DEEP SPATIAL MEMORY

Quantifying Architectural Spatial Experiences through Agent-driven Simulations and Deep Learning

SHUHAN MIAO[1], WENZHE PENG[2], DANIEL TSAI[3] and TAKEHIKO NAGAKURA[4]
[1]*Harvard Graduate School of Design*
[2,3,4]*Massachusetts Institute of Technology*
[1]*shuhan_miao@gsd.harvard.edu, 0009-0006-4522-6814*
[2]*pwz@alum.mit.edu, 0009-0008-0129-7578*
[3]*dtsai@mit.edu, 0000-0002-8082-6392*
[4]*takehiko@mit.edu, 0000-0002-3219-3930*

Abstract. In architectural theory, the spatial experience is dynamic, evolving from sequences of interconnected views shaped by past encounters and future expectations. Traditional computational methods such as Isovists provide geometric insights but fall short in representing their sequential nature. To address this gap, the paper introduces a novel methodology that combines agent-driven simulation, 3D Isovist sampling, and deep learning for quantitative analysis and comparison of spatial experiences in architecture. This approach leverages the Grasshopper plugin Pedsim for simulating pedestrian paths and a self-supervised video representation learning model MemDPC for processing depth panorama sequences and extracting numerical features for each sequence. The methodology is first validated through a controlled experiment with various sequence typologies, affirming its efficacy in recognizing typological similarities. A case study is conducted comparing Louis Kahn's designs with Roman architecture, quantitatively analysing their intertwined spatial experiences. This research offers a framework for quantitatively comparing spatial experiences across buildings and interpreting the nuanced impact of historical references on modern spaces.

Keywords. Deep Learning, Artificial Intelligence, Spatial Experience, Isovist, Agent-driven Simulation, Self-Supervised Learning

1. Introduction

In the field of architectural theory, the representation of spatial experience has been discussed extensively, particularly focusing on its dynamic and selective nature. Zevi views in *Architecture as Space* that architectural spatial experience as a time-imbued continuum, best captured through film rather than static images,

ACCELERATED DESIGN, Proceedings of the 29th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2024, Volume 1, 109-118.

to capture the full essence of navigating and interacting with spaces (Zevi, 1974, p. 59). Adding to this dynamic aspect is the selectivity of spatial experience. Ching suggests in *Architecture, Form, Space, and Order* that our experience of space is inherently selective, influenced by our navigation in the space (Ching, 1979, p. 280). This selectivity in perception means that we don't absorb all elements of a space equally; rather, our focus shifts as we navigate, leading to a subjective interpretation based on our position and movement (Hershberger, 1970, p. 43). Such subjectivity presents a challenge for architects and theorists who seek to understand and compare architectural spaces in a more objective lens. Traditional methods of quantitative representation—such as plans, sections, and static images—offer a fragmented view that misses the fluid continuum of experience as described above.

Addressing this gap, the research introduces a novel methodology for quantitatively representing and comparing architectural spatial experiences. It integrates agent-driven simulation, 3D Isovist techniques, and artificial neural networks for pedestrian path generation, spatial data capture, and feature extraction. The methodology is validated through two experiments. The first employs self-supervised learning for feature extraction from spatial sequence typologies, using unsupervised clustering to validate the features' self-clustered quality, then applies a supervised classifier to translate features into legible sequential types. The second experiment extends the methodology to a real-world application, examining the influence of Roman architecture on Louis Kahn's designs. The model, fine-tuned for complex architectural spaces, categorizes sequential types to initially understand typological similarities and differences. Subsequent in-depth feature analysis through unsupervised clustering and nearest neighbour methods uncovers latent patterns within same sequential type. This methodology converts subjective spatial experiences into quantifiable data, enabling an objective comparative analysis of historical and contemporary architectural elements.

2. Related Works

Machine learning involves computer systems that analyse and deduce patterns in data through algorithms and statistical models. This approach is particularly beneficial in architectural spatial analysis, facilitating the understanding and interpretation of complex spatial data. Machine learning includes categories such as supervised and unsupervised learning. Supervised learning trains models on labelled datasets for classifying data or predicting outcomes. In contrast, unsupervised learning explores unlabelled data to uncover its inherent structure (Deng & Yu, 2014). In architectural spatial analysis, both methods offer distinct benefits. The predefined labels of supervised learning enhance interpretability and specificity but lack generalizability across diverse architectural spaces. Unsupervised learning, which discerns patterns from unlabelled data, offers broader exploration capabilities. However, the lack of human-readable labels necessitates additional analysis for interpretation.

