

Article

FaSDiNet: An End-to-End Efficient Network for Oracle Bone Fragment Rejoining Driven by Full-Image Candidate Screening

Yang Hong ^{1,†}, Rui Tian ^{2,†}, Yuqing Zhou ³, Yunlong Feng ³ and Chuanming Song ^{2,4,*} 

¹ School of Chinese Language and Literature, Liaoning Normal University, Dalian 116081, China; hongy@lnnu.edu.cn

² School of Information Engineering, Dalian University, Dalian 116622, China; tianrui@s.dlu.edu.cn

³ School of Computer Science and Artificial Intelligence, Liaoning Normal University, Dalian 116081, China; 13840377125@163.com (Y.Z.); 15141333590@163.com (Y.F.)

⁴ Provincial Key Laboratory for Computer Information Processing Technology, Soochow University, Suzhou 215006, China

* Correspondence: songchuanming@dlu.edu.cn

† These authors contributed equally to this work.

Abstract

Oracle bone rejoining is central to oracle bone studies and digital cultural heritage curation, but many computer-assisted methods still rely on manually selected contour curves, candidate fracture edges, or local preprocessing. We introduce the Fast Siamese Diffusion Network for Oracle Bone Fragment Rejoining (FaSDiNet), a FasterNet-based Siamese framework with a latent-diffusion-based data augmentation module for full-fragment-image candidate screening and assisted rejoining of oracle bone fragments. FaSDiNet combines latent diffusion augmentation, a FasterNet-based Siamese feature encoder, and joint optimization of contrastive loss and binary cross-entropy loss to improve pairwise matching under limited training data. On the self-built OBF dataset, FaSDiNet achieves Top-1, Top-5, Top-10, and Top-15 recall values of 23.68%, 60.53%, 76.32%, and 84.21%, respectively. Among the evaluated OBF methods, its strongest comparative performance appears at broader candidate ranges, where it obtains the best Top-5, Top-10, and Top-15 recall values. On the public COBD dataset, FaSDiNet achieves Top-10 and Top-15 recall values of 93.8% and 96.9%, respectively, showing competitive candidate-retention performance at larger candidate ranges. These quantitative results indicate that FaSDiNet is especially effective for preserving truly rejoinable fragments within manageable ranked candidate lists, supporting efficient expert-assisted screening and offering a practical visual-computing workflow for oracle bone collation.

Keywords: oracle bone rejoining; digital cultural heritage; Siamese network; FasterNet; latent diffusion augmentation; candidate screening; visual retrieval



Academic Editor: Ion Sandu

Received: 23 May 2026

Revised: 1 July 2026

Accepted: 2 July 2026

Published: 20 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Oracle bone inscriptions are among the earliest mature writing systems discovered in China and remain important to historical studies, paleography, and digital cultural heritage. Yet many oracle bones survive only as scattered fragments because of long-term burial, weathering, excavation damage, and later dispersal. Rejoining these fragments is therefore more than a physical matching problem: it directly affects character restoration, textual reconstruction, and the recovery of historical information [1]. As shown in Figure 1, the value of oracle bone rejoining lies in restoring both fragment continuity and broken

textual relations. Figure 1a shows the separated fragment rubbings before rejoining, where each rubbing preserves only partial visual and textual evidence. After rejoining, Figure 1b presents the rejoined rubbing image, in which the spatial relationship between the fragments and the continuity of the inscriptions become clearer. Figure 1c further shows the traced rejoined image, where the reconstructed character structure provides a more reliable basis for later interpretation.

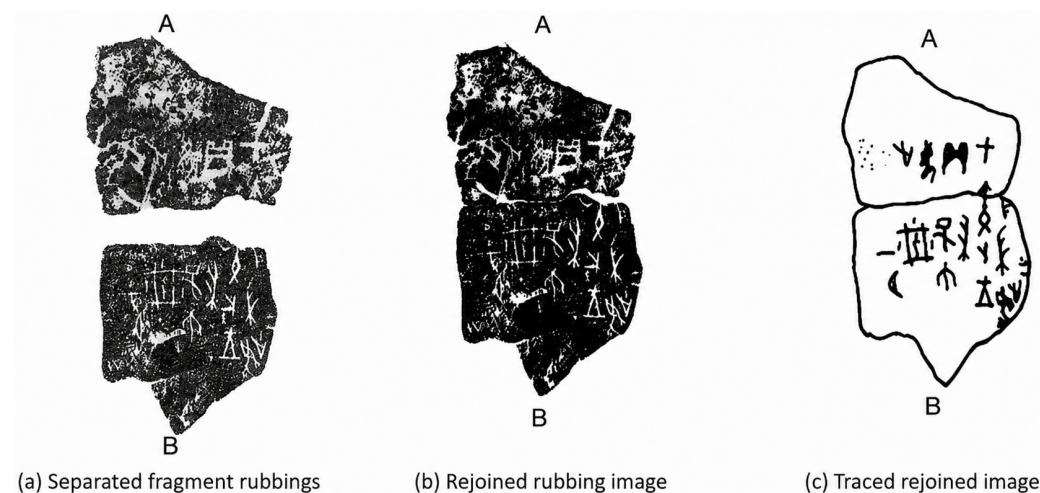


Figure 1. Example of oracle bone fragment rejoining. (a) Separated oracle bone fragment rubbings before rejoining, where each fragment preserves only partial visual and inscriptional evidence. (b) Rejoined rubbing image, showing the restored spatial relationship between the fragments and the improved continuity of the inscriptions. (c) Traced rejoined image, in which the reconstructed character structure and fragment relationship are further clarified for subsequent scholarly interpretation.

Traditional oracle bone rejoining relies heavily on expert knowledge and manual comparison, which makes the process labor-intensive, inefficient, and difficult to scale to large digital curation projects [1,2]. Recent computer-assisted studies have introduced image processing, pattern recognition, and data-driven analysis into this task, moving oracle bone rejoining beyond purely experience-based workflows [3–6]. However, practical limitations remain clear. Verified rejoin samples are still scarce [7], and many existing methods continue to depend on contour curves, candidate fracture edges, or manually prepared local regions [8–10]. Such designs can be effective in constrained settings, but they do not fully exploit the joint evidence carried by boundary shape, internal texture, and engraved traces. This weakens robustness when fragment boundaries are worn, incomplete, or visually complex, and it also limits the practical value of these methods for large-scale candidate screening in digital cultural heritage workflows.

To address these issues, this article presents FaSDiNet, a FasterNet-based Siamese framework with latent diffusion augmentation for full-fragment-image candidate screening and assisted rejoining of oracle bone fragments. Instead of requiring manually selected candidate edges, FaSDiNet works directly on full fragment images and combines latent diffusion augmentation, a FasterNet-based Siamese feature encoder, and joint optimization of contrastive loss and binary cross-entropy loss. The goal is not to replace expert judgment with a fully automatic final decision, but to preserve truly rejoinable fragments near the top of a ranked candidate list and thereby reduce manual review cost. In this way, FaSDiNet frames oracle bone rejoining as a practical visual-computing workflow for expert-assisted screening under limited training data.

The methodological contribution of FaSDiNet lies in its task-oriented coordinated design for full-fragment-image candidate screening in oracle bone rejoining, a specific cultural-heritage retrieval problem. This design is motivated by the practical characteristics

of oracle bone collation, where verified rejoining pairs are scarce, fracture boundaries are often worn or incomplete, and candidate search requires preserving plausible matches for expert verification. First, the proposed framework directly processes complete fragment images and jointly exploits fracture-boundary shape, internal texture, engraved strokes, and global fragment appearance without requiring manually selected candidate edges. Second, relation-preserving latent diffusion augmentation is introduced before Siamese training to enrich the visual variation of training fragments while retaining their original rejoining-group relationships. Third, the shared FasterNet-based Siamese encoder learns comparable full-fragment representations in a unified feature space, so that boundary-level and interior visual cues can be used jointly for candidate ranking. Fourth, contrastive distance supervision and binary classification supervision are jointly optimized to connect feature-space organization with the matching probabilities used for Top-K retrieval. Through this coordinated design, data augmentation, full-image representation learning, and similarity-based candidate screening form a unified workflow for expert-assisted oracle bone rejoining. Experiments on the self-built OBF dataset and the public COBD dataset demonstrate the effectiveness of the proposed framework under full-image screening and public benchmark settings.

2. Related Work

2.1. Oracle Bone Rejoining

Oracle bone rejoining is an important task in oracle bone studies and digital cultural heritage curation. Traditional rejoining work mainly relies on expert experience and rule-based judgment. Researchers usually examine fracture continuity, curvature variation, local shape complementarity, inscription continuity, and contextual clues from incomplete characters to determine whether separated fragments can be rejoined [1,2]. Although such manual analysis remains indispensable for scholarly verification, it is labor-intensive and difficult to scale to large fragment collections.

With the development of computer technology, image processing, pattern recognition, and data-driven analysis have gradually been introduced into oracle bone rejoining [3–6]. Existing computational studies mainly follow two technical routes. One route focuses on contour curves or fracture boundaries and evaluates rejoinability through curve matching, curve fitting, correlation analysis, or local contour comparison [8]. These methods are effective when fragment boundaries are relatively clear and well preserved, but their performance can be affected by noise, weathering, missing edges, and irregular fracture morphology. Another route introduces deep learning into oracle bone rejoining and shifts the task from hand-crafted matching rules to learned visual representation and similarity modeling [7,9–11].

For example, Zhang et al. [9] proposed deep rejoining models that combine edge localization, target-area cropping, and convolutional neural networks for fragment matching. Zhang et al. [7] employed self-supervised learning and data augmentation to improve contour-based representation. Yuan et al. [10] further proposed SFF-Siam, a Siamese-network-based method that strengthens contour-curve matching through similarity feature fusion and a forward feedback network. These studies have promoted the automation of oracle bone rejoining and demonstrated the potential of computer-assisted analysis.

However, several limitations remain. Verified rejoin samples are still scarce, which constrains stable training and generalization under small-sample conditions [1,7]. In addition, many existing methods still depend on annotated contour curves, candidate fracture edges, or cropped local regions [8–10]. As a result, they tend to emphasize local boundary compatibility while making limited use of complementary cues from internal

texture, engraved traces, and overall fragment appearance. This restricts their applicability to full-fragment candidate screening in large-scale digital heritage collections.

2.2. Deep Visual Networks for Oracle Bone Fragment Representation

Deep visual networks have become an important tool for learning discriminative representations from complex images. Compared with hand-crafted descriptors, convolutional neural networks and their variants can learn hierarchical features from raw visual data, including low-level edge patterns, mid-level texture structures, and high-level semantic representations. This ability is particularly relevant to oracle bone fragment analysis, because reliable rejoining cues may appear not only along fracture boundaries but also in internal texture continuity, engraved strokes, surface degradation patterns, and global fragment shape.

In oracle bone rejoining, deep learning methods have been used to improve feature extraction and fragment matching [7,9–11]. Nevertheless, some existing deep models still require manually selected edges, target-area cropping, contour extraction, or manually prepared local regions before similarity estimation [7,9,10]. Such preprocessing can reduce the complexity of local matching, but it may also discard useful full-image evidence. For practical candidate screening, a network should be able to extract robust representations directly from full fragment images while preserving both boundary-level and interior visual cues.

Efficiency is another important requirement for oracle bone fragment screening. In practical rejoining scenarios, the model is often expected to compare one query fragment with a large number of candidates. Therefore, the visual encoder should not only have sufficient representation capability, but also support repeated pairwise comparison with acceptable computational cost. FasterNet provides a favorable balance between feature representation and computational efficiency by reducing redundant computation and memory access through partial convolution [12]. This makes it suitable as a lightweight backbone for large-scale candidate screening, where high-recall ranking is more important than one-off image classification.

