

<https://doi.org/10.1038/s40494-025-01859-9>

An open benchmark for oracle bone rubbing image retrieval



Bang Li^{1,2}, Zengmao Ding², Yan Zhang³, Han Zhang^{1,2}, Jing Yang², Jiaqi Fang², Taisong Jin³✉, Yongge Liu^{1,2}, Zhenhao Song⁴ & Rongrong Ji³

Oracle bone inscriptions provide critical insights into ancient Chinese history. However, the retrieval and analysis of inscription rubbings remain challenging due to fragmentation, weathering, and non-standardized character forms. These challenges fundamentally limit the applicability of conventional image retrieval methods, an issue exacerbated by the lack of large-scale annotated datasets. To tackle these challenges, we introduce the first dataset and a Multi-step Strategy for Homologous Rubbing Retrieval (MSHRR). MSHRR employs a three-stage pipeline integrating character extraction, cross-rubbing matching, and similarity scoring, bypassing Optical Character Recognition (OCR) dependencies. This novel framework outperforms state-of-the-art methods in handling glyph structures through its morphology-aware paradigm. More importantly, MSHRR has found 276 new homologous sets, accounting for over 10% of documented cases in twenty years. Our benchmark also offers a reproducible evaluation framework for computational archeology and reveals new historical connections.

Oracle Bone Inscriptions (OBIs), dating back over 3600 years to the Shang dynasty, constitute one of the most invaluable cultural and historical relics in Chinese history. As the earliest known systematic form of Chinese writing, they offer invaluable insights into the political, social, religious, and linguistic landscapes of ancient China. These inscriptions not only document royal divinations and significant historical events but also serve as a cornerstone for understanding the evolution of Chinese characters and writing systems. The study of OBIs has been indispensable for historians and archeologists in reconstructing the ancient Chinese civilization, unveiling profound facets of its language, culture, and governance structures.

Despite their immense historical significance, there remains a pivotal challenge in the realm of OBI research: the identification and retrieval of **homologous rubbings**—rubbings derived from identical oracle bones. The capacity to compare these rubbings is indispensable for tracing the provenance and reconstructing the original context of the inscriptions. Historically, organizing such data has been an arduous task, often spanning several years and heavily reliant on manual image retrieval and comparison. A notable milestone in this endeavor was the compilation of the Collection of Oracle Bone Inscriptions (COBI)¹, which took an extraordinary 26 years to complete and stands as the most comprehensive archive of oracle bone rubbings to date. Despite the extensive 26-year compilation effort behind COBI, the assembly of homologous rubbings remains incomplete². Identifying homologous rubbings poses a formidable challenge due to the visual

disparities among them. These disparities, which may arise from substantial contour variations, varying degrees of completeness of the oracle bones, and differing rubbing techniques, significantly complicate the efficacy of image matching algorithms, as exemplified in Fig. 1.

For the homologous rubbing retrieval task, the textual content that is identical in imagery serves as the primary basis for determining their consanguinity. In recent years, Artificial Intelligence (AI) technology based on character recognition has played a significant role in the organization of ancient documents^{3–5}. However, this character recognition-based approach is not completely applicable to the collation of OBIs, as character recognition necessitates the annotation of characters, and the characters in OBI are only partially deciphered^{6–8}. The incompleteness of their decipherment renders character-level annotation impractical, as illustrated in Fig. 2. Due to this incompleteness, compounded by the lack of annotated datasets, the field faces significant challenges. Particularly when developing AI models, although character recognition has traditionally been a core component of text-based image retrieval tasks, AI models in the homologous rubbing retrieval task must effectively address the complexities of OBIs without relying on character recognition.

The evolution of AI-assisted OBI research has been propelled by specialized datasets, yet critical gaps remain. In 2015, Guo et al. proposed an OBI recognition model based on the dataset Oracle-20K, which, despite its comprehensiveness, was limited in scope due to the absence of a publicly

¹School of Computer & Information Engineering, Anyang Normal University, Anyang, Henan, China. ²Key Laboratory of Oracle Bone Inscriptions Information Processing, Ministry of Education of China, Anyang Normal University, No. 436 Xianghe Road, Anyang, 455000 Henan, China. ³Key Laboratory of Multimedia Trusted Perception and Efficient Computing, Ministry of Education of China, Xiamen University, No. 422 Siming South Road, Xiamen, 361005 Fujian, China.

⁴Institute of History, Chinese Academy of Social Sciences, Beijing, China. ✉e-mail: jintaisong@xmu.edu.cn

Fig. 1 | Visual Dissimilarity Among Homologous Rubbings. The left panel illustrates the matching feature points identified by our method, while the right panel displays the feature points extracted by SuperGlue²⁵. The cases result in visual dissimilarity: (a) Significant contour differences due to fragment assembly. (b) Inconsistencies in the completeness of the oracle bone pieces between two rubbing sessions. (c) Notable visual disparities in the images resulting from different rubbing techniques.

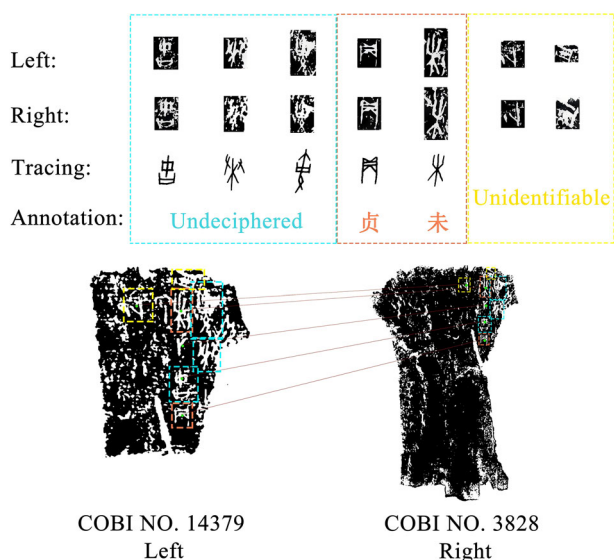
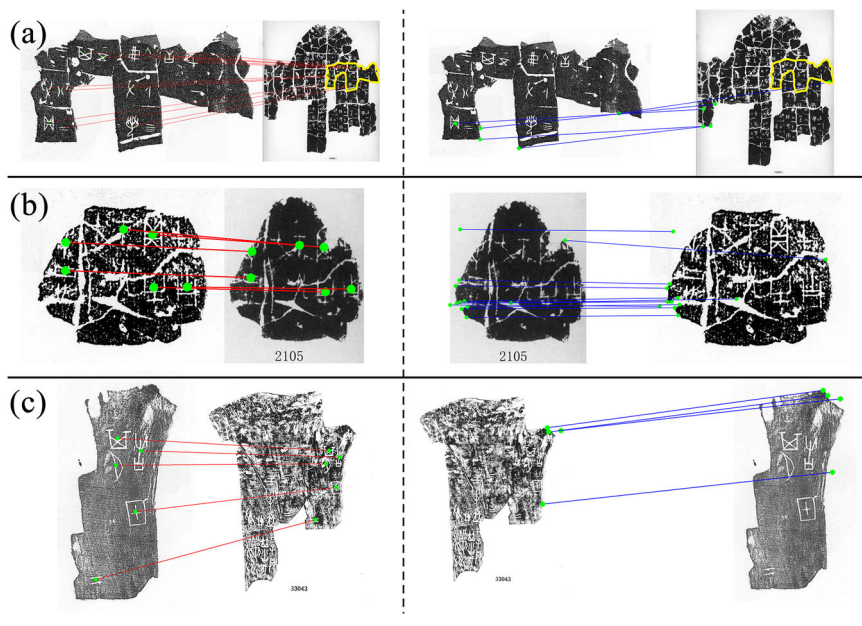


Fig. 2 | A comparison of homologous oracle bone rubbings from different periods, showcasing the challenges in recognition. The enlarged characters (top) exhibit variations due to distinct rubbing techniques. The blue areas denote three undeciphered characters lacking modern Chinese annotations. The orange areas indicate two deciphered characters with known modern equivalents. The yellow areas highlight two characters that are severely worn, making them indistinguishable and a matter of scholarly debate.

accessible dataset⁹. This constraint hindered further research until the introduction of the OBC306 dataset by Huang et al. in 2019, boasting 309,551 images across 306 classes and sparking a renewed interest in OBI recognition¹⁰. The OBC306 dataset's extensive collection and categorization provided a fertile ground for numerous subsequent studies^{11–13}. Subsequently, a succession of OBI datasets emerged, including the HWOBC, which features 3,881 classes of handwritten OBI characters¹⁴, and the Oracle-MNIST dataset, comprising 28×28 grayscale images of 30,222 ancient characters from 10 categories¹⁵. A significant contribution to the field was made by Liu et al. with the HUST-OBC dataset, encompassing 77,064 images of 1588 deciphered characters and 62,989 images of 9411 undeciphered characters, totaling 140,053 images¹⁶. Beyond character-level

datasets, Meng et al.¹⁷ pioneered the first unsupervised framework to organize Shirakawa's handwritten annotations, constructing a benchmark for neural network learning from unsegmented OBI documents. While these datasets have enabled advancements in OBI recognition, they usually consist of images that are cropped or denoised character images from rubbings and ignore the broader context crucial for comprehensive AI understanding. The only publicly available dataset with complete oracle bone rubbings is a detection dataset by Liu et al., which, despite including 9,200 rubbings, lacks character category annotations¹⁸.

Image retrieval of historical handwritten documents is a highly remarked issue, serving as a valuable tool to assist historical researchers in analyzing the content of images of handwritten documents. In ICDAR 2019 Competition on Image Retrieval for Historical Handwritten Documents¹⁹. Lai et al. proposed a novel method by encoding Pathlet and SIFT for feature extraction and clinched the best result²⁰. Compared to page retrieval, a more challenging task in the ICFHR 2020 competition is to retrieve document fragments from the same page. The winner of this competition uses a method based on training two different residual networks in terms of style and writer identification²¹. Peer et al. present a novel neural network architecture that combines a residual backbone with a feature mixing stage to improve retrieval performance²². Chammas et al. proposed a deep learning system with a feature descriptor that has shown high performance on two Latin historical datasets ICDAR 2019 and ICFHR 2020, and one Arabic non-historical KHATT dataset²³. Although the collation of homologous rubbings falls under the task of historical document retrieval, it differs from the aforementioned task domain. The retrieval target in this case is different versions of the same document rather than different fragments of the same document. Furthermore, the number of different versions of the same oracle bone fragment is not abundant, making it unsuitable for training deep models with supervised learning to retrieve images with similar features. Instead, it is more suitable for serving as a validation set for unsupervised learning to evaluate the performance of algorithms.

