Rejoining Precious Artifacts: Efficiently Bone Stick Rejoining Based Massive Fragment Images by Contour, Script, and Texture

Xingyi Wang[1], Wen Huang[1*], Mengqiang Hu[2], Junhui Chen[1], Weixin Zhao[1], Wenzheng Xu[1], Jian Peng[1]

[1]Sichuan University
[2]Tsinghua University
562421007@qq.com

Abstract

Rejoining fragment images of precious artifacts is a meaningful task because complete artifacts could provide valuable clues for the research of human civilization. However, existing rejoining methods face several challenges including time-consuming manual annotation, insufficient rejoining accuracy, and prohibitive computation cost. For rejoining fragment images of bone sticks (a precious artifact), we propose a lightweight vision graph neural network called RejoinViG to address these challenges. First, our method avoids time-consuming manual annotation of ballast contour data by experts. Specifically, our method directly takes a pair of fragment images as input and then determines whether the image pair is rejoinable. Second, our method improves rejoining accuracy by contour, script, and texture through dynamically constructing local and global graphs. Third, our method improves rejoining accuracy while reducing computation cost by introducing a new attention mechanism named node self-attention. Extensive experiments demonstrate that our method outperforms the state-of-the-art methods significantly. For example, the Top-1 accuracy of our method is 3.9 times that of SFF-Siam. Surprisingly, our method successfully rejoins a pair of previously unknown but rejoinable fragment images of bone sticks in a real-world scenario.

Code — https://github.com/Wang-XingYi/RejoinViG

Introduction

Rejoining fragment images of precious artifacts is a meaningful task. Precious artifacts record the social, economic, and political background of ancient times, so they are precious materials to appreciate the richness of human civilization (Liu et al. 2023). However, due to prolonged burial or environmental factors, unearthed artifacts usually break into many fragments. Rejoining fragment images of precious artifacts brings a second opportunity to appreciate the splendid human civilization.

Bone stick rejoining as shown in Figure 1(a) is a typical and valuable case of rejoining task. The scripts on bone sticks offer insight into the lives, beliefs, and achievements of people from the Western Han Dynasty (an ancient Chinese dynasty)(Liu et al. 2023). Thus, these scripts have significant historical and cultural values, which are essential materials for archaeologists around the world to research the history of the Western Han period (Du et al. 2022). Successfully rejoining fragment images of bone sticks can contribute to the research of ancient Chinese civilization and expand the research of human civilization.

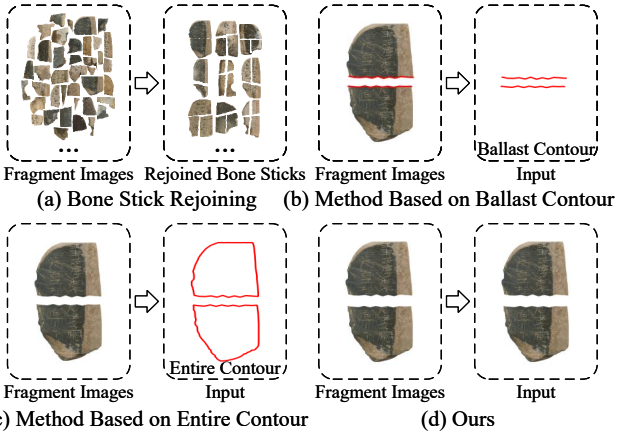

Figure 1: Examples of bone stick rejoining and existing methods' input.

The computer technology has emerged as a crucial method to complete the rejoining task. Existing methods typically take the fragment images' ballast contour (Yuan et al. 2023; Zhang et al. 2022; Tian et al. 2021; Zhang et al. 2021) (Figure 1(b)) or entire contour (Jin and Yang 2023) (Figure 1(c)) as input, and then utilize a matching algorithm to evaluate the rejoining probability of two fragment images. While these methods achieve good performance, they still face several challenges: *(i)* Ballast contour data needs to be generated manually by experts, so generating ballast contour data is very time-consuming. *(ii)* Existing methods have difficulty in utilizing script and texture information, resulting in insufficient rejoining accuracy. *(iii)* The number of artifacts' fragment images is usually massive, which leads to prohibitive computation cost. For example, the amount of bone stick fragments exceeds 60,000 (Zhang

*Corresponding author.

et al. 2024), resulting in that the rejoining method need to verify $60,000 \times 59,999 = 3,599,940,000$ possibly rejoinable pairs.

To address the above challenges, we propose a lightweight vision graph neural network called RejoinViG for rejoining fragment images of bone sticks. Specifically, to avoid manual annotation of ballast contour data, our method directly takes two fragment images as input and then determines whether the two fragment images are rejoinable, as shown in Figure 1(d). To further improve rejoining accuracy, our method dynamically constructs local and global graphs to more accurately extract features of contour, script, and texture. The local graph makes its nodes more semantic by aggregating node information within a local region. The global graph further aggregates semantic information of nodes globally, which improves the accuracy of extracted features. To improve rejoining accuracy and reduce computation cost at the same time, our method introduces a new kind of attention mechanism named node self-attention. Our node self-attention is inspired by self-attention and can adaptively adjust each node's focus to its potential neighbor nodes. In summary, our contributions are as follows:

