

Rejoining fragmented ancient bamboo slips with physics-driven deep learning

Jinchi Zhu^{†1}, Zhou Zhao^{‡5}, Hailong Lei^{§2}, Xiaoguang Wang³, Jialiang Lu², Jing Li², Qianqian Tang¹, Jiachen Shen¹, Gui-Song Xia^{*4}, Bo Du^{*1,4}, and Yongchao Xu^{*1,4}

¹School of Computer Science, Wuhan University

²Center of Bamboo and Silk Manuscripts, Wuhan University

³School of Information Management, Wuhan University

⁴School of Artificial Intelligence, Wuhan University

⁵School of Computer Science, Central China Normal University

Abstract

Bamboo slips are a crucial medium for recording ancient civilizations in East Asia, and offers invaluable archaeological insights for reconstructing the Silk Road, studying material culture exchanges, and global history. However, many excavated bamboo slips have been fragmented into thousands of irregular pieces, making their rejoining a vital yet challenging step for understanding their content. Here we introduce WisePanda, a physics-driven deep learning framework designed to rejoin fragmented bamboo slips. Based on the physics of fracture and material deterioration, WisePanda automatically generates synthetic training data that captures the physical properties of bamboo fragmentations. This approach enables the training of a matching network without requiring manually paired samples, providing ranked suggestions to facilitate the rejoining process. Compared to the leading curve matching method, WisePanda increases Top-50 matching accuracy from 36% to 52% among more than one thousand candidate fragments. Archaeologists using WisePanda have experienced substantial efficiency improvements (approximately 20 times faster) when rejoining fragmented bamboo slips. This research demonstrates that incorporating physical principles into deep learning models can significantly enhance their performance, transforming how archaeologists restore and study fragmented artifacts. WisePanda provides a new paradigm for addressing data scarcity in ancient artifact restoration through physics-driven machine learning.

Main

Bamboo slips, serving as a fundamental medium for documenting ancient East Asian civilizations, contain invaluable historical records spanning philosophy, law, and social life of the period [1–5]. Their durability has enabled these artifacts to survive millennia underground while retaining legible content, offering scholars unprecedented insights into historical societies. However, the excavation of these delicate artifacts presents a critical challenge [2, 6] - many bamboo slips have been fragmented into thousands of pieces (Figure 1), significantly complicating efforts to reconstruct and interpret their content. This fragmentation creates a fundamental obstacle in accessing the wealth of historical knowledge contained in these artifacts.

The rejoining of fragmented bamboo slips represents one of the most challenging problems in archaeological preservation and cultural heritage studies [7]. The difficulty stems from multiple factors: First, the enormous number of potential fragment combinations makes manual matching extremely time-consuming - for instance, The Qin bamboo slips from the Shuihudi site comprise fragments numbering in the tens of thousands, with each piece potentially matching any of the others [8]. Second, environmental factors like moisture and pressure have caused extensive physical deterioration [9], distorting the original shapes and surfaces. Third, the sparse recorded characters often make it difficult to capture textual remnants at the fracture curves created by transverse breakage, rendering text-based matching approaches largely ineffective and necessitating reliance primarily on the morphological patterns of fracture curves for identification. The unique fiber structure of bamboo creates complex fracture patterns that traditional curve-matching approaches struggle to analyze effectively. These challenges have made

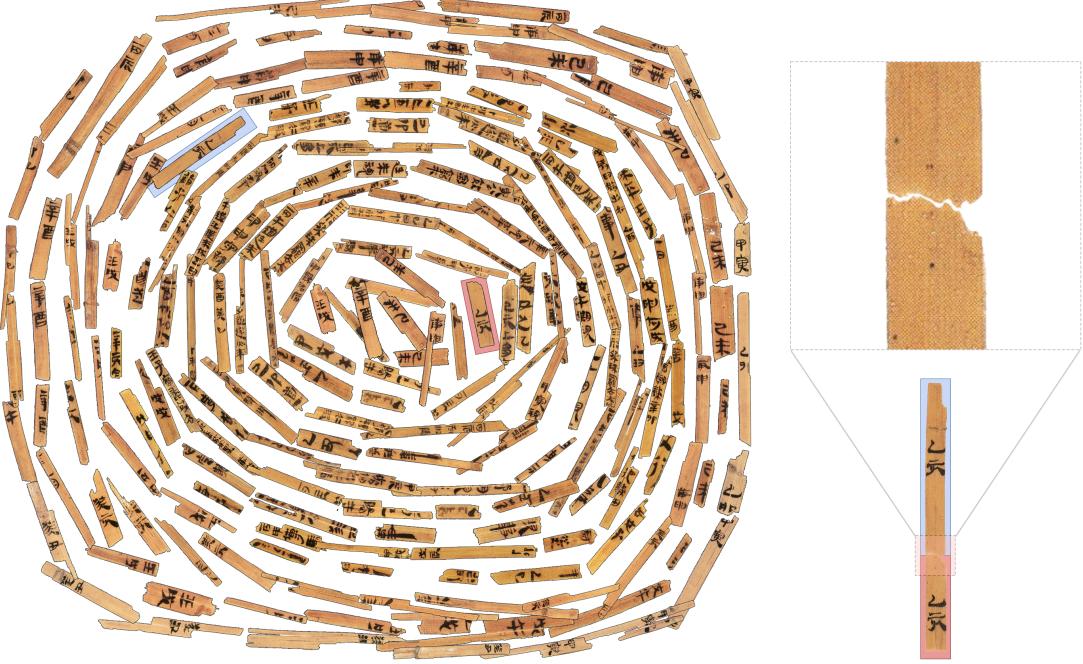


Figure 1: The challenge of bamboo slip rejoicing in archaeological research. Some excavated bamboo slip fragments from the Qin dynasty (221–206 BCE) arranged in a spiral pattern (left) illustrate the overwhelming scale of the rejoicing challenge faced by archaeologists—with thousands of fragments at single excavation sites, each potentially matching with any other piece. A zoomed view of a representative fragment (upper right) reveals the characteristic irregular breakage pattern along the fracture curve. An example of a reconstructed complete slip (lower right) demonstrates how proper rejoicing establishes morphological continuity, which is essential for the interpretation of ancient texts. The complex morphology of the fractures, coupled with deterioration from millennia underground, creates a matching problem of extraordinary complexity that has traditionally required intensive manual effort, with experts sometimes spending weeks to successfully match a single pair of fragments.

the restoration process highly labor-intensive, with experts sometimes requiring weeks to successfully match a single pair of fragments.

Deep learning for ancient artifact repair

Here we overcome the challenges of fragmented bamboo slip rejoicing through physics-driven deep learning. Our approach integrates principles from the physics of fracture with deep neural networks, enabling automated matching of fragmented pieces without relying on manually labeled training data. Deep learning [10] has been successfully used in restoring and attributing ancient texts [11] empowered by large training datasets. Beyond cultural heritage applications, deep learning has also demonstrated remarkable success across diverse scientific domains [12–18]. While recent years have also seen significant progress in fragmented cultural relic restoration [19] with labeled data - from oracle bone reconstruction [20, 21] to manuscript fragment matching [22, 23] - most existing approaches rely on traditional computer vision techniques and struggle with complex degradation patterns [24]. Some works have attempted to use generative models for cultural heritage restoration [21]. While achieving promising results, these methods are constrained by the scarcity of training data and often fail to capture the intricate physical properties of ancient materials [25, 26].

We propose WisePanda, a deep learning framework that leverages physical principles to overcome the data scarcity problem inherent in fragmented bamboo slip rejoicing task. Training WisePanda presents a unique paradox: while manual fragment rejoicing is prohibitively time-consuming. The very problem we aim to solve - this same process would traditionally be required to generate training data for the model. We resolve this dilemma by resorting to the physics of fracture [27]. By modeling the physical

properties of bamboo and the process that govern its degradation [9, 28], we generate extensive synthetic training data that captures the essential characteristics of real paired fragment slips (Figure 2). This physics-driven approach enables us to produce large-scale, realistic paired training data without requiring manual matching efforts, while ensuring the model learns meaningful patterns based on actual material properties rather than superficial features.

Data synthesis with physics of fracture

The generation of training data for bamboo slip rejoicing requires a comprehensive understanding of both fracture formation and degradation processes [29]. Our physical model builds on two key observations: First, bamboo slips exhibit distinct “transverse” and “longitudinal” fracture modes due to their unique fiber structure [29], as illustrated in Extended Figure E1, with transverse fractures accounting for approximately 70% of cases. Second, these fractures undergo complex corrosion processes over time, significantly altering their original patterns [2, 9, 30]. Building on these insights, we develop a systematic approach to generate synthetic fragment pairs for transverse fractures, which represent the majority of rejoicing challenges.

The physics of fracture in bamboo slips is governed by their distinctive vertical fiber arrangement [31–33], as detailed in Extended Figure E2. When stress extends horizontally across the slip, it creates characteristic fracture patterns as it propagates through consecutive fiber bundles [33, 34]. We model this process through a probabilistic framework where the fracture angle of each fiber bundle is influenced by the stress field generated at the fracture endpoint of the preceding fiber bundle (Figure 2a and b). This physical relationship can be expressed through a probability density function that determines likely fracture paths based on stress distribution and geometric configurations. The corrosion process is then simulated by analyzing geometric exposure - protruding areas and exposed surfaces are more susceptible to degradation through moisture absorption and microorganism activity [35]. To efficiently calculate the exposure area of each fiber bundle, we ingeniously employ ReLU functions that compare the height differences between adjacent fiber bundles, effectively capturing only the protruding portions that are more vulnerable to environmental deterioration (Figure 2c). This computational approach allows us to precisely model the progressive and differential deterioration patterns observed in archaeological samples.

To ensure the generated data accurately reflects real bamboo slip characteristics, we optimize model parameters using a genetic algorithm [36]. The algorithm evaluates parameter sets by comparing the distribution of generated fragments against a reference set of 200 real fragment curves through dimensionality reduction and Silhouette analysis [37]. Our analysis shows high similarity between the synthetic and real data distributions. The resulting pipeline enables us to generate extensive paired training data (Figure 2d and e) that captures both the physical properties of fracture formation and the effects of long-term degradation, effectively addressing the data scarcity challenge in rejoicing fragmented bamboo slips.

WisePanda is a ranking system

WisePanda’s architecture features an interpretable pipeline for fragmented bamboo slip rejoicing task (Figure 3), using synthesized data with physics of fracture for model training and providing Top-k candidates to assist archaeologists. At its core lies a TripletNet-based deep learning network [38], designed to learn effective feature representation that can distinguish between matching and non-matching fragments. The network processes each input as a 64-dimensional vector that captures the detailed features of the fracture curves. By leveraging a triplet loss function [38, 39], the network learns to minimize the distance between matching pairs while maximizing the distance between non-matching ones.