Matching images that depict the same scene or object are commonly based on sparse local features, which consist of key points along with their corresponding descriptors representing the local appearance. Learning-based descriptors, such as SuperPoint²⁴, often combine point detection and the description of the interest. Deep learning-based matchers, such as SuperGlue²⁵, are deep networks trained to perform joint matching of local features and outlier rejection given a pair of input images. SuperGlue operates on key points equipped with descriptors using a sparse matching network. Recent advances like DeDoDe²⁶ propose detector-free methods

that leverage dense networks to extract local features. For complex structural matching tasks, methods combining fine-grained feature association with the Hungarian algorithm^{27–29} have shown efficacy in establishing robust correspondences under ambiguities. Homologous rubbings of oracle bone fragments inherently contain a sufficient amount of approximate local features. Wu et al. employed a deep neural network-based feature extractor called “Diviner” for homologous rubbings retrieval (<https://www.xianqin.org/blog/archives/17264.html>). However, deep feature extractors typically focus on the texture features of the rubbings. Different versions of the same oracle bone inscription are often produced by different rubbers, leading to variations in image texture due to different rubbing techniques. Additionally, since the materials of oracle bones are primarily turtle shells or animal bones, the use of the same material results in similar texture features across different rubbings. Therefore, to achieve a more efficient comparison of oracle bone fragments, models need to possess an understanding of the information conveyed by the rubbings.

Despite the paramount significance of retrieving oracle bone rubbings, historians face considerable challenges in this endeavor, often necessitating decades of expertise. These challenges are further exacerbated by the absence of specialized datasets, the visual inconsistencies among homologous rubbings, and the practical infeasibility of performing character-level annotations.

In response to these concerns, we propose a dataset specifically tailored for homologous rubbing retrieval, accompanied by a comprehensive multi-step learning model that sets a new benchmark in the field.

1. We propose a comprehensive dataset for homologous rubbing retrieval, which is expected to significantly advance research in this field and encourage broader participation in the application of AI to historical studies.
2. We present the world’s first hybrid method for content-based image retrieval of OBI rubbings, which addresses the challenges posed by visual dissimilarity and the absence of character-level annotations. The proposed method can serve as a benchmark model for homologous rubbing retrieval, achieving a recall@5000 of 88.44% and a top-10 accuracy of 90.09%.
3. The proposed method identifies 276 new retrieval results, constituting approximately 10% of the total homologous rubbings over the past two decades. Furthermore, within the newly released OBI images, we discern 3991 pairs of images that are homologous to those in COBI.

Methods

Motivation

Homologous rubbing retrieval plays a pivotal role in two distinct applications within the study of OBIs. The first application is crucial during the extensive collation of oracle bone rubbings, where it is imperative to verify whether collected rubbing images have been previously published. It is noteworthy that even in comprehensive published collections, there is a possibility of overlooked homologous rubbings. Consequently, scholars persist in the identification of these missing pairs. In 2003, Wang et al. conducted a seminal analysis of homologous rubbings within the “Supplementary Collection of Oracle Bone Inscriptions (SCOBI)”³⁰, identifying 189 instances of internal homologous rubbings within SCOBI and 733 instances of mutual homologous rubbings between the COBI and SCOBI^{31,32}.

The second application arises during the study of a specific oracle bone, where the objective is to locate different versions of its rubbing. This is essential for tracing the historical dissemination of rubbings in various catalogs. Besides, the institutional curation process requires the documentation of the provenance and circulation of oracle bone collections. This methodical examination is vital for understanding the historical journey of these artifacts. In 2020, Zhao et al. identified 27 sets of oracle bones from the National Library of China that were unmarked in COBI, and 20 sets of homologous oracle bone rubbings that were redundantly recorded in both COBI and SCOBI due to the disparate sources of rubbings³³.

The following two applications correspond to two distinct image retrieval problems: (1) *Homologous Rubbing Alignment*. This task involves aligning homologous rubbings from two distinct collections of OBIs, often representing the most labor-intensive stage in the compilation of oracle bone catalogs. (2) *Single Rubbing Retrieval*. This task focuses on retrieving homologous rubbings of a specific target within a given collection, exemplified by those in COBI and SCOBI.

Both retrieval tasks rely on determining the homology between two rubbings by meticulously analyzing their content for similarities. Despite differing in the procedure, both tasks confront a shared obstacle: to efficiently process and compare a vast number of sample pairs to pinpoint the relatively few correct matches. In the instance of SCOBI, which comprises rubbings of 13,170 OBI fragments, the task of identifying merely about 300 internal homologous rubbing pairs within such an extensive dataset results in an exceedingly low retrieval ratio, approximately 600,000:1.

The details of dataset

The homologous rubbing dataset is curated from the images in the SCOBI for the homologous rubbing retrieval task. SCOBI, which serves as an expansion to the COBI, is recognized for its broader inclusivity. SCOBI is constructed by applying less stringent criteria for the exclusion of homologous rubbings³⁴. This more inclusive approach, while beneficial for the richness of the collection, introduces challenges for historical research on the Shang dynasty. As a result, the dataset is organized to facilitate the study and retrieval of homologous rubbings. It has been made accessible on the OBI data platform ‘yin qi wen yuan’ (<https://jgw.aynu.edu.cn/home/download/index.html?sysid=19>).

To establish a dataset tailored for homologous rubbing retrieval, we draw inspiration from the internal collation task of SCOBI. SCOBI has an impressive collection of rubbings derived from 13,170 oracle bone fragments, including separate rubbings from both sides of certain bones of a total of 14,834 images. Although Wang et al. analyzed the internal collation of homologous rubbings within the ‘Supplementary Collection’ and identified 189 instances^{31,32}, their efforts did not exhaustively uncover all the rubbings. Thus, we integrate information on homologous rubbings discovered by scholars over the past two decades. These sources encompass supplementary tables in OBI reference books^{35,36}, records from oracle bone organization reports^{33,34}, and authoritative website information (<https://www.xianqin.org/blog/archives/17264.html>). Despite undergoing multiple rounds of collation, the portion of the data does not comprehensively cover all pairs of homologous rubbings.

For the real-world retrieval tasks, it is challenging to explore all potential instances within the dataset. Consequently, we develop an unsupervised algorithm specifically designed for retrieval tasks. Utilizing this algorithm, we have identified 25 previously unrecorded pairs of homologous rubbings. Following verification by historians, these pairs have been incorporated into the dataset as additional ground truth, thereby enriching the existing body of knowledge.

The proposed dataset encompasses 14,834 rubbing images within SCOBI. Among these images, the authentic homologous rubbing pairs consist of 303 instances, meticulously curated from a multitude of resources, encompassing monographs, scholarly articles, online databases, and the foremost predictions generated by our methodology. Note that an additional 28 homologous pairs were excluded from the ground truth. The reason is the insufficiency of image data to verify their homology. These excluded pairs can be classified into three distinct categories. First, there are 11 instances where rubbings from spliced oracle bones were considered homologous to two separate fragment rubbings from the same oracle bone, but they lacked clear-cut visual cues indicating homology. Second, 16 cases involved non-rubbing artifacts that partially obscured or overlapped with the rubbing images. Finally, in a single case, two rubbings represented the front and back of the same oracle bone fragment, yet they failed to meet the inclusion criteria because of the lack of conclusive visual homology evidence as mentioned above.

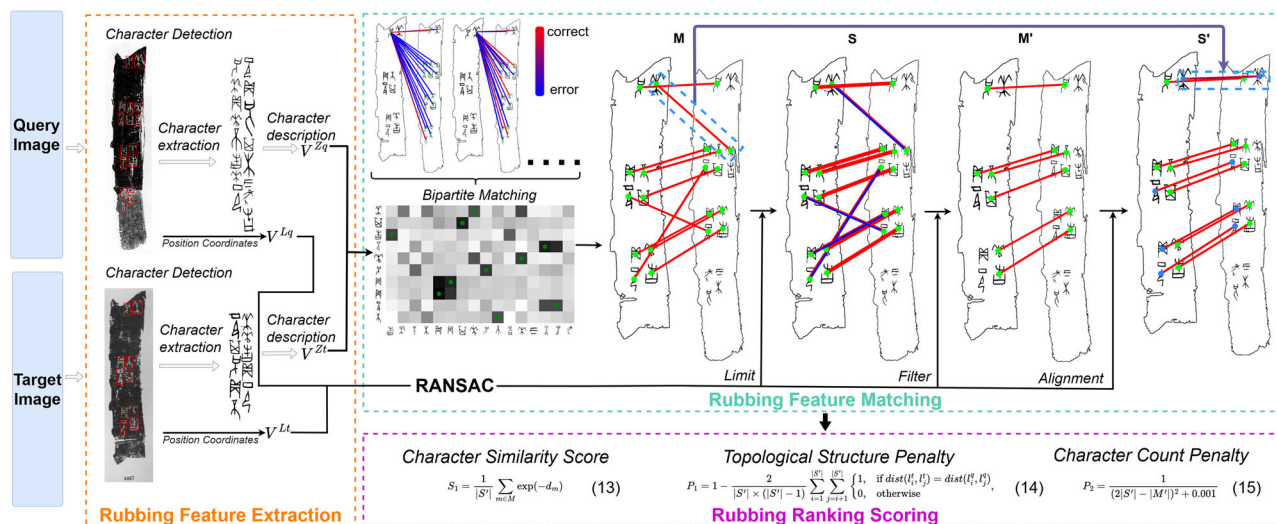


Fig. 3 | An illustration of the three steps in MSHRR. Rubbing feature extraction is performed using a cascaded network designed to extract character features from rubbings; Rubbing feature matching compares morphological and spatial features to

find analogous glyphs, bypassing OCR; Rubbing rank scoring ranks rubbings by glyph similarity, structure, and matched pairs to prioritize identical content.

In our dataset, the numerical identifiers embedded within the file names of the rubbing images correspond precisely to the SCOBI identification numbers assigned to each respective oracle bone. The absence of a letter suffix appended to the number in the filename signifies that a single rubbing is available for that particular oracle bone. Conversely, the inclusion of the letters 'f', 'b', or 's' as suffixes within the filename denotes rubbings derived from the front, back, and bone socket of the oracle bone, respectively. Additionally, sequential letters like 'i', 'j', 'k', and so forth within a filename signify that, although these rubbings originate from the same oracle bone, they represent discrete, non-contiguous sections and cannot be seamlessly assembled together. This structured naming convention greatly enhances the organization and traceability of the rubbing images within our dataset.