Data scarcity further increases the difficulty of training deep networks for oracle bone rejoining. Unlike common vision datasets, verified rejoin pairs are limited and usually require expert confirmation [1,7]. Therefore, data augmentation is important for improving representation robustness. Conventional transformations such as rotation, scaling, flipping, and cropping can simulate geometric variations, while diffusion-based augmentation can enrich local texture and surface-detail variations while better preserving the underlying image distribution [13]. These augmentation strategies provide additional support for training visual encoders under small-sample conditions.

Recent developments in visual representation learning provide a broader methodological context for image retrieval under limited annotations. Residual networks improve the optimization of deep convolutional encoders [14], whereas Vision Transformers model long-range visual dependencies through self-attention [15]. Masked autoencoders further demonstrate that reconstruction-based self-supervised pretraining can learn scalable visual representations [16]. In vision-language pretraining, CLIP learns transferable representations from image–text pairs [17], while Tip-Adapter adapts such representations to few-shot tasks using limited labeled samples [18]. More recently, DINOv2 has shown that self-supervised pretraining can produce general-purpose visual features transferable across image-level and dense prediction tasks [19]. These developments are relevant to oracle bone fragment retrieval because verified rejoining pairs are scarce and candidate screening requires transferable full-fragment representations.

2.3. Contrastive Learning for Pairwise Similarity Modeling

Oracle bone rejoining is naturally a pairwise matching problem. The key question is not to assign each fragment to a fixed semantic category, but to determine whether two fragments can form a valid rejoining relation. This makes pairwise similarity modeling more suitable than ordinary multi-class classification. Siamese networks provide a classical framework for this type of problem. By using two weight-sharing branches, a Siamese network maps two input samples into a shared feature space and estimates their similarity through feature distance or binary decision objectives [20,21].

Contrastive learning provides an effective optimization mechanism for pairwise similarity modeling. Its basic idea is to reduce the distance between positive pairs while enlarging the distance between negative pairs in the embedding space [21,22]. This is highly consistent with the requirement of oracle bone rejoining: fragments that truly belong to the same rejoinable relation should be close in feature space, whereas unrelated fragments should be separated. Compared with a single classification objective, contrastive constraints can directly shape the geometry of the learned embedding space and improve the discriminability of pairwise representations. Self-supervised frameworks such as SimCLR extend this principle by maximizing agreement between independently augmented views of the same image while contrasting other images within a minibatch, thereby learning transferable representations without class labels [23].

Siamese and contrastive learning have already shown potential in oracle bone rejoining. S³-Net introduced self-supervised learning and contrastive constraints for contour-shape representation, demonstrating the feasibility of Siamese-style modeling in this task [7]. SFF-Siam further incorporated similarity feature fusion and a forward feedback mechanism into a Siamese framework, achieving promising performance in contour-based rejoining [10]. These studies show that pairwise similarity learning is well suited to distinguishing rejoinable and non-rejoinable fragment pairs.

However, most existing Siamese-based oracle bone rejoining methods still focus on contour-level or candidate-edge-level matching. For full-fragment candidate screening, the model needs to combine robust visual representation, efficient feature extraction, and discriminative pairwise similarity learning. This motivates the design of a full-image-based Siamese framework that can jointly use fracture-boundary cues, internal texture, engraved traces, and global fragment appearance for candidate ranking and expert-assisted oracle bone rejoining.

To further clarify the relationship between existing oracle bone rejoining methods and the proposed framework, Table 1 provides a qualitative comparison in terms of input representation, core matching mechanism, dependence on localized contour or edge information, visual evidence, and task orientation. Because these methods use different input forms and levels of preprocessing, this comparison focuses on their methodological characteristics rather than treating them as identical task settings.

As summarized in Table 1, existing methods have progressively advanced oracle bone rejoining from handcrafted curve comparison to learned contour representation and Siamese similarity modeling. However, most representative methods still operate on localized contour curves, candidate fracture edges, or cropped target regions. FaSDiNet differs in its task formulation: it directly processes full fragment images and ranks potentially rejoinable candidates for subsequent expert verification. This design retains complementary evidence from fragment boundaries, internal textures, engraved strokes, and overall appearance, which is particularly important when fracture edges are worn, incomplete, or visually ambiguous.

Table 1. Qualitative comparison of representative oracle bone rejoining methods.

Method	Input and Preprocessing	Core Mechanism	Visual Evidence and Task Orientation
CAP&CFDA [8]	Contour curves; contour input is required	Curve fitting and Pearson correlation analysis	Uses fracture-boundary geometry for contour-level matching
Deep Rejoining Model [9]	Localized and cropped target regions	Edge localization, target-area cropping, and CNN-based matching	Uses local image regions and requires edge localization before fragment matching
S ³ -Net [7]	Contour-shape sequences; contour input is required	Self-supervised representation and Siamese similarity learning	Learns contour-shape representations for candidate retrieval
SFF-Siam [10]	Candidate contour-curve pairs	Similarity feature fusion and forward feedback	Strengthens local contour similarity for pairwise rejoining decisions
FaSDiNet	Full fragment images; no manually selected candidate fracture edges	Latent diffusion augmentation, FasterNet-based Siamese encoding, and joint loss learning	Jointly uses boundary shape, internal texture, engraved strokes, and global appearance for ranked candidate screening

3. Materials and Methods

3.1. Problem Formulation and Overall Framework

In this study, oracle bone fragment rejoining is formulated as a full-fragment candidate-screening problem rather than as a one-shot automatic reassembly task. Given an input fragment pair (X_1, X_2) , the model aims to estimate a rejoining similarity score or matching probability. A positive pair corresponds to two fragments that have a verified rejoining relation, whereas a negative pair corresponds to two fragments without a valid rejoining relation. Under this formulation, the key objective is not only to determine whether two fragments match in isolation, but also to preserve truly rejoinable fragments near the top of a ranked candidate list when the search space is large and the available training data are limited.

This formulation is consistent with the practical workflow of oracle bone collation. In real scholarly practice, automatic methods are expected to reduce the search space and provide credible candidates for expert verification, rather than directly replace final expert judgment. Accordingly, for each query fragment, FaSDiNet expresses its output as pairwise matching scores and a ranked Top-K candidate list. Therefore, a suitable model should satisfy three requirements. First, it should improve training robustness under limited verified rejoining samples. Second, it should extract discriminative full-fragment representations from boundary shape, internal texture, engraved traces, and global fragment appearance. Third, it should learn a pairwise similarity structure that supports candidate ranking and binary matching decisions.

Based on these requirements, we design FaSDiNet, a FasterNet-based Siamese framework with latent diffusion augmentation. The overall framework is shown in Figures 2 and 3. FaSDiNet contains three tightly coupled components. First, a latent diffusion augmentation module expands the visual distribution of oracle bone fragments while preserving their original pairwise relationships. Second, a FasterNet-based Siamese encoder maps full fragment images into a shared embedding space and captures both boundary-level and interior visual cues. Third, a pairwise similarity measurement module jointly optimizes feature-space separation and binary matching decisions. Compared with those methods that mainly operate on manually selected contour curves, candidate

fracture edges, or manually prepared local regions [8–10], FaSDiNet works directly on full fragment images and provides a unified workflow for full-image candidate screening and expert-assisted oracle bone rejoining.

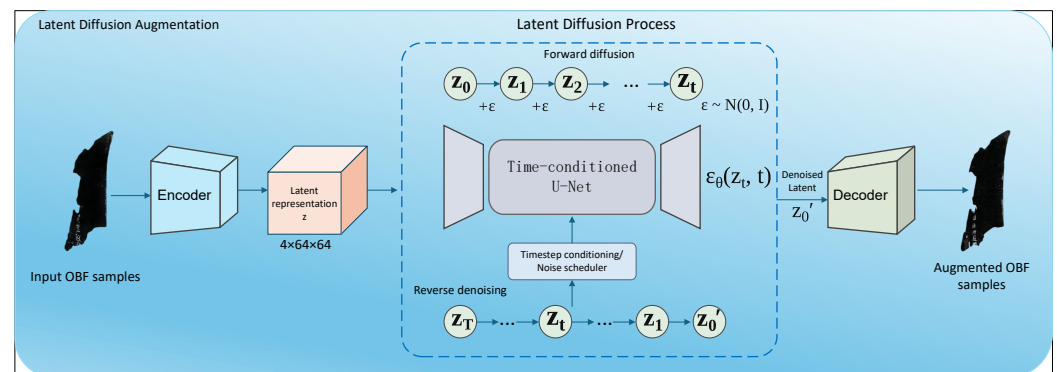


Figure 2. Architecture of the latent diffusion model.

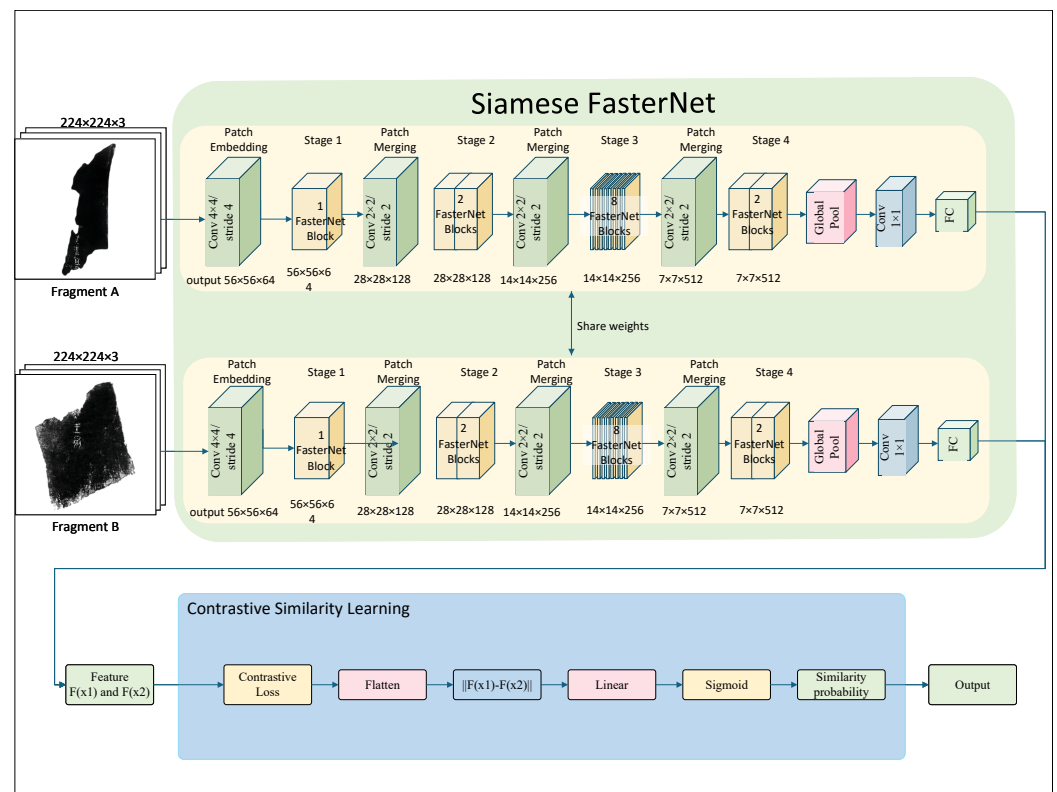


Figure 3. Overall framework of FaSDiNet.

