Development of Dataset Items for Commercial Space Design Applying AI

Jung Hwa SEO[1], Segeun CHUN[2], Ki-Pyeong, KIM[3]

Abstract

In this paper, the purpose is to create a standard of AI training dataset type for commercial space design. As the market size of the field of space design continues to increase and the time spent increases indoors after COVID-19, interest in space is expanding throughout society. In addition, more and more consumers are getting used to the digital environment. Therefore, If you identify trends and preemptively propose the atmosphere and specifications that customers require quickly and easily, you can increase customer trust and conduct effective sales. As for the data set type, commercial districts were divided into a total of 8 categories, and images that could be processed were derived by refining 4,009,30MB JPG format images collected through web crawling. Then, by performing bounding and labeling operations, we developed a 'Dataset for AI Training' of 3,356 commercial space image data in CSV format with a size of 2.08MB. Through this study, elements of spatial images such as place type, space classification, and furniture can be extracted and used when developing AI algorithms, and it is expected that images requested by clients can be easily and quickly collected through spatial image input information.

Keywords : Dataset, Commercial Space, Design, AI.

Major Classification Code: Mathematical Methods, Information, Knowledge, Uncertainty

1. Introduction

The size of the market in the field of spatial design continues to increase regardless of type, and the importance is increasing in proportion (Lee, et al., 2019; Yonhapnews, 2021). In specific, as the time spent indoors

1 First Author. Adjunct lecturer, Department of Spatial Design, Seoul Institute of the Arts, Korea. Email: saeam111@naver.com
2 Second Author, Professor, Department of Spatial Design, Seoul Institute of the Arts, Korea. Email: junechun@seoulart.ac.kr
3 Corresponding Author, Professor, Department of Logistics and Trade, Daejeon University, Korea. Email: gpkim2970@daum.net

increases after COVID-19, interest in space in society as a whole is increasing, while the amount of consumption for space improvement continues to rise((Lee, et al., 2020). According to the National Statistical Office and Samsung Fashion Research Institute, the size of the domestic lifestyle (home furnishing) market has grown from 8 trillion won in 2010 to 13.1 trillion won in 2016, nearly doubling in six years, and reaching 18 trillion won by 2023. was estimated to be in the billions of dollars.

In the private and public sectors, there is a growing interest in improving and supporting sales space (Lee, 2020). As more consumers are familiar with the digital environment, distribution channels are diversified according to online-based consumption patterns, and the role and change of commercial space are intensifying as non-face-to-face becomes common (Hankyung, 2022; Kim,

2022; Hwang & jiyoung, 2020; Sudiatmika, & Ariantini, 2021).

For this reason, it is very necessary for space designers to collect consumer requirements and deliver concepts and images according to the field composition and ratio when delivering design intentions, and space designers are using various technology strategies to plan smooth communication and efficient design in accordance with the rapidly changing consumption environment (He, 2019).

It takes a lot of time to collect design image data necessary for this process, and securing a large amount of spatial design references enables various suggestions according to the requirements of the customer, which is directly related to an increase in sales success rate(Kim, CHOI, Song, & Moon, 2022). If crawler is activated, anyone can easily and quickly obtain the latest information (Kim & Kim, 2021; Hong, 2020), and if a space design trend is easily and quickly identified and preemptive suggestions are possible based on the secured image, it is expected that sales will proceed effectively by increasing customer trust(Batini et al., 2009).

In this study, images are collected and processed to establish standards for AI learning datasets for commercial space design. Through this, we intend to develop a web crawler that automatically crawls the desired image by entering the desired conditions such as website address, space shape, design concept, and color.
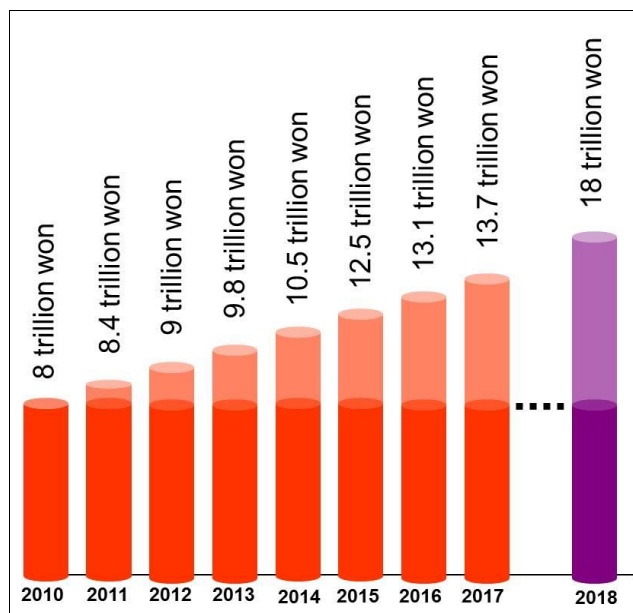
Source: Statistics Korea(2019), Samsung Fashion Research Institute (KMAC, 2019. 2)