Problem formulation

The problem formulation is predicated on the fundamental assumption that the features of rubbings from homologous sources follow a Gaussian distribution around a common mean. We consider each rubbing $x_i = \mu_x + \epsilon_x$, where ϵ_x is distributed according to a Gaussian distribution:

$$\epsilon_x \sim \mathcal{N}(0, \Sigma), \quad (1)$$

where μ_x and Σ_x represent the mean and covariance of the feature distribution, respectively. Thus, the expectation of the loss of homologous rubbings is formulated as:

$$E\|x_i - x_j\|_2^2 = E\|\epsilon_x^i - \epsilon_x^j\|_2^2. \quad (2)$$

The expectation for the loss of non-homologous rubbings is:

$$\begin{aligned} E\|x_i - y_k\|_2^2 &= E\|(\mu_x - \mu_y) + (\epsilon_x^i - \epsilon_y^k)\|_2^2 \\ &= E\|(\mu_x - \mu_y)\|_2^2 + 2(\mu_x - \mu_y)^T(\epsilon_x^i - \epsilon_y^k) + \|\epsilon_x^i - \epsilon_y^k\|_2^2. \end{aligned} \quad (3)$$

From Eq. (1), it is evident that the expectation of ϵ_x^i and ϵ_y^k are both 0, when μ_x and μ_y are both constant vector. Thus, Eq. (3) can be simplified as:

$$E\|x_i - y_k\|_2^2 = E\|(\mu_x - \mu_y)\|_2^2 + \|\epsilon_x^i - \epsilon_y^k\|_2^2. \quad (4)$$

For non-homologous rubbings, which are quite similar and easily misaligned, the Gaussian distribution is similar, meaning $\Sigma_x \approx \Sigma_y$. Thus, $E\|\epsilon_x^i - \epsilon_y^k\|_2^2 \approx E\|\epsilon_x^i - \epsilon_x^j\|_2^2$, and

$$E\|x_i - y_k\|_2^2 = E\|(\mu_x - \mu_y)\|_2^2 + E\|\epsilon_x^i - \epsilon_y^k\|_2^2 > E\|\epsilon_x^i - \epsilon_x^j\|_2^2. \quad (5)$$

Under ideal conditions, homologous rubbings and non-homologous rubbings can be easily distinguished.

Due to wear and tear on oracle bone rubbings caused by various factors, the disparity between homologous rubbings typically exceeds that of non-homologous rubbings, rendering the establishment of Eq. 5's conclusion challenging. To rectify the error, we propose a typesetting mechanism, e.g., *Topological Structure Penalty*. For the p th character of the i th rubbing, its feature is x_i^p and its location is l_i^p . Thus, the distance between homologous rubbing characters can be formulated as

$$E_{p,q}(\min_{m,n}(\|dist(l_i^p - l_i^q) - dist(l_j^m - l_j^n)\|_2^2)) \quad (6)$$

When the rubbings are homologous,

$$\min_{m,n}(\|dist(l_i^p - l_i^q) - dist(l_j^m - l_j^n)\|_2^2) = 0. \quad (7)$$

Thus,

$$\begin{aligned} E\|x_i - y_k\|_2^2 + E_{p,q}(\min_{m,n}(\|dist(l_i^p - l_i^q) - dist(l_j^m - l_j^n)\|_2^2)) &> E\|x_i \\ - y_k\|_2^2 &> E\|x_i - x_j\|_2^2 \end{aligned} \quad (8)$$

In other words, the typesetting mechanism increases the matching lower bound of non-homologous rubbings, which makes homologous rubbings easier to be recognized.

The overview of benchmark model

The benchmark model's primary objective is to evaluate the similarity between a query image and a set of target images within an unlabeled image training dataset. As shown in Fig. 3, we propose a Multi-step based method for Homologous Rubbing Retrieval (MSHRR), which includes (1) **Extraction of Rubbing Features**, (2) **Matching of Rubbing Features**, and (3) **Rubbing Rank Scoring**. For clarification, we list the important notations in Table 1.

Table 1 | List of important notations

Notation	Description of the Notations
Z^q, Z^t	Query image, Target image
X^q, X^t	The feature matrix of each character from a query image, target image
L^q, L^t	The location matrix of each character from a query image, target image
B^s, B^t	The bipartite matching graph calculated from $\{X^q, X^t\}, \{L^q, L^t\}$
E^s, E^p	The edges in the B^s, B^p
M	Matching characters set calculated from B^s
S	inliers set limited with $\{L^t, L^q\}$
M'	Optimal matching set filter from S
$T_\theta, T_{\Delta d}$	Angular threshold calculated from E^s, E^p
S'	Optimal matching set by the alignment with $T_\theta, T_{\Delta d}$

Rubbing feature extraction is a cascaded neural network engineered to extract key features from specific regions within rubbing image, bypassing the need for OCR. This module is composed of three integral components: an object detection network, a generated adversarial network, and a contrast network. This is represented as:

$$F(z; \theta_F) = \{X, L\}, \quad (9)$$

where X denotes a set of nodes and each node encapsulates the morphological features of an individual character. L represents another set of nodes, where each node represents the positional data of a character.

Rubbing feature matching employs feature matching under the premise that homologous rubbings share the same textual content. Given the limitations of OCR in the context of Oracle Bone Inscriptions (OBI), our method focuses on evaluating the morphological resemblance and the spatial configuration of characters on the rubbings. By calculating the features of both the query image X^q, L^q and the target image X^t, L^t , the proposed method is designed to precisely identify K analogous pairs of glyphs.

Rubbing rank scoring leverages the K glyph pairs matched through Rubbing Feature Matching to assign scores and rank the retrieval results. This method integrates three vital evaluation criteria: glyph similarity, congruence in topological structure, and the influence of the number of matching characters K on the previous two factors. By integrating these elements, it is tailored to address specific and potential challenges in the retrieval process. The ultimate objective of Rubbing Rank Scoring is to accurately assign priorities to rubbing pairs with identical content. In this way, it ensures that homologous rubbings, namely those having precisely the same content, are given the utmost precedence within the retrieval sequence.

Due to practical constraints in character recognition annotation, multiple deep learning models are trained to capture the intricacies of character morphological features and their positional distribution on oracle bone rubbings. The Rubbing Feature Extraction framework comprises three independently trained deep neural networks with distinct roles: *character detection* (localizing glyphs), *character extraction* (isolating text regions), and *character description* (generating morphological vectors).

Character detection identifies OBI glyphs within rubbing images and their spatial coordinates, enabling subsequent analysis of glyph distribution. A YOLOv8-based model³⁷ is used to first localize OBI glyphs in the rubbing image. This detection process converts the input image into a set of individual character images, denoted as z_i , where i spans from 1 to N (N is the number of the detected characters). Concurrently, the model extracts the positional coordinates l_i for each glyph, thereby establishing the coordinate set L .

Character extraction processes the detected glyph images z_i to remove non-textual artifacts. Unlike modern documents, oracle bone rubbings capture both inscriptions and surface indentations caused by the carving process. Variations in rubbing techniques lead to significant visual differences in glyph

representations, as depicted in Fig. 2. To ensure the content-based retrieval focuses on textual content, we employ a U-Net-based Generative Adversarial Network³⁸ to isolate textual regions. This step results in refined glyph images \hat{z}_i^c , containing only the characters from the original rubbing images.

Character description generates feature vectors for OBI characters. Due to incomplete OCR coverage, we train a siamese network-based matching model³⁹ to evaluate the glyph similarity. This training process results in morphological descriptors for each character, transforming \hat{z}_i^c into the character morphological vector x_i , which collectively forms the set X , enabling clustering of stylistically similar OBI characters in the feature space.

Rubbing feature matching aims to maximize the identification of characters sharing identical topological structures and spatial configurations, while explicitly addressing false positives (unique character mismatches) and false negatives (repeated character omissions). This is achieved by computing the correlation between cascade features $\{X^q, L^q\}$ (including morphological and spatial features) from query image Z^q and $\{X^t, L^t\}$ from target image Z^t .

Character morphological similarity is evaluated using a bipartite matching algorithm applied to a graph represented as $B^s = (X^q, X^t, E^s)$. In this graph, E^s denotes the set of edges connecting nodes from X^q to X^t . The weights of these edges are dynamically adjusted by incorporating both feature dissimilarity $d(x_i^t, x_j^q)$ and spatial inconsistency penalties, filtering false positives caused by visually similar glyphs in inconsistent positions. By leveraging the Hungarian algorithm⁴⁰, the maximum matching within the bipartite graph G is determined, yielding a matching set $M \subseteq E$ that prioritizes geometrically consistent pairs.

To address false negatives caused by repetitive characters, the RAN-SAC algorithm⁴¹ is utilized to validate correspondences through the Homography matrix H . Specifically, we re-evaluate all character coordinates in L^q by mapping them to \hat{L}^q via H , recovering legitimate matches missed by the global optimization constraint. A refined bipartite graph $\hat{B}^p = (\hat{L}^q, L^t, E^p)$ is then constructed to identify similar glyphs and spatial distributions. In this graph, E^p consists of vectors that represent the coordinate mapping relationships. Both \hat{L}^q and L^t are positioned within a three-dimensional coordinate system, with \hat{L}^q having a z-axis coordinate of 0 and L^t set to 1. By introducing the z-axis coordinate to distinguish different character instances, spatial ambiguities caused by repeated characters are explicitly resolved. The coordinate pairs (l_i^t, l_j^q) from these sets are transformed into vectors $\langle l_i^t, l_j^q \rangle$, which form the edges E^p . Simultaneously, the set of coordinate pairs S is converted into a set of vectors S' , which serves as a reference for filtering and identifying more accurate matching pairs within E^p .

Following the homography-based coordinate augmentation, the geometric verification employs dual dynamic thresholds to concurrently mitigate false positives (incorrect matches) and negatives (valid match omissions) caused by repetitive characters. The angular threshold T_θ (Eq. (11)) suppresses false positives by enforcing angular consistency over initial candidate pairs S , generating filtered matches $M' \subseteq S$. The distance threshold $T_{\Delta d}$ (Eq. (12)) then processes rejected pairs $B^s \setminus M'$ through residual error analysis $\Delta d_{ij} = |d_{ij} - \mu_d^S|$. The final match set S' is obtained through a two-stage refinement, formalized as:

$$S' = M' \cup \left\{ \langle x_i^t, x_j^q \rangle \in B^s \setminus M' \mid \Delta d_{ij} \leq T_{\Delta d} \wedge \theta_{ij} \leq T_\theta \right\} \quad (10)$$

where $d_{ij} = d(x_i^t, x_j^q)$, μ_d^S is the global distance mean from S , and B^s denotes the initial bipartite edge set. This dual-threshold mechanism synergistically resolves repetitive character ambiguities: T_θ eliminates directionally inconsistent matches via angular constraints, while $T_{\Delta d}$ compensates for overly-strict filtering by recovering geometrically valid pairs through adaptive residual thresholds.

$$T_\theta = \max_{(l_i^t, l_j^q) \in S} \arccos \left(\frac{\langle l_i^t, l_j^q \rangle \cdot \mu_l^S}{\| \langle l_i^t, l_j^q \rangle \| \| \mu_l^S \|} \right), \quad \mu_l^S = \frac{1}{|S|} \sum_S \langle l_i^t, l_j^q \rangle \quad (11)$$

$$T_{\Delta d} = \max_{(x_i^t, x_j^q) \in B^S \setminus M'} |d(x_i^t, x_j^q) - \mu_d^S|, \quad \mu_d^S = \frac{1}{|S|} \sum_S d(x_k^t, x_m^q) \quad (12)$$

OBI rubbing retrieval is conducted under the premise that oracle bone fragments originating from the same source are likely to have partially shared spatial distributions of characters. The set S' has been compiled to identify such character pairs. However, the presence of character pairs in S' does not confirm that they are from homologous rubbings. To mitigate this uncertainty, a scoring system has been applied to S' , which assigns lower scores to pairs that appear less likely to be part of homologous rubbings.