The three modules form a sequentially coupled processing chain. The latent diffusion module first expands the visual variation of verified training fragments while retaining their rejoining-group labels. The shared FasterNet-based Siamese encoder then transforms the complete fragment images into comparable representations containing both boundary and interior evidence. Finally, the joint similarity-learning objective organizes these representations and converts them into matching probabilities for candidate ranking. This coupling enables FaSDiNet to connect limited-data augmentation directly with full-image representation learning and Top-K candidate screening.

3.2. Latent Diffusion Augmentation

A central difficulty in automatic oracle bone rejoining is the limited number of real rejoin samples available for supervised learning [1,7]. Unlike common vision benchmarks, oracle bone data are scarce not only in total quantity, but especially in verified positive pairs, which usually require long-term curation and expert confirmation. This makes the training process vulnerable to overfitting and weak generalization. To mitigate this problem, FaSDiNet introduces a latent diffusion model (LDM) to enrich the visual variation of oracle bone fragment images while preserving their main structural characteristics [13].

From a methodological perspective, the augmentation process consists of three stages: image encoding, latent-space diffusion, and image decoding, as illustrated in Figure 2. Given an oracle bone fragment image x , the encoder $\mathcal{E}(\cdot)$ maps it into a compact latent representation $z = \mathcal{E}(x)$. The diffusion process is then performed in the latent space, where noise is progressively injected and removed through a time-conditioned denoising network. Finally, the decoder $\mathcal{D}(\cdot)$ reconstructs the generated latent representation into an augmented image sample. Compared with direct generation in pixel space, latent-space diffusion reduces computational cost while maintaining the main visual structure of oracle bone fragments.

For latent diffusion augmentation, this study uses the pretrained Stable Diffusion v1.5 inpainting checkpoint, `stable-diffusion-v1-5/stable-diffusion-inpainting`, as a general image inpainting prior rather than training a latent diffusion model from scratch on the OBF dataset. In this study, the pretrained model is used only for controlled image augmentation under structural constraints, including the original fragment image, edge or foreground-boundary masks, controlled local repainting, and generated-sample filtering. The generated samples are therefore treated as structurally constrained training variants rather than as outputs based on domain-specific cultural-heritage knowledge.

The model is used in an image-to-image inpainting manner rather than as an unconditional text-to-image generator. For each oracle bone fragment, the input to the inpainting pipeline consists of the original fragment image, an edge or foreground-boundary mask, a text prompt, and a negative prompt. The mask constrains the generation process to local boundary-related regions, so that the model can introduce controlled visual variations while preserving the global fragment contour and the rejoining-related structure.

The theoretical diffusion process underlying the augmentation module is performed in latent space. Let $z_0 = \mathcal{E}(x)$ denote the latent representation of an input fragment image x . At diffusion step t , the noisy latent variable is obtained as

$$z_t = \sqrt{\bar{\alpha}_t} z_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \quad (1)$$

where $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$, and β_t controls the amount of noise introduced at step t . The time-conditioned U-Net denoising network [13,24] predicts the added noise from z_t , the timestep t , and the inpainting condition c . Its noise-prediction objective is

$$\mathcal{L}_{\text{LDM}} = \mathbb{E}_{x,t,\epsilon} \left[\|\epsilon - \epsilon_\theta(z_t, t, c)\|_2^2 \right], \quad (2)$$

where c represents the conditioning information provided by the original fragment image, the boundary mask, and the text prompt. Minimizing this objective enables the denoising network to recover a structurally plausible latent representation. For oracle bone augmentation, the boundary mask and controlled denoising strength restrict visual changes to local regions, thereby enriching surface texture and crack-related details while preserving the global fragment contour and the original rejoining relationship.

Figure 4 further illustrates the forward noising and reverse denoising processes in latent space. The forward process progressively degrades the latent representation by adding

noise, while the reverse process uses the denoising network and timestep information to recover the representation and synthesize new image samples. For oracle bone fragment images, this mechanism helps retain the global contour and main structural appearance while introducing controlled variations in local visual details.

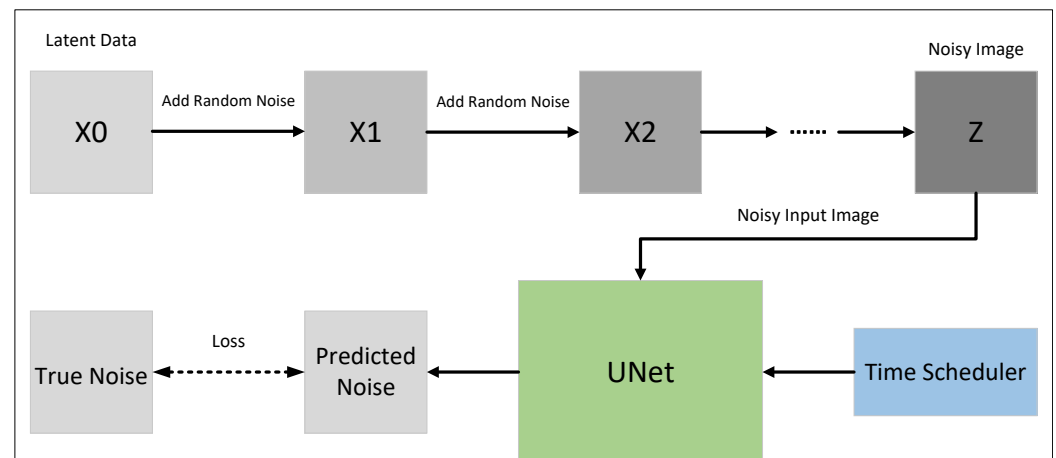


Figure 4. Forward diffusion and reverse denoising in latent space.

The resulting samples are further combined with geometric transformations such as rotation, scaling, flipping, and cropping to form the augmented dataset, thereby improving robustness to scale variation, pose disturbance, and local defects. The detailed augmentation procedure is summarized in Algorithm 1.

In the grouping stage, *index* denotes the local image index within each rejoining group. An index value of zero marks the beginning of a new group. *OPGroup* temporarily stores the images belonging to the current rejoining group, whereas *OP* stores all completed groups for subsequent augmentation.

The sampling method ζ , number of inference denoising steps λ , classifier-free guidance scale ρ , and random seed in Algorithm 1 are configurable generation parameters. For the implementation reported in this study, these parameters were selected from the corresponding choices or ranges and fixed throughout the augmentation process. Specifically, we used DDIM as the inference sampler, 50 denoising steps, a classifier-free guidance scale of 5.0, a denoising strength of 0.30, a random seed of 20250613, a boundary-mask width of 5 pixels, and two generated variants for each original training image. The official scheduler configuration of the pretrained checkpoint uses 1000 training timesteps and a scaled-linear beta schedule, with β increasing from 0.00085 to 0.012.

The pretrained inpainting checkpoint provides the upstream diffusion prior used in this study. The model documentation reports that the inpainting checkpoint was initialized from Stable Diffusion v1.2, trained for 595 k regular steps at a resolution of 512×512 , and then further trained for 440 k inpainting steps at the same resolution. In the OBF augmentation protocol, generated samples were accepted according to structural consistency and visual plausibility. Generated samples with excessive contour deformation were removed, while structurally valid samples were retained for training.

Figure 5 presents representative samples before and after augmentation. The displayed examples illustrate the local texture and surface-detail variations introduced by the augmentation process. Structural consistency was further assessed using foreground-mask IoU and manual visual inspection, as described below. The visual examples are therefore used to illustrate the augmentation results, while generated-sample acceptance is determined by the quantitative contour-consistency threshold and the visual quality-control criteria.

Algorithm 1: Latent diffusion-based data augmentation algorithm

Input: Original OBF dataset $\Gamma = \{(index, Image)\}$ ordered by rejoining group, where $index = 0$ denotes the first image of a new group

Output: Augmented oracle bone fragment image dataset RP

Initialize empty lists OP , $OPGroup$, and RP ;

Initialize the candidate sampling method set $\mathcal{S} = \{\text{DDIM}, \text{DPM2}, \text{DPM}++, \text{DPM fast}, \text{DPM adaptive}, \text{SDE}\}$, and configure $\xi \in \mathcal{S}$;

Configure the number of inference denoising steps $\lambda \in [1, 150]$;

Configure the classifier-free guidance scale $\rho \in [0, 30]$;

Configure the random seed $seed \in \mathbb{Z}^+$ for reproducible generation;

Set the total diffusion timestep T according to the scheduler of the pretrained model;

// Construct OP as a list of rejoining groups

foreach $(index, Image) \in \Gamma$ **do**

if $index \neq 0$ **then**

$OPGroup.append(Image)$;

else

if $OPGroup \neq \emptyset$ **then**

$OP.append(OPGroup)$;

$OPGroup \leftarrow [Image]$;

if $OPGroup \neq \emptyset$ **then**

$OP.append(OPGroup)$;

foreach $X \in OP$ **do**

$z_0 \leftarrow E(X)$;

for $t = T, \dots, 1$ **do**

if $t > 1$ **then**

 Sample noise $\epsilon_t \sim \mathcal{N}(0, I)$;

else

 Set $\epsilon_t = 0$;

$z_t \leftarrow \text{AddNoise}(z_0, \epsilon_t, t)$;

$new_image \leftarrow \text{diffusionmodel}(z_t, \xi, \lambda, \rho, seed)$;

$RP.append(new_image)$;

return RP ;

Latent diffusion augmentation was performed only after the OBF training, validation, and test subsets had been separated, and it was applied exclusively to the training subset. The training subset contained 566 training pair groups, corresponding to 1132 original training images. Two generated variants were produced for each original image, resulting in 2264 generated samples and a theoretical training-set size of 3396 images. The validation and test subsets retained only their original fragment images.

The quality-control procedure combined foreground-contour consistency with manual visual inspection. Foreground masks were extracted from the original and generated images, and their intersection-over-union (IoU) was calculated. An IoU below 0.90 indicated excessive contour deformation that could damage the original rejoining relationship. A total of 42 generated samples showing excessive contour deformation or unreliable visual structures were removed. Manual inspection further verified that each generated image preserved a single oracle bone fragment, contained no watermark or text artifacts, intro-

duced no additional objects, and avoided implausible changes to fracture edges, surface cracks, and engraved strokes. After quality control, the final augmented training subset contained 3354 images. All generated images inherited the rejoining-group labels of their corresponding original fragments and did not introduce new rejoining relationships. These checks were used to prevent synthetic artifacts from creating artificial fracture-edge or stroke cues.

