WEAKLY-SUPERVISED ORACLE BONE INSCRIPTION COMPONENT SEGMENTATION VIA HIERARCHICAL GRAPH ALIGNMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Oracle Bone Inscription (OBI) recognition is of great scientific importance for deciphering ancient Chinese characters, a large portion of which still remain undeciphered. Since complex OBI characters are usually composed of components with independent semantics, accurately uncovering the topological relationships between characters and components provides crucial support for character decipherment. However, existing methods suffer from high annotation costs, while significant structural variants of the same component make it challenging to match against standard forms. To address these issues, we propose HiGAS, a weakly supervised method for OBI component segmentation via hierarchical graph alignment. Using standard components as topological references, HiGAS constructs bottom-up hierarchical undirected graphs across segment, sub-stroke, and stroke levels for both input characters and reference components. Through within-level edge matching and cross-level topological priors, HiGAS effectively alleviates topological discontinuity and under-segmentation caused by local feature overfitting under weak supervision. Combined with a dynamic edge selection strategy and joint optimization of classification and segmentation completeness losses, the model predicts component presence probabilities while performing fine-grained edge-level segmentation. Extensive experiments on the public Component-20 and our extended Component-50 datasets demonstrate that HiGAS achieves a top-5 hit rate of 96.77% in multi-component recognition while delivering segmentation results closest to expert annotations.

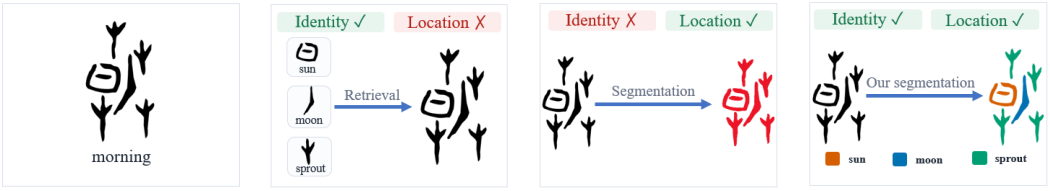
1 INTRODUCTION

Automatic recognition and analysis of Oracle Bone Inscriptions (OBIs), which are among the earliest logographic writing systems, represents a long-standing challenge in pattern recognition and computer vision Chen et al. (2026b). Among the approximately 4,500 discovered distinct OBI characters, only about one-third have been clearly deciphered Chang et al. (2022). Intelligent structural analysis of undeciphered characters is of vital value for ancient scripts decipherment Chen et al. (2025). An OBI character is typically composed of foundational sub-character units with independent semantics, termed components.

As illustrated in Fig. 1(a), a single OBI character often comprises multiple components, where identical components yield drastically distinct semantic characters through varied topological arrangements, such as two ”tree” components forming ”forest”, or a ”person” alongside a ”tree” constituting ”rest”. Accurately modeling the topological relationships and local semantics of constituent components is essential to assist palaeographers in deciphering unknown characters.

Prior studies have explored visual features to locate components Chi et al. (2024); Li et al. (2025a); Zhang et al. (2026). Early work introduced component-level retrieval to search for characters containing a query component Hu et al. (2024). However, it merely predicts component presence without spatial localization, as shown in Fig. 1(b). Subsequent work attempted component segmentation to isolate target components Hu et al. (2025), but simplified the problem to binary foreground-background masking. When multiple component categories coexist in a character, binary masking cannot separate their regions, still requiring manual expert intervention, as shown in Fig. 1(c). Fur-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



(a) Multi-component character (b) Retrieval without location (c) Segmentation without identity (d) Ours: identity and location

Figure 1: Comparison of OBI component analysis methods.

thermore, existing supervised methods rely heavily on costly pixel- or box-level annotations, strictly limiting their scalability across diverse glyph variants.

A promising avenue to bypass prohibitive annotation costs is to achieve fine-grained component segmentation guided solely by image-level presence labels, which specify component occurrences without spatial annotations. Nevertheless, this weakly supervised setting presents two fundamental challenges: (1) **Structural heterogeneity induced by glyph variations**. Lacking standardized writing rules, OBI characters exhibit non-rigid deformations, stroke additions or omissions, and mirror flipping, making cross-glyph component alignment extremely challenging. (2) **Topological discontinuity and under-segmentation**. Lacking explicit spatial supervision, models tend to overfit to highly discriminative local subgraphs while neglecting the global continuity of complete components, inevitably resulting in broken strokes and missing regions.

To address these challenges, we propose HiGAS, a weakly supervised method for OBI component segmentation via hierarchical graph alignment. Using solely image-level component presence labels, HiGAS jointly predicts component categories and achieves fine-grained component localization, as shown in Fig. 1(d). Specifically, HiGAS abstracts character images and reference components into bottom-up hierarchical undirected graphs, comprising a Segment Graph for local geometric details, a Sub-stroke Graph for continuous writing patterns across inflection points, and a Stroke Graph for complete semantic strokes across intersections. HiGAS establishes with-level edge matchings and utilizes higher-level continuous topological priors to enhance lower-level segment responses across levels. This cross-level enhancement significantly improves the matching confidence of less discriminative target segments, effectively mitigating weak supervision-induced topological discontinuity. Finally, coupled with a dynamic edge selection strategy and joint optimization objectives, HiGAS performs precise edge-level segmentation for high-confidence components. Our main contributions are summarized as follows:

- We propose HiGAS, a novel hierarchical graph alignment method that uses standard components as topological references, capturing core structural invariants across glyph variants via bottom-up Segment, Sub-stroke, and Stroke Graphs.
- We design a cross-level structure enhancement mechanism coupled with a joint optimization objective, explicitly enforcing topological response consistency across graph levels to alleviate discontinuity and under-segmentation under image-level weak supervision.
- Extensive experiments on the public Component-20 benchmark and the extended Component-50 dataset demonstrate that HiGAS achieves performance competitive with supervised methods in both multi-component presence prediction and fine-grained segmentation, offering reliable quantitative support for OBI decipherment.

