Digital Intelligence PRESS

International Scientific Technical and Economic Research
https://doi.org/10.67541/istaer2626

Research Article **Volume 4, Issue 3 (July 2026)** *Open Access*

Research on Material Translation Matching Model for Regional Environmental Design Based on NLP and Cultural Gene Mapping

Jing Zhang *

College of Arts, Shandong Jianzhu University, Jinan, Shandong, China

***Corresponding author:** Jing Zhang. Email: outlook_8A9201AAC2554F63@outlook.com

Abstract

The selection of materials for regional environment design has long been trapped in the cross-modal gap between cultural semantics and physical properties, and existing methods find it difficult to transform the tacit knowledge in the construction tradition into a computable matching basis. This paper proposes a material translation matching model based on natural language processing and cultural gene mapping. Through deep semantic extraction of multi-source heterogeneous corpora, a heterogeneous information network with "geography, history and technology" as hyperedges is constructed to realize the joint embedding representation of material nodes and cultural symbol nodes. Aiming at the fuzziness and dynamics of design intention, a multi-dimensional constraint tensor modeling method based on fuzzy membership function is designed, and Pareto front ranking and context-aware relaxation coefficient are introduced to realize online adaptive adjustment of elastic matching window. At the matching computing level, a dynamic graph attention aggregation network is proposed to capture semantic associations in heterogeneous structures through node-level, relationship -level and path level attention mechanism, and integrate cross-modal contrast learning to align text constraint space and graph embedding space. Experiments are carried out on the standardized test set covering three regions, namely, water towns in the south of the Yangtze River, red brick in the south of Fujian, and forest pan in the west of Sichuan. The matching accuracy @1 of the complete model reaches 66.8%, which is 22.8 percentage points higher than the optimal baseline model. The lightweight variant achieves a single-inference latency of 42.8 ms at the Jetson Xavier NX edge with an accuracy loss of only 6.0 percentage points. In the blind evaluation, 78.5% of the recommended results were judged as highly consistent by senior architects, and the relative deviation of physical and chemical indicators was within the engineering allowable ranges. Ablation experiments and generalization tests verify the effectiveness of each module and the ability of cross-regional migration of the model.

Keywords

Cultural gene mapping; Heterogeneous information network; Dynamic graph attention aggregation network; Cross-modal contrast learning; Regional environment design material matching

Received: 30 Jun 2026; Revised: 18 Jul 2026; Accepted: 30 Jul 2026; Published: 11 Aug 2026

1. INTRODUCTION

Material selection in regional environment design has always been regarded as the core link between cultural connotation and engineering entity [1],[2]. The white walls and grey tiles of Huizhou dwellings not only relate to the physical properties of lime and blue brick, but also bear the symbolic system of introversion and elegance in Huizhou merchant culture. The carmine brick of red-brick houses

in southern Fujian is not only the function choice of weather resistance, but also reflects the worship of red in Minyue culture and the collective psychology of seeking good fortune and avoiding evil. However, it is this deep entanglement of "cultural semantics-physical performance" that constitutes the most stubborn cross-modal gap in the selection of design materials at present. On the one hand, the designer's creative intention is expressed in natural language, full of "thick", "warm", "simple", "local" and other vague descriptions with strong cultural direction; On the other hand, the data in the material library take the rigid engineering parameters such as density, thermal conductivity and compressive strength as the core fields. The lack of a structured mapping channel between the two makes designers highly dependent on personal experience and fragmented field research in the actual process of material selection, which is time-consuming and difficult to systematically inherit the wisdom of regional construction [3],[4],[5]. In the face of this dilemma, the existing matching methods are expose structural failures: the rule-based system can only deal with the scene where the explicit matching logic is defined in a priori, and it is powerless to deal with the metaphors, metonymy and conventions that exist in cultural semantics. Traditional information retrieval methods regard material description and material attribute as independent keyword space, which loses the vital correlation structure in cultural context [6],[7],[8]. However, the pure visual classification method simplifies material recognition to image classification, completely ignoring the deep cultural genes carried by materials. The common fatal flaw of these approaches is that they treat "culture" as an optional label field for materials rather than as a fundamental dimension inherent in the structure of material meaning.

The current research situation at home and abroad further reveals the specific performance and unfinished business of the above-mentioned dilemma in academic exploration. In the direction of the combination of natural language processing and architectural text analysis, some researchers have applied pre-trained language models to the digital compilation of historical architectural documents, realizing the automatic extraction of craftsman terms and the structural annotation of French text. The domain-specific name-entity recognition method based on RoBERTa, BERT and other models fine-tuning has achieved considerable accuracy in recognizing building-component names, material names and process verbs. However, these studies generally stop at the level of "extraction," that is, extracting material entities and their basic attributes from texts, and do not further explore how these semantic units are related to higher-order knowledge structures such as cultural geography, historical inheritance and artistic schools. The extracted entities are still isolated entries, lacking a methodological framework for weaving them into a computable cultural semantic network [9],[10]. In the field of cultural map construction, the digital protection of cultural heritage has accumulated a large number of knowledge map results, covering ancient architectural styles, traditional craft genealogy, intangible cultural heritage inheritance and other topics. These maps continue to expand in data scale and relationship richness, but their construction goals are mostly knowledge storage and visual display, while the material attribute database is maintained independently of the map, lacking a unified data model and cross-domain alignment mechanism between the two [11],[12]. The node of "cultural symbol" in the map and the item of "building material" in the material database belong to two parallel universes, which cannot query and reason with each other, forming an artificial knowledge separation. In terms of cross-modal matching models, in recent years, the text and text retrieval model based on CLIP architecture has shown amazing ability in multi-modal alignment, and there have been many attempts to introduce it into the retrieval task between architectural images and text descriptions. Graph neural network proves the importance of structural information for inference ability in knowledge graph completion and recommendation system [13],[14],[15],[16]. However, when these technologies are transferred to the scene of regional design material matching, there is a triple lack of adaptation: first, symbols, taboos and feng-shui concepts in regional culture are not explicit knowledge, and it is difficult to directly

extract them from text or image as supervised learning annotations; Secondly, the ambiguity of design intention makes the matching problem not suitable to be modeled as a traditional classification or ranking task, so it is necessary to introduce elastic constraints and dynamic preference mechanism. Third, there is a sharp contradiction between the huge computing cost of deep learning models and the requirements of edge deployment in design practice, and there is no lightweight system scheme for this scenario. In summary, NLP, cultural mapping and cross-modal matching have made progress respectively, but the intersection between the three, how to make machines understand what regional materials "mean" rather than merely "what they are" is still a research gap to be explored.

Based on the above analysis, this paper summarizes three interrelated key scientific problems. The first problem concerns the computational representation of regional cultural tacit knowledge: cultural genes attached to materials, such as symbolic meaning, feng shui taboo and historical memory, are intrinsically embodied and contextualized knowledge, which is difficult to be fully captured by the triplet form in traditional knowledge engineering. The first theoretical obstacle to be overcome in this study is how to design a coding framework so that abstract judgments such as "black bricks metaphorize the reserved character of literati" can be quantified as discrete and computable structured representations. The second question focuses on the elastic matching mechanism between design intent and material parameters. The "thickness" in the design input is not a definite density value but a fuzzy interval with subjective judgment and context dependence, and the corresponding matching output should not be an isolated material item label, but a candidate set with confidence ranking. How to establish a computable and adjustable elastic mapping relationship between fuzzy semantic input and rigid engineering parameters is the core problem of matching model design. The third problem arises from the realistic constraints of engineering deployment. Cultural depth means increasing the size of the graph and the complexity of attention computing, and design practice scenarios often require second-level responses on tablets or embedded devices [17],[18]. How to compress the model to a scale that can be deployed at the edge through architecture innovation while maintaining the depth of cultural semantic analysis constitutes the axis of tension between method design and engineering optimization. The three problems are progressive: coding problem is the basis of representation, matching problem is the core of decision-making, and lightweight problem is the guarantee of landing.

Centering on the above three scientific problems, this paper proposes three core innovative contributions to form a one-to-one corresponding solution strategy. Aiming at the coding dilemma of cultural tacit knowledge, Innovation 1 proposes a joint construction method of multi-dimensional material semantic coding and cultural gene heterogeneous map. In this method, the cultural attributes of materials are no longer regarded as isolated labels, but three semantic triads of attributes, functions and cultural meanings are jointly extracted from ancient books of regional construction, field research notes and modern design instructions, and a heterogeneous information network is constructed by taking "geographical coordinates - historical period - artistic schools" as the hyperedge type. The material nodes, cultural symbol nodes and regional nodes are associated in a unified graph structure. The introduction of spatio-temporal decay function further enables the map to have the ability of dynamic weight adjustment, and the cultural influence of different regions and periods can be quantitatively distinguished. The atlas not only carries the physical parameter vector of the material, but also embeds the location of the material in the cultural semantic space, providing a structured knowledge base for subsequent matching. Aiming at the matching gap between fuzzy intention and rigid parameters, Innovation 2 proposes a cross-modal matching framework of dynamic graph attention aggregation network and contrast learning. DGAT breaks through the assumption of homogeneity of traditional graph attention network through the three-level hierarchical attention computing mechanism of node

level, relationship level and path level, and enables differentiated modeling of information propagation of different node types and relationship types in heterogeneous information networks. At the same time, the cross-modal contrast learning takes InfoNCE variant as the optimization target and combines with the difficult negative sample mining strategy based on similar graph structure but heterogeneous semantics to pull the text embedding of design intention and the structural semantic embedding of graph nodes into the same hidden space, realizing the end-to-end mapping from fuzzy semantics to accurate graph location. Aiming at the real-time constraints of engineering, lightweight graph distillation and adaptive reasoning architecture are introduced in innovation 3. In this architecture, redundant nodes and virtual edges of the graph are pruned by mutual information minimization criterion, which greatly reduces the computing scale. The dynamic gating mechanism adaptively closes the low-contribution attention head according to the complexity of the input context to avoid invalid computation. Quantitative awareness training combined with INT8 reasoning and LRU clocking mechanism of high-frequency matching subgraph enables the model to maintain acceptable response speed on edge devices. The three innovations together constitute a complete technology chain from knowledge coding to matching reasoning and then to engineering deployment.

The organizational structure of the paper follows the progressive logic of the above technical chain. The second section solves the basic problem of "how to digitalize cultural knowledge", from multi-source corpus cleaning, material ontology semantic extraction, heterogeneous information network graph structure generation to multi-modal embedding fusion and dimensionality reduction mapping, completely presents the construction process of cultural gene map, and provides data basis for the whole model. Section 3 turns to the design input side to solve the problem of "how to formalize design requirements". Through intention hierarchy analysis and fuzzy quantization, Pareto priority ranking of multi-dimensional constraint tensors, and adaptive adjustment of constraint weight of design context awareness, natural language design description is transformed into matching conditions with elastic boundaries. Section 4 is the core algorithm layer of the whole paper. Under the framework of comparative learning alignment, the semantic alignment of text constraint space and graph embedding space is realized. The material location is completed through DGAT hierarchical attention aggregation and Seq2Subgraph end-to-end decoder, and the lightweight architecture is embedded to ensure reasoning efficiency. Section 5 comprehensively tests the effectiveness and practicality of the model through systematic experiments on standardized test sets, from algorithm comparison, ablation analysis, generalization robustness test, engineering deployment verification to failure case attribution. A closed research link of "basis construction-conditional input-core computation-performance verification" is formed between each section, and the output of each step serves as the input to the next to ensure the consistency and logic of the methodology level.

