buildings

MDPI

Article

Can Visual Clusters Support Semantic Steering? A Cross-Scale Study of Responsive Façade Images

Rongrong Liu [1], Ye Lu [2] and Sheng-Yang Huang [2,*]

[1] School of Creative Industry Management and Performing Arts (SCIMPA), Universiti Utara Malaysia, Sintok 06010, Kedah Darul Aman, Malaysia; liu_rongrong2@ahsgs.uum.edu.my
[2] Bartlett School of Architecture, University College London, 22 Gordon Street, London WC1H 0QB, UK; ye.lu.18@ucl.ac.uk
* Correspondence: william.huang@ucl.ac.uk; Tel.: +44-7716-036320

Abstract

Active responsive façades are commonly classified by function or technology, while their visual organization across component and whole-façade scales remains underexplored. This study asks whether image features can organize cross-scale differences and whether cluster-derived semantics can improve generative steering. It introduces a traceable dual-scale workflow linking visual features, clustering, architectural semantics, and image generation. The dataset comprises 75 matched pairs of component close-ups (Set A) and whole-façade views (Set B). Each image was encoded by 43 features: SSIM, mean RGB values, a 32-bin gradient orientation histogram, and seven Hu moments. In two-dimensional PCA space, K-means yielded six scale-specific clusters, interpreted through feature evidence, metadata, and researcher-reviewed GPT-assisted descriptions. Cross-scale agreement was weak but nonrandom (ARI = 0.033; NMI = 0.194; p = 0.0065); however, this association became nonsignificant after feature-group equal weighting (p = 0.829). Tectonic type was associated with clustering at both scales. In a 144-image experiment, semantic prompts increased visual–prototype similarity by 0.0339 (p = 0.0215) and the own-prototype–strongest-competitor margin by 0.0306 (p = 0.0093). Top-1 identification rose from 13.9% to 44.4%, although the difference was nonsignificant (p = 0.0586). The contribution is a traceable visual–semantic process for exploratory generative steering, rather than a fixed taxonomy or stable class-specific control.

Keywords: responsive architecture; K-means clustering; visual feature analysis; semantic modeling; generative artificial intelligence

Check for updates

Received: 24 August 2026
Revised: 12 September 2026
Accepted: 14 September 2026
Published: 16 September 2026

1. Introduction

1.1. Research Background

Generative artificial intelligence is changing the expression of form, space, façades, and imagery in architectural design [1–4]. Architectural design, however, is not image generation in a general sense. Its outputs must respond coherently to tectonic logic, material expression, spatial order, environmental performance, and design intent. Although current models can produce realistic images, they remain limited in domain-specific semantics, logical interpretability, and process control. In particular, generic natural language alone rarely describes façade components, material textures, and environmental relationships with sufficient precision [1,3,5].

The building envelope is a critical interface between indoor space and the external environment, and its design directly affects daylighting, shading, ventilation, heat exchange, and building energy demand. Under climate change and increasingly uncertain environmental conditions, the capacity of static envelopes to respond to dynamic conditions is limited; adaptive envelopes can instead operate as systems that sense and respond to environmental change. From a biomimetic and climate-adaptive perspective, Sommese [6] situates this transition within a broader reconsideration of the building envelope and thereby provides a conceptual basis for understanding the environmental role and technological development of active responsive façades.

Within this broader climatic and technological context, active responsive building façades present a more demanding version of this problem. Their mechanical components, smart materials, sensors, or control systems respond to light, temperature, humidity, wind conditions, and occupant behavior [7–10]. As a result, their visual features are dynamic, temporal, and environmentally dependent. Conventional labels and experience-based descriptions often fail to provide clear, stable, and computable representations.

The visual presentation of an active responsive façade concerns more than dynamic effects and aesthetics. It also communicates environmental adaptation, performance modulation, and affordances in human–machine interaction [11]. These cues emerge jointly from component order, materials, light and shadow, environmental conditions, and response mechanisms. The façade thus becomes a coupled system of performance, environmental information, material behavior, component organization, and visual expression [7–9].

The south façade of the Institut du Monde Arabe in Paris is a representative example. Its mechanical shading components alter their degree of opening in response to light [12]. Such façades operate through relationships among materials, components, controls, and environmental responses, whereas a static image retains only an instantaneous state. This study therefore does not attempt to reconstruct a complete temporal process from a single image. It examines only how visible states and formal cues represent component behavior and potential mechanisms [7,12]. Two-dimensional static images cannot replace video, sensor records, or time-series data. Even so, they remain an important visual medium for disseminating cases, analyzing designs, and supporting generative-AI-assisted design in active responsive architecture. Images can reliably record component states, formal organization, light and shadow, and environmental relationships [3,13]. Structured feature extraction and semantic modeling of these images therefore remain basic tasks for linking architectural visual cognition with generative-AI-assisted design.

1.2. Research Gap, Research Question, and Contributions

Understanding the complex and dynamic façade expression of active responsive architecture requires returning to the ways architecture has traditionally described façade design. As an interface of architectural expression, a façade usually conveys several layers of information, including proportion, composition, material organization, openings, ornament, construction methods, and cultural meaning. Previous studies have commonly used stylistic analysis, formal induction, and typology to classify, abstract, and organize this information hierarchically [14]. Simon's account [15] of complex systems, classification, and hierarchy, together with architectural typology's synthesis of formal, spatial, and tectonic logic, provides a foundation for organizing this knowledge [14].

When an active response mechanism is introduced, façade design extends beyond static form to include component behavior, feedback control, environmental adaptation, and occupant interaction [7,9,16,17]. Conventional style or type labels cannot readily encompass this information. Analysis must combine visible component states with environmental cues and cautiously relate visual form to potential behavior and control mechanisms [7–9]. Without structured treatment of these visual cues, AI tends to remain at the level of surface similarity and struggles to connect potential response mechanisms, component logic, and design semantics.

Multimodal generative design requires mappings among images, language, semantic labels, and design intent [1,2,18]. Existing computer-vision studies of buildings and façades mainly address classification, element recognition, and pixel-level segmentation [13,19], while "multiscale" typically refers to image resolutions or feature levels rather than component-to-façade architectural comparison [20]. Recent studies have begun linking façade images to structured descriptions and generative prompts [3,21,22]. Accordingly, this study uses interpretable visual features and unsupervised clusters as a researcher-mediated visual–semantic layer for generative steering.

The authors' previous study clustered material, form, and climate attributes of wind-responsive passive façades and translated the resulting groups into a hierarchical prompt framework [23]. The present study extends that work in four ways: it examines active façades, uses 43 image-derived visual features, adopts paired component/whole-façade analysis, and evaluates cluster-derived prompts through controlled image generation. Its contribution is therefore not a fixed taxonomy, but a traceable visual–semantic workflow for precedent organization and cross-scale generative steering.

In response to this gap, the study asks whether image feature extraction and clustering can organize the complex visual differences in active responsive building façades at close-up and full-façade scales into comparable and interpretable groups, and whether those groups can be translated into cluster-level semantic labels and tested as prompts for generative steering.

The contribution of this study is an exploratory visual–semantic interface for active responsive building façades. The interface converts visual differences into structured semantics and connects design cognition, architectural information organization, and prompt-based generative steering.

2. Materials and Methods

The research framework comprises dataset construction, benchmark representation, visual feature clustering, robustness validation, semantic mapping, and generative validation (Figure 1). First, a paired dataset of close-up images (Set A) and full-façade images (Set B) was assembled and processed through a common preprocessing procedure. ResNet-50 then provided a benchmark representation. Forty-three interpretable visual features were extracted, and working clusters were obtained through standardization, PCA, and K-means. Cross-scale associations between Sets A and B were then analyzed. The clustering results underwent robustness and external validation through equal weighting and ablation of feature groups, sensitivity analysis of K, a same-pipeline comparison with ResNet-50, and associations with independent architectural attributes. Researcher-reviewed semantic labels were subsequently derived from within-cluster samples, feature profiles, and project documentation. Finally, controlled generation experiments compared a Free Prompt with a Semantic Prompt. CLIP semantic alignment, visual–prototype similarity, and class specificity among six same-scale prototypes were used to examine the generative steering effects of the semantic labels and the limits of their applicability.

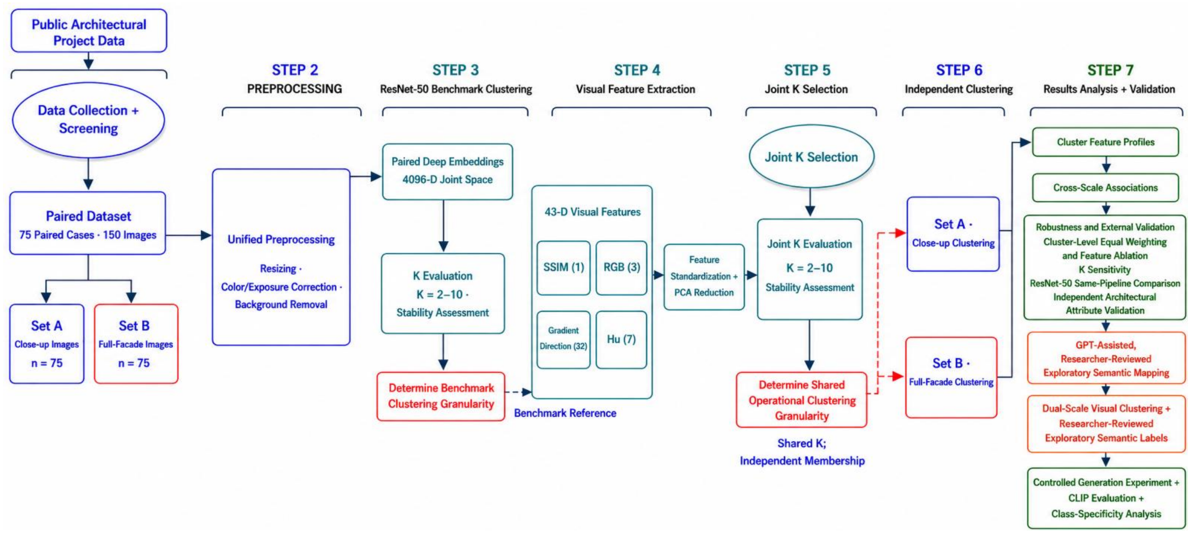

Figure 1. Workflow for Clustering Active Responsive Building Façade Images.

2.1. Dataset Construction

2.1.1. Keyword Extraction and Semantic Descriptor Design

Images were collected through manual searches of publicly accessible websites. Search terms were developed in four stages: theoretical synthesis, case analysis, semantic expansion, and researcher review (Figure 2). The vocabulary was expanded from seed terms into a final system to improve both search coverage and semantic coherence.

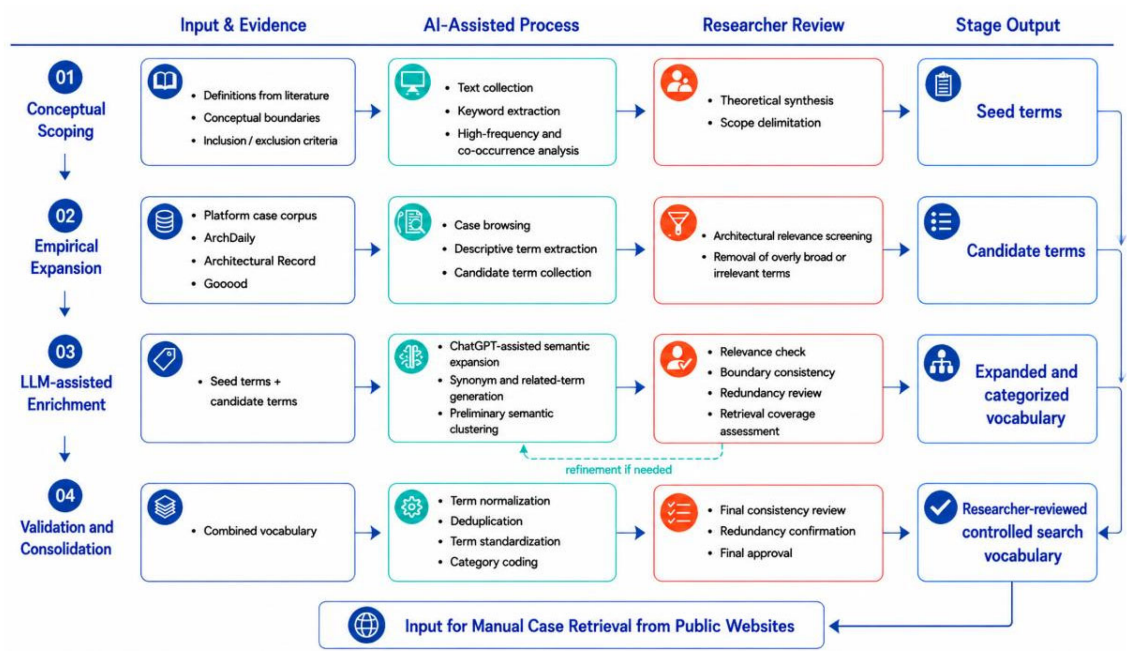

Figure 2. Semantic Workflow for Developing the Search Keywords.

2.1.2. Data Collection and Sample Criteria

The keywords described above were used to search public architecture platforms, including ArchDaily, Gooood, and Architectural Record. Case images, structural attributes, locations, and climate data were collected for feature extraction, cluster interpretation, semantic mapping, and contextual verification. Table 1 lists the collected fields and their purposes.

Table 1. Data Types and Sample-Field Structure of the Active Responsive Building Façade Dataset.