2 Related Work

2.1 OBI Recognition

OBI recognition has expanded beyond classical holistic classification to encompass topological modeling and multimodal decipherment. To handle heavy noise and font variations, early approaches focus on disentangling glyph geometry from background textures Wang et al. (2022) or fusing image pixels with point sets Jiang et al. (2023), supported by standardized benchmarks such

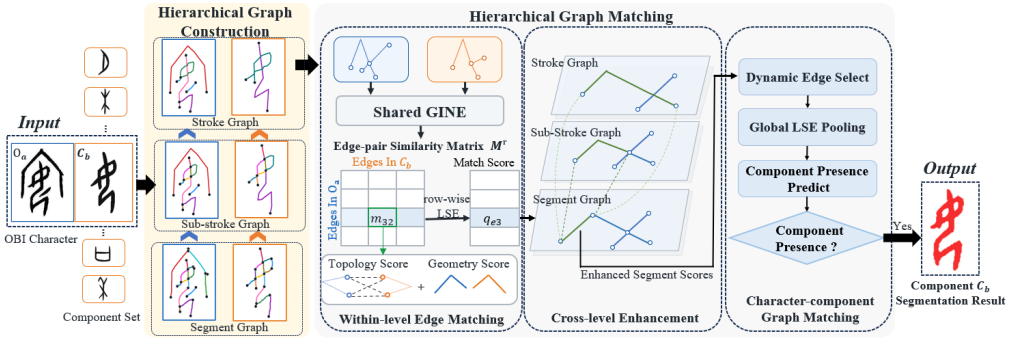


Figure 2: Overall architecture of HiGAS. Character and component inputs are abstracted into hierarchical graphs and aligned via within-level edge matching, cross-level enhancement, and character-component graph matching for joint component prediction and segmentation.

as Oracle-P15K Li et al. (2025b), HUST-OBC Wang et al. (2024b), and InteChar Diao et al. (2026). To capture complex glyph structures, recent works incorporate stroke-level priors and historical evolution, including stroke-guided semi-supervised learning Luo et al. (2026b), graph-based evolutionary modeling Chi et al. (2022); Jiao et al. (2025), character evolution benchmarks Guan et al. (2024a), and cross-font image retrieval Wu et al. (2025); Zhuo & Zhang (2024). Moreover, recent advancements leverage conditional diffusion models Guan et al. (2024b); Li et al. (2026), multimodal reasoning Wu et al. (2026), and large-scale multimodal benchmarks Song et al. (2026); Chen et al. (2026a). Despite these advances, existing methods primarily rely on holistic character-level representations, offering limited capability to explicitly deconstruct and analyze fine-grained sub-character components.

2.2 OBI Component Analysis

Deconstructing complex characters into components with independent semantics provides critical structural and graphemic clues for decipherment. Early supervised efforts employ dual-branch vision transformers Hu et al. (2024) to localize constituent radicals, supported by box- or pixel-annotated datasets such as ACCID Diao et al. (2023a). To reduce character-level supervision, recent works deconstruct glyphs into strokes for component reconstruction Wang et al. (2024a) or leverage knowledge graphs for relational reasoning Diao et al. (2023b). Recently, Large Multimodal Models and agent architectures have been adapted for component reasoning using directed acyclic graphs (DAGs), spatial grounding, or multimodal knowledge graphs Qiao et al. (2025); Li et al. (2025a); Qiao et al. (2024); Zhang et al. (2026); Chi et al. (2024). Meanwhile, initial weakly supervised explorations Hu et al. (2025) attempt pixel-level component segmentation via domain priors and topological rules. However, supervised methods are constrained by costly box/mask annotations, whereas weakly supervised approaches rely on binary masking, failing to separate different component categories.

2.3 Graph Matching and Topological Alignment

Graph matching and topological alignment map structural entities by establishing node- and edge-level correspondences. Abstracting character skeletons into graph structures allows models to align structural elements while remaining invariant to geometric deformations and writing style variations Zhang et al. (2025). To handle unstandardized ancient scripts, hybrid architectures combine vision backbones with graph neural networks to model global topology Yuan et al. (2024), while reinforcement learning-driven frameworks autonomously parse character bitmaps into coordinate-based line segment graphs Luo et al. (2026a). However, weakly supervised graph matching lacks spatial box or mask signals, causing models to overfit to discriminative local subgraphs and neglect global component continuity, which ultimately yields topological discontinuity and under-segmentation.

3 PROPOSED METHOD

3.1 OVERALL SCHEME

Given an OBI character O_a and a set of standard components $\mathcal{C} = \{C_b\}_{b=1}^{m}$, HiGAS takes any character–component pair (O_a, C_b) as input to jointly predict component presence and perform fine-grained segmentation using only presence labels. As illustrated in Fig. 2, the overall method consists of two main stages: hierarchical graph construction and hierarchical graph matching. First, we abstract character and component images into bottom-up hierarchical graphs comprising a Segment Graph, a Sub-stroke Graph, and a Stroke Graph, preserving both local geometric primitives and global semantic topology. Second, during hierarchical graph matching, HiGAS extracts structural features using GINE backbones and conducts edge-level alignment at each graph level. Crucially, top-down continuous topological priors from higher graph levels are propagated to enhance weak segment-edge responses at the lowest level. Finally, governed by a component-scale dynamic edge selection strategy and optimized via a joint loss function, HiGAS identifies component presence while isolating complete, fine-grained component regions within O_a.