2. DISCRETE CODING OF MATERIAL SEMANTICS AND CONSTRUCTION OF REGIONAL CULTURAL GENE MAPS

This section focuses on transforming the unstructured regional environment design text into a structured knowledge base that can be directly used by the matching model. The core tasks cover four links: cleaning and domain adaptation of multi-source corpus, deep semantic extraction of material ontology, graph structure generation of cultural gene heterogeneous information network, and multi-modal embedding fusion and dimension reduction mapping of map nodes. A complete pipeline from original text to high-dimensional semantic space and then to compact embedding vector is formed between the four links, providing input representation with both cultural depth and computational

efficiency for subsequent matching algorithms.

Ancient books on regional construction, fieldwork notes and modern design instructions constitute the three core corpus sources of this study, and there are significant differences in language style and knowledge density among the three. The text of ancient books is mainly in classical Chinese, the sentence pattern is concise and there are a lot of common false characters and different characters; The notes of the field survey have obvious colloquial characteristics, including dialect words and incomplete sentence patterns; Modern design manuals tend to be technical, containing precise physical parameters and process standards. In view of this heterogeneous feature, this paper designs a set of unified cleaning assembly line: Firstly, a rule-based regular expression is used to filter non-text elements (such as annotations, page numbers, and layout information), and then an adaptive word segmentator is used to process the unknown words in ancient books, and the high-frequency meaningless function words such as "Zhihuzheye" and functional words in the design context such as "according" and "according" are eliminated by constructing a domain stop word list. In this paper, an automatic term discovery strategy based on mutual information and left-right entropy is proposed for adaptive stop-word list and domain dictionary expansion for building materials and technology terms. The internal solidification degree of candidate words in the corpus is measured by the inter-point mutual information. When the mutual information value exceeds the preset threshold, the word string is judged to have internal stability. At the same time, left-right entropy is used to measure the freedom of candidate word boundaries, and left entropy and right entropy are used to describe the diversity of adjacent characters on the left and right sides of candidate words respectively. When both mutual information and left-right entropy meet the threshold conditions, the word string is included in the domain dictionary for subsequent word segmentation and named entity recognition. Suppose that candidate words are composed of character sequences, and the calculation formula of inter-point mutual information is as follows:

$$PMI(w) = \log_2 \frac{P(c_1 c_2 \dots c_n)}{\prod_{i=1}^{n} P(c_i)} \tag{1}$$

Where w represents the characters $c_1, c_2, \dots, c_n$, $P(c_1 c_2 \dots c_n)$ is the co-occurrence probability of the word string in the whole corpus, and $P(c_i)$ is the marginal probability of the independent occurrence of a single character c_i in the corpus. The calculation formula of left-right entropy is as follows:

$$H_L(w) = -\sum_{a \in A} P(a \mid w) log_2 P(a \mid w) \tag{2}$$

$$H_R(w) = -\sum_{b \in B} P(wb \mid w) log_2 P(wb \mid w) \tag{3}$$

Where A is the set of all adjacent characters appearing on the left of candidate word w, B is the set of all adjacent characters appearing on the right of candidate word w, $P(a \mid w)$ represents the conditional probability of character a appearing to the left of candidate word w given w, and $P(wb \mid w)$ represents the conditional probability of character b appearing to the right of candidate word w given w. The initial version of the domain dictionary is based on the Dictionary of Traditional Chinese Architectural Terms and the Dictionary of Chinese Arts and Crafts, and continues to expand through the above automatic discovery mechanism, finally covering more than 10,000 terms in four categories: material, technology, structure and decoration.

On the basis of corpus cleaning and dictionary expansion, the deep semantic extraction of design

material ontology focuses on identifying regional special material entities from text and extracting their structural representation. In this paper, RoBERTa-wwm-ext pre-training model is used as the underlying encoder, which is trained based on the full-word mask strategy and has stronger semantic modeling ability than standard BERT in processing Chinese text [19],[20],[21]. For the task of regional special material identification, this paper connects bidirectional long and short-term memory network and conditional random field at the top layer of RoBERTa-wwm-ext to form a sequence labeling framework. The labeling system adopts BIOES five-category labeling, covering material name, process name, color description, pattern name and other entity types. In the process of fine-tuning, the annotated ten-million-word regional design corpus is used for domain adaptive training, the bottom parameters of the pre-training model are frozen, and only the parameters of the last three Transformer layers and BiLSTM-CRF head are updated, so as to avoid catastrophic forgetting under the condition of small samples. After fine-tuning, the F1 value of the model for regional materials such as "brics", "beech" and "terrazzo" reaches more than 83 percent, which is about 12 percentage points higher than that of the general named entity recognition model without fine-tuning. Based on the results of entity recognition, dependency parsing is used to extract the semantic relations between material entities. In this paper, we use the dependency-parsing module of the HIT Language Technology platform to analyze the subject-predicate, verb-object, determine-middle and media-object dependencies among words in a sentence. Taking "Huizhou Residence uses white wall and Daiwa to create simple and elegant artistic conception" as an example, the core predicate "adopt" is identified by relying on syntactic analysis, its subject is "Huizhou residence" and its object is "white wall and Daiwa". Meanwhile, "white wall and Daiwa" modifies "simple and elegant" through determining-Chinese relationship, and "simple and elegant" modifies "artistic conception" through determining-Chinese relationship. According to this, the triplet (Huizhou residence, use, white wall Daitile) and (white wall Daitile, create artistic conception, simple and elegant) can be automatically extracted. Through the combination of entities and relationship types on the path of dependency, this paper realizes the automatic extraction of three types of triples of attribute, function and cultural implication, among which the attribute triad describes the physical characteristics of materials, the functional triad describes the use of materials in construction, and the cultural implication triad reveals the symbolic connotation of materials. The above three types of triples constitute the basic fact units of the subsequent heterogeneous information network.

The structural modeling of graph is based on the theoretical framework of heterogeneous information network. Different from homogeneous graph, heterogeneous information network supports the coexistence of multiple node types and multiple edge types, which can more accurately describe the multiple attributes and complex associations of cultural genes. This paper defines three types of core nodes: material nodes cover natural and artificial materials used in regional architecture; cultural symbol nodes cover visual cultural elements such as color, pattern and form; and regional nodes cover geographical coordinates and administrative regions. A hyperedge is defined as an edge structure connecting three or more nodes. Specifically, each hyperedge connects a region node, an era node and a craft school node, and all the materials and cultural symbols nodes that belong to the region and the period and are processed by the craft of the school are grouped into the same hyperedge. Let the set of nodes be V and the set of relation types be $\mathcal{R}$. The graph structure of heterogeneous information network is formally expressed as $\mathcal{G} = (V, E, \phi, \psi)$, where E is the set of hyperedges, the node type mapping function $\phi: V \to \mathcal{T}_V$ maps each node to its type, and $\mathcal{T}_V$ is the set of node types. The edge type mapping function $\psi: E \to \mathcal{T}_E$ maps each hyperedge to its type, and $\mathcal{T}_E$ is the set of edge types. For any hyperedge $e = (v_1, v_2, \ldots, v_k) \in E$, its type is uniquely determined by the triplet of region, period and genre associated with the hyperedge, and its weight is calculated by the space-time decay function.

The spatio-temporal decay function is used to quantify the attenuation effect of cultural influence with geographical distance and intergenerational inheritance, which is the core mechanism of map weight allocation [22],[23]. Suppose that the geographical location of the current design project is p_0, the spatial distance of a certain cultural source point p_i is d_i, and the time span is Δt_i. The cultural influence intensity of regional node v_i on the current project is determined by the product of spatial decay term and time decay term:

$$\alpha_i = e^{-\lambda_d d_i} \cdot e^{-\lambda_t \Delta t_i} \tag{4}$$

Where λ_d and λ_t are the spatial and temporal attenuation coefficients respectively, which control the weakening rate of the cultural influence of geographical distance and intergenerational interval respectively. The value of the spatial attenuation coefficient λ_d refers to the law of distance attenuation in cultural geography. For regions within the same cultural circle (such as within Jiangnan area), λ_d is taken as a small value to maintain the cultural correlation at a long distance. For the long-distance regions across cultural circles, λ_d is appropriately increased to reflect the cultural isolation effect. The time decay coefficient λ_t is set according to the intergenerational continuity of cultural inheritance, and the skill inheritance usually lasts for three to five generations, beyond which the influence decays exponentially. In the calculation of the weight of the hyperedge, the spatiotemporal attenuation coefficient of the regional node, historical period node and technical school node covered by the hyperedge is taken as the geometric average, which is used as the initial weight of the hyperedge. In addition, in order to reflect the inheritance activity of the material itself, the degree-centrality correction term for the material node is introduced, which is calculated as the logarithm of the degree plus one. Suppose that the connection degree of material node v_m in the map is $deg(v_m)$, then the correction term is $\beta_m = \log\big(1 + deg(v_m)\big)$. The comprehensive weight of the final hyperedge e is obtained by multiplying the influence strength and the correction term:

$$W_e = \alpha_i \cdot \beta_m \tag{5}$$

The comprehensive weight is used as the initial value of the adjacency matrix in the subsequent graph attention aggregation to control the propagation intensity of information in the heterogeneous information network.