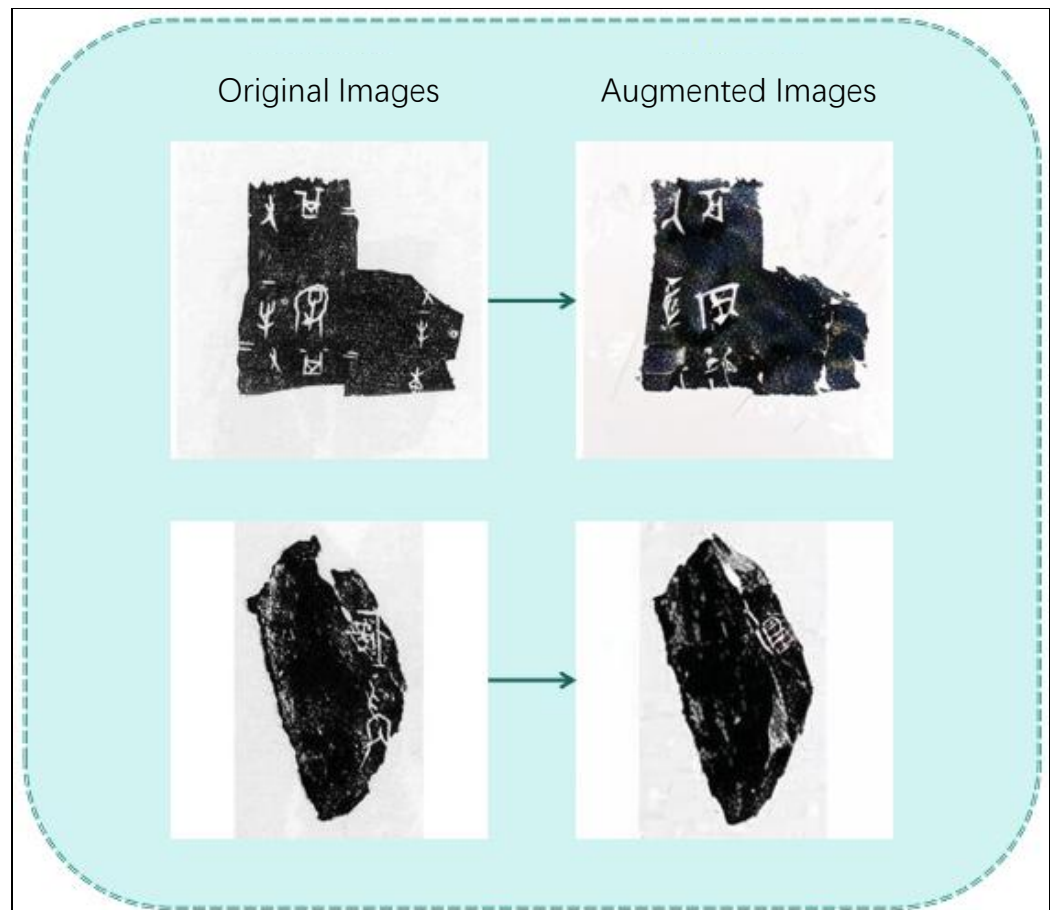


Figure 5. Comparison of samples before and after data augmentation.

The choice of two generated variants per original training image was based on the balance between data diversity and structural reliability. Oracle bone rejoining depends strongly on fracture boundaries, fragment contours, surface texture, and engraved strokes. Generating too few samples provides limited improvement in visual diversity, whereas generating too many samples increases the probability of artificial contours, distorted fracture edges, or unrealistic stroke patterns. Therefore, we used a moderate augmentation ratio to enlarge the training distribution while keeping the generated samples close to physically plausible oracle bone fragments.

3.3. FasterNet-Based Siamese Feature Encoder

After data augmentation, FaSDiNet requires a visual encoder that can extract discriminative representations from full oracle bone fragment images. In oracle bone rejoining, useful evidence is not limited to fracture boundaries. Internal texture continuity, engraved stroke traces, surface degradation patterns, and overall fragment shape may also provide important clues for judging whether two fragments can be rejoined. Therefore, an effective encoder should jointly capture boundary-level and interior visual information rather than relying only on manually selected contour curves or cropped local regions.

At the same time, full-fragment candidate screening requires repeated comparison between a query fragment and a large number of candidate fragments. This makes computational efficiency an important requirement for the feature extraction backbone. A heavy visual backbone may improve representation capacity, but it can also increase the cost of pairwise comparison in large candidate libraries. To balance representation capability and efficiency, FaSDiNet adopts FasterNet as the shared backbone of a Siamese encoder [12,20].

The Siamese encoder contains two branches with shared parameters. Given two input fragment images X_1 and X_2 , the two branches use the same FasterNet backbone to extract feature representations under a consistent mapping. The weight-sharing design encourages different fragments to be encoded in the same feature space, which is important for stable similarity measurement. It also reduces representation bias caused by imaging variation, local damage, and surface noise. After feature extraction, the two feature vectors are passed to the subsequent similarity measurement module for distance learning and binary matching.

As shown in Figure 3, each input fragment image is resized to $224 \times 224 \times 3$ before being fed into the Siamese encoder. The FasterNet backbone first uses a patch embedding layer to obtain low-level visual features, and then gradually reduces spatial resolution through several hierarchical stages. In these stages, patch merging operations increase the channel dimension while reducing the feature-map size, allowing the network to capture both local details and higher-level structural information. A global pooling layer and a fully connected layer are then used to produce the final fragment-level feature vector.

The core building unit of FasterNet is the FasterNet block, as illustrated in Figure 6. Instead of applying spatial convolution to all channels, FasterNet uses partial convolution to process only part of the channels and keeps the remaining channels unchanged [12]. This design reduces redundant computation and memory access while retaining sufficient spatial feature extraction capability. For oracle bone fragment screening, such an efficient design is valuable because the model must repeatedly compute and compare fragment representations during candidate ranking.

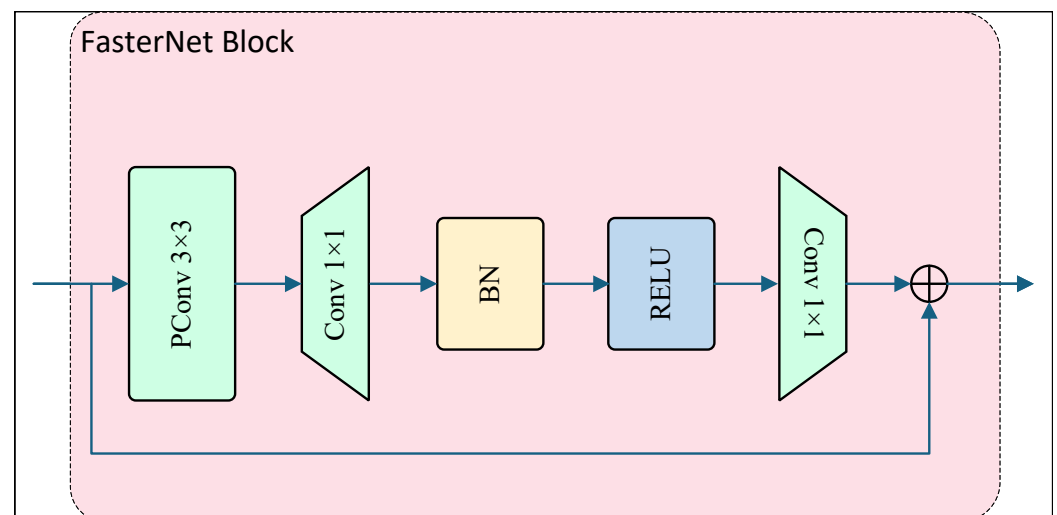


Figure 6. Architecture of the FasterNet block.

Formally, let $\mathbf{U} \in \mathbb{R}^{H \times W \times C}$ denote an input feature map. The channels are divided into a processed subset $\mathbf{U}_p$ with C_p channels and an unchanged subset $\mathbf{U}_r$ with $C - C_p$ channels. Partial convolution can be written as

$$\text{PConv}(\mathbf{U}) = \text{Concat}(\mathbf{K} * \mathbf{U}_p, \mathbf{U}_r), \quad (3)$$

where $\mathbf{K}$ is the spatial convolution kernel, and $*$ denotes convolution. Processing only part of the channels reduces redundant spatial computation, while the unchanged channels preserve complementary information. This is suitable for oracle bone fragment representation because candidate screening requires both localized responses to fracture boundaries and efficient preservation of texture, engraved-stroke, and global-shape information.

The visual behavior of the encoder is illustrated in Figure 7. The feature responses suggest that the FasterNet encoder responds not only to structural regions near fracture boundaries, but also to internal texture and engraved details. This indicates that the encoder can preserve global structural information while highlighting local discriminative cues that are directly relevant to rejoining. Such a representation is well aligned with the needs of oracle bone candidate screening, where reliable candidate ranking depends on both shape compatibility and interior continuity cues.

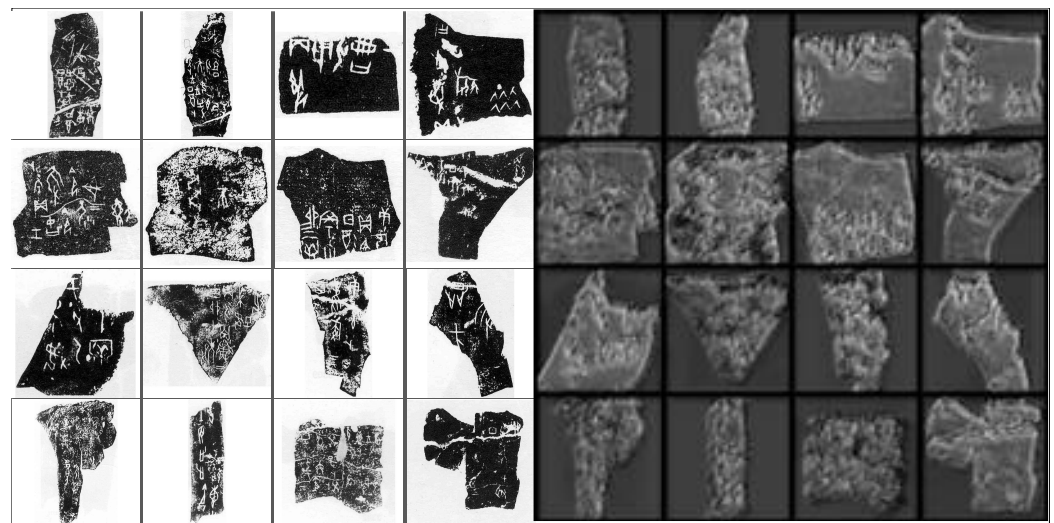


Figure 7. Feature responses extracted by the first FasterNet block.

Overall, the FasterNet-based Siamese encoder provides the representation basis of FaSDiNet. It enables the model to extract full-fragment features efficiently while preserving both boundary-related and interior visual evidence. The resulting feature vectors are then used by the contrastive similarity learning module to measure pairwise rejoining relationships.

3.4. Contrastive Similarity Learning

Once full-fragment representations have been extracted, the remaining problem is how to measure the rejoining relationship between two fragments. This differs from ordinary image classification, because oracle bone rejoining aims to determine whether two fragments can form a valid pair rather than assigning each fragment to a fixed category. In practical screening, the model should not only output a matching probability, but also rank true rejoinable fragments near the top of the candidate list. Therefore, the learned feature space should reflect the pairwise structure of the task: rejoin pairs should be close to each other, while non-rejoin pairs should be sufficiently separated. To achieve this, FaSDiNet uses a pairwise similarity measurement module that combines contrastive distance supervision in the embedding space with binary classification supervision at the decision level [21,22].

Let X_1 and X_2 denote the two input fragment images, and let Y denote their pair label, where $Y = 1$ indicates a verified rejoin pair, and $Y = 0$ indicates a non-rejoin pair. Because the two branches share the same encoder parameters W , their feature representations

are written as $\mathbf{f}_1 = F_W(X_1)$ and $\mathbf{f}_2 = F_W(X_2)$. The Euclidean distance between the two embeddings is defined as

$$D_W(X_1, X_2) = \|\mathbf{f}_1 - \mathbf{f}_2\|_2, \quad (4)$$

where a smaller distance indicates greater similarity in the shared feature space.

To organize the embedding space according to the rejoining relationship, the contrastive loss is defined as

$$\mathcal{L}_{\text{con}} = \frac{Y}{2} D_W^2 + \frac{1-Y}{2} [\max(0, m - D_W)]^2, \quad (5)$$

where $m > 0$ is the margin that controls the minimum desired separation between non-rejoin pairs. When $Y = 1$, minimizing the first term pulls the embeddings of a rejoin pair closer together. When $Y = 0$, the second term penalizes a non-rejoin pair only when its feature distance is smaller than m . The margin is set to $m = 1.0$ in the experiments.