3.2 HIERARCHICAL GRAPH CONSTRUCTION

To capture local geometric primitives and global semantic topology, we construct bottom-up hierarchical graphs, as shown in Fig. 2. Driven by keypoints extracted from the topological skeleton, we progressively build the Segment Graph, Sub-stroke Graph, and Stroke Graph. Nodes and edges across all levels are assigned explicit attribute vectors encoding local geometry, stroke continuity, and morphological topology. The same construction pipeline is applied symmetrically to both character and component images.

3.2.1 SKELETON KEYPOINT EXTRACTION

Given an OBI character O_a or a component C_b image, we first binarize it and apply the Zhang–Suen algorithm Zhang & Suen (1984) to extract a one-pixel-wide topological skeleton. For any skeleton pixel p, let $\mathcal{N}_8(p) = \{\xi_1, \ldots, \xi_8\}$ denote its 8-connected neighbors in clockwise order, where $\xi_i \in \{0, 1\}$ represents background (0) or skeleton (1). The topological state of p is defined by its degree $\deg(p) = \sum_{i=1}^{8} \xi_i$ and neighborhood state transitions $\Psi(p) = \sum_{i=1}^{8} |\xi_i - \xi_{i+1}|$ (with $\xi_9 = \xi_1$).

Based on $\deg(p)$ and $\Psi(p)$, skeleton pixels are categorized into four topological keypoints sets. **Endpoints($\mathcal{P}_{\text{end}}$)** are pixels satisfying $\deg(p) = 1$ and $\Psi(p) = 2$. **Junction points ($\mathcal{P}_{\text{junc}}$)** are pixels satisfying $\deg(p) \geq 3$ and $\Psi(p) \in \{6, 8\}$. **Inflection points ($\mathcal{P}_{\text{infl}}$)** are non-junction pixels whose local tangent orientation change exceeds a threshold θ_{infl}. **Loop anchors ($\mathcal{P}_{\text{anc}}$)** consist of three pixels uniformly sampled along each closed loop lacking natural endpoints or junction points to anchor its topology. The complete keypoint set is defined as $\mathcal{P} = \mathcal{P}_{\text{end}} \cup \mathcal{P}_{\text{junc}} \cup \mathcal{P}_{\text{infl}} \cup \mathcal{P}_{\text{anc}}$.

3.2.2 BOTTOM-UP GRAPH CONSTRUCTION

Let $r \in \{\text{seg}, \text{sub}, \text{str}\}$ index the segment, sub-stroke, and stroke levels, respectively, where each graph is denoted as $G^r = (V^r, E^r)$. The **Segment Graph** G^{seg} is constructed directly over the keypoint set $\mathcal{P}$, which serves as the node set V^{seg}. An edge $e = (u, v) \in E^{\text{seg}}$ is added if $u, v \in \mathcal{P}$ are connected by a skeleton path containing no intermediate keypoints. The **Sub-stroke Graph** G^{sub} is constructed over G^{seg} to model local geometric continuity. Inflection points $\mathcal{P}_{\text{infl}}$ are absorbed by merging adjacent segment edges into sub-stroke edges E^{sub}. The **Stroke Graph** G^{str} further abstracts G^{sub} to capture complete semantic strokes. Approximately collinear edges sharing junction points $w \in V^{\text{sub}}$ are merged, with w removed from the node set. We denote $\mathcal{U}(e) \subseteq E^{\text{seg}}$ as the set of underlying segment edges covered by a higher-level edge $e \in E^{\text{sub}} \cup E^{\text{str}}$.

For any node $v \in V^r$, its initial feature vector is formulated as $\mathbf{x}_v = [(\hat{x}_v, \hat{y}_v), \mathbf{t}_v, \deg(v)]$, concatenating normalized spatial coordinates $(\hat{x}_v, \hat{y}_v) \in [0, 1]^2$, a one-hot keypoint type encoding $\mathbf{t}_v$, and node degree $\deg(v)$. Edge attributes provide complementary geometric and topological descriptions across levels. The segment-edge vector $\mathbf{e}^{\text{seg}}$ models local geometry, encoding normalized path length, endpoint Euclidean distance, path efficiency defined as the distance-to-length ratio, curvature statistics via extreme values and variance, and axial direction $(\cos 2\alpha, \sin 2\alpha)$ for chord angle α.

Higher-level attributes $\mathbf{e}^{\text{sub}}$ and $\mathbf{e}^{\text{str}}$ aggregate geometric features from underlying segments $\mathcal{U}(e)$ and incorporate global stroke features, including covered segment count $|\mathcal{U}(e)|$, cumulative and maximum inflection angles, boundary-closure proximity, and a loop factor distinguishing closed loops from open curves.

3.3 Hierarchical Graph Matching

Given a pair of hierarchical graphs $(G^r_{O_a}, G^r_{C_b})$, HiGAS infers structural correspondences between character O_a and component C_b. Within-level edge matching first establishes basic structural alignments. Top-down topological priors from higher-level strokes are then introduced to enhance segment-edge matching, mitigating local overfitting under weak supervision. Finally, a component-scale dynamic edge selection strategy aggregates global topological consistency to predict component presence and generate fine-grained segmentation results.