In practical deployment, WisePanda operates by predicting matching probabilities between fragments and offers archaeologists a ranked list of the top potential matches for each fragment. This significantly reduces the search space from thousands of candidates to a manageable set [40, 41], thereby streamlining the rejoicing process. The network’s output is a value between 0 and 1, indicating the match probability between a given pair of fragments, where a value closer to 1 signifies a higher likelihood of a correct match. This ranking approach effectively mitigates the overwhelming task of manually matching fragments by narrowing down the possibilities to the most probable matches.

WisePanda as an archaeologist adjunct

To bridge the gap between computational methods and archaeological research practices, we have developed an intuitive computer-assisted tool that implements WisePanda’s capabilities in a practical

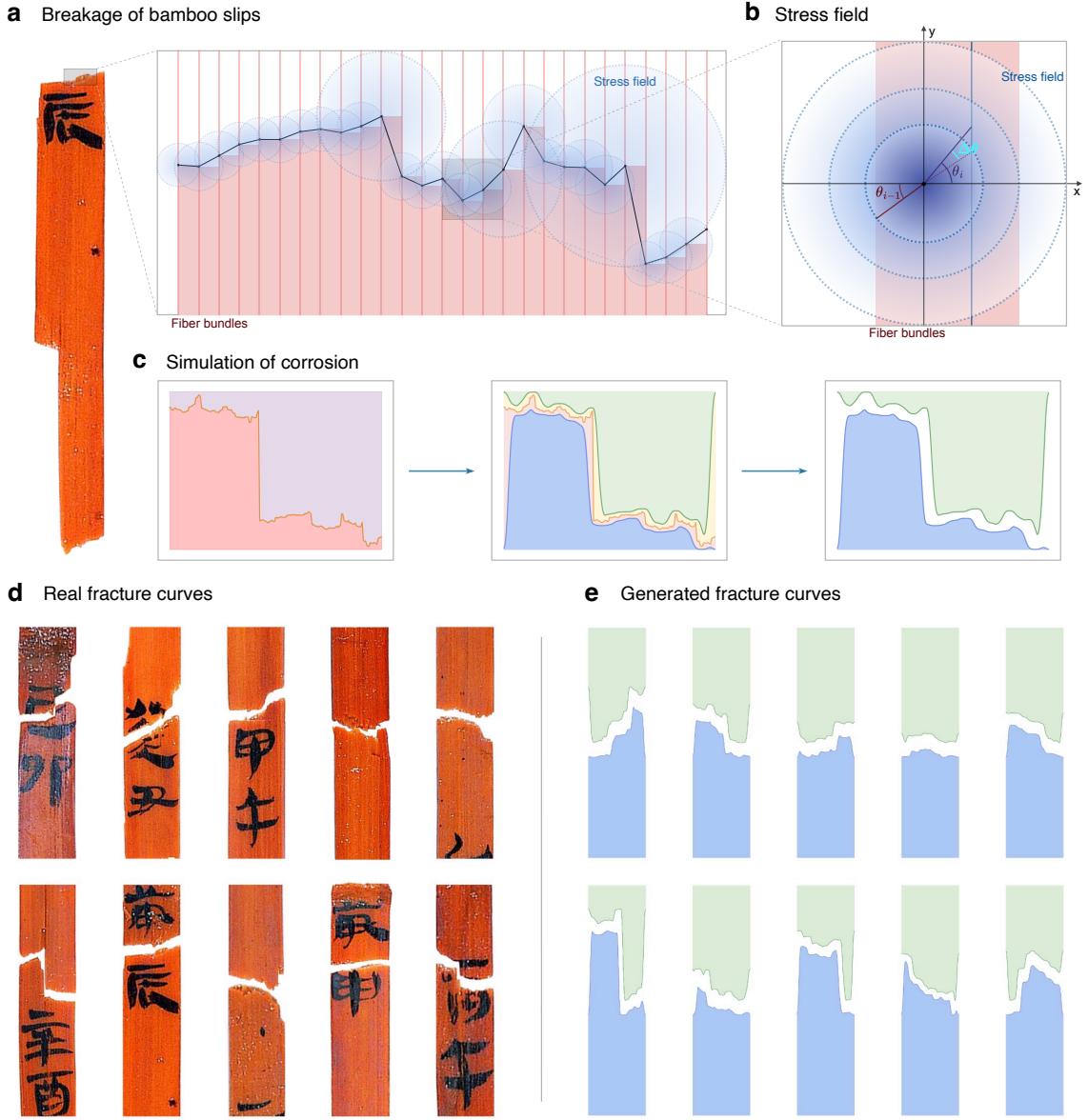


Figure 2: Physics-driven modeling of bamboo slip fracture and deterioration. **a**, The breakage process of bamboo slips showing how fracture propagates across the bamboo's fiber structure, with the resulting irregular curve composed of black line segments and the corresponding stress field distribution (blue gradient). **b**, Detailed stress field model illustrating the mathematical relationship between the fracture angles (θ_{i-1} , θ_i) and the stress propagation in the x - y coordinate system, where the blue dotted circles represent the stress field emanating from the fracture endpoint, showing how stress radiates outward and concentrates at fiber boundaries. **c**, Time-sequential simulation of the corrosion process, demonstrating how the original fracture pattern (left) changes through environmental exposure (middle) to produce the final deteriorated curve morphologies (right), with protruding areas experiencing accelerated degradation. **d**, Collection of real bamboo slip fracture curves extracted from archaeological samples, exhibiting diverse breakage patterns that serve as reference for model validation. **e**, Synthetically generated fracture curves produced by our physics-driven model, displaying morphological characteristics highly similar to the real samples in panel d. This approach enables the generation of extensive paired training data that captures both the physical properties of bamboo fragmentation and the effects of long-term degradation, effectively addressing the data scarcity challenge inherent in ancient artifact restoration.

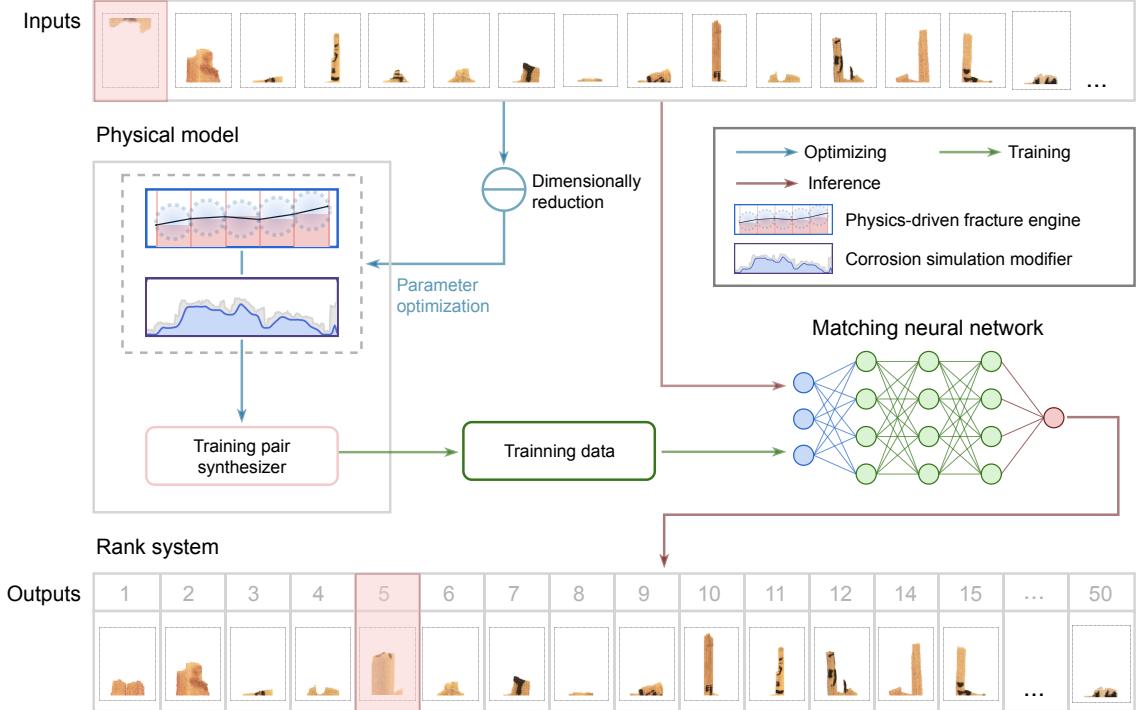


Figure 3: Framework of WisePanda integrating physics-driven deep learning for bamboo slip rejoicing. The workflow diagram illustrates the complete pipeline of our system. The process begins with input bamboo slip fragments (top row), where the target fragment is highlighted in pink. In the physical model stage, we implement two key components: a physics-driven fracture engine that generates initial breakage patterns, and a corrosion simulation modifier that replicates long-term degradation effects. The dimensionality reduction module (center) enables comparison between real and synthetic fragments through statistical visualization techniques, facilitating parameter optimization to ensure realistic fracture simulation. The optimized physical model feeds into the training pair synthesizer, which generates matching fragment pairs for training the matching network. This matching network learns to effectively represent and compare fragment curves through a multi-layer architecture with specialized embedding capabilities. During inference (red arrows), the system compares the target fragment against all candidates, producing a ranked list of potential matches (bottom row). The ranking transforms thousands of possible matches into a manageable set, with the system highlighting the most probable candidates (position 5 shown in pink represents a correct match). This approach enables archaeologists to focus their verification efforts on the most promising candidates, significantly accelerating the traditionally time-consuming rejoicing process.

workflow (Extended Figure E5). When working with this tool, archaeologists first select a query fragment from the collection. The system then automatically generates a ranked list of potential matching candidates, significantly narrowing down the search space from thousands to dozens of fragments. Each suggested match is accompanied by a matching score given by WisePanda’s matching network, allowing archaeologists to prioritize their examination of the most promising candidates.

The tool provides a dedicated workspace where archaeologists can visually manipulate and compare fragments through intuitive operations such as dragging, rotating, and fine-tuning alignments. To facilitate detailed examination, the system offers specialized visualization features like layer swapping and curve enhancement, enabling careful inspection of fracture patterns, ink traces, and textures (*e.g.*, fiber patterns). All successful matches are saved and documented in a centralized database, creating a growing repository of verified rejoicing that contributes to the broader restoration effort. This computer-assisted tool has significantly accelerated the traditional rejoicing workflow [41], ensuring archaeologists maintain full control over the final verification of matches while benefiting from AI-powered suggestions.