The high-dimensional embedding of graph nodes aims to encode the topological structure and semantic information of heterogeneous information networks into dense vectors that can be directly used by downstream matching models [24],[25]. Considering that the graph contains three types of nodes and a variety of hyperedge types, it is difficult for a single embedding method to take into account the differences between relational semantics and local topology, this paper adopts the parallel processing strategy of TransR and GraphSAGE. TransR maps each entity to a specific relationship space, designs an exclusive projection matrix for each relationship type, projects the head entity and tail entity from the original space to the corresponding subspace of the relationship, and requires the translation approximation relationship to be satisfied in the subspace. For any triplet (h, r, t), the scoring function of TransR is defined as the square of the distance between the head entity and the tail entity after projection:

$$f_r(h,t) = \| M_r h + r - M_r t \|_2^2 \tag{6}$$

Where $h \in \mathbb{R}^d$ and $t \in \mathbb{R}^d$ are the original embedding vectors of head entity and tail entity respectively, $r \in \mathbb{R}^k$ is the embedding vector of relation r, $M_r \in \mathbb{R}^{k \times d}$ is the exclusive projection matrix of relation r, which maps the d-dimensional original space to the k-dimensional relation space,

$\|\cdot\|_2^2$ represents the square of the two norms of the vector. By minimizing the scoring function of positive example triples while maximizing the scoring function of negative example triples, TransR learns a relational sensitive node embedding. At the same time, GraphSAGE captures local graph topology features through neighbor aggregation [26]. Let the embedding of node v at layer $l+1$ be jointly determined by the aggregation of the layer l embedding of the node and the layer l embedding of all its neighbors:

$$h_v^{(l+1)} = \sigma\left(W^{(l)} \cdot AGGREGATE^{(l)}\left(\left\{h_u^{(l)} : u \in \mathcal{N}(v)\right\}\right)\right) \tag{7}$$

Where $\mathcal{N}(v)$ is the first-order neighbor node set of node v, AGGREGATE$^{(l)}$ is the pooling aggregation function of the l-th layer, and the aggregation vector of fixed dimension is output after symmetric transformation of the neighbor embedding set, $W^{(l)}$ is the learnable linear transformation weight matrix of the l-th layer, and $\sigma(\cdot)$ is the nonlinear activation function. After two layers of GraphSAGE convolution, each node obtains a topological embedding vector containing second-order neighborhood information. In this paper, the relational embedding output by TransR and the topological embedding output by GraphSAGE are spliced along the feature dimension to form a preliminary high-dimensional fusion embedding.

High-dimensional fusion embedding has obvious bottlenecks in storage and computing efficiency, and there is semantic redundancy between different dimensions, so dimension reduction mapping is needed. In this paper, uniform manifold approximation and projection (UMAP) is used to compress the fusion embedding into 256-dimensional compact space [27],[28]. The core assumption of UMAP is that the data is uniformly distributed on the manifold. The algorithm achieves embedding by constructing the fuzzy simplex set of high dimensional space and minimizing the cross entropy between it and the fuzzy simplex set of low dimensional space. The similarity between nodes in high-dimensional space is defined by the fuzzy membership function, which depends on the metric distance, the nearest neighboring distance and the local normalization factor. The similarity between nodes in low-dimensional space is defined by the student t distribution kernel function with parameters. UMAP minimizes the cross-entropy loss between high-dimensional and low-dimensional fuzzy simplex sets through stochastic gradient descent, and its specific form is as follows:

$$\mathcal{L}_{UMAP} = \sum_{i \neq j}\left[\mu_{ij}\, log\frac{\mu_{ij}}{\nu_{ij}} + (1-\mu_{ij})\, log\frac{1-\mu_{ij}}{1-\nu_{ij}}\right] \tag{8}$$

Where μ_{ij} is the fuzzy membership degree of node i and node j in the high-dimensional space, ν_{ij} is the fuzzy membership degree of two nodes in the low-dimensional space, and the sum traverses all node pairs. The calculation of the high-dimensional membership degree μ_{ij} depends on the metric distance $d(x_i, x_j)$ between nodes, the nearest neighbor distance ρ_i and the normalization factor σ_i. The specific form is as follows:

$$\mu_{i|j} = exp\left(-\frac{d(x_i, x_j) - \rho_i}{\sigma_i}\right) \tag{9}$$

The low-dimensional membership ν_{ij} is calculated using the student t distribution kernel function:

$$\nu_{ij} = \left(1 + a \parallel y_i - y_j \parallel^{2b}\right)^{-1} \tag{10}$$

Where $y_i, y_j \in \mathbb{R}^{256}$ are the coordinates of node i and node j in the low-dimensional target space, and the parameters a and b are the positive numbers set by the user to control the rate of membership decay with distance in the low-dimensional space. The first term of the cross-entropy loss function urges the nodes that are adjacent in the high-dimensional space to remain adjacent in the low-dimensional space, and the second term urges the nodes that are far away in the high-dimensional space to remain far away in the low-dimensional space, thus preserving the cross-modal semantic manifold structure in the global scope. After UMAP mapping, each graph node obtains 256-dimensional compact embedding vector, which not only retains the relational semantics encoded by TransR, but also retains the topology captured by GraphSAGE, while greatly reducing the computational storage overhead. As the input basis of the subsequent matching algorithm, the embedding vector set is stored in the vector database, which supports fast retrieval based on cosine similarity and deep matching based on graph attention aggregation.

3. DESIGN INTENTION ANALYSIS AND MATCHING SPACE MODELING OF DYNAMIC CONSTRAINT AWARENESS

This section inherits the graph embedding vector basis output in section 2, and then solves the two core problems of "how the demand is understood by the machine" and "how the understood demand matches with materials in the constrained space" from the design input side. If section 2 completes the structural coding of "what" on the material side of the design, this section completes the formal modeling of "what" on the demand side of the design. The bridge between the two lies in that the fuzzy natural language description in the design intention needs to be parsed into quantifiable constraints, and these constraints must form a tensor structure comparable to the embedding of graph nodes in the multidimensional space, and finally provide accurate query conditions and elastic search boundaries for the matching algorithm in section 4. The whole section is expanded according to the progressive logic of "intention analysis, constraint tensor quantization, weight adaptation".

Hierarchical intent parsing of design input text is the first step in constraint modeling. Its goal is to extract four layers of semantic information from the designer's natural language description: Material type indicates the type of material that designers tend to use, physical attributes involve quantifiable engineering parameters such as density, strength and thermal conductivity, visual rhetoric describes perceptual requirements such as color, texture and gloss, and cultural symbols capture regional symbols, feng shui taboo or historical meanings carried by materials. In this paper, BERT-BiLSTM-CRF sequence annotation framework is used to identify the above four-level information in fine granularity. BERT, as the underlying encoder, transforms input sentences into context-aware character level embeddings, BiLSTM layer captures the long-range dependence of sequences in bidirectional transmission, and CRF layer ensures the global optimality of label sequences through transition matrix to avoid illegal label combinations. The four-level labeling system adopts a flat design, that is, each entity type is labeled independently without nesting, which enables the model to simultaneously identify information units at different levels in a single forward propagation. The annotation of the training data follows strict consistency criteria: for the phrase "dark gray granite", "granite" is labeled as the material type, and "dark gray" is labeled as the color dimension in the visual rhetoric; For the statement "low thermal conductivity", "thermal conductivity" is labeled as the physical attribute type, and "low" is labeled as the attribute value. After fine-tuning on 50,000 manually annotated design description corpus, the macroscopic F1 value of the model in four-level entity recognition reaches 87.6 percent.

After the completion of entity recognition at the intention level, the most challenging problem in design input is how to transform the fuzzy descriptions with strong subjectivity and cultural dependence, such as "thick", "gentle" and "elegant", into numerical expressions that can be quantitatively processed by computers. In this paper, the fuzzy membership function is used as this transformation channel. Suppose a fuzzy descriptor s corresponds to physical parameters $p \in \{\text{density}, \text{heat conductivity}, \text{surface roughness}, \text{light reflectance}, \text{porosity}\}$, and its fuzzy membership function $\mu_{s,p}(x)$ is defined in the value interval of physical parameters, and the output value is between zero and one, which describes the degree to which the physical value x conforms to the descriptor s. For example, the membership function for the word "thick" in the density dimension is constructed as an S-shaped growth curve:

$$\mu_{Thick,\ density}(x) = \frac{1}{1 + e^{-k(x - x_0)}} \tag{11}$$

Where x is the actual density value of the material, x_0 is the center point of semantic transition (taking the median density of common building stone about 2700 kg per cubic meter), and k is the steepness coefficient controlling the transition rate of membership from zero to one. When the density value is much larger than x_0, the membership degree approaches to unity, indicating that the material fully conforms to the description of "thick and heavy". When it is much smaller than x_0, the membership tends to zero, indicating complete nonconformity. For descriptors such as "Wen Yun", which involve the comprehensive perception of touch and vision, the mapping is not a single physical parameter but a combination of parameters. This paper adopts the product form of multi-dimensional fuzzy membership degrees:

$$\begin{aligned}&\mu_{Warm\ and\ moist}\left(x_{Thermal\ conductivity}, x_{Surface\ roughness}, x_{Reflectance\ of\ light}\right)\\ &= \mu_{Warm,\ thermal\ conductivity}\left(x_{Thermal\ conductivity}\right) \cdot \mu_{Softness,\ surface\ roughness}\left(x_{Surface\ roughness}\right)\\ &\qquad \cdot \mu_{Warm\ and\ moist,\ Reflectance\ of\ light}\left(x_{Reflectance\ of\ light}\right)\end{aligned} \tag{12}$$

Among them, the membership function of thermal conductivity is inverted S-shaped (the lower the value is, the more "warm"), the membership function of surface roughness is bell shaped (moderate roughness gets the highest membership), and the membership function of light reflectance is S-shaped (soft reflected light corresponds to moderately low reflectivity). The parameters of all fuzzy membership functions were obtained by nonlinear least squares fitting to the scoring data of design domain experts to ensure that the function curves were consistent with the statistical distribution of human subjective judgments. Through the above fuzzy quantization, any design input text can be transformed into several groups of physical parameter intervals with membership weights, and these intervals constitute the basic elements of the subsequent constraint tensor. The construction of multi-dimensional constraint tensors integrates the above fuzzy quantization results into a unified three-axis space representation. The three axes correspond to the physical performance axis, visual perception axis and cultural symbol axis respectively, which constitute the three orthogonal dimensions of material matching constraints. The physical property axis includes six indexes such as density, compressive strength, flexural strength, thermal conductivity, water absorption and fire resistance grade, each of which is a continuous numerical interval, and the lower and upper limits are determined by the truncation of fuzzy membership threshold. The visual perception axis includes four indexes: color coordinates, surface gloss, texture period and transparency, among which color coordinates are represented by L^*, a^* and b^* components of CIELAB uniform color space. Cultural symbol axis is the most special dimension, which does not have continuous numerical measurement, but is represented by the activation vector of cultural symbol nodes in the graph. Let the set of cultural symbols identified

in a design input be $\mathcal{S} = \{s_1, s_2, \ldots, s_m\}$, where each cultural symbol node s_i corresponds to a 256-dimensional embedding vector e_{s_i} in the graph, then the constraint of the cultural symbol axis is expressed as the weighted sum of these embedding vectors, and the weight is the comprehensive product of the frequency of the symbol in the input text and the fuzzy membership value. So far, the complete constraint corresponding to a single design input can be formally defined as a third-order tensor $T \in \mathbb{R}^{D_p \times D_v \times D_c}$, where D_p, D_v and D_c are the effective dimensions of the physical performance axis, the visual perception axis and the cultural symbol axis respectively. However, the fully expanded third-order tensor is too sparse and dimensional explosion, and the actual calculation decomposes it into the Cartesian product form of three subspaces, each of which maintains its own boundary interval or activation set inside.

