ARCHISEARCH: A TEXT-IMAGE INTEGRATED CASE-BASED SEARCH ENGINE IN ARCHITECTURE USING SELF-ORGANIZING MAPS

CHENYI CAI[1*], XIAO WANG[2], BIAO LI[3], and PIETER HERTHOGS[4]
[1,4]*Singapore-ETH Centre, Future Cities Lab Global Programme, CREATE campus, 1 Create Way, #06-01 CREATE Tower, Singapore 138602.*
[2,3]*School of Architecture, Southeast University, Nanjing, China, 210096.*
[1]*chenyi.cai@sec.ethz.ch, 0000-0002-6249-1816*
[2]*wangxiao_seu@seu.edu.cn, 0000-0001-9471-5370*
[3]*jz_studio@seu.edu.cn*
[4] *pieter.herthogs@sec.ethz.ch, 0000-0003-2562-8374*
[*] *corresponding author*

Abstract. Case-based study and reasoning are considered fundamental meta-practices of architectural design. There are many online platforms to share architectural projects, which serve as data sources for case studies. However, search and retrieval capabilities offered by such platforms often do not cater to the professional needs of architects and do not integrate or synthesize the semantics present in text and images. We propose a systematic methodology to develop a search engine for architectural and urban projects using text and image data from one such online project-sharing platform: Chinese platform Gooood. Our approach automatically collects and extracts features from data, and integrates figurative and descriptive retrieving methods using word2vec, deep learning, and clustering algorithms. Our methodology provides a flexible approach for developing case-based search engines for architectural projects that take into consideration both images and texts and could be applied to any (Chinese language) platform. It offers architects augmented case-based workflows, which enrich design inspiration and accelerate decision-making meta-practices by unlocking the semantic search and retrieval of existing projects in novel ways.

Keywords. Case-based Design, Text-image Semantic Search, Case Representation, Self-organizing Map

1. Introduction

Case-based studies are considered fundamental meta-practices in architectural design (Ching, 1999). Learning and reasoning from similar cases in (dis)similar contexts is important for synthesis and decision-making in architectural design. Case-based

ACCELERATED DESIGN, Proceedings of the 29th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2024, Volume 3, 19-28.

reasoning (CBR) and one of its applications, case-based design (CBD), accelerate the creation of design concepts by retrieving and adapting evidence from existing cases to address new problems (Kolodner, 1992; Maher et al., 2014). The ongoing accumulation of architectural cases and advancement of computer aided design technology underscores the significance of developing suitable digital CBD tools for architecture and urban design.

Currently, many online platforms that showcase architectural and urban projects play a crucial role in browsing architectural cases. Regarding case retrieval, such platforms typically offer keyword-based semantic filter functions. Images and text are arguably the most frequently featured media on these platforms, and the meanings conveyed by these two different media complement each other. The search options commonly provided by online platforms lack semantic flexibility, image-semantic links, and context-based search. Firstly, there is a lack of semantic flexibility. The menu-based semantic search often relies on limited pre-defined categories and binary cut-off filters, limiting its suitability for CBD. Searching content by semantic proximity is often not supported, with search relying on exact word matching within titles or pre-determined tags. Secondly, image content is often not searchable, and image content is not linked to image semantics. Thirdly, context-based search is often unavailable, meaning users cannot parse queries according to the context (the database) and rank the searched results by relevance. In addition, existing platforms often do not allow interlinked search between different content formats (e.g., text + image). Therefore, we argue the search capabilities of existing online design project platforms do not cater to the professional needs of case-based search and case retrieval.

Hence, in this paper we introduce a systematic methodology to develop flexible and context-based search applications to offer architects augmented case and example-based workflows to enrich design inspiration and accelerate decision-making meta-practices. In this paper, we demonstrate our methodology on one such Chinese platform, Gooood (gooood.cn). In section 3, we introduce our methodology to create case-based search engines, which builds on figurative and descriptive retrieving methods. Our methodology automates data collection from gooood.cn. It extracts image features using deep learning. It adds text semantics using low-cost feature extraction for Chinese language texts using Markov Chains and TF-IDF. Our methodology also enables figurative and descriptive retrieving methods using two self-organizing maps (SOMs). Section 4 presents the search options and the results by our search engine, which allows rich and flexible search by finding text synonyms and interlinking two types of media. Our application provides architects and designers with a navigable explorer space consisting of the built cases and present semantics in texts and images, and hence supports design condition synthesis and design decision-making and facilitates interactions between humans and machines.