Data Type	Example Fields	Purpose
Image data	close-up images; full-façade images	Extract visual features of the building exterior
Basic building information	Building name, year of completion, floor area, number of stories, and building type	Provide basic information about the building
Designer information	Architect's name and nationality	Identify the design background and cultural-geographical distribution
Response-mechanism attributes	Response target, actuation mechanism, tectonic type, and material type	Record the response strategy and support classification of the response type
Geographical and climate information	Country, city, latitude and longitude, minimum temperature, maximum temperature, annual mean temperature, annual precipitation, mean daily precipitation, and humidity	Cluster or association analysis between environmental variables and building responses

Because active responsive architecture is a relatively recent building type, the number of cases and the completeness of available documentation remain limited [9,12]. The study therefore used purposive sampling across the open web, architectural media, design-firm websites, and project platforms. Built, competition, conceptual, and research cases with clear active response characteristics were included to broaden both the sources and the coverage of the sample [24]. Candidate images were divided by observation scale into close-up images (Set A) and full-façade images (Set B). After excluding images of nontarget buildings, corrupted files, severe defocus, and samples in which the primary subject could not be identified, 75 cases remained. Each case contributed one Set A image and one Set B image, yielding 150 images in total. For each case, the Set A close-up was cropped and enlarged from the corresponding Set B image, ensuring the same façade state, source, media type, and environmental conditions; the primary difference was the field of view and observation scale. The complete metadata are publicly available in the GitHub repository specified in the Data Availability Statement.

2.2. Data Preprocessing

Image preprocessing comprised six steps: exposure and color-cast correction, size standardization, geometric correction, manual cropping, edge-preserving denoising, and format unification. These procedures improved consistency in color, scale, composition, and quality across images from different sources. Figure 3 presents the logic of the preprocessing sequence.

To control preprocessing bias rigorously, one researcher performed the procedure and another reviewed it independently. Color and exposure corrections addressed only evident deviations; no cross-sample homogenization was applied. Perspective correction preserved building proportions and component relationships, while cropping removed only nonarchitectural and temporary distractions. No geometry, material, or texture was added, removed, or redrawn, and color outliers were retained. Complete clustering of the raw and preprocessed images produced identical memberships, indicating that preprocessing did not alter sample assignment.

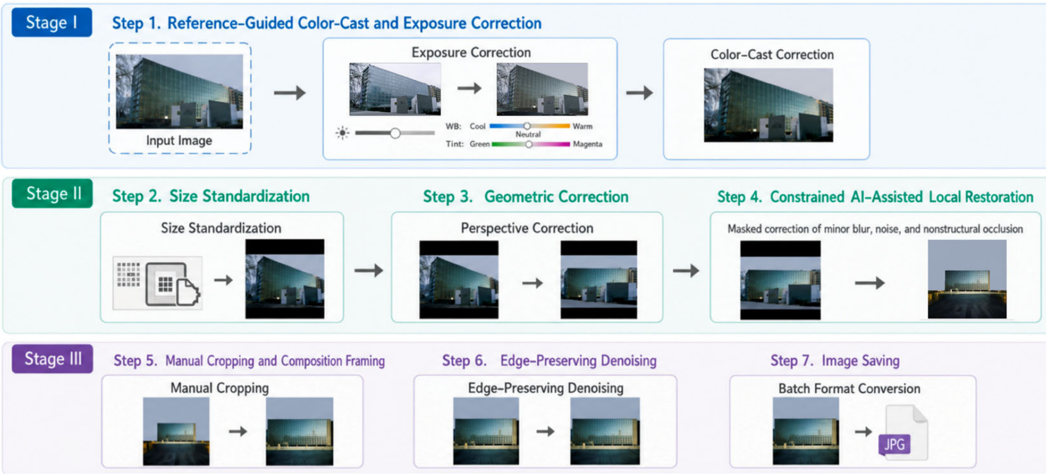

Figure 3. Image Preprocessing Workflow for the Active Responsive Building Façade Dataset.

2.3. Benchmark Clustering Based on ResNet-50 Deep Image Representations

To test the dependence of cluster structure on the form of representation, an ImageNet-1K-pretrained ResNet-50 was used as a reference for general deep representation, rather than as an algorithmic performance competitor to the subsequently handcrafted visual features [25]. The model was implemented in PyTorch (version 2.6.0) [26]. From the 75 raw images in each of Sets A and B, it extracted 2048-dimensional embeddings after global average pooling of layer4 and normalized them with the L2 norm.

The two sets of 2048-dimensional vectors were concatenated case by case with equal weighting to form a 75 × 4096 joint matrix. K-means solutions for K = 2–10 were compared in the full space. Clustering and evaluation were implemented in scikit-learn [27], and the working granularity was determined jointly from the silhouette coefficient, the Calinski–Harabasz index, minimum cluster size, and initialization stability.

As Table 2 and Figure 4 show, the ResNet-50 benchmark selected K = 2 after internal validity, stability, cluster size, and parsimony were considered together. Mean silhouette coefficients for K = 2–10 were all below 0.05, indicating weak separation in the full embedding space. This result serves only as a benchmark for general deep representation. The subsequent 43-dimensional features were not designed to improve clustering performance. They were constructed as domain-interpretable representations for architectural visual analysis and were used to examine how structural similarity, color, gradient orientation, and geometric form organize the data. The results of the two representations therefore compare alternative organizations of the data; they do not constitute a direct ranking of algorithmic performance.

Table 2. Internal Validity Metrics and Minimum Cluster Sizes for K = 2–10 in the Joint ResNet-50 Embedding Space.

K	Mean Silhouette Coefficient	CH Index	Minimum Cluster Size	Mean ARI
2	0.0471	4.5484	30	0.6652
3	0.0486	3.8143	16	0.5276
4	0.0443	3.4291	5	0.3862
5	0.0385	2.9284	7	0.2785
6	0.0298	2.7665	5	0.2733
7	0.0308	2.6253	4	0.2743
8	0.0301	2.5259	5	0.2605
9	0.0328	2.465	5	0.2797
10	0.0373	2.4459	3	0.3015

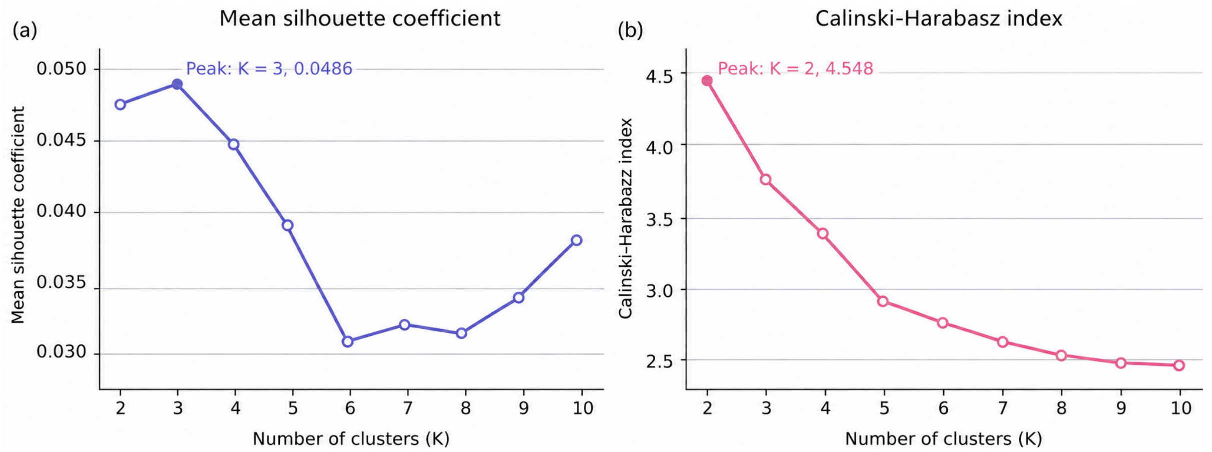

Figure 4. Cluster-validation metrics across K in the joint ResNet-50 feature space: (**a**) mean silhouette coefficient; (**b**) Calinski–Harabasz index.

2.4. Visual Feature Extraction

To convert unstructured façade images into computable, comparable, and interpretable objects [13,28], the preprocessed images in Sets A and B were quantified at four levels: structural similarity, color distribution, edge orientation, and geometric form.

The Structural Similarity Index Measure (SSIM) serves as a proxy for differences in grayscale structure, compositional proportion, light and shadow, and layout [29,30]. Mean values in the red, green, and blue (RGB) channels provide a coarse description of overall tone and environmental atmosphere without identifying specific materials [28,31]. A gradient orientation histogram represents edge direction, rhythm, and compositional order [32–34]. Hu invariant moments add cues about global contour and form, although they do not directly measure architectural symmetry or complexity [35]. Together, these measures provide the feature basis for K-means clustering and semantic interpretation. Figure 5 presents the feature extraction pathway.

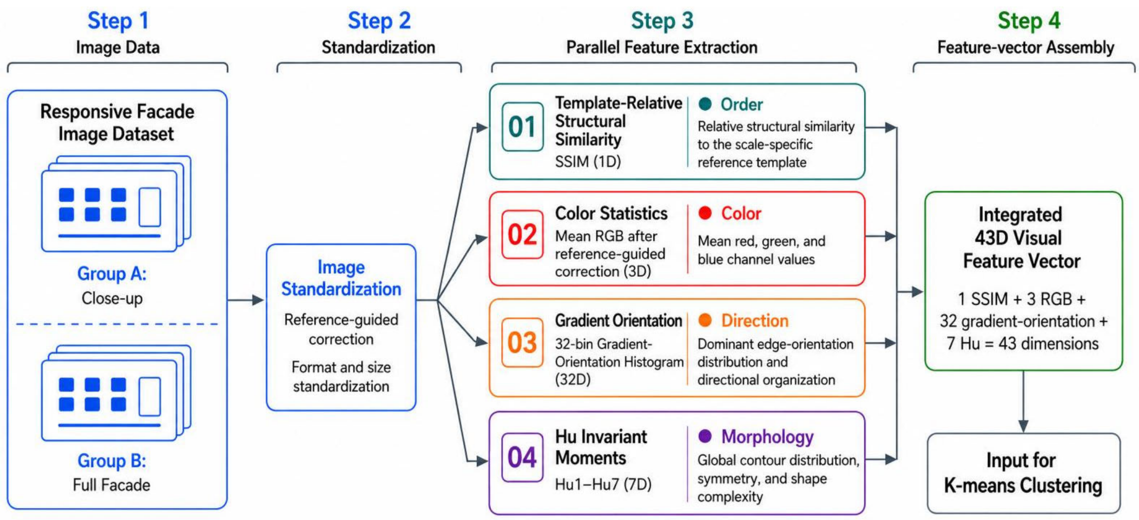

Figure 5. Workflow for Extracting Visual Features from Active Responsive Building Façades.

2.4.1. Extraction of Structural Similarity Features (SSIM)

All images were converted to 256 $\times$ 256 grayscale images, and a within-set pixel-wise mean image was constructed for each set. The SSIM between each image and its corresponding template represents similarity in grayscale structure [29,30]. Table 3 reports the principal statistics.

Table 3. SSIM Statistics for Sets A and B.

Image Set	Minimum SSIM	Maximum SSIM	SSIM Range	Median SSIM
Set A	0.029764	0.681468	0.030–0.681	0.17
Set B	0.025533	0.604031	0.026–0.604	0.29

In summary, typical Set A samples have lower structural similarity to their within-set mean template, whereas Set B samples are generally closer to their template. Differences between the two sets in central tendency and dispersion of structure provide a quantitative basis for the subsequent clustering.

2.4.2. Extraction of Color-Statistical Features

Mean values in the red, green, and blue channels provide a coarse description of the overall tone, brightness, and color tendency of Sets A and B without identifying specific materials [28]. The three-dimensional distribution of channel means across the 150 images is available in an online interactive plot https://rgb3d-0716-image-marks.netlify.app (accessed on 13 September 2026). Figure 6 presents a static version.

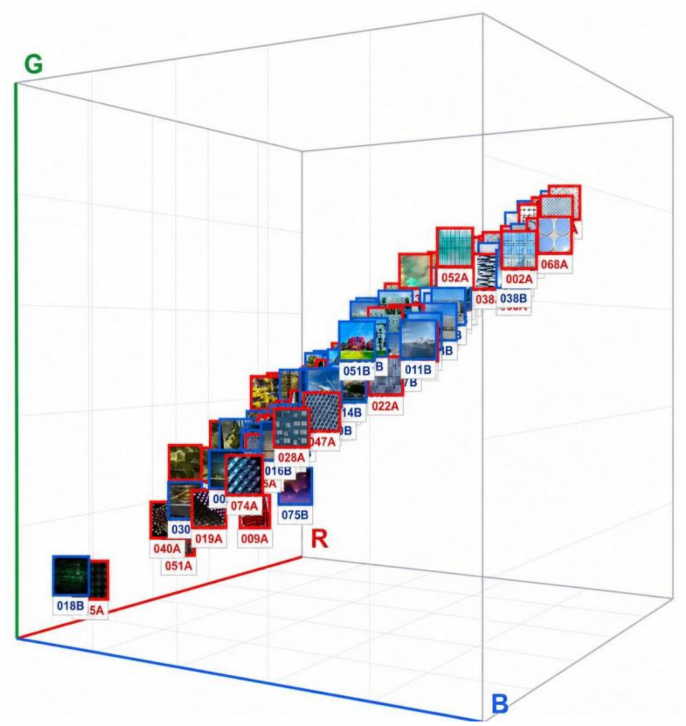

Figure 6. Three-Dimensional Distribution of Mean RGB Values in Sets A and B.

The RGB distribution of Set A is broader, and some samples have higher mean red values, possibly because of material texture, component surfaces, and warm tones. Set B is more concentrated, and some samples have higher mean green and blue values, which may reflect the influence of sky, vegetation, glass reflections, and environmental background. These different distributions of composite color provide quantifiable color variables for subsequent K-means clustering.

