

Machine Learning-Based Classification and Prediction of Residential Building Architectural Styles

Authors: Xia Bing, Shi Hui, Chen Sichong, Chen Jiamei, Li Xin, Xia Bing

Date: 2020-01-08T17:04:19+00:00

Abstract

[Purpose] This paper proposes a method for the classification and prediction of residential property styles, building upon a review of existing research on architectural form styles.

[Method] Through structured interviews and questionnaire surveys with front-line designers and project planners, the study identifies key formal elements that influence architectural style, as well as site economic factors that may affect style positioning. Based on machine learning (including cluster analysis, correspondence analysis, and discriminant analysis, etc.), data analysis is conducted on 372 cases of new residential properties in Hangzhou from the past decade.

[Results] The study reveals that currently common property styles can be classified into eight major categories; the most important formal elements for style classification are the presence or absence of curves in form, roof form, and color tone richness; while building height is the site economic factor with the greatest influence on residential style. Inputting the selected five site economic factors into a neural network model for training to predict formal variables and style categories yields an average accuracy of 77.2%.

[Limitations] If the machine learning classification methods adopted in this study could be further combined with and validated against research on human perception of different styles, the conclusions would be more comprehensive and possess greater practical application value.

[Conclusion] The conclusions of this study can assist project designers and planners in enhancing the scientific rigor of architectural style selection and positioning.

Full Text

Research on Style Classification and Prediction of Residential Buildings Based on Machine Learning

Xia Bing¹, Shi Hui¹, Chen Sichong¹, Chen Jiamei¹, Li Xin²

¹ Architecture Department, Zhejiang University, Hangzhou 310058, China

² College of Urban Design, Wuhan University, Wuhan 430072, China

Abstract:

[Objective] Based on a review of existing architectural style research, this paper proposes a method for the classification and prediction of residential building styles. **[Methods]** Through structured interviews and questionnaire surveys of front-line designers and project planners, the study identifies key morphological elements that influence architectural style classification and positioning, as well as site economic factors that may affect style determination. Using machine learning techniques (including cluster analysis, correspondence analysis, and discriminant analysis), the study analyzes data from 372 newly built residential projects in Hangzhou over the past decade. **[Results]** The research finds that contemporary residential styles can be categorized into eight major types. The most important morphological elements for style classification are the presence of curved forms, roof type, and color richness, while building height is the most influential site economic factor affecting residential style. When five selected site economic factors are used to train a neural network model for predicting morphological variables and style categories, the average accuracy reaches 77.2%. **[Limitations]** If the machine learning classification method employed in this research could be further combined with and validated against studies of human perception of different styles, the conclusions would be more comprehensive and practically applicable. **[Conclusions]** The findings of this study can help project designers and planners improve the scientific basis for architectural style selection and positioning.

Keywords: Architectural style; Machine learning; Classification; Prediction; Residential buildings; Statistical analysis

Style serves as the signature of architecture, reflecting its morphological characteristics and cultural connotations [1][2][3][4], and constitutes an essential aspect of architectural design. Architectural style manifests across multiple dimensions, including exterior form, spatial composition, color schemes, material selection, and interior environments [5][6][7][8]. Among these, the external stylistic expression not only reveals a building's function, value, and cultural background [9], but also represents a critical element through which people perceive and evaluate cities [10]. Consequently, in residential project planning, the positioning of architectural appearance and stylistic expression is of paramount importance, requiring comprehensive consideration of historical context, regional characteristics, environmental coordination, and other factors.

However, in practice, style determination remains complex and ambiguous. First, the concept of style is inherently subjective [11][12], and the question of how specific morphological elements combine to constitute a particular style, as well as how different style categories are distinguished, lacks complete scientific rigor in architectural design. In other words, both the constituent elements of style and classification methods involve uncertainty [13]. Second, stylistic expression is influenced by numerous external conditions [14][15], yet no scientific methodology has been established for selecting appropriate style types based on different design constraints. Moreover, in real design scenarios, styles are often hybridized. Residential building styles, in particular, differ from historical styles, appearing more flexible, diverse [16], and heavily influenced by commercial factors [17]. Many projects fail to exhibit characteristics consistent with their nominal style [18][19], creating confusion that hinders professional communication and easily leads to public misunderstanding. To address this, we need to propose a new research approach to understand the relationships between morphological elements and different styles, and to assist designers and planners in creating distinctive and appropriate architectural styles under specific conditions.

The objectives of this research are: (1) to identify key morphological variables that constitute residential building styles; (2) to propose a classification method for residential commercial building styles and compare different styles; (3) to identify key site economic factors that influence the determination of residential commercial building styles and assess their correlation with architectural styles and stylistic variables; and (4) to establish a predictive model for forecasting or positioning residential building styles.

2. Literature Review

2.1 Overall Analysis of Architectural Style Research

Although style research constitutes an important component of architectural studies, relatively few scientific papers address this topic. Using the keywords “architecture” and “style,” we retrieved 489 journal articles and conference papers from the Web of Science database published before 2019. Employing Citespace software [20] for co-word analysis, combined with critical reading of key literature, we examined research frontiers, evolutionary paths, and content in architectural style studies. The analysis reveals [Figure 1: see original paper] that research hotspots concentrate on “style characteristics” and “style history,” with primary content including historical background, morphological features, analytical comparisons, and subjective evaluations of styles. Methodologically, studies predominantly employ historical, sociological, typological, and morphological approaches. Overall, research topics remain fragmented, case data sources are relatively singular, quantitative research is scarce, and insufficient attention is paid to style determination issues in design practice.

2.2 Historical Dimension of Architectural Style

The fundamental styles of architectural art are products of history [21][22], representing comprehensive expressions of culture, economy, and politics in a particular region during a specific historical period, and reflecting shared tendencies in formal creation among artists [23]. Basic classification methods for styles can be organized according to historical eras [24][25], cultural spheres [26][27], or geographic regions [28][29][30]. For instance, styles categorized by era include Classicism, Gothic, Renaissance, Rationalism, Art Deco, Modernism, and others; regional classifications encompass British, German, American, Nordic, Mediterranean styles, and so forth. Typical identification methods rely on differences in spatial configuration, details, materials, and construction techniques, with basic morphological elements (such as roofs, doors, windows, columns, domes, towers, arches, etc.) serving as the most distinctive features for recognizing architectural styles [31][32][33].