2. Background

During the 1990s, Case-Based Design (CBD) - the application of case-based reasoning (CBR) in design - emerged as a popular computer-aided design approach for creating intelligent design assistance (Kolodner, 1992). CBD systems identify novel design solutions by drawing upon analogous experiences from the past (Domeshek & Kolodner, 1992). The significance of digital tools in CBD has been underscored by the ongoing accumulation of architectural cases, the evolutions of technology, and the suitability of the approach in architecture and urban design.

Web-based case-sharing platforms for architectural projects have become primary sources for case studies. Mainstream online platforms, such as ArchDaily (archdaily.com), Dezeen, or Gooood (gooood. cn), archive and index a wealth of architectural project documents, including floor plans, photographs, and project descriptions. ArchDaily and Gooood are architecture and design platforms that offer content in both English and Chinese language. The Boolean-based and keyword-based search options are commonly found on mainstream architecture web-platforms. As the provided categories are pre-defined and limited and the search options are binary cut-off filters, such case search approaches lack flexibility.

CBD applications have been widely adopted in computer-aided architectural design (CAAD) since the 1990s. From 1990 to 2000, Archie-II, CADRE, and SEED were considered the popular CBD applications for architecture (Heylighen & Neuckermans, 2001). Public web-based architectural case search tools can be traced back to DYNAMO, which was developed as an online design education system for architecture students and designers (Heylighen & Neuckermans, 2000). Archie-II was one of the earliest prototypes for case-based design. By interlinking blueprints and annotated text stories, it enabled the search for design cases by inputting user stories (Domeshek & Kolodner, 1992). Since then, there have been continuous research efforts to develop CBD and decision support applications, applied in architecture design and construction fields (Hu et al., 2016; Maher et al., 2014).

The case representation in CBR approaches helps users eliminate the complicated details of raw data, enabling them to concentrate on valuable, and potentially useful knowledge. The computed numbers carry distinct information depending on different representation methods. Therefore, the crucial question during digital tool development lies in determining how to encode and decode the case information carried by multimodal data. Extracting semantics and features from various forms of architectural cases is the fundamental process for the subsequent context-based case retrieval. In feature learning of images, high-dimensional feature vectors are extracted from image contents through learning from the context (Bengio et al., 2013). For text, natural language processing methods such as TF-IDF, word embedding, and word2Vec are used for translating text to feature vectors, primarily for the English language. The Markov Chain method is often used in language models to predict the next character. These feature extraction methods are context-based and highly dependent on the dataset, thus have the potential to support context-based search.

Multimodal learning methods linking images and text are being developed (Zhang et al., 2017), but hardly for developing tools for case-based search. Such methods are crucial to enrich case retrieval. The self-organizing map method contains topologically trained grids, hence allowing both clustering and indexing during the training process (Kohonen, 2012). The flexible nature of the algorithm enables it to have a diverse set of applications, including case-based search according to multiple data sources (Cai et al., 2021), segments of floorplans (Marinčić, 2017), and hybrid urban spaces (Tomarchio et al., 2023). Therefore, SOM has the potential to interlink multimodal data for integrated and context-based search.

3. Methodology

This section introduces our methodology for developing a context-based search engine for architectural cases. It is a systematic pipeline for collecting, analysing, and retrieving architectural projects, resulting in a multimodal (text and image) search engine. We set out the implementations based on the following scopes and aims: 1) The architecture and scripts are intended to be useable for any online architectural project platform; 2) In order to offer a flexible platform for architects to address multiple types of design tasks, our current implementation focuses on context-based feature extraction methods to be able to deal with multi-modal data; 3) Our methodology aims to be computationally fast in comparison. 4) We set out to enable cross-search of multimodal contents (by linking multiple self-organizing maps); 5) In order to offer comprehensive evidence of architectural cases to support CBD, we aim to integrate multi-contents (text and image in this case) and result in any input contents being searchable.

3.1. DATA COLLECTION AND PREPROCESSING

We developed scripts to automatically open a webpage (for this paper, gooood.cn) and parse its HTML to download corresponding elements using JavaScript. We collected a dataset of 8182 project articles featuring 203,350 images. Since we focused on project descriptions in Chinese, we filtered out the advertisements and English texts. We also omitted the punctuation from the articles for the subsequent feature extraction.