Isovist analysis is essential in architectural studies for quantifying spatial

environments. It maps visible points from a selected perspective, translating spatial geometry into perceptual experiences, capturing both phenomenological and morphological aspects of spaces (Benedikt, 1979). This approach has evolved over decades in fields such as analysing plan visibility and urban space quality (Dawes & Ostwald, 2014; Leduc et al., 2011; Turner et al., 2001). Recent years have seen a surge in the integration of Isovist methods with breakthroughs in deep learning. Peng et al. (2017) use 2D depth map images of 3D Isovists to train a deep convolutional neural network in a supervised learning manner to classify spatial typologies. Conversely, Johanes & Huang (2021) applied self-supervised learning on 2D Isovist data, using an unsupervised clustering algorithm to categorize spatial plans. Additionally, recent studies have merged 2D Isovist analysis with pedestrian trajectory simulations, analysing Isovist data along simulated paths using unsupervised clustering to uncover spatial properties (Feld et al., 2020; Sedlmeier & Feld, 2018). Building on these developments, the proposed experiments seek to integrate the strengths of both supervised and unsupervised learning methods, combining human-readable categorization of supervised learning with the flexible pattern discovery of unsupervised learning. This hybrid method is designed to create a comprehensive and adaptable framework for architectural analysis.

3. **Methods**

This study employs a comprehensive methodology combining agent-driven simulation (Section 3.1), 3D Isovist sampling (Section 3.2), and self-supervised representation learning (Section 3.3). It aims to capture the complexity of spatial experiences by simulating pedestrian behaviour, sampling spatial geometry through 3D Isovists, and extracting latent representations of these experiences using self-supervised learning techniques (Figure 1). This approach is designed to systematically encode and analyse spatial sequences in architecture.

3.1. AGENT-DRIVEN SIMULATION

To effectively capture the subjective nature of spatial sequences, the study utilizes PedSim, a Grasshopper plug-in that simulates pedestrian paths using the social force model and anticipatory collision avoidance (Riise, 2022). While actual pedestrian movement data could be sourced from real-world sensors, the controlled simulation ensures consistent input for analysis purpose. Agents in the simulation algorithm will move from designated start to end points, visiting various points of interest within their field of vision. For consistency and comparability across different spaces, each trajectory is established with a single start and end point, along with ten interest points strategically selected to represent architectural features such as columns and arches. The intentional inclusion of expert knowledge ensures a realistic simulation of pedestrian behaviour. Paths taken by these simulated agents are recorded and converted into 3D polylines at a standard eye level of 1.6 meters for 3D Isovist sampling.

By simulating various paths within a single trajectory, the approach reflects diverse individual experiences in a unified spatial sequence. This variety also serves as data augmentation, ensuring robustness in subsequent model training.