Historical styles are often associated with specific aesthetic meanings. The accumulation of architectural images leads to style formation, with repetition deepening aesthetic perception. Meaning and form continuously reflect each other in practice, ultimately resulting in their inseparability [34][35]. Architectural typology [36][37][38], through typological deduction and diagrammatic thinking processes [39], enables the analysis, refinement, and recombination of stylistic elements [40][41], thereby achieving critical inheritance of traditional styles.

Nevertheless, such qualitative classification methods inevitably appear subjective and ambiguous. First, most styles overlap considerably, making differences difficult to distinguish. For example, both Minimalism and Modernism employ simple rectangular forms [42], while Baroque and Rococo both feature complex curves [43]. Second, architectural historians propose different theoretical and philosophical perspectives on form, such as Constructivism [44], Deconstructivism [45], and Functionalism [46], without unified standards. Moreover, these theories do not align with design methods in current engineering practice, as most residential designs do not follow specific stylistic conventions but rather borrow and recombine elements from different styles [47]; renowned architects also tend to create their own distinctive styles.

2.3 Perception of Architectural Style

Style cognition is fundamentally subjective. Studying human visual perception and evaluation of styles can help urban designers and planners better consider public preferences for architectural style positioning in urban development [48]. Through subjective evaluations of various architectural styles—examining cognitive factors such as clarity, complexity, friendliness, originality, robustness, and meaningfulness [49]—and combining these with cognitive responses like preference (like-dislike), arousal (awake-drowsy), naturalness (natural-artificial), and relaxation (relaxed-painful) [50], or even brainwave reactions [51], key visual elements and their meanings can be identified. Combining

these elements appropriately—through legibility, coherence, and harmonious repetition [52][53]—can enhance the image characteristics of a development project or district. Zhao and Xu [54][55] determined urban architectural styles through a process of selecting influence factors, assigning factor weights, and weighted comprehensive overlay, transforming conscious individual perceptions of architectural style into an unconscious objective existence at the group level.

2.4 Rational Cognition of Architectural Style

Unlike perceptual evaluation studies, shape grammar provides an abstract mathematical explanation of style from a geometric perspective [56]. Stiny and Mitchell [57] defined style analysis as a mathematical set problem. By establishing logical rules for combining stylistic elements (grammar), formal grammars can not only create new designs but also evaluate style similarity. Duarte's study of Siza's work demonstrates this feasibility [58]. Other researchers have developed similar analytical techniques, such as using fractal geometry box-counting methods to assess visual complexity in building facades [59][60][61][62][63], applying graph theory principles and betweenness centrality metrics to reveal hierarchical structures controlling facade systems [64], and utilizing space syntax indicators to compare similarities in building plan layouts across different architectures [65]. Although mathematical analysis models and tools such as fuzzy inference systems [66], neural networks [67], genetic algorithms [68], and region segmentation algorithms [69] have been applied to morphological feature analysis—producing results independent of architects' experience and intuition that possess auxiliary reference value—style understanding and evaluation ultimately depend on human subjective judgment. How to combine mathematical models to aid our understanding of style and integrate subjective judgment with objective logic represents an important issue in style research. For example, Banaei et al. [70] manually labeled interior space images and used AI algorithms to group them, achieving classification and comparison of interior architectural styles from different historical periods. Han [71] explored computational methods for measuring morphological similarity through expert-weighted similarity models for image comparison.

3. Methodology

3.1 Research Framework

Compared with previous studies, this research's innovation lies in its departure from the operational level of architectural morphological design practice, comprehensively employing machine learning techniques to construct a systematic research framework and propose methods for analyzing, comparing, and predicting residential building styles that are diverse and difficult to define.

The research first reviews existing literature to extract commonly used morphological elements, then uses questionnaire surveys to rank and filter the importance of these elements and site economic factors. Based on 372 residential

cases built in Hangzhou between 2009-2019, the study forms style categories and their characteristics through variable assignment, two-step clustering, and correspondence analysis, then employs discriminant analysis to propose methods for identifying style categories. Additionally, through cross-tabulation analysis, site economic variables strongly correlated with style variables are used as dependent variables to construct a neural network model exploring the possibility of style prediction (Figure 2 [Figure 2: see original paper]).

3.2 Structured Interviews and Hypotheses

According to the research framework, the study conducted face-to-face or telephone interviews with 13 practitioners from five Hangzhou institutions (including architects from design institutes, project managers from real estate companies, and university faculty) between May and June 2019. Based on interview feedback (Table 1) and literature review, the research proposes the following basic hypotheses: (1) Residential design should endow the overall building form with certain stylistic characteristics; (2) The overall style of residential developments results from the combined effect of different morphological elements; and (3) Style formation derives not only from architects' or project planners' subjective concepts but is also potentially influenced by objective conditions.

Table 1. Examples of Structured Interview Questions and Summarized Responses

Question	Summarized Responses
Is style positioning important for developments? / Does style significantly influence residential design? / Do residences need clear, distinctive stylistic features?	Interviewees 1, 4, 5: Style positioning is very important; different styles convey different value orientations. Interviewees 3, 8: Style positioning requires careful consideration as it creates the first impression of a development. Interviewees 10, 12: Clear, striking, and cohesive stylistic features are advantageous for project marketing.
How is the overall style formed? / What techniques and strategies do you generally employ in stylistic design?	Interviewees 2, 3, 8, 9: Synthesize various visual morphological elements such as materials, details, and volumetric forms to create harmonious overall compositions. Interviewees 1, 8, 11, 12: Reference excellent cases, extract key morphological features and compositional rules of corresponding styles, then adapt and transform them.