To reveal further differences in color structure among individual images, Figures 7 and 8 juxtapose grayscale contours with swatches representing mean RGB values for both sets. Within each set, Mahalanobis distance was also used to detect global observations that departed from the center of the composite color distribution [36], and the local outlier factor was used to test for anomalies in local color density [37].

https://doi.org/10.3390/buildings16183682

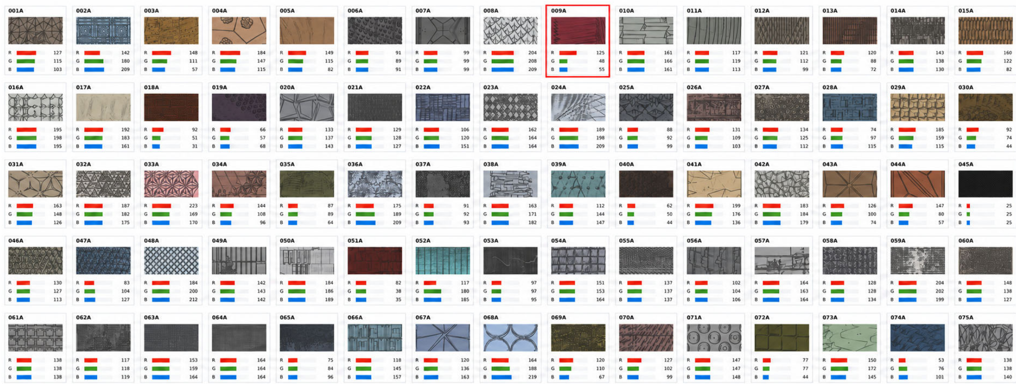

Figure 7. Grid of Mean RGB Color Swatches for Set A Images. The red box identifies the global color outlier detected by CIELAB Mahalanobis distance.

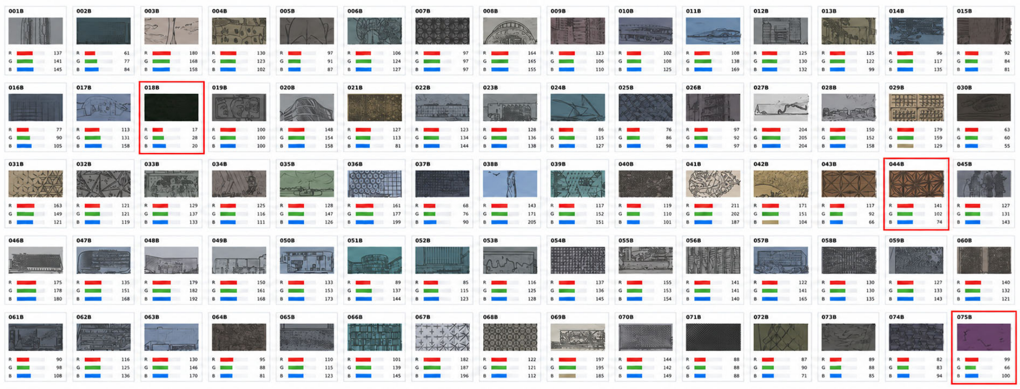

Figure 8. Grid of Mean RGB Color Swatches for Set B Images. The red boxes identify global color outliers detected by CIELAB Mahalanobis distance.

Overall, Set A is dominated by gray-brown, brown, and pale-blue tones. CIELAB Mahalanobis distance identified 009A as a global outlier. Its swatch is redder and darker than the set center and gives greater visual prominence to materials, component textures, and local construction [37]. Set B has a more dispersed distribution of hue and saturation, reflecting the composite coloration of buildings and their surroundings to a greater degree. Images 018B, 044B, and 075B were global outliers: 018B was dark and greenish, while the latter two departed from the overall within-set distribution in other ways. The detection of one global outlier in Set A and three in Set B demonstrates that the RGB features possess a distributional structure and sample-level discrimination suitable for clustering.

2.4.3. Extraction of Gradient Orientation and Compositional Features

Canny edge detection, gradient orientation calculation, and orientation quantization were used to generate a 32-dimensional orientation histogram (hist_0–hist_31) that quantified edge directions and geometric structure in the architectural images [32,33]. As Figure 9 shows, the normalized gradient peaks for Sets A and B occurred at 2.8° (9.11%) and 92.8° (8.35%), respectively. Set A had higher values near 0° and 180°, whereas Set B had a dominant peak near 90°. Because gradient orientation is perpendicular to contour orientation [33], the former pattern emphasizes vertical boundaries and local components, while the latter captures horizontal layering and lateral composition. This difference in directional distributions between the sets also supports later semantic modeling and prompt-based generative steering using orientation distributions.

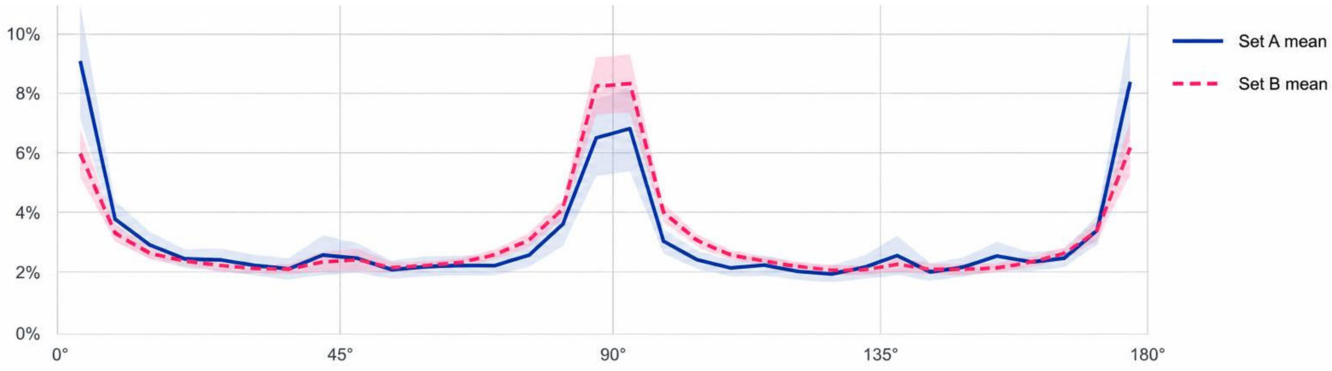

Figure 9. Mean Distribution of Gradient Orientations in Sets A and B.

Paired differences in orientation features for the 75 cases with matching identifiers further showed that most Set A samples were dominant near 0° and 180°, whereas Set B samples were usually dominant near 90° (Figure 10). This tendency was not produced by a small number of extreme cases. Gradient orientation therefore exhibits a stable distribution together with sample variation and can serve as an input variable representing differences in geometric structure in the subsequent K-means clustering.

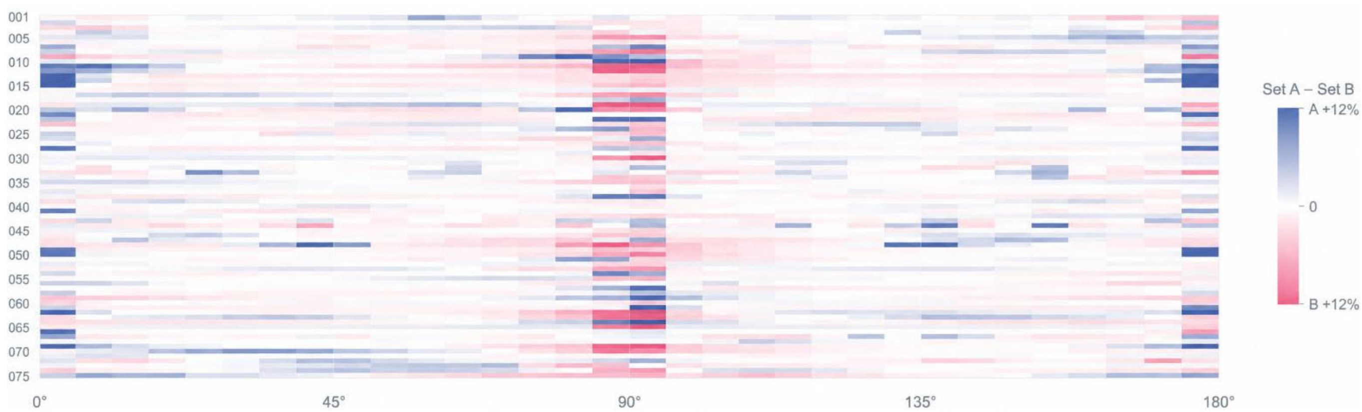

Figure 10. Case-by-Case Paired Differences in Gradient Orientation Between Sets A and B.

2.4.4. Extraction of Hu Invariant-Moment Features

Seven Hu invariant moments were used to describe the global distribution of shapes in contours and regions. These measures are invariant to translation, scale, and rotation, but they do not directly measure architectural symmetry or formal complexity [35]. Table 4 and Figure 11 summarize their statistical behavior and interpretive limits.

Table 4. Distribution, Mathematical Meaning, and Architectural Interpretive Limits of Hu Invariant Moments in Sets A and B.

Hu Feature	Statistical Pattern	Mathematical Interpretation	Limits of Architectural Interpretation
Hu_1 (φ_1)	Similar distributions in Sets A and B	Reflects the second-order overall dispersion of shape relative to its centroid; affected by contour extraction and the proportion occupied by the primary subject	Can help represent the global shape distribution and subject proportion, but does not correspond directly to architectural scale or a specific formal type
Hu_2 (φ_2)	More concentrated in Set A and slightly more dispersed in Set B	Reflects directional differences in second-order moments and shape anisotropy	Can help represent shape directionality and regularity, but does not directly measure architectural symmetry

Table 4. *Cont.*

Hu Feature	Statistical Pattern	Mathematical Interpretation	Limits of Architectural Interpretation
Hu_3 (φ_3)	Less variable in Set A and slightly more variable in Set B	A combination of third-order moments that is sensitive to contour skew, asymmetric distribution, and local perturbations	Can help represent asymmetry in the contour distribution, but is not equivalent to boundary clarity
Hu_4 (φ_4)	A slightly more pronounced skewed distribution in Set B	Represents higher-order differences in shape distribution and can reflect imbalance in the contour distribution	Can help represent imbalance in formal distribution, but cannot identify specific occlusion or background factors directly
Hu_5 (φ_5)	Both sets span positive and negative values, with more extreme values in Set B	A combination of higher-order moments that is sensitive to complex contours and local perturbations	Can help represent changes in complex shapes, but does not correspond directly to curved surfaces, component deployment, or dynamic mechanisms
Hu_6 (φ_6)	Both sets separate into positive and negative values, with greater variation in Set B	A combination of higher-order moments that can reflect complex changes in the global shape distribution	Can help represent contour complexity, but does not directly measure detail complexity or the degree of environmental mixture
Hu_7 (φ_7)	Both sets span positive and negative values and show substantial variation	Sensitive to mirror transformations and capable of representing mirror relationships and antisymmetry in a shape distribution	Can help represent mirrored or antisymmetric forms, but does not directly represent dynamic complexity

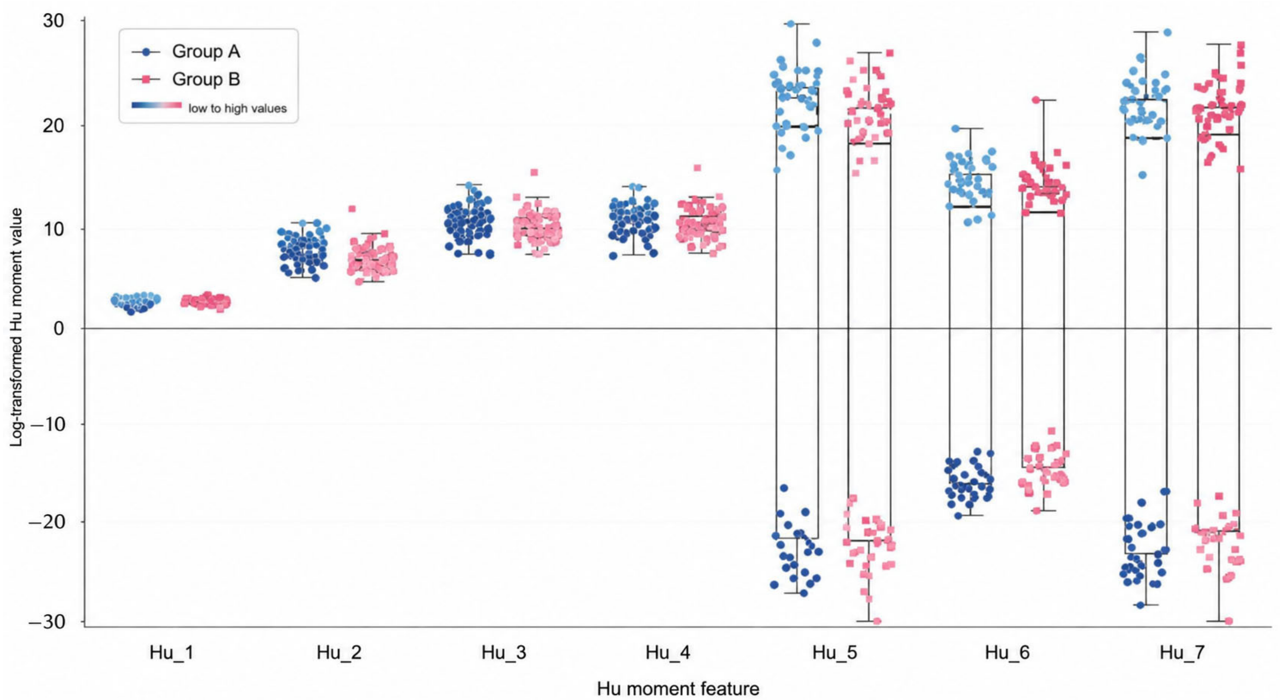

Figure 11. Boxplots With Overlaid Scatterplots of Hu Invariant-Moment Features in Sets A and B.

