Architectural Visualization of Knowledge Spaces: An Image-Based Study of Contemporary Library Design and Information Flow

Dr. S. Shyam Sunder Rao[1*], Dr Ar Abhilasha Pathak[2], Alivia Mondal[3], Rajendra Kumar Ganiya[4]

[1]Deputy Librarian, Nalanda University, Rajgir, Bihar - 803116. Department of Library and Information Science ORCID ID: orcid.org/0000-0003-2697-7824 Email ID: deputylibrarian@nalandauniv.edu.in
[2]Associate Professor and Architect, School of Architecture and Design, DIT University Dehradun Uttrakhand, Dehradun 248009 ORCID ID: https://orcid.org/0000-0001-7163-9256 Email ID: drar_jmits@yahoo.com
[3]Assistant Professor, Department of Architecture, Techno India University, West Bengal, 700091 Specialization: Urban Design & Energy Email ID: alivia.m@technoindiaeducation.com
[4]Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, A.P, 522302, India ORCID ID: 0000-0002-9959-5985 Email ID: rajendragk@kluniversity.in

Abstract
The application of artificial intelligence in architectural studies has mostly focused on the field of geometric reconstruction and generative design, whereas the semantic explanation of knowledge distribution in the interior spaces has not been studied in detail. This paper will suggest a computing model to explain how spatial arrangements of objects in the architectural environment encode knowledge. With the MIT Indoor-67 dataset (15,620 images and 67 categories), a ResNet50 deep learning model was trained on scene classification, with a Top-1 and Top-5 accuracy of 77.01 and 94.6, respectively, on 1,340 test samples. The matrices of scene-object frequency were constructed with the help of object-level annotations, and a Knowledge Density Index (KDI) was proposed to measure the knowledge-related object concentration. The spaces with high knowledge density (bookstores: KDI = 0.609 and libraries: 0.484) were significantly greater than the domestic spaces (bedrooms: 0.049). The similarity modeling based on TF-IDF weights produced a Knowledge Flow Network with high semantic relationships such as bookstore-library (0.8847 similarity). Object-centric model reasoning was validated by the Grad-CAM visualization. The findings indicate that the spatial knowledge distribution can be computedally measured, structured and can be interpreted using a combination of deep learning and semantic modeling methods. The framework leads to the development of AI-supported interpretive spatial intelligence in architectural image analysis.

Keywords: Architectural Image Analysis; Deep Learning; Knowledge Density Index; Indoor Scene Recognition; Semantic Network Modeling.

1. Introduction

The recent high rate of artificial intelligence (AI) introduction into the world of architecture and practice has revolutionized the processes of environment representation, analysis, and design. The latest developments in deep learning have made 3D reconstructions and urban modeling of images much more efficient, allowing spatial interpretation over large scales by visual data (Akhavi Zadegan et al., 2025). Simultaneously, image generation and diffusion-based design systems powered by AI are transforming the process of creativity, providing new concepts on automated interior and spatial design (Beyan and Rossy, 2023; Chen et al., 2023; Hu et al., 2025). This is evidenced by the increasing use of computational vision systems in the production of architectural knowledge.

In addition to generative design, image-based recognition technologies are becoming applications in spatial classification, house-type recognition and reconstruction of built environments (Chang et al., 2025). Systematic reviews also show that the latest digital technologies, such as machine learning and visual computing are redefining the process of analysis and decision-making in architecture (Fang et al., 2025). In addition to visual computing, AI-based digital twin systems are also making real-time monitoring, semantic integration of data, and intelligent decisions in physical spaces within built environments a reality, between knowledge systems, physical spaces, and computational systems (Gautam et al., 2025).

Although these developments have been made, most of the current literature focuses on geometric reconstruction, generative synthesis, or visual realism, and the semantic explanation of architectural knowledge structures has been relatively understudied. At the same time, information science research highlights that visual information is at the center of the perception and interpretation of the environment by the users (Cho et al., 2022). Spatial image representations in tourism and urban perception studies have been revealed to affect cognitive comprehension of place identity (Jia et al., 2025). In the framework of academic and public libraries, the changing spatial structures are more and more influenced by digital technologies, the hybrid knowledge infrastructure, and flexible service paradigms (Jinendran Jain and Kumar Behera, 2023). Nonetheless, there is scanty literature that quantitatively studies how architectural spaces encode and distribute knowledge using object arrangement and spatial semantics based on computer vision approaches.

Despite the progress of AI-based visual technologies in architectural reconstruction, design generation, and optimization of facilities layout (Hu et al., 2025), a methodological gap in comprehending the way of how the knowledge is reflected and conveyed in the visual composition of the indoor spaces persists. The semantic aspect of information density in architectural spaces in object-based form is under-researched, most of the current research focuses on geometric reconstruction or creative synthesis. Little empirical data is available on whether deep learning models recognize indoor spaces using object-centric semantic features, as opposed to using global spatial layout features. Also, it is not clear whether there are object-density signatures of knowledge-intensive environments, including libraries, classrooms, and offices, that can be measured and differentiate them among other spatial typologies. Moreover, there is a lack of research studies on the ability to formalize semantic links between various kinds of spaces as structured information-flow networks. To overcome these gaps, a unified computational system based on scene recognition, object-level annotation analysis, and semantic similarity modeling is needed. The paper will concentrate on indoor architectural scenes based on the MIT Indoor-67 data. It only analyzes the static RGB images and the related object annotations; no spatial geometry, temporal dynamics, and user behavioral information are included. The Knowledge Density Index is operationalized with the help of pre-defined categories of objects related to knowledge, which might not be able to embrace all the spatial knowledge manifestations.