Figure 1: Growth trend of the domestic lifestyle (home furnishing) interior market

2. Dataset Generation Procedure

First, an overall plan to secure, process, and analyze data was established, and criteria for classification were established. Second, commercial space design was analyzed and classified into 8 types of businesses. Thirdly, image data were collected by dividing them into seven subdivided columns. Fourth, we organized our own data and collected additional data such as public data through web crawling to perform bounding and labeling tasks.

Results data were derived through data processing such as additional types of concreteness and data purification. Based on this, the final result data set was completed through a procedure to verify, modify, and supplement the derived data((Pipino, 2002).
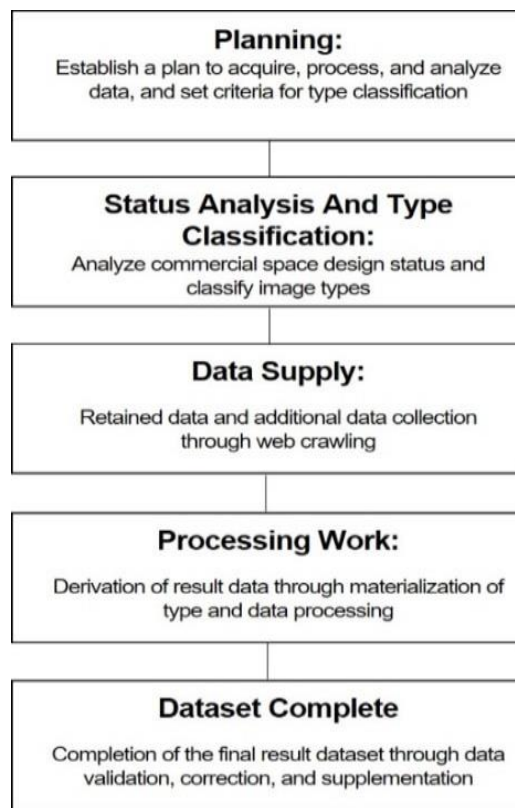

Figure 2: Process of Dataset Development

3. Data Acquisition Process

In this study, before securing data, a plan to secure, process, and analyze data was established in consultation with a company specializing in data processing, and a plan for the processing process, such as specifying and selecting

columns, was also established (Kim, et al., 2020). Afterwards, through the analysis of previous studies, it was confirmed that the main industries that small businesses engage in are catering, sales, and lodging, and the commercial space design was analyzed by focusing on the major industries.

Table 1: Place type of commercial space

classification	Sectors	classification	Sectors
Catering	Restaurant	Store	Mart
	cafe		Supermarket
	Bakery Cafe		
	Food Court		Convenience Store
	Pub		

Afterwards, the restaurant industry was classified into 5 categories, namely restaurants, cafes, bakery cafes, food courts, and taverns, and the retail stores were classified into 3 categories, namely marts, supermarkets, and convenience stores. The final 8 types of places designated in this way were divided into 5 letter-type columns, respectively, into space form, design concept, interior material, color, and industry. In addition, the image data was collected and organized based on the criteria of classifying the bounding position and bounding size into two numeric type columns. In this process, the necessary data was reinforced, and one 'accessibility' column was added for efficient offline sales, and it was collected and organized into a total of eight columns.

Table 2: How to use spatial image search AI

Input information for spatial imagery		Input result
Space form	Living room	
Design concept	Nature	
Interior material	Glass, iron, marble	
Color	Black, white	
Sectors	Accommodation	

Image data collected through web crawling other than self-held images were excluded if there were no target objects or were out of focus, and the final processable images were derived through pre-quality inspection such as image quality and format, and bounding and labeling of location, space, and furniture types.

