A Behaviorally Evidence-based Method for Computing Spatial Comparisons of Image Scenarios

Ziyang Weng[1], Shuhao Wang[1], Ziyu Zhang[1], Renyi Liu[2]*

[1] School of Information Management, Wuhan University, China
[2] School of Earth Sciences, Zhejiang University, China
weng_ziyang@whu.edu.cn, wangshuhao@whu.edu.cn, zhang_ziyu@whu.edu.cn, liurenyi@zju.edu.cn

Abstract

Large amounts of noise and a lack of contextual domain knowledge lead to slow and inefficient cross-domain image learning. This paper proposes an image scenario spatial data classification model based on evidence-based behavioral logic, intervenes in image annotation through evidence-based dynamic knowledge graphs, and uses spatial similarity measurement to evaluate the effectiveness and robustness of the method. The results show that: 1) Organizing the dynamic knowledge graphs of contextual domain knowledge by behavioral logic can significantly improve the association efficiency of each model. 2) The calculation method of image scenario space comparison based on behavior evidence can decrypt the implicit knowledge of images and significantly improve the effectiveness of image scenario space interpretation. The research results are helpful to guide the design and implementation of cross-domain image interpretation systems and improve the efficiency of information sharing.

Keywords: Behavioral logic, Image interpretation, Evidence-based computing, Image scenario, Narrative inference

1 Introduction

In the field of image interpretation intelligence, it is an important task to recognize the spatial information in ancient landscape paintings. However, remote sensing digital elevation models (DEM) usually lack scenario description information, and experts in the art field seldom pay attention to the geographical information of landscape painting. As a result, it is difficult to identify and locate the scenario space of ancient landscape paintings with a single domain knowledge, resulting in a blank in this research field.

In this paper, we propose a spatial mapping method for Chinese landscape painting scenarios based on the mountain and hydrological features, establish an evidence-based mechanism of spatial-temporal constraints through the movement track and behavior habits of the spatial recorder (the painter), and use relevant algorithms to process multi-level continuous complex structural features, thus realizing the geographical positioning and boundary discrimination of landscape painting images.

The complex behaviors involved in the landscape images are generated into adversarial networks [1] and sparse detection networks [2], combined with the DEM spatial datasets [3], to convert the deep network into a series of heat maps representing probabilities through which key coordinate points are captured. A self-focused generative adversarial network is used to extract key location nodes instead of the common generative adversarial network to achieve composite results of spatially comparable similarity evidence [4], and the effectiveness is verified by an evaluation mechanism.

The paper is organized as follows. In Section 2, we propose the motivation generated by the research gap of scenario recognition in traditional Chinese landscape paintings. Our approach is described in detail in Section 3, where we propose an Image Scene Spatial Interpretation model and describe the process of acquiring 2.5D key features with the intervention of contextual domain knowledge. In the fourth and fifth sections, our case studies complete the mapping between 2D landscape paintings and 3D remote sensing (RS) spaces and verify the validity of the mapping by using the similarity index. Section 6 discusses the limitations. In section 7, we summarize the approach and describe the challenges ahead.

2 Motivation

2.1 Related Studies

In the current studies of deep learning of Chinese landscape painting based on adversarial generative networks, the top research trends are in machine original art, and their main work focuses on image edge extraction and technical processing, such as GAN landscape image generation based on landscape photo edge extraction [5] and the painting styles migration based on deep learning of landscape image painting styles [6]. The above studies have been technically successful, but they lack emphasis on the content of the image scenarios. Uncertainty about the location of scenarios depicted in Chinese landscape paintings leads to a large number of semantic deficits and inefficiencies.

Therefore, given the problem that the spatial positioning of landscape painting cannot be carried out by single-domain knowledge due to the deformation of mountains and rivers caused by the art processing of the paintings, we put forward the idea that the contextual-domain dynamic

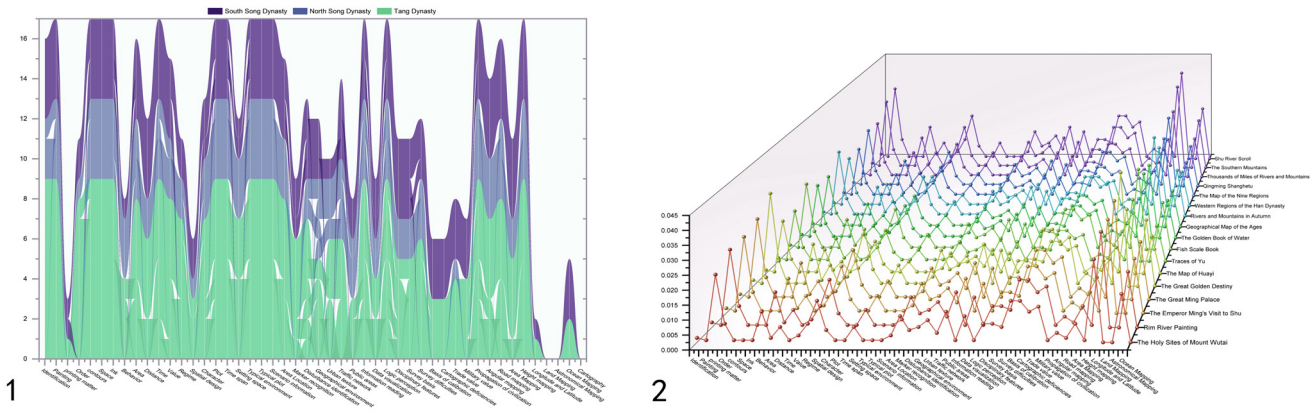
*Corresponding Author: Renyi Liu; E-mail: liurenyi@zju.edu.cn
DOI: 10.53106/160792642023092405009