Moreover, the model is based on the controlled learning and pre-trained convolutional networks that can cause bias in the dataset and restrict the generalization to the categories other than the benchmark. The semantic similarity network is founded upon the patterns of co-occurrence of objects and does not deduce causal or functional relation among spatial typologies.The study is relevant to the study of architectural images by changing the analysis emphasis of geometric reconstruction and generative design to semantic interpretation of knowledge dispersion in built environments. The study combines the methods of deep learning classification, object-frequency modeling, and network analysis and proves that architectural typologies can be quantitatively separated on the basis of information density.

The study methodologically presents a Knowledge Density Index (KDI) that is measurable to determine the intensity of knowledge through spatial typologies. It also integrates object-grounded validation with attention visualization to determine whether model prediction is guided by semantically meaningful object cues. Besides, the research develops a semantic Knowledge Flow Network that reflects the spatial proximity and association with each other in the pattern of distribution of objects.

This work can be seen in the wider framework of AI-powered architectural systems and digital spatial infrastructures as an extension of computational approaches to reconstruction and generation to interpretive spatial intelligence.

In response to the identified gaps, this study aims to:

- Develop and evaluate a deep learning–based indoor scene classification model capable of learning high-dimensional representations of architectural spaces.
- Quantify object-based semantic distributions across scene categories using polygon-based annotations.
- Construct a Knowledge Density Index (KDI) to measure the concentration of knowledge-related objects within different spatial typologies.
- Model semantic proximity between spatial categories using TF-IDF–weighted similarity networks to identify structured knowledge relationships.
- Validate object-centric reasoning through attention-based visualization (Grad-CAM) to ensure interpretability of model predictions.

Together, these objectives establish a computational framework for analyzing architectural knowledge spaces through image-based semantic modelling.

2. Literature Review

Artificial intelligence in architecture has been used with increased pace, and it includes design automation, digital reconstruction, intelligent monitoring, and multimodal generative systems. Currently, according to reviews, AI-based systems are becoming more efficient in the design, faster in decision-making, and efficient in performance assessment among architectural processes (Li et al., 2025). In particular, generative and diffusion-based models have already been able to speed up the process of interior design and provide creative flexibility (Rahmoun & Bozkurt, 2025). Multimodal systems like FUGenerator also incorporate textual, visual and parametric inputs to facilitate adaptive architectural design systems (Xu et al., 2025). Although the mentioned developments demonstrate the disruptive power of AI in architectural production, the focus is mostly on synthesis, optimization, and creativity, instead of semantic interpretation of spatial meaning.

The similar progress in digital twin systems and smart systems has also reinforced the connection between physical and computational world. The frameworks based on digital twins allow real-time tracking and synchronization of data in the manufacturing and infrastructure setting (Mahdi et al., 2025). The IoT-based systems in conjunction with machine learning have been suggested to be used in library settings to optimize the performance of the environment and user experience, which will be part of intelligent spatial management (Mammadov and Kucukkulahli, 2025). Likewise, domain-adaptive learning methods have been implemented to corrosion detection and BIM mapping, and it has been shown that AI has the ability to improve structural diagnostics and digital building intelligence (Maharjan et al., 2025). Despite the fact that such systems widen the operational intelligence of built environments, they are largely focused on performance monitoring and structural modelling as opposed to the object-level analysis of spatial typologies in terms of its semantics.

Intelligent information modeling frameworks have also been applied to image-based and laser-based surveys in heritage and documentation. Historic Building Information Modelling (HBIM) is a technology that combines the image-based data with the organized semantic layers to increase preservation intelligence (McGovern and Pavia, 2025). Massive digital platforms have also been created to store and centralize multidisciplinary heritage information, i.e. in the example of Notre Dame de Paris (Néroulidis et al., 2024). These strategies underscore the increased significance of the structured semantic metadata in the architectural archives. Nevertheless, it is concerned with documentation and preservation and not with computational interpretation of the knowledge encoded by object configurations in interior environments.

Geospatially, the research of information representation has been conducted on the effects of spatial data structures on perception and decision-making, especially in disaster and crisis scenarios (Li et al., 2022). This literature shows that the systems of visual representation play an important role in cognitive knowledge about space. Adding to this point of view, neuroscientific research on the concept of biophilic design in learning corridors and stairwells has used physiologic sensors and artificial intelligence (AI)-generated stimuli to assess the perception of space (Kim and Park, 2026). These observations highlight the importance of the fact that spatial environments convey meaning not only in terms of geometry but also in terms of visual and material signals. However, these studies mainly look at the outcomes of perceptions or behavior and not a computational modeling of semantic object distributions.

Indoor scene recognition foundations give a significant methodological precedent to semantic spatial analysis. Quattoni and Torralba (2009) showed that indoor scene recognition involves the combination of global spatial layout information with local object information, which makes interior classification tasks complicated. Further developments in deep learning have increased classification accuracy significantly, but the usefulness of learned representations is a continuing problem. Modern AI studies of architecture still use scene recognition methods, although frequently without making the explicit link between classification performance and object-level semantic reasoning.

Overall, the literature indicates three key directions of AI-enabled architecture research, namely generative design and multimodal creativity (Li et al., 2025; Rahmoun and Bozkurt, 2025; Xu et al., 2025), intelligent monitoring and the integration of a digital twin (Mahdi et al., 2025; Mammadov and Kucukkulahli, 2025; Maharjan et al., 2025), and semantic documentation and spatial perception modeling (McGovern and Although these areas show a remarkable technological progress, there is a clear gap in quantitatively modeling the way interior architectural spaces encode and distribute knowledge in the form of object-level configurations.