Overall, Hu_1–Hu_4 largely overlap between the sets, although their dispersion and skewness differ. Hu_5–Hu_7 separate into positive and negative values and vary more widely, making them more sensitive to complex shapes, mirror relationships, and image perturbations. The Hu moments therefore distinguish contours and form along different dimensions and can be used as shape-feature inputs in the subsequent K-means clustering.

https://doi.org/10.3390/buildings16183682

2.5. Clustering and Dimensionality-Reduction Analysis

2.5.1. Feature Standardization, PCA, and the Clustering Analysis Space

Sets A and B each produced a 75 × 43 feature matrix comprising one SSIM value, three mean RGB values, a 32-dimensional L1-normalized orientation histogram, and seven signed-log Hu moments. To calibrate a common value of K, the two sets were concatenated case by case into a 75 × 86 joint matrix. Principal component analysis (PCA) was performed after Z-score standardization to reduce the effects of feature scale, correlation, and distance in high-dimensional space. Variable-wise standardization can still bias weights across feature groups, however, so equal weighting at the feature-group level was also examined later.

The study compared two-, three-, five-, and ten-dimensional PCA spaces with the full standardized feature space. As shown in Table 5, the working K and clustering performance varied across these spaces. The two-dimensional PCA space provided relatively good separation and stability. The cumulative explained variance of the first two principal components was 31.0%, 36.3%, and 48.2% for the joint matrix, Set A, and Set B, respectively. This indicates that the two-dimensional projection retained only part of the variance in the original standardized feature space. The complete feature loadings for PC1 and PC2 are publicly available in the associated GitHub repository. Therefore, two-dimensional PCA and K = 6 were used only as an exploratory working space and granularity for cross-scale semantic comparison and should not be interpreted as representing the intrinsic structure of the full feature space.

Table 5. Clustering Evaluation Across PCA Dimensions and the Full Standardized Feature Space.

Analysis Space	K	Silhouette	CH	DB	Minimum Cluster	Initialization ARI	Resampling ARI
Two-dimensional PCA	6	0.4357	73.8157	0.7279	9	0.9981	0.88
Three-dimensional PCA	4	0.3237	38.742	1.0094	16	0.9389	0.7541
Five-dimensional PCA	3	0.2232	22.7049	1.519	24	0.8824	0.6379
Ten-dimensional PCA	2	0.1609	15.1369	2.081	37	0.804	0.6148
Full feature space	2	0.1116	10.3609	2.5461	37	0.7318	0.6137

2.5.2. Determination of the Number of Clusters (K)

To determine a common number of clusters for Sets A and B, solutions for K = 2–10 were compared without prior labels. K-means used k-means++ initialization and Lloyd's algorithm [38,39], and clustering and evaluation were implemented in scikit-learn [27]. The silhouette coefficient, Calinski–Harabasz (CH) index, and Davies–Bouldin (DB) index respectively measure compactness and separation, the ratio of between-cluster to within-cluster dispersion, and the overall relationship among cluster separations [40–42].

At K = 6, the mean silhouette coefficient and CH index reached the highest values within the candidate range, at 0.4357 and 73.8157, respectively (Table 6 and Figure 12). The six joint clusters contained 20, 14, 11, 11, 10, and 9 cases, so no extremely small cluster was formed. Considering validity, stability, and cluster size together, the study treats K = 6 as a common working granularity in the two-dimensional space rather than as the only intrinsic number of categories in the full space.

The applicability of the common K = 6 solution was also examined separately in the two subsets. With K fixed at 6, neither Set A nor Set B formed an extremely small cluster, and both retained reproducible principal cluster structures. These results support K = 6 as a working granularity for cross-scale comparison (Table 7).

Table 6. Internal Validity Metrics and Minimum Cluster Sizes for Reference Clustering Solutions, K = 2–10.

K	Mean Silhouette Coefficient	CH Index	Minimum Category Size
2	0.3454	45.6158	28
3	0.3915	61.4867	20
4	0.3758	58.115	12
5	0.4077	63.3448	10
6	0.4357	73.8157	9
7	0.4054	69.7118	7
8	0.4039	67.8481	1
9	0.4053	68.2325	1
10	0.4137	68.3483	1

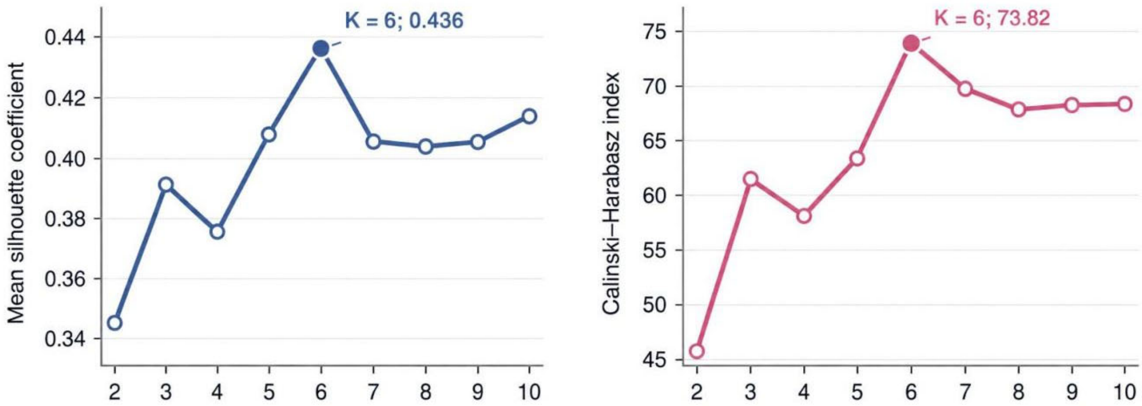

Figure 12. Mean Silhouette Coefficients and Calinski–Harabasz Indices Across K Values.

Table 7. Clustering Validity and Stability of Sets A and B at K = 6.

Subset	Silhouette	Initialization ARI	80% Subsampling ARI	Minimum Cluster
Set A	0.3764	0.7828	0.6842	7
Set B	0.3527	0.6848	0.5982	8

2.5.3. Cluster Robustness and External Validation

To examine the influence of feature-group weighting on the clustering results, SSIM, RGB, orientation-histogram, and Hu-moment variables were divided by $\sqrt{1}$, $\sqrt{3}$, $\sqrt{32}$, and $\sqrt{7}$, respectively, after Z-score standardization. This procedure gave equal weight to the four feature groups. PCA–K-means was then rerun, followed by stability, feature-group ablation, and K-sensitivity analyses (Table 8).

Table 8. Results of Feature-Group Equal Weighting and Robustness Tests.

Test	Result
Agreement between original and revised partitions (ARI/NMI)	Set A: 0.022/0.151; Set B: 0.083/0.261
Validity at K = 6 (silhouette coefficient/minimum cluster)	Set A: 0.393/8; Set B: 0.375/7; joint: 0.369/8
Stability (initialization/80% subsampling ARI)	Set A: 0.922/0.781; Set B: 0.727/0.692
Ablation ARI (lower values indicate greater influence; SSIM/RGB/orientation/Hu)	Set A: 0.521/0.302/0.930/0.786; Set B: 0.574/0.285/0.557/0.793
Sensitivity to K in the joint analysis (silhouette coefficient)	K = 3/6/10: 0.397/0.369/0.401; minimum cluster ≤ 3 when K ≥ 8
Cross-scale and semantic stability	ARI/NMI/p: −0.009/0.086/0.829; Jaccard: 0.125–0.333

Because RGB features in full-façade images may be influenced by non-façade content, we conducted an additional façade-masking sensitivity analysis for Set B. Set A was not included because its close-up images contain little visible context. For Set B, binary masks excluded identifiable sky, ground, adjacent buildings, and foreground vegetation. Visible façade surfaces and their reflections were retained. Visible content within transparent or perforated envelopes was also retained. Only the mean RGB features were recalculated within the valid façade regions defined by the masks; the remaining non-RGB features were retained unchanged. We then repeated clustering using the updated RGB feature matrix. The original feature set, standardization procedure, two-dimensional PCA space, and K-means setting (K = 6) were retained.

The equal-weight clusters agreed poorly with the original clusters: ARI was only 0.022 in Set A and 0.083 in Set B. The equal-weight solution itself nevertheless retained a degree of reproducibility. Ablation results showed that RGB had the greatest influence on the partitions at both scales, while Hu moments had the least. The joint analysis did not support K = 6 as the unique optimum, so this value was retained only as a working granularity for comparing alternative specifications. After equal weighting, the cross-scale association was not significant, and Jaccard overlap between matched clusters from the original and revised solutions was only 0.125–0.333. In summary, the two scale-specific partitions show low agreement and are not interchangeable. The K = 6 structure and its semantic interpretation should be understood as exploratory results produced under a particular weighting specification.

We compared the cluster assignments obtained from the original and masked images. Only 3 of the 75 Set B images (026B, 049B, and 061B) changed cluster assignment. These images represented 4.0% of the set. The other 72 images retained their original assignments. They represented 96.0% of the set. The Set B partition was therefore largely stable under the masking procedure used in this study. The result suggests that visible sky, ground, vegetation, and adjacent buildings were not the main factors determining the original partition. Figure 13 presents three examples of the masking procedure.

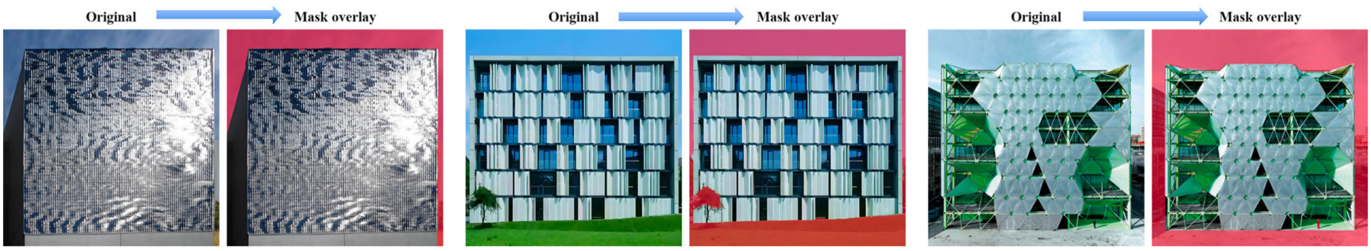

Figure 13. Examples of Façade Masking in Set B.

3. Results

3.1. K-Means Clustering Results and Feature Analysis

3.1.1. Clustering Results for Close-Up Images in Set A

Figure 14 shows varying degrees of local aggregation and boundary transition in Set A at K = 6. Component scale, material texture, edge orientation, and local geometry in the close-up images may therefore influence group membership jointly. Table 9 summarizes the spatial characteristics.

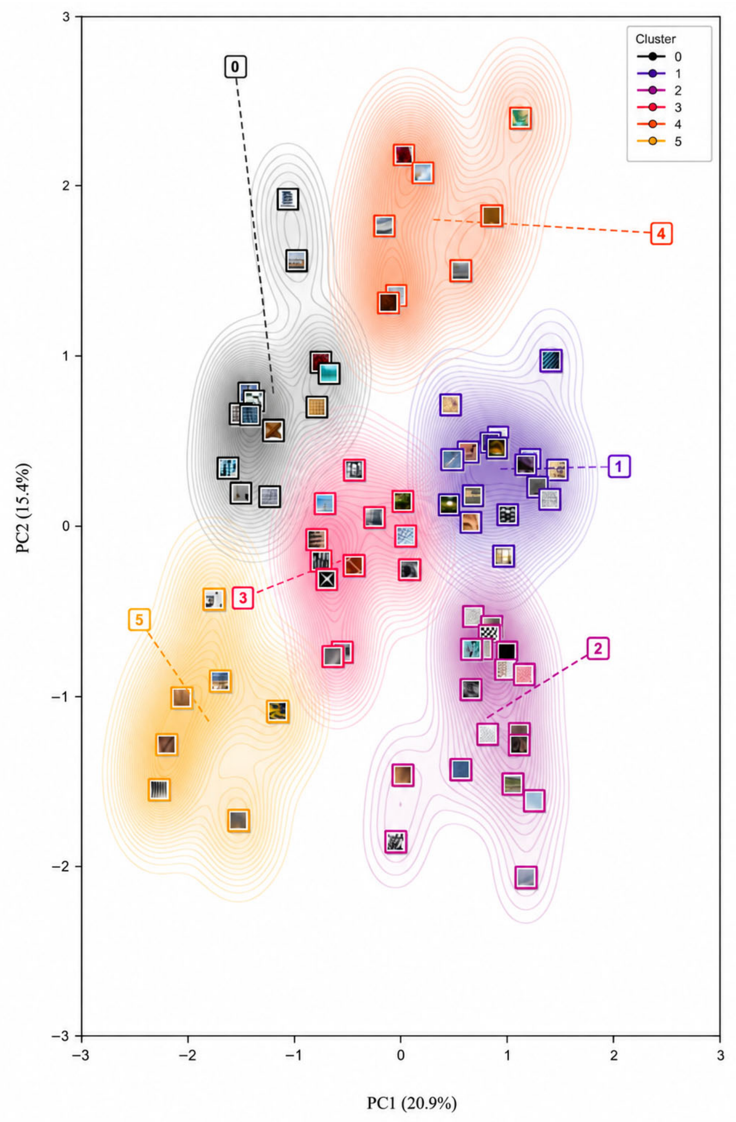

Figure 14. Two-Dimensional Visualization of K-Means Clustering for Close-Up Images in Set A (K = 6).

Table 9. Spatial Distribution Characteristics of Set A Clusters.

