BP has more outdoor space compared to the EH, yet the EH has a greater proportion of walls (Table 4).



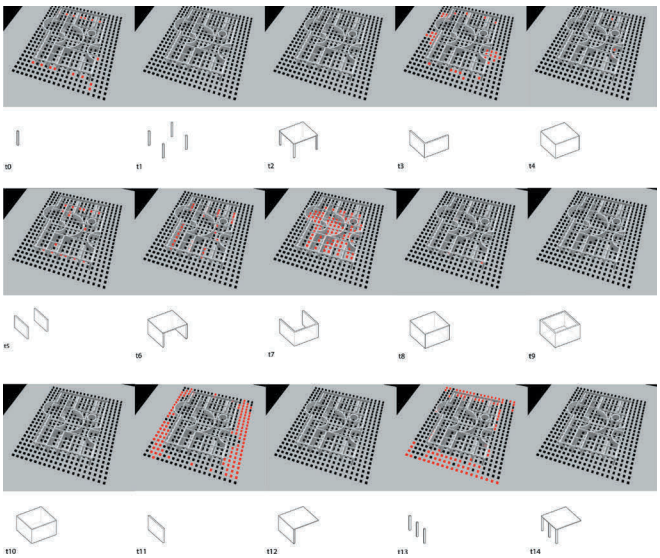
8



9

The main distribution of Seed-Spaces display similar trends for the EH and the BP. The PVAVE case shows different proportions of Seed-Spaces compared with the previous two, although all three designs are one-story buildings composed of mainly walls (t11). The PVAVE case returns more pocket walls (t7) and one-sided columns (t13), but no shaded one-sided columns (t14).

Last but not the least, it's also important to point out that for each sampling, the network actually produces a probability distribution of the Seed-Spaces. In this perspective, each spatial composition can be considered as a "Hybrid Space" of the Seed-Spaces. For example, if the result indicates that the input has a 60% probability to be space t1 and a 40% probability to be space t2, this input space T` is 60% similar to t1 and 40% similar to t2. It can be considered as a hybrid space (0.6t1 + 0.4t2) of the two. This soft assignment allows the system to be applied to analyze space in more flexible scenarios, allowing for the description of transitional space types.



10

CONCLUSION

Although the methodology is only tested with limited spatial composition types and artificial training sets, we see the potential of this methodology in architectural studies. The results of the three building tests show that our system extracts features from space samplings, and produces results that are similar to human perception. By sampling the whole building, we get style-related spatial composition statistics. The following discusses possible future improvements and developments.

As for the system itself, the network may not be deep enough, meaning there might still be the potential to acquire better feature extraction performance with better-crafted networks. A

8 The result of applying the system to the sampled locations in the Barcelona Pavilion model. The black dots in each figure stand for negative samplings for the space composition indicated by the legend below, based on the top 1 presumption result. The red dots stand for the positive ones.

9 The result of applying the system to the sampled locations in the Exhibition House, Berlin model. The black dots in each figure stand for negative samplings for the space composition indicated by the legend below, based on the top 1 presumption result. The red dots stand for the positive ones.

10 The result of applying the system to the sampled locations in the Paviljoen van Aldo van Eyck model. The black dots in each figure stand for negative samplings for the space composition indicated by the legend below, based on the top 1 presumption result. The red dots stand for the positive ones.

space type	t0	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	t11	t12	t13	t14
Sample Amount	6	5	20	115	44	0	41	46	12	0	10	445	42	115	170
Percentage	0.6 %	0.5 %	1.9 %	10.7 %	4.1 %	0 %	3.8 %	4.3 %	1.1 %	0 %	0.9 %	41.5 %	3.9 %	10.7 %	15.9 %

Table 1 Barcelona Pavilion Stats (1071 samples in total)

Table 1

space type	t0	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	t11	t12	t13	t14
Sample Amount	29	23	4	153	83	5	108	36	5	0	73	322	97	29	173
Percentage	2.5 %	2.0 %	0.4 %	13.4 %	7.3 %	0.4 %	9.5 %	3.2 %	0.4 %	0 %	6.4 %	28.2 %	8.5 %	2.5 %	15.2 %

Table 2 Exhibition House, Berlin Stats (1140 samples in total)

Table 2

space type	t0	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	t11	t12	t13	t14
Sample Amount	16	0	0	40	3	17	31	107	3	1	0	115	0	102	1
Percentage	3.7 %	0 %	0 %	9.2 %	0.7 %	3.9 %	7.1 %	24.5 %	0.7 %	0.2 %	0 %	26.4 %	0 %	23.4 %	0.2 %

Table 3 Paviljoen van Aldo van Eyck Stats (436 samples in total)

Table 3

larger training set with greater variety can also help in the generalization of the network, and would help to avoid overfitting.

Considering our experiments, we conduct evenly distributed samplings in the models, acquiring a distribution of Seed-Spaces. In reality, architects may design a building considering the sequences of spatial experiences. It can also be interesting to run sampling on a designated walk-through in the space, which not only gives a space-type distribution, but the variation of different spaces along the sequence. This can lead to interesting research topics in circulation design.