Specifically, very few studies have been done which combine scene classification, object annotation analysis and semantic similarity modeling to understand knowledge-rich settings like libraries, classrooms and offices. Current literature is focusing on geometry and reconstruction, performance monitoring, or creative synthesis, without quantitatively measuring the information density of objects and semantically modeling inter-spatial relationships. This is where the necessity to have a computational framework that can be used to go beyond recognition accuracy to a semantically interpretable analysis lies.

The gap that is filled by the present study is the combination of deep learning-based scene classification with object-frequency modeling and network-based similarity analysis. The study takes a step further in applying object-centric reasoning to indoor scene recognition by proposing a Knowledge Density Index and verifying this methodology by visualizing attention in object-centric reasoning. By so doing, it adds to the changing discourse of AI in architecture by moving the focus away of generative production and structural monitoring towards semantic interpretation of knowledge spaces.

3. Methodology

3.1 Research Design

The research design adopted in this study combines a deep learning-based scene classification, object-level semantic analysis, and network modeling to examine the relationship between knowledge encoding and distribution in the indoor architectural environment in terms of spatial object configurations. The research design is designed in such a way that it shifts the predictive modeling into semantic interpretation and validation.

The framework consists of three interconnected analytical layers:

- Controlled Scene Classification with a convolutional neural network (ResNet50) to acquire high-dimensional spatial feature of typologies within the interior.
- Object-Semantic Modeling based on polygon-based annotations in order to build scene-object signatures and count object distributions.
- Knowledge Flow Network Analysis using TF-IDF similarity and a Knowledge Density Index (KDI) to model semantic proximity and knowledge concentration across space types.

Grad-CAM (Gradient-weighted Class Activation Mapping) is added to make the interpretations of the attention to be interpretable. This enables it to be validated that the discriminative regions of the model are related to annotated knowledge bearing objects. A triangulated methodology is created by the combination of predictive performance metrics, semantic object statistics, and attention-based explainability.

3.2 Data Collection Methods

3.2.1 Dataset

The experiment uses the MIT Indoor-67 dataset, which is an open-source benchmark dataset that has been extensively used in the research of indoor scene recognition. The database has 67 categories of indoor scenes representing different typologies of architecture, such as libraries, bookstores, offices, corridors, classrooms, computer rooms, and other functional interior spaces. One of them is libraries and archival settings, which have been identified as spatial expressions of knowledge infrastructures, with the principles of architectural form intersecting the principles of information organization and heritage design (Zou et al., 2025).

The dataset has a total of 15,620 RGB images, which are in JPEG (.jpg) format. The count of images per category is different, but there are at least 100 images per category, which guarantees the minimum representation of classes and prevents the severe imbalance of the classes when the learning is supervised. The data is assigned as research use only and all images were utilized in accordance with its academic use terms.

In this research, the conventional training and testing splits that were given with the dataset were preserved so that the study could be reproducible and comparable with the previous studies.

In order to perform object-level semantic analysis, polygon-based annotations in the LabelMe XML format were used. Such annotations represent the instances of objects in terms of polygon coordinates and semantic labels. The matching of image filenames and annotation files was programmatically done to guarantee that the visual data and object metadata are aligned. The total count of image-annotation pairs that were matched and parsed to extract objects was 2,739, and there were only a few unmatched files, because of inconsistencies in naming.
Such pairs of matched image and annotation were used to model the frequency of scenes and objects, TF-IDF semantic similarity analysis, and Knowledge Density Index computation.

3.2.2 Object Extraction and Preprocessing
For each annotation file:
• Object instances were parsed using polygon coordinates.
• Object labels were extracted and aggregated per scene category.
• A scene–object frequency matrix was constructed.
• Frequencies were normalized per scene to control for varying object counts.
To reduce sparsity and noise, only object labels meeting minimum occurrence thresholds were retained for semantic similarity modeling.

3.3 Population and Sampling
The population consists of all annotated indoor images within the Indoor-67 dataset. The unit of analysis operates at two levels:
• Image-level analysis for classification, feature extraction, and attention modeling.
• Scene-category-level analysis for object aggregation and semantic network construction.
Sampling follows the dataset's predefined training and test splits:
• Training set: Used for model learning and feature extraction.
• Test set: Used for performance evaluation and generalization analysis.
In the case of semantic similarity modeling, statistics of objects were merged together in all annotated images in each category of the scene. To validate Grad-CAM, test images whose correct classification is based on key knowledge-related scenes (i.e. library, bookstore, office, classroom, computer room) as well as transitional scenes (e.g. corridor) were chosen to achieve interpretive clarity with minimum classification bias.
TF-IDF modeling was done with a minimum object occurrence threshold (min_count ≥ 50) to provide statistical strength on the construction of the network.

3.4 Data Analysis Techniques
3.4.1 Scene Classification
The fine-tuning of a ResNet50 convolutional neural network pretrained on ImageNet was performed on indoor scene classification with 67 classes. The cross-entropy loss and Adam optimizer were used to train the model in several epochs. Model performance was evaluated using:
• Training and test loss curves
• Top-1 accuracy
• Top-5 accuracy
• Confusion matrix analysis
Divergence between training and test loss curves was examined to assess overfitting behavior and generalization limits.

3.4.2 Feature Space Visualization
UMAP (Uniform Manifold Approximation and Projection) was used to reduce high-dimensional feature embeddings of the penultimate model trained. This dimensionality reduction method made it possible to visualize cluster structure, showing semantic proximity and overlap in the categories of scenes in representation space.