3.3.1 Within-Level Edge Matching

To establish structural alignment, we independently compute edge-level topological and geometric matching scores at each graph level. A three-layer GINE encoder Hu et al. (2020) serves as the structural backbone to jointly encode topological and geometric features during message passing. For character node $u \in V^r_{O_a}$ and component node $i \in V^r_{C_b}$, let $\mathbf{h}^r_{O_a,u}, \mathbf{h}^r_{C_b,i} \in \mathbb{R}^{D_h}$ denote their D_h-dimensional structural representations. Their structural similarity z^r_{ui} is formulated via a projected dot product:

$$z^r_{ui} = \frac{\left(\mathbf{W}^r_O \mathbf{h}^r_{O_a,u} + \mathbf{b}^r_O\right)^\top \left(\mathbf{W}^r_C \mathbf{h}^r_{C_b,i} + \mathbf{b}^r_C\right)}{\sqrt{d_r}}, \tag{1}$$

where $\mathbf{W}^r_O, \mathbf{W}^r_C \in \mathbb{R}^{d_r \times D_h}$ and $\mathbf{b}^r_O, \mathbf{b}^r_C \in \mathbb{R}^{d_r}$ denote learnable projection parameters, and $\sqrt{d_r}$ scales the inner product for numerical stability in d_r-dimensional subspace.

For an undirected character edge $e_O = (u, v) \in E^r_{O_a}$ and component edge $e_C = (i, j) \in E^r_{C_b}$, we aggregate the two candidate endpoint alignment configurations using LogSumExp to define the topological correspondence score:

$$p^r_{e_O,e_C} = \log\left[\exp\left(z^r_{ui} + z^r_{vj}\right) + \exp\left(z^r_{vi} + z^r_{uj}\right)\right]. \tag{2}$$

Concurrently, geometric consistency $f^r_{e_O,e_C}$ is evaluated as the negative mean L_1 distance between edge attribute vectors over a selected feature index set $\mathcal{D}_r$:

$$f^r_{e_O,e_C} = -\frac{1}{|\mathcal{D}_r|} \sum_{k \in \mathcal{D}_r} \left| e^r_{u,v,k} - e^r_{i,j,k} \right|. \tag{3}$$

Combining topological alignment and geometric similarity yields the joint edge similarity:

$$m^r_{e_O,e_C} = p^r_{e_O,e_C} + \lambda_f f^r_{e_O,e_C}, \tag{4}$$

where λ_f is a hyperparameter to balance the geometric constraint. Computing $m^r_{e_O,e_C}$ across all edge pairs forms the pairwise similarity matrix $\mathbf{M}^r = [m^r_{e_O,e_C}]_{e_O \in E^r_{O_a}, e_C \in E^r_{C_b}}$. For each character edge $e_O \in E^r_{O_a}$, its level-r edge matching score $q^r_{e_O}$ is obtained via normalized LogSumExp pooling over its joint similarities with all candidate component edges $e_C \in E^r_{C_b}$, avoiding score scale variation caused by differing component graph sizes.

3.3.2 Cross-Level Structural Enhancement

While the Segment Graph G^{seg} captures precise local topology, its restricted receptive field renders local edges susceptible to background clutter and ambiguous matches under weak supervision. Conversely, higher-level graphs G^{sub} and G^{str} aggregate segment edges into stroke-level structural edges, providing richer contextual constraints and stronger structural discriminability. We propagate top-down matching priors from higher levels to enhance segment edges with low local confidence.

For a character segment edge $e_O = (u, v) \in E^{\text{seg}}_{O_a}$, we enhance its matching score by fusing positive structural evidence from its covering parent edges $\pi_{\text{sub}}(e_O) \in E^{\text{sub}}_{O_a}$ and $\pi_{\text{str}}(e_O) \in E^{\text{str}}_{O_a}$

$$\widetilde{q}_{e_O} = q^{\text{seg}}_{e_O} + \eta_{\text{sub}} \left[q^{\text{sub}}_{\pi_{\text{sub}}(e_O)} \right]_+ + \eta_{\text{str}} \left[q^{\text{str}}_{\pi_{\text{str}}(e_O)} \right]_+, \tag{5}$$

where $[\cdot]_+ = \max(\cdot, 0)$ suppresses negative clutter, and η_{sub} and η_{str} control the enhancement intensities across hierarchical levels.

3.3.3 CHARACTER–COMPONENT GRAPH MATCHING

To measure global structural correspondence between character O_a and component C_b, we compute a global topological consistency score $S_{a,b}$ from the enhanced segment score pool $\{\widetilde{q}_{e_O}^{seg} \mid e_O \in E_{O_a}^{\text{seg}}\}$. Since a component typically aligns with only a local subgraph of the character, unconstrained pooling over all character edges introduces noise from irrelevant structures, while fixed-cardinality selection fails to accommodate varying component scales and topological complexities.

To resolve this, we propose a component-scale dynamic edge selection strategy. Let $N_{O_a} = |E_{O_a}^{\text{seg}}|$ and $N_{C_b} = |E_{C_b}^{\text{seg}}|$ denote the segment edge counts of the character and component graphs. The target selected size $K_{a,b}$ is determined by:

$$K_{a,b} = \max\left\{1, \min\left\{N_{O_a}, \operatorname{round}\left(\rho N_{C_b}\right)\right\}\right\}, \tag{6}$$

where the scaling factor ρ accommodates local glyph deformations. Ranking character segment edges in descending order of $\widetilde{q}_{e_O}^{a,b}$, the top $K_{a,b}$ edges form the alignment subset $\mathcal{T}_{a,b}$. The global topological consistency score $S_{a,b}$ is then obtained by applying normalized LogSumExp pooling over the scores in $\mathcal{T}_{a,b}$.

Finally, applying the sigmoid function to $S_{a,b}$ predicts the component presence probability $\widehat{y}_{a,b} = \sigma\left(S_{a,b}\right)$ for candidate $C_b \in \mathcal{C}$. For high-confidence components satisfying $\widehat{y}_{a,b} > \delta_{\text{cls}}$, HiGAS retains all segment edges with probability $P_{e_O}^{a,b} = \sigma\left(\widetilde{q}_{e_O}^{a,b}\right) > \delta_{\text{seg}}$ to construct the final segmentation mask, where δ_{cls} and δ_{seg} denote the presence and edge segmentation thresholds, respectively.