The character similarity score is primarily designed to address bad cases where the content differs, but the characters on rubbings are visually similar. This score is derived from the average weight of the edges within the matching set S' , which is determined using the Hungarian matching algorithm. The character similarity score is mathematically defined as follows:

$$S_1 = \frac{1}{|S'|} \sum_{m \in M} \exp(-d_m), \quad (13)$$

where d_m represents the distance between the matched characters, and $|S'|$ denotes the number of elements in the matching set S' .

The topological structure penalty primarily focuses on bad cases where a common geometric distribution is formed by discontinuous oracle bone characters with high frequency. To tackle this, triangulation of the character matrices L^q and L^t is employed, leading to the generation of character relationship matrices G^q and G^t . For any pair of homologous oracle bone characters within the matched character set S' , the shortest distances within the corresponding G^q and G^t are equal. Consequently, the topological structure score is calculated as follows:

$$P_1 = 1 - \frac{2}{|S'| \times (|S'| - 1)} \sum_{i=1}^{|S'|} \sum_{j=i+1}^{|S'|} \begin{cases} 1, & \text{if } \text{dist}(l_i^t, l_j^t) = \text{dist}(l_i^q, l_j^q) \\ 0, & \text{otherwise} \end{cases}, \quad (14)$$

where l_i^q and l_j^q represent two vertices in G^q corresponding to characters in S' . l_i^t and l_j^t represent two vertices in G^t corresponding to characters in S' . The function $\text{dist}()$ denotes the shortest path between the two vertices.

The influence of character similarity and topological structure on the assessment of rubbing content similarity is significantly diminished, especially when the number of characters in the matched set S' is less than four. This highlights the reduced impact of these factors as the number of matching characters decreases. Consequently, a penalty term has been introduced to mitigate the impact of high scores in these instances, thereby enhancing the fairness and precision of the scoring system. The formula for this penalty term is formulated as

$$P_2 = \frac{1}{(2|S'| - |M'|)^2 + 0.001} \quad (15)$$

The aforementioned adjustment ensures that the scoring system remains robust across a variety of scenarios.

The primary goal of homologous rubbing retrieval is to facilitate historians in systematically organizing rubbings. In the design of the scoring mechanism, emphasis is placed on repositioning samples that do not conform to the criteria of homologous rubbings towards the end of the ranking. This consideration has been integral to the development of the scoring mechanism. The scoring formula is defined as:

$$S_{\text{total}} = S_1 - P_1 - P_2. \quad (16)$$

This formula effectively distinguishes rubbings that are likely to be homologous from those that are not, ensuring that the most relevant results are given priority in the retrieval process.

Results

Datasets

We encompass the comprehensive training data essential for the Rubbing Feature Extraction and propose a specially curated test dataset tailored for homologous rubbing retrieval. Rubbing Feature Extraction consists of a three-stage supervised learning framework, where each stage undergoes individual training via a distinct supervised learning dataset. Regarding the task of homologous rubbing retrieval, it has been meticulously scrutinized under the rigorous lens of historical research. By harnessing the supplementary tools presented in this paper, and through extensive studies spanning numerous years dedicated to organizing oracle bone materials, we compile an exhaustive annotation set for homologous rubbings.

The dataset for the detection of OBI, sourced from ref. 18, comprises 9500 high-resolution scans of OBI rubbings. Each character within this dataset is meticulously labeled with its precise upper-left and lower-right coordinate boundaries. A critical observation regarding the dataset is that it provides annotations for the locations of OBIs but does not include categorical labels for the inscriptions. Despite this limitation, the dataset is well-suited for our research objective, which is exclusively the detection of OBI. The oracle bone inscription denoising data is annotated by us. We select 4000 pairs of rubbing-facsimile images. The annotators reference the corresponding facsimile to delineate each pixel on the rubbing, resulting in 4000 pairs of pixel-level corresponding facsimile-rubbing pairs. Subsequently, using an OBI detection model, single-character samples from the rubbings and facsimile characters corresponding to the pixels are obtained including a total of 20,000 pairs.

The model utilized for feature extraction of oracle bone characters employs the comprehensive OBI reference, "Compilation of Oracle Bone Inscriptions"⁴², which meticulously documents 50,050 individual facsimile images of characters, systematically classifying them into 4378 distinct categories. Leveraging this authoritative resource, the annotation team meticulously captured screenshots of each character, subsequently organizing them into a dedicated dataset tailored for metric learning-based morphological feature matching of individual characters. This meticulous approach guaranteed that characters belonging to the same category in the OBI would exhibit sufficiently close proximity in the feature space, thereby enhancing the accuracy and reliability of the feature extraction process.

We evaluate the retrieval performance of our algorithm on two tasks: homologous rubbing alignment and single rubbing retrieval. For the homologous rubbing alignment task, considering that homologous rubbing alignment poses an algorithmic complexity of $O(n^2)$ for retrieval, using the entire test set directly would involve numerous comparisons between entirely unrelated images, which can be computationally wasteful for algorithm evaluation. To ensure efficient and targeted evaluation, we carefully selected a subset of 3000 rubbings with correlated content for this task. Thus, the goal of homologous rubbing alignment is to accurately retrieve 303 homologous pairs from 9,000,000 possible matches. For the single rubbing retrieval task, we directly used all images from our SCOBI-based retrieval dataset for retrieval.

Experiment setup

Homologous rubbing alignment needs that the algorithm maximizes the presence of precise matches within these top 5000 pairs, underscoring the significance of accurate retrieval in the initial search results. To quantitatively evaluate the algorithm's efficacy, we leverage Recall@5000 as a evaluation metric. Additionally, to evaluate the model's proficiency in ranking accurate results, we utilize Mean Rank (MR) and Mean Reciprocal Rank (MRR) metrics. MR determines the average position of correct results within the ranked list, whereas MRR evaluate the quality of the highest-ranked accurate match. Since retrieval algorithms may not have strong discriminatory ability in extreme cases, calculating MR and MRR for all samples can only partially reflect the algorithm's ability to rank search results. Therefore, we compare the MR and MRR of the top 100 positive examples identified by the algorithm.

Table 2 | Comparison of homologous rubbing alignment methods on recall@5000, MR@50, MR@150, MR@250, MRR@50, MRR@150, and MRR@250

Category	Method	Recall@5000	MR@50	MR@150	MR@250	MRR@50	MRR@150	MRR@250
End-to-end	MeCoq ¹²	4.95%	4228.56	4743.52	4846.512	0.0003	0.0002	0.0002
	HiHPq ⁴³	3.63%	4514.06	4838.6	4903.61	0.0003	0.0002	0.0002
	Dino ⁴⁴	36.63%	71.78	1959.74	3176.248	0.0649	0.0222	0.0134
Feature-based	SIFT ⁴⁵	85.47%	30.30	81.56	156.76	0.0656	0.0286	0.0189
	SuperGlue ²⁵	87.12%	25.50	76.76	138.30	0.0899	0.0371	0.0241
	SGMnet ⁴⁶	83.82%	25.58	76.19	203.94	0.0899	0.0371	0.0241
	LightGlue ⁴⁷	82.83%	25.92	81.53	396.48	0.0897	0.0368	0.0229
	DeDoDe ²⁶	77.23%	25.50	75.50	525.24	0.0899	0.0372	0.0238
	MSHRR (ours)	88.44%	25.50	76.63	160.45	0.0899	0.0372	0.0239

The bold values indicate the best-performing methods for each evaluation metric, highlighting the highest scores across all compared approaches.

Table 3 | Comparison of single rubbing retrieval methods on Top-1 accuracy, Top-10 accuracy, MR, and MRR

Category	Method	Top-1	Top-10	MR	MRR
End-to-end	MeCoq ¹²	3.13%	10.23%	9.42	0.1389
	HiHPq ⁴³	2.47%	11.38%	9.37	0.1376
	Dino ⁴⁴	35.47%	53.13%	5.86	0.4522
Feature-based	SIFT ⁴⁵	81.35%	85.97%	2.35	0.8488
	SuperGlue ²⁵	84.48%	89.27%	2.05	0.8760
	SGMnet ⁴⁶	79.70%	86.46%	2.41	0.8348
	LightGlue ⁴⁷	78.38%	89.60%	2.22	0.8354
	DeDoDe ²⁶	79.37%	83.99%	2.54	0.8278
	MSHRR (ours)	84.81%	90.09%	1.96	0.8817

The bold values indicate the best-performing methods for each evaluation metric, highlighting the highest scores across all compared approaches.

Single rubbing retrieval is a common retrieval task by historians, wherein the correct retrieval of homologous rubbings is crucial for the study of oracle bone provenance. Therefore, we selected 303 oracle bone rubbings with homologous images as queries to evaluate the algorithm's performance in finding homologous rubbings among the remaining 14,834 images. To comprehensively assess the algorithm's performance, we leverage several evaluation metrics, including top-1 accuracy, top-10 accuracy, MRR, and MR, for comparative analysis.

Performance Metrics

Homologous rubbing alignment is shown in Table 2 lists the comparative results of different methods, including MeCoq¹², HiHPq⁴³, Dino⁴⁴, SIFT⁴⁵, SuperGlue²⁵, SGMnet⁴⁶, LightGlue⁴⁷, DeDoDe²⁶, and MSHRR(ours).

Upon scrutinizing the data, among End-to-end methods, Dino demonstrates a notable Recall@5000 of 36.63%, significantly outperforming MeCoq (4.95%) and HiHPq (3.63%) in the end-to-end category, yet still falling short of the feature-based methods and content-base methods. The proposed MSHRR exhibits the highest Recall@5000 at 88.44 %, showing that the text-based methods can more effectively identify the features for homologous rubbing retrieval. This higher recall rate indicates a robust capability to find the correct matches, which is essential for the applications that require comprehensive retrieval of rubbing features.

Among the feature-based methods, an analysis of the first 50 to 150 sample retrievals reveals a highly commendable performance. With a remarkably high Mean Reciprocal Rank at 50 (MR@50) of 25.5, it indicates near-perfect accuracy within the top 50 retrievals. This accomplishment underscores the robustness of these methods in promptly and accurately identifying correct matches at an early stage.