3.2. FEATURE EXTRACTION FOR IMAGES

We used a pre-trained deep convolutional neural network (DCNN) "Wide ResNet-50-2 Trained on ImageNet Competition Data" for image feature extraction using the Wolfram Mathematica platform. We extracted the 22nd layer output of the neural network, which translates each image to 2048-dimensional feature vectors. The DCNN structure is shown in Figure 1 (left). The feature vector values varying between 0-1. We take two images and visualize the feature vector by transferring values to a black to white colour gradient (Figure 1).

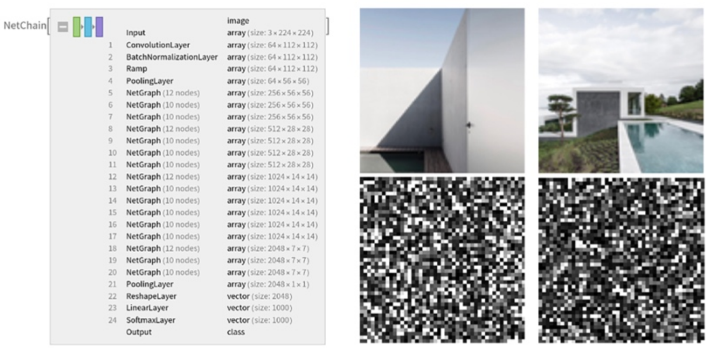

Figure 1. The DCNN structure (left) and the colorized visualization of image feature vectors (right)

3.3. FEATURE EXTRACTION FOR TEXTS

We used the Markov chain method to extract keywords of Chinese texts and the term frequency–inverse document frequency (TF-IDF) method for text feature extraction.

The methods are suitable for processing Chinese because the algorithms are consistent with the combinatorial nature of the Chinese language. Chinese words are formulated by combining Chinese characters, mostly in 2 characters, sometimes in 3 characters, and very few in 4 characters. Therefore, if we see common words as characters that are combined with high possibility, we can use the Markov Chain to extract the frequently combined characters, making it computationally fast in comparison.

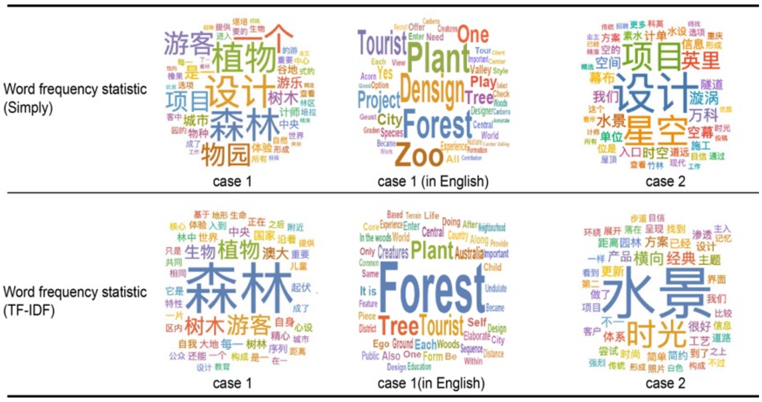

Figure 2. Tow showcases of the extracted keyword cloud from an article using Markov Chain plus word frequency statistic, and Markov Chain plus TF-IDF

We sliced the text from the whole library into sequences, and used the 1868 top frequent 2-character sequences and 180 top frequent 3-character sequences, resulting in a per word feature vector that is 2048 dimensional. However, a highly frequent word is not necessarily a keyword for a particular article. Therefore, we use the TF-IDF method to extract the actual keywords that are high-frequency and special. This feature extraction method extracts special keywords of one article in the context of the whole library. For instance, in Figure 2, with simple word frequency statistics, the word with the highest frequency is 'Design' which should not be the project's keyword because this word has also the highest frequency in other projects. After TF-IDF, the word with the highest value is 'Forest', which is an important keyword for this project (Figure 2).

3.4. SOMS TRAINING

We trained both a text SOM and an image SOM with the extracted features, to be used for clustering and indexing projects according to their figurative and descriptive characteristics. SOMs are a general-purpose nonlinear data transformation method that offer solutions for data clustering and 2D dimension reduction. During the training process, an index of a cluster can be assigned for each sample. The index is also called the best-matching unit (BMU) (Cai et al., 2022). Our SOMs are 30x30 units in size. We trained 10 iterations for the image SOM and 12 iterations for the text SOM, both with a batch size of 5000 samples.