Experimental evaluation

Datasets. For evaluating WisePanda’s effectiveness, we collected several datasets of bamboo and wooden slip fragments. Our primary test set (Bamboo236) consists of 118 pairs of matched bamboo slip fragments, carefully verified and provided by archaeological experts who successfully rejoined these fragments through meticulous manual work (achieved by three archaeological experts within half a year). These bamboo slip artifacts are part of an ongoing archaeological research project, with the cultural relics yet to be formally published. To simulate more challenging real-world scenarios, we created an extended dataset (Bamboo1350) by introducing 1114 additional interference fragments from the same set, expanding the candidate pool to 1,350 total fragments while maintaining the original matched pairs. To assess generalization capabilities across different materials, we also collected wooden slip datasets. The base wooden slip test set (Wood670) contains 335 expert-verified pairs, while its extended version (Wood3833) includes 3163 additional interference fragments. These diverse datasets enable comprehensive evaluation of fragment matching methods under varying conditions and material types.

Evaluation methodology. The evaluation process involved computing similarity scores between each fragment’s fracture curve and all potential matching candidates from the pool. Fragments were pre-classified as upper or lower parts, with matching only performed between these distinct groups. We extracted features from both fracture curves of each fragment, ranked the computed similarity scores, and measured the probability of finding the correct match within the Top-k predictions. We conducted comprehensive experiments comparing our WisePanda with several classical methods, including Manual Random Search, curve matching techniques (using Dynamic Time Warping (DTW)) [42], Fast Matching Method (FMM) [43], and Scale Invariant Signature (SIS) [44]. This comparative framework allows us to assess the relative performance of our physics-driven approach against established techniques in the field of fragment matching.

Results and analysis. The performance variations among different methods reveal insights about the nature of the fragmented bamboo slip matching problem. In our original test set (Bamboo236), WisePanda demonstrated superior performance across all metrics, achieving a Top-1 accuracy of 12.29%, Top-5 accuracy of 35.17%, Top-10 accuracy of 52.54%, Top-20 accuracy of 69.07%, and Top-50 accuracy of 94.07%. Among traditional methods, SIS [44] achieved the second-best performance with a Top-1 accuracy of 10.59% and Top-50 accuracy of 72.88%, followed by DTW [42] with a Top-50 accuracy of 75.42%. FMM [43] performed less effectively, with Top-50 accuracies of 65.25%. To simulate real-world archaeological scenarios where numerous unrelated fragments exist, we expanded our test dataset by introducing additional interference fragments. As shown in Extended Figure E4, when the candidate pool was increased by a factor of approximately 10, WisePanda maintained a Top-50 accuracy of 52.54% compared to the original 94.07%, demonstrating robust performance even with significantly increased search space.

Interpretability of WisePanda

The fundamental challenge in generating training data for fragmented bamboo slip matching lies in the irreversible nature of the physical degradation process [9, 24]. When archaeologists uncover bamboo slip fragments, they observe only the final degraded curves [2] (Figure 2c). The original fracture patterns are permanently altered by centuries of degradation and cannot be directly inferred from the preserved fragments [7]. This irreversibility poses a significant limitation for conventional data-driven approaches [21, 25, 26], which typically require paired samples of “before-degradation” and “after-degradation” states for effective training data that is inherently unavailable in archaeological contexts. Even generative models like generative adversarial networks and diffusion models suffer from distribution bias when learning only from real fragments, failing to incorporate the underlying physical principles that govern fracture formation and degradation.

Our physics-driven model addresses this challenge by modeling the forward process of fragmentation and degradation. We first simulate the initial fracture patterns based on the physics of bamboo fiber structures [28, 29, 45, 46], then apply physical degradation processes to these patterns. This sequential modeling captures both the immediate fracture characteristics and their subsequent transformation through environmental exposure. The physical processes governing bamboo deterioration follow specific geometric principles, where protruding areas and exposed surfaces are more susceptible to degradation [9, 24]. By incorporating these physical constraints, WisePanda generates realistic degradation patterns that closely mirror natural deterioration processes, effectively bridging the gap between the unknown original fracture patterns and the observed degraded edges.

Table 1: Comparison of fragment matching methods across different test scenarios

Method	Top-1(%)	Top-5(%)	Top-10(%)	Top-20(%)	Top-50(%)	Top-100(%)
Bamboo236	Manual Random Search	0.80	4.22	8.40	17.00	42.30
	DTW [42]	7.20	22.46	30.93	49.58	75.42
	FMM [43]	8.90	18.64	26.27	38.14	65.25
	SIS [44]	10.59	22.03	30.93	47.03	72.88
	WisePanda	12.29	35.17	52.54	69.07	94.07
Bamboo1350	Manual Random Search	0.08	0.44	0.89	1.79	4.48
	DTW [42]	3.81	12.29	14.41	20.34	32.20
	FMM [43]	6.36	11.44	13.14	16.10	25.00
	SIS [44]	7.20	15.25	19.07	24.15	36.02
	WisePanda	5.93	18.22	24.58	37.29	52.54
Wood	DTW [42][Wood670]	2.69	6.72	11.04	15.82	23.43
	FMM [43][Wood670]	2.24	6.72	9.25	13.28	21.94
	SIS [44][Wood670]	2.24	7.91	10.75	15.22	23.43
	WisePanda[Wood670]	1.94	10.00	17.31	25.22	32.99
	WisePanda[Wood3833]	1.19	2.84	5.67	9.25	15.67

Results show Top-k matching accuracy (percentage of correctly identified matches within the top k candidates) across bamboo and wooden slip datasets. The original bamboo test set (Bamboo236) contains 118 paired fragments, while the expanded dataset (Bamboo1350) expands the candidate pool with 1,114 additional interference fragments. Similarly, for wooden slips, we compare performance on the original test set (Wood670) and an expanded dataset (Wood3833). As demonstrated in Extended Figure E4, WisePanda consistently outperforms traditional curve matching approaches in both material types and across varying complexity levels, demonstrating the effectiveness of our physics-driven approach in handling diverse archaeological fragments.

These physical principles not only enhance the quality of synthetic training data but also provide interpretability to our WisePanda’s behavior. The combination of authentic fracture patterns and physically-based degradation enables WisePanda to capture fundamental characteristics of bamboo slip fragmentation, generating abundant realistic training data and leading to more robust and generalizable performance. Unlike pure data-driven approaches that may overfit to superficial patterns in limited training samples, our physics-driven model learns meaningful features grounded in the actual physical processes of fracture and degradation, allowing it to generalize effectively to unseen fragments with diverse deterioration patterns.

Rejoin fragmented ancient wooden slips

To examine WisePanda’s generalization capability, we applied our approach to wooden slips - another important medium for ancient text preservation that shares similar characteristics with bamboo slips but exhibits distinct material properties. While wooden slips served similar documentary purposes, their different fiber arrangement and material structure present unique challenges for fragment matching [47]. We evaluated our method on a dataset of 335 pairs of wooden slip fragments, following the same experimental protocol as with bamboo slips.

As shown in Table 1, WisePanda demonstrated promising cross-material generalization with 32.99% Top-50 accuracy on wooden slips. When tested with extended interference fragments in both materials, performance predictably decreased to 52.54% for bamboo slips (Bamboo1350) and 15.67% for wooden slips (Wood3833).

This performance difference can be attributed to two key factors. First, the fiber structure of wooden slips differs substantially from bamboo slips, leading to distinct fracture patterns that deviate from our physics-driven model’s assumptions [48]. Second, the degradation process varies significantly due to diverse preservation conditions - factors such as burial duration, soil moisture levels, microbial activity, and regional soil composition all affect how materials deteriorate over time [49]. Even artifacts of the same material excavated from different sites or time periods may exhibit varying degradation patterns, requiring different physical parameters for accurate modeling. These variations are particularly pronounced when comparing wooden and bamboo materials, as their distinct physical properties interact differently with environmental factors, leading to material-specific degradation processes that need to be carefully considered in the modeling approach.

Discussion

In this work, we present WisePanda, the first physics-driven deep learning framework designed for rejoining fragmented ancient bamboo slips. By incorporating principles from the physics of fracture into the generation of synthetic training data, our approach overcomes the fundamental challenge of paired data scarcity in AI for rejoining fragmented ancient bamboo slips, and significantly improves the efficiency of archaeologists. The effectiveness of this approach is demonstrated through superior matching accuracy and the development of practical tools that are now assisting archaeologists in their reconstruction work. Our framework's success in handling both fracture pattern generation and material degradation simulation highlights the advantage of combining physical principles with deep learning, particularly in scenarios where traditional data-driven approaches are constrained by limited training samples.

Beyond its current implementation, this research opens new horizons in bamboo slip reconstruction and broader cultural heritage preservation. For bamboo slips specifically, our framework can be extended to handle more complex scenarios by incorporating additional physical parameters that account for regional variations and different preservation conditions. Furthermore, our approach establishes a new paradigm for fragment matching where paired training data is scarce. The incorporation of physics-driven mechanism can be adapted to other archaeological materials such as wooden slips, ceramics, metals, and pottery, by modeling their specific physics of fracture and degradation processes.

In conclusion, WisePanda represents not merely a technical advancement in fragmented bamboo slip rejoining, but rather a transformative approach to preserving and studying cultural heritage. By combining physical principles with artificial intelligence, our method provides a robust solution when paired training data is limited but physical principles are well understood. This synergy of physics and deep learning opens new possibilities for archaeological fragment matching, potentially revolutionizing how we recover and interpret artifacts from our past.

References - Main

1. Allan, S. *Buried ideas: Legends of abdication and ideal government in early Chinese bamboo-slip manuscripts* (State University of New York Press, 2015).
2. Liu, G. *Approaching the Tsinghua Bamboo Slips* (Taylor & Francis, 2024).
3. Liu, X. From Bamboo Slips to Received Versions: Common Features in the Transformation of the "Laozi". *Harv. J. Asiat. Stud.* **63**, 337–382 (2003).
4. Ling, L. Formulaic structure of Chu divinatory bamboo slips. *Early China* **15**, 71–86 (1990).
5. Zheng, W. A brief explanation of the Chu bamboo slips Hengxian. *Front. Philos. China* **1**, 418–431 (2006).
6. Johnson, M. *Archaeological theory: an introduction* (John Wiley & Sons, 2019).
7. Zhou, J. *et al.* Identifying designs from incomplete, fragmented cultural heritage objects by curve-pattern matching. *J. Electron. Imaging* **26**, 011022–011022 (2017).
8. *Shuihudi Qin Mu Zhujian (Shui hu di qin mu zhu jian)* (ed Group, S. Q. T. B. S. C.) Title translates to: Bamboo Slips from the Qin Tomb at Shuihudi. [in Chinese] (Wenwu Chubanshe (Cultural Relics Publishing House), Beijing, 1990).
9. Cha, M. Y., Lee, K. H. & Kim, Y. S. Micromorphological and chemical aspects of archaeological bamboos under long-term waterlogged condition. *Int. Biodeterior. Biodegrad.* **86**, 115–121 (2014).
10. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
11. Assael, Y. *et al.* Restoring and attributing ancient texts using deep neural networks. *Nature* **603**, 280–283 (2022).
12. Silver, D. *et al.* Mastering the game of go without human knowledge. *Nature* **550**, 354–359 (2017).
13. Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
14. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
15. Dafoe, A. *et al.* Cooperative AI: machines must learn to find common ground. *Nature* **593**, 33–36 (2021).