knowledge graphs, based on the evidence of author behavior trajectories, can participate in image annotation. The context domain knowledge of images is defined as data sets related to cartographic behaviors, such as geopolitics, hydrology, mountain topography, economic ecology, etc. Based on the evidence-based mechanism, the mapping behavior of the painter on the mountains and rivers is deduced, and the 2.5D key feature sets between the 2D images and the 3D remote sensing spaces are constructed to fill the semantic gap.

The integrated analysis of multi-flow, multi-temporal contextual data requires spatial autocorrelation theory for spatial and temporal constraints. The smart city construction with digital twin proposed by Deren Li provides a research blueprint for space humanities research [7]; the spatial location superposition protocol proposed by Jianya Gong enables the synergy of multi-scale, multi-stream, and multi-temporal geographic data and provides a digital environment construction standard for superposition of spatial humanities datasets [8]; the construction of geographic multi-flow network proposed by Kun Qin becomes a theoretical tool for understanding time-varying patterns and correlation patterns of human behaviors [9]; Jiansong Li's theory of reliable dynamic monitoring for national geographical state provides technical support for understanding ancient geographic changes under time-varying conditions [10]; Xudong Lai's high-precision noise removal technology based on point cloud feature selection [11] provides an inspiring perspective for the 3D spatial data denoising work in this paper.

2.2 Preliminary Work

To build a contextual domain knowledge foundation for semantic annotation of landscape painting image scenario spaces, we selected all typical landscape paintings and maps from the Sui Dynasty to the Song Dynasty in China to form a testing dataset. The images in this dataset were first conducted for eye-tracking experiments. And the key feature label sets were constructed from 50 indexes (Figure 1), such as visual perception recognition, scenario placement demand analysis, social limitations of dynasties, and mapping technology limitations [12], to establish the basis for DEM scene reconstruction of landscape spaces [13].



(a) The number of features collected from landscape images of different Chinese dynasties

(b) TF- IDF weights of different landscape images on each of the 50 feature labels

Figure 1. Mapping norms and changing patterns of ancient landscape paintings and maps discovered by the TF-IDF algorithm

3 Our Approach

In this section, we propose an overall understanding path for spatial interpretation of complex image contexts. The process of image scenario space construction includes 1) Data acquisition for DEM scene reconstruction; 2) Semantic segmentation for image scenario spaces; 3) Target feature detection; 4) Multi-target tracking; 5) Model weighted fusion.

3.1 The Image Scenario Space Interpretation Model

The Image Scenario Space Interpretation Model (ISSI) is proposed (Figure 2). Robust and efficient DEM scene reconstruction allows professional image mapping intelligence to effectively interact with the RS data scenes. The amount of DEM data grows with the size and accuracy of the acquisition area. Similarly, due to the unique cartographic rules of landscape painting, the changes in the resolution of the input landscape images also cause changes in network depths. Therefore, a direct comparison is not feasible. This paper continues to explore the application of behaviorally logic-based evidence for coupled distribution and similarity comparison understanding tasks of landscape images and DEM scenes.