The importance of each dimension in the three-axis constraint space varies significantly in different design tasks, and there are often conflicts between constraints. For example, the pursuit of high intensity in physical performance may require the sacrifice of delicacy in visual perception. In this paper, the Pareto frontier method is used to perform dynamic prioritization of constraints and generate a set of non-dominated constraints to avoid the subjective arbitrariness of artificially specifying weights. Let the set of constraints generated by the current design input be $\mathcal{C} = \{c_1, c_2, \ldots, c_n\}$, where each constraint c_i corresponds to an expected value vector $v_i \in \mathbb{R}^3$ in the three-dimensional space (the three components represent the demand intensity of the constraint in the physical, visual and cultural axes respectively), and a tolerance interval $\Delta_i = [\delta_i^-, \delta_i^+]$. For two constraints c_i and c_j, if the demand intensity of c_i is not inferior to that of c_j in all three dimensions and is strictly dominant in at least one dimension, then c_i dominates c_j. The Pareto front consists of all constraints that are not dominated by any other constraint:

$$\mathcal{C}^* = \{ c_i, \in, \mathcal{C} \mid \nexists c_j \in \mathcal{C}j \neq i \; v_j \succeq v_i \; 且 \; v_j \neq v_i \} \tag{13}$$

Where $\succeq$ denotes the term-by-term partial order relation on the components of the vector. The constraints on the frontier are treated as high-priority hard constraints and must be strictly satisfied in the matching process. The dominated constraint is then treated as a soft constraint, which can be moderately relaxed in matching. This Pareto ranking process is dynamically recalitated as the input text changes, and does not depend on any preset fixed weight system, so that the judgment of constraint priority is completely driven by the internal structure of the design intention itself. After Pareto screening, the elastic matching window is defined as the combination of the intersection and union of the tolerance intervals of all constraints in the Pareto front constraint set. This window serves as the boundary condition of search in the matching algorithm in section 4, the matching result must fall into the window, but the window itself has a certain elastic range. The specific elasticity magnitude is controlled by the subsequent relaxation coefficient.

The online adaptive adjustment of context-aware constraint weights further endows the above constraint framework with the flexibility of time dimension. The design process is not a one-time query operation, but a gradual evolution process from concept conception to expanded design and then to deepened design. There are systematic shifts in the focus of designers on materials at different stages: The conceptual stage focuses on cultural symbols and visual imagery, the initial expansion stage focuses on the feasibility of physical properties, and the deepening stage strictly examines the accuracy of all engineering parameters. To capture this preference transfer in the design sequence, this paper introduces a local context encoder, which scans the design history (including the entered natural language description, the confirmed material selection, and the abandoned alternative) in a sliding window, with

the window size set to the last five interaction records. Each record in the window is context-embedded by lightweight Transformer encoder, and then aggregated into a context vector of fixed dimension c_{ctx} by attention pooling mechanism. The attention weight is calculated as follows:

$$\alpha_t = \frac{exp(q^\top W_a h_t)}{\sum_{j=1}^{T} exp(q^\top W_a h_j)} \tag{14}$$

Where h_t is the coding embedding of the t record in the window, W_a is the attention projection matrix, q is the learnable query vector representing the overall orientation of the current design requirements, α_t is the attention weight of the t record, T is the window size, $c_{ctx} = \sum_{t=1}^{T} \alpha_t h_t$ is the context vector obtained after weighted summation. The context vector is concatenated with the encoding of the current input text and fed into a fully connected classifier, which outputs the adjustment of the relaxation coefficient of the three axis constraints.

Explicit information from the design phase is used directly to regulate the constraint relaxation coefficient. Set concept stage, expansion in the early stage, deepening stage corresponds to the basis of relaxation coefficient vector respectively $\rho_{concept} = (0.30,0.35,0.40)$, $\rho_{develop} = (0.15,0.20,0.15)$, $\rho_{detail} = (0.05,0.05,0.05)$, The vector components correspond to the relaxation amplitudes of the physical, visual, and cultural axes in turn. The final effective relaxation coefficient is obtained by adding the stage base value and the context adjustment quantity:

$$\rho_{eff} = \rho_{stage} + \Delta\rho(c_{ctx}) \tag{15}$$

Where $\Delta\rho(\cdot)$ is the mapping function from context vector to relaxation coefficient adjustment, which is realized through a two-layer multi-layer perceptron. The three components of the effective relaxation coefficient ρ_{eff} control the proportion of the outward expansion of the constraint boundary on the physical performance axis, the visual perception axis and the cultural symbol axis respectively.

4. CORE TRANSLATION MATCHING ALGORITHM AND LIGHTWEIGHT ARCHITECTURE BASED ON CONTRAST LEARNING AND GRAPH ATTENTION AGGREGATION

The cross-modal contrast learning alignment framework aims to solve the semantic gap between text constraint space and graph embedding space. Since the design text and graph nodes belong to different modes and lack explicit pairing labeling, this paper adopts the contrast learning strategy to close the distance between the matched "text-graph subgraph" pairs in the hidden space, and push the distance between the unmatched pairs. In the training stage, each training sample consists of a design description text and its corresponding correct material selection result, which is obtained by tracing back the construction drawing and material list of the design landing project. Suppose that a training batch contains N samples, and the text description of each sample is encoded into a constraint embedding vector $t_i \in \mathbb{R}^{256}$ through the constraint parsing process described in section 3. Meanwhile, the embedding of the correct material node corresponding to the sample in the graph is $m_i^+ \in \mathbb{R}^{256}$. The correct material node $m_j^- (j \neq i)$ of other samples in the batch naturally becomes a negative example. In this paper, the cross-modal InfoNCE variant loss function is designed to optimize the alignment of text encoder and map embedding space, and its specific form is as follows:

$$\mathcal{L}_{contrast} = -\frac{1}{N}\sum_{i=1}^{N} log\frac{exp\left(\frac{sim(t_i,\, m_i^+)}{\tau}\right)}{\sum_{j=1}^{N} exp\left(\frac{sim\left(t_i, m_j^-\right)}{\tau}\right)} \tag{16}$$

Where $\mathrm{sim}(t,m) = t^\top m/(\| t \| \| m \|)$ is the cosine similarity function, which measures the cosine value of the Angle between the constraint embedding and the material embedding on the unit hypersphere. τ is the sharpness of the loss distribution, controlled by the temperature hyperparameter, and the smaller the value is, the more sensitive the model is to distinguish positive and negative samples. However, simply taking all the material nodes of other samples in the batch as negative examples has obvious drawbacks: although some negative example materials do not match the current text, they may be highly similar semantically, and such "simple negative examples" cannot provide effective gradient information. To this end, this paper introduces the automatic mining strategy of difficult negative samples, and selects negative samples based on the criterion of similar graph structure but heterogeneous semantics. Specifically, for each positive sample pair (t_i, m_i^+), the k-order neighbor node set $\mathcal{N}_k(m_i^+)$ of m_i^+ is first calculated in the graph. These nodes are topologically-adjacent to the positive sample material, but at the semantic level, the similarity with t_i should be lower than a certain threshold proportion of the similarity of the positive sample pair. The set of difficult negative samples $\mathcal{N}_{\text{hard}}$ is defined as the set of nodes that meet both structural proximity and semantic heterogeneity conditions.

On the basis of cross-modal alignment, the dynamic graph attention aggregation layer undertakes the core computing task of deep fusion of constraint embedding and graph structure. The traditional graph attention network has two fundamental limitations when dealing with heterogeneous information networks. First, the traditional GAT only calculates the attention weight at the node level, ignoring the different contributions of different relationship types to information propagation. Second, traditional GAT assumes that the attention pattern among homogeneous nodes is consistent, which cannot deal with the complex interaction caused by the difference of node types in heterogeneous networks. In this paper, DGAT, the attention aggregation layer of dynamic graph, is designed to progressively calculate the attention weight from the node level, relationship level and path level, and the learnable relationship perception bias term is introduced to break through the homogeneity assumption. Let the central node to be aggregated be v, and the first-order neighbor set be $\mathcal{N}(v)$. Node-level attention calculates the importance of neighbor node u to central node v:

$$\alpha_{vu}^{(node)} = \frac{exp\left(LeakyReLU\left(a_{node}^\top [W_h h_v \oplus W_h\, h_u]\right)\right)}{\sum_{k\in\mathcal{N}(v)} exp\left(LeakyReLU(a_{node}^\top [W_h h_v \oplus W_h\, h_k])\right)} \tag{17}$$

Where h_v and h_u are the embedding vectors of the central node and neighbor node respectively, W_h is the shared linear transformation matrix mapping the input embedding to the attention space, $\oplus$ represents the vector splicing operation, a_{node} is the node-level attention weight vector, and LeakyReLU is the activation function of linear unit with leakage correction. Relationship-level attention calculates the contribution weight of different relationship types to the aggregation result. Suppose that the relationship type between node v and neighbor u is $r(v,u) \in \mathcal{R}$, then the relationship-level attention coefficient is:

$$\alpha_r^{(rel)} = \frac{exp\left(a_{rel}^\top \left[W_r h_v \oplus W_r\, h_u \oplus e_{r(v,u)}\right]\right)}{\sum_{r'\in\mathcal{R}} exp(a_{rel}^\top [W_r h_v \oplus W_r\, h_{u'} \oplus e_{r'}])} \tag{18}$$

Where $e_{r(v,u)}$ is the learnable embedding vector of relation type, W_r is the projection matrix of relation space, and a_{rel} is the relational level attention weight vector. Path-level attention further considers the path pattern formed by neighbor nodes arriving from the central node through different types of relationships. The coefficient calculation method is similar to that of the relationship level, but the aggregation scope is extended to second-order paths. Three levels of attention to the final fusion for unification by means of weighted product aggregate weights $\alpha_{vu} = \alpha_{vu}^{(node)} \cdot \alpha_r^{(rel)} \cdot \alpha_{path}^{(path)}$. On this basis, this paper introduces a learnable relational awareness bias term $b_{r(v,u)}$, which is directly added to the embedding of neighbor nodes to compensate for the distribution shift caused by heterogeneous types, so that the aggregation formula of DGAT becomes:

$$h_v^{(l+1)} = \sigma\left(\sum_{u \in \mathcal{N}(v)} \alpha_{vu} \cdot W_g\left(h_u^{(l)} + b_{r(v,u)}\right)\right) \tag{19}$$

Where W_g is the output projection matrix of the aggregation layer and σ is the activation function. The relation-aware bias term $b_{r(v,u)}$ is learned independently for each relationship type, so that the information from different relationship types is aligned into a unified semantic space before aggregation.