16. Merchant, A. *et al.* Scaling deep learning for materials discovery. *Nature* **624**, 80–85 (2023).
17. Lu, M. Y. *et al.* A multimodal generative AI copilot for human pathology. *Nature* **634**, 466–473 (2024).
18. Bodnar, C. *et al.* A foundation model for the Earth system. *Nature*, 1–8 (2025).
19. Bickler, S. H. Machine learning arrives in archaeology. *Proc. of Adv. Archaeol. Pract.* **9**, 186–191 (2021).
20. Meng, L., Kamitoku, N. & Yamazaki, K. Recognition of oracle bone inscriptions using deep learning based on data augmentation. in *MetroArchaeo* (2018), 33–38.
21. Zhang, C. *et al.* Data-driven oracle bone rejoining: A dataset and practical self-supervised learning scheme. in *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining* (2022), 4482–4492.
22. Zhu, L., Zhou, Z. & Hu, D. Globally consistent reconstruction of ripped-up documents. *IEEE Trans. on Pattern Anal. and Mach. Intell.* **30**, 1–13 (2007).
23. Zheng, Y., Li, X. & Weng, Y. Reunion helper: an edge matcher for sibling fragment identification of the Dunhuang manuscript. *Heritage Science* **12**, 52 (2024).
24. Li, M.-Y. *et al.* Investigation into the deterioration process of archaeological bamboo strips of China from four different periods by chemical and anatomical analysis. *Polym. Degrad. Stab.* **109**, 71–78 (2014).
25. Eslami, D., Di Angelo, L., Di Stefano, P. & Pane, C. Review of computer-based methods for archaeological ceramic sherds reconstruction. *Virtual Archaeol. Rev.* **11**, 34–49 (2020).
26. Hou, M. *et al.* Novel method for virtual restoration of cultural relics with complex geometric structure based on multiscale spatial geometry. *ISPRS Int. J. Geo-Inf.* **7**, 353 (2018).
27. Anderson, T. L. & Anderson, T. L. *Fracture mechanics: fundamentals and applications* (CRC press, 2005).
28. Tan, T. *et al.* Mechanical properties of functionally graded hierarchical bamboo structures. *Acta Biomater.* **7**, 3796–3803 (2011).
29. Behera, D. *et al.* A review on bamboo fiber reinforced composites and their potential applications. *Emergent Mater.* **8**, 157–172 (2025).
30. Chen, Z., Ma, R., Du, Y. & Wang, X. State-of-the-art review on research and application of original bamboo-based composite components in structural engineering. in *Structures* **35** (2022), 1010–1029.
31. Dixon, P. G. & Gibson, L. J. The structure and mechanics of Moso bamboo material. *J. R. Soc. Interface* **11**, 20140321 (2014).
32. Ramful, R. & Sakuma, A. Investigation of the effect of inhomogeneous material on the fracture mechanisms of bamboo by finite element method. *Materials* **13**, 5039 (2020).
33. Chen, M. *et al.* Fracture modes of bamboo fiber bundles in three-point bending. *Cellulose* **26**, 8101–8108 (2019).
34. Ramful, R. Investigation of the transverse fracture mechanisms of bamboo by the finite element method. *J. Mater. Sci.* **57**, 6233–6248 (2022).
35. Blanchette, R. A., Haight, J. E., Koestler, R. J., Hatchfield, P. B. & Arnold, D. Assessment of deterioration in archaeological wood from ancient Egypt. *J. Am. Inst. Conserv.* **33**, 55–70 (1994).
36. Goldberg, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, Reading, MA, 1989).
37. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9** (2008).
38. Hoffer, E. & Ailon, N. Deep metric learning using triplet network. in *Proc. Int. Worksh. on Simil.-Based Pattern Recognit.* (2015), 84–92.
39. Tian, Y., Barroso Laguna, A., Ng, T., Balntas, V. & Mikolajczyk, K. HyNet: Learning local descriptor with hybrid similarity measure and triplet loss. *Proc. of Adv. Neural Inf. Process. Syst.* **33**, 7401–7412 (2020).
40. Funkhouser, T. *et al.* Learning how to match fresco fragments. *ACM J. Comput. Cult. Herit.* **4**, 1–13 (2011).
41. Brown, B. *et al.* Tools for virtual reassembly of fresco fragments. *Int. J. Herit. Digit. Era* **1**, 313–329 (2012).

42. Berndt, D. J. & Clifford, J. Using dynamic time warping to find patterns in time series. in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* (1994), 359–370.
43. Frenkel, M. & Basri, R. Curve matching using the fast marching method. in *Proc. Int. Worksh. on Energy Minimiz. Methods in Comput. Vis. and Pattern Recognit.* (2003), 35–51.
44. Calabi, E., Olver, P. J., Shakiban, C., Tannenbaum, A. & Haker, S. Differential and numerically invariant signature curves applied to object recognition. *Int. J. Comput. Vis.* **26**, 107–135 (1998).
45. Amada, S., Ichikawa, Y., Munekata, T., Nagase, Y. & Shimizu, H. Fiber texture and mechanical graded structure of bamboo. *Compos. B: Eng.* **28**, 13–20 (1997).
46. Low, I., Che, Z. & Latella, B. Mapping the structure, composition and mechanical properties of bamboo. *J. Mater. Res.* **21**, 1969–1976 (2006).
47. Staack, T. Reconstruction of Early Chinese Bamboo and Wood Manuscripts: A Review (1900–2010). *CSCMC Occas. Pap.* **5** (2016).
48. Haiyan, T. Damage of bamboo and wooden materials based on linear elastic fracture mechanics in garden design. *Fractur. Struct. Integr.* **10**, 472–480 (2016).
49. Singh, A. P., Kim, J. S., Möller, R., Chavan, R. R. & Kim, Y. S. The pivotal role of microscopy in unravelling the nature of microbial deterioration of waterlogged wood: A review. *Forests* **15**, 889 (2024).

Methods

Previous work

In recent years, various approaches have been proposed for restoring and analyzing fragmented ancient artifacts [1, 2]. Traditional methods have primarily focused on geometric feature analysis, particularly curve matching methods. Modern approaches mainly leverage artificial intelligence techniques [3, 4]. Notable recent work includes the restoration of Pompeii’s frescoes using a robotic system for automated fragment matching [5], Ithaca, a deep neural network system that successfully restored and attributed ancient Greek texts [6], and the reunion helper for Dunhuang manuscript fragments [7, 8]. For oracle bone fragments, researchers have explored using GAN to generate training samples for restoration tasks [9].

The manual restoration of bamboo and wooden slips has a long history in China [10–12]. Wang Guowei’s pioneering work *Liusha Zhuijian* [13] systematically documented traditional techniques for matching and rejoicing ancient bamboo and wooden manuscripts, establishing foundational methodologies that are still valuable today. However, manual restoration remains an extremely time-consuming and labor-intensive task, often requiring weeks or months to successfully match a single pair of fragments. With recent technological advancements, computational approaches and artificial intelligence have emerged as powerful tools to assist archaeologists in this complex process, significantly accelerating the restoration workflow while maintaining the essential role of expert knowledge. While deep learning approaches have shown promise in cultural relic restoration, the application to fragmented bamboo slips faces significant challenges due to the lack of sufficient paired training data and the complex physical degradation patterns unique to bamboo materials. Our work addresses these limitations by introducing physics-driven data generation. Unlike previous approaches, we uniquely integrate fracture mechanics and material degradation principles into the restoration process. This novel approach allows us to generate physically plausible paired training data that reflects the actual physical processes of how bamboo slips break and degrade over time.

Moreover, our physics-driven approach not only provides interpretable results that align with archaeological domain knowledge, but also demonstrates strong extensibility to various archaeological fragment restoration tasks. The framework we propose may open a new paradigm for applying deep learning in archaeological restoration [14], particularly valuable when paired training data is scarce but physical principles are well understood. This methodology can potentially be adapted to restore other types of archaeological fragments by modeling their specific physical properties and degradation processes [15, 16], offering a generalizable solution for the broader field of cultural heritage preservation.

Physical model for data generation

Fracture modeling. Bamboo slip fragments exhibit two distinct fracture modes: transverse and longitudinal, with transverse fractures accounting for approximately 70% of cases [17]. Due to bamboo’s

unique fiber structure, the transverse fractures occur when stress cuts across the bamboo slip's cross-section, creating uneven, wavy edge curves. The longitudinal fractures happen along the fiber direction, resulting in straight edges, as illustrated in Extended Figure E1. In this work, we focus on transverse fractures due to their prevalence and complexity. To systematically model the bamboo slip structure, we establish a coordinate system where the x -axis represents the horizontal direction along the bamboo slip's width and the y -axis represents the vertical direction corresponding to fiber bundle heights. We abstract the bamboo slip as a series of N_f vertically aligned fiber bundles, as shown in Extended Figure E2a. Each fiber bundle is represented as a discrete structural unit with a uniform width of δ_x , forming a continuous arrangement along the horizontal axis. The initial height for the front of i -th fiber bundle before environmental degradation is denoted as y_i^{init} . The final height after corrosion processes is represented as y_i^{final} . This discretization allows us to precisely model the propagation of fracture across the material's anisotropic structure. The vertical positions of these fiber bundles determine the morphology of the fracture curve, with the height of each bundle representing its position on the fracture curve. When modeling the fracture process, we observe that bamboo slips were typically buried in a loosely rolled form, causing them to experience predominantly Mode III (out-of-plane shear) stress during burial compression and subsequent deterioration. This specific stress mode occurs when opposing forces act parallel to each other but in different planes, creating a tearing effect particularly relevant to rolled manuscripts. We model the bamboo slip structure as consecutive fiber bundles [18–21], where stress propagates from one bundle to another one (Figure 2a). At the terminal point of each fractured fiber, a stress component K_{III} continues to propagate [22], following the formula:

$$K_{III} = \sqrt{2\pi}\sigma\sqrt{\alpha}, \quad (1)$$

where σ represents the far-field shear force and α denotes the cumulative stress extension path length. The stress field at the fracture propagation front $\sigma_{yz}(r, \theta)$ can be described by:

$$\sigma_{yz}(r, \theta) = \frac{K_{III}}{\sqrt{2\pi r}} \cos\left(\frac{\theta}{2}\right), \quad (2)$$

where r is the radial distance from the fracture propagation front and θ is the angular position relative to the direction of stress propagation, as shown in Extended Figure E2b. This equation describes how the stress force decreases with distance from the fracture propagation front while also varying with direction.