Question	Summarized Responses
How is style determined? / What external factors influence style? / What factors are considered in style determination?	Interviewees 3, 4: Must integrate with development positioning, location, and surrounding environment. Interviewees 5, 6: Reference and borrow stylistic features from similar or previously developed projects, then make fine adjustments based on specific design conditions. Interviewee 9, 13: Jointly decided by designers and developers.

3.3 Questionnaire Survey

Based on existing literature, this study compiled 41 elements influencing residential style (Table 2) and, according to in-depth interview results, selected 18 economic factors affecting project style planning (Table 3). The study then conducted an importance survey of the preliminarily selected stylistic morphological elements and site economic elements. The questionnaire survey was administered from May 15 to June 10, 2019, and from July 3 to August 6, 2019. A total of 132 questionnaires were distributed, with 113 valid questionnaires retained after excluding invalid responses due to mismatched professional backgrounds or insufficient work experience, yielding an 85.6% validity rate. Participants comprised design professionals, real estate planning personnel, and non-professionals. The questionnaire measured element importance through a scoring system: “very important” = 2 points, “generally important” = 1 point, “uncertain” = 0 points. Elements scoring above the average for all three participant groups were designated as important.

Table 2. Preliminary Key Morphological Elements

Color and Material: Main material type, color temperature, color lightness, color richness, material richness, surface smoothness, surface transparency, surface reflectivity, obvious material contrast, double-skin facade

Details: Line form, line material, balcony form, balcony position, balcony shape, balcony railing form, window opening shape, window opening uniformity, window-to-wall ratio, presence of bay windows, presence of shading devices, shading device form, presence of columns, presence of complex moldings, presence of decorative patterns, presence of decorative frameworks

Form: Complexity of upper and lower components, plan symmetry, elevation symmetry, horizontal form graduation, vertical form graduation, form segmentation, roof form, roof complexity, corner treatment, obvious openings, ground-level pilotis, multi-story composition, presence of podium, podium connectivity

Table 3. Key Economic and Site Factors Influencing Style Positioning

Planning Indicators: Building combination form, site topography, floor area ratio, building density, height restrictions, green space ratio

Surrounding Environment: Surrounding building functions, surrounding architectural styles, surrounding building density, surrounding building spacing, presence of large green spaces or water bodies nearby, road hierarchy, transportation convenience, surrounding supporting facilities

Project Characteristics: Development timeline, project location, project grade positioning, unit type requirements, real estate company size

3.4 Data Collection

Using Python web scraping technology [72], we collected data and photographs of 1,024 newly built residential developments in Hangzhou from 2009-2019 on the Anjuke website [73], from which 390 were randomly selected as research subjects for manual annotation. Annotation was completed by three architecture professionals who received simple training. After excluding samples with incomplete data, the final sample totaled 372 cases. To ensure annotation objectivity, most options employed a clear binary classification. Detailed annotation content and assignment rules are shown in Table 4. Economic data such as price, location, floor area ratio, building height, density, and green space ratio were sourced from online platforms including the 5i5j website and Baidu Maps.

Table 4. Assignment Methods for Key Stylistic Morphological Elements

Category	ID	Element
Color	F4	Color richness
Details	D1	Decorative line directionality
Details	D7	Presence of complex moldings
Form	V1	Form (plan) complexity
Form	V2	Presence of curved form (plan)
Form	V3	Obvious segmentation
Form	V4	Varied roof forms
Form	V6	Corner treatment method

3.5 Two-Step Clustering

Compared with traditional clustering methods, two-step clustering (TSC) can simultaneously accommodate continuous and discrete variables, making it more suitable for studies with categorical variables. Second, algorithmic improvements enable faster clustering speeds, offering advantages when processing large samples. Third, TSC genuinely employs statistical measures as distance metrics for clustering while automatically recommending optimal cluster numbers based on statistical criteria, rendering classification results more reliable. The TSC calculation method consists of two main steps:

- 1) **Build a Cluster Feature (CF) tree:** The first record in the dataset is placed in a leaf node originating from the root. Using distance measure-

ment as a similarity criterion, subsequent records are merged with existing nodes or generate new nodes based on similarity, recursively building the cluster feature tree. The distance measurement model employs log-likelihood similarity, calculated as follows [74]:

$$d(j, s) = \xi_j + \xi_s - \xi_{\langle j, s \rangle}$$

where $\xi_j = -N_j \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{jk}^2) + \sum_{k=1}^{K^B} \hat{E}_{jk} \right)$, with K^A representing the total number of continuous variables, K^B the total number of categorical variables, L_k the category number of the k th categorical variable, N_k the number of records in class k , $\hat{\sigma}_k^2$ the estimated variance of the k th continuous variable in the entire dataset, $\hat{\sigma}_{jk}^2$ the estimated variance of the k th continuous variable in class j , \hat{E}_{jk} the estimated mean of the k th continuous variable in class j , N_{jkl} the number of records in class j for the l th category of the k th categorical variable, N_{kl} the number of records for the l th category of the k th categorical variable, $d(j, s)$ the distance between classes j and s , and $\langle j, s \rangle$ the index of the combined class of j and s .

- 2) **Combine leaf nodes using agglomerative clustering:** This produces a set of clustering solutions with different numbers of clusters. The Bayesian Information Criterion (BIC) is then used to compare these solutions and automatically select the optimal number of clusters. The BIC is calculated as [75]:

$$\text{BIC} = -2 \log L(\hat{\theta}) + m_j \log N$$

where N represents the number of records in the dataset, $\sum \zeta$ is the maximum value of the likelihood function, and m_j is the number of model parameters.

3.6 Correspondence Analysis

Correspondence Analysis explains the intrinsic relationships between variables by analyzing interaction summary data composed of attribute variables, while simultaneously revealing differences among categories of the same variable and correspondence relationships between different variables. The advantage of this method becomes more pronounced as the number of variable categories increases. The main calculation steps are:

First, given n samples with m observed variables each, the original data matrix is $(x_{ij})_{m \times n}$, where x_{ij} represents the j th variable value of the i th sample.

Second, row sums X_i , column sums X_j , and total sum T are calculated.

Third, the probability matrix P of the original data is computed.