3.4.3 Scene–Object Semantic Modeling
Object frequencies were transformed using TF-IDF weighting:

$$TF\text{-}IDF_{s,o} = TF_{s,o} \times \log \frac{N}{DF_o} \quad (1)$$

where $TF_{s,o}$ represents normalized object frequency in scene s, DF_o denotes the number of scenes containing object o, and N is the total number of scenes.
To build on an Information Flow Network, cosine similarity between scene TF-IDF vectors was calculated, and edges between two scene types are interpreted as similarity in semantics. A threshold filtering was used to keep only strong semantic connections.

3.4.4 Knowledge Density Index (KDI)
The Knowledge Density Index (KDI) was defined as:

$$KDI_s = \frac{\text{Number of knowledge-related objects in scene} s}{\text{Total object count in scene } s} \quad (2)$$

Knowledge-related objects were operationally defined as elements associated with information production, storage, or exchange (e.g., books, bookcases, monitors, desks).

KDI values were used to weight node sizes in network visualizations and to interpret differences in semantic knowledge concentration across spatial typologies.

3.4.5 Attention-Based Validation (Grad-CAM)

The last convolutional layer was subjected to grad-cam to produce spatial attention heatmaps of correctly classified test images. The heatmaps were normalized and superimposed on original pictures and compared to polygon annotations. Qualitative inspection revealed the high consistency between high-activation in high-KDI scenes (e.g. bookshelves in libraries, monitors in computer rooms, desks in offices). Conversely, transitional spaces like corridors showed attention focused on geometric depth cues and not on clusters of objects as they assisted in the semantic interpretation based on KDI and TF-IDF modeling.

3.5 Ethical Considerations

The research uses publicly available image data, but does not use personally identifiable information. There were no human subjects involved. The entire analysis was done on academic grounds and the use of the data was in accordance with the terms of the license. Environment-based access controls were used to secure authentication credentials. The results of the model performance such as overfitting measures and misclassification trends were also reported explicitly to control the interpretive bias. The paper recognizes the risks of bias that are presented by pretrained networks and the dataset structure.

4. Results

4.1 Data Presentation

4.1.1 Scene Classification Performance

The ResNet50 model with fine-tuning had an overall Top-1 and Top-5 test accuracy of 77.01 % and 94-95 % respectively on the Indoor-67 benchmark (see Figure 1 and Figure 2). The macro-averaged precision (0.7773) and recall (0.7711) show the equal performance among categories.

The training curves show that there is a rapid convergence at the initial stages of epochs, and then the performance levels off after a few epochs (around epoch 15).

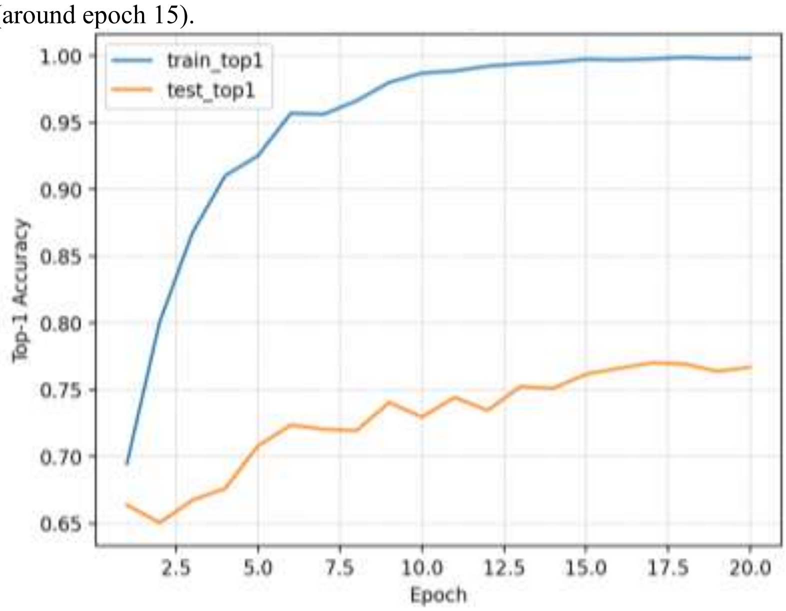

Figure 1. Top-1 Accuracy Curves (Train vs Test)

It is observed that the model is convergent and the training accuracy is close to 100 % whereas the test accuracy is close to 76-77 % (Figure 1). The distance between the curves shows slight overfitting but, the performance of generalization is constant with regard to unseen data.

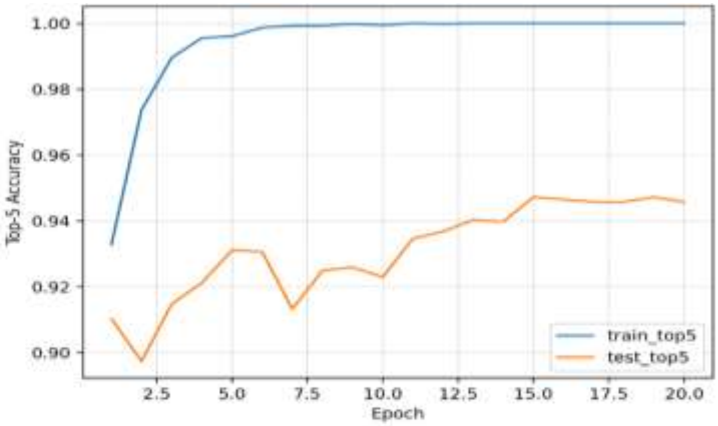

Figure 2 Top 5 Accuracy Curves (Train vs Test)

As Figure 2 demonstrates, Top-5 accuracy is almost perfect on the training set, and it levels off at 94-95 % on the test set. This shows good learning of semantic features and the correct labels are always found in the top predictions of the model.