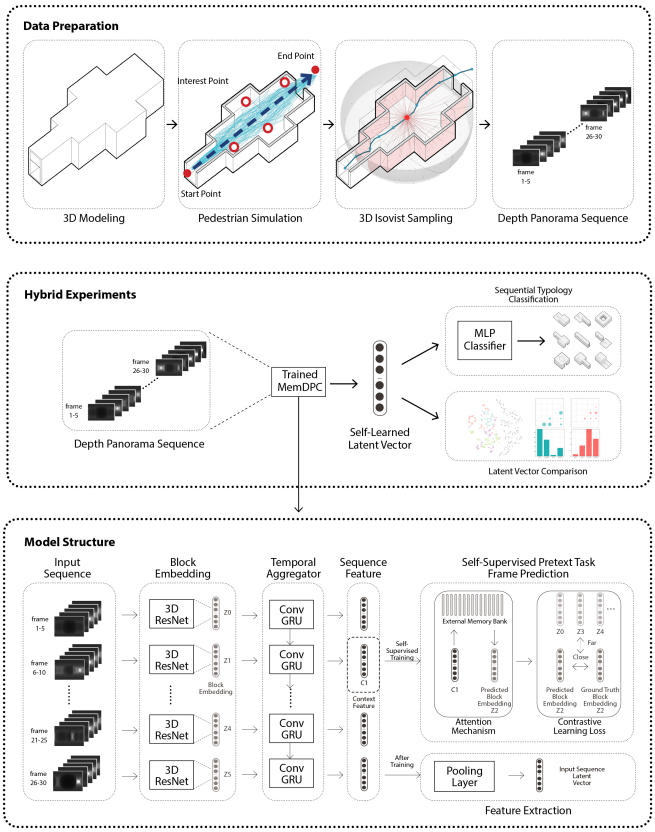

Figure 1. Illustration of Proposed Pipeline

3.2. 3D ISOVIST SAMPLING

This study's 3D Isovist sampling employs a custom Python-scripted component in Grasshopper to process simulated pedestrian paths. Each path is segmented into 30 equally spaced points. At each point, rays are projected onto a sphere's surface, and the distances at intersection points are computed and transformed into grayscale values. These values are then remapped into depth panorama images, conforming to the resolution parameters established by Peng et al. (2017). A consistent maximum sampling distance of 20 meters is set to accommodate the monumental scale of the buildings in the following case studies.

It is vital to recognize that while these depth map images effectively capture spatial geometrical boundaries facing rays, they represent a reduced version of the complete spatial experience. This study specifically focuses on spatial geometry information, presenting it as a simplified form of spatial experience. Therefore, the complexity of representing spatial experience is distilled into 30 sequential depth panorama images for each simulated path. These image sequences, formatted as videos, are compatible with existing self-supervised video representation learning models, enabling a more nuanced analysis of spatial perception.

3.3. SELF-SUPERVISED REPRESENTATION LEARNING

Self-supervised learning (SSL) is a subset of unsupervised learning. It trains models on large datasets without the need for predefined labels, thus overcoming the limitations of supervised learning (Jing & Tian, 2019). SSL employs 'pretext tasks', derived directly from unlabelled data, such as video reconstruction or frame order prediction. Solving such complex pretext tasks requires the model to have a high-level understanding of the training data to learn generalizable features (Schiappa et al., 2023). SSL in this study involves employing MemDPC, a model that uses frame prediction as its pretext task (Han et al., 2020). This model divides each video into blocks, encoding them into embeddings. These embeddings are time-aggregated using RNNs (Recurrent Neural Networks) to extract the context feature for each block. A global attention mechanism is initialized as a learnable memory bank for hypothesis formation and future block prediction. After training, an average pooling layer condenses the context features of each block into a 256-dimensional feature vector, representing the latent spatial experience of a simulated path. The memory bank approach conceptually resembles how humans accumulate spatial experiences. For humans, this includes visual impressions and understanding of spatial relationships. However, unlike human cognition, neural networks identify patterns without semantic understanding, necessitating further analysis. Subsequent experiments use sequential typologies as labels to translate these SSL-acquired features into human-readable terms and analyses these features to discern underlying similarities.

4. **Experiments**

This section details experiments that validate our framework for studying sequential spatial experiences, combining a supervised approach for identifying typological similarities and differences, and an unsupervised method for in-depth pattern discovery. The first experiment, utilizing the 'typology dataset', validates the model's capability in autonomously clustering features through self-supervised learning. A supervised classifier is then applied to these features to translate the 256-dimensional vector into nine human-readable sequential typologies. The second experiment, with the 'case study dataset', fine-tuned this model to adapt to the complexities of real-world architectural spaces. Here, the

classifier initially categorizes sequential typologies, facilitating a preliminary organization of trajectories. This is followed by a detailed feature analysis using similarity measures and clustering, designed to reveal deeper patterns and nuances within architectural spaces categorized under the same typology.

4.1. EXPERIMENT I: SEQUENTIAL TYPOLOGIES

The first experiment examines the SSL model's ability to effectively extract features, using a typology dataset of 2,700 path sequences with three space types: 'room', 'passage', and 'exterior' at both start and end, creating nine distinct combinations (Figure 2). These sequence data undergo augmentation during SSL training, including adjustments in brightness, playback speed, cropping, and horizontal flips. After training, each sequence is transformed into a 256-dimensional feature vector. The model's performance in feature extraction is initially assessed by its validation accuracies in the SSL pretext task of frame prediction, achieving 75% top-1 and 99% top-5 accuracies. These results indicate a robust feature learning capability without the need for labelled data. To further evaluate the learned features, unsupervised clustering is applied. Using the K-means algorithm (Lloyd & S., 1982) to cluster these 256-dimensional feature vectors and comparing the results against the ground truth of sequential typology labels yielded a 96% accuracy. Additionally, the clustering's relevance and precision are confirmed by an Adjusted Mutual Information score (Vinh et al., 2010) of 0.95, reflecting a high degree of correspondence between the unsupervised clustering outcomes and the actual labels. This demonstrates that the SSL-extracted features inherently possess a self-clustered organization that significantly correlates with human-defined categories.