For the xy plane of the writing surface, we develop a probabilistic model to describe how fracture propagates across consecutive fiber bundles. As illustrated in Extended Figure E2b, when a fracture reaches point P_2 at the end of fiber bundle f_{i-1} , it generates a stress field that determines the probable direction of continued fracture propagation into the next fiber bundle f_i . The blue dotted circles emanating from point P_2 represent this stress field distribution, with point P_1 marking the beginning of the fracture on bundle f_{i-1} and point P_3 indicating the potential endpoint of the fracture on bundle f_i . The stress field on the xy plane is described by:

$$\sigma_{xy}(r, \theta) = V_{xy} \cdot \frac{K_{III}}{\sqrt{2\pi r}} \cos\left(\frac{\theta}{2}\right), \quad (3)$$

where V_{xy} is a dimensionless constant representing the projection of out-of-plane (Mode III) stress onto the writing surface plane, r is the radial distance from the fracture propagation front at point P_2 , and θ is the angular position in the stress field. This formulation accounts for bamboo's anisotropic fiber structure in translating three-dimensional stress into planar fracture propagation. The angles θ_{i-1} and θ_i , measured relative to the horizontal axis, define the fracture directions in consecutive fiber bundles, with $\theta_i = \theta_{i-1} + \Delta\theta$. The stress field determines the probability distribution of angle change $\Delta\theta$, which governs how the fracture direction transitions between fiber bundles. This probability density function can be expressed as:

$$p(\Delta\theta|\theta_{i-1}) = \frac{V_{xy} \cdot K_{III} \cdot \cos\left(\frac{\Delta\theta}{2}\right)}{\sqrt{2\pi}(\delta_x \cos(\theta_{i-1} + \Delta\theta))}. \quad (4)$$

By substituting $K_{III} = \sqrt{2\pi}\sigma\sqrt{\alpha}$, we simplify the above equation to:

$$p(\Delta\theta|\theta_{i-1}) = \frac{V_{xy} \cdot \sigma\sqrt{\alpha} \cdot \cos\left(\frac{\Delta\theta}{2}\right)}{\delta_x \cos(\theta_{i-1} + \Delta\theta)}. \quad (5)$$

While V_{xy} and σ are unmeasurable constants that scale the probability distribution, they do not affect the relative probabilities of different fracture paths. Once the fracture angle is determined, the initial

vertical height relationship between adjacent fiber bundles follows:

$$y_{i+1}^{init} - y_i^{init} = \tan(\theta_i) \cdot \delta_x. \quad (6)$$

This mathematical modeling enables us to construct a Markov chain-like model for fracture propagation, where each state represents a fiber bundle's initial front height y_i^{init} , and transitions between states are governed by the probability density function $p(\Delta\theta|\theta_{i-1})$. This results in a physically plausible model that captures how Mode III stress-induced fracture propagates through bamboo's anisotropic structure, allowing us to generate diverse yet authentic synthetic fracture patterns that serve as the foundation for subsequent corrosion simulation to obtain the final degraded fracture curves.

Corrosion process simulation. After the initial fracture curves are generated, the bamboo slip fragments undergo long-term environmental degradation that significantly alters their original morphology. The degradation process of bamboo slip fragments is highly dependent on their geometric features, where sharp edges and protrusions are more susceptible to corrosion compared to smooth surfaces [23, 24], even with identical perimeter lengths. This geometric dependency is primarily due to stress concentration at sharp points and increased surface area exposure to environmental factors [23]. We observe that in archaeological samples, protruding areas of bamboo slip edges show significantly more deterioration than recessed areas, consistent with their greater environmental exposure. To mathematically capture this geometric dependency in our corrosion model, we employ the ReLU function to calculate the structural exposure area for each fiber bundle. We define the structural exposure area S_i as:

$$S_i = \text{ReLU}(y_i^{init} - y_{i-1}^{init}) + \text{ReLU}(y_i^{init} - y_{i+1}^{init}). \quad (7)$$

The erosion process is then modeled by reducing the fiber height based on the structural exposure area and a corrosion rate coefficient c_e . The final degraded height for the front of i -th fiber bundle y_i^{final} is given by:

$$y_i^{final} = y_i^{init} - (S_i \times c_e). \quad (8)$$

This approach ensures that regions with higher geometric exposure experience accelerated degradation, accurately simulating how moisture and microorganisms propagate through the bamboo structure [23–25]. To capture the progressive nature of long-term degradation, we implement an iterative corrosion process where the degradation is applied over multiple time steps, with each iteration representing a period of environmental exposure. The synchronized update mechanism ensures that all fiber heights are calculated based on the current state before updating to the next iteration, preventing cascade effects where early updates could disproportionately influence later calculations within the same corrosion cycle. This multi-step simulation accurately reflects the physical degradation process observed in actual bamboo slips, where environmental factors affect the entire surface concurrently rather than sequentially, ultimately generating realistic synthetic fragment pairs that capture both the initial fracture characteristics and their subsequent transformation through centuries of environmental exposure.

Parameter optimization. To ensure our physical model accurately reflects real bamboo slip characteristics, we develop a parameter optimization strategy driven by comparison between real and synthetic data distributions. We first select a set of 200 real fragment curves (set A) extracted from archaeological bamboo slips, which represent the final state after both fracture formation and corrosion processes, with curves characterized by their y^{final} coordinates along the fragment curve. Concurrently, we generate a set of 200 simulated fragment curves (set B) using randomly initialized parameters in our physical model, where the synthetic fragments undergo the complete process from initial fracture generation (y^{init} heights) through corrosion simulation to obtain final degraded curve profiles (y^{final} heights). Both sets are subjected to t-SNE (t-Distributed Stochastic Neighbor Embedding) [26] dimensionality reduction, projecting the high-dimensional curve data into a 2D space to facilitate comparison of their distributions. The silhouette score s for these reduced data sets is calculated as:

$$s(j) = \frac{b(j) - a(j)}{\max\{a(j), b(j)\}}, \quad (9)$$

where $a(j)$ represents the mean distance between point j and all other points within its own set (either set of real fragments A or set of synthetic fragments B), and $b(j)$ is the mean distance between point j and all points in the other set. A silhouette score approaching zero indicates that the boundary between real and synthetic data distributions becomes increasingly indistinguishable, suggesting optimal similarity

between our simulated fragmented curves and actual archaeological fragments. To find the optimal set of parameters for our physical model, we employ a genetic algorithm [27] with the fitness function F as:

$$F = \frac{1}{|s|} + \lambda \cdot \text{diversity}, \quad (10)$$

where λ is a weight coefficient that balances optimization precision (minimizing silhouette score) with exploration capability (maintaining genetic diversity). The genetic algorithm manages a population of parameter configurations $\mathcal{P} = \{p_1, \dots, p_n\}$, where each parameter set p_i contains specific values for both fracture physics and corrosion simulation components of our model. During each optimization iteration, new candidate parameter sets are generated through controlled recombination:

$$p_{\text{new}} = \beta \cdot p_1 + (1 - \beta) \cdot p_2 + \epsilon, \quad (11)$$

where p_1 and p_2 are parent parameter sets selected based on their fitness scores, $\beta \in [0, 1]$ is a randomly generated crossover weight that varies with each recombination operation, and ϵ represents a small random mutation noise term added to maintain genetic diversity. This randomized crossover mechanism allows the algorithm to explore different combinations of parameter values while gradually converging towards configurations that produce simulated fracture patterns closely matching the real archaeological data [27].

Data preparation and augmentation

Fracture curve generation. Based on our analysis of Mode III (out-of-plane shear) stress-induced fracture mechanisms in bamboo slips [28], which typically occurs when rolled bamboo manuscripts experience perpendicular forces during burial and subsequent deterioration, we implement a probabilistic framework to generate initial fracture patterns. The generation process leverages Eq. (5), Eq. (6), and Eq. (7) to construct a Markov chain-like structure, where each state represents the initial height y_i^{init} of i -th fiber bundle (Algorithm 1). This algorithm generates fracture curves that closely mimic the physical behavior of bamboo fiber bundles under stress. The probabilistic nature of the generation process, governed by the Probability Density Function (PDF) derived from stress field analysis in Eq. (5), ensures that each generated curve exhibits realistic variations while maintaining the fundamental characteristics of bamboo slip fracture. These generated curves serve as the foundation for subsequent corrosion simulation.

Algorithm 1: Physics-Driven Fracture Pattern Generation

```

Input :  $\delta_x$ : Width of each fiber bundle
         $N_f$ : Number of vertically aligned fiber bundles
         $\theta_{\text{init}}$ : Initial fracture angle at  $(x_0, y_0) = (0, 0)$ 
Output:  $P^{\text{init}} = [(x_0, y_0^{\text{init}}), (x_1, y_1^{\text{init}}), \dots]$ 

// Initialization
 $P^{\text{init}} \leftarrow \text{new List}()$ 
 $current\_x \leftarrow 0.0$ 
 $current\_y^{\text{init}} \leftarrow 0.0$ 
 $current.\theta \leftarrow \theta_{\text{init}}$ 
Append  $(x_0 = current\_x, y_0^{\text{init}} = current\_y^{\text{init}})$  to  $P^{\text{init}}$ 
 $i \leftarrow 0$ 

// Iterative generation across the width
while  $i < N_f$  do
    // Sample the change in angle based on Eq. (5)
     $\Delta\theta \leftarrow \text{SampleAngleChange}(current.\theta)$ 
    // Calculate next state
     $next.\theta \leftarrow current.\theta + \Delta\theta$ 
     $\Delta y^{\text{init}} \leftarrow \tan(next.\theta) \times \delta_x$ 
     $next\_y^{\text{init}} \leftarrow current\_y^{\text{init}} + \Delta y^{\text{init}}$ 
     $next\_x \leftarrow (i + 1) \times \delta_x$ 
    // Update current state for next iteration
     $current\_x \leftarrow next\_x$ 
     $current\_y^{\text{init}} \leftarrow next\_y^{\text{init}}$ 
     $current.\theta \leftarrow next.\theta$ 
     $i \leftarrow i + 1$ 
    // Record the new point
    Append  $(x_i = current\_x, y_i^{\text{init}} = current\_y^{\text{init}})$  to  $P^{\text{init}}$ 
return  $P^{\text{init}}$ 

```

Corrosion simulation. Building upon Eq. (6), Eq. (7) and Eq. (8), we developed a multi-scale corrosion model that quantifies degradation based on the geometric exposure of individual fibers and their spatial relationships with adjacent fibers. The ReLU function effectively captures only the protruding portions that are more vulnerable to environmental degradation (Algorithm 2). Algorithm 2 implements a synchronized update mechanism where all fiber heights' expected changes are calculated first, and then all heights are updated simultaneously at each time step. This approach prevents cascade effects where early fiber updates could disproportionately influence later calculations within the same corrosion cycle. The number of iterations N_c and corrosion rate c_e are determined through parameter optimization to match observed degradation patterns in archaeological samples. This two-phase process - height calculation followed by simultaneous update - ensures that the corrosion simulation accurately reflects the physical degradation process observed in actual fragmented bamboo slips, where environmental factors affect the entire surface concurrently rather than sequentially. The algorithm transforms the initial fracture curves into realistic degraded fragments, completing the synthetic data generation pipeline.