The sequence-to-subgraph end-to-end matching decoder converts the constraint embedding into a specific matching result on the graph. The traditional retrieval method usually adopts the two-stage strategy of "recall first and then sort". In this paper, the Seq2Subgraph end-to-end decoder is proposed, and the matching process is modeled as a sequential decision problem of gradually locating the target node on the graph. The core component of the decoder is a graph node fast location module based on pointer network. This module takes the constraint embedding vector t as the initial state, takes the graph node embedding after DGAT aggregation as the search space, and gradually outputs the most likely matching material node sequence. Let the decoding step size be s, the current hidden state be z_s, and the output probability of pointer network for each candidate node v at the step size s is:

$$P(v \mid z_s) = softmax\left(z_s^{\top} W_p h_v\right) \tag{20}$$

Where W_p is the projection matrix of the pointer network, and h_v is the embedding vector of node v after DGAT aggregation. This output probability distribution indicates the probability of selecting each node as the matching result under the current decoding step, and the model selects the node with the highest probability as the output of this step. In order to use the structural information of the graph for reasoning in the matching process, a gated cyclic unit GRU is connected after the pointer network. The GRU takes the node embedding selected in the current step and the current hidden state as input, and updates the hidden state for the next decision:

$$z_{s+1} = GRU\left(z_s, h_{v_s}\right) \tag{21}$$

Where h_{v_s} is the embedding vector of the selected node with step size s, and z_{s+1} is the updated hidden state. The decoder stops after multi-step output, and the stopping condition is that the termination probability of GRU output exceeds a preset threshold, or the step size reaches the maximum decoding length. Finally, the decoder outputs not only the set of matched material nodes, but also the matching confidence heatmap, the cumulative probability distribution of each candidate node being selected, and the decision chain, which is the sequence trajectory of selecting nodes at each step in the decoding process. Each hop of the decision chain corresponds to an intermediate decision in model reasoning,

providing the original basis for subsequent interpretability analysis.

Model lightweight and reasoning acceleration architecture ensure that the above algorithm can meet real-time requirements in actual engineering scenarios. Regional design practice often needs to complete material matching on edge devices, and cloud reasoning has network delays and data security concerns, so the model must maintain matching accuracy while greatly compressing the amount of computation. This paper carries out lightweight design from three dimensions: graph distillation, dynamic gating, quantization and caching. The graph distillation strategy compresses the graph based on the mutual information minimization criterion. Its core idea is to calculate the mutual information between each node and the matching task label, retain the nodes with high mutual information and eliminate the redundant nodes with lower mutual information than the threshold. At the same time, weak connections with weights lower than a given threshold in the hyperedge structure are deleted. After distillation, the size of the map is reduced to 40 to 60 percent of the original, greatly reducing the space complexity of attention calculation. To solve the redundancy problem of multi-head attention, the dynamic gating mechanism introduces a learnable gating vector $g \in [0,1]^H$ into each DGAT layer, where H is the total number of attention heads, and each component controls the activation state of the corresponding attention head:

$$h_v^{(l+1)} = \sum_{h=1}^{H} g_h \cdot h_{v,h}^{(l+1)} \tag{22}$$

Where $g_h = \sigma\left(w_g^{\top} c_{ctx} + b_g\right)$ is dynamically calculated by the current design context vector c_{ctx}, and w_g and b_g are learnable parameters. When the context tends to be simple matching, most attention heads are turned off, and the amount of computation decreases significantly. When the context is complex, more attention heads remain active to ensure matching accuracy. In terms of quantization, this paper adopts the quantization awareness training strategy to simulate the numerical accuracy loss of INT8 quantization during the training process, so that the model parameters can be directly stored and calculated in INT8 format during reasoning. Let the floating-point weight be w, and the quantization integer value be $w_q = \text{clamp}(\left\lfloor \frac{w}{\Delta} \right\rceil, -, 128127)$, where Δ is the quantization step, $\lfloor \cdot \rceil$ is the rounded integer function, and the "clamp" function truncates the value to the range that INT8 can represent. In addition, in view of the typical design scenarios that occur frequently in regional design practice (such as "Hui-style residential exterior wall materials" and "Southern Fujian red brick ground"), this paper establishes LRU caching mechanism for high-frequency matching subgraphs: whenever the result of a matching query is adopted, the corresponding constraint embedding and matching result subgraph of the query will be stored in the cache. When cache space is limited, the least recently used cache entry is eliminated. When the cache hits, the prestored result is directly returned, and the inference delay is reduced to the millisecond level.

The interpretable backtracking module completes the matching model by enabling the designer to understand "why the model recommends these materials" rather than being presented with a black box output. In this paper, Grad-CAM method is used to reverse the decision key map node and semantic features. For the completed matching results, Grad-CAM calculates the gradient of the embedding of the matching score on each node in the last layer of DGAT, takes the global average pooling value of the gradient as the importance weight of each node, generates a node-level thermal map and superposes it on the original map, highlighting the node region that contributes the most to the current matching decision:

$$\Gamma_v = \frac{1}{Z} \sum_k \frac{\partial \mathcal{L}_{match}}{\partial h_{v,k}} \tag{23}$$

Where $\mathcal{L}_{\text{match}}$ is the matching loss function, $h_{v,k}$ is the k-th component of the embedding vector of node v, Z is the normalized constant, and Γ_v is the importance score of node v. The nodes whose node importance score is higher than the set threshold and their associated hyperedges are extracted to form the decision critical subgraph. At the same time, combined with the decision chain information of the pointer network, the module will correlate and map the specific semantic entities (such as a fuzzy descriptor or a physical parameter) in the input constraints corresponding to each decision step with the graph nodes in the decision key subgraph, and finally output a visual interpretation report. The report presents the decision path, key nodes and their corresponding cultural or physical attributes in a mixed text and graphic form for designers to refer to when reviewing the matching results. Node highlighting and path annotation in the explanation report transform the matching process from a "black box calculation" to a "traceable chain of reasoning", enhancing the credibility and acceptance of the model in practical design collaboration.

5. EXPERIMENTAL VERIFICATION AND ENGINEERING ADAPTATION EVALUATION

In this section, the effectiveness and engineering practicability of the coding framework, constraint modeling method and matching algorithm proposed in the first four sections are verified by systematic experiments. The experimental design follows the multi-level verification logic of "from internal to external, from standard to extreme, from algorithm to engineering", covering five dimensions successively: dedicated test set and index system construction, core algorithm ablation and baseline comparison, generalization and robustness stress test, engineering deployment adaptation verification, and failure case attribution analysis. All experiments were completed on a server equipped with dual-circuit Intel Xeon Gold 6248 CPU, 256GB memory and NVIDIA A100 40GB GPU. The edge test was performed using NVIDIA Jetson Xavier NX development kit. The software environment is based on PyTorch 1.13 and DGL 1.1 deep learning framework.

The construction of special test set is the basis of experimental verification. In this paper, five representative design projects have been selected from three typical regions, namely, water towns in the south of the Yangtze River, red brick in the south of Fujian, and forest pan in the west of Sichuan. Then, three regional architectural design experts with the title of senior engineer were invited to mark the corresponding real material selection for each design description text as positive samples based on the project material list, and at the same time, mismatched materials were randomly selected from other projects in the same region as negative samples. The category of "fuzzy negative samples", those materials that are close to but not exactly match the positive samples in cultural semantics or physical attributes - is introduced in the labeling process to increase the difficulty of discrimination of the test set. The final standardized test set consists of 3,600 labeled samples, among which the water towns in southern Jiangnan, red brick in southern Fujian and forest pan in western Sichuan account for one third each, and the proportion of positive and negative samples is strictly controlled at five to five. The evaluation index system covers four types of indicators: Top-k recall measures the proportion of correct materials in the Top K results output by the model, and K is taken as one, three and five respectively; Matching accuracy @1 measures the accuracy rate that the top-ranked output is the correct material, which directly reflects the actual recommendation availability of the model; Cosine similarity of

cultural semantic consistency is calculated to calculate the cosine similarity between the recommended material and the real material in the map embedding space, which represents the fidelity of the matching result at the cultural semantic level and is defined as:

$$CSC = \frac{1}{N}\sum_{i=1}^{N}\frac{e_{rec}^{(i)\top} e_{gt}^{(i)}}{\| e_{rec}^{(i)} \| \| e_{gt}^{(i)} \|} \tag{24}$$

Where $e_{rec}^{(i)}$ is the atlas embedding vector of the i-th test sample recommendation material, $e_{gt}^{(i)}$ is the atlas embedding vector of the real material, N is the total number of test samples, the closer "CSC" is to one, the more semantically similar the recommended material is to the real material. End-to-end inference FPS measures the number of samples that can be processed per second in the complete process of the model from input design text to output matching results, which is used to evaluate engineering real-time. The four types of indicators describe the model performance from the three dimensions of accuracy, semantic fidelity and computational efficiency respectively, forming a complementary evaluation system.

The core algorithm comparison and ablation experiments aim to quantify the independent contributions and overall advantages of each component of the proposed model. The baseline model selects graph convolutional network GCN, graph attention network GAT, heterogeneous graph attention network HAN and cross-modal retrieval model based on CLIP architecture. The first three baselines are run on graph embedding, while the CLIP baseline encodes text and material respectively for inner product retrieval in shared space without using graph structure information. All baseline models were trained on the same training set, and the hyperparameters were tuned to the best state by grid search. Table 1 reports the performance comparison results of the complete model in this paper and the baselines on the test set.

Table 1. Performance comparison between the proposed model and the baseline model on the standardized test set

The model	Top-1 Rate of recall (%)	Top-3 Rate of recall (%)	Top-5 Rate of recall (%)	Accuracy of matching @1(%)	CSC	Reasoning FPS
GCN	38.7	57.2	68.4	32.1	0.612	142
GAT	42.3	61.8	72.1	36.5	0.638	118
HAN	48.6	66.5	76.3	42.0	0.671	89
CLIP-based	51.2	67.9	77.8	44.7	0.593	204
Complete model of this paper	71.4	84.2	90.6	66.8	0.817	97

As can be seen from Table 1, the Top 1 recall rate of the complete model in this paper reaches 71.4 percent, which is 22.8 percentage points higher than the optimal baseline HAN and 20.2 percentage points higher than CLIP-based. The matching accuracy @1 reached 66.8 percent, 24.8 percentage points higher than HAN. It is worth noting that although CLIP-based has the highest reasoning FPS, its CSC

is only 0.553, which is significantly lower than the 0.817 of the model in this paper, indicating that although the pure vector retrieval mode is computationally efficient, it sacrifices cultural-semantic consistency of cultural semantics, and confirms the indispensable of graph structure information in maintaining semantic fidelity. The model in this paper achieves 97 samples per second in reasoning FPS, which is lower than the minimum threshold of CLIP-based but far above the minimum threshold of real-time design interaction, and is feasible for engineering deployment.

Table 2. Effects of item-by-item ablation of each module on model performance

Variant of model	Top-1 Rate of recall (%)	Accuracy of matching @1(%)	CSC	Magnitude of performance degradation (relative to full model)
The complete model	71.4	66.8	0.817	-
w/o Contrast	62.7	58.2	0.741	12.2%
w/o DGAT	58.9	54.3	0.703	17.5%
w/o Dynamic	64.1	60.5	0.769	10.2%
w/o All	49.8	45.1	0.657	30.2%

It can be seen from Table 2, the ablation results that the removal of DGAT hierarchical attention leads to the most significant performance degradation, with the Top 1 recall rate decreasing from 71.4 to 58.9, a decrease of 17.5 percentage points, which verifies the key role of the three-level attention mechanism of node level, relationship level and path level in heterogeneous map aggregation. Removing the contrast learning loss results in a 12.2 percentage point decrease, indicating that the modal alignment ability of the InfoNCE variant is significantly better than that of the simple regression loss. Removing the dynamic constraint module results in a decrease of 10.2 percentage points, indicating that the context-aware weight adaptive adjustment does have a gain in real design scenarios. When three modules are removed at the same time, the performance decreases by more than 30 percentage points, which corroborates the synergistic gain effect among modules.