Cluster	Outline Color	Spatial Distribution Characteristics (K = 6)
Cluster 0	Black	The cluster contains n = 13 samples, located mainly in the middle-left portion of the two-dimensional projection and extending distinctly in the vertical direction. Its upper end forms a clear vertical tail. Its lower boundary lies close to Cluster 3, while its right side adjoins Clusters 1 and 2.
Cluster 1	Dark purple	The cluster contains n = 17 samples, located mainly in the middle-right portion of the two-dimensional projection. Samples are generally dense within the cluster, with slight horizontal dispersion. Its lower edge adjoins Cluster 2, and its left boundary forms a transitional zone with Cluster 3.
Cluster 2	Magenta	The cluster contains n = 18 samples and occupies mainly the lower-right portion of the two-dimensional projection. It is the largest cluster in Set A. The cluster has a pronounced vertical extension; its upper part adjoins Cluster 1, and a small number of samples along its lower-left edge transition toward the center.
Cluster 3	Red	The cluster contains n = 12 samples, located mainly in the central portion of the two-dimensional projection. Samples are relatively concentrated. Its boundaries lie close to Clusters 0 and 1, and it forms a spatial transition toward Cluster 5 in the lower-left direction, giving it a comparatively clear intermediary position.

Table 9. *Cont.*

Cluster	Outline Color	Spatial Distribution Characteristics (K = 6)
Cluster 4	Orange	The cluster contains n = 8 samples, located mainly from the upper-central to the upper-right portion of the two-dimensional projection. It remains visibly separated from the central and lower clusters and is relatively independent. Samples extend in both horizontal and vertical directions.
Cluster 5	Yellow	The cluster contains n = 7 samples and lies mainly in the lower-left portion of the two-dimensional projection. It is the smallest cluster. Its upper-right edge adjoins Cluster 3. Overall, it is relatively independent and extends diagonally from lower left to upper right.

3.1.2. Clustering Results for Full-Façade Images in Set B

Figure 15 shows substantial overlap and interpenetration among some Set B clusters at K = 6. Full-façade groupings may therefore be influenced simultaneously by building massing, environmental background, lighting, and long-range composition. Table 10 summarizes the spatial characteristics.

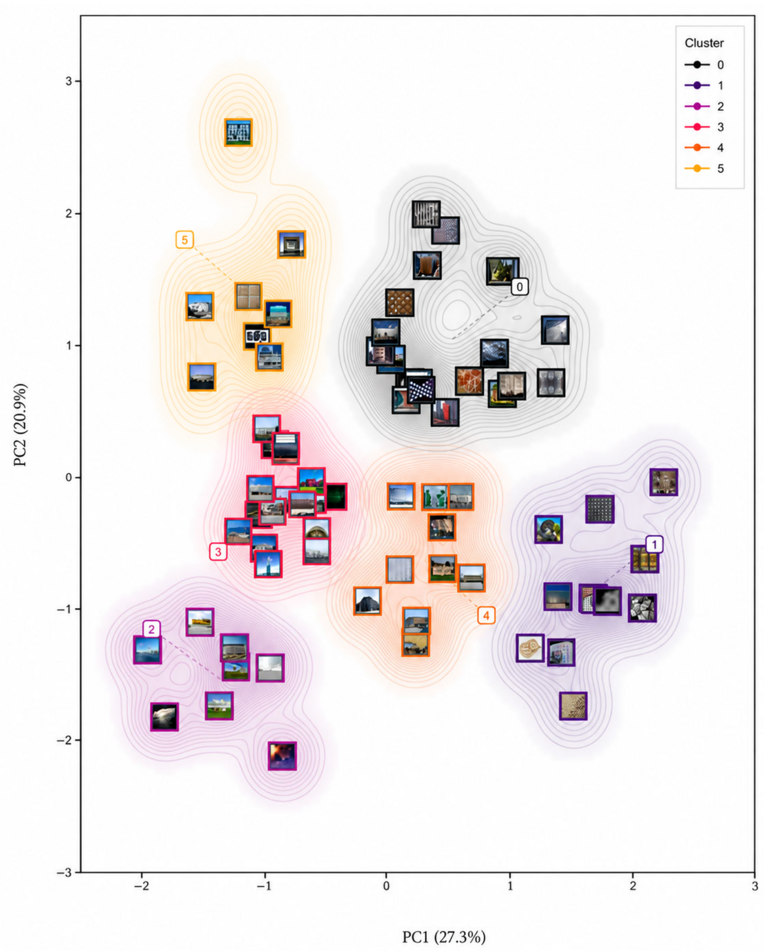

Figure 15. Two-Dimensional Visualization of K-Means Clustering for Full-Façade Images in Set B (K = 6).

These clustering results show that, within the selected working space, visual differences in material expression, component organization, formal features, and environmental presentation collectively provide the basis for grouping the samples. Clustering therefore functions as an intermediate layer between manually selected, domain-oriented visual features and architectural semantic interpretation.

Table 10. Spatial Distribution Characteristics of Set B Clusters.

Cluster	Outline Color	Spatial Distribution Characteristics (K = 6)
Cluster 0	Black	The cluster contains n = 20 samples and is the largest of the six clusters. It is located mainly from the upper-central to the upper-right portion of the two-dimensional projection, covering a broad area and extending in both horizontal and vertical directions. The cluster core is relatively concentrated, while its lower-left boundary extends toward Cluster 4.
Cluster 1	Dark purple	The cluster contains n = 12 samples, located mainly in the lower-right portion of the two-dimensional projection and extending distinctly in the vertical direction. Its left boundary lies close to Cluster 4.
Cluster 2	Magenta	The cluster contains n = 8 samples and is the smallest cluster. It is located mainly in the lower-left portion of the two-dimensional projection and extends horizontally to diagonally. Its upper-right boundary adjoins Cluster 3.
Cluster 3	Red	The cluster contains n = 17 samples, located mainly in the middle-lower portion on the left side of the two-dimensional projection. Samples are relatively concentrated. Its upper edge connects with Cluster 5, and its lower-right boundary adjoins Cluster 4.
Cluster 4	Orange	The cluster contains n = 10 samples, located mainly just below the center of the two-dimensional projection. Its left boundary adjoins Cluster 3, while its right side approaches Cluster 1. The cluster therefore has a spatially connective and transitional character.
Cluster 5	Yellow	The cluster contains n = 8 samples, located mainly in the upper-left portion of the two-dimensional projection. It spans a substantial vertical range and includes one relatively isolated sample at the top. Its lower edge lies close to Cluster 3.

3.1.3. Standardized Visual Feature Profiles of the Six Clusters in Each Set

To determine whether semantic interpretation depended excessively on representative images, cluster means were calculated for all 43 standardized variables using K = 6 memberships. The 32-dimensional orientation histogram was then consolidated into four directional sectors and combined with SSIM, RGB, and seven Hu moments to form a 15-item cluster-level profile. As Figure 16 shows, the clusters have different combinations of standardized features. For example, A-C5 is high in the 0° sector, B-C2 is high in SSIM and the 90° sector, and B-C4 is high across the RGB variables. The differences therefore arise from aggregate within-cluster features rather than from a single representative image. These profiles provide cluster-level evidence for semantic mapping and form an intermediate analytical step in translating computational clusters into architectural semantics.

Figure 16. Heatmap of Mean Standardized Visual Feature Profiles for the Six Clusters in Sets A and B.

3.1.4. Cross-Scale Category Associations Within the Same Cases

To examine category associations and semantic substitutability between the two observation scales, the K = 6 labels for Sets A and B were paired by identifier across the 75 cases to construct a 6 × 6 contingency table (Table 11).

Table 11. Paired Contingency Table of Set A and Set B Cluster Categories (n = 75; Cells Show Frequencies and Row Percentages).

Set A	B-C0	B-C1	B-C2	B-C3	B-C4	B-C5	Total
A-C0	3 (23.1%)	0 (0.0%)	0 (0.0%)	6 (46.2%)	0 (0.0%)	4 (30.8%)	13
A-C1	2 (11.8%)	5 (29.4%)	3 (17.6%)	3 (17.6%)	3 (17.6%)	1 (5.9%)	17
A-C2	4 (22.2%)	5 (27.8%)	1 (5.6%)	4 (22.2%)	4 (22.2%)	0 (0.0%)	18
A-C3	6 (50.0%)	2 (16.7%)	0 (0.0%)	1 (8.3%)	2 (16.7%)	1 (8.3%)	12
A-C4	2 (25.0%)	0 (0.0%)	2 (25.0%)	3 (37.5%)	1 (12.5%)	0 (0.0%)	8
A-C5	3 (42.9%)	0 (0.0%)	2 (28.6%)	0 (0.0%)	0 (0.0%)	2 (28.6%)	7
Total	20	12	8	17	10	8	75

In summary, the two sets of categories have a many-to-many correspondence. Similar local features do not necessarily correspond to a single overall context. Agreement between the partitions is low (ARI = 0.033, NMI = 0.194), precluding a stable one-to-one correspondence [43,44]. A permutation test with 10,000 iterations showed that the observed NMI was greater than would be expected from random pairing, p = 0.0065 [45]. The two scales therefore have a limited but nonrandom selective association. Figure 17 shows the specific pairing paths among categories. The low cross-scale agreement but above-chance association indicates that local construction and the overall façade have only limited, nonrandom selective links; they should be treated as complementary rather than interchangeable descriptions.

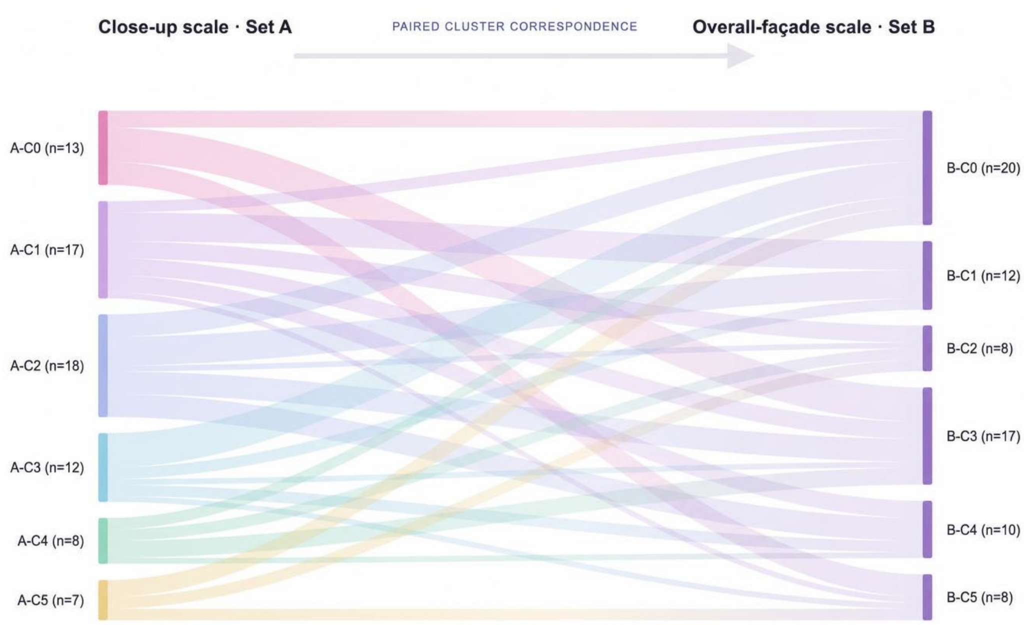

Figure 17. Cross-Scale Correspondence Between Set A and Set B Cluster Categories for the 75 Paired Cases.

3.1.5. Same-Pipeline Validation with ResNet-50 Representations

To examine whether clustering results depended on the form of feature representation, ResNet-50 was evaluated through the same validation procedure used in the principal analysis of the 43 handcrafted features. The paired joint data again selected K = 2, consistent with the full-space result in Section 2.3. The two clusters in the two-dimensional PCA space, however, contained 72 and 3 cases and did not form a size-balanced multiprototype structure.

Table 12 presents a comparison with the 43-dimensional solution when K was fixed at 6. ResNet-50 produced single-sample clusters in both the joint data and Set A, while the smallest Set B cluster contained only three samples. Although ResNet-50 showed comparable or higher values on some internal validity and stability metrics, its cluster-size distribution was substantially more uneven. The 43-dimensional handcrafted feature representation was therefore retained for the principal analysis because it produced a more balanced six-cluster structure that could be used consistently for subsequent cluster-level semantic labeling and cross-scale comparison.

Table 12. Same-Pipeline Comparison of the 43-Dimensional Handcrafted Features and ResNet-50 at K = 6.

Data	Feature Representation	Mean Silhouette Coefficient	Minimum Cluster Size	Initialization-Stability ARI	80% Subsampling ARI
Paired joint data	43-dimensional handcrafted features (original result)	0.4357	9	0.9981	0.88
Paired joint data	ResNet-50	0.4078	1	0.9483	0.6542
Set A	43-dimensional handcrafted features (original result)	0.3764	7	0.7828	0.6842
Set A	ResNet-50	0.4833	1	0.8101	0.6856
Set B	43-dimensional handcrafted features (original result)	0.3527	8	0.6848	0.5982
Set B	ResNet-50	0.4422	3	0.8517	0.6389

3.1.6. Validation Results for Independent Architectural Attributes

To test associations between visual clusters and independent architectural attributes, the study compared Set A and Set B clusters in terms of response target, actuation mechanism, tectonic type, material type, and climate conditions. Categorical attributes were tested with Monte Carlo permutation chi-square tests, while climate variables were tested with permutation Kruskal–Wallis tests. Cramér's V and ε^2 were reported as effect sizes. Multiple labels were combined into a "multiple" category, and missing values were not imputed. p values were adjusted with the Benjamini–Hochberg false discovery rate procedure, with $q < 0.05$ treated as significant.