Algorithm 2: Multi-Step Corrosion Simulation

```

Input :  $Y^{init}$ : Array of initial fiber heights  $[y_1^{init}, y_2^{init}, \dots, y_{N_f}^{init}]$ 
         $N$ : Number of corrosion simulation iterations
         $N_f$ : Number of vertically aligned fiber bundles
         $c_e$ : Corrosion rate coefficient
Output:  $Y^{final}$ : Array of fiber bundle heights after  $N$  iterations  $[y_1^{final}, y_2^{final}, \dots, y_{N_f}^{final}]$ 

// Initialization
 $Y^{current} \leftarrow \text{copy}(Y^{init})$  // Start with initial heights
 $N_f \leftarrow \text{length}(Y^{current})$  // Number of fiber bundles

// Perform  $N$  iterations of corrosion
for iteration from 1 to  $N$  do
     $Y^{next} \leftarrow \text{new Array}(N_f)$ 
    for  $i$  from 1 to  $N_f$  do
         $y_{curr} \leftarrow Y^{current}[i]$ 
        if  $i > 1$  then
            |  $y_{prev} \leftarrow Y^{current}[i - 1]$ 
        else
            |  $y_{prev} \leftarrow y_{curr}$ 
        if  $i < N_f$  then
            |  $y_{next} \leftarrow Y^{current}[i + 1]$ 
        else
            |  $y_{next} \leftarrow y_{curr}$ 
        // Calculate structural exposure using  $\text{ReLU}(x) = \max(0, x)$ 
         $exposure \leftarrow \max(0, y_{curr} - y_{prev}) + \max(0, y_{curr} - y_{next})$ 
        // Calculate height after erosion
         $eroded\_height \leftarrow y_{curr} - (exposure \times c_e)$ 
        // Store result for the next state
         $Y^{next}[i] \leftarrow eroded\_height$ 
    // Update the current state simultaneously for the next iteration
     $Y^{current} \leftarrow Y^{next}$ 
 $Y^{final} \leftarrow Y^{current}$ 
return  $Y^{final}$ 

```

Parameter tuning through genetic algorithm. To ensure our physical model accurately reflects real bamboo slip characteristics, we developed a parameter optimization strategy driven by comparison between real and synthetic data distributions. The parameters in our physical model have physical interpretations that can be loosely connected to real-world properties [29, 30], such as bamboo fiber structure, material characteristics, and environmental degradation factors. While these parameters are not direct measurements of physical properties, they collectively contribute to modeling the complex fracture and corrosion processes observed in archaeological bamboo slip fragments [31, 32]. We collect curves from real fragmented bamboo slips, which represent the final state after both fracture and corrosion processes, characterized by their y^{final} coordinates along the fragment boundary. These curves undergo dimensionality reduction using t-SNE to create a reference distribution. Concurrently, we generate synthetic data using randomly initialized parameters and apply the same dimensionality reduction process. Our synthetic fragments undergo the complete physical simulation pipeline: starting with initial fracture generation to obtain y^{init} coordinates, followed by corrosion simulation to produce final degraded profiles with y^{final} coordinates that can be directly compared with real archaeological data. The optimization objective is to minimize the distributional difference between real and synthetic data, quantified by the

silhouette coefficient in Eq. (9). Our genetic algorithm iteratively refines the parameter set by maintaining a population of parameter configurations $\mathcal{P} = \{p_1, \dots, p_n\}$, where each p_i contains parameters for both fracture and corrosion models [27]. The visualization in Extended Figure E3 confirms that optimizing these physical parameters directly improves the model’s ability to identify correct fragment matches. The fitness function follows Eq. (10), which combines the silhouette score with a diversity term for balancing between distribution matching and population diversity. New parameter sets are generated through Eq. (11), where p_1 and p_2 are selected based on fitness scores, β is a random crossover weight, and ϵ represents mutation noise. The effectiveness of our parameter optimization approach is visually demonstrated in Extended Figure E3, where the distribution of synthetic data (blue points) progressively aligns with the real data distribution (green points) through iterations. The corresponding heatmap of matching quality shows stronger correlation patterns as the silhouette coefficient improves, validating our parameter optimization strategy [33]. This optimization ensures that our synthetic fragments accurately capture both the initial fracture characteristics and the subsequent degradation patterns, enabling the deep learning model to learn features grounded in actual physical processes.

WisePanda architecture

Input representation. Each input to WisePanda consists of bamboo slip fragment curve data encoded as 64-dimensional vectors. We extract these vectors through a straightforward sampling process: after detecting and isolating the fracture curve from each fragment image, we normalize the curve length and sample 64 equidistant points along the curve from left to right. The resulting sequence of 64 heights capture the distinctive geometric profile of each fracture curve while providing a standardized input format for the following neural network. This fixed-dimensional representation ensures consistent processing across fragments of different sizes and complexities while preserving the key morphological features that distinguish potential matching pairs.

Network design. The core of WisePanda’s architecture is a TripletNet-based deep learning network designed for similarity learning between fragment pairs. The network processes triplets of fragments: an anchor sample c_a (a fragmented curve), a positive sample c_p (a matching fragment), and a negative sample c_n (a non-matching fragment). Each branch of the TripletNet shares identical weights to maintain consistency in feature extraction, and comprises three key components: a local feature encoder that captures fine-grained geometric patterns using 1D convolutions, a self-attention mechanism for modeling long-range dependencies in the curve, and a cross-fragment attention module that learns to align corresponding patterns between potential matches [34]. This network is trained using a triplet loss function:

$$\mathcal{L}_{\text{triplet}} = \max(0, d(c_a, c_p) - d(c_a, c_n) + m), \quad (12)$$

where $d(c_a, c_p)$ represents the distance between anchor and positive samples in the learned feature space, $d(c_a, c_n)$ is the distance between anchor and negative samples, and m is the margin parameter that enforces a minimum separation between matching and non-matching pairs. This loss function effectively guides the network to learn a feature space where matching fragments are pulled closer together while pushing non-matching fragments apart [34]. The network outputs a similarity score within [0,1] for each potential match, where values closer to 1 indicate higher match probability.

Ranking mechanism. To facilitate archaeological fragment rejoicing, WisePanda implements a ranking-based matching system that provides archaeologists with a manageable set of potential matches [35, 36]. For each input fragment, the network computes similarity scores with all candidate fragments in the dataset. These scores are then sorted to generate a ranked list of the Top- k most probable matches, typically $k = 50$, significantly reducing the search space from thousands of possibilities to a practical subset for further manual verification. The ranking process is guided by the predicted similarity metric from our TripletNet network architecture. When a new fragment is presented, the system: 1) extracts the fracture curve and generates the 64-dimensional feature vector; 2) computes pair-wise similarity scores with existing fragments through the trained network; 3) ranks all candidates based on their similarity scores; and 4) returns the Top- k candidates with the highest matching probabilities. This approach transforms WisePanda from a theoretical matching model into a practical archaeological tool that effectively guides the rejoicing process while involving human expertise in the final verification steps.

Training details

WisePanda was trained using two NVIDIA RTX 4090 GPUs with a batch size of 100. The network was trained using the Adam optimizer with an initial learning rate of 1e-3 and a one-cycle learning rate

scheduler [37]. We implemented a combined loss function that extends the concept of triplet loss to a more direct optimization objective:

$$\mathcal{L} = E[(d_{\text{pos}} - 0)^2 + (d_{\text{neg}} - 1)^2] \quad (13)$$

where d_{pos} represents the distance between matching pairs (which we want to push toward 0) and d_{neg} represents the distance between non-matching pairs (which we want to push toward 1). This squared error formulation provides smoother gradients compared to the standard triplet loss while maintaining the same conceptual goal of maximizing the separation between matching and non-matching fragments. We utilized PReLU activation functions for non-linear transformations, which adapt to the data by learning the optimal negative slope parameter rather than using a fixed value as in standard ReLU.

The input fragmented curves were normalized by rescaling both x and y^{final} coordinates to fall within the $[0, 1]$ range before being sampled and fed into the network. The training process ran for 150 epochs to ensure convergence, with model parameters being updated approximately 100,000 times. To ensure reproducibility, model training was conducted with fixed random seeds. For inference, we employ an efficient ranking strategy where each fragment is compared with the entire database using the trained similarity metric. The matches are then sorted by matching probabilities to generate the Top- k recommendations for archaeological verification.

Baseline methods

To evaluate WisePanda’s effectiveness, we compare it against three traditional methods widely used in fragment matching and curve alignment tasks.

Traditional curve matching baseline. The traditional curve matching approach employs dynamic time warping (DTW) [38] to measure the geometric similarity between fragmented curves. This method directly compares the normalized (x, y^{final}) coordinate sequences of two fragments, characterizing their alignment cost as a measure of potential matching.

Scale Invariant Signature (SIS). SIS [39] transforms fragmented curves into scale-invariant representations by computing local curvature features along the fragmented curve. This transformation generates signatures that are independent of fragment size and orientation, enabling comparison through simple distance metrics.

Fast Matching Method (FMM). FMM [40] accelerates the matching process by decomposing fragmented curves into key geometric features and employing a hierarchical matching strategy. It first identifies potential matches using coarse features, then refines the results using more detailed geometric information.

Evaluation metrics

Top- k accuracy. To assess WisePanda’s performance in bamboo slip rejoicing, we employ the Top- k accuracy metric, which measures the percentage of correctly identified matches within the k highest-ranked suggestions. For each query fragment, a match is considered successful if the true matching fragment appears among the Top- k candidates proposed by the model. We report results for $k \in 1, 5, 10, 20, 50, 100$, providing a comprehensive view of the model’s ranking capability at different thresholds of consideration. This metric directly reflects the system’s practical utility in archaeological workflows, where experts typically examine a limited number of the most probable matches.

Future work

While our current work demonstrates the effectiveness of physics-driven deep learning for fragmented bamboo slip rejoicing, several promising directions remain for future exploration. First, our present approach primarily addresses transverse fractures, which account for approximately 70% of bamboo slip breakages. Extending WisePanda to handle longitudinal fractures (Figure E1)—which occur along the fiber direction and produce distinctly different breakage patterns—represents an important next step. For these longitudinal fractures, where the straight physical breakage features are less distinctive, focusing on text continuity and character alignment would be more effective than attempting to model the physical fracture properties. This would entail developing complementary methods that leverage optical character recognition and textual coherence to guide the matching process when geometric features alone are insufficient.