Fourth, the data transformation matrix Z is calculated.

Fifth, the covariance matrices for variables and samples are computed.

Sixth, **R-type factor analysis**: Calculate eigenvalues of the variable covariance matrix, $\lambda_1 > \lambda_2 > \dots > \lambda_m$, typically selecting the first k eigenvalues with cumulative contribution $> 70\%$ - 90% , compute corresponding eigenvectors u_1, u_2, \dots, u_k , and obtain the R-type factor loading matrix $F = [F_1, F_2, \dots, F_k]$, where $F_k = u_k \sqrt{\lambda_k}$. **Q-type factor analysis**: For the first k eigenvalues from R-type analysis, compute corresponding eigenvectors of matrix B , where $v_1 = Zu_1, v_2 = Zu_2, \dots, v_k = Zu_k$, then normalize them to obtain the Q-type factor loading matrix $G = [G_1, G_2, \dots, G_k]$, where $G_k = v_k \sqrt{\lambda_k}$.

Seventh, plot on two-dimensional factor axes, simultaneously displaying both samples and variables on the same diagram, reflecting both R-type and Q-type analysis results.

3.7 Discriminant Analysis

Unlike cluster analysis, discriminant analysis establishes discriminant functions based on known classification criteria to determine which category a research object belongs to. Fisher's discriminant method imposes no special requirements on population distribution or variance. Its fundamental concept involves sequentially extracting canonical variables, using them to calculate centroid coordinates of each category in low-dimensional space, establishing discriminant functions to compute coordinate values for each sample, and finally making classification judgments based on proximity to these centroids. The general form of Fisher's linear discriminant function is:

$$y = c_{1x}1 + c_{2x}2 + \dots + c_{kx}k$$

where y is the discriminant value, x_1, x_2, \dots, x_k are variable values reflecting object characteristics, and c_1, c_2, \dots, c_k are discriminant coefficients for corresponding variables, with k being the number of variables.

To effectively distinguish samples from different populations, between-group variance should be maximized while within-group variance minimized, meaning λ should be as large as possible:

$$\lambda = \frac{\sum_{i=1}^m n_i (\bar{y}_i - \bar{y})^2}{\sum_{i=1}^m n_i \sigma_i^2}$$

where \bar{x} is the overall mean vector, m is the number of populations, n_i is the sample size of the i th population, and \bar{y}_i and σ_i^2 are the sample mean and variance of discriminant values in the i th population, respectively.

3.8 RBF Neural Network

The Radial Basis Function Neural Network (RBF-NN) is an artificial neural network that employs local receptive fields to perform function mapping, drawing on biological knowledge of local regulation and overlapping receptive regions. It surpasses BP neural networks in approximation capability, classification ability, and learning speed. Since RBF neural networks can approximate any nonlinear function with fast learning convergence, they have been successfully applied to nonlinear function approximation, pattern recognition, and information processing [76]. Similar to BP networks, RBF neural networks are typical three-layer feedforward networks consisting of an input layer, hidden layer, and output layer [Figure 3: see original paper].

Let the RBF network input be a k -dimensional vector $x = [x_1, x_2, \dots, x_k]$. The transformation from input to hidden layer uses radial basis functions to achieve nonlinear transformation of input information. The most commonly used RBF is the Gaussian function:

$$\phi_k(x) = \exp\left(-\frac{\|x - c_k\|^2}{2\sigma_k^2}\right)$$

where parameter σ_k is the width of the RBF. The network output is the weighted sum of hidden layer outputs, with the linear mapping relationship:

$$y_j = \sum_{k=1}^N w_{jk} \phi_k(x), \quad j = 1, 2, \dots, m$$

where x is the input vector, ϕ_k is the RBF function, $\|\cdot\|$ denotes Euclidean norm, w_{jk} are network weights, N is the number of hidden layer neurons, and c_k is the RBF center in the input vector space.

4. Results

4.1 Style Feature Comparison

(1) Key Factors

According to survey results [Figure 4: see original paper][Figure 5: see original paper], the average importance scores for 41 stylistic morphological elements were 1.05 (architects), 1.20 (clients), and 1.11 (non-professionals). For all three respondent groups, 18 morphological elements scored above the average: “material type” (F1), “color temperature” (F2), “color lightness” (F3), “color richness” (F4), “decorative line directionality” (D1), “balcony form” (D2), “balcony shape” (D3), “window opening form” (D4), “bay window” (D5), “shading devices” (D6), “complex moldings” (D7), “decorative frameworks” (D8), “form complexity” (V1), “curved form” (V2), “form segmentation” (V3), “roof form”

(V4), “varied roof” (V5), and “corner treatment” (V6). Seven site economic elements scored above the average for both architects and clients: “project grade” (I), “completion time” (II), “location” (III), “building height” (IV), “floor area ratio” (V), “total building area” (VI), and “green space ratio” (VII).

(2) Style Categories

The 18 selected morphological elements were assigned values according to Table 4 rules and analyzed using two-step clustering. Table 5 presents the automatic clustering table, showing the process of selecting the number of clusters. According to two-step clustering rules for automatically determining optimal cluster numbers, the best solution requires not only the smallest BIC value but also relatively large rates of change in BIC and distance measures. Therefore, eight clusters represents the optimal choice, with a silhouette coefficient of 0.25, indicating acceptable classification (generally, silhouette coefficients > 0.2 suggest good clustering results). Among 372 cases, 35 were assigned to Cluster 1 (9.4%), 48 to Cluster 2 (12.9%), 37 to Cluster 3 (9.9%), 76 to Cluster 4 (20.4%), 82 to Cluster 5 (22.0%), 35 to Cluster 6 (9.4%), 30 to Cluster 7 (8.1%), and 29 to Cluster 8 (7.8%), with no missing values.

Table 5. Automatic Clustering

Number of Clusters	Schwarz’s Bayesian Criterion (BIC)	BIC Change	BIC Change Ratio	Distance Measure Ratio
...

Note: a. Change based on previous cluster number in table. b. Ratio change relative to two-cluster solution. c. Distance measure ratio based on current rather than previous cluster number.