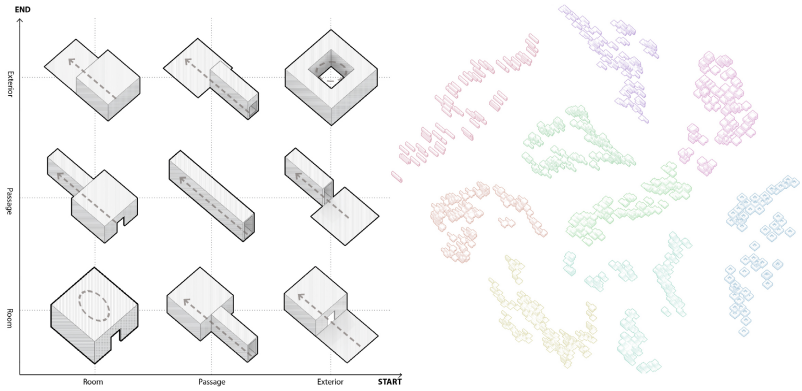

Figure 2. Sequential typologies(left) and t-SNE visualization of K-means Clustering (right)

To further interpret these self-organized features in a subsequent experiment, a Multi-Layer Perceptron (MLP) classifier was trained on these features to translate them into human-readable labels. The MLP outputs a 6-D probability vector, with the first three dimensions predicting the start type and the last three the end type. This classifier's high validation (99%) and test set (98%) accuracies, where accuracy implies correctly matching both start and end types, confirm its efficacy

in accurately categorizing sequential typologies from SSL-extracted features.

4.2. EXPERIMENT II: COMPARATIVE ARCHITECTURAL CASE STUDY

The second experiment aims to explore the influence of Roman architecture on Louis Kahn's designs, fine-tuning the SSL model and MLP classifier from the first experiment. This approach mirrors human cognitive processes, where new spatial experiences are interpreted based on prior knowledge. The strength of such a framework in this comparative study lies in its ability to analyse complex spatial relationships and discern nuanced design influences, tasks typically challenging for traditional methods. By adapting a framework that already has a baseline understanding of various spatial sequences, the models are equipped to interpret intricacies of real-world architectural spaces.

The choice of case study was motivated by the well-documented influence of Roman architecture on Kahn's design philosophy. Kahn's 1950 visit to Rome as a resident architect left a profound and well-known impact on his subsequent designs. This is evidenced in his own writings, accounts from his family and colleagues, and analyses by historical theorists (Barizza, n.d.; Rabifard, n.d.; Scully, 1992). The Indian Institute of Management (IIM) was chosen for its intentional design references to Roman ruins, aiming to invoke a sense of monumentality. Similarly, the Pantheon, Trajan's Market, and Baths of Caracalla were acknowledged by Kahn as influential in his design process. The study aims to quantitatively analyse these interactions using the proposed framework, seeking insights into Kahn's design strategy and philosophy.

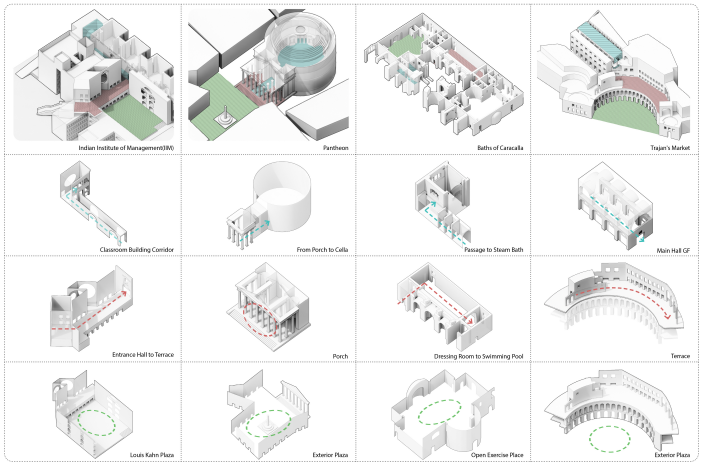

Figure 3. Sampled Trajectories in Four Case Study Buildings

The sampling strategy acknowledges the challenges of representing the universal average experience, given the subjectivity inherent in spatial perception. The criteria for selecting trajectories prioritized architectural significance, ensuring that each path chosen highlighted either unique design elements, areas frequently

engaged by visitors, or spaces with notable architectural details. The final selection of 12 trajectories was carefully curated to ensure that they were not only encompassing a spectrum of experiences but also provided the most valuable data for the study's comparison objectives as shown in Figure 3. Each trajectory is simulated with 100 paths to enrich the dataset.