- We design a lightweight vision graph neural network called RejoinViG to avoid manual annotation of ballast contour data. RejoinViG takes fragment images as input instead of the ballast contour, so it avoids manual annotation by experts.
- We dynamically construct a local graph and a global graph to accurately capture the features of contour, script, and texture. By progressively constructing local and global graphs, our method reduces the noise interference caused by color and texture similarity, resulting in the improvement of rejoining accuracy.
- We propose a node self-attention that improves rejoining accuracy while reducing computation cost. Node self-attention accurately captures the correlation relationship between nodes, resulting in the improvement of rejoining accuracy. Meanwhile, node self-attention only calculate correlations between each node and its potential neighbor nodes instead of all nodes, resulting in the reduction of computation cost.
- We conduct extensive experiments to verify the effectiveness of our method. The experiment results indicate that our method significantly outperforms state-of-the-art methods. For example, the Top-1 accuracy of our method is 3.9 times that of SFF-Siam. Surprisingly, our method successfully rejoins a pair of previously unknown but rejoinable fragment images of bone sticks in a real-world scenario.

Related Works

Rejoining Fragment Images

Early rejoining methods heavily rely on experts to manually analyze fragment images, so these methods are time-consuming and inefficient (Zhang et al. 2022). With advances in computer algorithms, many methods utilize ballast contour-based matching algorithms to rejoin fragment images (Yuan et al. 2023; Zhang et al. 2022; Tian et al. 2021; Zhang et al. 2021). However, the annotation cost of ballast contour data grows rapidly when the number of fragment images increases. To reduce annotation cost, Jin et al. (Jin and Yang 2023) extract the entire contour of fragment images by algorithms and then utilize the entire contour to rejoin fragment images. Although the entire contour can be automatically extracted by algorithms, the script and texture information in the fragment images are neglected. Recent studies (Zhang, Guo, and Li 2022; He et al. 2024; Zhang et al. 2024) attempt to predict similarity scores directly from fragment image pairs, but they do not consider the rejoining of multiple fragment images. To efficiently and accurately rejoin multiple fragment images, we propose a method to rejoin fragment images by contour, script, and texture.

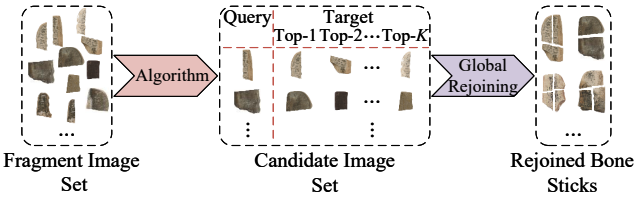

Figure 2: An intuitive presentation of the problem formulation for the rejoining task.

Graph for Computer Vision

Although graph neural networks (GNNs) are typically used for graph-based data (Wu et al. 2020), they are also widely applied to computer vision tasks now, such as image classification (Lin et al. 2022; Dong et al. 2022; Zhong et al. 2023) and image segmentation (Yang et al. 2021). With the emergence of vision GNNs (ViG) (Han et al. 2022), GNNs are gradually becoming an alternative to convolutional neural networks (He et al. 2016; Krizhevsky, Sutskever, and Hinton 2012; Liu et al. 2022) and vision transformers (ViTs) (Dosovitskiy et al. 2020). Unfortunately, ViG suffers from high computation complexity. Recent works (Han et al. 2023; Wu et al. 2023; Spadaro et al. 2024; Munir, Avery, and Marculescu 2023; Munir et al. 2024) aim to solve this problem. In particular, GreedyViG (Munir et al. 2024) introduces dynamic axial graph construction to reduce computation cost. However, GreedyViG only constructs a global graph. Its nodes lack semantic information and are easily affected by noisy neighbor nodes with similar colors and textures. To rejoin fragment images of bone sticks, we improve GreedyViG by constructing both local and global graphs.

Problem Statement

Figure 2 illustrates the example of the rejoining fragment images of bone sticks. As shown in Figure 2, rejoining task is to identify all fragment images of an artifact from massive fragment images. Formally, assume that the set of all fragment images is denoted by $F = \{I_m\}_{m=1}^{u}$ and the set of all bone sticks is denoted by $B = \{B_n\}_{n=1}^{v}$, where u is the number of images and v is the number of bone sticks. Our target is to identify multiple sets of images $S_k = \{I_j\}$,