4. Results Deployment

4.1. Collected data (image)

The collected data is commercial space image crawling data as of 2022. It includes 3,356 indoor images of commercial spaces, and the image file name is written in English and the type of place is described. The size of the data is 4,009.30MB, and the file format is jpg. The renewal cycle of data is once, and the method of provision was downloaded from the 'Data Voucher Business Management System'.
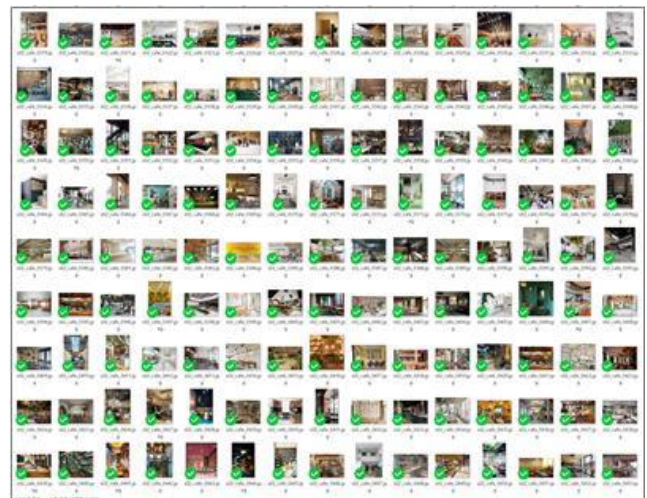

Figure 3: Collected Data

4.2. Result data (dataset)

The range of result data is a dataset of 15,5000 annotation results, including data file ID, figure shape, annotation details, and class. This is the actual result data of 'Establishment of an AI learning dataset using spatial images' and consists of a total of 8 columns: 'space shape, design concept, interior material, color, industry, bounding location and size, and accessibility'. The size of the data is 2.08MB, and the file format is CSV. As with the collected data, the renewal cycle of the data is once, and the method of provision was downloaded from the 'Data Voucher Business Management System'.

Image data collected through web crawling, in addition to images owned by the company, were excluded from images that do not have target objects or are out of focus. went through Afterwards, work rules were established by defining classes and items, and an AI learning dataset was built by labeling the results of annotation work. As a result, after deriving a final processable image through preliminary quality inspection such as image quality and format, the type of place, space classification, and furniture type were bound and labeled.

As a result, it is in the planning stage to develop AI algorithms and select learning models and to modularize them, and algorithm development and web crawler planning using datasets have been visualized.

Table 3: Outcome Data (Dataset)

| No. | Data file name | | Data item | | | | | | | | |
	Data file name (Korean)	Data file name (English)	Item name (Korean)	Item name (English)	Data type	length	Primary key	NULL	Include personal information	Including affiliate information	Code group value
2	결과데이터_서정화	result_data_sjh_csv	이미지 파일 아이디	fileid	Text	20	N	Y	N	N	—
2	결과데이터_서정화	result_data_sjh_csv	도형 형태	shape_name	Text	10	N	Y	N	N	—
2	결과데이터_서정화	result_data_sjh_csv	어노테이션 객체 아이디	region_id	Text	20	N	Y	N	N	—
2	결과데이터_서정화	result_data_sjh_csv	폴리곤의 x 좌표	all point_x	Text	—	N	Y	N	N	—
2	결과데이터_서정화	result_data_sjh_csv	폴리곤의y 좌표	all point_y	Text	—	N	Y	N	N	—
2	결과데이터_서정화	result_data_sjh_csv	rect의 좌측x좌표	x	number	—	N	Y	N	N	—
2	결과데이터_서정화	result_data_sjh_csv	rect의 좌측x좌표	y	number	—	N	Y	N	N	—
2	결과데이터_서정화	result_data_sjh_csv	rect의 가로길이	width	number	—	N	Y	N	N	—
2	결과데이터_서정화	result_data_sjh_csv	rect의 세로길이	height	number	—	N	Y	N	N	—
2	결과데이터_서정화	result_data_sjh_csv	클레스	class	Text	100	N	Y	N	N	—

5. Conclusion

Before data application and utilization, it took a lot of time to collect data by listening to the client's requirements and consulting with the company's internal personnel to collect data, and to refine image data, collect additional data, and provide it to the client.

Since a basic dataset has been established after data application and utilization, it is expected that the image required by the client can be collected quickly and easily through spatial image input information. We then expect that it will be easy to find similar types of images even if the desired image is not finally implemented on the dataset we have(Lee, et al., 2022). In addition, it is expected that new customers who need space design can be introduced in

addition to customers who have requested space design consulting from existing companies through external promotion after establishing the service, and sales can be increased by securing customers through various business channels.