3.4 JOINT OPTIMIZATION OBJECTIVE

To train HiGAS end-to-end under weak supervision without pixel-level annotations, we design a joint optimization objective combining component presence classification loss and component segmentation completeness. During training, the model takes a character–component pair (O_a, C_b) with a binary label $y_{a,b} \in \{0, 1\}$ indicating whether O_a contains C_b, where $y_{a,b} = 1$ and $y_{a,b} = 0$ denote positive and negative pairs, respectively.

3.4.1 COMPONENT PRESENCE CLASSIFICATION LOSS

To provide hierarchical supervision and boost component discriminability across all levels, we obtain single-level topology alignment scores $\{S_{a,b}^{r} \mid r \in \{\text{seg}, \text{sub}, \text{str}\}\}$ by using the same normalized LogSumExp pooling at each level. Combining single-level scores with the topology-enhanced consistency score $S_{a,b}$, the classification loss $\mathcal{L}_{\text{cls}}$ is formulated via binary cross-entropy (BCE):

$$\mathcal{L}_{\text{cls}} = \lambda_{\text{mil}}\text{BCE}(\sigma(S_{a,b}), y_{a,b}) + \lambda_{\text{seg}}\text{BCE}(\sigma(S_{a,b}^{\text{seg}}), y_{a,b}) + \frac{\lambda_{\text{h}}}{2}\sum_{r\in\{\text{sub},\text{str}\}}\text{BCE}(\sigma(S_{a,b}^{r}), y_{a,b}), \tag{7}$$

where $\sigma(\cdot)$ denotes the sigmoid function, and λ_{mil}, λ_{seg}, λ_{h} balance loss terms across hierarchical levels.

3.4.2 COMPONENT SEGMENTATION COMPLETENESS LOSS

The segmentation completeness objective encourages extracting structurally complete component edge sets via a structural topology completeness loss $\mathcal{L}_{\text{st}}$ and a hierarchical group consistency loss $\mathcal{L}_{\text{grp}}$.

To mitigate dependence on sparse local features and recover less discriminative component edges under weak supervision, $\mathcal{L}_{\text{st}}$ ensures every segment edge in component C_b receives matching support within character O_a. Specifically, we apply normalized LogSumExp pooling to each column of the segment similarity matrix and average the column scores to obtain structural completeness score $S_{\text{st},a,b}$, and define $\mathcal{L}_{\text{st}}$ as:

$$\mathcal{L}_{\text{st}} = y_{a,b}\,\text{softplus}\left(\tau_{+} - S_{\text{st},a,b}\right) + \left(1 - y_{a,b}\right)\text{softplus}\left(S_{\text{st},a,b} - \tau_{-}\right), \tag{8}$$

where $\tau_{+} = 0$ and $\tau_{-} = -1$ set the margins for positive and negative pairs, respectively.

For positive pairs, $\mathcal{L}_{\text{grp}}$ enforces consistent responses among segment edges belonging to the same higher-level stroke. We define an adaptive weight $w_{e_O} = 2\sigma(\text{sg}(\widetilde{q}_{e_O}^{a,b} - q_{e_O}^{\text{seg}})) - 1 \in [0, 1)$, where

$\mathrm{sg}(\cdot)$ stops gradients. A larger w_{e_O} reflects stronger top-down support from higher levels. Summing over character segment edges $e_O \in E_{O_a}^{\mathrm{seg}}$, the group consistency loss is formulated as:

$$\mathcal{L}_{\mathrm{grp}} = \frac{\sum_{e_O} w_{e_O} \,\mathrm{softplus}(\gamma - \widetilde{q}_{e_O}^{a,b})}{\sum_{e_O} w_{e_O} + \varepsilon}, \tag{9}$$

where γ denotes the expected lower activation bound for enhanced segment edges, and ε prevents division by zero. When strong higher-level priors are propagated but covered segment edges exhibit scores $\widetilde{q}_{e_O}^{a,b} < \gamma$, $\mathcal{L}_{\mathrm{grp}}$ penalizes these unaligned edges, thereby smoothing local segment responses and maintaining stroke topological continuity.

Overall Optimization Objective. Combining classification objective and segmentation completeness constraints yields the overall training loss:

$$\mathcal{L} = \mathcal{L}_{\mathrm{cls}} + \mathcal{L}_{\mathrm{st}} + \lambda_{\mathrm{grp}} y_{a,b} \mathcal{L}_{\mathrm{grp}}, \tag{10}$$

with hyperparameter λ_{grp} balancing the group consistency constraint.

4 Experiment

4.1 Dataset

Our experiments are conducted on two OBI datasets. **OBI Component-20 Dataset** Hu et al. (2024) contains $9,245$ character images across 20 component categories, each annotated with the presence of a single target component. Additionally, OBI experts annotated $1,012$ selected images with pixel-level ground truth to benchmark segmentation performance. To support multi-component recognition and expand component diversity, we construct the **OBI Component-50 Dataset**. Leveraging character–component correspondences from Yinqi Wenyuan[1], we extend the original 20 categories with 30 high-frequency components, establishing a set of 50 standard components. Corresponding character images are retrieved from EVOBC Guan et al. (2024a), compiling $78,819$ images across $1,188$ categories with multi-component annotations. The dataset comprises $37,238$ single-component, $33,283$ two-component, $7,504$ three-component, and 794 four-component images.