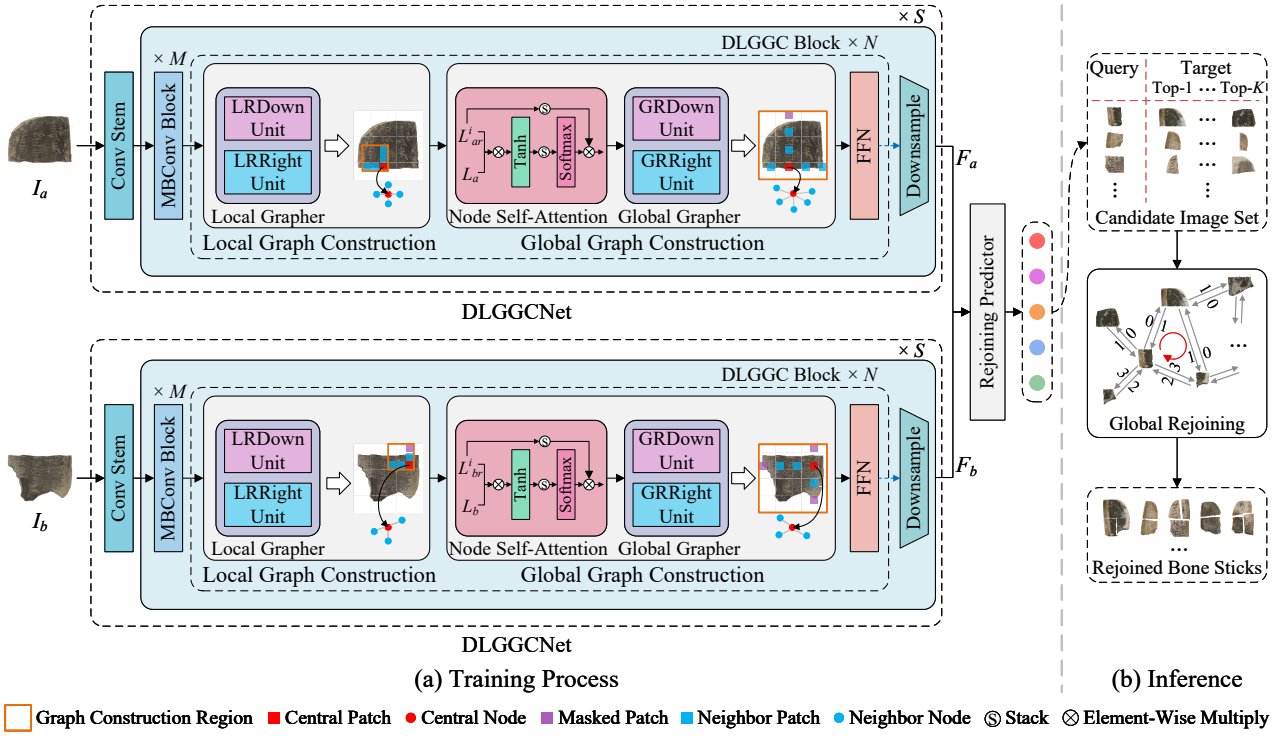

Figure 3: The pipeline of RejoinViG. DLGGCNet predicts the rejoining category of fragment image pairs. During inference, the global rejoining module is performed to rejoin multiple bone sticks based massive mixed fragment images. This pipeline effectively improves training efficiency and rejoins multiple bone sticks.

$1 \leq k \leq v$, $1 \leq j \leq u$ such that all $I_j \in S_k$ belong to the same bone stick.

Methodology

Overview

Our method aims to rejoin fragment images of bone sticks (a precious artifact). Specifically, many bone sticks are broken into many fragments due to prolonged burial or environmental factors, and the fragments of all bone sticks are mixed. Our method takes the fragment images of all bone sticks as input and outputs multiple sets of fragment images such that all fragment images in each set are from the same bone stick.

As shown in Figure 3, our RejoinViG consists of two main modules: a dynamic local and global graph construction network (DLGGCNet) module and a global rejoining module. DLGGCNet module is used to determine whether two fragment images of bone sticks are rejoinable and, if so, how the two fragment images should be rejoined such as top-bottom rejoinable, bottom-top rejoinable, left-right rejoinable and right-left rejoinable. The global rejoining module is used to identify all fragment images from the same bone sticks through predictions from DLGGCNet.

DLGGCNet

The structure of DLGGCNet is shown in Figure 3(a). Specifically, given two fragment images I_a and I_b as input, the Conv stem transforms each image into multiple graph nodes for the construction of the local and global graphs. Next, the local graph is constructed by local rolling down (LRDown) and local rolling right (LRRight) units such that node information within a local region is aggregated. Global graph is constructed through node self-attention and global grapher such that accurate node information within a global region is aggregated. Finally, rejoining predictor determines whether the two fragment images are rejoinable.

Graph Node Initialization To construct the graph structure, our method divides the fragment image into multiple graph nodes. Specifically, given a fragment image of size $H \times W \times 3$ as input, it is first processed by the Conv stem and divided into multiple patches. Each patch is treated as a graph node in the subsequent dynamic graph construction.

Instead of relying on annotated ballast contours, our method processes directly on fragment images. This design enables RejoinViG to take a pair of fragment images as input and predict whether they are rejoinable, without any manual annotation. In contrast, existing rejoining methods (Yuan et al. 2023; Zhang et al. 2022) typically require experts to manually annotate the ballast contours of fragment images, which are time-consuming and difficult to scale. By bypassing this requirement, RejoinViG not only simplifies the pipeline but also enhances practicality in rejoining scenarios of large-scale fragment images.

Dynamic Local and Global Graph Construction To accurately capture the contour, script, and texture features of

fragment images, we design a dynamic local and global graph construction (DLGGC) inspired by GreedyViG (Munir et al. 2024), as shown in Figure 3(a). The key structural and semantic information of fragment images is contained in contour, script, and texture features, which serve as important cues for determining whether two fragments are rejoinable. While GreedyViG can capture these features, it is easily affected by noisy nodes with similar colors or textures. To address this issue, we first enhance the semantics of nodes using a local graph, and then construct the global graph using these semantic nodes.