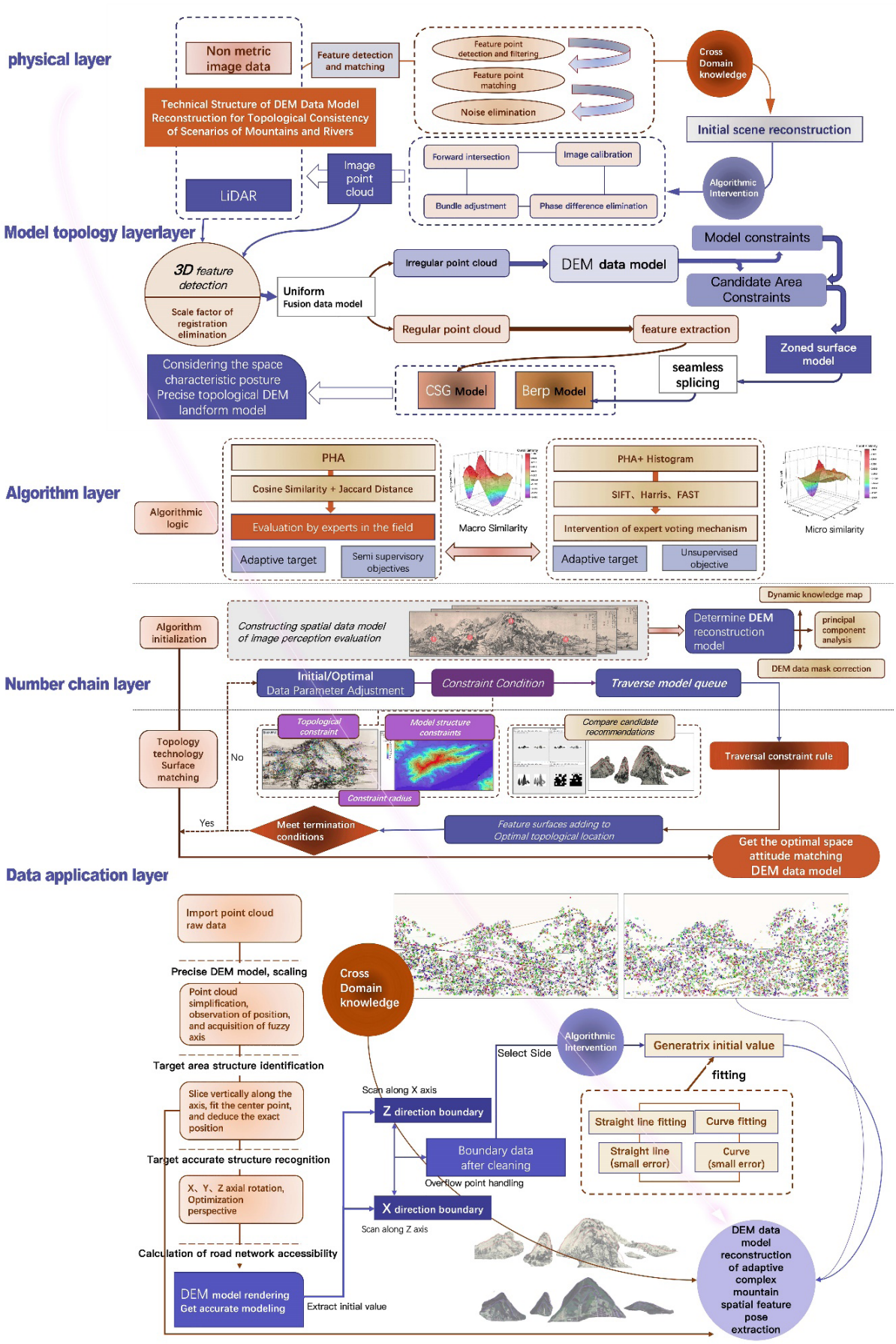


Figure 2. The image scenario space interpretation model

3.2 Evidence-based Spatial Comparative Similarity of Scenarios

The group conducted several experiments over years, focusing on the following research questions: 1) To explore the effect of recommendation labels on the cross-domain annotation characteristics of researchers by comparing groups sensing the image sets with and without the recommendation labels. 2) Experiments on the topological efficacy of spatial-temporal location labels were conducted to investigate the effects of different recommendation labels on the annotation process and the study findings of researchers.

University students majoring in social sciences became the participants. Image contextual domain knowledge recommendation label sets were established by domain experts. Eye-tracking experiments based on the tags were designed, including unsupervised groups, semi-supervised groups, and strongly supervised groups. With the intervention of dynamic domain knowledge (Figure 3. 1-1), through the results and the analysis model of behavioral geography under temporal and spatial constraints (Figure 3. 2-1), we transform the analysis results into the basis of image scenario calculation. Semantic computing (Figure 3. 3-1) is used to interpret the scenario space of the images [14].

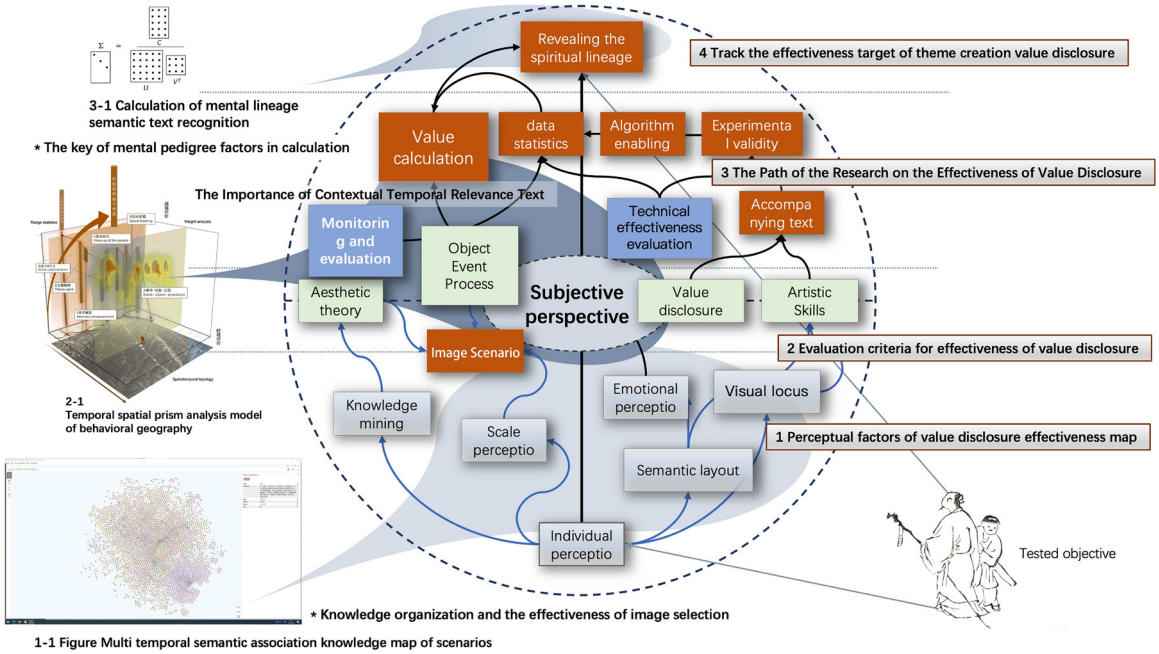


Figure 3. Dynamic knowledge link complementary to support the spatial traceability of landscape scenarios