(3) Style Characteristics

Figure 6 [Figure 6: see original paper] displays variable importance for clustering, where “curved form” (V2) is the most important clustering variable with an importance value of 1.0, followed by “roof form” (V4), “color richness” (F4), “color temperature” (F2), and “balcony form” (D2) with importance values of 0.95, 0.88, 0.73, and 0.72, respectively. Additional influential factors include “material” (F1), “corner treatment” (V6), “curved balcony” (D3), “color lightness” (F3), and “decorative framework” (D8).

Table 6 shows distribution values of different variable categories across the eight style types. Examining the three most important variables: In Cluster 1, 97.1% have no curved forms, all have flat roofs, and feature rich colors; Cluster 2 has 89.6% without curves, 95.8% with flat roofs, and 75% with monochromatic colors; Cluster 3 contains no curved forms, all have rich colors, with 97.3% featuring flat roofs; Cluster 4 has no curved forms, with 98.7% having flat roofs and rich colors; Cluster 5 cases all have monochromatic colors, with the vast

majority featuring non-curved, flat-roof forms (98.8%, 92.7%); Cluster 6 all have pitched roofs, mostly without curves, with half featuring combined colors; Cluster 7 has curved forms in 96.7% of cases, flat roofs in 90%, and mostly monochromatic colors; Cluster 8 has no curved forms, mostly flat roofs, and generally rich colors.

The study conducted multiple correspondence analysis between 18 morphological variables and style category variables obtained from two-step clustering to further analyze the eight style characteristics. Correspondence analysis simultaneously reflects variables and their relationships on a two-dimensional scatter plot, concentrating closely related category points and dispersing distant ones. Two principles guide interpretation: (1) Different categories of the same variable falling in similar directions and regions from the origin have similar properties; (2) Categories from different variables in similar directions and regions may be related. Following these principles, Figure 7 [Figure 7: see original paper] reveals: (1) Cluster 1's main feature is recessed balconies; (2) Cluster 2 primarily features glass curtain wall enclosures; (3) Clusters 3 and 4 mainly feature non-curved forms; (4) Cluster 5's main features are uniform window openings and regular plan forms; (5) Cluster 6 features pitched roofs and obvious shading devices; (6) Cluster 7 features curved forms and curved balconies; (7) Cluster 8 features varied roofs; (8) Regarding inter-category differences, Clusters 3, 4, and 5 are relatively similar, while they differ significantly from Clusters 3, 4, and 5.

(4) Style Comparison

The study performed discriminant analysis on different style categories using morphological elements as independent variables and tested discriminant function effectiveness. More variables are not necessarily better in discriminant analysis, as excessive variables increase computational load and may reduce overall discrimination effectiveness due to interference from less discriminating variables. Based on clustering results, 372 cases were used as training samples with corresponding group labels as grouping variables, employing stepwise discriminant analysis to select variables. After multiple steps, 15 variables entered the discriminant equations while 3 were excluded. Based on Table 6 results, eight discriminant functions were derived. For example, Cluster 1's discriminant function is:

$$f = 30.893F1 + 15.765F2 + 10.221F3 + 32.137F4 + 15.480D2 + 51.152D3 + 26.537D5 + 17.876D6 + 11.413D7 + 20.27$$

Variance testing of discriminant functions yielded $P < 0.005$, indicating that category differentiation by discriminant functions reaches significant levels. Combined with discriminant functions and scores of 15 indicators, the reliability of clustering results can be tested. For classification, sample indicator scores are substituted into the eight discriminant functions, with the largest function value determining style category. Overall, discriminant analysis group labels

are largely consistent with clustering results, with an 89.5% overall consistency rate between discriminant and clustering results for 372 developments. This discrimination effect is satisfactory given that all variables are discrete. This simultaneously demonstrates that the eight style categories from two-step clustering are reasonable, and Fisher discriminant functions can effectively identify categories for new developments.

Table 7. Classification Function Coefficients

											Fisher's			
											Linear			
											Discriminant			
Two-Step	F1	F3	F4	D2	D3	D5	D6	D7	D8	V1	V3	V5		
Cluster	Material	Color	Lightness	Richness	Form	Balance	Curve	Bay	ing	Complex	Form	Segment	Var-	nant
ID	Color	Lightness	Richness	Form	Balance	Curve	Bay	ing	Complex	Form	Segment	Var-	nant	Function
...	12.25468	115

The study selected standard cases of several popular styles from China's largest real estate development companies, analyzed their morphological characteristics, and used discriminant functions to calculate their relationships with two-step clustering categories. Table 8 shows that among common styles, Modern style belongs to Cluster 1, Modern Minimalist to Cluster 2, Modern Elegant to Cluster 3, International, International Elegant, and European Minimalist are relatively similar, all belonging to Cluster 4, Classical Minimalist to Cluster 5, European, Modern Chinese, European, and French styles to Cluster 6, and Modern Minimalist and Postmodern styles belong to Clusters 7 and 8, respectively.

Table 8. Morphological Analysis of Popular Styles

4.2 Style Prediction

(1) Correlation Between Morphological Elements and Site Economic Factors

Since most style morphological factors are categorical while site economic factors are continuous, cross-tabulation chi-square tests were used to analyze their relationships, with continuous variables converted to interval variables. Results are shown in Table 9 . Overall, site economic factors have some influence on morphological elements, with five factors significantly correlated: "completion time," "location," "building height," and "floor area ratio." Among seven site factors, "building height" and "completion time" most significantly influence style morphological elements and classification. Cross-tabulation analysis shows Cluster 6 style appears more frequently in low-rise and multi-story residences, while other style categories concentrate in high-rise buildings. Style categories 4 and 5 emerged later, while categories 7 and 8 appeared earlier. Among the

seven site factors, “project grade” and “total building area” show no correlation with style morphological elements or categories. Table 8 also reveals that different morphological elements are influenced by economic factors to varying degrees: roof form (V4) is most affected, primarily by “completion time,” “location,” “building height,” and “floor area ratio,” while bay window (D5) is almost unaffected by site conditions.