4.2 Evaluation Metrics

Model performance is evaluated on component presence prediction and fine-grained stroke segmentation. Candidate components are ranked by predicted probabilities $\widehat{y}_{a,b}$. **Hit@k** measures the proportion of characters with at least one ground-truth component retrieved in the top-k candidates. **Complete Recall CR@$(n+m)$** calculates the proportion of characters where all n ground-truth components are covered within top $n+m$ candidates (m is a tolerance budget). Since expert annotations in OBI Component-20 cover only central stroke skeletons, we expand the ground truth via 12-pixel topological dilation. Using this enhanced ground truth, segmentation accuracy is evaluated using **Intersection over Union (IoU)**, **Precision**, **Recall**, and **F1-Score**.

4.3 Comparison Methods

For component presence prediction, we compare HiGAS against Component-Level OBI Retrieval (CL-OBIR) Hu et al. (2024), an image-based retrieval model, and a DINOv2-based baseline Zhang et al. (2026) with a frozen backbone and a trained multi-label classification head. All models are trained and evaluated on the Component-50 dataset, which contains $66,996$ training and $11,823$ test images.

For component segmentation, we compare against Component-Level OBI Segmentation (CL-OBIS) Hu et al. (2025), a stroke-rule augmented model, and Puzzle Pieces Picker (P^3) Wang et al. (2024a), a decipherment model via component decomposition. While both baselines require pixel-level annotations and are trained on 708 pixel-annotated images from Component-20, HiGAS is trained without using any Component-20 images or annotations. All methods are evaluated on the same held-out set of 304 pixel-annotated Component-20 images.

[1] https://www.jgwlbq.org.cn/

Table 1: Baseline comparison for component presence prediction and component segmentation.

(a) Presence prediction

Method	Hit@1	Hit@3	Hit@5
CL-OBIR	54.28%	76.01%	83.46%
DINOv2-based baseline	67.68%	87.96%	92.92%
HiGAS (Ours)	**80.97%**	**94.07%**	**96.77%**

(b) Segmentation

Method	IoU	Prec.	Rec.	F1
CL-OBIS	34.81%	72.04%	41.08%	50.93%
P^3	42.25%	63.16%	56.22%	57.53%
HiGAS (Ours)	**64.35%**	**92.65%**	**68.92%**	**77.58%**

Table 2: Complete recall across varying component counts on Component-50.

# Components	Method	CR@n	CR@$(n+1)$	CR@$(n+2)$	CR@$(n+3)$	CR@$(n+4)$
$n=1$	CL-OBIR	47.95%	61.50%	68.31%	72.70%	76.47%
	DINOv2-based baseline	59.20%	74.57%	81.63%	85.68%	88.25%
	HiGAS (Ours)	**71.53%**	**83.99%**	**89.56%**	**92.15%**	**93.87%**
$n=2$	CL-OBIR	10.85%	22.27%	30.57%	38.26%	44.29%
	DINOv2-based baseline	32.54%	51.07%	62.34%	69.77%	75.26%
	HiGAS (Ours)	**59.75%**	**76.54%**	**84.52%**	**88.97%**	**91.69%**
$n=3$	CL-OBIR	3.93%	7.69%	10.81%	15.64%	20.55%
	DINOv2-based baseline	17.52%	31.55%	42.54%	51.47%	58.71%
	HiGAS (Ours)	**47.27%**	**68.99%**	**81.05%**	**88.56%**	**91.33%**
$n=4$	CL-OBIR	0.00%	5.88%	11.76%	16.81%	23.53%
	DINOv2-based baseline	10.92%	27.73%	36.13%	43.70%	51.26%
	HiGAS (Ours)	**47.90%**	**62.18%**	**78.99%**	**84.87%**	**88.24%**

4.4 Implementation Details

HiGAS is implemented using PyTorch 1.11 and PyTorch Geometric 2.1 and trained on a single NVIDIA A100 GPU (40GB) for 200 epochs with AdamW, a batch size of 256, an initial learning rate of 10^{-3}, and weight decay of 10^{-4}. Automatic mixed precision, gradient clipping at 5.0, and random seed 42 are used; all positive components are retained while negative components are dynamically sampled each epoch. We set $\theta_{\text{infl}} = 0.80$, $\lambda_f = 0.35$ at all three levels, $\eta_{\text{sub}} = \eta_{\text{str}} = 0.15$, and $\rho = 1.25$. The loss weights are $(\lambda_{\text{mil}}, \lambda_{\text{seg}}, \lambda_{\text{h}}, \lambda_{\text{grp}}) = (1.0, 0.1, 0.1, 0.05)$, and $\delta_{\text{cls}} = \delta_{\text{seg}} = 0.5$.

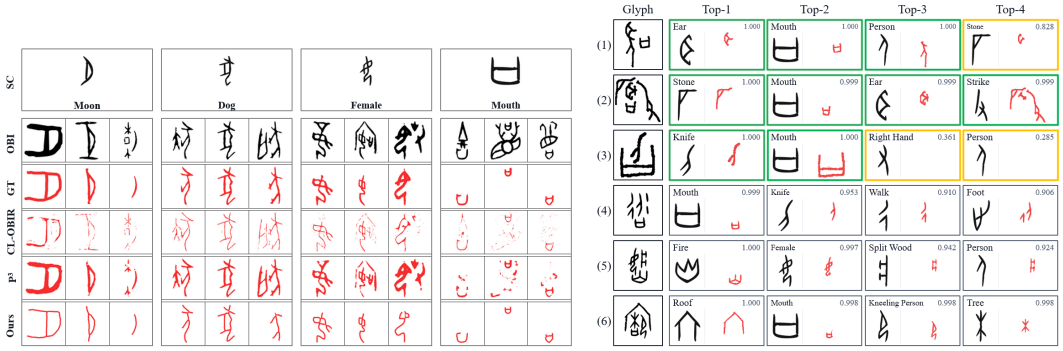
4.5 Accuracy of Component Presence Prediction

As shown in Table 1(a), HiGAS outperforms both baselines across all Hit@k metrics. Hit@1 reaches 80.97%, exceeding CL-OBIR and DINOv2 by 26.69 and 13.29 percentage points, respectively, confirming high fidelity of the top-ranked prediction; hierarchical graph representations further reduce structural ambiguity among visual components. Hit@5 reaches 96.77%, placing almost all true components within the top five candidates and reliably feeding subsequent segmentation.