This paper evaluates the eye-trajectory clustering results at three levels. For labeling features, the present study focused on the length of the visual trajectories, i.e., the number of consecutive nodes in the location labels. For label semantics, this study focused on label sources and label semantic similarity metrics. For gazing trajectories, this study focused on the number of labels added and labeled locations in multiple sets of experimental data of the subjects.

3.3 Mapping Method between 2D Landscape Image and 3D Spatial Data Based on Adversarial Relative Depth Constraint Network

A method for mapping 2D landscape images and 3D spatial data based on adversarial relative depth constrained network, the technical process of which includes:

(1) Input 2D pixel coordinates of 2D landscape image nodes (including the first type of coordinate points: peaks, ridges, valleys, cliffs, etc.; the second type of coordinate points: road network intersections, water systems, human settlement areas, agriculture areas, etc.) and perform normalized pre-processing.

(2) Input the marked geo-survey coordinate nodes identified by the domain expert. The normalized pre-processed 2D pixel coordinates of each node obtained in the previous step are input to the depth prediction network composed of three modules for predicting the depth values of the mountain feature nodes, including the following steps:

1. Each node's normalized pre-processed 2D pixel coordinates are input to the feature extraction module to extract features, consisting of a fully connected layer containing 1024 neurons and a linear rectification activation function layer.

2. The features extracted by the feature extraction module are input to the sparse network module for feature learning, which consists of two sparse blocks. Each sparse block inputs the output values of the previous layer of the neural network to a fully connected layer containing 1024 neurons and a linear rectification activation function layer and outputs the preliminary feature values. The preliminary feature values are then fed to a fully connected layer containing 1024 neurons and further feature values are output. The further feature values are then added to the input values which have been

input to the sparse block. Finally, the summed feature values are input to a layer of linear rectification activation function layer and output the sparse block eigenvalues to the next layer of the neural network.

3. Input the output feature values of the sparse network module to the depth-value regression module, which consists of a fully connected layer containing the corresponding number of neurons. The depth-value regression module inputs the output features of the sparse network module and outputs the depth values of the corresponding number of nodes.

4. Reconstruct the 3D coordinates of the nodes using the depth values of the nodes and the 2D pixel coordinates to obtain the reconstructed sparse feature geographical gestures of the 2D images in a 2.5D viewpoint.

5. Input the DEM data model, select the matching viewpoint, and then match the 3D coordinates of the tagged nodes to obtain the reconstructed sparse feature geographical gestures of the 3D spaces in a 2.5D viewpoint.

6. The authenticity error is calculated by using the discriminator of the generative adversarial network. The relative depth error is calculated using the reconstructed geographical feature gestures and the relative depth information of the nodes corresponding to the images.

7. The authenticity and relative depth errors are summed to get the total error and fed back to the depth prediction network. Constraints are placed on the depth prediction network to predict more accurate depth values to more accurate mapping evaluations and achieve mapping discovery.

After completing the sampling using the same-region temporal images at the preset scale, sampling can be performed at larger scales to calculate the spatial semantic similarity at different scales. Meanwhile, the size of the pre-defined scales can be scaled to increase the size of the same-region temporal images. After completing the calculation of the spatial semantic similarity at each scale, it can be stored as a multi-scale semantic similarity model. Thus, the spatial semantic similarity can be analyzed, clustered, and compared at specific scales and different scales.

3.4 Weighted Integrated Learning

In this paper, algorithms in the framework are treated as contributing models, which are fused into a generic model for image scenario space mapping. We choose CRITIC [15] as the algorithm weighting method, which is an objective weighting method for evaluation indexes proposed by Diakoulaki in 1995. Its basic ideas are:

(1) Normalize each indicator to be assigned with the formula:

$$x_{ij}^{'} = \frac{X_{ij} - \min(X_{1j}, X_{nj}, ..., X_{nj})}{\max(X_{1j}, X_{nj}, ..., X_{nj}) - \min(X_{1j}, X_{nj}, ..., X_{nj})}. \quad (1)$$

For negative indicators, the formula is:

$$x_{ij}^{'} = \frac{\max(X_{1j}, X_{nj}, ..., X_{nj}) - X_{ij}}{\max(X_{1j}, X_{nj}, ..., X_{nj}) - \min(X_{1j}, X_{nj}, ..., X_{nj})}. \quad (2)$$

(2) Determining contrast intensity of indicators: the variability is expressed in terms of standard deviation, and S_j denotes the standard deviation of the jth indicator.