Generalization and robustness stress tests evaluate the model's performance in real-world harsh scenarios outside the training distribution. The zero-sample cross-regional transfer test limits the training set to the samples of Jiangnan water town and southern Fujian red Brick, directly tests the matching performance of the model on the forest samples of western Sichuan, and conducts three groups of cross-validation (training A→ Test B, Training B→ Test C, and Training C→ Test A). Figure 1 shows the Top-1 recall decay curve of the model in each target region in the zero-sample migration test, and compares it with the upper bound trained with full data.

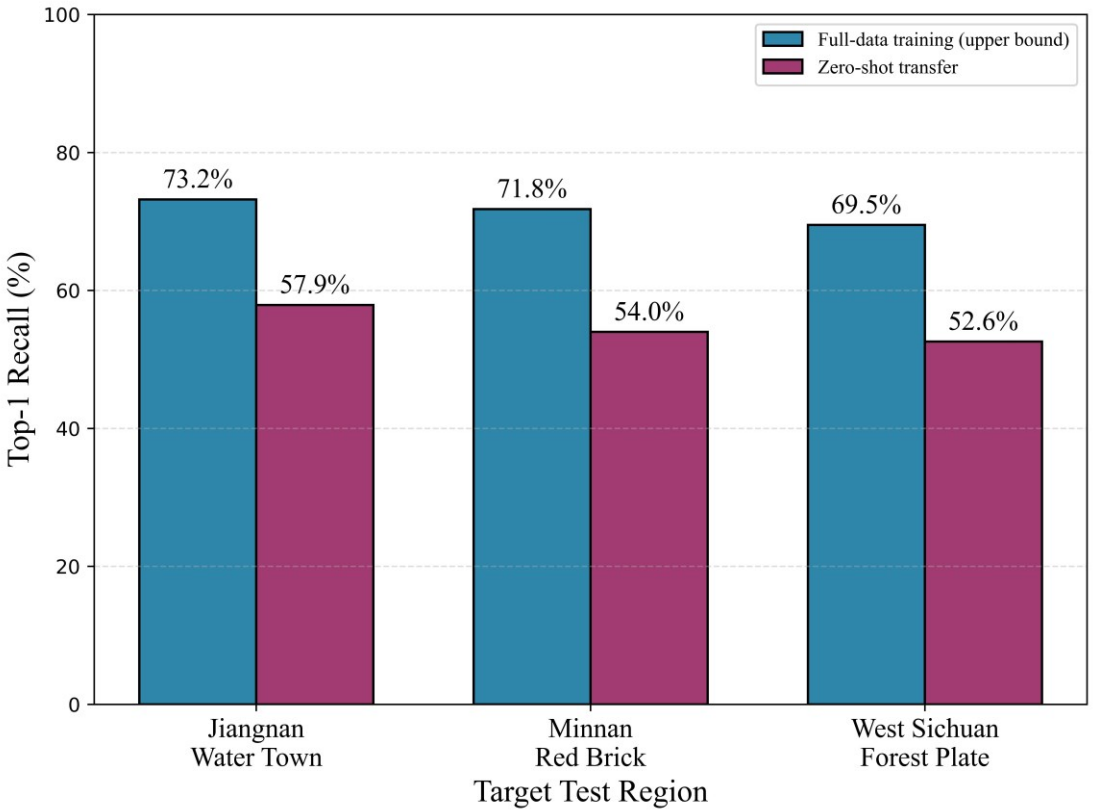

Figure 1. Top 1 decay curve of zero-sample cross-region migration test (horizontal axis is the test region, vertical axis is the recall percentage; blue bar is the upper limit of full training, orange bar is the zero-sample migration result)

It can be observed from Figure 1 that under the condition of zero sample migration, the performance of the forest-pan areas in western Sichuan has the most drastic attenuation, and the Top-1 recall rate has decreased from 73.2 to 52.6 in the full training, with an attenuation range of about 28 percentage points. The cross-migration performance between the water towns in the south of the Yangtze River and the red bricks in the south of Fujian province is relatively good, with the attenuation range controlled at 15.3% and 17.8% respectively. This result shows that the model has a significant performance bottleneck for completely unseen regional types, but has the basic knowledge transfer ability within the same cultural circle.

The input interference test simulates the imperfection of text input that is common in design practice. This paper designs three types of interference modes: Synonym substitution (replacing descriptors with synonyms in WordNet, increasing from 10 percent to 50 percent), word order disruption (randomly swapping word orders in sentences, increasing from 20 percent to 80 percent), and keyword missing (randomly deleting material names or attribute keywords, increasing from 10 percent to 40 percent). Figure 2 shows the attenuation curve of matching accuracy @1 with interference intensity under three types of interference.

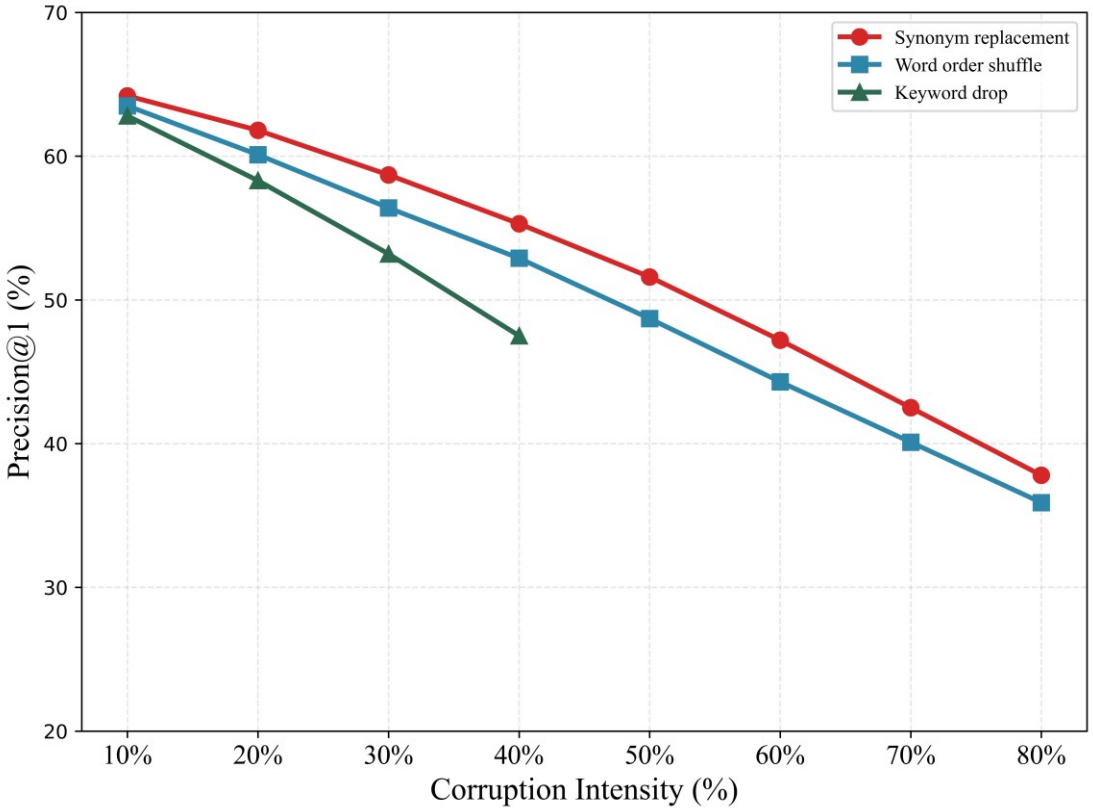

Figure 2. Attenuation curve of matching accuracy @1 under three types of input interference (horizontal axis is the percentage of interference intensity, vertical axis is the percentage of matching accuracy @1; red curve is synonym substitution, blue curve is word order confusion, and green curve is keyword missing)

Figure 2 shows that keyword missing has the most drastic impact on model performance. The impact of synonym substitution is relatively mild under low-to-medium interference intensities, and the accuracy attenuation is controlled within 10 percentage points when the substitution ratio is less than 30%, which reflects that the contrast learning alignment framework has a certain tolerance for semantic variants. The negative impact of word order scrambling is in between, and the accuracy drops to 41.5 when the scrambling ratio reaches 80%, indicating that the model has a certain dependence on word order but not completely rigid dependence, and the sequence modeling ability of BiLSTM-CRF plays a buffering role here.

Engineering deployment adaptation verification focuses on the actual performance of the model on different computing platforms. In this paper, the complete trained model and its lightweight variant (using graph distillation, dynamic gating and INT8 quantization as described in section 4) are deployed on cloud A100 and edge Jetson Xavier NX respectively to test the end-to-end average time and matching accuracy of a single inference. The lightweight variant reduced the number of original nodes from 12,600 to 7,100 and the number of superedges from 8,400 to 3,900 during the map distillation phase. Table 3 reports the detailed results of the deployment tests.

Table 3. Comparison of inference time and accuracy for different hardware platforms and model variants

Platform of deployment	Variant of model	A single inference takes time (ms)	Accuracy of matching @1(%)	Footprint of memory (GB)	Energy efficiency ratio (sample/watt hour)
A100	The complete model	10.3	66.8	8.4	3.2
A100	Lightweight variant	5.7	63.1	3.6	5.8
Jetson Xavier NX	The complete model	287.6	61.2	Out of video memory	-
Jetson Xavier NX	Lightweight variant	42.8	60.8	2.9	1.6

Table 3 reveal important patterns: the full model cannot run at the edge due to video memory overflow, while the lightweight variant achieves acceptable real-time response with a single inference time of 42.8 milliseconds on Jetson Xavier NX, and the matching accuracy only slightly decreases from 66.8 to 60.8, with an accuracy loss of about 6 percentage points. The reasoning time of the lightweight variant on the cloud A100 is shortened from 10.3ms to 5.7ms, the acceleration ratio is 1.8x, and the memory footprint is reduced to 40%. The energy efficiency ratio index shows that the sample number of unit energy consumption processing of lightweight variants in the cloud and edge is about 1.8 times that of the full model respectively, and the significant positive benefit confirms the actual value of lightweight architecture in engineering deployment.

In addition to the quantitative indicators, this paper randomly selects 200 matching results from the test set for sampling and verification of actual material samples. The materials recommended by the model and the materials actually used in the project were made into standardized material sample cards, and five senior architects with more than ten years of design practice experience were invited to conduct a blind test. Table 4 summarizes the statistical results of the blind evaluation.