Table 9. Correlation Analysis Matrix of Economic Factors on Morphological Factors

...	.087	.072	.136	.000	.243	.000
-----	------	------	------	------	------	------

(2) Prediction Accuracy

Based on chi-square test results, five site economic elements—“completion time,” “location,” “building height,” “floor area ratio,” and “green space ratio”—were selected to predict residential development style morphological elements and style categories using a radial basis artificial neural network. Training and testing groups were split 80% and 20%, with neuron numbers and layers optimized automatically. Figure 11 [Figure 11: see original paper] shows that the average prediction accuracy for 18 elements and style categories is 77.2% (78.5% for training group, 75.9% for testing group), though predictive capability varies across elements. For binary categorical variables, prediction accuracy is relatively high for material (F1), balcony form (D3), and window form (D4), exceeding 90% for both training and testing groups. Conversely, site economic factors show lower predictive capability for color richness (F4), form complexity (V1), and roof variation (V5). Overall, using radial basis neural networks, site economic elements demonstrate certain predictive power for style types, with training and testing accuracies of 34.6% and 31.7%, respectively. Although not achieving high prediction rates, this conclusion still offers valuable reference for design practice given the inherent ambiguity and uncertainty of style determination.

Figure 11. Comparison of RBF Neural Network Prediction Accuracy

5. Discussion and Conclusion

5.1 Limitations

The main limitations and areas for improvement of this study are: First, the selection of morphological style factors and site economic factors partially relies on human judgment, introducing some subjectivity. Second, architectural forms are infinitely variable, and the different aspects of the same morphological element proposed in this study may not be exhaustive, and other classification methods are also feasible. Third, due to limited sample size, the style

classification and prediction methods and conclusions proposed require comparative studies to verify their applicability to other building types and regions. Fourth, since style understanding is closely related to human subjective perception, combining the machine learning classification method used in this study with research on human perception of different styles for validation would make conclusions more comprehensive and practically applicable. Furthermore, while facade morphology most significantly influences style, style expression extends beyond this. Space is the protagonist of architecture, and spatial organization represents the essential distinction between styles [77]. Therefore, integrating facade morphology with spatial volume to find optimal methods for distinguishing and determining architectural styles remains an issue requiring further exploration.

5.2 Future Directions

Understanding and recognition of architectural and urban styles are gradually deepening with technological and urban development. Several aspects merit attention:

First, the relationship between human subjective perception evaluation and objective composition of different morphological and spatial elements—whether objective stylistic differences also cause differences in human subjective perception—requires clarifying correlations between environmental psychological variables and spatial morphological variables.

Second, recent AI technology developments, particularly image recognition for style feature extraction, provide new methods for architectural style research. Although mathematical image parsing differs from generative design thinking in morphology, it can offer important reference value [78][79][80][81].

Third, architecture always exists within a specific environmental context, and urban style evolution and harmony result from the collective effect of every building's style. How to scientifically determine style and organize corresponding morphological elements within certain historical contexts remains an unresolved scientific question [82].

Style determination is currently an important yet ambiguous aspect of architectural design. Based on 372 residential development cases in Hangzhou, this study employs machine learning methods to scientifically explore style classification and prediction, yielding the following conclusions:

- 1) According to questionnaire results ranking 41 morphological elements and 18 site economic factors, 18 important morphological elements and 7 economic elements were identified.
- 2) Through two-step cluster analysis, the sample developments are appropriately divided into eight style categories. The most important morphological factors for classification are “presence of curved form,” “roof form,” and “color richness.” Among the eight categories, Category 5 is

the largest, characterized by monochromatic colors, predominantly non-curved flat-roof forms, uniform window openings, and regular plan forms.

- 3) Discriminant analysis reveals that among common styles, International, International Elegant, and European Minimalist styles are relatively similar, all belonging to Category 4; Classical Minimalist belongs to Category 5; European, Modern Chinese, European, and French styles belong to Category 6.
- 4) The study finds certain correlations between site economic factors and morphological elements, with “building height” having the greatest influence on morphological element selection, while “project grade” and “total building area” show no influence on morphological elements or style categories.
- 5) Finally, using five site economic factors as independent variables, a radial basis neural network model was established to predict 18 morphological elements and style categories, achieving an average accuracy of 77.2%.

References

- [1] John Milnes Baker. *American House Styles: A Concise Guide*[M]. Courtryman Press, 2018
- [2] Davidson Cragoe, Carol. *How to Read Buildings: A Crash Course in Architectural Styles*[M]. Rizzoli, 2008
- [3] John C. Poppeliers, S. Allen Chambers. *What Style Is It?: A Guide to American Architecture*. Wiley; Revised edition (October 6, 2003)
- [4] Nicole Bridge. *Architecture 101: From Frank Gehry to Ziggurats, an Essential Guide to Building Styles and Materials*. Adams Media, 2015
- [5] A. Trystan Edwards. *Style and Composition in Architecture - An Exposition of the Canon of Number Punctuation and Inflection*. Forbes Press, 2016
- [6] Sarah Cunliffe. *A Journey Through World Architectural Styles: From Classic Temples to Modern Skyscrapers*[M]. China Machine Press, 2015
- [7] Hiroyuki Suzuki. *Illustrated Chronology of Western Architectural Styles*[M]. Tsinghua University Press, 2013
- [8] Qian Zhengkun. *History of World Architectural Styles*[M]. Shanghai Jiao Tong University Press, 2005
- [9] Abbas Zadeh, F. and Sulaiman, A.B. (2010) Dynamic street environment. *Justice and Sustainability* 15(5): 433–452.
- [10] Nasar, J.L., Stamps III, A.E. and Hanyu, K. (2005) Form and function in public buildings. *Environmental Psychology* 25(2): 159–165.
- [11] Steen Eiler Rasmussen. *Experiencing Architecture*[M]. Beijing: Intellectual Property Publishing House, 2008
- [12] Sigfried Giedion. *Space, Time and Architecture: The Growth of a New Tradition, Fifth Revised and Enlarged Edition*. Harvard University Press, 2009
- [13] Davidson Cragoe, Carol. *How to Read Buildings: A Crash Course in Architectural Styles*[M]. Rizzoli, 2008