Local graph construction makes nodes more semantic by aggregating nodes within a local region. Specifically, the input feature map X is first rolled down or right within a local region to obtain the rolled feature map X_r. The size of the local region is $(D+1) \times (D+1)$, where D is the maximum distance of the connection. Rolling step K is set to 1. After each rolling operation, the Euclidean distance between the input feature map X and the rolled feature map X_r is calculated. When the distance between X and X_r is smaller than $\mu - \sigma$ (where μ and σ stand for the mean and standard deviation of the Euclidean distance between nodes, respectively), the mask is set to 1 to maintain the connection between nodes. The mask is multiplied by $X_r - X$ to preserve the max-relative scores of the connected nodes. Next, the max operation is applied to retain the result in X_{final}. After the above operations, the $Conv2d$ is applied. By enhancing the semantics of nodes, the interference from noisy neighbor nodes is effectively reduced.

Global graph construction generates a global graph in a similar way as local graph construction except for two points. First, the global graph is generated after the node self-attention operation adaptively adjusts each node's attention weight to its potential neighbor nodes. Second, the global graph is generated by rolling the feature map globally, where the rolling step K decreases as the size of the feature map decreases. The detailed algorithmic process can be found in the supplementary material.

It is worth noting that our dynamic local and global graph construction differs significantly from GreedyViG. GreedyViG constructs only a global graph, where nodes lack semantics and are easily disturbed by noisy nodes. In contrast, our dynamic local and global graph construction builds both local and global graphs. It accurately captures the contour, script, and texture features of fragment images, effectively reducing noise and improving rejoining accuracy.

Node Self-Attention More accurately capturing correlation relationships between nodes can improve the rejoining accuracy of fragment images. The reason is that nodes within cracked or broken regions often exhibit strong semantic correlations. However, existing vision GNNs (Han et al. 2022; Munir, Avery, and Marculescu 2023; Munir et al. 2024) often do not pay enough attention to correlation relationships between nodes. Introducing self-attention into vision GNNs can more accurately capture correlation relationships between each node and all other nodes, but it also brings high computation cost.

To reduce the computation cost and accurately capture correlation relationships between nodes, we design a node self-attention. As shown in Figure 4, self-attention needs to calculate correlations between the red patch and all blue patches. However, our node self-attention only needs to calculate correlations between the red patch and partial blue patches. Thus, computation cost of our node self-attention is significantly smaller than that of self-attention.

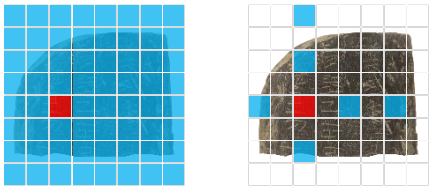

Figure 4: Comparison of computation cost between self-attention and node self-attention.

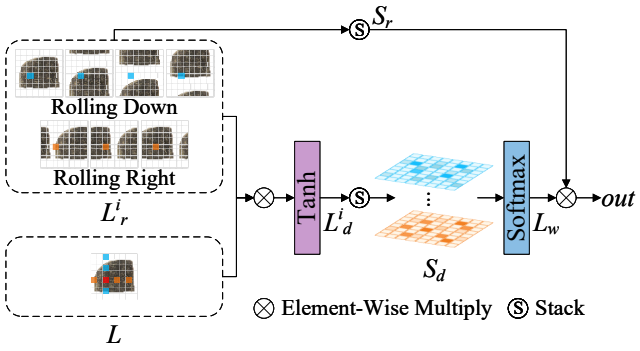

Figure 5: The structure of node self-attention. The blue and orange patches represent potential neighbor nodes obtained by rolling the red patch (node) down and right, respectively.

Figure 5 illustrates the structure of the node self-attention. Specifically, given an feature map L, node self-attention generates the rolled feature map L^i_r by rolling the feature map down and right. This process can be expressed as:

$$L^i_r = rolled(L), \quad i = 1, 2, \cdots, N \tag{1}$$

where $rolled(\cdot)$ denotes rolling operation. N represents the total number of times the feature map is rolled.

Then, node self-attention calculates the element-wise product between the feature map L and each rolled feature map L^i_r, which effectively captures the correlation relationships between each node and its potential neighbor nodes. To stabilize the training process and mitigate the risk of gradient explosion, node self-attention applies normalization and a Tanh activation:

$$L^i_d = Tanh(\frac{L \cdot L^i_r}{N}), \quad i = 1, 2, \cdots, N. \tag{2}$$

Next, node self-attention stacks the feature map L^i_d of size $W \times H \times C$ to obtain stacked feature map S_d of size $N \times W \times H \times C$. By performing a summation operation and applying softmax normalization, node self-attention obtains

attention weights L_w between each node and its potential neighbor nodes. This process can be expressed as:

$$L_w = softmax(sum(S_d)), \tag{3}$$

where $softmax(\cdot)$ denotes the softmax function. $sum(\cdot)$ is the summation operation along the channel dimension.

Node self-attention also stacks the rolled feature map L_r^i to obtain stacked feature map S_r. Finally, the weighted sum of S_r is calculated by the attention weights L_w. This process can be formulated as:

$$out = sum(S_r \cdot L_w), \tag{4}$$

where $sum(\cdot)$ is the summation operation on dimension N.