Beyond bamboo slips, the methodology established in this work opens avenues for broader applications in archaeological artifact restoration. Particularly promising is the extension to pottery and ceramic fragment matching, where material properties and degradation processes differ significantly from bamboo. Ceramics exhibit distinctive fracture mechanics governed by brittleness, firing conditions, and microstructure, which could be modeled using adaptations of our physics-driven approach. The translational potential of our framework to these domains could significantly advance digital archaeology tools across diverse cultural heritage contexts.

From a methodological perspective, integrating expert knowledge more deeply into the learning process presents another promising direction. Future systems might implement active learning paradigms where archaeologist feedback continuously refines the rejoicing models, creating a virtuous cycle of human-AI collaboration. Additionally, exploring self-supervised or few-shot learning approaches could further reduce the dependence on synthetic data, potentially capturing even more nuanced aspects of artifact deterioration that are challenging to model explicitly.

These future directions share a common goal: bridging the gap between physical understanding and computational methods in cultural heritage preservation. By continuing to combine domain-specific physical principles with advanced machine learning techniques, we envision creating increasingly powerful tools that amplify rather than replace human expertise in the vital work of preserving and understanding our shared cultural past.

Competing interests

The authors declare no competing interests.

Author contributions

J.Z. and Y.X. conceived the study. H.L., X.W., J.L.L. and J.L. provided archaeological data and domain knowledge. J.Z. and Z.Z performed experiments, analyzed data, wrote the code and manuscript. J.Z., Q.T. and J.S. contributed to the literature review and software implementation. G.X. participated in methodology discussions and manuscript organization. B.D. participated in methodology discussions, data analysis, and manuscript editing. Y.X. participated in methodology discussions, experimental design, manuscript organization and revision.

Data availability

The datasets used for training and evaluation are publicly available at GitHub <https://github.com/zhujinchi/wisepanda> under Apache License 2.0. The repository includes fragment matching datasets: Bamboo236 (118 paired bamboo fragments), Wood670 (335 paired wooden fragments), and their extended versions Bamboo1350 and Wood3833 with additional 1,114 and 3,163 interference fragments respectively. Due to confidentiality requirements for unpublished archaeological materials, the original high-resolution fragment images cannot be released. Instead, the repository provides fracture curves, cropped edge images, and synthetic training data generated by our physics-driven model. The complete data processing workflow for extracting fracture curves from archaeological fragments and generating synthetic paired data is included in the codebase. All morphological features and curve data necessary for reproducing the fragment matching experiments are available in the repository.

Code availability

The source code for WisePanda framework is publicly available on GitHub at <https://github.com/zhujinchi/wisepanda> under Apache License 2.0. The repository contains the implementation of our physics-driven deep learning model for bamboo slip fragment rejoicing, including the fracture modeling algorithms, corrosion simulation processes, neural network architecture, and user interface. Users can install the required dependencies by following the installation instructions provided in the repository. The codebase is designed to be accessible to researchers, featuring an intuitive graphical user interface for fragment selection, comparison, and verification with integrated AI assistance. Additionally, we have established a project website at <https://wisepanda.tech/> that provides demonstration videos, visual examples, and detailed documentation of the fragment rejoicing and visualization tools discussed in this paper. This open-source release aims to facilitate further research in digital archaeology and enable the application of our methods to other types of archaeological fragment rejoicing challenges.

References - Methods

1. Zhou, J. *et al.* Identifying designs from incomplete, fragmented cultural heritage objects by curve-pattern matching. *J. Electron. Imaging* **26**, 011022–011022 (2017).
2. Zhang, Y., Li, K., Chen, X., Zhang, S. & Geng, G. A multi feature fusion method for reassembly of 3D cultural heritage artifacts. *J. Cult. Herit.* **33**, 191–200 (2018).
3. Das, B. R., Maringanti, H. B. & Dash, N. S. in *Digital. Cult. Technol.* 92–97 (Routledge, 2022).
4. Tsatsanashvili, A. Artificial Intelligence In The Protection Of Intangible Cultural Heritage. *Eur. J. Turk. Stud.* **12**, 163–178 (2024).
5. Tsesmelis, T. *et al.* Re-assembling the past: The RePAIR dataset and benchmark for real world 2D and 3D puzzle solving. *Proc. of Adv. Neural Inf. Process. Syst.* **37**, 30076–30105 (2024).
6. Assael, Y. *et al.* Restoring and attributing ancient texts using deep neural networks. *Nature* **603**, 280–283 (2022).
7. Zheng, Y., Li, X. & Weng, Y. Reunion helper: an edge matcher for sibling fragment identification of the Dunhuang manuscript. *Heritage Science* **12**, 52 (2024).
8. Zhang, Y. *et al.* LLMCO4MR: LLMs-Aided Neural Combinatorial Optimization for Ancient Manuscript Restoration from Fragments with Case Studies on Dunhuang. in *Proc. Eur. Conf. on Comput. Vis.* (2024), 253–269.
9. Li, B. *et al.* Oracle Bone Inscriptions Multi-modal Dataset. *arXiv preprint arXiv:2407.03900* (2024).
10. Staack, T. Bindings of Ancient Chinese Bamboo and Wood Scrolls. *Stud. Manusc. Cult.*, 39 (2023).
11. Chang, C.-s. *Written on Bamboo and Silk: The Beginning of Chinese Books and Inscriptions* 1964.
12. Needham, J. *Science and Civilisation in China: Volume 5, Chemistry and Chemical Technology, Part 5, Spagyrical Discovery and Invention: Physiological Alchemy* (Cambridge University Press, 1983).
13. Luo, Z. & Wang, G. *Liusha Zhuijian (Liu sha zhui jian)* (Shangyu Luo shi yin xing, 1914).
14. Ghaith, K. & Hutson, J. A qualitative study on the integration of artificial intelligence in cultural heritage conservation. *Metaverse* **5** (2024).
15. Aldakheel, F., Elsayed, E. S., Heider, Y. & Weeger, O. Physics-based machine learning for computational fracture mechanics. *Mach. Learn. Comput. Sci. Eng.* **1**, 18 (2025).
16. LeSar, R. *Introduction to computational materials science: fundamentals to applications* (Cambridge University Press, 2013).
17. Ramful, R. Investigation of the transverse fracture mechanisms of bamboo by the finite element method. *J. Mater. Sci.* **57**, 6233–6248 (2022).
18. Dixon, P. G. & Gibson, L. J. The structure and mechanics of Moso bamboo material. *J. R. Soc. Interface* **11**, 20140321 (2014).
19. Chen, M. *et al.* Fracture modes of bamboo fiber bundles in three-point bending. *Cellulose* **26**, 8101–8108 (2019).
20. Ramful, R. & Sakuma, A. Investigation of the effect of inhomogeneous material on the fracture mechanisms of bamboo by finite element method. *Materials* **13**, 5039 (2020).
21. Huang, Y. & Fei, B. Comparison of the mechanical characteristics of fibers and cell walls from moso bamboo and wood. *BioResources* **12**, 8230–8239 (2017).
22. Gdoutos, E. E. *Fracture mechanics: an introduction* (Springer Nature, 2020).
23. (US), F. P. L. *Wood handbook: wood as an engineering material* **72** (The Laboratory, 1987).
24. Blanchette, R. A., Haight, J. E., Koestler, R. J., Hatchfield, P. B. & Arnold, D. Assessment of deterioration in archaeological wood from ancient Egypt. *J. Am. Inst. Conserv.* **33**, 55–70 (1994).
25. Liu, L. *et al.* Cracks Repairing and Resistance to Water Penetration Properties of Microbial Self-Healing Cement. *Eng. Life. Sci.* **25**, e70010 (2025).
26. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9** (2008).
27. Goldberg, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, Reading, MA, 1989).

28. Anderson, T. L. & Anderson, T. L. *Fracture mechanics: fundamentals and applications* (CRC press, 2005).
29. Hussain, A., Sakhaei, A. H. & Shafiee, M. Machine learning-based constitutive modelling for material non-linearity: A review. *Mech. Adv. Mater. Struct.*, 1–19 (2024).
30. Kutz, J. N. Machine learning for parameter estimation. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2300990120 (2023).
31. Brynjarsdottir, J. & O'Hagan, A. Learning about physical parameters: The importance of model discrepancy. *Inverse Probl.* **30**, 114007 (2014).
32. Matthiesen, H. Detecting and quantifying ongoing decay of organic archaeological remains: A discussion of different approaches. *Quaternary International* **368**, 43–50 (2015).
33. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
34. Hoffer, E. & Ailon, N. Deep metric learning using triplet network. in *Proc. Int. Worksh. on Simil.-Based Pattern Recognit.* (2015), 84–92.
35. Funkhouser, T. *et al.* Learning how to match fresco fragments. *ACM J. Comput. Cult. Herit.* **4**, 1–13 (2011).
36. Di Angelo, L., Di Stefano, P. & Guardiani, E. A review of computer-based methods for classification and reconstruction of 3D high-density scanned archaeological pottery. *J. Cult. Herit.* **56**, 10–24 (2022).
37. Smith, L. N. & Topin, N. Super-convergence: Very fast training of neural networks using large learning rates. in *AI ML Multi-Domain Oper. Appl.* **11006** (2019), 369–386.
38. Berndt, D. J. & Clifford, J. Using dynamic time warping to find patterns in time series. in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* (1994), 359–370.
39. Calabi, E., Olver, P. J., Shakiban, C., Tannenbaum, A. & Haker, S. Differential and numerically invariant signature curves applied to object recognition. *Int. J. Comput. Vis.* **26**, 107–135 (1998).
40. Frenkel, M. & Basri, R. Curve matching using the fast marching method. in *Proc. Int. Worksh. on Energy Minimiz. Methods in Comput. Vis. and Pattern Recognit.* (2003), 35–51.

Extended Figure



Figure E1: **Typical fracture patterns of ancient bamboo slips.** (a), The upper row shows bamboo samples with transverse fractures, the most common fracture mode (approximately 70%), characterized by irregular wavy breakage patterns. Most transverse fracture curves contain no text, with only occasional fragments showing partial characters, making morphological features the primary basis for rejoining. (b), The lower row exhibits longitudinal fractures, characterized by relatively smooth fracture curves along the fiber direction. These fragments typically preserve complete text, making textual continuity the primary criterion for matching. This diversity of fracture patterns and text preservation highlights the complexity of ancient bamboo slip rejoining.