Table 13 presents the results. After adjustment, only tectonic type was significantly associated with both partitions (Set A: q = 0.002, V = 0.426; Set B: q = 0.018, V = 0.390). The clusters therefore reflect differences in tectonic organization to some extent. Response target, actuation mechanism, material type, and climate variables showed no significant associations ($q \geq 0.387$). The clusters partly capture constructional differences, but they cannot be interpreted as being determined directly by the other architectural attributes.

Table 13. Association Tests Between Independent Architectural Attributes and the K = 6 Partitions (Categorical Variables $n = 75$; Climate Variables $n = 73$; BH-FDR Adjusted).

Architectural Attribute	Set A: q (Effect Size)	Set B: q (Effect Size)	Result
Response target	0.745 (V = 0.247)	0.442 (V = 0.272)	Not significant
Actuation mechanism	0.745 (V = 0.227)	0.628 (V = 0.241)	Not significant
Tectonic type	0.002 (V = 0.426)	0.018 (V = 0.390)	Significant at both scales
Material type	0.745 (V = 0.262)	0.417 (V = 0.280)	Not significant
Climate variables (annual mean temperature, annual precipitation, and mean humidity)	0.387–0.745 (ε^2 = 0–0.059)	0.417–0.521 (ε^2 = 0–0.033)	Not significant

3.2. Selection of Representative Images and Visual–Semantic Mapping

To make the clustering results more intuitive and interpretable, the real sample with the shortest Euclidean distance to the K-means cluster centroid was selected from each cluster in Sets A and B as the primary representative image. This procedure linked the abstract groups to observable and comparable visual samples.

A representative image cannot independently reveal the full architectural semantics and is used only for visual presentation. Semantic mapping was therefore researcher-led, supported by ChatGPT suggestions for candidate wording and by project documentation; labels were not generated automatically from the features. Candidate descriptions first combined visible characteristics, cluster-level feature profiles, and documented information about mechanisms or context. The researchers then verified, consolidated, and condensed those descriptions to produce the exploratory semantic mappings in Figures 18 and 19. This process links visual evidence, contextual documentation, and generative prompts, while avoiding a direct equation of static cues with material performance or response mechanisms [1,3].

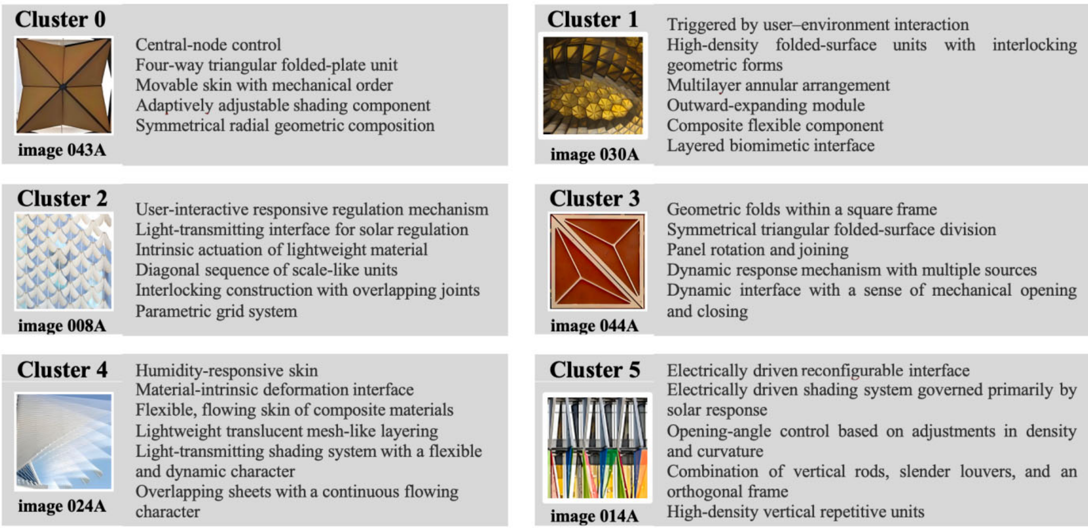

Figure 18. Exploratory Visual–Semantic Mapping for Set A (K = 6).

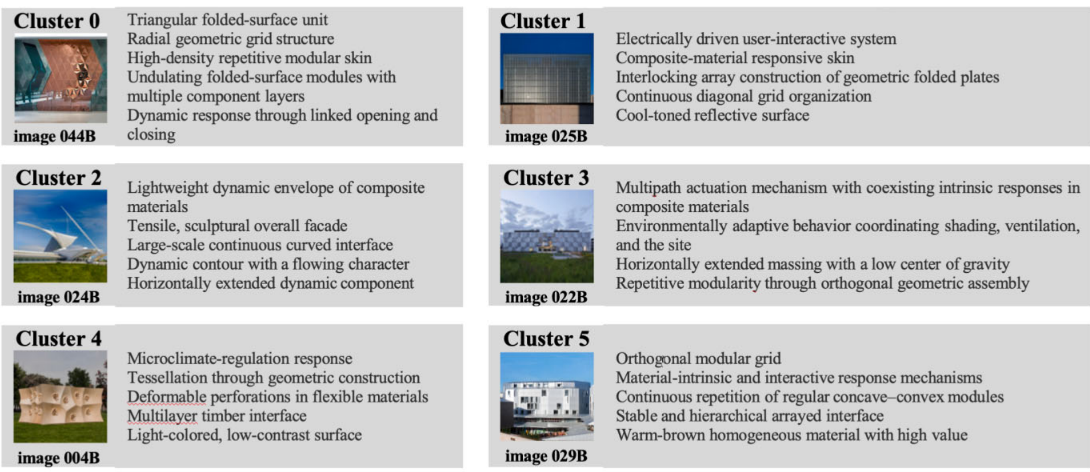

Figure 19. Exploratory Visual–Semantic Mapping for Set B (K = 6).

Set A contains close-up images that primarily show local components, material textures, open or closed states, and other visible cues in active responsive building envelopes. Figure 18 presents the representative images and exploratory semantic mapping for each cluster.

Set B contains full-façade images and places greater emphasis on overall composition, proportion, spatial assembly, and building–environment relationships. Its semantic mapping also addresses massing, façade layering, environmental reflection, site relationships, and landmark expression. Figure 19 presents the results for each cluster.

3.3. Controlled Generation Experiment with Cluster Semantic Labels

3.3.1. Experiment and Image Generation

To test the effectiveness of semantic labels in generative steering, the study established two conditions: a Free Prompt and a Semantic Prompt. Both prompts used the same building type, observation scale, composition, style, lighting, and negative constraints. Only the latter added a structured cluster-level semantic description.

The experiment covered 12 semantic targets across Sets A and B and generated 144 images in two independent batches. Evaluation used two CLIP metrics: target–text similarity and visual–prototype similarity [46]. The 12 semantic clusters served as the units of inference. Cluster-level Semantic–Free paired differences were pooled with equal weighting, evaluated with a two-sided exact sign-flipping test, and accompanied by cluster-level bootstrap confidence intervals. To examine class specificity further, each generated image was compared with the six visual prototypes at the same scale using CLIP similarity. The resulting own-prototype–strongest-competitor margin, Top-1 prototype identification rate, and nearest-prototype confusion matrix were used to evaluate class specificity, as shown in Figure 20.

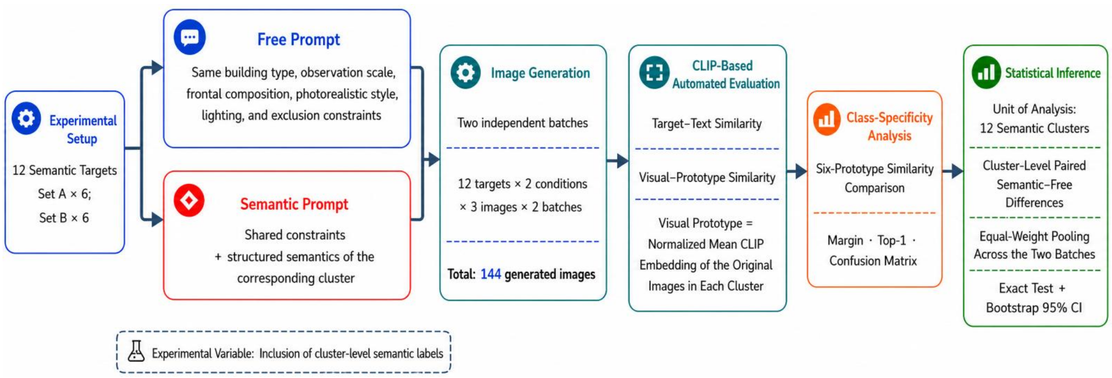

Figure 20. Workflow of the Controlled Generative-Steering Experiment.

3.3.2. Results of the Controlled Generation Experiment with Cluster Semantic Labels

Figure 21 presents representative outputs under the two prompt conditions. The target visual features became more explicit after the semantic labels were added. The figure is provided only to illustrate experimental differences; statistical conclusions are based on all 144 generated images.

CLIP evaluation showed that the Semantic Prompt outperformed the Free Prompt on both metrics. After equal-weight pooling, target–text similarity increased by 0.0355 (95% CI [0.0266, 0.0447], $p = 0.0005$), and visual–prototype similarity increased by 0.0339 (95% CI [0.0097, 0.0569], $p = 0.0215$). Figure 22 and Table 14 show effects in the same direction and of similar magnitude across both batches, providing preliminary evidence of reproducibility. The results support improved visual–semantic alignment, but they do not imply gains in constructability, responsive performance, or overall design quality.

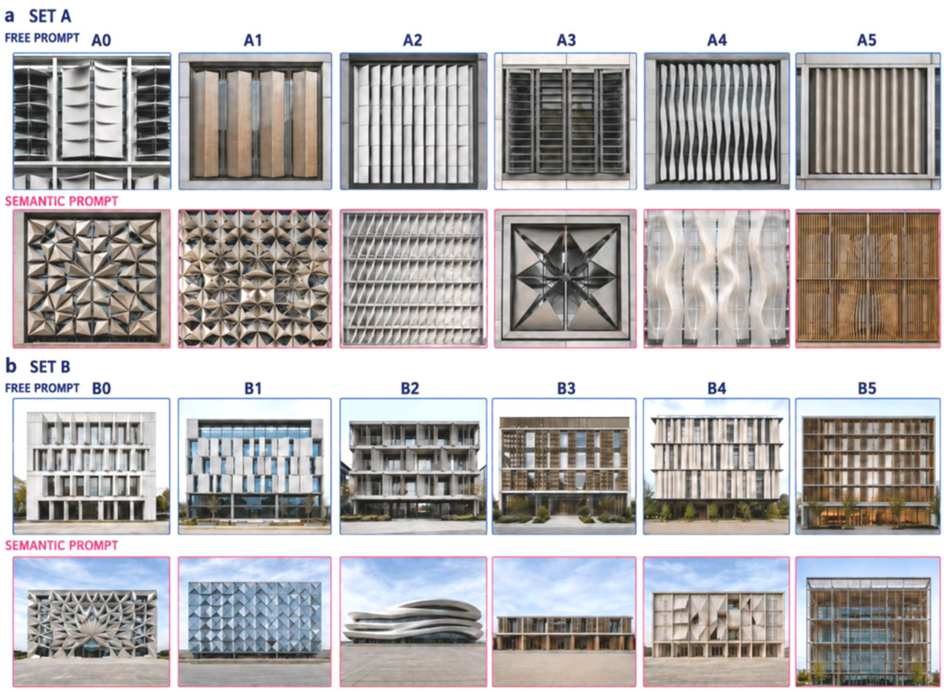

Figure 21. Representative Results Under the Free Prompt and Semantic Prompt Conditions: (**a**) Set A (close-up scale), where A0–A5 denote its six semantic clusters; (**b**) Set B (full-façade scale), where B0–B5 denote its six semantic clusters.

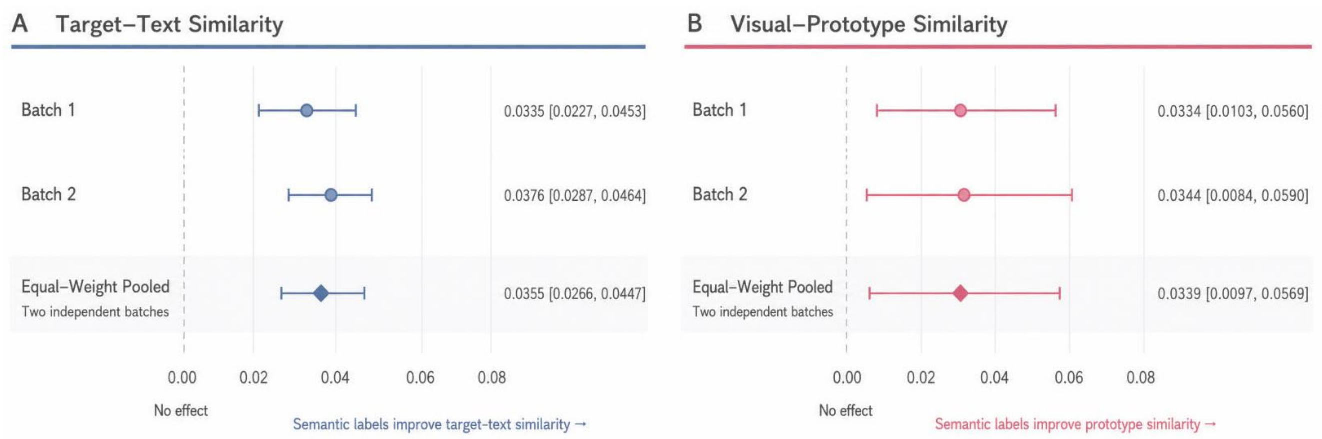

Figure 22. Cluster-Level CLIP Effects for the Two Batches and Their Equal-Weight Pooled Estimate.