When considering the top 250 retrievals, SuperGlue stands out with the lowest MR@250. This implies its superior precision in achieving accurate matches among the highest-ranked results. The proposed MSHRR closely trails behind. It has a Mean Reciprocal Rank at 250 (MRR@250) of 0.0239, which is on a par with that of the best-performing method. The performance of MSHRR in terms of MRR@250 highlights its outstanding ability to precisely rank correct matches. This is a crucial factor in applications where the retrieval sequence holds utmost importance.

The ranking precision of the proposed MSHRR method, especially its high Recall@5000, is vital for prioritizing the most relevant rubbing features in search results. As a result, it significantly improves the efficiency and accuracy of the retrieval process. MSHRR's precision in ranking plays an instrumental role in ensuring that search results prominently showcase the most relevant rubbing features, thereby enhancing the effectiveness of the retrieval process.

Single rubbing retrieval is shown in Table 3 provides an in-depth analysis of single rubbing retrieval methods, underscoring the paramountcy of retrieval consistency with the input rubbing's content. From Table 3, we observe the following:

MSHRR distinguishes itself by achieving the highest top-1 and top-10 accuracies, underscoring its precision and reliability in retrieving homologous rubbings. With a top-1 accuracy of 84.81% and a top-10 accuracy of 90.09%, MSHRR ensures that the historical and cultural nuances of rubbings are preserved, aligning search outcomes with the intrinsic features of the artifacts.

MSHRR exhibits an impressively low Mean Reciprocal Rank (MR) of 1.96, which indicates its efficiency in elevating relevant rubbings to the top of retrieval results. Furthermore, MSHRR's MRR of 0.8817 is the highest among the evaluated methods, highlighting its unwavering accuracy in ranking homologous rubbings. This metric is crucial for reflecting the relevance of search outcomes, thereby enhancing the efficacy of the retrieval process.

MSHRR's ranking precision is instrumental in accentuating the most pertinent rubbing features, reinforcing its status as the premier choice for single rubbing retrieval tasks that necessitate stringent content consistency and accuracy.

Computational efficiency analysis is shown in Table 4 lists experimental results comparing the efficiency of various retrieval methods.

The end-to-end methods, such as MeCoq (38 seconds), HiHPq (72 seconds), and Dino (57 seconds), are capable of completing pairwise comparisons within tens of seconds. However, their recall@5000 scores are still insufficient for homologous rubbing retrieval (see Table 2).

Notably, the proposed MSHRR achieves a 165 × pairwise acceleration (0.2ms vs. 27.449ms) and 19 × total speedup (2.13h vs. 34.32h) over feature-based counterparts like SuperGlue. These significant improvements are vividly demonstrated through comparable retrieval metrics, as presented in Table 2, with the best results highlighted in bold.

Table 4 | Comparison of homologous rubbing alignment methods on time efficiency

Category	Method	Match Points	Feature Extract	Feature Match	Total Time
End-to-end	MeCoq ¹²	N/A	N/A	N/A	38s
	HiHPq ⁴³	N/A	N/A	N/A	72s
	Dino ⁴⁴	N/A	N/A	N/A	57s
Feature-based	SIFT ⁴⁵	4,927	34ms (CPU)	15.805ms (CPU)	19.77h
	SuperGlue ²⁵	850	35ms (GPU)	27.449ms (GPU)	34.32h
	LightGlue ⁴⁷	767	42ms (GPU)	18.653ms (GPU)	23.34h
	DeDoDe ²⁶	10,000	172ms (GPU)	13.921ms (GPU)	17.53h
	SGMnet ⁴⁶	998	11ms (GPU)	8.231ms (CPU)	10.29h
Content-based	MSHRR (ours)	16	2263ms (GPU)	0.200ms (CPU)	2.13h

Both the feature-based methods and the proposed MSHRR method involve feature extraction ($O(n)$ time complexity) and pairwise matching ($O(kn^2)$ complexity, where k denotes the average number of feature points per image). The feature-based methods extract an average of 767–10,000 match points per image, resulting in high $O(kn^2)$ matching costs (8–27ms/pair). *In contrast*, the proposed MSHRR's paleography-aware feature selection extracts only a mean of 16 textual features per image, increasing $O(n)$ extraction time (2263ms/img vs. 11–35ms) but drastically reducing $O(kn^2)$ matching to 0.2ms/pair. This architecture reduces dominant $O(kn^2)$ costs from SuperGlue's mean $850 \times n^2$ to $16 \times n^2$, outweighing linear $O(n)$ extraction overhead.

Content-based MSHRR demonstrates comparable effectiveness to feature-based methods while achieving considerably higher efficiency. Although MSHRR does not outperform SuperGlue in Table 3, this superficial similarity in performance masks their fundamental methodological divergence. As substantiated in Fig. 4, MSHRR's semantic-driven paradigm exhibits essential complementarity with feature-based matching: SuperGlue prioritizes local keypoint matching for ranking, whereas MSHRR leverages semantic understanding of holistic visual patterns. These opposing strategies produce complementary blind spots, with each method excelling in the exact scenarios that confound its counterpart. For instance, SuperGlue struggles with rubbings containing densely inscribed yet texture-deficient characters, where sparse discriminative features hinder reliable keypoint matching. Conversely, MSHRR underperforms when characters become semantically ambiguous due to degradation.

These observations highlight the necessity of task-specific method selection and reveal the potential synergistic advantages of hybrid approaches combining local features with global semantics. Motivated by this complementary relationship, we investigate the integration and analyse the synergistic effects of integrating MSHRR, a content-based retrieval method, with image feature-based methods such as SIFT, SuperGlue, SGMnet, LightGlue, and DeDoDe. This integration is designed to merge the textual insight of MSHRR with the visual acuity of these established algorithms, as demonstrated in Tables 5 and 6.

These results, as indicated by the “MSHRR+[Method]” notation, show a notable enhancement in performance across various metrics. Specifically, the integration leads to improved scores in Recall@5000, Top-1 Accuracy, and MRR, while optimizing MR scores. These improvements highlight the strength of the combined approach in delivering a more accurate and efficient retrieval process.

The image feature-based methods in question utilize feature extractors to pinpoint salient image characteristics. They also employ Procrustes analysis, which involves rotation and translation matrices, to determine the maximum number of inliers. This number serves as a critical metric for image matching quality. To complement this, MSHRR introduces an optimized penalty term P_{2i} , within the scoring system. This term is designed to mitigate the character count penalty, especially when dealing with fragmented oracle bone inscriptions as opposed to

complete rubbings. The revised penalty term is formulated as:

$$P_{2i} = \frac{1}{\max(2|S'| - |M'|, 0) * C_{inlier}}, \quad (17)$$

where C_{inlier} is the number of inliers.

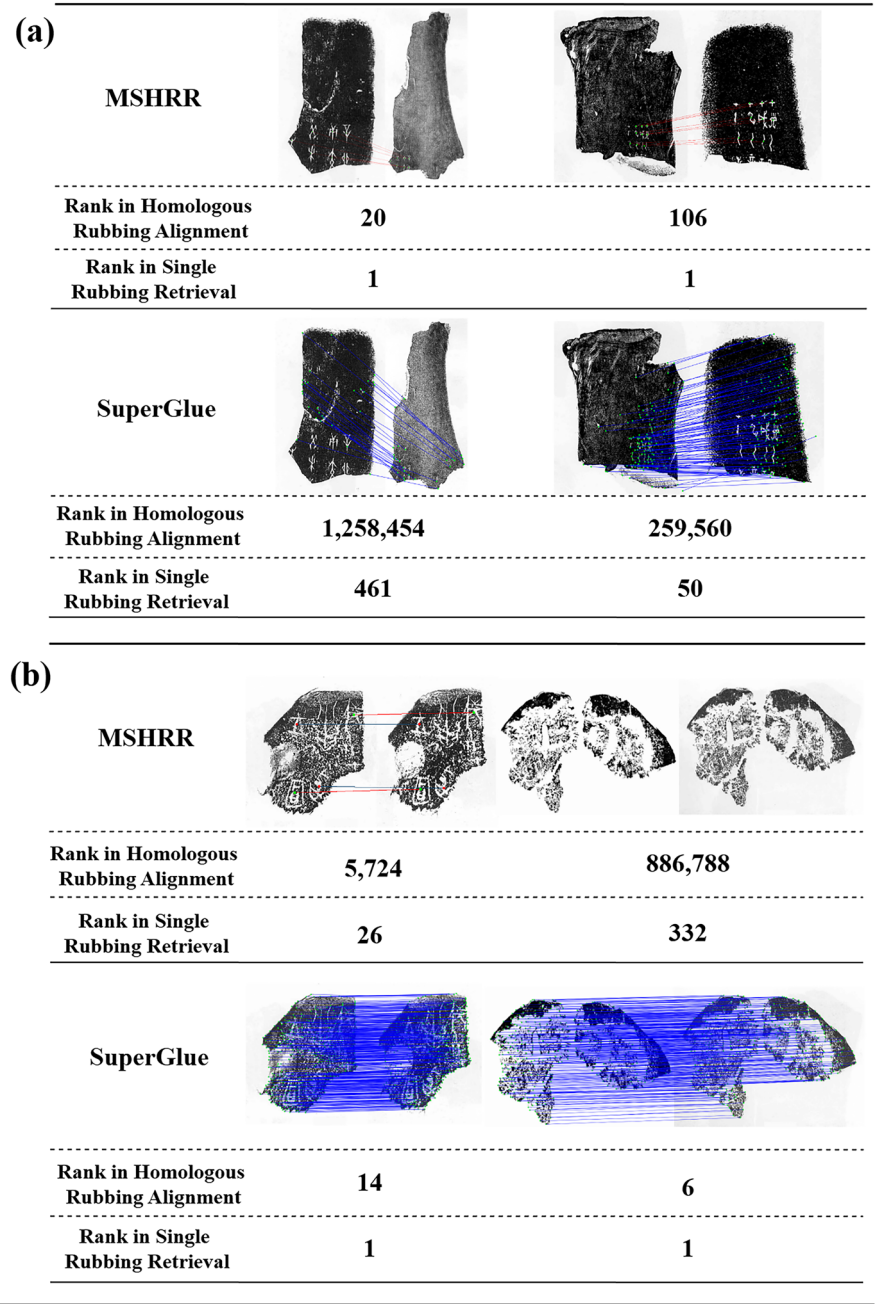
We conduct a detailed ablation study to analyze the effectiveness and contribution of various components in our proposed model as shown in Table 7. By incrementally adding the topological structure penalty P_1 and character count penalty P_2 to the similarity score S , as well as optimizing P_2 by replacing it with the image match method, denoted as P_{2i} , we observe changes in model performance and provided an in-depth discussion of the results.

When P_1 and P_2 are independently applied as penalty terms, they improve Recall@5000 by 31.68% and 52.80%, respectively. Considering the role of S in calculating the similarity of character morphological features, these results demonstrate that the penalty terms endow the model with the capability to distinguish the distribution of similar character morphological features, thereby enabling highly similar homologous rubbings to obtain higher retrieval scores. When the S score is high, it indicates the presence of a significant number of overlapping identical glyphs in the retrieval results. P_1 targets the bad cases where the content continuity of these identical glyphs is inconsistent, while P_2 addresses the bad cases where there are inconsistencies among glyphs with the same distribution.