After training, the weight (a high-dimensional vector) of each unit of the SOM is optimized according to the training data. In Figure 3, we visualized the trained SOMs and some selected clusters. We see a gradient of the images from different clusters.

We also see content clusters according to the word clouds, such as clusters with main topics such as 'Hotel' ('酒店'), 'Restaurant' ('餐厅') and so on.

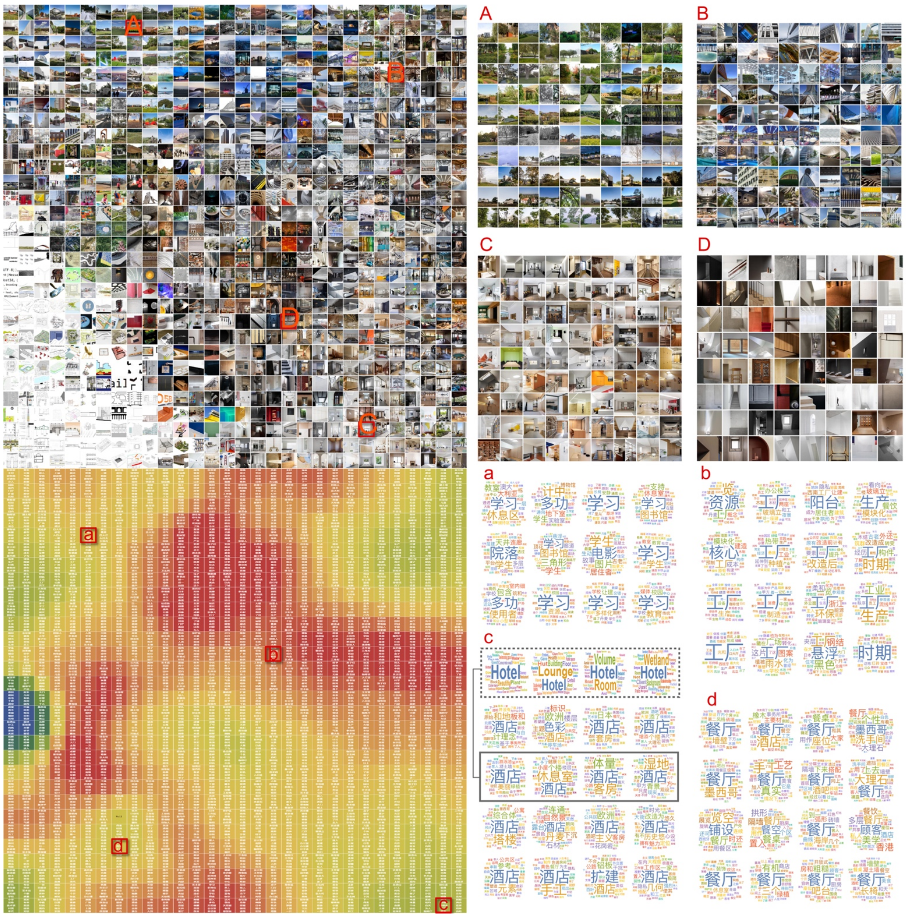

Figure 3. The trained results of image SOM (top left) visualized with one sample in each unit and demonstrations of four clusters visualized with 100 samples (top right), text SOM (bottom left) visualized by colouring according to unit weights and demonstrations of four clusters visualized with 12-16 sample word clouds.

4. Results

We show the search options provided by our search engine from three aspects: 1) inputting images as queries, to search for semantics and similar images; 2) inputting long phrase queries, to search for similar semantics and projects with similar word clouds; 3) inputting text and image integrated packages, to search for projects that are semantically similar in terms of images, word clouds, or both.

4.1. IMAGE AS QUERY

Our application extracts high-dimensional feature vectors of the input image using the DCNN and then assigns the image with a Best Matching Unit (BMU) index according to clustering. During the retrieval process, one can search for similar images, with similarity ranked according to the Euclidean distance between the feature vectors of the input image with those in the sample library. Alternatively, one can search for a defined number of images that are in the same BMU. Figure 4 demonstrates that our search engine performs well at providing images of similar cases. Based on the retrieved similar images, we can link the input image with possible semantics. For instance, we see the retrieved keyword sets: green plants; residence; garden, amusement park; game; forest; heritage, and so on (Figure 4).