$$\begin{cases} \overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \\ S_j = \sqrt{\frac{\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2}{n-1}} \end{cases}. \quad (3)$$

(3) Determining the conflicting character of indicators: the correlation coefficient is used to express the variability. r_{ij} denotes the correlation coefficient between indicator i and indicator j.

$$R_j = \sum_{i=1}^{p}(1 - r_{ij}). \quad (4)$$

(4) Amount of information: the greater the amount of information C_j , the greater the role of its indicators in the overall evaluation index system, and the more weight should be assigned to it. The formula is:

$$C_j = S_j \sum_{i=1}^{p}(1 - r_{ij}) = S_j \times R_j. \quad (5)$$

(5) Weight: the objective weight W_j of the indicator is calculated. The formula is:

$$W_j = \frac{C_j}{\sum_{j=1}^{p} C_j}. \quad (6)$$

4 Measure Evaluation

In this section, we use four experiments to evaluate the comprehensive efficacy of our method by testing single and multiple images.

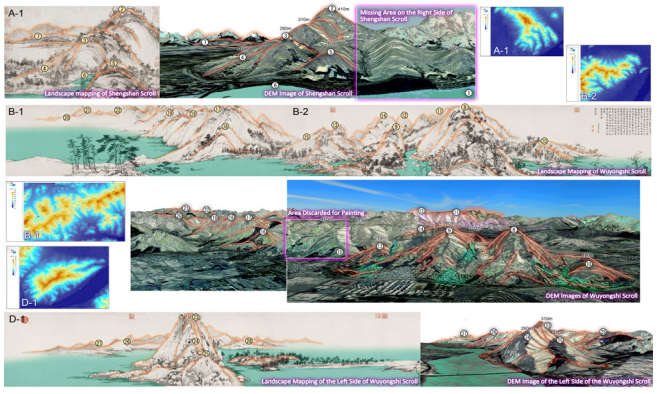


Figure 4. Evaluation of computational effectiveness of image scenario space decoding composite based on the intervention of expert voting mechanism

4.1 Experiment Setup

We selected the famous Chinese landscape painting "Dwelling in the Fuchun Mountains (Wuyongshi - Shengshan Scroll)" as an experimental case. The samples are divided into two categories (Figure 4). Category A represents the sample slices of the ancient painting; Category B represents the slices of DEM images in Fuyang, China (Table 1).

Table 1. Experimental datasets

Categories	Type	Amount	Used in
A	Origin image	4	Experiment 1
B	Origin image	4	Experiment 1
A	Origin image	4	Experiment 2
B	Origin image	4	Experiment 2
A	Featured profiles	12	Experiment 2
B	Featured profiles	12	Experiment 2
A	Featured profiles	6	Experiment 4
B	Featured profiles	3	Experiment 4
A	Featured points	2	Experiment 4
B	Featured points	1	Experiment 4

Among them, the samples were adjusted to the same size, black and white, without noises such as the inscriptions, seals, and small damages. The pictures were saved in JPG format with a resolution of 600 dpi or more.

4.2 Research Questions

(1) RQ1: How to obtain key feature nodes by eye-movement experiments with deep semantic annotation?

(2) RQ2: How to compute and verify the mapping effectiveness of 2D images to 3D space?

(3) RQ3: How to perform fusion evaluation of mapping efficacy?

(4) RQ4: How to stitch physically fragmented 2D landscape images by 3D spatial images?

RQ1 is designed to obtain key feature points and contours, and RQ2 and RQ3 determine the validity of key features by the mapping results of 2D images and 3D RS scenes. RQ4 enables data association of different images based on the method of geo-localization performance. These four questions cover the process and application of the method.

4.3 Evaluation Indicators

Many RS image classification methods have been proposed by scholars. The classification methods can be classified into supervised and unsupervised classification according to the need of selecting labeled samples or not. Based on the way of expressing and learning features, the methods can be classified into manual feature-based, machine-learning-based, and deep learning-based, which are not strictly different but learn from each other.

The method used in this paper is a semi-supervised feature-based classification method. To compare the matching efficiencies, through repeated experiments on various distance and similarity algorithms, we choose the following three similarity algorithms as the metrics of this paper: Cosine similarity and Jaccard distance as precise feature similarity indexes and Perceptual Hash similarity as overall shape similarity index (RQ1, RQ2 and RQ4) (Table 2). In the comprehensive evaluation, we chose the weighted model fusion method, CRITIC. By convention, a similarity of 0.6-0.69 is considered acceptable, 0.7-0.79 is good, and 0.8-1 is excellent.

Table 2. Evaluation indicators and reasons

Subject	Indicator	Perspective
Precise features	Cosine similarity	Reduce the effect of the size distance of the two images and calculate the vector angle [16]
	Jaccard similarity	Calculate the coordinate similarity of two images of the same size [17]
Overall shape	Perceptual Hash similarity	Calculating the similarity of subjectively perceived shapes from the viewer's visual system [18]

5 Results and Analysis

This section shows the experimental results. The experimental materials can be found at https://github.com/Ziyang0304/A-Behaviorally-Evidence-based-Method-for-Computing-Spatial-Comparisons-of-Image-Scenarios.git.