Table 14. CLIP Evaluation Results for Two Batches and Their Equal-Weight Pooled Estimate.

Endpoint	Analysis	Free	Semantic	Difference	95% CI	Exact p	Positive Clusters
Target–text similarity	Batch 1	0.2864	0.3199	0.0335	[0.0227, 0.0453]	0.0005	12 of 12
	Batch 2	0.2865	0.3241	0.0376	[0.0287, 0.0464]	0.0005	12 of 12
	Equal-weight pooled	0.2865	0.322	0.0355	[0.0266, 0.0447]	0.0005	12 of 12
Visual–prototype similarity	Batch 1	0.8057	0.8391	0.0334	[0.0103, 0.0560]	0.0234	10 of 12
	Batch 2	0.8074	0.8418	0.0344	[0.0084, 0.0590]	0.0293	10 of 12
	Equal-weight pooled	0.8066	0.8404	0.0339	[0.0097, 0.0569]	0.0215	10 of 12

To examine class specificity further, the 144 generated images were compared with the six prototypes at the same scale. As Table 15 and Figure 23 show, the Semantic Prompt increased own-prototype similarity from 0.8066 to 0.8404 (p = 0.0215). The margin over the strongest competing prototype improved from −0.0425 to −0.0119 (p = 0.0093). The Top-1 identification rate rose from 13.9% to 44.4%, and identification exceeded the chance level obtained by permuting labels. The difference in Top-1 rates between the two conditions did not reach statistical significance (p = 0.0586). Semantic labels therefore improved prototype consistency and category discrimination in the generated outputs. The negative strict margin and Top-1 rate below 50% nevertheless indicate that stable class-specific control has not yet been demonstrated.

Table 15. Same-Scale Six-Prototype Class-Specificity Analysis.

Metric	Free Prompt	Semantic Prompt	Difference	95% CI	p Value
Own-prototype similarity	0.8066	0.8404	0.0339	[0.0097, 0.0569]	0.0215
Margin over strongest competing prototype	−0.0425	−0.0119	0.0306	[0.0137, 0.0484]	0.0093
Top-1 prototype identification rate	13.90%	44.40%	30.6 percentage points	[5.6, 55.6] percentage points	0.0586

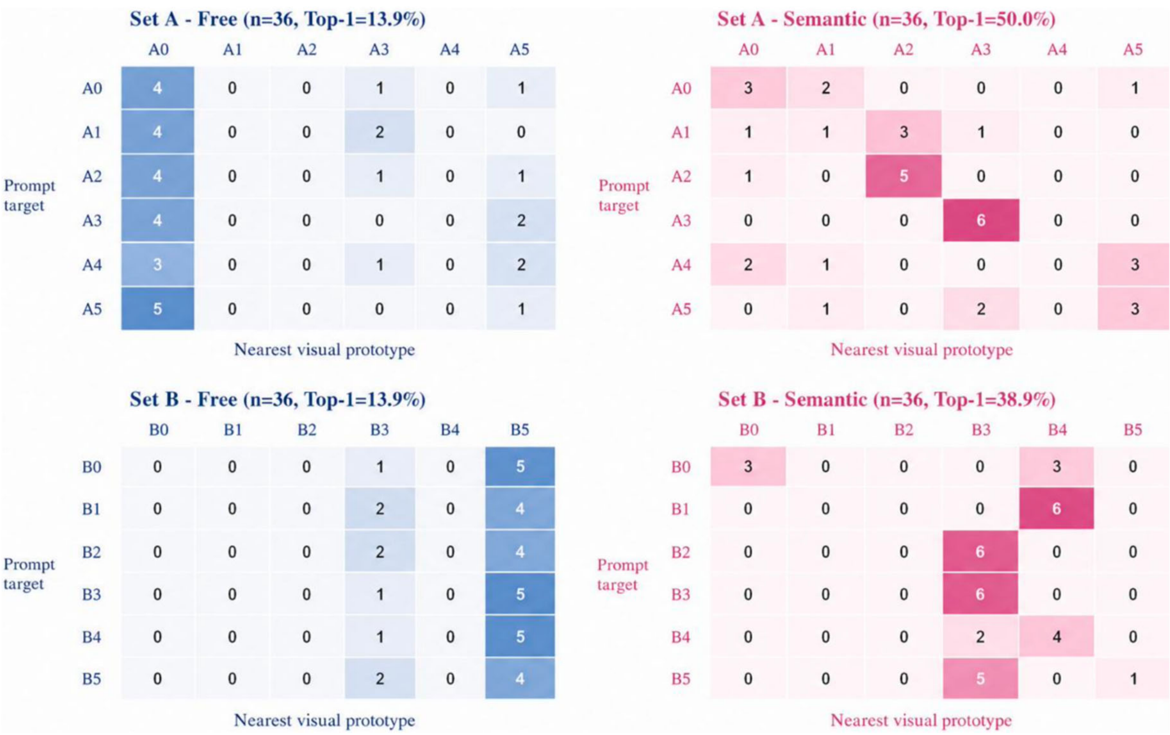

Figure 23. Same-Scale Nearest-Prototype Confusion Matrices for the 144 Generated Images. Blue matrices show the Free Prompt condition and pink matrices show the Semantic Prompt condition; darker cells indicate higher counts.

4. Discussion

Previous computational studies of architectural imagery have largely focused on classifying conventional buildings, recognizing styles, synthesizing types, and quantifying visual features. That work established a basis for computational treatment of architectural visual data, but its endpoint has usually remained classification, recognition, or interpretation of model results [13]. The present study instead examines active responsive building façades. It uses structural, color, orientation, and form features to cluster close-up and full-façade images without supervision. The following discussion interprets the visual–semantic structures revealed by those clusters and considers their potential significance for understanding architectural imagery, organizing architectural information, and supporting generative-AI-assisted design.

4.1. Differentiation of Visual Semantics Across Observation Scales

The clustering results show that observation scale changes how visual information from active responsive building façades is organized. Close-up groupings in Set A primarily reflect material texture, component units, connection methods, states, and repetitive order. Differences among these clusters therefore arise mainly from local construction and repeated forms. Full-façade images in Set B are also influenced by massing, proportion, background, lighting, and building–environment relationships, producing more continuity and overlap. Agreement between the two partitions is low (ARI = 0.033, NMI = 0.194), but their association exceeds chance (p = 0.0065). Local construction and the overall façade therefore have a limited, nonrandom, selective relationship.

On this basis, the study uses tectonic visual description to characterize local façade composition, component states, and repetitive order, and contextual visual description to characterize their appearance within the building and its environment. Neither description can replace the other; instead, they provide complementary levels of interpretation. The visual semantics of an active responsive façade are shaped jointly by variable components, material expression, and environmental relationships and cannot be covered by a single style label [14,15].

The visual semantics formed jointly by local component states and the broader environmental context can also be understood through the concept of affordance. Gibson emphasizes real relational properties between the environment and an actor's capabilities, whereas Norman focuses on how users perceive possible actions through visible cues such as appearance, position, and movement [11,47]. The opening and closing of components, folded forms, and material changes identified in static images are therefore better understood as perceptual cues to potential responsive states.

4.2. Interpretive Translation from Visual Features to Visual–Semantic Prototypes

The study develops a three-layer translation from visual features to cluster structure and then to interpretive labels. This approach accords with Simon's account [15] of organizing complex information through classification, abstraction, and hierarchy. Its purpose is to create a comparable and interpretable analytical structure, not to prescribe an architectural typology in advance.

Computational features represent differences in structure, color, orientation, and form. K-means identifies visual similarities. Researchers then derive interpretive labels from cluster members, representative samples, feature profiles, and project documentation. This process allows similarities identified through computer vision to be compared cautiously with architectural concepts such as material, construction, form, visible component states, and environmental imagery [28–30,33,35,39]. Compared with research that uses clustering to organize architectural precedents [48], this study also compares the grouping and semantic translation of the same cases at close-up and full-façade scales. The approach reduces the constraint of prior style labels, but clustering still depends on the representation. Changes in PCA dimensionality alter the working value of K and some membership assignments. K = 6 is therefore an analytical granularity in an exploratory two-dimensional PCA space rather than a dimension-invariant natural category [49]. Robustness analysis further shows that cluster structure changes with feature-group weighting and visual representation. K = 6 should consequently be understood as a working granularity under the present feature system and analytical settings, not as a fixed category. Validation against independent architectural attributes found a significant association only between tectonic type and clustering at both scales. The visual clusters therefore reflect differences in tectonic organization in part, but they cannot be equated directly with response mechanism, material type, or climate conditions.

K-means identifies groups of similar observations; it does not define a vocabulary of shapes or transformation rules. The prototypes in this study are therefore not equivalent to a shape grammar [50]. For other building types, cultural regions, or image sources, the features, standardization, cluster granularity, and semantic interpretation must be validated again [51]. What can be transferred from this study is the methodological procedure rather than its present results.

Compared with the previous passive-façade study [23], extending the workflow to active responsive façades and paired component–whole-façade scales increased its descriptive scope and made scale-dependent visual relationships directly observable. The dual-scale analysis reveals how some visual features persist or change between component details and the whole façade, providing additional information for precedent organization and cross-scale semantic analysis. However, this extension did not produce a more stable façade taxonomy. The low cross-scale agreement and the disappearance of the association under feature-group equal weighting indicate that the active, dual-scale setting is more sensitive to image representation and feature-weighting choices. Because the two studies used different samples, feature representations, and evaluation procedures, their effectiveness cannot be compared through a single performance measure. Overall, the extension strengthens the method for exploratory precedent organization, scale-relation analysis, and semantic prompt construction, while its capacity to support stable clustering or a fixed taxonomy remains limited.

4.3. Application Potential and Transfer Pathways of Visual–Semantic Prototypes

The controlled generation experiment provides preliminary evidence for applying visual–semantic prototypes. Both independent batches showed positive effects in the same direction for target–text similarity and visual–prototype similarity. The class-specificity analysis also showed improved prototype identification, but the Top-1 rate remained below 50%. The evidence therefore supports improved visual–semantic consistency but is not sufficient to establish stable class-specific control. Visual–semantic prototypes are better treated as candidate semantic mediators between image analysis and generative design than as generative controllers with fixed class boundaries.

In generative architectural design, prompts act as semantic mediators between design intent and image generation [3,52,53]. Active responsive architecture involves materials, component states, form, and environmental relationships simultaneously, and broad language can leave the dimensions of control unclear. The visual–semantic prototypes developed here can therefore provide a semantic source for structuring prompts and organizing design variables. They still need to be combined and adjusted for the particular design task rather than used directly as fixed generation instructions.

From the perspective of second-order cybernetics, the semantic intermediate layer can also serve as a cognitive interface through which an observer participates in generative feedback. It connects feature extraction, label construction, prompt generation, output feedback, and semantic revision, turning a one-way operation into an ongoing human–AI cognitive interaction [54–56]. This process differs from the physical control loop of an actual responsive façade [57]. Its central concern is the designer's continuing interpretation of generated results and revision of their semantics. Accordingly, the workflow is intended primarily for precedent organization, early conceptual design, façade-language exploration, and human–AI ideation, rather than for direct performance prediction, constructability assessment, or responsive-control design.

The method can also complement existing architectural information frameworks. Conventional frameworks primarily organize geometry, materials, construction, performance, and spatial relationships [58]. The semantic mapping in this study adds an interpretable visual–semantic layer that connects image features with generative AI. Its transferable value lies in the sequence from visual features to cluster structure, semantic mapping, and generative feedback, rather than in direct reuse of the present K = 6 categories or specific semantic labels.

4.4. Limitations and Directions for Future Research

The first limitation concerns the sample. Few active responsive buildings have been completed, and public images vary in documentation quality, viewing angle, clarity, and scene consistency [9,12]. To broaden coverage as far as possible, the study assembled a purposive sample of 75 cases from the open web, architectural media, design firms, project platforms, competitions, and research projects. This sample reflects the information currently available and cannot represent all active responsive buildings [24]. Differences in completion status, photography or rendering method, and source may also affect the representation, causing some clusters to reflect differences in image presentation. For the exploratory objectives of visual feature analysis, clustering tendencies, and semantic mapping, the present sample nevertheless supports a preliminary analysis within its defined scope. The masking analysis showed that cluster assignments were largely stable, but it could not remove all effects of photographic conditions. Lighting differences, façade reflections, and background visible through transparent or perforated envelopes may still affect the visual features.

The second limitation concerns static images. The defining characteristic of active responsive architecture is a dynamic process that changes with environmental conditions and patterns of use [7,9]. Because the present analysis relies mainly on static two-dimensional images, it cannot fully represent the continuous transition of a façade from its initial to its responsive state. Two-dimensional images nevertheless remain important visual media in architectural design, case dissemination, competition presentations, conceptual comparison, and generative design. They record component states, envelope organization, and formal features at particular moments [3,13]. Although static images cannot replace analysis of dynamic processes, they can capture key visual states of an active responsive building at a given moment and provide a computable basis for analyzing its formal features, envelope organization, and semantic expression. Future research should incorporate video, time-series data, motion trajectories, and synchronized environmental data.

The third limitation is dependence on representation and weighting. Equal weighting at the feature-group level and same-pipeline validation with ResNet-50 show that cluster structure changes with feature composition and the form of representation. The present K = 6 solution and its semantic prototypes should therefore not be treated as fixed categories. Future studies could test their stability using larger samples, multiview feature fusion, and consensus clustering.

The fourth limitation concerns generative validation. Semantic labels improved CLIP alignment and prototype consistency, but stable class-specific control has not been demonstrated. The next stage should expand the case library of active responsive architecture, introduce blinded expert evaluation, and conduct focused research on generative-AI prompt construction for active responsive buildings.