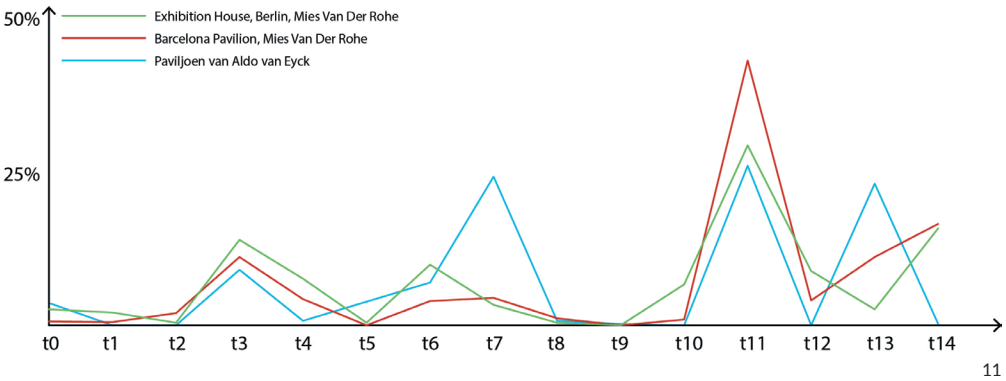
Due to the difficulty and comparatively higher cost of real building sampling, this experiment is conducted with artificial data generated from virtual models. In this way, we can easily generate huge datasets; however, compared to real space, artificial data encapsulates limited features and simplified shapes, which may lead to overfitting and other issues. With the aid of 3D-scanning techniques and photogrammetric sampled models, we can construct training data based on real space data, resulting in a more convincing and better generalized prediction model.

Comparing the machine's perception and a human's perception would also be interesting, and would provide us feedback for improving our system. One approach would be to use VR equipment to test a human subject with an identical space, and compare the result of the human subject with the presumed result of the system. A VR experiment would provide an immersive experience for a subject, yielding a more convincing spatial feeling.

Last but not least, the fifteen local spatial composition types (Seed-Spaces) are used in this experiment to test the classification methodology. However, the Seed-Spaces can be fully customized to a specific context of a given research topic. One potential direction is to research the relationship between function and space configuration. Meanwhile, it would also be interesting to apply the system to evaluate spaces acquired with generative systems, thus allowing the methodology to be further employed in an architectural generative design process.

REFERENCES

Abadi, Martín, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado et al. 2016. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems." arXiv preprint arXiv:1603.04467.



11

11 Plot of Seed-Space distribution in all three buildings.

Space Type	Wall Elements (t3, t4, t5, t6, t7, t8, t9, t10, t11, t12)	Column Elements (t0, t1, t2, t13, t14)	Walls with Shade (t4, t6, t8, t10, t12)	Walls with no Shade (t3, t5, t7, t9, t11)	Columns with Shade (t2, t14)	Columns with no Shade (t0, t1, t13)
Barcelona Pavilion	70.4%	29.6%	13.8%	56.6%	17.8%	11.8%
The Exhibition House, Berlin	77.3%	22.6%	32.1%	45.2%	15.6%	7%
Paviljoen van Aldo van Eyck	72.7%	27.3%	8.5%	64.2%	0.2%	27.1%

Table 4

Table 4 Categorized sampling result for all three buildings.

Benedikt, Michael L. 1979. "To Take Hold of Space: Isovists and Isovist Fields." *Environment and Planning B: Planning and Design* 6 (1): 47–65.

Ching, Francis D. K. 2014. *Architecture: Form, Space, and Order*, 4th ed. Hoboken, NJ: John Wiley & Sons.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. "Imagenet Classification with Deep Convolutional Neural Networks." In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, 1097–1105. Lake Tahoe, NV: NIPS.

Kheradpisheh, Saeed Reza, Masoud Ghodrati, Mohammad Ganjtabesh, and Timothée Masquelier. 2016 "Deep Networks Can Resemble Human Feed-Forward Vision in Invariant Object Recognition." *Scientific Reports* 6: 32672.

Zeiler, Matthew D., and Rob Fergus. 2014. "Visualizing and Understanding Convolutional Networks." In *Proceedings of the 13th European Conference on Computer Vision*, edited by David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, 818–33. Zurich, Switzerland: ECCV.

Zhang, Fan, Fabio Duarte, Ruixian Ma, Dimitrios Milioris, Hui Lin, and Carlo Ratti. 2016. "Indoor Space Recognition using Deep Convolutional Neural Network: A Case Study at MIT Campus." arXiv preprint arXiv:1610.02414.

IMAGE CREDITS

All drawings and images by the authors.

Wenzhe Peng is a Master's student at the MIT Computation Group. He has a Bachelor of Architecture degree from Southeast University, Nanjing, and a Master of Architecture degree from UC Berkeley. His interest lies in the intersection of architectural design, artificial intelligence and data visualization.

Fan Zhang is a Ph.D. Candidate in Geo-Information Science at The Chinese University of Hong Kong (CUHK), advised by Prof. Hui Lin. He got an MSc in Earth System Science from CUHK in November 2013 and BSc in Electronic Engineering from Beijing Normal University, Zhuhai in June 2012. His research interest lies in Virtual Geographic Environments, big spatial data mining, and deep learning.

Takehiko Nagakura is Associate Professor and director of the Computation Group at MIT. He teaches courses related to computer-aided design, and his research focuses on the representation and computation of architectural space and formal design knowledge.