5.1 RQ1: Feature Acquisition Evaluation

The eye-tracking experiments were established. A Tobii X30-120 eye-tracker with a sampling rate of 120Hz was used. The key feature nodes and contours of landscape images and RS images are obtained respectively.

A total of 72 eye trajectory maps and 72 heat maps were obtained from nine experimental groups through eye movement experiments. The result maps were placed in Photoshop, and the same layers were overlaid to confirm the key feature coordinates and gestures by superposition calculation.

Table 3. Image scenario data acquisition pseudo-code

Algorithm 1. Image scenario data acquisition pseudo-code

Input: Images with .png suffix, eye-tracking observation data
Output: Objective and subjective similarity

```
Function CreateRelation (pic_dem, num)
  a=cv.imread(pic_dem) //read image
  Get the point in image
  While a<pic_dem.shape[1]
    While a<pic_dem.shape[0]
      x.append(a)
    End while
  End while
  Get point based on the spirit
  a=len(pic_dem)
  While i<a
    If d is_key()
      y.append[d]
    End if
  End while
  Calculate the similarity of feature points
  RS= 1 - spatial.distance.cosine(vec1, vec2)
  Return RS
```

In Table 3, line 2 reads the image, lines 3-8 extract all the coordinates in the image, lines 9-16 get the corresponding feature points based on the possibility, and line 16-17 calculates the similarity. Line 18 outputs the data.

Answer to RQ1: Through experiment and calculation, the key feature nodes and contours are captured, which form the 2.5D key feature dataset.

5.2 RQ2: Mapping Performance Evaluation

The experiments of RQ2 determine whether the 2D image scenarios and the 3D RS scenes are the same geographical entity.

Table 4. The similarity of the overall shapes of A and B

Comparison\ Group	Slice 1	Slice 2	Slice 3	Slice 4	Average
Perceptual hash similarity	0.8594	0.9063	0.7500	0.8750	0.8477

Table 5. The similarity of the precise features of A and B

Group	Slice	Cosine sim	Jaccard sim	Perceptual hash	Group average
G1	Mountain1	0.8305	0.7576	0.9863	0.8729
	Water1	0.8917	0.7667	0.9492	
	Both1	0.8292	0.9289	0.9160	
G2	Mountain2	0.8190	0.8723	0.9971	0.9013
	Water2	0.9609	0.7930	0.9727	
	Both2	0.8627	0.8739	0.9600	
G3	Mountain3	0.7153	0.8512	0.9775	0.8597
	Water3	0.9280	0.7393	0.9287	
	Both3	0.7536	0.9468	0.8965	
G4	Mountain4	0.8283	0.9338	0.9961	0.9063
	Water4	0.9570	0.7513	0.9629	
	Both4	0.8492	0.9367	0.9414	

It can be seen in Table 4 that the overall shape similarities between the two types of samples are all good. Despite the art processing and painting errors in the ancient paintings, the similarity still reaches between 0.85 and 0.9. The lower similarity of the third group is due to the RS image of Class B, which has a farther perspective and cannot accurately restore the perspective of the author's mountain data collection in the painting.

Table 5 shows the similarities between mountain coordinates, hydrological coordinates, and mountain-water superimposed coordinates by using the superposition protocol of the two types of samples. Data shows that the mapping between 2D images and 3D images can be completed under the double tests of overall shape coordinates and precise gesture coordinates.

Answer to RQ2: Good matching results of 2D and 3D images show that the 2.5D feature annotation method can effectively confirm the 3D geographic location in the scenario space of 2D images.

5.3 RQ3: Convergence Assessment

The experiments of RQ3 evaluate the mapping efficacy calculated by different models through the CRITIC method and provide a mapping efficacy score for each type of samples (Table 6).

Table 6. Weight results of each model

Model	Cosine sim.	Jaccard sim.	Perceptual Hash	RGB sim.
Weight	0.2250	0.3363	0.2351	0.2036

Table 7. Mapping performance score

Group	Media	Score	Group	Media	Score
M5	Outline	94.9745	W3	Outline	79.7215
R2	Outline	92.3737	W6	Outline	76.0738
M4	Outline	92.1698	M3	Outline	74.9550
W2	Outline	89.0940	R1	Outline	74.6676
M2	Outline	84.3853	W5	Outline	74.6479
W4	Outline	84.0292	A3	Outline	68.1869
M7	Outline	82.6007	M6	Nodes	65.4924
A2	Outline	82.1556	O3	Image	51.8571
A4	Outline	80.8850	O1	Image	50.2066
A1	Outline	80.5684	M8	Nodes	48.2964
W1	Outline	80.5290	O2	Image	44.2500
M1	Outline	80.2896	O4	Image	43.3439

Table 7 reports the comprehensive mapping performance of selected samples. We find that those mapping with key feature contours as the medium is the best, and the key feature point mapping effect is the second best. O1 (50.2066), O2 (44.2500), and O4 (43.3439) are the original mapping without any key features, and we can find that their performance is not acceptable. By comparison, it shows that the 2.5D key features method constructed in this paper can significantly improve the matching, association, and sharing effectiveness of data.