Since spatial design is three-dimensional, it has temporal and spatial constraints and diversity. Therefore, since information such as continuous trend changes and new material development is vast, it is difficult to accommodate all information with the current amount of datasets. Thus, more data and images should be reinforced to increase smooth crawling and accuracy. However, there is a significant meaning that consumers wanted to provide customers and sales and sales points that designers can easily present the foundation of web chromatography.

In future studies, we intend to develop an AI algorithm with this dataset. In addition, it is necessary to establish a dataset to design other industries or residential spaces. It is expected that this will expand the scope of customers by pioneering sales in all fields such as individuals, public, and companies, and will not only secure continuous sales but also help designers in the same industry(Kim et al., 2020).

References

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM computing surveys (CSUR), 41*(3), 1-52.

Hankyung (2022). It's the last minute from the login ... Older people are 'non-face-to-face stress'. Retrieved Auguest 30, 2022, from https://www.hankyung.com/society/article/2021031413121

Kim, J., Woo, H., Yang, H., Kim, M., Kim, S., Lee, S., Kim, B., Kim, Y., Kwak, J., Choi, H., Jung, I., Lee, Y., & Lee, W. (2020). Information and standard curriculum proposals for the 2022 amendments. *The Journal of Korean Association of Computer Education, 23*(1), 1-28. DOI:10.32431/kace.2020.23.1.001

Kim, J. Y. (2022). A Study on the Use of Technology as a Spatial Design Marketing Strategy. *Journal of Korean Space Design Society*, *17*(1), 165-178.

Kim, S. E., MUN J. H., KIM, K. S., & KANG, M. S. (2020). A Development of Healthcare Monitoring System Based on Internet of Things Effective, *KJAI. 8*(1), 1-6.

Kim, S. K., & Kim, T. Y. (2020). *Study on Educational Standard Dataset Requirements for Artificial Intelligence Education.* Dept. of Computer Education, *Korea National University of Education, 25*(2), 43-46.

Kim S. W., CHOI D. J., Song Y. M., & Moon I. Y. (2022). Design and Implementation of Voice-based Interactive Service KIOSK. *Journal of Practical Engineering Education,14*(1), 99-108.

KMAC (2019. 2). *Sell value, electric style rather than goods or services!* Retrieved October 20, 2022, from https://m.blog.naver.com/kmacforever/221479774108

Lee, M. N., Park, H. K., Oh, C. O., Baik, J. W., & Kim, S. T. (2019). A Meta Analysis on Domestic Research Trends of Healthcare Design in Spatial Design Field. *KISD*, *14*(7), 37-50. Doi: www.kisd.or.kr/10.35216/kisd.2019.14.7.37

HWANG, J. Y. (2020). Re: Store. Seoul: Influential.

Lee, J. Y., Shin, S. J., Yoon, S. H., & Kim, T. Y. (2022). Survey on Artificial Intelligence & Machine Learning Models and Datasets for Network Intelligence. *The Journal of Korean Institute of Communications and Information Sciences, 47*(4), 625-643. https://doi.org/10.7840/kics.2022.47.4.625

HE, Y. G. (2019). The Effect of Artificial Intelligence on Economic Growth: Evidence from Cross-Province Panel Data. *Korean Journal of Artificial Intelligence, 7*(2), pp. 9-12.

HONG, J. Y., & KIM, Y. S. (2020). Development of an AI data science education program to foster data literacy in elementary school students. *Journal of Information Education Society, 24*(6), 633-641.

Lee, W. R, CHOI, J.W., CHOI, J. H., & KANG M. S. (2022). Research related to the development of an age-friendly convergence system using AI. *Korean Journal of Artificial Intelligence, 10*(2), 1-6.

Pipino, L. L, Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM, 45*(4), 211-218.

Son, W. (2020). Development of SW education class plan using artificial intelligence education platform : focusing on upper grade of elementary school. *Journal of The Korean Association of Information Education. 24*(5), 453-462. DOI:10.14352/jkaie.2020.24.5.453

Statistics Korea (2019). Retrieved October 15, 2022, from https://kostat.go.kr/unifSearch/search.es

Sudiatmika, I., & Ariantini, M. S. (2021). Digital Marketing Chatbot Using API Dialog Flow Case Studi ITB Stikom Bali, Jimbaran Campus. Retrieved October 20, 2022, from https://www.researchgate.net/scientific-contributions/Made-Suci-Ariantini-2167036810

Yonhapnews (2021). The digital signage market grows… Growth 23% this year over Corona. Retrieved October 27, 2022, from https://www.yna.co.kr/view/AKR20210511013200003