In addition to feature-space distance learning, FaSDiNet produces a probability-based rejoining decision. The element-wise absolute difference between the two embeddings is denoted by $\mathbf{d} = |\mathbf{f}_1 - \mathbf{f}_2|$. A trainable decision head $h_\phi(\cdot)$ maps this difference vector to a scalar logit x , which is converted into a rejoining probability p by the sigmoid function:

$$x = h_\phi(\mathbf{d}), \quad p = \sigma(x) = \frac{1}{1 + e^{-x}}. \quad (6)$$

Here, $p \in [0, 1]$ represents the predicted probability that the two input fragments can be rejoined. The corresponding binary cross-entropy loss is

$$\mathcal{L}_{\text{bce}} = -[Y \log(p) + (1 - Y) \log(1 - p)]. \quad (7)$$

The contrastive loss and binary cross-entropy loss provide complementary supervision. The former constrains the relative geometry of fragment embeddings, whereas the latter directly optimizes the probability-based rejoining decision used for candidate ranking. They are combined as

$$\mathcal{L}_{\text{Total}} = W_1 \mathcal{L}_{\text{con}} + W_2 \mathcal{L}_{\text{bce}}, \quad (8)$$

where $W_1 + W_2 = 1$. In the experiments, W_1 and W_2 are both set to 0.5, so that feature-distance learning and binary decision learning contribute equally to the overall objective. The Adam optimizer minimizes $\mathcal{L}_{\text{Total}}$ to jointly update the shared encoder and the decision head. Algorithm 2 summarizes the computational procedure of this similarity module.

Algorithm 2: Contrastive learning-based similarity measurement algorithm

Input : Batch image features

$$\Phi_{\text{BatchImagesFeatures}} = [\delta_{1\text{MatchImageFeature}}, \delta_{2\text{MatchImageFeature}}], Y_{\text{Label}}$$

Output: $\xi_{\text{Similarity}} \in [0, 1], \mathcal{L}_{\text{TotalLoss}}$

foreach $(\delta_{1\text{MatchImageFeature}}, \delta_{2\text{MatchImageFeature}}) \in \Phi$ **do**

$$\mathcal{L}_{\text{ContrastiveLoss}} \leftarrow \text{ContrastiveLoss}(\delta_{1\text{MatchImageFeature}}, \delta_{2\text{MatchImageFeature}});$$

$$\lambda_{1\text{Flattened}}, \lambda_{2\text{Flattened}} \leftarrow \text{Flattening}(\delta_{1\text{MatchImageFeature}}, \delta_{2\text{MatchImageFeature}});$$

$$\gamma \leftarrow \text{L1Norm}(\lambda_{1\text{Flattened}} - \lambda_{2\text{Flattened}});$$

$$\xi_{\text{Similarity}} \leftarrow \text{MLP}(\gamma);$$

$$\mathcal{L}_{\text{BCELoss}} \leftarrow \text{BCELoss}(\xi_{\text{Similarity}}, Y_{\text{Label}});$$

$$\mathcal{L}_{\text{TotalLoss}} \leftarrow W_1 \times \mathcal{L}_{\text{ContrastiveLoss}} + W_2 \times \mathcal{L}_{\text{BCELoss}};$$

return $\xi_{\text{Similarity}}, \mathcal{L}_{\text{TotalLoss}};$

Overall, this similarity module does not rely on a single distance function or a single classification function. Instead, it uses composite-loss learning to jointly model feature closeness and matching feasibility. This improves the model's sensitivity to interior details and local engraved cues while increasing the stability of final rejoining decisions. As a result, FaSDiNet provides a more reliable basis for candidate ranking and full-image screening in oracle bone rejoining.

3.5. Datasets

Automatic oracle bone rejoining is highly dependent on data, yet the amount of oracle bone imagery that can be directly used for model training and public comparison remains limited [1,7]. On the one hand, oracle bone fragment images are costly to collect and curate [1]. On the other hand, verified rejoin samples usually require long-term manual screening and expert confirmation [1,7], which makes the number of available positive pairs far smaller than that in common visual tasks. To validate the proposed method, this study uses a self-built OBF dataset together with the public COBD dataset. The former is used to evaluate full-image candidate screening under weak manual intervention, while the latter serves as a public benchmark for comparison with existing oracle bone rejoining methods.

3.5.1. OBF Dataset

The OBF dataset is built for raw oracle bone fragment images and is intended to support automatic rejoining retrieval directly from full fragment images. Its positive pairs are constructed through manual screening and verification based on published oracle bone collation results [1] and are standardized according to data organization practices used in prior computer-assisted oracle bone rejoining studies [7]. Figure 8 shows representative rejoinable oracle bone fragment pairs in the OBF dataset.

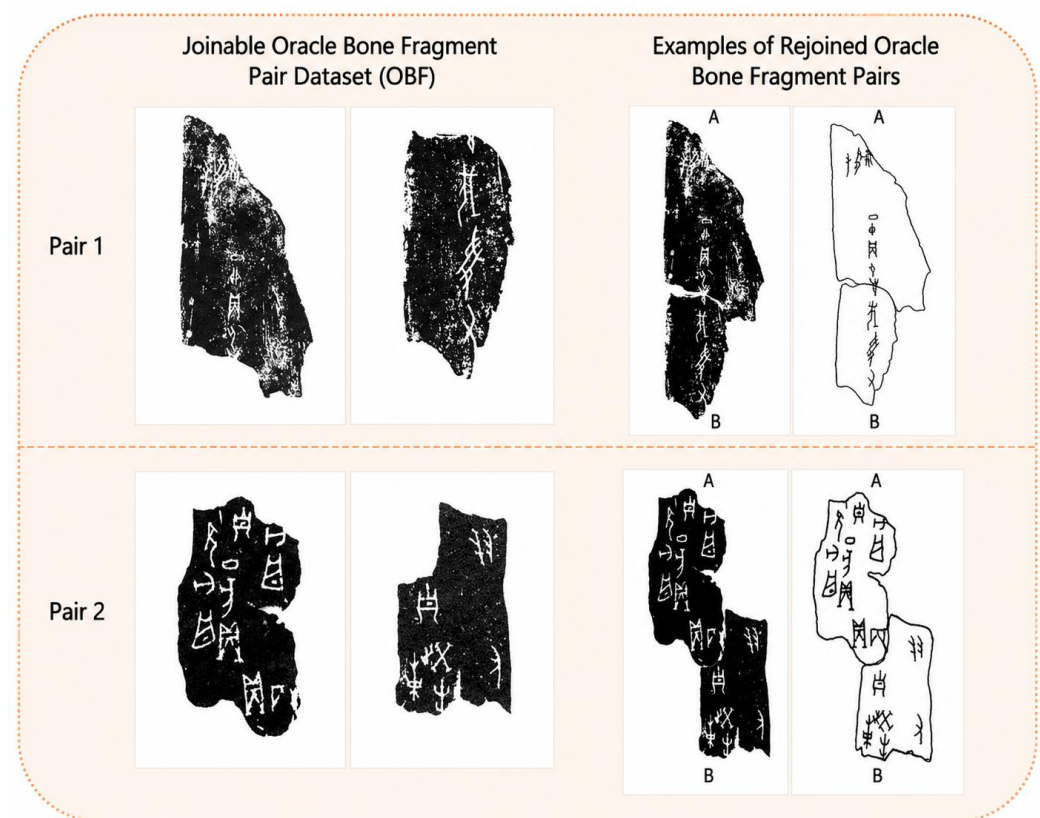


Figure 8. Representative rejoinable oracle bone fragment pairs in the OBF dataset.

The final OBF dataset contains 1254 oracle bone fragment images, including 587 rejoinable relationships that are further split into 627 positive pairs. The rejoining labels in OBF are derived from published oracle bone rejoining records, including the *Jia Gu Pin He Ji* series. These records provide literature-confirmed rejoinable relationships and are used as the ground-truth basis for constructing positive pairs. The dataset is organized by rejoining pair groups. Images belonging to the same rejoining group are regarded as positive candidates, whereas images from different groups are used to construct negative pairs. After cropping, denoising, restoration, and scale normalization, all images are re-sized to 224×224 pixels. Unlike methods that mainly rely on contour curves or candidate fracture-edge fragments [8–10], OBF preserves the complete visual information of full fragment images and therefore better supports the workflow emphasized in this paper: full-image input, automatic candidate screening, and subsequent similarity ranking.

During preprocessing, each fragment image retains the orientation provided by its source record, and no additional rotation normalization is applied. This preserves the natural visual variation of oracle bone fragments and supports full-image candidate screening.

3.5.2. COBD Dataset

In addition to the self-built OBF dataset, the public COBD dataset is used as a benchmark platform for comparison. In existing oracle bone rejoining research, contour-curve-based evaluation has remained an important setting for public comparison [7,8,10]. Unlike OBF, which focuses on raw fragment images, COBD emphasizes contour-level rejoining cues, and the two datasets therefore differ in both data form and task setting.

Experiments are conducted on both OBF and COBD. The former is used to evaluate the automatic screening capability of FaSDiNet under full-image conditions, while the latter supports comparison with existing public methods. Their combination allows the proposed method to be evaluated from two complementary perspectives: practical full-image candidate screening and public benchmark comparison.

3.6. Implementation Details

All experiments are implemented in PyTorch 1.13.1 with CUDA 11.7. The principal hyperparameters used for FaSDiNet training and latent diffusion augmentation are summarized in Table 2. The network-training settings follow the experimental configuration of the FasterNet-based Siamese model, while the augmentation settings correspond to the final configuration used to generate the OBF training samples.

The shared Siamese backbone uses FasterNet-T1 initialized with pretrained weights. For each input fragment, the final $7 \times 7 \times 512$ feature map is aggregated by global pooling and further transformed by a 1×1 convolution and a fully connected layer to obtain a fixed-length fragment representation. All encoder parameters are shared between the two Siamese branches. For binary matching, the fragment representations are flattened, and their element-wise absolute difference is mapped to a scalar logit by a linear decision layer. The logit is then converted into a matching probability through the sigmoid function.

The experiments are conducted on a workstation equipped with an NVIDIA GeForce RTX 3060 GPU with 12 GB memory, an 8-core Intel Core i7-12700K CPU @ 3.6 GHz, and 16 GB RAM. The same implementation environment is used for the comparative, ablation, and case-study experiments to ensure consistency across different experimental settings.

Before latent diffusion augmentation, the OBF data were divided by rejoining-pair group into training, validation, and test subsets. The split contains 566 training pair groups, 42 validation pair groups, and 19 test pair groups. All images associated with the same original fragment identity and rejoining group were assigned exclusively to one subset. Therefore, an original fragment could not appear in different subsets through different

pair combinations. The split was completed before augmentation, and latent diffusion augmentation was applied only to the training subset. All generated variants remained in the same training subset as their corresponding original fragments, while the validation and test subsets retained only their original images.

Table 2. Principal hyperparameters used in FaSDiNet training and latent diffusion augmentation.