Although P_1 and P_2 can handle different bad cases, both scenarios coexist when search results represent the same type of divination content as the query, but differ in details such as the diviner, divination time, or sacrificial objects. However, the situations they tackle are independent of each other. For instance, P_1 struggles with different types of divination content from the same period due to the presence of only a few commonly distributed characters, while P_2 encounters difficulties with repeatedly divined content because of the recurrent appearance of identically distributed characters representing the same content. Therefore, when P_1 and P_2 work together, Recall@5000 increases to 88.44%. Furthermore, when the penalty term P_2 is replaced by P_{2i} , which combines the similarity of image texture features with content similarity, in the task of Homologous Rubbing Alignment, the introduction of image texture features has led to a significant improvement of 3.96% in Recall@5000. Additionally, in the Single Rubbing Retrieval task, the Top-1 accuracy has increased by 3.14%, and the Top-10 accuracy has risen by 2.31%. This enhancement can be attributed to the fact that image texture features have boosted the scoring of different versions of rubbings from the same oracle bone fragment, assuming no changes have occurred to the fragments themselves. However, it's important to note that if the oracle bone fragments have undergone fragmentation or been recombined in different versions, the image texture features can have an adverse effect, explaining the increase of 1.55 in MR@150.

Furthermore, we conduct an ablation study to compare the retrieval performance with and without the feature extraction as shown in Table 8.

Fig. 4 | Representative examples comparing our method and SuperGlue on two complementary tasks. Homologous Rubbing Alignment (i.e., identifying the rank of correct matching pair from a large candidate set) and Single Rubbing Retrieval (i.e., the left image is treated as a query, and the rank denotes the position of its corresponding right image within the retrieved list). In our method, the red connected points represent correctly matched OBIs, while in SuperGlue, they indicate successfully matched feature points. **a** Examples where our method outperforms SuperGlue in both alignment and retrieval. **b** Examples where SuperGlue outperforms our method in both alignment and retrieval. The comparison highlights the complementary nature of the two methods.



The experimental results demonstrate that the introduction of the character extraction module significantly enhances the learning performance of the proposed model. For homologous rubbing alignment, the Recall@5000 performance increases from 35.31% to 92.40%, while the mean rank (MR@150) drops sharply from 2540.93 to 78.18. For the single rubbing retrieval, the accuracy of Top-1 and Top-10 increases by 35.7 and 13.86 percentage, respectively. These experimental results demonstrate that character extraction plays a critical role in overcoming domain differences between rubbing and transcribed copies, as feature disentanglement effectively enhances the robustness of cross-domain character representation. Note that the character description model employs distinct training datasets for rubbing characters and extracted characters. This separation is necessitated by the significant domain discrepancy between the two character types. To ensure a fair comparison as much as possible, we selected the dataset–OBC306, as it currently represents the most comprehensive publicly available collection of rubbing characters, containing 309,551 validated character samples.

Discussion

In this section, we elaborate on the significance of our proposed image retrieval task in historical research. Specifically, the retrieval outcomes facilitate the identification of diverse versions of rubbings. Furthermore, they prove to be instrumental in various data curation endeavors, including oracle bone reconstruction, restoration of incomplete characters, and sentence completion, thereby enhancing the overall efficiency and accuracy of historical research.

Utilizing our method, we have meticulously organized and verified a substantial collection of homologous rubbing pairs, totaling 5988 groups. With the assistance of our method, we organized the homologous rubbings around SCOBI. Referring to ref. 31, we searched for homologous rubbings within SCOBI and for rubbings in SCOBI and COBI that are homologous to each other. Additionally, we collected and organized all published records of homologous rubbings. The search results are shown in Fig. 5. With the aid of our method, 286 pairs of homologous rubbings were retrieved within SCOBI, among which 25 pairs had not appeared in published articles or

Table 5 | Enhanced performance of MSHRR in combination with image-based methods on homologous rubbing alignment metrics

Method	Recall@5000	MR@50	MR@150	MR@250	MRR50	MRR150	MRR250
SIFT ⁴⁵	85.47%	30.30	81.56	156.76	0.0656	0.0286	0.0189
MSHRR (ours)+SIFT	91.41%	109.00	182.42	254.86	0.0664	0.0252	0.0163
SuperGlue ²⁵	87.12%	25.50	76.76	138.30	0.0899	0.0371	0.0241
MSHRR (ours)+SuperGlue	92.40%	26.96	78.18	144.26	0.0880	0.0364	0.0236
SGMnet ⁴⁶	83.82%	25.58	76.19	203.94	0.0899	0.0371	0.0241
MSHRR (ours)+SGMnet	87.45%	25.50	76.49	156.57	0.0899	0.0372	0.0240
LightGlue ⁴⁷	82.83%	25.92	81.53	396.48	0.0897	0.0368	0.0229
MSHRR(ours)+LightGlue	91.41%	71.28	133.7	207.16	0.0717	0.0280	0.0182
DeDoDe ²⁶	77.23%	25.50	75.50	525.24	0.0899	0.0372	0.0238
MSHRR (ours)+DeDoDe	91.42%	83.06	150.25	229.48	0.0686	0.0266	0.0172

Table 6 | Synergistic effects of MSHRR integration with image-based methods on single rubbing retrieval accuracy and ranking

Method	Top-1	Top-10	MR	MRR
SIFT ⁴⁵	81.35%	85.97%	2.35	0.8488
MSHRR (ours)+SIFT	84.81%	91.25%	1.88	0.8859
SuperGlue ²⁵	84.48%	89.27%	2.05	0.8760
MSHRR(ours)+SuperGlue	87.95%	92.40%	1.73	0.9083
SGMnet ⁴⁶	79.70%	86.46%	2.41	0.8348
MSHRR (ours)+SGMnet	83.33%	87.45%	2.17	0.8654
LightGlue ⁴⁷	78.38%	89.60%	2.22	0.8354
MSHRR (ours)+LightGlue	83.00%	91.41%	1.88	0.8759
DeDoDe ²⁶	79.37%	83.99%	8.66	0.8153
MSHRR(ours)+DeDoDe	84.98%	91.42%	5.30	0.8803

Table 7 | Performance of various penalties on homologous rubbing alignment and single rubbing retrieval

Method	Homologous Rubbing Alignment		Single Rubbing Retrieval	
	Recall@5000	MR@150	Top-1	Top-10
S	33.33%	2696.28	37.62%	68.15%
S + P ₁	65.01%	712.59	52.14%	83.82%
S + P ₂	86.13%	79.16	83.33%	90.26%
S + P ₁ + P ₂	88.44%	76.63	84.81%	90.09%
S + P ₁ + P _{2l}	92.40%	78.18	87.95%	92.40%

Table 8 | The necessity of character extraction on homologous rubbing alignment and single rubbing retrieval

Method	Homologous Rubbing Alignment		Single Rubbing Retrieval	
	Recall@5000	MR@150	Top-1	Top-10
w/o Character Extraction	35.31%	2540.93	52.25%	78.54%
w/ Character Extraction	92.40%	78.18	87.95%	92.40%

books. Combining with other literature, a total of 331 pairs of homologous rubbings were organized. Meanwhile, 1711 pairs of homologous rubbings were retrieved between SCOB1 and COB1, of which 251 pairs had not appeared in any published articles. Combining with other literature, a total

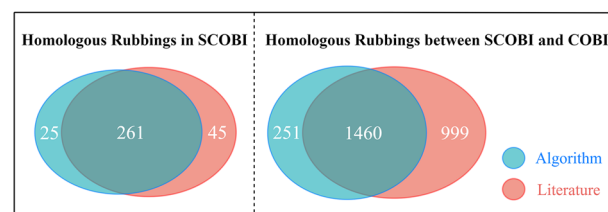


Fig. 5 | Algorithms and experts search for homology in SCOB1 and homology search between SCOB1 and COB1. a A total of 331 pairs of homologous images in SCOB1 have been found, and the data are from 286 pairs retrieved by the algorithm and 306 pairs sorted out by the literature, of which 261 pairs are jointly confirmed, 25 pairs are independently identified by the algorithm, and 45 pairs are independently sorted out by the literature. **b** A total of 2710 pairs of homologous images in SCOB1 and COB1 have been found, and the data are derived from 2459 pairs sorted out by the literature and 1711 pairs identified by the algorithm, of which 1460 pairs are jointly confirmed, 251 pairs are independently identified by the algorithm, and 999 pairs are independently sorted out by the literature.

of 2710 pairs of homologous rubbings were organized. Despite nearly 20 years of manual homologous retrieval and organization, our algorithm still retrieved nearly 10% of homologous rubbings that had not been discovered in previous efforts, fully demonstrating the important role of our algorithm in the organization of homologous oracle bone rubbings.

Moreover, we highlight a significant advancement in the study of oracle bones housed at the National Library of China. The proposed method has played a pivotal role in this discovery. Out of the 35,651 pieces of oracle bones, 8021 are cataloged in COB1, yet they lack specific corresponding records. The National Library of China has made 12,832 pieces of these oracle bones publicly accessible (<http://read.nlc.cn/specialResource/jiaguIndex>). By applying our method to this subset, our algorithm has successfully identified an impressive 3991 pairs of homologous rubbing pairs (<https://www.xianqin.org/blog/archives/21034.html>). This achievement underscores the transformative impact of our algorithm in uncovering previously unrecognized connections among these historical artifacts.

Manual homologous retrieval of oracle bone rubbings is impeded by the intricate and fragmented nature of these artifacts. Figure 6 shows examples that are frequently neglected during manual sorting but are correctly identified by our benchmark method. These include small fragments, such as those in Fig. 6a, which are reassembled fragments, or early rubbings with partial text as shown in Fig. 6b. Despite minor outline differences, certain fragments like those without visible join in Fig. 6c or the excessively blurred ones in Fig. 6d are often intuitively overlooked.

Homologous rubbing retrieval is not solely used to trace the origin of oracle bones by retrieving different versions of rubbings. It also plays a crucial role in enhancing the rubbings. By integrating fragment retrieval with oracle bone morphology analysis, incorrect assembly instances can be

Fig. 6 | Several examples were identified by historians with assistance from our method. a The retrieved oracle bone fragments have a high similarity score compared to the images stitched together with the oracle bone fragments. **b** The retrieved pairs consist of the rubbings that only capture the text and the corresponding rubbings that capture the complete fragment edges' outlines. **c** The retrieved pairs consist of rubbings and their corresponding rubbing that captures the smaller fragments that occurred after the initial rubbing was made. **d** The retrieved pairs consist of rubbings with varying levels of clarity.