The SSL model and MLP classifier were fine-tuned using the combination of typology dataset and a subset of the case study dataset. The selected subset, constituting 31% of the total case study dataset, included 5 trajectories with clearly distinguishable sequential types, carefully chosen to avoid ambiguous labels that might confound the model. This fine-tuning aimed to retain the model's foundational knowledge while adapting to the architectural complexities of the case study.

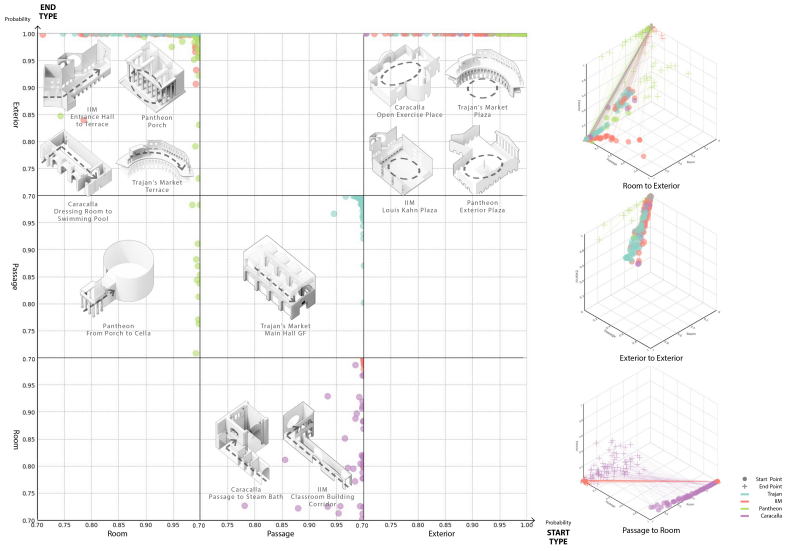

Figure 4. Visualization of Output of MLP Classifier. Left is a nested scatter plot illustrating the predicted probabilities for 12 sampled trajectories across 9 sequential types. On the right, 3D plots highlight the probability vectors for specific start and end types, connecting pairs of points for each building

The fine-tuned models, with 96% validation accuracy on the combined dataset and 94% on the labelled case study subset, effectively grouped trajectories with similar sequence typologies in the case study dataset (Figure 4). Such a typological grouping is crucial for in-depth comparative studies, as it initially sorts sequences into similar typologies for further comparison of interest. Additionally, this approach highlights the framework's efficiency in processing large-scale datasets, a typically time-consuming and challenging task for manual analysis. In the 'exterior to exterior' category, the K-means clustering and nearest neighbour similarity analysis on the feature vectors revealed a notable similarity between the IIM and Trajan's Market, while distinctly differentiated from the Pantheon,

Caracalla's Baths, and a standard courtyard (Figure 5). This pattern suggests shared architectural elements such as terraces and colonnades in both buildings and the absence of arches in the Pantheon's Plaza, exemplifying the framework's ability to uncover nuanced connections not immediately apparent to human observers. By quantitatively assessing how IIM's feature vector diverges from those of Roman buildings and a standard courtyard, the framework enables a data-driven exploration of Kahn's design choices, thereby complementing traditional qualitative analyses.

Figure 5.t-SNE Visualization of K-means Clustering of 'Exterior-to-Exterior' Feature Vectors and Nearest Neighbour Similarity Analysis

5. Conclusion

This research showcases a robust analytical pipeline effectively merging deep learning's computational pattern recognition with the detailed interpretation required in architectural studies. This research's contributions extend from enabling immediate comparative analysis to establishing a foundation for future methodological exploration. Moving forward, the aim is to enhance the spatial perception capabilities of the model by incorporating additional sensory channels beyond grayscale, such as integrating semantic segmentation or object detection layers. Further refinement of Isovist data resolution will allow for a more detailed understanding of architectural elements like ornamentation. Such enhancements will enable the direct visualization of correlations between the model's features and specific architectural elements, thereby greatly enriching the interpretive value. In addition, expanding the dataset and utilizing clustering algorithms will enable the quantification of common qualities in spatial experiences, such as 'monumentality' in the case study experiment, thereby paving the way for a new paradigm in architectural analysis that bridges qualitative assessment with quantitative precision.