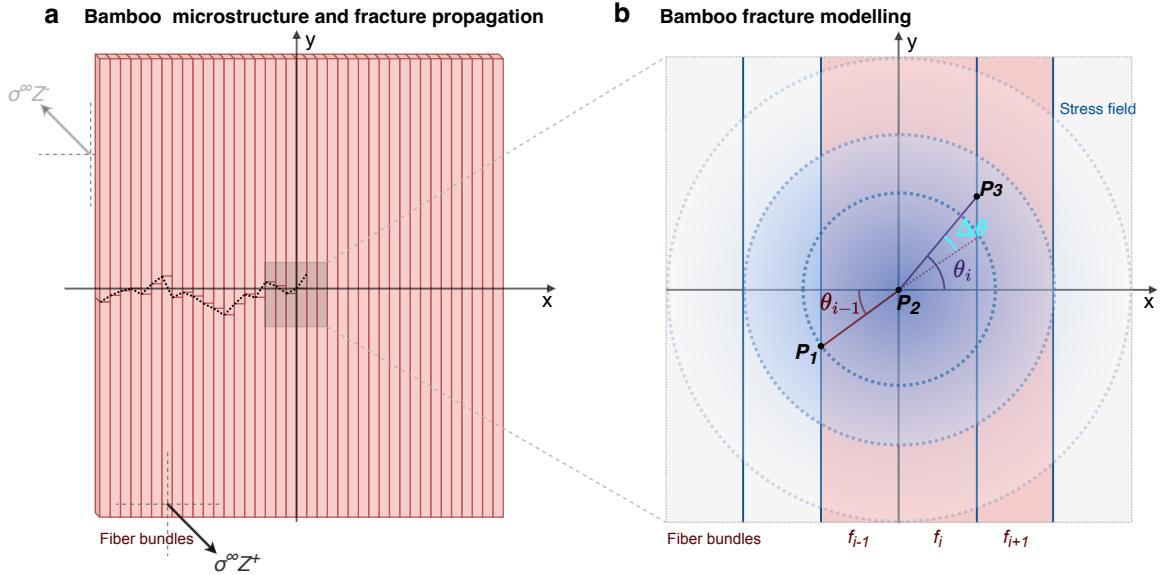


Figure E2: **Physics-driven model of bamboo fracture propagation.** **a**, Bamboo microstructure showing vertical fiber bundles (red blocks) and a possible fracture path (composed of dotted black segments) traversing across fibers. The black arrows indicate shear forces ($\sigma^\infty Z^+$) inducing Mode III fracture. **b**, Bamboo fracture modeling with stress field propagation between fibers. Points P_1 , P_2 , and P_3 represent left fracture height on consecutive fibers f_{i-1} , f_i and f_{i+1} , with angles θ_{i-1} and θ_i showing fracture directions of fibers f_{i-1} and f_i . Blue dotted circles illustrate the stress field emanating from point P_2 , determining the probability distribution of potential fracture paths.

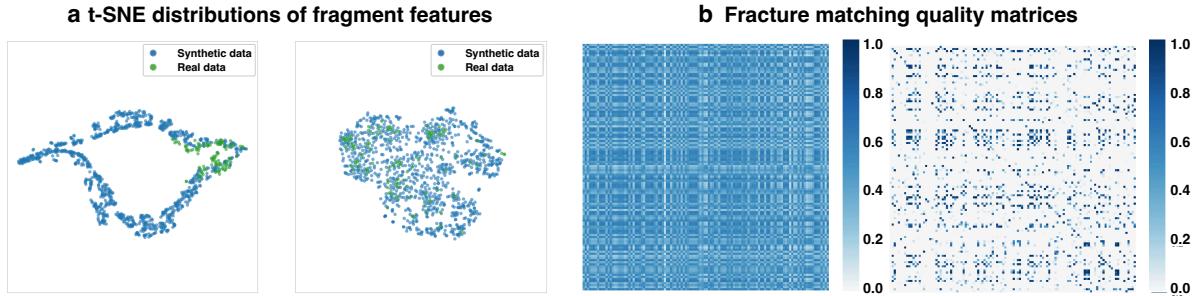


Figure E3: **Data distribution analysis and matching performance.** **a**, t-SNE distributions of fragment curve features showing the relationship between synthetic data (blue) and real bamboo fragments (green) across two parameter configurations. Convergence indicates effective parameter optimization in model. **b**, Fracture matching quality matrices for 118 real fragment pairs. Deeper blue indicates higher similarity, with diagonal elements representing correct matches. The lower-right matrix shows good discriminative performance with clear contrast between prominent diagonal (blue) and off-diagonal elements (white), corresponding to the most effective parameter setting (right of **a**).

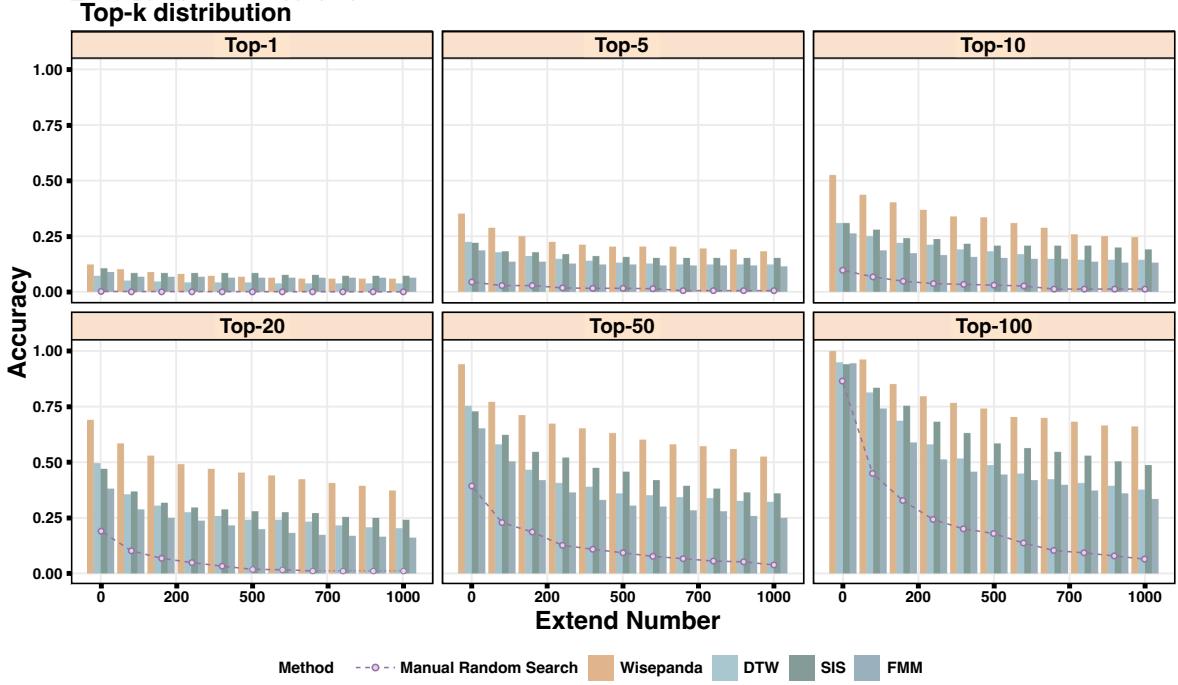


Figure E4: **Robustness analysis of different fragment matching algorithms with increasing interference fragments.** Performance comparison across six accuracy levels (Top-1 to Top-100) as candidate pool expands from 0 to 1000 fragments. WisePanda (orange bars) maintains relatively stable performance despite increasing interference, while Manual Random Search (dotted purple line) shows rapid accuracy decline. Other methods (DTW, SIS and FMM) show lower performance compared to WisePanda across different interference levels. At maximum interference, WisePanda maintains 50-65% accuracy at Top-50 and Top-100 levels, while competing methods achieve below 36% at Top-50, demonstrating the physics-driven approach's advantage in real archaeological scenarios with thousands of potential matches.

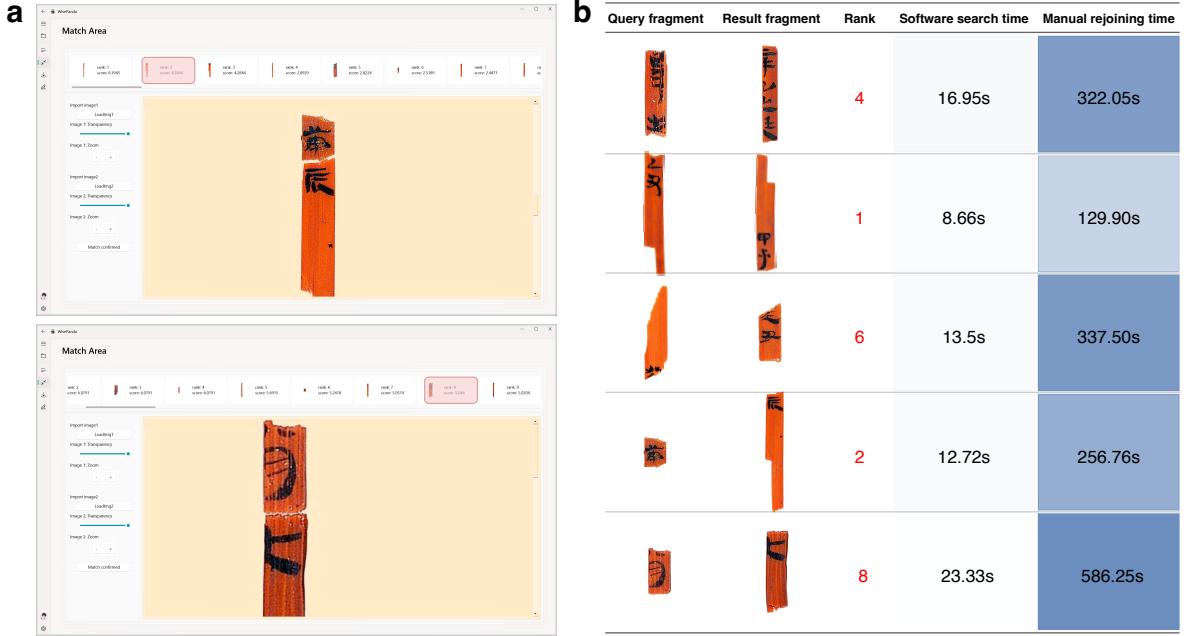


Figure E5: Practical application of WisePanda. **a**, WisePanda software interface showing two successful matching cases: target bamboo slip fragments with their corresponding matching results, demonstrating the system’s verification capabilities. **b**, Efficiency comparison between WisePanda-assisted and traditional manual matching across five test cases. The table records query (top fragment) and result fragments, system recommendation rank, and time measurements for both methods. WisePanda reduces average matching time from 326.49 to 15.03 seconds—approximately a 20-fold efficiency improvement. These results validate the physics-driven approach’s practical value for large-scale bamboo slip fragment rejoining.