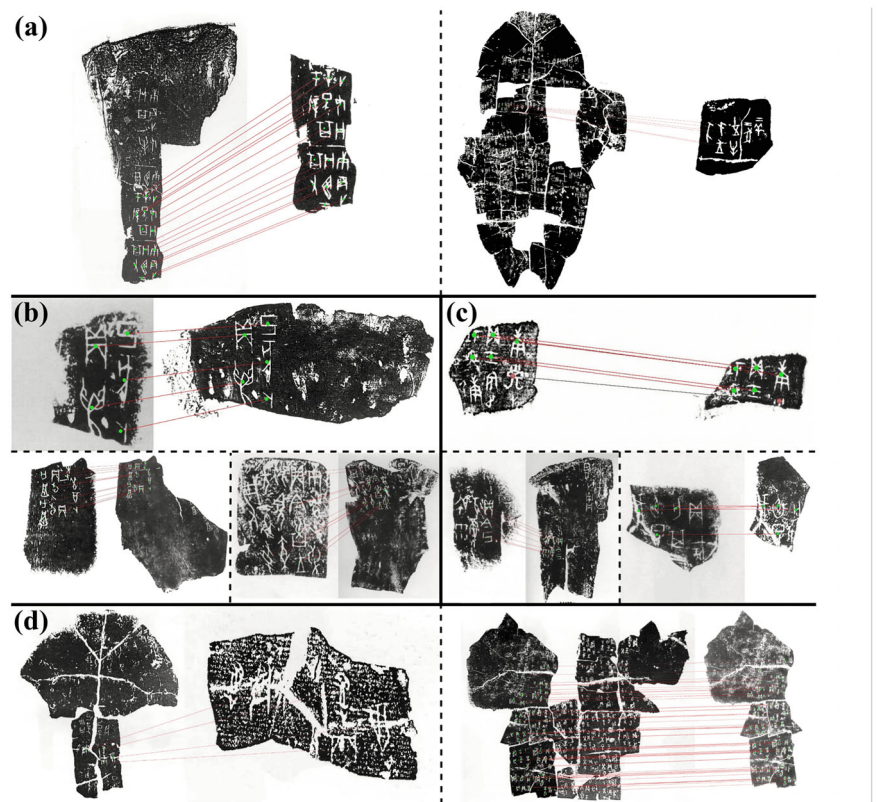
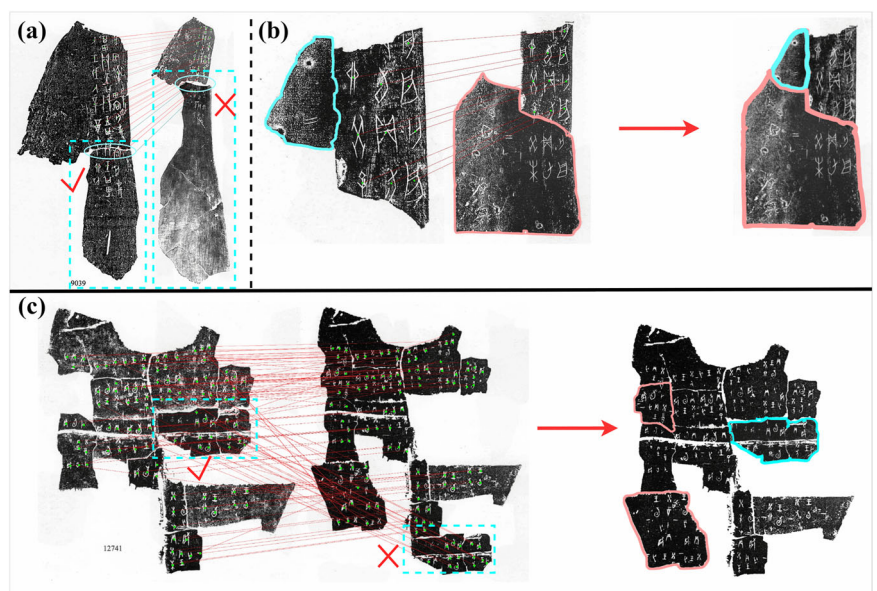


Fig. 7 | Examples of the impact of homologous rubbing retrieval in the correction of oracle bone fragment assembly. a Correction of incorrect rejoin. The blue boxes represent two different assemblies of the same fragment. The blue ellipse represents the morphological analysis of the edges of the fragments. The red hook represents the correct rejoin, and the red cross represents the wrong rejoin assembly. **b** More complete assembly. The disparate regions of the two different assembly results of the homologous rubbings are indicated by blue and orange outlines, respectively, which together assemble a more complete rubbing indicated by the red arrow. **c** Correction of incorrect splicing and more complete assembly, represented by combining the symbols from **a** and **b**.



identified. Figure 7a displays instances of incorrect assembly identification through combined retrieval. On the other hand, when a matched fragment is retrieved as a repeated rubbing in two pieced-together rubbings, a more complete oracle bone can be reconstructed. As illustrated in Fig. 7b, an oracle bone fragment can be pieced together with fragments from different directions. Moreover, as shown in Fig. 7c, incorrect fragment assembly and oracle bone fragment reconstruction may be discovered within the same pair of retrieval results.

The enhancement of rubbings is not limited to fragment assembly but also manifests in obtaining clearer glyphs or more complete edges.

Considering that variations in the handling of cracks during the rubbing process can lead to blurred inscriptions. This may result in dissimilar characters with the same distribution appearing in a homologous rubbing pair. Different versions of the rubbings may contain the clearest version of a particular character, and combining them can yield a sharper rubbing image. This is illustrated in Fig. 8a. Additionally, the assembled fragments may not represent the clearest version of the glyphs in the rubbing, and their outer contours may be incomplete, as shown in Fig. 8b. Based on the retrieval results of homologous rubbings, the information contained in the assembled rubbing images can be further enhanced.

Fig. 8 | Examples of the impact of homologous rubbing retrieval in oracle bone research. a Two homologous rubbings with clearer inscriptions were selected for comparison. The inscriptions with higher clarity were marked with blue rectangles. **b** the rubbings retrieved of fragments with clearer inscriptions, which can be used to update the assembled images.

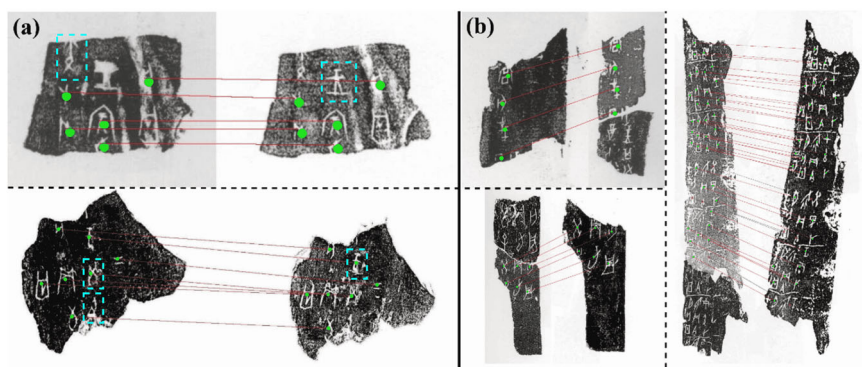
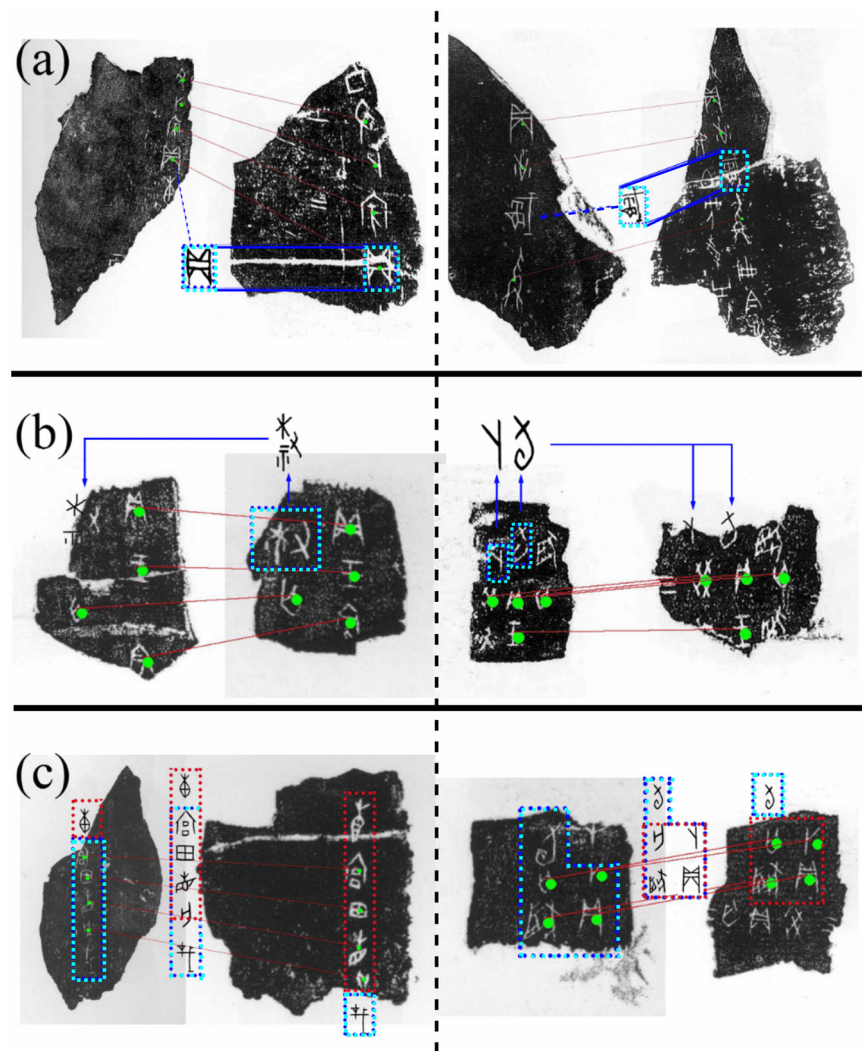


Fig. 9 | Examples of the impact of non-homologous rubbing retrieval with high content similarity. a The blue box on the black oracle bone represents the unclear glyph caused by the cracking and wear of the oracle bone, and the same glyph of the homologous rubbing pointed by the blue line completes the glyph, and the combination of the two gives the clear glyph in the blue box in the blank area. **b** The blue box represents the complete glyph, which completes the incomplete glyph on the homologous rubbing, which is pointed by the blue arrow. **c** The blue box is used to indicate the words in the homologous oracle bone fragment on the left, and the red box is used to represent the words in the homologous oracle bone fragment on the right. They have a complementary relationship, enabling the assembly of a more complete example of inscription groups.