Hyperparameter	Setting
FaSDiNet training	
Backbone and initialization	FasterNet-T1 initialized with pretrained weights
Input resolution	$224 \times 224 \times 3$
Optimizer and batch size	Adam; batch size 2
Training schedule	500 epochs in total
Initial learning rates	1×10^{-4} for stage 1 (epochs 1–50); 1×10^{-5} for stage 2 (epochs 51–500)
Learning-rate scheduler	StepLR; step size 1 epoch; $\gamma = 0.96$
Weight decay	0.001 in the second training stage
Contrastive margin m	1.0
Loss weights	$W_1 = 0.5$ for contrastive loss; $W_2 = 0.5$ for BCE loss
Latent diffusion augmentation	
Pretrained checkpoint	Stable Diffusion v1.5 inpainting
Inference sampler and iterations	DDIM; 50 inference iterations per generated image
Guidance scale and denoising strength	5.0 and 0.30, respectively,
Random seed and boundary-mask width	20250613 and 5 pixels, respectively,
Generated variants	Two variants per original training image
Pretrained scheduler configuration	1000 training timesteps; scaled-linear beta schedule from 0.00085 to 0.012

Pair construction follows the rejoining-group organization of the OBF dataset. During training, two images sampled from the same rejoining group form a positive pair with label $Y = 1$. A negative pair with label $Y = 0$ is formed by combining one image from the selected group with an image randomly sampled from a different rejoining group. Diffusion-augmented images inherit the group labels of their corresponding original fragments and therefore do not create new rejoining relationships. Each training sampling unit contains one positive pair and one negative pair, producing a balanced positive-to-negative ratio of 1:1. For validation, the positive and negative pairs are constructed in advance and remain fixed across epochs, thereby reducing randomness in validation loss and checkpoint selection. No explicit hard-negative mining is used in the current implementation; all negative pairs are randomly sampled from different rejoining groups. During test-time retrieval, each query fragment is ranked against all other candidate fragments rather than against an artificially balanced set of test pairs.

3.7. Evaluation Metrics

Because the goal of this work is not to output a single fully automatic final rejoining result, but to perform candidate screening and ranking in a large fragment library, Top-K recall is adopted as the core evaluation metric. For a given query fragment, retrieval is considered successful if the true matching fragment appears within the top K returned candidates.

Let Q denote the number of query fragments, let $\mathcal{C}_q^K$ denote the set of the first K candidates returned for query q , and let $\mathcal{G}_q$ denote the set of its verified rejoinable fragments. Top-K recall is defined as

$$\text{Recall@K} = \frac{1}{Q} \sum_{q=1}^Q \mathbf{1}(\mathcal{C}_q^K \cap \mathcal{G}_q \neq \emptyset), \quad (9)$$

where $\mathbf{1}(\cdot)$ is the indicator function. A query is counted as successfully retrieved when at least one verified rejoinable fragment occurs within its first K candidates.

Top-K recall is used because oracle bone rejoining in this study is formulated as a ranked candidate-screening task rather than an isolated binary classification task. Top-1 reflects whether the true rejoinable fragment is returned as the first candidate, whereas Top-5, Top-10, and Top-15 measure whether it is preserved within progressively broader candidate ranges for expert verification. These settings correspond to the practical workflow in which an automatic system narrows a large fragment collection to a manageable candidate list, while the final rejoining decision remains subject to scholarly examination. Top-K recall therefore directly evaluates the candidate-preservation capability that FaSDiNet is designed to provide.

In the experiments on the OBF dataset, Top-1, Top-5, Top-10, and Top-15 recall are reported to evaluate the full-image candidate-screening capability of the proposed method. In the experiments on the public COBD dataset, Top-1, Top-3, Top-5, Top-10, and Top-15 recall are reported following the comparison protocol of existing oracle bone rejoining studies. Based on these metrics, comparative and ablation experiments are conducted to evaluate the automatic screening capability of FaSDiNet, its performance on a public benchmark, and the contribution of its key components.

4. Results

This section reports the experimental results of FaSDiNet based on the datasets, implementation details, and evaluation metrics described above. The results include comparison on the self-built OBF dataset, benchmark evaluation on the public COBD dataset, ablation analysis, and a representative rejoining case study.

4.1. Results on the OBF Dataset

To strengthen the comparison under full-fragment-image conditions, six methods were evaluated on the self-built OBF dataset. In addition to the typical Siamese network [20,21] and typical triplet network [25], three representative modern baselines were introduced: ResNet50-Siamese as a conventional deep CNN baseline [14], ViT-Tiny-Siamese as a lightweight Vision Transformer baseline [15], and ImageNet-initialized SimCLR-based retrieval as a self-supervised representation-learning baseline [23]. The newly added baselines were evaluated using the same OBF data split and all-candidate test-time ranking protocol. Because OBF directly uses full fragment images without manually annotated candidate fracture edges, this comparison evaluates the candidate-screening capability of different visual representation architectures under the intended task setting. Table 3 presents the corresponding Top-K recall results.

Table 3. Top-K recall on the OBF dataset (%).

Method	Top-1	Top-5	Top-10	Top-15
Typical Siamese Network	7.89	28.95	42.11	57.90
Typical Triplet Network	5.26	26.32	39.47	50.00
ResNet50-Siamese	21.05	42.11	52.63	78.95
ViT-Tiny-Siamese	26.32	47.37	60.53	68.42
SimCLR-based Retrieval	31.58	55.26	68.42	76.32
FaSDiNet	23.68	60.53	76.32	84.21

As shown in Table 3, SimCLR-based retrieval achieves the highest Top-1 recall of 31.58%, followed by ViT-Tiny-Siamese at 26.32% and FaSDiNet at 23.68%. This result reflects the different representation and ranking characteristics of the compared methods.

The SimCLR-based baseline uses instance-discriminative contrastive learning and ranks normalized representations by cosine similarity, which is beneficial for nearest-candidate retrieval. ViT-Tiny-Siamese uses global self-attention to capture long-range visual context and overall fragment appearance. By contrast, FaSDiNet combines FasterNet-based spatial feature modeling with contrastive and binary-classification supervision, enabling fracture-boundary shape, internal texture, engraved strokes, and overall fragment appearance to contribute jointly to candidate ranking.

At broader candidate ranges, FaSDiNet achieves the highest Top-5, Top-10, and Top-15 recall values of 60.53%, 76.32%, and 84.21%, respectively. Compared with the strongest newly added baseline at each candidate range, the corresponding improvements are 5.27, 7.90, and 5.26 percentage points. This performance pattern is consistent with the intended role of FaSDiNet: preserving true rejoinable fragments within a manageable ranked candidate list for subsequent expert verification. Its stronger performance at Top-5, Top-10, and Top-15 demonstrates more reliable candidate retention as the expert-review range expands.

To further evaluate training stability, FaSDiNet was independently trained ten times using different random seeds under the same data split and experimental configuration. Table 4 reports the mean and standard deviation of the Top-K recall results. The mean results are consistent with the primary results reported in Table 3, while the relatively limited standard deviations indicate stable performance across independent runs.

Table 4. FaSDiNet performance over ten independent runs on the OBF dataset (%).

Metric	Mean $\pm$ SD (%)
Top-1 Recall	22.89 $\pm$ 2.50
Top-5 Recall	59.47 $\pm$ 2.22
Top-10 Recall	75.53 $\pm$ 1.27
Top-15 Recall	83.42 $\pm$ 2.50

Beyond candidate-retention performance, computational efficiency is also important for practical expert-assisted screening of large oracle bone fragment collections. We therefore profiled the final FaSDiNet model on an NVIDIA GeForce RTX 3090 using FP32 precision, an input resolution of 224×224 , and a batch size of one. The model contains 25.18 million parameters. Single-image feature extraction requires 1.72 GFLOPs, while a complete Siamese forward pass for one fragment pair requires 3.46 GFLOPs.

For the retrieval-efficiency evaluation, candidate features were precomputed from the 38 OBF test images, and each query was ranked against the remaining 37 candidates. After 50 warm-up iterations, 200 measurements were performed with CUDA synchronization. The median feature-extraction time was 6.38 ms, and the median end-to-end retrieval latency was 6.60 ms, including query-feature extraction, similarity computation, and candidate sorting. The peak GPU memory consumption was 233.41 MB, including the model and candidate feature bank. These results demonstrate that FaSDiNet supports millisecond-level full-image candidate retrieval with moderate memory consumption and a low computational cost, making it suitable for practical oracle bone fragment screening.

4.2. Results on the COBD Dataset

In addition to the self-built dataset, further comparative experiments are conducted on the public COBD dataset. Unlike OBF, which is based on full fragment images, COBD emphasizes contour-curve-level rejoining cues. This experiment is therefore mainly used to evaluate the comparability of FaSDiNet against representative contour-based and Siamese-

based oracle bone rejoining methods [7,8,10]. Table 5 presents the Top-K recall results of different methods on COBD.

Table 5. Top-K recall on the COBD dataset (%).

Method	Top-1	Top-3	Top-5	Top-10	Top-15
CAP&CFDA	27.5	48.1	65.0	80.0	85.5
S^3 -Net _{BCE}	51.8	71.5	80.3	87.2	91.6
S^3 -Net _{Focal}	56.2	77.5	83.1	91.2	92.8
SFF-Siam	74.4	83.5	90.5	92.4	94.8
FaSDiNet	50.0	71.9	81.3	93.8	96.9

As shown in Table 5, different methods show different performance characteristics under different Top-K settings.

First, the CAP&CFDA method uses curve fitting and Pearson correlation analysis to complete oracle bone fragment rejoining based on contour-curve matching [8]. It achieves 27.5%, 48.1%, 65.0%, 80.0%, and 85.5% on Top-1, Top-3, Top-5, Top-10, and Top-15 recall, respectively. This method can effectively describe the geometric similarity of fragment contours when the boundaries are clear. However, because it mainly depends on contour curves, its robustness is insufficient when oracle bone fragment boundaries are worn, incomplete, or locally damaged. Therefore, its recall is lower than that of learning-based methods on all Top-K metrics.

Second, S^3 -Net_{BCE} transforms the edge information of oracle bone fragments into a sequential shape representation through self-supervised learning and uses binary cross-entropy loss to train the network [7]. Compared with CAP&CFDA, its Top-1, Top-3, Top-5, Top-10, and Top-15 recalls increase by 24.3%, 23.4%, 15.3%, 7.2%, and 6.1%, respectively. This shows that self-supervised contour representation can significantly improve the discriminative ability of oracle bone fragment matching. However, this method still mainly relies on contour-level information, and its use of internal texture, engraved traces, and global fragment appearance is limited.

Third, S^3 -Net_{Focal} keeps the same network architecture as S^3 -Net_{BCE} but uses focal loss to complete network training [7]. It achieves 56.2%, 77.5%, 83.1%, 91.2%, and 92.8% on Top-1, Top-3, Top-5, Top-10, and Top-15 recall, respectively. Compared with S^3 -Net_{BCE}, the recall improves by 4.4%, 6.0%, 2.8%, 4.0%, and 1.2%, respectively. This indicates that a more suitable loss function can improve the rejoining performance of the network, especially by enhancing the model's attention to difficult contour-matching samples. Nevertheless, because its input still focuses on contour representation, its ability to use full-fragment visual evidence remains limited.

Fourth, SFF-Siam uses a similarity feature fusion module to enhance the feature extraction ability of the Siamese network for rejoinable fragment images and candidate fragment images, and uses a forward feedback network to calculate the rejoining probability between fragment images [10]. It achieves 74.4%, 83.5%, and 90.5% on Top-1, Top-3, and Top-5 recall, respectively, which are the best results among all compared methods at smaller K values. Compared with S^3 -Net_{Focal}, SFF-Siam improves Top-1, Top-3, and Top-5 recall by 18.2%, 6.0%, and 7.4%, respectively. This demonstrates that similarity feature fusion and forward feedback are effective for local contour-based similarity modeling. However, at Top-10 and Top-15, SFF-Siam achieves 92.4% and 94.8%, which are lower than FaSDiNet. This suggests that although SFF-Siam is strong in returning the most likely contour candidate at the very top, its candidate retention ability becomes less advantageous when the candidate range is expanded.

Fifth, FaSDiNet achieves 50.0%, 71.9%, 81.3%, 93.8%, and 96.9% on Top-1, Top-3, Top-5, Top-10, and Top-15 recall, respectively. Compared with CAP&CFDA, FaSDiNet improves Top-10 and Top-15 recall by 13.8% and 11.4%, respectively. Compared with S^3 -Net_{BCE}, the improvements at Top-10 and Top-15 are 6.6% and 5.3%. Compared with S^3 -Net_{Focal}, the improvements are 2.6% and 4.1%. Compared with SFF-Siam, FaSDiNet further improves Top-10 and Top-15 recall by 1.4% and 2.1%, respectively. These results show that FaSDiNet has stronger high-recall candidate preservation ability at larger candidate ranges.