Table 1. Overall Classification Performance Summary

Metric	Value
Top 1 Accuracy	77.01%
Top 5 Accuracy	94.6%
Macro Precision	0.777
Macro Recall	0.771
Macro F1-score	0.768
Test Samples	1340

According to Table 1, the model has Top-1 accuracy of 77.01% and Top-5 accuracy of 94.6% on 1,340 test samples. The precision (0.777), recall (0.771), and F1-score (0.768) are macro-averaged, which means that the data is balanced among the 67 scene types. The confusion matrix also indicates that the majority of the misclassifications are not between environments of different semantics, like living room-bedroom and bookstore-library, but rather indicates the overlap in object configurations instead of error.

High-performing classes included:

- casino (F1 = 0.97)
- bowling (F1 = 0.95)
- greenhouse (F1 = 0.95)
- cloister (F1 = 0.93)

Lower-performing classes included:

- museum (F1 = 0.47)
- children_room (F1 = 0.51)
- artstudio (F1 = 0.55)

This suggests that visually distinctive typologies outperform semantically overlapping domestic or cultural interiors.

4.1.2 Attention-Based Validation

Grad-CAM visualizations indicate the agreement between model attention and knowledge-carrying objects in knowledge-rich scenes.

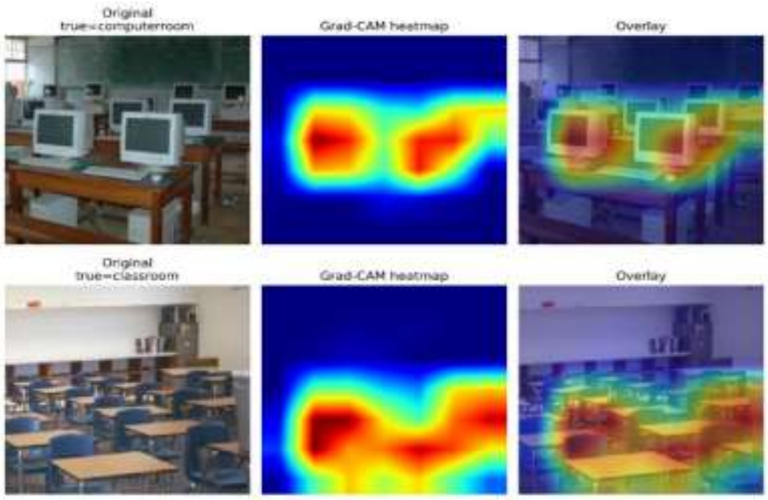

Figure 3. Grad-CAM overlays for selected scenes.

Figure 3 also demonstrates Grad-CAM attention alignment in computer room and classroom scenes with high-activation areas matching closely with the clumped monitors, desks and sitting arrangements. In the computer room scene, the focus is made on monitor groupings and workstation arrangements whereas in the classroom scene, the focus is made on desk rows and instructional areas. This regular overlap of attention heat maps with semantically meaningful clusters of objects supports the idea that the predictions of the model are directed by the object level structure itself and not by abstract spatial geometry itself.

In:

- Library scenes, activation concentrates on bookshelves and book clusters.
- Computer rooms, attention centers on monitors and desk rows.
- Gym scenes, focus is on exercise equipment clusters.

This alignment validates that the model's discriminative reasoning is object-grounded rather than purely geometric.

4.1.3 Knowledge Density and Semantic Modeling

Object-level analysis revealed significant variation in knowledge-related object concentration across scene categories.

Table 2. Top Knowledge-Dense Scenes (KDI Ranking)

Scene	Total Objects	Knowledge Objects	KDI
Bookstore	1856	1131	0.609
Library	523	253	0.484
Office	2519	792	0.314
Computer room	542	160	0.295
Classroom	1108	269	0.243

Table 2 demonstrates that bookstore has the highest Knowledge Density Index (0.609) which means that more than 60 percent of the annotated objects are knowledge-related. Library comes next with a KDI of 0.484 and office, computer room and classroom are characterized by moderate knowledge concentration. Domestic settings, in turn, including living room (0.095) and bedroom (0.049) show much less knowledge density, and there are evident quantitative differences between knowledge-intensive and residential spatial typologies.

4.1.4 Semantic Similarity and Knowledge Flow Networks

TF-IDF modeling produced scene-to-scene similarity structures visualized as a Knowledge Flow Network (Figure 4.4).

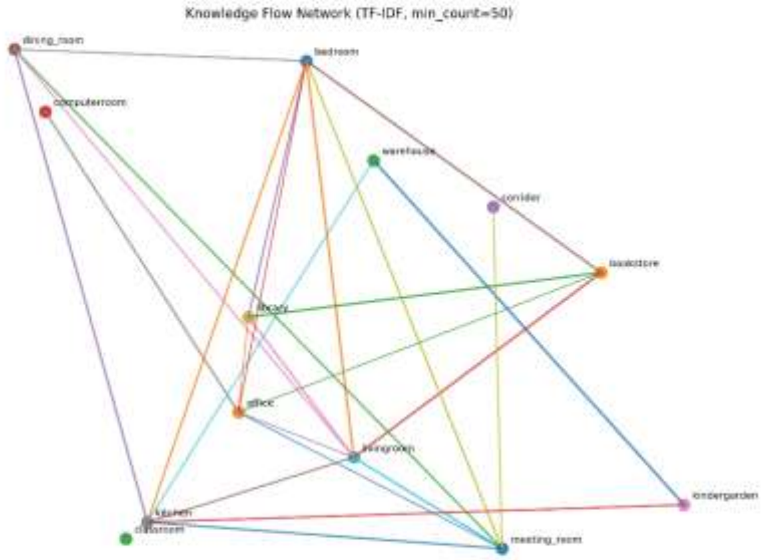

Figure 4. Knowledge Flow Network (TF-IDF similarity, min_count ≥ 50).