Although both self-attention and graph attention networks (GATs) (Veličković et al. 2017) could model correlation relationships between nodes, our node self-attention shows better advantages in visual tasks. In particular, GAT is primarily designed for structured data such as text, and thus it is less effective on unstructured data like images (Yang et al. 2025). GAT also typically uses an adjacency matrix to restrict each node to compute correlations only with its neighbor nodes. In contrast, our node self-attention is tailored for visual tasks. It computes correlations between each node and its potential neighbor nodes using the rolling operation and does not rely on an adjacency matrix. The node self-attention enables each node to adaptively adjust its attention to potential neighbor nodes, thereby effectively improving rejoining accuracy.

Loss Function In the rejoining task for bone sticks, we employ focal loss (Lin et al. 2017) as the loss function. Focal loss is defined as:

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t), \tag{5}$$

where α_t is the weighting factor and γ is the focusing parameter. In the experiment, α_t is set to [0.67, 0.67, 0.89, 0.89, 0.89], and γ is set to 1.5. p_t is the predictive probability.

Global Rejoining

To completely rejoin multiple bone sticks from large-scale mixed fragment images, we propose a global rejoining module inspired by JigsawNet (Le and Li 2019). As shown in Figure 3(b), the global rejoining module first uses DLGGC-Net to construct a Top-K rejoinable candidate image set for each fragment image. Based on these candidate image sets, a directed graph is build, where edge labels represent the rejoining category to ensure directional consistency along the path. To improve the accuracy of bone stick rejoining, the global rejoining module only chooses fragment image pairs satisfying a bidirectional rejoining condition. Then, the global rejoining module performs a depth-first search starting from each node to explore all possible loop paths. To eliminate redundancy, the global rejoining module filters the paths based on the uniqueness of node sets, ensuring that each loop corresponds to one individual bone stick.

Unlike JigsawNet, which relies on loop consistency, our method introduces directional constraints and bidirectional validation during directed graph construction. It performs a loop search from each node. This design reduces the computation cost of consistency checks and is more suitable for the rejoining scenario of large-scale fragment images.

		Train	Validation	Test
Fragments		29,016	1,608	1,657
Pairs	Top-Bottom	20,968	1,154	1,216
	Bottom-Top	20,968	1,154	1,216
	Left-Right	7,203	397	422
	Right-Left	7,203	397	422
	Unrejoinable	7,283	405	2,740,716
	Total	63,625	3,507	2,743,992

Table 1: Dataset division.

Experiments

Experiment Settings

Dataset To solve the problem of lacking suitable datasets, we create a fragment image dataset of bone sticks of the Han Dynasty to develop and evaluate rejoining methods. Specifically, we collect 8,093 original images from 90 volumes of published data in Han Chang'an City Weiyang Palace Bone Slips (2018–2020) to build the dataset. To simulate real-world bone stick fragments, we randomly generate splitting curves on the original bone stick images to create fragments. 32,281 fragment images are obtained in total. The division of training, validation and test set is shown in Table 1. Additional details about the dataset are provided in the supplementary material.

Implementation Details Our network is implemented in PyTorch 2.1.1 and trained on an NVIDIA GeForce RTX 4090. We use the AdamW optimizer with an initial learning rate of 2×10^{-3} and a cosine annealing scheduling strategy. Training is conducted for 300 epochs with a batch size of 64. The number of stages S is set to 4. From stage 1 to 4, the number of MBConv blocks and DLGGC blocks are set to $M = [2, 2, 6, 2]$ and $N = [2, 2, 2, 2]$, respectively. In the local graph construction, the maximum distance of the connection is set to $D = [4, 2, 1, 1]$. In the global graph construction, the rolling step is set to $K = [8, 4, 2, 1]$.

Evaluation Metric In this paper, we adopt Top-K accuracy (Zhang et al. 2023, 2022) as the evaluation metric. In the experiments, the maximum value of K is set to 15.

Comparison Methods Our method is compared with the state-of-the-art methods, including ConvNeXt (Liu et al. 2022), InceptionNeXt (Yu et al. 2024), ViT (Dosovitskiy et al. 2020), PoolFormer (Yu et al. 2022), BiFormer (Zhu et al. 2023), ViG (Han et al. 2022), WiGNet (Spadaro et al. 2024), MobileViG (Munir, Avery, and Marculescu 2023), and GreedyViG (Munir et al. 2024). We also compare our method with SFF-Siam (Yuan et al. 2023), which is a method for rejoining fragment images of oracle bones.

Experiment Results

Quantitative Comparison Table 2 reports the quantitative results between our method and comparison methods. The results show that our method not only significantly outperforms state-of-the-art methods in accuracy, but also

Method	Type	Parameters (M)	MACs (G)	Top-1 (%)	Top-3 (%)	Top-5 (%)	Top-10 (%)	Top-15 (%)
SFF-Siam	CNN	32.3	7.5	8.3	20.3	29.9	46.7	57.7
ConvNeXt	CNN	27.8	4.5	23.1	48.3	62.0	79.1	86.5
InceptionNeXt	CNN	25.8	4.2	14.1	35.4	50.8	72.6	84.2
ViT	Transformer	86.4	17.7	<u>31.7</u>	<u>54.8</u>	<u>66.0</u>	79.5	85.6
PoolFormer	Transformer	20.9	3.4	12.6	30.7	43.2	63.4	75.5
BiFormer	Transformer	25.0	4.5	8.2	20.5	34.8	59.4	73.8
ViG	GNN	28.0	4.6	23.7	49.4	64.0	<u>82.9</u>	<u>91.4</u>
WiGNet	GNN	26.4	5.8	23.6	48.0	62.0	80.6	88.7
MobileViG	CNN-GNN	6.3	1.0	11.8	28.5	40.3	57.2	68.4
GreedyViG	CNN-GNN	10.5	1.6	14.1	31.0	44.4	63.9	75.2
Ours	CNN-GNN	12.2	3.7	**32.3**	**60.9**	**76.0**	**90.0**	**94.6**

Table 2: Quantitative results of accuracy, parameters, and MACs between our method and comparison methods. Bold and <u>underline</u> represent the best and second-best results, respectively.