Answer to RQ3: The model fusion mechanism can give a comprehensive evaluation of each matching method, and the matching efficacy of feature contours is better than the matching of feature points, and far better than the original images matching.

5.4 RQ4: Scalability Validation

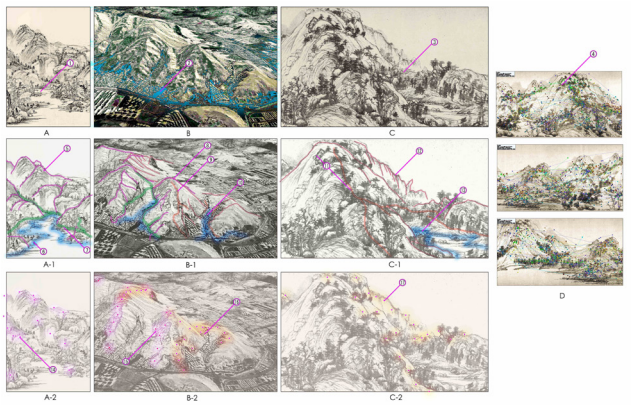


Figure 5. Graphical structural stitching findings based on comparative evidence

In Figure 5, Column B includes the RS images of the Fuyang district, China. Column A is a famous Chinese painting "Scenery of JiuZhu Peaks", and Column C is a slice from "Dwelling in the Fuchun Mountains". The two paintings are physically independent, but it is suspected that the scenarios in their images were continuous in the real world.

Table 8. Matching results

Samples	Cosine sim	Jaccard sim	Perceptual hash
Fuyang Mountain	0.9340	0.8828	0.9727
Fuyang Points	0.8814	0.1911	0.9973
Fuyang Road	0.9271	0.5215	0.9473
Fuyang River	0.8071	0.9695	0.7832
Jiuzhu Mountain	0.7876	0.9457	0.9424
Jiuzhu Points	0.7396	0.1756	0.9980
Jiuzhu Road	0.8766	0.9806	0.9219
Jiuzhu River	0.7996	0.9924	0.7598

Table 8 shows that the mountains, roads, and hydrological features depicted in Column C are highly matched with a certain Peak 1 in Fuyang. Meanwhile, those features depicted in Column A are highly matched with a certain Peak 2 in Fuyang. Since Peak 1 and Peak 2 have continuity between mountain entities in 3D geographic space, the scenery drawn by the two maps has a very high possibility of stitching (Figure 5).

Answer to RQ4: The method in this paper can not only solve the geolocation problem of single landscape paintings but also deal with the geographic classification problem of multiple landscape paintings and enhance the semantic association between images.

6 Limitation

This study may have limitations. First, the experimental results may be influenced by manual annotation from domain experts. To reduce this influence, a complete Image Scenario Space Interpretation process is proposed as a guide, which is strictly followed throughout the experimental period. Second, the cases may be typical. To eliminate the stereotype of semantic annotation caused by the notoriety of the case images, we cut each image into several slices until the participants could not recognize the image by local details, rather than based on the annotation and matching of the whole image. This helps to increase the generalizability of the method.

7 Conclusion

This paper proposed a behaviorally evidence-based spatial similarity evaluation method to apply a complex network linguistic labeling classification model involved in image scenario space to the automatic filtering of DEM data, assembles multi-class homogeneous target interpretation methods and verifies the effectiveness of research path reuse and conclusion reuse after taking their best balance by using a weighted model fusion mechanism.

This paper breaks the semantic gap of image learning from the perspective of a single discipline, connects traditional Chinese landscape painting with geographic space, expands the semantic capacity of such images, improves the efficiency of image learning, enables image scenario analysis of cross-domain knowledge, and improves the efficiency of information exchange based on multi-domain knowledge.

Facing the fierce competition in image interpretation, valuing effective feedback from researchers' innovative experiments, and maintaining alertness to the existing pool of techniques are key to identifying the evolution of research paths. Both feedback from research and the differences among similar studies are important sources to secure the need for research evolution, where the following issues remain in urgent need of resolution:

(1) The algorithm models of different targets cannot be directly coupled. Due to a large amount of noise in the data set, manual screening of useful information and extraction of evolutionary requirements is a heavy workload and low efficiency. Therefore, how to automatically identify effective data forms and extract evolutionary needs is an urgent problem to be solved.

(2) The requirement of image interpretation is usually related to specific scenes, while the description of scenes in the DEM data model is often incomplete or even missing, which makes it difficult for RS experts to accurately interpret the real intention of the image recording process by using a single data resource. Therefore, how to aggregate scene information to better understand the real intention (implied semantics) in the image scene space is another problem worth continuous attention.

Acknowledgements

Great thanks for the suggestions from Prof. Deren Li, Prof. Jiansong Li, Prof. Kun Qin, and Prof. Xudong Lai. We also want to give thanks to the Key Laboratory of Semantic Publishing and Knowledge Services and the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing. This paper is supported by the National Social Science Foundation of China (Grant No. 21 & ZD332).

References

[1] L. Jin, F. Tan, S. Jiang, Generative Adversarial Network Technologies and Applications in Computer Vision, *Computational Intelligence and Neuroscience*, Vol. 2020, pp. 1-17, August, 2020.