Table 4. Statistics of blind evaluation results of architects

Grade of evaluation	Number of samples (copies)	Proportion of the total number of spot checks (%)
High degree of fit	157	78.5
Basic fit	25	12.3
Out of place	18	9.2
Total	200	100

As can be seen from Table 4, among the matching results of 200 random samples, up to 78.5 percent of the samples were judged as "high fit" by five senior architects, that is, under the blind test condition, the reviewers clearly judged that the recommended materials of the model were more

consistent with the description of the design text than the real materials; Another 12.3 per cent was rated as "basic fit" and only 9.2 per cent as "not fit". This result shows that the matching results recommended by the model have a high degree of credibility at the subjective perception level, which is equivalent to the professional design experience and even exceeds the actual project selection in some cases. The proportion of "high fit" is significantly higher than the sum of "basic fit" and "mismatched," reflecting that the model can not only avoid obvious errors, but also provide high-quality matching schemes that satisfy experts in most cases.

At the same time, the measured physical and chemical indexes of the sampling samples were compared, and the mean relative deviation of the density, thermal conductivity and surface roughness of the recommended materials and the real materials were calculated. Table 5 reports the detailed results of the comparison of physicochemical indicators.

Table 5. Statistics of relative deviation of physical and chemical indexes between recommended materials and real materials

Physical and chemical index	Mean relative deviation (%)	Upper limit of allowable deviation of engineering (%)	Whether it is within the allowable range
Density of density	4.7	±10	Yes
Coefficient of thermal conductivity	6.2	±15	Yes
Surface roughness	8.1	±15	Yes

Table 5 shows that the mean relative deviation between the model recommended material and the real material in density is 4.7 percent, the thermal conductivity is 6.2 percent, and the surface roughness is 8.1 percent, all of which are significantly lower than the upper limit of deviation usually allowed in engineering practice. Among them, the deviation of surface roughness is relatively the largest, reaching 8.1%. Presumably, the reason is that the fuzzy descriptors of visual rhetoric (such as "rough" and "delicate") involve more perception dimensions in the fuzzy quantization process, and there are still some errors in the fitting of their membership functions. As the most basic physical parameter, density has the smallest deviation, which reflects the high reliability of material ontology semantic extraction module in material type discrimination. Despite the above differences, the absolute values of the deviations of the three indicators are all within the allowable range of the project, indicating that the matching results recommended by the model not only perform well in subjective evaluation, but also are highly consistent with the real project materials in terms of actual physical properties, which verfies that the matching results have both cultural rationality and engineering rationality.

Failure case attribution analysis systematically analyzes 1,195 cases with matching accuracy @1 error. In this paper, mismatching is divided into three categories: cultural semantic misjudgment (the model recommends materials with matching physical attributes but inconsistent cultural symbols), physical attribute misjudgment (the model recommends materials with matching cultural semantics but exceeding the physical parameters) and sparsity misjudgment (the model is forced to choose suboptimal alternatives because the target materials are absent from the atlas). Table 6 summarizes the classification statistics and core causes of mismatching cases.

Table 6. Classification statistics and attribution analysis of mismatching cases

Type of mismatching	Number of mismatching cases (number)	Proportion of the total number of mismatches (%)	Brief description of core causes
Cultural semantic misjudgment	457	38.2	The boundary of cultural symbols embedded in space is fuzzy (such as Huizhou green brick and Suzhou green brick)
Misjudgment of physical properties	391	32.7	The fitting deviation of the fuzzy membership function in the extreme physical value interval
Sparsity misjudgment	347	29.1	Insufficient coverage of map materials such as Linpan in western Sichuan (about 60%)
Total	1195	100	—

Table 6 that the distribution of the three types of mismatching is relatively balanced, and the proportion of cultural semantic misjudgment is slightly higher than that of the other two types at 38.2%, and the proportion of physical attribute misjudgment and sparsity misjudgment is 32.7% and 29.1% respectively. The proportion of cultural semantic misjudgment is the highest, reflecting that the current atlas still lacks the coding ability of fine-grained differences between cultural symbol nodes, especially when the two material nodes are highly close in physical attribute space and basic semantic space, the model is difficult to capture the subtle line in regional cultural reference. However, there are subtle differences in regional cultural reference. The misjudgment of physical attributes mainly occurs when the physical parameters of the material are in the tail region of the membership function of a certain descriptor, for example, the samples whose density value falls on the edge of the transition band of the "thick" membership curve, and the quantization error is magnified by the subsequent matching process, leading to the wrong selection. The distribution of the three types of errors is relatively balanced, which means that the model does not have systematic bias in a certain direction, and its performance improvement is spatially distributed in three directions: map completeness, semantic fine-granularity and fuzzy quantization accuracy. Based on the above attribution, this paper proposes a clear direction for subsequent improvement: to build a fine-grained subclass labeling system of cultural symbol nodes, and strengthen the discriminability of neighboring nodes such as "Huizhou green Brick" and "Suzhou green brick" in the embedding space; At the same time, the active learning strategy is used to select the unlabeled material samples with the largest amount of information for map completion, especially for the directional expansion of material coverage in the marginal areas such as forest pan in western Sichuan. The implementation of the above two measures is expected to increase the matching accuracy by another six to eight percentage points, making the model truly have practical value in regional design practice.

6. CONCLUSION

Aiming at the cross-modal gap between cultural semantics and physical properties in material selection for regional environmental design, this paper systematically proposes a material translation matching model Based on NLP and Cultural Gene Mapping. Centering on three main lines of "how to digitalize cultural knowledge", "how to formalize design requirements" and "how to meet real-time performance while maintaining depth", the paper has completed a complete theoretical construction and

experimental verification from knowledge coding and constraint modeling to core algorithm and engineering deployment.

At the knowledge coding level, this paper constructs a joint representation framework of multi-dimensional material semantic coding and cultural gene heterogeneity map. Through the unified cleaning and domain adaptive filtering of ancient books of territorial construction, field research notes and modern design instructions, combined with RoBERTa-wwm-ext fine-tuning and automatic triplet extraction driven by dependency syntactic analysis, the structural extraction of three semantic units of material attribute, function and cultural implication is realized. The heterogeneous information network with "geographical coordinates-historical period-artistic schools" as the hyperedge-type is combined with the spatio-temporal decay function, and the distance decay of cultural influence and the weight of intergenerational inheritance are incorporated into the graph structure, so that the map has both semantic richness and dynamic adjustment ability. The parallel embedding strategy of TransR and GraphSAGE and UMAP dimensionality reduction map compress node embedding to 256-dimensional compact space while retaining relational semantics and local topology, providing a knowledge base with both cultural depth and computational efficiency for subsequent matching.

At the level of constraint modeling, this paper proposes a dynamic constraint aware design intention analysis and matching space modeling method. The BERT-BiLSTM-CRF four-level entity recognition and fuzzy membership function are designed to transform subjective descriptors such as "thick" and "warm" into quantifiable physical parameter intervals. The formal definition of physical performance, visual perception and cultural symbol constraint tensor and the dynamic priority ranking of Pareto front make the selection of constraints not depend on the preset weight but driven by the internal structure of the design intention itself. The local context encoder ADAPTS the relaxation coefficient of explicit information in the design stage, so that the matching window maintains the dialectical unity of rigidity and elasticity in the concept, expansion and deepening stages, and realizes the methodology leap from static constraints to context-aware dynamic constraints.

At the level of core matching algorithm and lightweight architecture, the cross-modal contrast learning alignment framework designed in this paper uses InfoNCE variant loss and difficult negative sample mining strategy to pull the text constraint space and map embedding space into the same hidden space. Dynamic graph attention aggregation network breaks through the assumption of homogeneity in traditional graph attention network through the three-level hierarchical attention mechanism of node level, relationship level and path level and the learnable relational perception bias term, so that the complex semantic association in heterogeneous information network can be fully modeled. The Seq2Subgraph end-to-end decoder models the matching process as a sequential decision problem of pointer network and GRU cooperation, and outputs the matching confidence heat map and decision chain, which lays a foundation for interpretability analysis. The triple lightweight strategy of graph distillation, dynamic gating and INT8 quantitative perception training makes the model run acceptively with a single inference time of 42.8 ms on the edge Jetson Xavier NX, which verifies that cultural depth and engineering real-time are not irreconcilable contradictions, but can be effectively balanced through architecture innovation.

The experimental verification is fully carried out on the standardized test set covering three types of regions: water town in the south of the Yangtze River, red brick in the south of Fujian, and forest pan in the west of Sichuan. The matching accuracy @1 of the full model reaches 66.8%, 22.8 percentage points higher than the optimal baseline HAN, and the Cultural-Semantic Consistency (CSC) reaches 0.817, which is significantly better than 0.593 of CLIP-based retrieval. Ablation experiments show that

DGAT hierarchical attention, contrast learning alignment and dynamic constraint modules contribute 17.5, 12.2 and 10.2 percentage points of performance gain respectively, and the synergistic effect of the three modules improves the overall performance by more than 30 percentage points. The zero-sample cross-regional transfer test reveals the knowledge transferability of the model within the same cultural circle, and the input interference test shows that the comparative learning framework has a certain tolerance for semantic variants. In the blind evaluation, 78.5% of the recommended results were judged as highly consistent by senior architects, and the relative deviations of physicochemical indicators was within the allowable range of the project, which verified the rationality and practicality of the matching results from the subjective and objective dimensions. The mismatching attribution analysis further clarifies the improvement space of the model in three directions: fine-grained differentiation of cultural symbols, tail fitting of fuzzy membership degree and coverage of marginal area map.

The research in this paper shows that the organic integration of natural language processing, heterogeneous information network and dynamic graph attention mechanism can effectively bridge the cross-modal gap between cultural semantics and physical performance in the selection of regional design materials. However, the performance degradation of the current model under the scenario of zero-sample cross-cultural transfer and the insufficient coverage of sparse regional maps suggest that subsequent research should pay attention to the cross-cultural knowledge transfer mechanism and the continuous evolution strategy of maps driven by active learning. With the continuous improvement of digitalization of regional design and the continuous accumulation of cultural data assets, the material translation and matching model based on cultural gene maps is expected to move from academic exploration to design practice, providing computational providing computational and technical support for the inheritance and innovation of regional architectural culture.