5. Conclusions

Using 150 images forming 75 matched pairs from 75 active responsive building cases, this study developed a dual-scale visual analysis framework for close-up components (Set A) and full façades (Set B). Set A primarily reflects component form, materials, connection methods, and modular order. Set B additionally integrates massing, composition, lighting, and environmental relationships. Although agreement between the scale-specific partitions is low (ARI = 0.033, NMI = 0.194), their association is significantly greater than chance ($p = 0.0065$). Local construction and overall context in active responsive building façades therefore form selective and mutually complementary visual representations.

Methodologically, the study establishes a traceable translation from visual features through cluster structure and interpretive labels to generative steering. In a two-dimensional PCA space, the 43 interpretable visual features support K = 6 as a working granularity for cross-scale comparison. Robustness and interpretive limits were examined further through feature-group equal weighting, feature ablation, same-pipeline comparison with ResNet-50, and validation against independent architectural attributes. The results show clear dependence of cluster structure on representation and weighting, although each analytical specification retained a degree of internal stability. Tectonic type was significantly associated with clustering at both scales, suggesting that the resulting visual partitions have some architectural and tectonic relevance rather than reflecting visual groupings alone.

The controlled generation experiment further demonstrated the application potential of the visual–semantic intermediate layer. The Semantic Prompt increased both target–text similarity and visual–prototype similarity in two independent batches; the pooled increase in visual–prototype similarity was 0.0339 ($p = 0.0215$). Analysis against six same-scale prototypes also showed a significant improvement in the margin between the own prototype and the strongest competitor, while the Top-1 identification rate increased from 13.9% to 44.4%. Structured semantic labels therefore improved visual–semantic alignment in the generated images and strengthened their tendency to approach the target visual prototype. The evidence is not yet sufficient, however, to demonstrate stable class-specific control.

Taken together, the principal contribution of this study is not a fixed 12-category semantic system. It is a traceable and testable visual–semantic analytical process that makes its dependence on representation explicit. The method provides a structured semantic basis for organizing visual information and constructing generative-AI prompts for active responsive architecture. Future research should combine larger samples, multiple generative models, independent evaluation by architectural experts, and dynamic time-series data to test the robustness, transferability, and architectural validity of the visual–semantic method in generative design for active responsive buildings.

Author Contributions: Conceptualization, R.L. and S.-Y.H.; methodology, R.L.; software, R.L.; validation, R.L. and Y.L.; formal analysis, R.L.; investigation, R.L.; data curation, R.L.; writing—original draft preparation, R.L.; writing—review and editing, S.-Y.H. and Y.L.; visualization, R.L.; supervision, S.-Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The case metadata, visual feature matrices, and cluster membership data used in this study are publicly available in the GitHub repository: https://github.com/Rhon0410/responsive-facade-visual-semantic-data (accessed on 13 September 2026).

Acknowledgments: OpenAI ChatGPT (GPT-5.6) was used to assist with search-term expansion and to suggest candidate wording for cluster-level semantic descriptions. All search terms and semantic labels were reviewed, verified, and finalized by the authors using source literature, images, quantitative features, and project documentation. OpenAI GPT Image 2 (gpt-image-2) was used to generate images for the controlled experiment and to assist with non-data conceptual diagrams. All AI-generated outputs were reviewed and edited by the authors. Generative AI was not used

for research design, statistical analysis, or the formulation of conclusions. The authors take full responsibility for the content and accuracy of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jang, S.; Roh, H.; Lee, G. Generative AI in architectural design: Application, data, and evaluation methods. *Autom. Constr.* **2025**, *174*, 106174. [CrossRef]
2. Li, C.; Zhang, T.; Du, X.; Zhang, Y.; Xie, H. Generative AI models for different steps in architectural design: A literature review. *Front. Archit. Res.* **2025**, *14*, 759–783. [CrossRef]
3. Paananen, V.; Oppenlaender, J.; Visuri, A. Using text-to-image generation for architectural design ideation. *Int. J. Archit. Comput.* **2024**, *22*, 458–474. [CrossRef]
4. Huang, S.-Y. The Connectionist Turn: How Contemporary Generative AI Reshapes Architectural Rationality. *Architecture* **2025**, *5*, 132. [CrossRef]
5. Odiah, A.; Gosling, S.D. Laying the foundations for using generative AI images in architectural research: Do images convey the intended spaces and ambiances? *Archit. Intell.* **2024**, *3*, 35. [CrossRef]
6. Sommese, F. *Re-Thinking the Building Envelope: Lessons from Nature in the Era of Climate Change*; Springer: Cham, Switzerland, 2025.
7. Loonen, R.C.G.M.; Trčka, M.; Cóstola, D.; Hensen, J.L.M. Climate adaptive building shells: State-of-the-art and future challenges. *Renew. Sustain. Energy Rev.* **2013**, *25*, 483–493. [CrossRef]
8. Tabadkani, A.; Roetzel, A.; Li, H.X.; Tsangrassoulis, A. A review of occupant-centric control strategies for adaptive facades. *Autom. Constr.* **2021**, *122*, 103464. [CrossRef]
9. Usman, U.; Yaro, A.I.; Asfour, O.S. Responsive building façades: A systematic review of technologies, challenges, and future directions. *Archit. Eng. Des. Manag.* **2026**, *online ahead of print*. [CrossRef]
10. Zhangabay, N.; Tursunkululy, T.; Ibraimova, U.; Abdikerova, U. Energy-efficient adaptive dynamic building facades: A review of their energy efficiency and operating loads. *Appl. Sci.* **2024**, *14*, 10979. [CrossRef]
11. Norman, D.A. Affordance, conventions, and design. *Interactions* **1999**, *6*, 38–43. [CrossRef]
12. Cocho-Bermejo, A. Adaptive architectural facades: Review 1985–2024. A comparative analysis of Media-TIC, the Arab Institute, and Al Bahar Towers dynamic facades approaches. *Nexus Netw. J.* **2025**, *27*, 925–945. [CrossRef]
13. Starzyńska-Grześ, M.B.; Roussel, R.; Jacoby, S.; Asadipour, A. Computer vision-based analysis of buildings and built environments: A systematic review of current approaches. *ACM Comput. Surv.* **2023**, *55*, 1–25. [CrossRef]
14. Kelbaugh, D. Typology—An architecture of limits. *Archit. Theory Rev.* **1996**, *1*, 33–52. [CrossRef]
15. Simon, H.A. *The Sciences of the Artificial*, 3rd ed.; MIT Press: Cambridge, MA, USA, 1996.
16. Arauz, R.; Filipov, E.; Fascetti, A.; Clifford, D.T.; Brigham, J.C. Evaluation of a Kirigami-inspired double-skin adaptive façade for natural ventilation and solar harvesting to enhance indoor environment and energy performance. *Energy Build.* **2024**, *324*, 114927. [CrossRef]
17. Niazy, D.; Metwally, E.A.; Rifat, M.; Awad, M.I.; Elsabbagh, A. A conceptual design of circular adaptive façade module for reuse. *Sci. Rep.* **2023**, *13*, 20552. [CrossRef]
18. Horvath, A.-S.; Pouliou, P. AI for conceptual architecture: Reflections on designing with text-to-text, text-to-image, and image-to-image generators. *Front. Archit. Res.* **2024**, *13*, 593–612. [CrossRef]
19. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
20. Xu, F.; Wong, M.S.; Zhu, R.; Heo, J.; Shi, G. Semantic segmentation of urban building surface materials using multi-scale contextual attention network. *ISPRS J. Photogramm. Remote Sens.* **2023**, *202*, 158–168. [CrossRef]
21. Qin, W.; Chen, L.; Zhang, B.; Chen, W.; Luo, H. NeoDescriber: An image-to-text model for automatic style description of neoclassical architecture. *Expert Syst. Appl.* **2023**, *231*, 120706. [CrossRef]
22. Law, S.; Valentine, C.; Kahlon, Y.; Seresinhe, C.I.; Tang, J.; Morad, M.G.; Fujii, H. Generative AI for architectural façade design: Measuring perceptual alignment across geographical, objective, and affective descriptors. *Buildings* **2025**, *15*, 3212. [CrossRef]
23. Liu, R.; Ismail, A.I.B.; Lu, Y. Research on enhancing the generalization ability of generative AI based on K-means clustering algorithm: Text-to-image task for wind-responsive passive building façade. In *Proceedings of the 2025 International Conference on Generative Artificial Intelligence and Digital Media (GADM 2025)*; Association for Computing Machinery: New York, NY, USA, 2025; pp. 140–148. [CrossRef]
24. Palinkas, L.A.; Horwitz, S.M.; Green, C.A.; Wisdom, J.P.; Duan, N.; Hoagwood, K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Adm. Policy Ment. Health Ment. Health Serv. Res.* **2015**, *42*, 533–544. [CrossRef]
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2016; pp. 770–778. [CrossRef]

26. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2019; Volume 32, pp. 8024–8035. Available online: https://proceedings.neurips.cc/paper_files/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html (accessed on 13 September 2026).
27. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
28. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 4th ed.; Pearson: Harlow, UK, 2018.
29. Bakurov, I.; Buzzelli, M.; Schettini, R.; Castelli, M.; Vanneschi, L. Structural similarity index (SSIM) revisited: A data-driven approach. *Expert Syst. Appl.* **2022**, *189*, 116087. [CrossRef]
30. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
31. Xie, J.; Li, M.; Wu, J.; Zhang, X.; Zhang, J. Semantic segmentation of building façade materials and colors for urban conservation. *npj Herit. Sci.* **2025**, *13*, 378. [CrossRef]
32. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *PAMI-8*, 679–698. [CrossRef]
33. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2005; Volume 1, pp. 886–893. [CrossRef]
34. Wang, B.; Zhang, J.; Zhang, R.; Li, Y.; Li, L.; Nakashima, Y. Improving facade parsing with vision transformers and line integration. *Adv. Eng. Inform.* **2024**, *60*, 102463. [CrossRef]
35. Hu, M.-K. Visual pattern recognition by moment invariants. *IRE Trans. Inf. Theory* **1962**, *8*, 179–187. [CrossRef]
36. Mahalanobis, P.C. On the generalised distance in statistics. *Proc. Natl. Inst. Sci. India* **1936**, *2*, 49–55.
37. Breunig, M.M.; Kriegel, H.-P.; Ng, R.T.; Sander, J. LOF: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*; Association for Computing Machinery: New York, NY, USA, 2000; pp. 93–104. [CrossRef]
38. Arthur, D.; Vassilvitskii, S. k-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2007; pp. 1027–1035.
39. Lloyd, S.P. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [CrossRef]
40. Caliński, T.; Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **1974**, *3*, 1–27. [CrossRef]
41. Davies, D.L.; Bouldin, D.W. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *PAMI-1*, 224–227. [CrossRef]
42. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
43. Hubert, L.; Arabie, P. Comparing partitions. *J. Classif.* **1985**, *2*, 193–218. [CrossRef]
44. Vinh, N.X.; Epps, J.; Bailey, J. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *J. Mach. Learn. Res.* **2010**, *11*, 2837–2854.
45. Good, P.I. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*, 3rd ed.; Springer: New York, NY, USA, 2005. [CrossRef]
46. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*; PMLR: Cambridge, MA, USA, 2021; Volume 139, pp. 8748–8763. Available online: https://proceedings.mlr.press/v139/radford21a.html (accessed on 13 September 2026).
47. Gibson, J.J. *The Ecological Approach to Visual Perception*; Houghton Mifflin: Boston, MA, USA, 1979.
48. Alymani, A.; Jabi, W.; Corcoran, P. Machine learning methods for clustering architectural precedents: Classifying the relationship between building and ground. In *Anthropologic: Architecture and Fabrication in the Cognitive Age—Proceedings of the 38th eCAADe Conference*; TU Berlin: Berlin, Germany, 2020; Volume 1, pp. 643–652. Available online: https://orca.cardiff.ac.uk/id/eprint/134938/ (accessed on 13 September 2026).
49. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
50. Stiny, G.; Gips, J. Shape grammars and the generative specification of painting and sculpture. In *Information Processing 71: Proceedings of the IFIP Congress 1971*; North-Holland: Amsterdam, The Netherlands, 1972; pp. 1460–1465.
51. von Luxburg, U. Clustering stability: An overview. *Found. Trends Mach. Learn.* **2010**, *2*, 235–274. [CrossRef]
52. Ma, H.; Zheng, H. Text semantics to image generation: A method of building facades design base on Stable Diffusion model. In *Phygital Intelligence: Proceedings of the 5th International Conference on Computational Design and Robotic Fabrication (CDRF 2023)*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 24–34. [CrossRef]
53. Pektaş, Ş.T.; Sağlam, B. Semiotics-based prompt engineering for architectural text-to-image generation processes. *Estoa* **2025**, *14*, 121–135. [CrossRef]

54. Amershi, S.; Weld, D.; Vorvoreanu, M.; Fourney, A.; Nushi, B.; Collisson, P.; Suh, J.; Iqbal, S.; Bennett, P.N.; Inkpen, K.; et al. Guidelines for human–AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–13. [CrossRef]
55. Glanville, R. Try again. Fail again. Fail better: The cybernetics in design and the design in cybernetics. *Kybernetes* **2007**, *36*, 1173–1206. [CrossRef]
56. Pask, G. *Conversation Theory: Applications in Education and Epistemology*; Elsevier: Amsterdam, The Netherlands, 1976.
57. Wiener, N. *Cybernetics: Or Control and Communication in the Animal and the Machine*, 2nd ed.; MIT Press: Cambridge, MA, USA, 1961.
58. Sacks, R.; Eastman, C.M.; Lee, G.; Teicholz, P. *BIM Handbook: A Guide to Building Information Modeling for Owners, Designers, Engineers, Contractors, and Facility Managers*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2018. [CrossRef]