Table 2 evaluates complete recall under different ground-truth component counts n. Baseline recall degrades sharply as n increases, whereas HiGAS remains robust across multi-component groups. With modest candidate expansion, HiGAS gains substantial recall at every complexity level, indicating that missed true components remain highly ranked and comprehensive component recall is achieved with minimal candidate redundancy.

4.6 Component Segmentation Performance

Conditional segmentation is evaluated on high-confidence characters with presence probabilities above δ_{cls}, covering 93.09% of the test set, with all methods evaluated on the same subset. Table 1(b) shows that HiGAS reaches 64.35% IoU and 77.58% F1. CL-OBIS misses local strokes and has only 41.08% recall, while P^3 attains 63.16% precision, indicating over-segmentation. Cross-level structural enhancement and the completeness loss therefore recover more complete strokes while maintaining high precision and recall.



(a) Component-variant segmentation comparison (b) Top-4 prediction and segmentation

Figure 3: Qualitative segmentation results. In (b), Rows 1–3 are deciphered characters and Rows 4–6 are undeciphered characters, green/yellow borders indicate correct/additional predictions.

Figure 3(a) shows robust segmentation of component variants despite changes in stroke trajectories, inflection points, and connectivity: HiGAS reconstructs target components, whereas CL-OBIS omits continuous strokes and P^3 includes non-target strokes, supporting standard components as effective topological references. Figure 3(b) reports top-4 results on deciphered and undeciphered multi-component characters; each candidate box shows the standard component and predicted probability. HiGAS prioritizes true components, and the average probability drops from 91.8% to 73.1% after all ground-truth components are recalled, revealing a confidence gap that filters irrelevant candidates while providing plausible hypotheses for undeciphered characters.

4.7 Ablation Study

Table 3: Ablation results for structural enhancement and completeness losses (%).

Method	Hit@1	Hit@3	Hit@5	IoU	Prec.	Rec.	F1
HiGAS	80.97	94.07	96.77	**64.35**	**92.65**	**68.92**	**77.58**
w/o G^{sub}	79.38	93.58	96.50	57.70	87.85	63.52	71.06
w/o G^{str}	79.79	93.61	96.42	54.30	82.79	60.12	67.00
only G^{seg}	77.48	92.64	95.88	51.25	86.41	54.76	64.20
w/o $\mathcal{L}_{\text{st}}$	**81.05**	**94.44**	**96.90**	51.38	83.92	54.26	63.86
w/o $\mathcal{L}_{\text{grp}}$	79.57	93.55	96.25	49.59	86.46	52.79	62.87
only $\mathcal{L}_{\text{cls}}$	80.37	94.29	96.74	44.47	87.27	46.46	57.61

Table 3 evaluates structural enhancement and completeness losses. Removing G^{sub}, G^{str}, or both higher-level graphs degrades performance; using only G^{seg} gives 77.48% Hit@1 and 51.25% IoU, while removing G^{str} lowers IoU by 10.05 percentage points, showing its importance for global topology. G^{sub} complements it with fine-grained stroke-turn cues, and the full three-level model performs best overall. For completeness losses, removing $\mathcal{L}_{\text{st}}$ or $\mathcal{L}_{\text{grp}}$ reduces IoU from 64.35% to 51.38% and 49.59%, respectively. Omitting $\mathcal{L}_{\text{st}}$ slightly improves Hit@k by relaxing segmentation constraints but harms completeness; using only $\mathcal{L}_{\text{cls}}$ further lowers IoU to 44.47% and recall to 46.46%, confirming that both completeness terms are needed to recover full component strokes.

5 Conclusion

HiGAS performs weakly supervised OBI component presence prediction and fine-grained segmentation by aligning hierarchical graphs of characters and standard components. Cross-level structural enhancement, dynamic edge selection, and joint optimization enable accurate multi-component prediction and segmentation using only image-level presence labels. Future work will extend HiGAS to broader component categories and more complex glyph variants to better assist expert decipherment.

REFERENCES

Xiang Chang, Fei Chao, Changjing Shang, and Qiang Shen. Sundial-gan: A cascade generative adversarial networks framework for deciphering oracle bone inscriptions. In *Proceedings of the 30th ACM international conference on multimedia*, pp. 1195–1203, 2022.

Zijian Chen, Wenjun Zhang, Guangtao Zhai, et al. Obi-bench: Can lmms aid in study of ancient script on oracle bones? In *International Conference on Learning Representations*, volume 2025, pp. 102851–102881, 2025.

Zijian Chen, Wenjie Hua, Jinhao Li, Lirong Deng, Fan Du, Tingzhu Chen, and Guangtao Zhai. Pictobi-20k: Unveiling large multimodal models in visual decipherment for pictographic oracle bone characters. In *ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 11937–11941. IEEE, 2026a.

Zijian Chen, Wenjie Hua, Jinhao Li, Yucheng Zhu, Xiaona Zhi, Zhiji Liu, Tingzhu Chen, Wenjun Zhang, and Guangtao Zhai. Oracle bone inscriptions information processing: a comprehensive survey. *npj Heritage Science*, 14(1):220, 2026b.

Yang Chi, Fausto Giunchiglia, Daqian Shi, Xiaolei Diao, Chuntao Li, and Hao Xu. Zinet: Linking chinese characters spanning three thousand years. In *Findings of the association for computational linguistics: ACL 2022*, pp. 3061–3070, 2022.