Figure 4. The results of a search by inputting an image (top left), a search for possible semantics (top right), a search for the most similar 144 images according to Euclidean distance by feature vectors.

For a 300-image query, our search engine took 8 seconds on the Wolfram Mathematica platform, with 1.5 seconds being computation time and the rest image visualization time. Hence, the results demonstrate the efficiency of our workflow, even for a proof of concept using Wolfram Mathematica.

4.2. LONG PHRASE AS QUERY

Our application allows search by long-phrase-query. It extracts the features and semantics of the input text in the same way the platform data was processed. We can locate the BMUs from the text SOM and then retrieve projects and images associated with the BMUs.

In Figure 5, we showcase a query of 'flat renovation regeneration old residential area' (literal translation of the input), for which we only get 3 search results on Gooood. With our application, we retrieved a wealth of projects by matching them with word clouds created from both titles and article content. From the retrieved cases, we also identified renovation projects where keywords related to ‘renovation’ are not explicitly mentioned in the titles or predefined tags but are found in the contents: for instance, the project 'Base Zhangjiang, Shanghai' (the box with green border in Figure 5).

Our search engine gives projects and provides the URLs, images and keywords. Hence, it provides richer possibilities for inspiration by linking images based on the text query and presenting other possibly relevant topics with the query semantics.

Figure 5. The results of a search by inputting a long phrase (top left), BMUs in the text SOM with similar topics (top right), a search for projects with the topics mentioned in the query (showcasing with 8 projects), and a result of 3 search results from gooood (in the box with grey border).

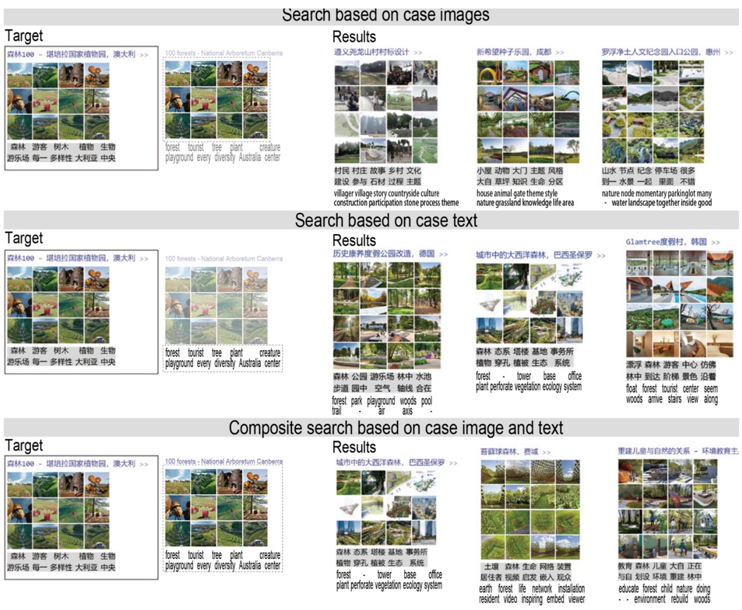

Figure 6. The results of a search by inputting projects (target), searching for projects with similar images (top row), with similar word clouds (middle row), and with a pre-set similarity blend number: 50/50% (bottom row).

4.3. TEXT-IMAGE INTEGRATED QUERY

Regarding the queries with both text (word cloud) and an image set, our application allows searching for projects that are similar in word clouds, similar in images, or a number of pre-set similarity blends (e.g., 50/50% image and text similarity). The result in Figure 6 shows one such query.

5. Discussion:

Our methodology improves CBD in case representation and retrieval processes. The efficient and professional case-based search engine has the potential to enrich both design inspiration and decision-making meta-practices, hence it can improve CBD.

We discuss our contributions in four aspects. 1) Our case representation methods are low-cost and general. The feature extraction methods of DCNN, MC, and TF-DIF can be applied to any image and Chinese text without the need for changing representation methods when the database changes. 2) Our methodology allows for the capacity of context-based search based on the specific database without pre-defining labels or categories. Hence, it can avoid the biases caused by the mismatch of pre-defined topics and the covered topics from the actual database. 3) Our methodology allows efficient case retrieval within a large number of cases and informs users with comprehensive case information. It provides more relevant projects that are ranked with similarity, while only three search results are provided by gooood.cn. 4) Our methodology allows flexible query to enable user-machine interaction and provide adequate information for improved CBD. Users can input text, images, and blends of both as a query and search for cases in any of those formats. Through searching according to semantic proximity and the word clouds, users not only get (dis)similar cases but also get other potentially relevant topics to inspire CBD. Hence it supports CBD by providing architects with rich and relevant cases through multimodal case-based search and it has promising potentials for subsequent evidence learning during further design tasks or scientific studies.