- [14] Leland M. Roth, Amanda C. Roth Clark. Understanding Architecture: its elements, history, and meaning[M]. Westview Press, 2013
- [15] Aldo Rossi. The architecture of the city[M]. MIT Press, 1984
- [16] Guangzhou Tangyi Culture Communication Co., Ltd. Fusion and Innovation of Styles 1[M]. Huazhong University of Science and Technology Press, 2013
- [17] Lorenzo García-Moruno, Julio Hernández-Blanco. A site planning approach for rural buildings into a landscape using a spatial multi-criteria decision analysis methodology[J]. 2013, 32: 108-118.
- [18] Wu Jiayi. Research on stylization of Collective Residence & Related Structural Application[D]. Beijing: Tsinghua University, 2011
- [19] Xu Ying, Yan Jinquan. Reflections on the Crisis of Chinese Urban Characteristics from a Global Perspective[J]. Urban Problems, 2006
- [20] Chen Caomei. CiteSpace: A Practical Guide for Mapping Scientific Literature[M]. Nova Science Pub Inc, UK ed., 2016
- [21] Jonathan Glancey. Architecture: A Visual History[M]. DK, 2017
- [22] Carol Davidson Cragoe. How to read buildings: a crash course in architectural styles. Rizzoli,
- [23] Eric Fernie. Romanesque architecture: the first style of the european age. Yale university press
- [24] Francis D.K. Ching, Mark M. Jarzombek, Vikramaditya Prakash. A global history of architecture[M]. Wiley, 2017
- [25] Cyril M. Harris. Illustrated dictionary of historic architecture[M]. Dover Publications, 1983
- [26] Leo Hansen. Culture and architecture: an integrated history[M]. Cognella academic publishing,
- [27] Sir Banister Fletcher. A History of Architecture on the Comparative Method: For the Student, Craftsman, and Amateur[M]. CreateSpace Independent Publishing Platform, 2018
- [28] Sarah Cunliffe. A Journey Through World Architectural Styles: From Classic Temples to Modern Ferris Wheels[M]. Beijing: China Machine Press, 2015
- [29] Wang Ruizhu. History of World Architecture[M]. Beijing: China Architecture & Building Press, 2018
- [30] Owen Hopkins. Architectural style: a visual guide. Laurence king publishing, 2014
- [31] Stephen Calloway, alan powers, elizabeth cromley. The elements of style: an encyclopedia of domestic architectural detail. Firefly Books, 2012
- [32] Wilfried Koch. Architectural Styles[M]. Shenyang: Liaoning Science and Technology Publishing House, 2006
- [33] Leland M. Roth, Amanda C. Roth Clark. Understanding Architecture: its elements, history, and meaning[M]. Westview Press, 2013
- [34] Robert Hillenbrand. Islamic Architecture: Form, Function, and Meaning[M]. New York: Columbia University Press, 2004
- [35] Aldo Rossi. The architecture of the city[M]. MIT Press, 1984
- [36] Rob Kerier. Town space[M]. Nanjing: Jiangsu Science and Technology Press, 2016

- [37] Tian Yinsheng. Urban Morphology, Architectural Typology, and Cities in Transition[M]. Beijing: Science Press, 2014
- [38] Gottfried Semper, Harry Mallgrave, Michael Robinson. Style in the technical and tectonic arts[M]. Getty Research Institute, 2004
- [39] White, Edward T., 1975. Concept Sourcebook: a Vocabulary of Architectural Forms. Architectural Media; Tucson, Ariz, USA.
- [40] Colin Rowe, Fred Koetter. Collage city[M]. MIT Press, 1984
- [41] Sam Jacoby. Typal and typological reasoning: a diagrammatic practice of architecture[J]. The Journal of Architecture, 2016(7): 938-963
- [42] Wang Shouzhi. History of Modern World Architecture[M]. Beijing: China Architecture & Building Press, 2012
- [43] Claudia Zanlun, Danila Tara. Baroque and Rococo Architecture[M]. Beijing: Beijing Art and Photography Publishing House, 2019
- [44] Frampton, Kenneth, Cava, John. Studies in tectonic culture: the poetics of construction in nineteenth and twentieth century[M]. MIT Press
- [45] Xue Enlun. Deconstructivism and Dynamic Composition: Exploration of Architectural Form and Space[M]. Beijing: China Architecture & Building Press, 2019
- [46] Gao Feng. Functionalist Architecture[M]. Tianjin: Tianjin University Press, 2009
- [47] Yang Feng. Breaking Through Style and Replication: Super High-rise and High-rise Buildings[M]. Nanjing: Jiangsu People's Publishing House, 2012
- [48] Askari, Amir Hossein; Dola, Kamariah Binti; Soltani, Soha. An evaluation of the elements and characteristics of historical building facades in the context of Malaysia[J]. URBAN DESIGN INTERNATIONAL, 2014(19): 113-124
- [49] Brown, G. and Gifford, R. (2001) Architects predict lay evaluations of large contemporary buildings: Whose conceptual properties? Environmental Psychology 21(1): 93-99.
- [50] Cubukcu, E. and Kahraman, I. (2008) Hue, saturation, lightness, and building exterior preference: An empirical study in Turkey comparing architects' and non-architects' evaluative and cognitive judgments. Color Research & Application 33(5): 395-405
- [51] Cheng Chezi, Li Huan. A comparative study of chinese and foreign architectural styles based on EEG Technology[J]. Huazhong Architecture, 2018(4): 39-42
- [52] Akalin, Aysu; Yildirim, Kemal; Wilson, Christopher; et al. USERS' EVALUATIONS OF HOUSE FACADES: Preference, Complexity and Impresiveness[J]. OPEN HOUSE INTERNATIONAL, 2010(35)
- [53] IMAMOGLU, Ç. 2000, Complexity, preference and familiarity: Architecture and non-architecture Turkish students' assessments of traditional and modern house façades. Journal of Environmental Psychology, 20(1): 5-16
- [54] Zhao Ning, Ge Dandong, Hua Chen. application and exploration of quantitative analysis methods for determining the architecture styles[J]. Journal of Zhejiang University (Science Edition), 2009, 36(1): 109-115
- [55] Xu Xiaoyan, Chen Haomiao, Gu Tingting, et al. Rational method for the orientation of urban architectural style[J]. Journal of Hefei University of