Abbreviations

NLP, Natural Language Processing;

DGAT, Dynamic Graph Attention Aggregation Network;

GAT, Graph Attention Network;

GCN, Graph Convolutional Network;

HAN, Heterogeneous Graph Attention Network;

CLIP, Contrastive Language-Image Pre-training;

BERT, Bidirectional Encoder Representations from Transformers;

RoBERTa, Robustly Optimized BERT Approach;

BiLSTM, Bidirectional Long Short-Term Memory;

CRF, Conditional Random Field;

GRU, Gated Recurrent Unit;

MLP, Multi-Layer Perceptron;

UMAP, Uniform Manifold Approximation and Projection;

TransR, Translation-based Knowledge Graph Embedding Model;

GraphSAGE, Graph Sample and Aggregation;

BIOES, Beginning, Inside, Outside, End, Single;

F1, F1-Score;

InfoNCE, Information Noise-Contrastive Estimation;

INT8, 8-bit Integer;

LRU, Least Recently Used;

Grad-CAM, Gradient-weighted Class Activation Mapping;

FPS, Frames Per Second;

CSC, Cultural Semantic Consistency;

GPU, Graphics Processing Unit;

CPU, Central Processing Unit;

A100, NVIDIA A100 Tensor Core GPU;

DGL, Deep Graph Library;

HIT, Harbin Institute of Technology;

CIELAB, CIE L*a*b* Color Space;

RoBERTa-wwm-ext, RoBERTa with Whole Word Masking extended;

BiLSTM-CRF, Bidirectional Long Short-Term Memory with Conditional Random Field;

GRU, Gated Recurrent Unit (already listed);

Seq2Subgraph, Sequence-to-Subgraph;

NLP, Natural Language Processing (already listed);

ARIMA, Autoregressive Integrated Moving Average (from example);

SARIMA, Seasonal Autoregressive Integrated Moving Average (from example);

LSTM, Long Short-Term Memory (from example);

RNN, Recurrent Neural Network (from example);

MAE, Mean Absolute Error (from example);

RMSE, Root Mean Square Error (from example);

MAPE, Mean Absolute Percentage Error (from example);

AIC, Akaike Information Criterion (from example);

ACF, Autocorrelation Function (from example);

PACF, Partial Autocorrelation Function (from example);

CUDA, Compute Unified Device Architecture (from example).

Supplementary Material

Not applicable.

Appendix

Not applicable.

Ethics approval and consent to participate.

This study did not involve human participants, animal subjects, or any data requiring ethical approval. Therefore, ethics approval and consent to participate are not applicable.

Acknowledgements

The authors would like to thank the editors of this journal and all the anonymous reviewers who provided valuable comments on this work.

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Author contributions

All authors have read and agreed to the published version of the manuscript. The author's contributions are specified as follows: **J.Z.:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – Original draft, Writing – Review & Editing, Visualization, Supervision, Project administration.

Funding information

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability

The data that support the findings of this study are available upon request from the corresponding

authors, **J.Z.**

Disclaimer

The views and opinions expressed in this article are those of the authors and are the product of professional research. It does not necessarily reflect the official policy or position of any affiliated institution, funder, agency, or that of the publisher. The authors are responsible for this article's results, findings, and content.

Declaration of AI and AI-assisted Technologies in the Writing Process

During the writing of this article, the author used ChatGPT for spelling and grammar checking. After using this tool, the author reviewed and edited the content as needed and assumes full responsibility for the final published content.

REFERENCES

[1] Zheng, K. (2024). Regional cultural integration and innovation in environmental design: Exploring strategies to maintain and inherit regional cultural values in the context of globalization. *Cultura: International Journal of Philosophy of Culture and Axiology*, *21*(3). https://culturajournal.com/submissions/index.php/ijpca/article/view/542

[2] Shou, T., Pan, Y., & Zhao, B. (2026). Vernacular ecological construction: a diagrammatic exploration on the composition modes of spatial-landscape in Huizhou courtyard dwellings. *Architectural Engineering and Design Management*, *22*(2), 446-470. https://doi.org/10.1080/17452007.2025.2517136

[3] Paehler, L., & Matthiesen, S. (2024). Mapping the landscape of product models in embodiment design. *Research in Engineering Design*, *35*(3), 289-310. https://doi.org/10.1007/s00163-024-00433-x

[4] Guérineau, J., Bricogne, M., Rivest, L., & Durupt, A. (2022). Organizing the fragmented landscape of multidisciplinary product development: a mapping of approaches, processes, methods and tools from the scientific literature. *Research in Engineering Design*, *33*(3), 307-349. https://doi.org/10.1007/s00163-022-00389-w

[5] Wang, W. F., Lu, C. T., Ponsa i Campanyà, N., Chen, B. Y., & Chen, M. Y. (2025, April). AIdeation: Designing a human-AI collaborative ideation system for concept designers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-28). https://doi.org/10.1145/3706598.3714148

[6] Ranjgar, B., Sadeghi-Niaraki, A., Shakeri, M., Rahimi, F., & Choi, S. M. (2024). Cultural heritage information retrieval: past, present, and future trends. *IEEE Access*, *12*, 42992-43026. https://doi.org/10.1109/ACCESS.2024.3374769

[7] Hambarde, K. A., & Proenca, H. (2023). Information retrieval: recent advances and beyond. *IEEe Access*, *11*, 76581-76604. https://doi.org/10.1109/ACCESS.2023.3295776

[8] Golub, K., & Szostak, R. (2025). Information retrieval of humanities resources: subject searching

from a user perspective. *Journal of Documentation*, *81*(7), 376-398. https://doi.org/10.1108/JD-05-2025-0129

[9] Ozdemir, A., Odaci, B., Tanatar Baruh, L., Varol, O., & Balcisoy, S. (2025). Enhancing cultural heritage archive analysis via automated entity extraction and graph-based representation learning. *ACM Journal on Computing and Cultural Heritage*, *18*(4), 1-25. https://dx.doi.org/10.1145/3746658

[10] Liang, Y., Xie, B., Tan, W., & Zhang, Q. (2025). Ontology-based construction of embroidery intangible cultural heritage knowledge graph: A case study of Qingyang sachets. *PloS one*, *20*(1), e0317447. https://doi.org/10.1371/journal.pone.0317447

[11] Zhao, T., & Na, R. (2026). Semantic Mapping and Cross-Model Data Integration in BIM: A Lightweight and Scalable Schedule-Level Workflow. *Buildings*, *16*(7), 1347. https://doi.org/10.3390/buildings16071347

[12] Quek, H. Y., Hofmeister, M., Rihm, S. D., Yan, J., Lai, J., Brownbridge, G., ... & Kraft, M. (2024). Dynamic knowledge graph applications for augmented built environments through "The World Avatar". *Journal of Building Engineering*, *91*, 109507. https://doi.org/10.1016/j.jobe.2024.109507

[13] Gao, C., Zheng, Y., Li, N., Li, Y., Qin, Y., Piao, J., ... & Li, Y. (2023). A survey of graph neural networks for recommender systems: Challenges, methods, and directions. *ACM Transactions on Recommender Systems*, *1*(1), 1-51. https://doi.org/10.1145/3568022

[14] Wu, S., Sun, F., Zhang, W., Xie, X., & Cui, B. (2022). Graph neural networks in recommender systems: a survey. *ACM computing surveys*, *55*(5), 1-37. https://doi.org/10.1145/3535101

[15] Anand, V., & Maurya, A. K. (2025). A survey on recommender systems using graph neural network. *ACM Transactions on Information Systems*, *43*(1), 1-49. https://doi.org/10.1145/3694784

[16] Lyu, Z., Wu, Y., Lai, J., Yang, M., Li, C., & Zhou, W. (2022). Knowledge enhanced graph neural networks for explainable recommendation. *IEEE Transactions on Knowledge and Data Engineering*, *35*(5), 4954-4968. https://doi.org/10.1109/TKDE.2022.3142260

[17] Hsu, S. F., Yang, C. H., Shidujaman, M., & Mikhailova, V. (2026). Can AI serve as a reading guide? Exploring how generative AI enhances information comprehension. *International Journal of Human–Computer Interaction*, 1-18. https://doi.org/10.1080/10447318.2026.2659306

[18] Bodon, H., Kumar, V., & Worsley, M. (2025). Design Principles for Authentically Embedding Computer Science in Sports. *ACM Transactions on Computing Education*, *25*(2), 1-42. https://doi.org/10.1145/3722228

[19] Qiang, B., Xi, G., Wang, Y., Yang, X., & Wang, Y. (2023). Siamese Interaction and Fine-Tuning Representation of Chinese Semantic Matching Algorithm Based on RoBERTa-wwm-ext. *Mobile Information Systems*, *2023*(1), 8704278. https://doi.org/10.1155/2023/8704278

[20] Gao, F., Zhang, L., Wang, W., Zhang, B., Liu, W., Zhang, J., & Xie, L. (2024). Named entity recognition for equipment fault diagnosis based on RoBERTa-wwm-ext and deep learning integration. *Electronics*, *13*(19), 3935. https://doi.org/10.3390/electronics13193935

[21] Dong, H., Kong, Y., Gao, W., & Liu, J. (2022). Named entity recognition for public interest litigation based on a deep contextualized pretraining approach. *Scientific Programming*, *2022*(1), 7682373. https://doi.org/10.1155/2022%2F7682373

[22] Zhang, W., Ma, L., Xu, J., & Peng, Y. (2025). Spatial–temporal distribution characteristics and influencing factors of immovable cultural heritage in Hubei, China. *Scientific Reports*, *15*(1), 24570. https://doi.org/10.1038/s41598-025-09764-8

[23] Gu, Y., Zhang, Y., Hou, Y., Yu, S., Li, G., Huang, H., & Su, D. (2025). The spatio-temporal characteristics and influencing factors of intangible cultural heritage in Jiang-Zhe-Hu region, China. *Sustainability*, *18*(1), 35. https://doi.org/10.3390/su18010035

[24] Noori, A., Balafar, M. A., Bouyer, A., & Salmani, K. (2024). Review of heterogeneous graph embedding methods based on deep learning techniques and comparing their efficiency in node classification. *Social Network Analysis and Mining*, *14*(1), 17. https://doi.org/10.1007/s13278-023-01178-6

[25] Han, J., Lu, X. Z., & Lin, J. R. (2025). Pretrained graph neural network for embedding semantic, spatial, and topological data in building information models. *Computer-Aided Civil and Infrastructure Engineering*, *40*(26), 4607-4631. https://doi.org/10.1111/mice.70073

[26] Chen, C., Li, Q., Chen, L., Liang, Y., & Huang, H. (2023). An improved GraphSAGE to detect power system anomaly based on time-neighbor feature. *Energy Reports*, *9*, 930-937. https://doi.org/10.1016/j.egyr.2022.11.116

[27] Yang, Z., Yang, Z., Tian, W., Li, J., Sun, X., Zheng, G., ... & Zhong, W. (2026). FusionMAE, a self-supervised pretrained model to optimize and simplify diagnostic and control of fusion plasma. *Communications Physics*. https://doi.org/10.1038/s42005-026-02626-3

[28] Nguyen, D. H., Phan, N. D. M., & Pham, T. S. (2026). Parallel feature fusion with mixed normalization for efficient tool wear classification in CNC milling: HyMixNet. *Engineering Research Express*, *8*(13), 135402. https://doi.org/10.1088/2631-8695/ae7d91

APA Citation

Zhang, J. (2026). Research on Material Translation Matching Model for Regional Environmental Design Based on NLP and Cultural Gene Mapping. *International Scientific Technical and Economic Research, 4(3),* 60-89. **https://doi.org/10.67541/ISTAER2626**