6. Conclusion

Our methodology for developing search engines for project platforms enables an automatic workflow for case-based search from data collection to cross-search for CBD by allowing context-based and text-image integrated search. We have enriched the case retrieval by searching according to the word clouds extracted from the whole project article, and by allowing long phrasal input in Chinese. The demonstrated search engine enables bidirectional and integrated search of text and images by interlinking two SOMs. The resulting search applications can enrich design inspirations and decision-making through the enhanced workflow of case-based reasoning and design.

Future work can extend our demonstration to other popular architectural platforms. We can include other advanced text and image analysis methods to build up more detailed connections of architectural objects for more accurate case retrieval. We could also test the search engine in specific design tasks with professionals to validate its contributions in CBD. The workflow can be repeated in other context beyond architectural platforms.

Acknowledgements

Part of this research was conducted at the Future Cities Lab Global at Singapore-ETH Centre. Future Cities Lab Global is supported and funded by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme and ETH Zurich (ETHZ).

References

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798–1828. https://doi.org/10.1109/TPAMI.2013.50

Cai, C., Guo, Z., Zhang, B., Wang, X., Li, B., & Tang, P. (2021). Urban Morphological Feature Extraction and Multi-Dimensional Similarity Analysis Based on Deep Learning Approaches. Sustainability, 13(12), Article 12. https://doi.org/10.3390/su13126859

Cai, C., Zaghloul, M., & Li, B. (2022). Data Clustering in Urban Computational Modeling by Integrated Geometry and Imagery Features for Probabilistic Navigation. Applied Sciences, 12(24), Article 24. https://doi.org/10.3390/app122412704

Ching, C. C. (1999). It's not just programming: Reflection and the nature of experience in learning through design. https://repository.isls.org//handle/1/4304

Domeshek, E. A., & Kolodner, J. L. (1992). A Case-Based Design Aid for Architecture. In J. S. Gero & F. Sudweeks (Eds.), Artificial Intelligence in Design '92 (pp. 497–516). Springer Netherlands. https://doi.org/10.1007/978-94-011-2787-5_25

Heylighen, A., & Neuckermans, H. (2000). DYNAMO: A Dynamic Architectural Memory On-line. Journal of Educational Technology & Society, 3(2), 86–95.

Heylighen, A., & Neuckermans, H. (2001). A case base of Case-Based Design tools for architecture. Computer-Aided Design, 33(14), 1111–1122. https://doi.org/10.1016/S0010-4485(01)00055-0

Hu, X., Xia, B., Skitmore, M., & Chen, Q. (2016). The application of case-based reasoning in construction management research: An overview. Automation in Construction, 72, 65–74. https://doi.org/10.1016/j.autcon.2016.08.023

Kohonen, T. (2012). Self-Organizing Maps. Springer Science & Business Media.

Kolodner, J. L. (1992). An introduction to case-based reasoning. Artificial Intelligence Review, 6(1), 3–34. https://doi.org/10.1007/BF00155578

Maher, M. L., Balachandran, M. B., & Zhang, D. M. (2014). Case-Based Reasoning in Design. Psychology Press.

Marinčić, N. (2017). Towards Communication in CAAD: Spectral Characterisation and Modelling with Conjugate Symbolic Domains [Doctoral Thesis, ETH Zurich]. https://doi.org/10.3929/ethz-b-000216502

Tomarchio, L., Herthogs, P., & Tunçer, B. (2023). Hybrid art space typologies in Singapore from social media data. City, Culture and Society, 34, 100519. https://doi.org/10.1016/j.ccs.2023.100519

Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D. N. (2017). StackGAN: Text to Photo-Realistic Image Synthesis With Stacked Generative Adversarial Networks. 5907–5915. https://openaccess.thecvf.com/content_iccv_2017/html/Zhang_StackGAN_Text_to_ICCV_2017_paper.html