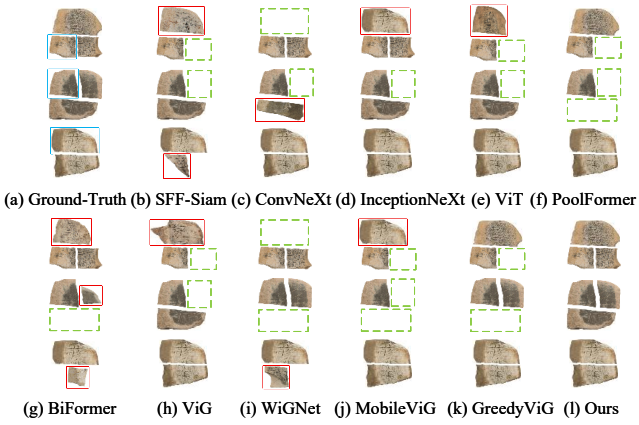

Figure 6: Qualitative results of our method and comparison methods. The blue boxes indicate the query fragment image. The red boxes denote the fragment image that is incorrectly identified as rejoinable. The green dashed boxes highlight a rejoinable region that is missed.

achieves the best balance between model complexity and accuracy. Specifically, the Top-5 and Top-10 accuracies of our method are 10% and 10.5% higher than those of the second-best method, ViT. The Top-1 accuracy of our method is 3.9 times that of SFF-Siam. Although our method has slightly more parameters and higher MACs compared to MobileViG and GreedyViG, our method consistently outperforms both methods in all Top-K accuracies. Detailed comparisons of parameters and MACs can be found in the supplementary material.

Qualitative Comparison Figure 6 reports the qualitative results between our method and comparison methods. All methods perform global rejoining to visualize the rejoining results of bone sticks. The results show that our method successfully rejoins bone sticks from a large set of fragment images, while the comparison methods have difficulty in rejoining bone sticks. In particular, when a bone stick is broken into three pieces, none of the comparison methods can correctly rejoin the bone stick. The supplementary material also provides additional comparisons of the Top-5 prediction results before global rejoining.

Figure 7: An example of a successfully rejoined bone stick in a real-world scenario. The blue box indicates the query fragment image, and the red box shows the rejoinable fragment image. For better visualization, the query fragment image is scaled. Due to scaling issue, the rejoinable fragment appears in the Top-6 prediction.

Evaluation on Real-World Scenario To evaluate the performance of our method in real-world scenarios, we randomly select 900 fragment images from bone sticks excavated at Weiyang Palace. These images are different from those used in dataset. By expert verification, our method successfully rejoins a pair of rejoinable fragments of bone sticks, which is previously unknown, as shown in Figure 7.

Visualization Results

To visualize the regions that our method focuses on within fragment images, we use Grad-CAM (Selvaraju et al. 2017) to generate activation maps, as shown in Figure 8. The results show that, given an image pair as input, our method effectively focuses to key regions required for determining rejoining relationships, such as contour, script, and texture areas. In particular, the rejoinable ballast contour regions receive significant attention.

1)	Modification	Parameters (M)	MACs (G)	Top-1 (%)	Top-3 (%)	Top-5 (%)	Top-10 (%)	Top-15 (%)
2)	Change to ResNet50 backbone	25.3	8.3	21.8	50.4	66.5	84.3	90.8
3)	Change to ViG backbone	28.7	9.2	23.8	55.0	70.5	86.4	93.1
4)	Change to GreedyViG backbone	10.7	3.2	15.1	37.7	55.8	78.8	89.2
5)	w/o local graph construction	11.5	3.5	**32.8**	60.8	74.4	89.3	94.4
6)	Change to the local branch	12.3	3.7	13.4	36.8	52.8	77.1	88.1
7)	w/o global graph construction	10.7	3.2	15.5	41.2	56.8	79.9	90.3
8)	Move to local graph construction	12.2	3.7	30.1	59.1	72.4	88.2	93.9
9)	Add to local graph construction	13.0	4.0	26.4	58.8	75.0	89.7	94.5
10)	w/o node self-attention	11.5	3.5	30.3	58.9	73.0	87.3	93.2
11)	Change to parameter $\gamma = 1$	12.2	3.7	19.4	44.4	62.3	83.1	91.2
12)	Change to parameter $\gamma = 2$	12.2	3.7	25.0	52.5	65.7	81.2	88.7
13)	Ours	12.2	3.7	32.3	**60.9**	**76.0**	**90.0**	**94.6**

Table 3: Results of ablation studies, including the backbone, local graph construction, global graph construction, node self-attention, and focal loss.