The differences at smaller and larger Top-K values should be interpreted in relation to the input preparation and intended functions of the compared methods. SFF-Siam and other contour-oriented methods use contour curves or candidate fracture-edge regions that have been prepared or localized before similarity calculation. This prior localization allows these methods to concentrate directly on local boundary compatibility and is particularly beneficial for returning the most likely contour match within the first few ranking positions. Their stronger Top-1 and Top-5 recall is therefore consistent with the use of localized contour information.

FaSDiNet is designed for full-fragment-image candidate screening without requiring experts to interactively select candidate fracture positions or annotate candidate edge segments before retrieval. Its intended output is a ranked candidate list for expert verification rather than a single automatic final decision. Accordingly, OBF is used to evaluate the intended full-image candidate-screening workflow, whereas COBD provides a complementary public comparison with existing contour-oriented methods. On COBD, FaSDiNet achieves Top-10 and Top-15 recall values of 93.8% and 96.9%, exceeding the 92.4% and 94.8% achieved by SFF-Siam. This performance pattern reflects the different input preparation and task orientation of the compared methods rather than an overall superiority claim. The larger-K results demonstrate the candidate-retention capability of FaSDiNet and its intended role in reducing preliminary manual screening while preserving true rejoinable fragments for expert verification.

4.3. Ablation Study

To validate the contributions of the major components of FaSDiNet, ablation experiments are conducted on the OBF dataset. Two aspects are examined: the effect of latent diffusion augmentation and the contributions of different backbone, loss, and augmentation combinations.

First, Table 6 reports the recall changes before and after applying latent diffusion augmentation [13] under $K = 1, 5, 10$, and 15.

Table 6. Top-K recall before and after data augmentation (%).

Top-K	Before Data Augmentation	After Data Augmentation
Top-1	15.79	23.68
Top-5	55.26	60.53
Top-10	63.16	76.32
Top-15	73.68	84.21

As shown in Table 6, the model improves on all Top-K metrics after data augmentation: Top-1 increases from 15.79% to 23.68%, Top-5 from 55.26% to 60.53%, Top-10 from 63.16% to 76.32%, and Top-15 from 73.68% to 84.21%. These results indicate that latent diffusion augmentation effectively expands the visual variation of training samples and improves both representation quality and candidate retention under small-sample conditions, with especially clear gains at larger candidate ranges.

Second, five module combinations are tested to further analyze the contribution of each component. Specifically, V0 consists of a VGG16-based Siamese network [20,26] with cross-entropy loss; V1 adds the composite loss to V0; V2 further adds conventional data augmentation to V1; V3 replaces the feature backbone with FasterNet [12] while retaining the composite loss and data augmentation; and V4 further introduces latent diffusion augmentation [13] on top of V3, corresponding to the complete FaSDiNet framework. Figure 9 presents the Top-5 recall changes of these five variants. This progressive ablation focuses on Top-5 recall because Top-5 corresponds to a compact candidate list that can be efficiently examined by experts while still testing whether the true rejoinable fragment is retained near the top of the ranking. The full Top-K behavior of the complete model is reported in the main comparison experiments and in the augmentation ablation analysis, whereas the module-combination ablation is used to show the stepwise contribution of the major design choices.

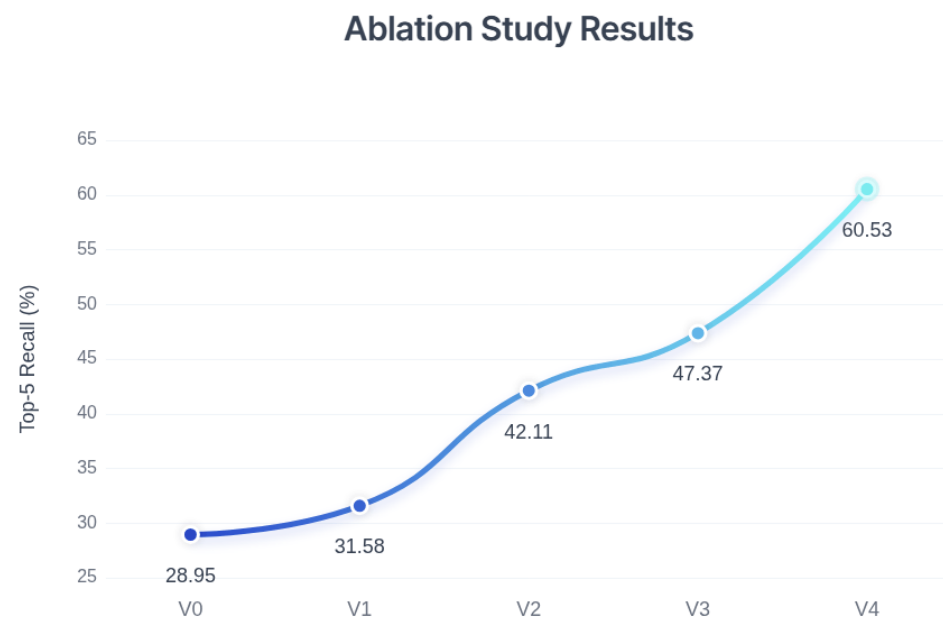


Figure 9. Ablation results of different module combinations.

As shown in Figure 9, performance increases steadily as the module design becomes more complete. The improvement from V0 to V1 indicates that a single cross-entropy loss is insufficient to constrain pairwise feature structure, whereas the composite loss enables better discrimination between rejoin and non-rejoin pairs. The improvement from V1 to V2 shows that conventional augmentation enhances robustness to scale changes, pose variation, and local defects. The change from V2 to V3 suggests that replacing VGG16 with FasterNet yields a more efficient feature representation while maintaining expressive power. Finally, V4 achieves the best performance after adding latent diffusion augmentation, indicating clear complementarity among diffusion augmentation, FasterNet backbone, and composite similarity learning. Taken together, the ablation study verifies the effectiveness of the proposed method from two aspects: latent diffusion augmentation significantly improves candidate recall under small-sample conditions, and the progressive combination of FasterNet, composite loss, and augmentation steadily enhances similarity modeling performance.

4.4. Rejoining Case Study

To further illustrate the practical performance of FaSDiNet from a qualitative perspective, a representative sample from the OBF dataset is selected for case analysis. For

the input fragment No. 28035 from *Jiaguwen Heji*, the top-ranked result among the Top-5 candidates retrieved by FaSDiNet is fragment No. 26885, which is the true rejoinable counterpart, with a similarity score of 0.908. The similarity score is used as a relative model-based ranking value for ordering candidates rather than as a direct numerical representation of expert confidence. The final validity of a proposed rejoining relationship remains subject to expert examination of fracture boundaries, surface textures, engraved strokes, inscriptional content, and the broader scholarly context. Further observation shows that the two fragments exhibit not only good contour complementarity along the fracture boundary, but also high consistency in internal texture continuity and the trajectory of incomplete character strokes. This result indicates that FaSDiNet does not rely solely on local geometric boundary matching, but instead learns higher-level integrated rejoining cues across oracle bone fragments.

Figure 10 presents the full visualization of this case. Figure 10a shows sample images after latent diffusion augmentation on the OBF dataset. Figure 10b presents the Top-5 retrieval results from the 1254 fragment images in OBF when fragment No. 28035 is used as input, with fragment No. 26885 ranked first. Figure 10c further shows the actual rejoining effect of the two fragments. The two pieces form a valid upper-lower rejoining relation, with natural contour continuity at the fracture boundary and continuous internal texture and stroke structure.

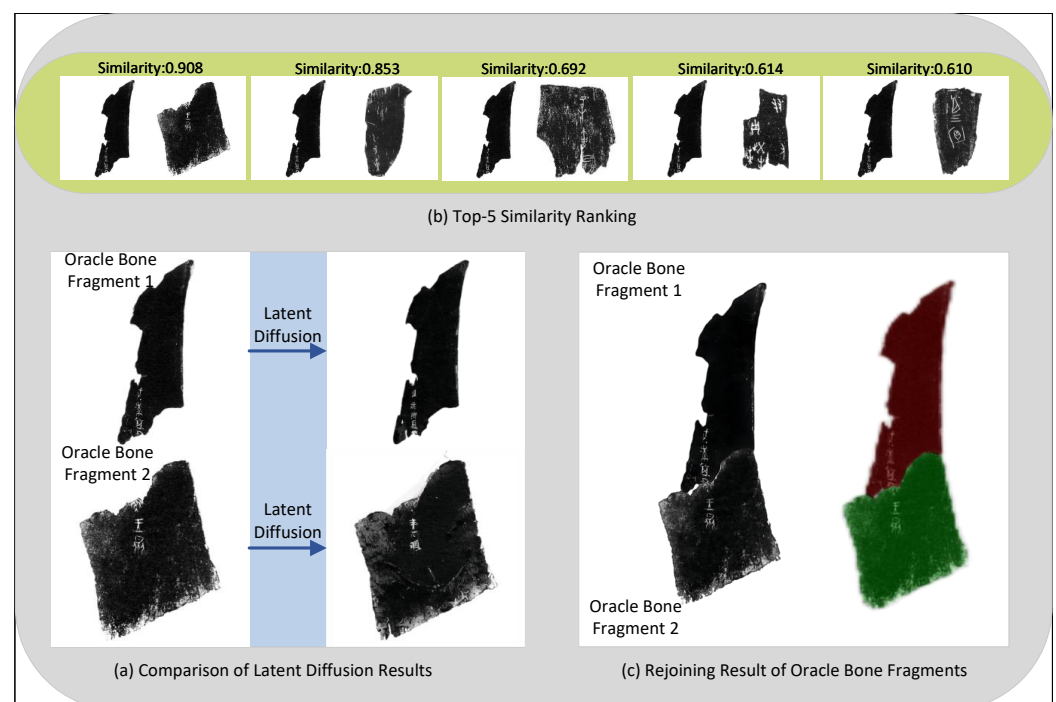


Figure 10. Oracle Bone rejoining example of FaSDiNet on the OBF dataset: (a) Samples after latent diffusion augmentation; (b) Top-5 retrieval results for the input fragment; (c) Rejoining Effect of the True Rejoinable Pair, where the red and green colors distinguish the two rejoined fragments.

From an application perspective, this case indicates that the value of FaSDiNet does not lie in directly replacing experts in making the final rejoining judgment, but in its ability to rapidly retain highly credible candidates from a large-scale fragment database and consistently rank truly rejoinable samples near the top. For oracle bone inscription collation, this means that the system can narrow the scope of manual retrieval at the front end, thereby supporting expert verification and subsequent scholarly judgment. Combined with the quantitative results on OBF and COBD discussed above, this representative case illustrates how the ranked candidates produced by FaSDiNet can be examined together

with fracture-boundary, texture, and engraved-stroke evidence. A more systematic expert-validation protocol and detailed analysis of incorrect retrieval cases are still needed to further evaluate its practical reliability in cultural-heritage applications.

5. Discussion

The experimental results indicate that the effectiveness of FaSDiNet is closely related to the practical task setting of oracle bone rejoining. On the self-built OBF dataset, FaSDiNet performs better than the typical Siamese network and the typical triplet network at all reported Top-K levels. Because OBF directly uses full fragment images rather than manually selected contour curves or candidate fracture-edge regions, this improvement suggests that the proposed method can use complementary evidence from boundary shape, internal texture, and engraved traces. This is important for practical oracle bone collation, where the key requirement is not a one-shot final decision but effective candidate screening from a large fragment library.