Figure 4 illustrates the semantic associations among the categories of scenes according to TF-IDF-weighted distributions of objects (minimum count 50). There are strong associations between bookstore and library, bedroom and living room, computer room and office, and dining room and kitchen, which are indicative of high cosine similarity due to similar vocabularies of objects as opposed to space layout. The network form depicts clustered knowledge-intensive and functionally relevant spaces, proving that the patterns of object co-occurrences create quantifiable semantic proximity between architectural typologies.

Strong semantic links were observed between:

- bookstore ↔ library (similarity = 0.8847)
- bedroom ↔ livingroom (0.9184)
- computerroom ↔ office (0.8001)
- dining_room ↔ kitchen (0.8355)

These links reflect shared object vocabularies rather than spatial geometry alone.

4.2 Key Findings

4.2.1 Deep Learning Captures Object-Centric Semantics

The fact that the Top-5 accuracy is high (around 95%) with Grad-CAM alignment proves that the CNN does not simply classify the scene based on the global layout but rather uses clustered object structures.
Knowledge-dense scenes exhibit:

- Higher classification stability
- Stronger attention-object correspondence
- More distinct TF-IDF semantic signatures

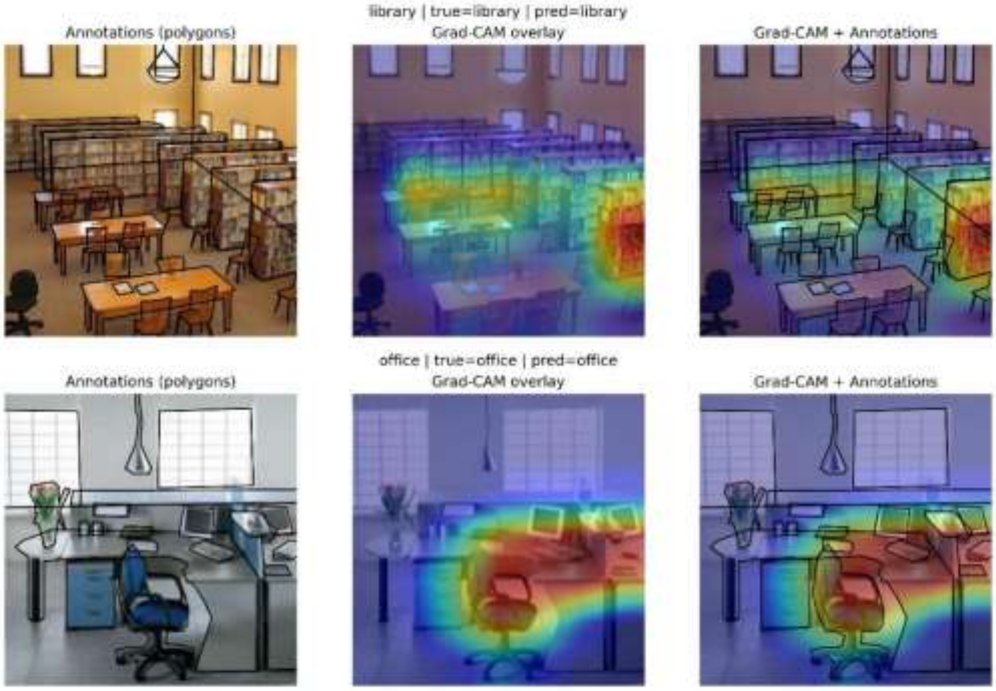

Figure 5. Grad-CAM attention alignment with annotated knowledge-bearing objects in library and office scenes

Figure 5 provides qualitative data of object-based attention in knowledge-laden scenes. Grad-CAM activation in both library and office cases is highly correlated with objects that bear annotated knowledge like bookshelves, desks, and monitors. The fact that high-activation areas overlap with the annotations of object polygons proves that the model works to produce the predictions based on semantically meaningful clusters of objects instead of global layout characteristics.

4.2.2 Knowledge-Dense Spaces Form a Distinct Semantic Cluster

The scenes, which have a high Knowledge Density Index (KDI) value, i.e., bookstore, library, office, computer room, and classroom, create a subnetwork with strong ties in the TF-IDF similarity graph. This structural clustering suggests that knowledge-intensive spaces have repeatable object configurations such as books, desks, monitors and storage objects. These objects are repeated and concentrated to increase semantic cohesion across these spatial typologies, which allows them to be quantitatively differentiated with other scene categories on the basis of object-based information density.

4.2.3 Transitional and Domestic Spaces Show Semantic Overlap

Reduced Knowledge Density Index (KDI) spaces, including living room, bedroom and corridor, have a relative weak semantic distinctiveness and a high misclassification. In contrast to knowledge-intensive spaces, these spaces do harbor wider and more diffused object vocabularies of chairs, tables, lamps, windows, and decorative features, which are not specifically suggestive of a single spatial activity. Consequently, their TF-IDF signatures are more similar across categories and as a consequence, there is semantic blending in the similarity network and low separability in the learned feature space. This interpretation is also supported by the analysis of Grad-CAM. In corridor shots, the focus is often given to the cues of linear perspective, depth gradients and vanishing points as opposed to the cluster of objects.
This is an indication that classification choices in transitional space are more dependent on global geometric structure than on semantically rich groupings of objects. Likewise, the distribution of attention in living rooms and bedrooms can be seen in the spread of diffuse furniture furnishings instead of closely packed knowledge-bearing objects. Taken together, these results point to the fact that the geometric dominance and overlap of objects characterize transitional and domestic settings, which leads to reduced semantic density and categorical distinctiveness than knowledge-based spatial typologies.