Technology, 2011(2): 282-285

[56] Li, Andrew I-Kang. Computing Style[J]. NEXUS NETWORK JOURNAL, 2011(13): 183-193

[57] Stiny and Mitchell. Pictorial and formal aspect of shape and shape grammars[M]. Birskauser,

[58] DUARTE, José Pinto. 2005. Towards the mass customization of housing: the grammar of Siza's houses at Malagueira. Environment and Planning B: Planning and Design 32, 3: 347–380.

[59] Samper, Albert; Herrera, Blas. A Study of the Roughness of Gothic Rose Windows[J]. NEXUS NETWORK JOURNAL, 2016(18): 397-417

[60] Ostwald, M. J. and Ö. Ediz, 2015. Measuring Form, Ornament and Materiality in Sinan's Kilic, Ali Pasa Mosque: an Analysis Using Fractal Dimensions. Nexus Network Journal 17(1): 15–22.

[61] Vaughan, J. and M. J. Ostwald, 2014. Measuring the Significance of Facade Transparency in Australian Regionalist Architecture: A computational Analysis of 10 Designs by Glenn Murcutt. Architectural Science Review 57(4): 249–259

[62] Lee, Ju Hyun; Ostwald, Michael J.; Gu, Ning. A Combined Plan Graph and Massing Grammar Approach to Frank Lloyd Wright's Prairie Architecture[J]. NEXUS NETWORK JOURNAL, 2017(19):

[63] Lorenz, Wolfgang E. Measurability of Loos' rejection of the ornament Using box-counting as a method for analysing facades//[C] 32nd International Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe). Northumbria Univ, Fac Engn & Environm, Dept Architecture & Built Environm, Newcastle upon Tyne, ENGLAND, SEP 10-12, 2014. FUSION: DATA INTEGRATION AT ITS BEST, VOL 1: 495-504

[64] Kenza Boussora. Formal Style of Medersas Buildings in North Africa[J]. Nexus Network Journal, 2015(17): 103-115

[65] Behbahani, Peiman Amini; Ostwald, Michael J.; Gu, Ning. A syntactical comparative analysis of the spatial properties of Prairie style and Victorian domestic architecture[J]. JOURNAL OF ARCHITECTURE, 2016(21):

[66] Arabacioglu, Burcin Cem, 2010. Using fuzzy inference system for architectural space analysis. Appl. Soft Comput. 10(3), 926–937.

[67] Hsiao, Shih-Wen, Tsai, Hung-Cheng, 2005. Applying a hybrid approach based on fuzzy neural network and genetic algorithm to product form design[J]. Int. J. Ind. Ergon. 35(5), 411–428.

[68] Pal, Pralay, Tigga, A.M., Kumar, A., 2005. Feature extraction from large CAD databases using genetic algorithm. Computer-Aided Des. 37(5), 545–558.

[69] Sunil, V.B., Pande, S.S., 2008. Automatic recognition of features from free form surface CAD models. Comput.-Aided Des. 40(4), 502–517.

[70] Banaei, Maryam; Ahmadi, Ali; Yazdanfar, Abbas. Application of AI methods in the clustering of architecture interior forms[J]. FRONTIERS OF ARCHITECTURAL RESEARCH, 2017(6): 360-373

[71] Han Subin. Research on evaluation method of similarity of product morphology[D]. Lanzhou University of Technology, 2018

[72] Ryan Mitchell. Web scraping with Python: collecting more data from the

- modern web[M]. New York, O'Reilly Media, 2018
- [73] Anjuke. Residential property information in Hangzhou.[OB/EL][2019.09.26]. <https://hangzhou.anjuke.com/>
- [74] Chiu T, Fang D, Chen J, Wang Y, Jeris C. A robust and scalable clustering algorithm for mixed type attributes in large database environment[J]. Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, 2001: 263-268
- [75] M Bogdan, J.K. Ghosh, R.W. Doerge. Modifying the Schwarz Bayesian information criterion to locate multiple interacting quantitative trait loci[J]. Genetics Society of America. 2004, 167(2): 989-999
- [76] Sridhar Seshagiri, Hassan K. Khalil. Output feedback control of nonlinear systems using RBF neural networks[J]. IEEE Transactions on Neural Networks, 2000, 11(1): 69-79
- [77] Francis D.K. Ching. Architecture: Form, Space, & Order[M]. New York: Wiley, 2014
- [78] Bao F., Schwarz M., Wonda P. Procedural facade variations from a single layout[J]. ACM Transactions on Graphics, 2013, 32(1): 1-13
- [79] Fan L., Musialski P., Liu L, et al. Structure completion for facade layouts[J]. ACM Transactions on Graphics, 2014, 33(6): 1-11
- [80] Yu Kailong, Cai Yuanzheng, Han Xiaodong et al. Recognition algorithms of architectural style image classification of the university[J]. Journal of Fujian Computer, 2019(5): 8-11
- [81] XU Z, TAO D, ZHANG Y, et al. Architectural Style Classification Using Multinomial Latent Logistic Regression[C]//Computer Vision–ECCV 2014. Berlin and Heidelberg: Springer, 2014:
- [82] Doersch C., Singh S., GUPTA A., et al. What makes Paris look like Paris?[J]. ACM Transactions on Graphics, 2012, 31(4)

(Corresponding author: Xia Bing, E-mail: 0016217@zju.edu.cn)

Author Contributions:

Xia Bing: Conceived research ideas, designed research protocol; drafted manuscript.

Shi Hui, Chen Sichong, Chen Jiamei: Conducted experiments; collected, cleaned, and analyzed data.

Li Xin: Revised final manuscript version.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.