The results on COBD further reveal the task-oriented characteristics of the proposed method. FaSDiNet shows its strongest comparative performance at broader candidate ranges, particularly Top-10 and Top-15, which is consistent with its design goal. Contour-oriented methods concentrate directly on localized edge compatibility at smaller candidate ranges, whereas FaSDiNet is designed for full-image-based candidate screening and emphasizes high-recall preservation of credible matches for subsequent expert review. Its performance at broader Top-K ranges is therefore consistent with the practical needs of assisted oracle bone rejoining.

The ablation study helps explain these results. The gains brought by latent diffusion augmentation show that limited verified rejoin pairs remain a major bottleneck in oracle bone rejoining. Rather than creating new semantic relations, the augmentation strategy enriches the visual variation of existing samples and improves robustness under small-sample conditions. At the same time, the progressive improvement from V0 to V4 suggests that the final performance comes from the combination of three factors: the efficient FasterNet-based Siamese encoder, the joint optimization of contrastive loss and binary classification loss, and the diffusion-enhanced training distribution. Their interaction makes the model more suitable for large-scale candidate screening.

Despite these advantages, several limitations remain. First, the scale of the self-built OBF dataset is still limited. The ten-run stability analysis provides an initial assessment of training robustness under the current group-level evaluation protocol. To further strengthen the evaluation from the perspective of data partitioning and dataset diversity, future work will quantify the robustness of FaSDiNet through repeated group-wise cross-validation across different data partitions or through larger independently curated oracle bone datasets, while maintaining strict separation of verified rejoining identities between subsets. Second, the current framework is still limited to pairwise matching and does not explicitly model multi-fragment global consistency. In real oracle bone collation, one fragment may be related to multiple surrounding fragments, and the final rejoining result may depend on global structural constraints among several pieces. Third, although full-image screening improves practical applicability, it may also introduce irrelevant visual information in some cases, especially when the fragment contains complex background texture, severe surface damage, or visually similar but non-rejoinable regions. Future extensions will also incorporate structured evaluation by oracle bone specialists, systematic analysis of false-positive and false-negative retrieval cases, and attention- or activation-based visualization of model evidence to further examine the relationship between learned visual representations and expert reasoning.

Future work may therefore focus on three directions. First, pairwise screening can be extended to multi-fragment reasoning so that local matching results can be further constrained by global rejoining consistency. Second, finer-grained geometric verification and more interpretable rejoining analysis can be incorporated after candidate screening to improve the reliability of final expert review. Third, larger-scale digital oracle bone resources, interactive expert feedback, and openly shared data and code platforms can be introduced to support more sustainable digital curation tools for oracle bone studies.

6. Conclusions

FaSDiNet shows that oracle bone rejoining should not be treated only as a local boundary-matching problem. For practical digital curation, this study formulates rejoining as a visual retrieval and candidate-screening task that helps experts narrow a large fragment library to a manageable set of plausible candidates. Under this formulation, the model supports candidate screening rather than confirmed rejoining: it ranks potentially rejoinable fragments, while the final confirmation of rejoining relationships remains the responsibility of domain experts. Full fragment images are therefore valuable because they allow the model to jointly use boundary shape, internal texture, and engraved traces instead of relying only on manually selected contour curves or candidate fracture edges.

In this study, FaSDiNet combines latent diffusion augmentation, a FasterNet-based Siamese feature encoder, and joint optimization of contrastive loss and binary cross-entropy loss to support full-fragment-image candidate screening for oracle bone rejoining. Experiments on the self-built OBF dataset and the public COBD dataset show that the proposed method is most effective in broader Top-K candidate-retention settings, where the goal is to preserve truly rejoinable fragments within a manageable ranked list for subsequent expert verification. These results suggest that full-image-based screening can provide a practical visual-computing workflow for oracle bone collation and expert-assisted digital cultural heritage curation, whereas confirmed rejoining still requires domain-expert examination of fracture continuity, surface texture, engraved strokes, inscriptional content, and the broader scholarly context.

Several directions remain worth pursuing in future work. First, the current framework mainly focuses on pairwise fragment matching and can be extended to multi-fragment joint ranking and more complex rejoining relations. Second, although the proposed method already achieves high-recall candidate screening at the full-image level, there is still room to improve finer-grained geometric alignment, rejoining interpretation, and result visualization. Third, if larger-scale digital oracle bone resources, interactive expert feedback, and openly shared data and code platforms can be incorporated in the future, such methods will be more likely to evolve from isolated algorithmic validation into sustainable digital curation tools.

Author Contributions: Y.H.: Conceptualization, Methodology, Investigation, Resources, Writing—review and editing, Funding acquisition. R.T.: Conceptualization, Methodology, Investigation, Software, Writing—original draft. Y.Z.: Methodology, Data curation, Validation, Visualization. Y.F.: Formal analysis, Validation, Visualization. C.S.: Conceptualization, Supervision, Writing—review and editing, Funding acquisition, Project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Social Science Fund of China's Major Projects (No. 23&ZD308), the National Social Science Fund of China Special Project for Rare and Endangered Disciplines (Academic Team Project) (No. 25VJXT002), the General Projects of Humanities and Social Sciences Research of the Ministry of Education (No. 21YJAZH075), and the General Program of Liaoning Provincial Natural Science Foundation (No. 2025-MS-302).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The public COBD dataset used in this study is available from its original source. The source code used for model training and evaluation, model configuration files, data-split identifiers, metadata schema, preprocessing protocol, and other unrestricted experimental materials are available from the corresponding author upon reasonable request. The full public release of the self-built OBF dataset and associated fragment images requires further authorization and permission checks because some oracle bone fragment images may be subject to copyright, collection-related, or cultural-heritage data-use restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

DDIM: Denoising Diffusion Implicit Models; DPM2: second-order Diffusion Probabilistic Model Solver; DPM++: improved Diffusion Probabilistic Model Solver; DPM fast: fast Diffusion Probabilistic Model Solver; DPM adaptive: adaptive Diffusion Probabilistic Model Solver; SDE: Stochastic Differential Equation.

References

- Huang, T. The Academic Significance and Methods of Oracle Bone Rejoining. *Palace Mus. J.* **2011**, *156*, 7–13. [\[CrossRef\]](#)
- Chou, H.H.; Opstad, D.G. Computer Matching of Oracle Bone Fragments: A Preliminary Report on a New Research Method. *Archaeology* **1973**, *26*, 176–181.
- Gu, S. A Review of Research on Digital Processing of Oracle Bone Inscriptions. *J. Xihua Univ. (Nat. Sci. Ed.)* **2010**, *29*, 38–42.
- Wang, A.; Zhong, L.; Ge, Y.; Liu, G. Research on Key Technologies of Intelligent Rejoining of Oracle Bone Fragments. *J. Wuhan Univ. Technol.* **2010**, *32*, 194–199.
- Wang, A.; Liu, G.; Ge, W.; Zhou, H.; Wang, D. Design of a Computer-Aided Oracle Bone Inscription Rejoining System. *Comput. Eng. Appl.* **2010**, *46*, 59–62.
- Gu, S. A Glyph Restoration Method for Oracle Bone Rubbings Based on Mathematical Morphology. *Comput. Technol. Dev.* **2018**, *28*, 176–178. [\[CrossRef\]](#)
- Zhang, C.; Wang, B.; Chen, K.; Zong, R.; Mo, B.; Men, Y.; Almpandis, G.; Chen, S.; Zhang, X. Data-Driven Oracle Bone Rejoining: A Dataset and Practical Self-Supervised Learning Scheme. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*; Association for Computing Machinery: New York, NY, USA, 2022; pp. 4482–4492. [\[CrossRef\]](#)
- Tian, Y.; Gao, W.; Liu, X.; Chen, S.; Mo, B. The Research on Rejoining of the Oracle Bone Rubbings Based on Curve Matching. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2021**, *20*, 97. [\[CrossRef\]](#)
- Zhang, Z.; Wang, Y.T.; Li, B.; Guo, A.; Liu, C.-L. Deep Rejoining Model for Oracle Bone Fragment Image. In *Pattern Recognition: ACPR 2021*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2022; Volume 13189, pp. 3–15. [\[CrossRef\]](#)
- Yuan, J.; Chen, S.; Gao, W.; Peng, M.; Jiang, L. SFF-Siam: A New Oracle Bone Rejoining Method Based on Siamese Network. *IEEE Comput. Graph. Appl.* **2023**, *43*, 22–32. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhang, C.; Zong, R.; Cao, S.; Men, Y.; Mo, B. AI-Powered Oracle Bone Inscriptions Recognition and Fragments Rejoining. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)*; IJCAI: Occitania, France, 2020; pp. 5309–5311. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chen, J.; Kao, S.H.; He, H.; Zhuo, W.; Wen, S.; Lee, C.-H. Run, Do not Walk: Chasing Higher FLOPS for Faster Neural Networks. In *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 17–24 June 2023; pp. 12021–12031. [\[CrossRef\]](#)
- Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-Resolution Image Synthesis with Latent Diffusion Models. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 18–24 June 2022; pp. 10684–10695. [\[CrossRef\]](#)
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [\[CrossRef\]](#)
- Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. In *Proceedings of the International Conference on Learning Representations*, Virtual Event, 3–7 May 2021.

16. He, K.; Chen, X.; Xie, S.; Li, Y.; Dollár, P.; Girshick, R. Masked Autoencoders Are Scalable Vision Learners. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 16000–16009. [\[CrossRef\]](#)
17. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning Transferable Visual Models From Natural Language Supervision. In Proceedings of the 38th International Conference on Machine Learning, Virtual Event, 18–24 July 2021; Volume 139, pp. 8748–8763.
18. Zhang, R.; Fang, R.; Zhang, W.; Gao, P.; Li, K.; Dai, J.; Qiao, Y.; Li, H. Tip-Adapter: Training-Free Adaption of CLIP for Few-Shot Classification. In *Computer Vision—ECCV 2022*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2022; Volume 13695, pp. 493–510. [\[CrossRef\]](#)
19. Oquab, M.; Darcet, T.; Moutakanni, T.; Vo, H.V.; Szafraniec, M.; Khalidov, V.; Fernandez, P.; Haziza, D.; Massa, F.; El-Nouby, A.; et al. *DINOv2: Learning Robust Visual Features Without Supervision*; Transactions on Machine Learning Research: New York, NY, USA, 2024.
20. Bromley, J.; Guyon, I.; LeCun, Y.; Säckinger, E.; Shah, R. Signature Verification Using a “Siamese” Time Delay Neural Network. In *Advances in Neural Information Processing Systems*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1993; Volume 6, pp. 737–744.
21. Chopra, S.; Hadsell, R.; LeCun, Y. Learning a Similarity Metric Discriminatively, with Application to Face Verification. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 539–546. [\[CrossRef\]](#)
22. Hadsell, R.; Chopra, S.; LeCun, Y. Dimensionality Reduction by Learning an Invariant Mapping. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06), New York, NY, USA, 17–22 June 2006; Volume 2, pp. 1735–1742. [\[CrossRef\]](#)
23. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. In Proceedings of the 37th International Conference on Machine Learning, Virtual Event, 13–18 July 2020; Volume 119, pp. 1597–1607.
24. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*; Springer: Cham, Switzerland, 2015; pp. 234–241. [\[CrossRef\]](#)
25. Hoffer, E.; Ailon, N. Deep Metric Learning Using Triplet Network. In *Similarity-Based Pattern Recognition*; Springer: Cham, Switzerland, 2015; pp. 84–92. [\[CrossRef\]](#)
26. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.