4.3 Patterns and Trends

The integrated analysis brings about three key trends. To begin with, there is an increase in classification performance with object distinctiveness. Scene categories with very distinct and visually stable object vocabularies e.g. casino, greenhouse and bowling had the best F1-scores, which shows that discriminative clusters of objects can improve model separability and predictive stability. Second, network centrality in the semantic similarity network seems to be predicted by knowledge density. The high KDI value scenes are the central nodes of the Knowledge Flow Network, which implies that the richness of the object and the concentration of this richness increase the semantic connectivity of the spatial typologies. Third, there is semantic mixing in domestic spaces. Bedroom and living room are another category that has high mutual similarity and high confusion rates, which is in favor of the interpretation that domestic settings have overlapping object configurations and relatively diffuse semantic boundaries. Taken together, these tendencies support the finding that object structure and knowledge concentration are very important in determining both classification performance and semantic network organization.

The findings indicate that the deep-learning-based scene recognition is highly consistent with object-level semantics, which means that classification choices are based on significant spatial features. Object frequency distributions can be used to quantitatively model knowledge density to show quantifiable differences between architectural typologies. Moreover, TF-IDF similarity and network modeling demonstrate that spatial categories have well-structured relations of information flow. A combination of classification performance, semantic modeling, and attention visualization proves that knowledge distribution in architectural space can be computationally detected and analytically measured.

5. Discussion

The results show that the indoor architectural environments represent quantifiable semantic structures which can be extracted and read out computationally. This work is a step forward in predictive accuracy to interpretive spatial intelligence through the combination of scene classification, object-frequency modeling, Knowledge Density Index (KDI) calculation, and semantic network analysis. The findings support the hypothesis that deep learning models can be used as recognition systems and as an analytical tool in the study of architectural knowledge. The classification accuracy (77.01% Top-1 and 94.6% Top-5 accuracy) suggests that convolutional neural networks are capable of distinguishing between different typologies of interiors. What is more important, semantic insights can be seen in performance differences. The good categories (casino, greenhouse, and bowling) indicate that visual distinct and closely clustering object vocabularies promote separability. Contrary to that, classes with lower performance as museum and children_room emphasize the challenge of separating spaces with similar object arrangements. The patterns are in line with the claim that object distinctiveness is a major determinant of stability in spatial classification.

The Knowledge Density Index also gives additional empirical data that some spatial topologies have semantic signatures that are concentrated. Bookstore (KDI = 0.609) and library (KDI = 0.484) have much higher proportions of knowledge-related objects than domestic spaces (bedroom 0.049) and living room (0.095). This quantitative difference supports the assumption that the knowledge-intensive environments could be described by the concentration of the object density. Instead of using qualitative interpretation, the KDI operationalizes knowledge intensity as a quantifiable ratio based on annotated distributions of objects as providing a replicable measure of semantic spatial analysis.

The similarity modeling using TF-IDF also demonstrates the organized connections between the categories of scenes. Close associations like bookstore-library and computerroom-office are examples of similar object vocabularies whereas bedroom-livingroom similarity is an example of domestic semantic blending. The Knowledge Flow Network illustrates that spatial proximity is the result of co-occurrence of objects, but not due to similarity by geometric patterns. It is worth noting that the high-KDI scenes constitute a closely knit subnetwork, which implies that knowledge-intensive spaces have common semantic setups. This supports the reading that the architectural knowledge spaces have coherent object ecosystems, which makes them stand out of transitional or residential typologies. These findings are reinforced by grad-cam attention analysis which confirms the object-centric reasoning. Activation maps in high-KDI workspace like libraries and computer rooms correspond well with bookshelves, desks and groups of monitors. This correspondence validates that object groupings based on semantically relevant meanings are the basis of classification decisions as opposed to abstract layout characteristics. On the contrary, corridors scenes have a focus on depth indicators and vanishing points, which imply the prevalence of geometry in the transitional space. The comparison of the object-driven and the geometry-driven attention gives us an understanding of the way various spatial categories are computed.

Collectively, these findings indicate that architectural settings are organized semantic systems. The cluster of objects that create high semantic cohesion and network centrality characterize knowledge-dense spaces. In contrast, transitional and domestic settings have distributed and overlapping object vocabularies and hence reduced distinctiveness and increased confusion. Combination of quantitative measurements (KDI), similarity modeling and attention-based validation shows that the distribution of spatial knowledge can be detected computationally and structured patterned. There are a number of shortcomings that should be recognized. It is based on the analysis of static RGB images and predetermined categories of knowledge objects, which cannot reflect such intangible dimensions as social interaction or digital infrastructure. The trained architecture can be biased in representation and co-occurrence similarity modeling of objects does not mean that spatial types are causally related in a functional way. Future studies may use multimodal sensing, time based information or embodied perception models to expand semantic interpretation.

Irrespective of these limitations, this paper redefines the role of indoor scene recognition as an interpretive analysis approach as opposed to a strictly predictive one. It measures the intensity of knowledge, models semantic cohesion, and verifies the alignment of attention, which contribute to the computational framework of knowledge about architectural knowledge spaces and to the design of AI-powered spatial intelligence based on object-based meaning.