[2] C. Li, Z. Zhang, W. Yang, Salient object detection method by combining Boolean map and grayscale rarity, *Journal of Image and Graphics*, Vol. 25, No. 2, pp. 267-281, February, 2020.

[3] J. Xu, X. Wen, H. Zhang, D. Luo, J. Li, L. Xu, M. Yu, Automatic extraction of lineaments based on wavelet edge detection and aided tracking by hillshade, *Advances in Space Research*, Vol. 65, No. 1, pp. 506-517, January, 2020.

[4] H. Gao, D. Yao, M. Wang, C. Li, H. Liu, Z. Hua, J. Wang, A Hyperspectral Image Classification Method Based on Multi-Discriminator Generative Adversarial Networks, *Sensors*, Vol. 19, No. 15, Article No. 3269, August, 2019.

[5] L. Zhou, Q.-F. Wang, K. Huang, C.-H. Lo, An Interactive and Generative Approach for Chinese Shanshui Painting Document, *2019 International*

Conference on Document Analysis and Recognition (ICDAR), Sydney, Australia, 2019, pp. 819-824.
[6] R. Wang, H. Huang, A. Zheng, R. He, Attentional Wavelet Network for Traditional Chinese Painting Transfer, *2020 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy, 2021, pp. 3077-3083.
[7] D. Li, W. Yu, Z. Shao, Smart city based on digital twins, *Computational Urban Science*, Vol. 1, Article No. 4, March, 2021.
[8] J. Gong, W. Huang, Z. Chen, Y. Liu, L. Li, W. Tang, Q. Zhang, J. Chen, B. Chen, P. Yue, J. Liu, J. Xiao, Global Location Information Superposition Protocol and Location-based Service Network Technology: Progress and Prospects, *Journal of Geo-information Science*, Vol. 24, No. 1, pp. 2-16, January, 2022.
[9] K. Qin, X. Yu, Y. Zhou, K. Zhang, D. Liu, Q. Wang, J. Tao, R. Xiao, B. Lu, G. Xu, Y. Yu, Q. Meng, Networked Mining and Association Analysis of Geographical Multiple Flows at a Global Scale, *Journal of Geo-information Science*, Vol. 24, No. 10, pp. 1911-1924, October, 2022.
[10] W. Shi, K. Qin, J. Chen, P. Zhang, Y. Yu, X. Zhang, Q. Hu, C. Huang, C. Yu, W. Zhang, X. Tang, Q. Meng, L. Meng, J. Li, Key theories and technologies on reliable dynamic monitoring for national geographical state, *Chinese Science Bulletin*, Vol. 57, No. 24, pp. 2239-2248, August, 2012.
[11] Y. Huang, X. Lai, E. Zhang, A Method of Power Line Extraction Based on Point Cloud Feature Selection, *Journal of Geomatics*, pp. 1-5, April, 2022, DOI: 10.14188/j.2095-6045.2021885.
[12] R. Dai, Zero-Shot Image Classification Algorithm Based on SIF Fusion Semantic Tags, *Automatic Control and Computer Sciences*, Vol. 56, No. 4, pp. 364-373, August, 2022.
[13] A. Wang, Y. Wang, X. Song, Y. Iwahori, Remote Sensing Image Super-Resolution Reconstruction based on Generative Adversarial Network, *International Journal of Performability Engineering*, Vol. 15, No. 7, pp. 1783-1791, July, 2019.
[14] J. G. Thanikkal, A. K. Dubey, M. T. Thomas, Deep Learning based Aquatic and Semi Aquatic Plants Morphological Features Extraction and Classification, *International Journal of Performability Engineering*, Vol. 18, No. 10, pp. 702-709, October, 2022.
[15] D. Diakoulaki, G. Mavrotas, L. Papayannakis, Determining objective weights in multiple criteria problems: The critic method, *Computers & Operations Research*, Vol. 22, No. 7, pp. 763-770, August, 1995.
[16] B. Kim, J. Oh, C. Min, Investigation on Applicability and Limitation of Cosine Similarity-Based Structural Condition Monitoring for Gageocho Offshore Structure, *Sensors*, Vol. 22, No. 2, Article No. 663, January, 2022.
[17] X. Tian, G. Zhou, M. Xu, Image copy-move forgery detection algorithm based on ORB and novel similarity metric, *The Institution of Engineering and Technology Image Processing*, Vol. 14, No. 10, pp. 2092-2100, August, 2020.
[18] X. Zhang, H. Yan, L. Zhang, A Perceptual Hash Algorithm for DEM Data Authentication and Tamper Localization, *Journal of Geo-information Science*, Vol. 22, No. 3, pp. 379-388, March, 2020.

Biographies

Ziyang Weng is an associate professor at School of Information Management, Wuhan University. His research interests include spatial humanity and historical geographic remote sensing.

Shuhao Wang was born in Shenyang City, China. She focuses on digital humanities and image interpretation.

Ziyu Zhang was born in Changsha City, China. She focuses on digital humanities.

Renyi Liu is a professor at the School of Earth Sciences, Zhejiang University, China. He is an expert in the development and application of spatial-temporal big data related to ocean, land, and mapping.