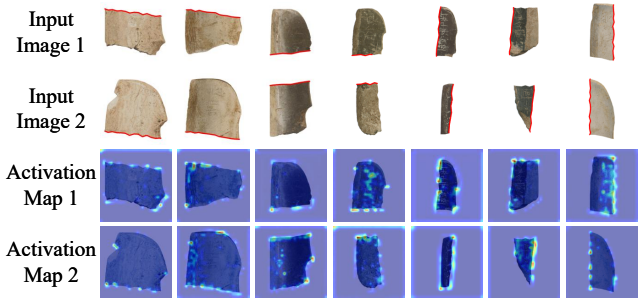

Figure 8: Comparison of Grad-CAM activation maps. The red curves indicate the rejoinable ballast contours.

Ablation Study

Backbone. To validate the effectiveness of the proposed DLGGCNet module, we replace it with several popular backbones. The experiment results are presented in rows 2-4 of Table 3. The results show that DLGGCNet module achieves the best performance and significantly outperforms the other backbones. For example, compared to GreedyViG, our method enhances Top-1 accuracy by 113.9%.

Local Graph Construction. To validate the effectiveness of the proposed local graph construction, we conduct two ablation experiments: removing the local graph construction and replacing it with the local branch (Wu et al. 2023). The experiment results are illustrated in rows 5 and 6 of Table 3. The experiment results show that removing the local graph construction reduces accuracy. For example, the Top-5 accuracy of our method decreases by 2.1%. Second, replacing local graph construction with the local branch not only decreases accuracy but also increases the number of parameters. In particular, all Top-K accuracy decreases, and the number of parameters increases by 0.8%. The reason is that the local graph construction introduces fewer neighbor nodes, decreasing the interference of noisy nodes. In summary, the proposed local graph construction is effective.

Global Graph Construction. To verify the effectiveness of proposed global graph construction, we remove all related operations from our method, as shown in row 7 of Table 3. The experiment results indicate that removing the global graph construction leads to a significant drop in all Top-K accuracies. For example, the Top-1 and Top-5 accuracies decrease by 52.0% and 32.4%, respectively.

Node Self-Attention. To evaluate the effectiveness of the proposed node self-attention, we conduct three experiments: moving node self-attention from global graph construction to local graph construction, adding extra node self-attention in local graph construction, and removing node self-attention from global graph construction. The experiment results are presented in rows 8-10 of Table 3. Our method consistently outperforms all three variants in all Top-K accuracies. For example, when adding an extra node self-attention, Top-1 accuracy declines by 18.3%. When operations related to node self-attention are removed, Top-1 accuracy decreases by 6.2%. In summary, the proposed node self-attention makes our method more accurate.

Focal Loss. To verify the effectiveness of $\gamma = 1.5$, we perform experiments on different values of the parameter γ. The experiment results are presented in rows 11-12 of Table 3. When γ is set to 1.5, the performance of our method is best.

Conclusion

To rejoin fragment images of bone sticks, we propose a lightweight vision graph neural network, termed RejoinViG. Specifically, we construct local and global graphs to improve rejoining accuracy by contour, script, and texture features. In addition, we propose node self-attention to improve rejoining accuracy and reduce computation cost at the same time. Our method not only performs well on the created dataset, but also successfully rejoins a pair of rejoinable fragment images which are previously unknown to experts. The success demonstrates the potential of our method to be successfully applied in other tasks of rejoining artifacts.

Acknowledgments

This work was supported in part by the Advanced Materials-National Science and Technology Major project under Grant 2024ZD0608200, in part by the National Natural Science Foundation of China under Grant 82474394, in part by the Key R&D Program of Sichuan Province under Grants 2023YFG0115 and 2023YFG0112, in part by the Basic Research Projects of the Sichuan Provincial Industrial Development Fund: "Container Cloud Computing Infrastructure Platform" under Grant 2023JB03, "Next-generation Big Data Visualization Based Decision-making and Analysis Platform" under Grant 2023JB06, in part by the China Postdoctoral Science Foundation under Grant 2024M752210, in part by the Sichuan University Postdoctoral Interdisciplinary Innvation Fund, and in part by the MOE (Ministry of Education in China) Liberal arts and Social Sciences Foundation under Grant 24XJCZH004.

References

Dong, Y.; Liu, Q.; Du, B.; and Zhang, L. 2022. Weighted Feature Fusion of Convolutional Neural Network and Graph Attention Network for Hyperspectral Image Classification. *IEEE Transactions on Image Processing*, 31: 1559–1572.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Du, M.; Wang, H.; Liu, R.; Wang, K.; and Wang, Z. 2022. Research on Bone Stick Text Recognition Method with Multi-Scale Feature Fusion. *Applied Sciences*, 12(24): 12507.

Han, K.; Wang, Y.; Guo, J.; Tang, Y.; and Wu, E. 2022. Vision GNN: An Image is Worth Graph of Nodes. In *Advances in Neural Information Processing Systems*, 8291–8303.

Han, Y.; Wang, P.; Kundu, S.; Ding, Y.; and Wang, Z. 2023. Vision HGNN: An Image is More than a Graph of Nodes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 19878–19888.