References

Barizza, E. (n.d.). *Rome and the legacy of Louis I. Kahn*. Routledge, an imprint of the Taylor & Francis Group.

Benedikt, M. (1979). To Take Hold of Space: Isovists and Isovist Fields. *Environment and Planning B: Planning and Design*, *6*(1), 47–65. https://doi.org/10.1068/b060047

Ching, F. D. K. (1979). Circulation. In *Architecture: Form, Space, and Order* (pp. 280–334). John Wiley & Sons, Inc., Hoboken, New Jersey.

Dawes, M. & Ostwald, M. (2014). Prospect-Refuge theory and the textile-block houses of Frank Lloyd Wright: An analysis of spatio-visual characteristics using isovists. *Building and Environment*, *80*, 228–240. https://doi.org/10.1016/j.buildenv.2014.05.026

Deng, L. & Yu, D. (2014). Deep Learning: Methods and Applications. *Found. Trends Signal Process.*, *7*, 197–387. https://api.semanticscholar.org/CorpusID:53304118

Feld, S., Illium, S., Sedlmeier, A. & Belzner, L. (2020). Trajectory annotation using sequences of spatial perception. *ArXiv*. https://doi.org/10.48550/arxiv.2004.05383

Han, T., Xie, W. & Zisserman, A. (2020). Memory-augmented Dense Predictive Coding for Video Representation Learning. *ArXiv*. https://doi.org/10.48550/arxiv.2008.01065

Hershberger, R. G. (1970). Architecture and Meaning. *The Journal of Aesthetic Education*, *4*(The Environment and the Aesthetic Quality of Life (Oct., 1970), pp. 37–55), 37–55. https://www.jstor.org/stable/3331285

Jing, L. & Tian, Y. (2019). Self-supervised Visual Feature Learning with Deep Neural Networks: A Survey. *ArXiv*. https://doi.org/10.48550/arxiv.1902.06162

Johanes, M. & Huang, J. (2021). Deep Learning Isovist: Unsupervised Spatial Encoding in Architecture. *Proceedings of the 41st Annual Conference of the Association of Computer Aided Design in Architecture (ACADIA)*, 134–141. https://doi.org/10.52842/conf.acadia.2021.001

Leduc, Chaillou, T. and, Ouard, F. and & Thomas. (2011). *Towards a ``typification'' of the Pedestrian Surrounding Space: Analysis of the Isovist Using Digital processing Method*. 275–292. https://doi.org/10.1007/978-3-642-19789-5_14

Lloyd & S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. https://doi.org/10.1109/tit.1982.1056489

Peng, W., Zhang, F. & Nagakura, T. (2017). Machines' Perception of Space. *Proceedings of the 37th Annual Conference of the Association of Computer Aided Design in Architecture (ACADIA)*, 474–481.

Rabifard, M. (n.d.). *The Integration of Form and Structure in The Work of Louis Kahn*. https://api.semanticscholar.org/CorpusID:147416394

Riise, J. (2022). *PedSim* [Grasshopper Script]. https://github.com/julianriise/pedsim

Schiappa, M. C., Rawat, Y. S. & Shah, M. (2023). Self-Supervised Learning for Videos: A Survey. *ACM Computing Surveys*, *55*(13s), 1–37. https://doi.org/10.1145/3577925

Scully, V. (1992). Louis I. Kahn and the Ruins of Rome. *MoMA (New York, N.Y.)*, *12*, 1–13.

Sedlmeier, A. & Feld, S. (2018). Learning indoor space perception. *Journal of Location Based Services*, *12*(3–4), 179–214. https://doi.org/10.1080/17489725.2018.1539255

Turner, A., Doxa, M., O'Sullivan, D. & Penn, A. (2001). From Isovists to Visibility Graphs: A Methodology for the Analysis of Architectural Space. *Environment and Planning B: Planning and Design*, *28*, 103–121. https://doi.org/10.1068/b2684

Vinh, N. X., Epps, J. & Bailey, J. (2010). Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance. *J. Mach. Learn. Res.*, *11*, 2837–2854.

Zevi, B. (1974). The Representation of Space. In B. Zevi, *Architecture as Space* (pp. 45–60). Horizon Press.