6. Conclusion

This paper shows that image based computational techniques can be used to study architectural interior environments as structured semantic systems. The study combines deep learning classification of scenes, object level annotation modeling, Knowledge Density Index (KDI) computation, and TF-IDF similarity networks to create a framework of how spatial typologies encode and distribute knowledge using object configurations. The results indicate that knowledge-intensive spaces like bookstores and libraries have concentrated clusters of objects, greater semantic cohesion and high centrality of the network compared to domestic or transitional spaces. The results of the classification suggest that convolutional neural networks distinguish between the indoor setting by using object-based semantic patterns as opposed to the global layout features. The interpretability of model decisions is supported by the fact that high-activation areas on grad-CAM attention maps are associated with knowledge-bearing objects in high-KDI spaces. The Knowledge Flow Network also shows that spatial proximity may be measured by common object vocabularies, indicating that architectural relationships may be generated by information density, not geometry. Despite the fact that it is restricted to the use of static pictures and object categories, the study provides a repeatable approach to semantic architectural analysis. Future studies might include multimodal data, spatial layouts or real-time sensing to model dynamic knowledge flow and project the framework to evolving and intelligent spatial ecosystems.

References

1. Akhavi Zadegan, A., Vivet, D., & Hadachi, A. (2025). Challenges and advancements in image-based 3D reconstruction of large-scale urban environments: a review of deep learning and classical methods. Frontiers in Computer Science, 7, 1467103.
2. Beyan, E. V. P., & Rossy, A. G. C. (2023). A review of AI image generator: influences, challenges, and future prospects for architectural field. Journal of artificial intelligence in architecture, 2(1), 53-65.
3. Chang, J., Lv, Y., Wang, J., Pang, H., & Liu, Y. (2025). Raster Image-Based House-Type Recognition and Three-Dimensional Reconstruction Technology. Buildings, 15(7), 1178.
4. Chen, J., Shao, Z., & Hu, B. (2023). Generating interior design from text: A new diffusion model-based method for efficient creative design. Buildings, 13(7), 1861.
5. Cho, H., Pham, M. T., Leonard, K. N., & Urban, A. C. (2022). A systematic literature review on image information needs and behaviors. Journal of Documentation, 78(2), 207-227.
6. Fang, L., Wang, X., Fang, C., Li, J., & Wang, S. J. (2025). Advanced technologies in creative architectural design: A systematic review. Advanced Intelligent Discovery, 1(1), 2400035.
7. Gautam, A., Aryal, M. R., Deshpande, S., Padalkar, S., Nikolaenko, M., Tang, M., & Anand, S. (2025). IIoT-enabled digital twin for legacy and smart factory machines with LLM integration. Journal of Manufacturing Systems, 80, 511-523.
8. Hu, F., Wang, C., & Wu, X. (2025). Generative artificial intelligence-enabled facility layout design paradigm. Applied Sciences, 15(10), 5697.
9. Jia, M., Chen, J., Chen, Y., Ge, Y., Zheng, L., & Yang, S. (2025). Coupling Relationship Between Tourists' Space Perception and Tourism Image in Nanxun Ancient Town Based on Social Media Data Visualization. Buildings, 15(9), 1465.
10. Jinendran Jain, S., & Kumar Behera, P. (2023). Visualizing the academic library of the future based on collections, spaces, technologies, and services. International Journal of Information Science and Management (IJISM), 21(1), 219-243.
11. Kim, J. Y., & Park, S. J. (2026). Assessment of Biophilic Design in Educational Corridors and Stairwells Using fNIRS and GSR with Generative AI Stimuli. Sensors, 26(3), 985.
12. Li, W., Zhu, J., Pirasteh, S., Zhu, Q., Fu, L., Wu, J., ... & Dehbi, Y. (2022). Investigations of disaster information representation from a geospatial perspective: Progress, challenges and recommendations. Transactions in GIS, 26(3), 1376-1398.
13. Li, Y., Chen, H., Yu, P., & Yang, L. (2025). A review of artificial intelligence in enhancing architectural design efficiency. Applied Sciences, 15(3), 1476.
14. Maharjan, S., Inadomi, S., Itakura, K., & Chun, P. J. (2025). Domain-adaptive self-supervised learning for corrosion detection and 3D building information model mapping in steel tunnels. Computer-Aided Civil and Infrastructure Engineering, 40(26), 4425-4447.
15. Mahdi, M. M., Bajestani, M. S., Do Noh, S., & Kim, D. B. (2025). Digital twin-based architecture for wire arc additive manufacturing using OPC UA. Robotics and computer-integrated manufacturing, 94, 102944.
16. Mammadov, S., & Kucukkulahli, E. (2025). A User-Centric Smart Library System: IoT-Driven Environmental Monitoring and ML-Based Optimization with Future Fog–Cloud Architecture. Applied Sciences, 15(7), 3792.
17. McGovern, E., & Pavia, S. (2025). Historic Building Information Modelling-Adding Intelligence to Laser and Image Based Surveys. Journal of Academic Research and Innovation, 2(1), 2.
18. Néroulidis, A., Pouyet, T., Tournon, S., Rousset, M., Callieri, M., Manuel, A., ... & De Luca, L. (2024). A digital platform for the centralization and long-term preservation of multidisciplinary scientific data belonging to the Notre Dame de Paris scientific action. Journal of Cultural Heritage, 65, 210-220.
19. Quattoni, A., & Torralba, A. (2009, June). Recognizing indoor scenes. In 2009 IEEE conference on computer vision and pattern recognition (pp. 413-420). IEEE.
20. Rahmoun, A., & Bozkurt, A. N. (2025). Enhancing Efficiency and Creativity in Interior Design Through Diffusion Models. Ejons International Journal on Mathematic, Engineering and Natural Sciences, 9(3), 309-320.
21. Xu, X., Feng, T., Zhang, Y., He, Z., & Yuan, P. F. (2025). FUGenerator: multimodal-AI platform for architectural design. Architectural Intelligence, 4(1), 14.
22. Zou, H., Wang, X., & Guo, D. (2025). Architectures of Knowledge: Archival Spaces at the Intersection of Library Science and Heritage Design–A Documentary Case Study Series. African Journal of Library, Archives and Information Science, 35(1).