He, J.; Wang, H.; Liu, R.; Mao, L.; Wang, K.; Wang, Z.; and Wang, T. 2024. Research on Rejoining Bone Stick Fragment Images: A Method Based on Multi-Scale Feature Fusion Siamese Network Guided by Edge Contour. *Applied Sciences*, 14(2): 717.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Jin, Y.; and Yang, X. 2023. Interactively Rejioning 2D Oracle Bone Fragments Based on Contour Matching. In *2023 9th International Conference on Virtual Reality (ICVR)*, 163–170.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*.

Le, C.; and Li, X. 2019. JigsawNet: Shredded image reassembly using convolutional neural network and loop-based composition. *IEEE Transactions on Image Processing*, 28(8): 4000–4015.

Lin, D.; Lin, J.; Zhao, L.; Wang, Z. J.; and Chen, Z. 2022. Multilabel Aerial Image Classification With a Concept Attention Graph Neural Network. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–12.

Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; and Dollar, P. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2980–2988.

Liu, C.; Wang, H.; Mao, L.; Liu, R.; Wang, Z.; and Wang, T. 2023. Image Stitching Method of Bone Stick Fragment Based on Similarity Freeman Code Matching. *IEEE Access*, 11: 23073–23084.

Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; and Xie, S. 2022. A ConvNet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11976–11986.

Munir, M.; Avery, W.; and Marculescu, R. 2023. MobileViG: Graph-Based Sparse Attention for Mobile Vision Applications. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2211–2219.

Munir, M.; Avery, W.; Rahman, M. M.; and Marculescu, R. 2024. GreedyViG: Dynamic Axial Graph Construction for Efficient Vision GNNs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6118–6127.

Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; and Batra, D. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626.

Spadaro, G.; Grangetto, M.; Fiandrotti, A.; Tartaglione, E.; and Giraldo, J. H. 2024. WiGNet: Windowed Vision Graph Neural Network. *arXiv preprint arXiv:2410.00807*.

Tian, Y.; Gao, W.; Liu, X.; Chen, S.; and Mo, B. 2021. The research on rejoining of the oracle bone rubbings based on curve matching. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 20(6): 1–17.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Wu, J.; Li, J.; Zhang, J.; Zhang, B.; Chi, M.; Wang, Y.; and Wang, C. 2023. PVG: Progressive Vision Graph for Vision Recognition. In *Proceedings of the 31st ACM International Conference on Multimedia*, 2477–2486.

Wu, Y.; Lian, D.; Xu, Y.; Wu, L.; and Chen, E. 2020. Graph convolutional networks with markov random field reasoning for social spammer detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 01, 1054–1061.

Yang, K.; Zuo, L.; Jing, M.; Tian, X.; He, K.; and Ding, Y. 2025. Flexible ViG: Learning the Self-Saliency for Flexible Object Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*.

Yang, L.; Zhuang, J.; Fu, H.; Wei, X.; Zhou, K.; and Zheng, Y. 2021. SketchGNN: Semantic Sketch Segmentation with Graph Neural Networks. *ACM Transactions on Graphics*, 40(3).

Yu, W.; Luo, M.; Zhou, P.; Si, C.; Zhou, Y.; Wang, X.; Feng, J.; and Yan, S. 2022. MetaFormer Is Actually What You Need for Vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10819–10829.

Yu, W.; Zhou, P.; Yan, S.; and Wang, X. 2024. InceptionNeXt: When Inception Meets ConvNeXt. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5672–5683.

Yuan, J.; Shanxiong, C.; Weize, G.; Maling, P.; and Lihua, J. 2023. SFF-Siam: a new oracle bone rejoining method based on Siamese network. *IEEE Computer Graphics and Applications*, 43(6): 22–32.

Zhang, C.; Wang, B.; Chen, K.; Zong, R.; Mo, B.-f.; Men, Y.; Almpanidis, G.; Chen, S.; and Zhang, X. 2022. 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*, 4482–4492.

Zhang, C.; Zong, R.; Cao, S.; Men, Y.; and Mo, B. 2021. AI-powered oracle bone inscriptions recognition and fragments rejoining. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 5309–5311.

Zhang, X.; Wang, H.; Mao, L.; Liu, R.; Wang, Z.; and Wang, K. 2024. Bone Stick Image Matching Algorithm Based on Improved ConvNeXt and Siamese Network. *IEEE Access*, 12: 60028–60038.

Zhang, Y.; Fang, Z.; Yang, X.; Zhang, S.; He, B.; Dou, H.; Yan, J.; Zhang, Y.; and Wu, F. 2023. Reconnecting the Broken Civilization: Patchwork Integration of Fragments from Ancient Manuscripts. In *Proceedings of the 31st ACM International Conference on Multimedia*, 1157–1166.

Zhang, Z.; Guo, A.; and Li, B. 2022. Internal similarity network for rejoining oracle bone fragment images. *Symmetry*, 14(7): 1464.

Zhong, X.; Gu, C.; Ye, M.; Huang, W.; and Lin, C.-W. 2023. Graph Complemented Latent Representation for Few-Shot Image Classification. *IEEE Transactions on Multimedia*, 25: 1979–1990.

Zhu, L.; Wang, X.; Ke, Z.; Zhang, W.; and Lau, R. W. 2023. BiFormer: Vision Transformer With Bi-Level Routing Attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10323–10333.