Non-homologous fragments with high content similarity also play an important role in oracle bone research, and such images often receive high scores when conducting rubbing retrieval based on glyph relationships. Figure 9a demonstrates the role of non-homologous rubbings with high content similarity in the repair of incomplete characters. Unlike homologous rubbings, which directly provide clear character forms, images with high content similarity offer the same textual content, aiding in identifying the specific character for unclear forms. Furthermore, when partial incomplete character forms exist at the edges of fragments, non-homologous rubbings with high content similarity provide hints for

completing the missing characters, as shown in Fig. 9b. Moreover, the prediction of characters beyond the fragment's edges that are missing due to fragmentation is an important direction in oracle bone research, known as the "reconstruction of missing words." Figure 9c showcases examples of correct predictions of characters beyond the edges through the comparison of fragments with the same content. It is essential to note that "reconstruction of missing words" serves as crucial clues for fragment assembly in oracle bone inscriptions.

In this paper, we have proposed a dataset that includes 14,834 images and 303 sets of homologous rubbings specifically tailored for the oracle bone

rubbing retrieval task. Furthermore, we propose a Multi-step based Homologous Rubbing Retrieval method as a benchmark model. The proposed method achieves a recall@5000 of 88.44% for rubbing alignment and a top-10 accuracy of 90.09% for single rubbing retrieval. Specifically, the proposed method identify 276 new homologous rubbing pairs, which represent approximately 10% of the total data set. Moreover, the proposed method identifies 3991 unrecorded pairs of homologous rubbings between rubbings published by the National Library of China and rubbings in COBI. In summary, our research not only elevates the technical sophistication within this niche area but also underscores the invaluable and significant applications of our methodology in advancing historical research and archeological investigations.

Data availability

The dataset is accessible on the OBI data platform—'yin qi wen yuan' (<https://jgw.aynu.edu.cn/home/down/detail/index.html?sysid=19>).

Received: 18 February 2025; Accepted: 4 June 2025;

Published online: 24 June 2025

References

- Guo, M. *Collection of Oracle Bone Inscriptions* (Zhonghua Publishing House, Beijing, China, 1978).
- Huang, T. Collation of "collection of oracle bone inscriptions". *Archaeol. Cult. Relics* **03**, 92–93 (1995).
- Assael, Y. et al. Restoring and attributing ancient texts using deep neural networks. *Nature* **603**, 280–283 (2022).
- Yuan, J. et al. R-gnn: recurrent graph neural networks for font classification of oracle bone inscriptions. *Herit. Sci.* **12**, 30 (2024).
- Fu, X., Zhou, R., Yang, X. & Li, C. Detecting oracle bone inscriptions via pseudo-category labels. *Herit. Sci.* **12**, 107 (2024).
- Wang, M., Deng, W. & Su, S. Oracle character recognition using unsupervised discriminative consistency network. *Pattern Recognit.* **148**, 110180 (2024).
- Li, J. et al. Towards better long-tailed oracle character recognition with adversarial data augmentation. *Pattern Recognit.* **140**, 109534 (2023).
- Liu, G. et al. Gca-pvt-net: group convolutional attention and pvt dual-branch network for oracle bone drill chisel segmentation. *Herit. Sci.* **12**, 260 (2024).
- Guo, J., Wang, C., Roman-Rangel, E., Chao, H. & Rui, Y. Building hierarchical representations for oracle character and sketch recognition. *IEEE Trans. Image Process.* **25**, 104–118 (2015).
- Huang, S., Wang, H., Liu, Y., Shi, X. & Jin, L. Obc306: A large-scale oracle bone character recognition dataset. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 681–688 (IEEE, 2019).
- Liu, M., Liu, G., Liu, Y. & Jiao, Q. Oracle bone inscriptions recognition based on deep convolutional neural network. *J. Image Graph.* **8**, 114–119 (2020).
- Wang, J., Zeng, Z., Chen, B., Dai, T. & Xia, S.-T. Contrastive quantization with code memory for unsupervised image retrieval. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, 2468–2476 (2022).
- Guo, Z. et al. An improved neural network model based on Inception-v3 for oracle bone inscription character recognition. *Sci. Program.* **2022**, 7490363 (2022).
- Li, B. et al. Hwobc-a handwriting oracle bone character recognition database. *J. Phys.: Conf. Ser.* **1651**, 012050 (2020).
- Wang, M. & Deng, W. A dataset of oracle characters for benchmarking machine learning algorithms. *Sci. Data* **11**, 87 (2024).
- Wang, P. et al. An open dataset for oracle bone character recognition and decipherment. *Sci. Data* **11**, 976 (2024).
- Yue, X., Wang, Z., Ishibashi, R., Kaneko, H. & Meng, L. An unsupervised automatic organization method for professor shirakawa's hand-notated documents of oracle bone inscriptions. *International Journal on Document Analysis and Recognition (IJDAR)* 1–19 (2024).
- Liu, G., Xing, J. & Xiong, J. Spatial pyramid block for oracle bone inscription detection. In *Proceedings of the 2020 9th International Conference on Software and Computer Applications*, 133–140 (2020).
- Christlein, V., Nicolaou, A., Seuret, M., Stutzmann, D. & Maier, A. Icdar 2019 competition on image retrieval for historical handwritten documents. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 1505–1509 (IEEE, 2019).
- Lai, S., Zhu, Y. & Jin, L. Encoding pathlet and sift features with bagged VLAD for historical writer identification. *IEEE Trans. Inf. Forensics Secur.* **15**, 3553–3566 (2020).
- Seuret, M., Nicolaou, A., Stutzmann, D., Maier, A. & Christlein, V. Icfhr 2020 competition on image retrieval for historical handwritten fragments. In *2020 17th International conference on frontiers in handwriting recognition (ICFHR)*, 216–221 (IEEE, 2020).
- Peer, M. & Sablatnig, R. Feature mixing for writer retrieval and identification on papyri fragments. In *7th International Workshop on Historical Document Imaging and Processing (HIP '23)*, 31–36 (2023).
- Chammas, M., Makhoul, A., Demerjian, J. & Dannaoui, E. A deep learning based system for writer identification in handwritten Arabic historical manuscripts. *Multimed. Tools Appl.* **81**, 30769–30784 (2022).
- DeTone, D., Malisiewicz, T. & Rabinovich, A. Superpoint: Self-supervised interest point detection and description. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 224–236 (2018).
- Sarlin, P.-E., DeTone, D., Malisiewicz, T. & Rabinovich, A. Superglue: Learning feature matching with graph neural networks. In *2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4938–4947 (2020).
- Edstedt, J., Bökman, G., Wadenbäck, M. & Felsberg, M. Dedode: Detect, don't describe – describe, don't detect for local feature matching. In *2024 International Conference on 3D Vision (3DV)*, 148–157 (2023).
- Miao, B. et al. Referring human pose and mask estimation in the wild. In *Advances in Neural Information Processing Systems (NeurIPS)* (2024).
- Miao, B., Bennamoun, M., Gao, Y. & Mian, A. Spectrum-guided multi-granularity referring video object segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 920–930 (2023).
- Miao, B., Bennamoun, M., Gao, Y. & Mian, A. Region aware video object segmentation with deep motion modeling. *IEEE Transactions on Image Processing* (2024).
- Peng, B., Xie, J. & Ma, J. *Supplementary Collection of Oracle Bone Inscriptions* (Chinese Language and Literature Publishing House, Beijing, China, 1999).
- Wang, Y. & Qi, H. Collation of recorded rubbings of "supplementary collection of oracle bone inscriptions" (part 1). *J. Yin Cap. Stud.* **1**, 8 (2003).
- Wang, Y. & Qi, H. Collation of recorded rubbings of "supplementary collection of oracle bone inscriptions" (part 2). *J. Yin Cap. Stud.* **2**, 4–13 (2003).
- Zhao, A. The collation records of oracle bones in the national library, including "collection of oracle bone inscriptions" and "supplementary compilation". *Doc. Res.* **1**, 37 (2020).
- Hu, H. Highlights of collation of oracle bones in the national library and "supplementary compilation of oracle bone inscriptions". *Doc. Res.* **5**, 15 (2017).
- Cai, Z. *Continuation of Oracle Bone Inscription Compilation* (Wenjin Publishing House, Beijing, China, 2004).
- Cao, J. & Shen, J. *Collected Annotations on Oracle Bone Inscriptions* (Shanghai Lexicographical Publishing House, Shanghai, China, 2006).

37. Varghese, R. & M., S. Yolov8: A novel object detection algorithm with enhanced performance and robustness. In *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, 1–6 (2024).
38. Torbunov, D. et al. Uvcgan: Unet vision transformer cycle-consistent gan for unpaired image-to-image translation. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 702–712 (2023).
39. Jiang, W., Huang, K., Geng, J. & Deng, X. Multi-scale metric learning for few-shot learning. *IEEE Trans. Circuits Syst. Video Technol.* **31**, 1091–1102 (2021).
40. Carion, N. et al. End-to-end object detection with transformers. In *European conference on computer vision*, 213–229 (Springer, 2020).
41. Seki, A. & Pollefeys, M. Sgm-nets: Semi-global matching with neural networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6640–6649 (2017).
42. Li, Z. *Compilation of Oracle Bone Inscriptions* (Zhonghua Publishing House, Beijing, China, 2012).
43. Qiu, Z., Liu, J., Chen, Y. & King, I. Hihpq: Hierarchical hyperbolic product quantization for unsupervised image retrieval. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, 4614–4622 (2024).
44. Caron, M. et al. Emerging properties in self-supervised vision transformers. In *Proceedings of the International Conference on Computer Vision (ICCV)* (2021).
45. Ng, P. C. & Henikoff, S. Sift: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
46. Chen, H. et al. Learning to match features with seeded graph matching network. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 6281–6290 (2021).
47. Lindenberger, P., Sarlin, P.-E. & Pollefeys, M. Lightglue: Local feature matching at light speed. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 17627–17638 (2023).

Acknowledgements

This research was supported by the Natural Science Foundation of Henan Province (Grant No. 242300420680), the Paleography and Chinese Civilization Inheritance and Development Program (Grant nos. G1807, G1806, G2821), the Henan Province Science and Technology Research Project (Grant Nos. 242102210116, 252102321071), the Open Research Topic of the Key Laboratory of Oracle Information Processing, Ministry of Education (Grant no. OIP2024E002), the Key Technology Project of Henan Educational Department of China (Grant No. 22ZX010), and the Henan

Province High-Level Talents International Training Program (Grant no. GCC2025028).

Author contributions

Conceptualization B. L. and T. J.; methodology, B. L., T. J. and H. Z.; software, Z. D., and; B. L. validation, B. L. and Y. Z.; formal analysis, B. L. and J. Y.; data curation, J. F. and Z. S.; writing—original draft preparation, B. L. and Y. Z.; writing—review and editing, Y. Z. and T. J.; visualization, J. Y.; project administration, R. J.; funding acquisition, Y. L. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Taisong Jin.

Reprints and permissions information is available at

<http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025