Yang Chi, Fausto Giunchiglia, Chuntao Li, and Hao Xu. Ancient chinese glyph identification powered by radical semantics. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 12065–12074, 2024.

Xiaolei Diao, Daqian Shi, Jian Li, Lida Shi, Mingzhe Yue, Ruihua Qi, Chuntao Li, and Hao Xu. Toward zero-shot character recognition: a gold standard dataset with radical-level annotations. In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 6869–6877, 2023a.

Xiaolei Diao, Daqian Shi, Hao Tang, Qiang Shen, Yanzeng Li, Lei Wu, and Hao Xu. Rzcr: Zero-shot character recognition via radical-based reasoning. In Edith Elkind (ed.), *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, pp. 654–662. International Joint Conferences on Artificial Intelligence Organization, 8 2023b. doi: 10.24963/ijcai.2023/73. URL https://doi.org/10.24963/ijcai.2023/73. Main Track.

Xiaolei Diao, Zhihan Zhou, Lida Shi, Ting Wang, Ruihua Qi, Daqian Shi, and Hao Xu. Intechar: A unified oracle bone character list for ancient chinese language modeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 211–219, 2026.

Haisu Guan, Jinpeng Wan, Yuliang Liu, Pengjie Wang, Kaile Zhang, Zhebin Kuang, Xinyu Wang, Xiang Bai, and Lianwen Jin. An open dataset for the evolution of oracle bone characters: EVOBC, 2024a.

Haisu Guan, Huanxin Yang, Xinyu Wang, Shengwei Han, Yongge Liu, Lianwen Jin, Xiang Bai, and Yuliang Liu. Deciphering oracle bone language with diffusion models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 15554–15567, 2024b.

Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. In *International Conference on Learning Representations*, 2020.

Zhikai Hu, Yiu-ming Cheung, Yonggang Zhang, Peiying Zhang, and Pui-ling Tang. Component-level oracle bone inscription retrieval. In *Proceedings of the 2024 International Conference on Multimedia Retrieval*, pp. 647–656, 2024.

Zhikai Hu, Yiu-ming Cheung, Yonggang Zhang, Zhang Peiying, and Tang Pui Ling. Component-level segmentation for oracle bone inscription decipherment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 28116–28124, 2025.

Runhua Jiang, Yongge Liu, Boyuan Zhang, Xu Chen, Deng Li, and Yahong Han. Oraclepoints: A hybrid neural representation for oracle character. In *Proceedings of the 31st ACM international conference on multimedia*, pp. 7901–7911, 2023.

Qingju Jiao, Jingwen Wu, Qi Liu, Han Zhang, Zhan Zhang, Bang Li, Jing Xiong, Guoying Liu, and Yongge Liu. A graph-based evolutionary dataset for oracle bone characters from inscriptions to modern chinese scripts. *npj Heritage Science*, 13(1):369, 2025.

Caoshuo Li, Zengmao Ding, Xiaobin Hu, Bang Li, Donghao Luo, AndyPian Wu, Chaoyang Wang, Chengjie Wang, Taisong Jin, Seven Shu, et al. Oraclefusion: Assisting the decipherment of oracle bone script with structurally constrained semantic typography. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 19893–19902, 2025a.

Jing Li, Qiufeng Wang, Siyuan Wang, Rui Zhang, Kaizhu Huang, and Erik Cambria. Diff-oracle: Learning styles and contents to augment realistic oracle characters in diffusion model. *ACM Transactions on Multimedia Computing, Communications and Applications*, 22(6):1–24, 2026.

Jinhao Li, Zijian Chen, Runze Jiang, Tingzhu Chen, Changbo Wang, and Guangtao Zhai. Mitigating long-tail distribution in oracle bone inscriptions: Dataset, model, and benchmark. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 7729–7738, 2025b.

Fuwen Luo, Zihao Wan, Ziyue Wang, Yaluo Liu, Pau Tong Lin Xu, Xuanjia Qiao, Xiaolong Wang, Peng Li, and Yang Liu. Enabling stroke-level structural analysis of hieroglyphic scripts without language-specific priors. In Maria Liakata, Viviane P. Moreira, Jiajun Zhang, and David Jurgens (eds.), *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 27794–27810, San Diego, California, United States, July 2026a. Association for Computational Linguistics. ISBN 979-8-89176-395-1. doi: 10.18653/v1/2026.findings-acl.1383. URL https://aclanthology.org/2026.findings-acl.1383/.

Tingzhang Luo, Yichao Liu, Yulin Xu, Ruizhong Liu, Xin Wang, Haijin Zeng, Shaoguang Huang, and Hongyan Zhang. Stroke-based perception: Discover novel oracle characters. *IEEE Transactions on Multimedia*, 2026b.

Runqi Qiao, Lan Yang, Kaiyue Pang, and Honggang Zhang. Making visual sense of oracle bones for you and me. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12656–12665, 2024.

Runqi Qiao, Qiuna Tan, Guanting Dong, MinhuiWu MinhuiWu, Jiapeng Wang, Yifan Zhang, Zhuoma GongQue, Chong Sun, Yida Xu, Yadong Xue, et al. V-oracle: Making progressive reasoning in deciphering oracle bones for you and me. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 20124–20150, 2025.

Rui Song, Lida Shi, Ruihua Qi, Yingji Li, and Hao Xu. Enhancing multimodal large language models for ancient chinese character evolution analysis via glyph-driven fine-tuning. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 22789–22804, 2026.

Mei Wang, Weihong Deng, and Cheng-Lin Liu. Unsupervised structure-texture separation network for oracle character recognition. *IEEE Transactions on Image Processing*, 31:3137–3150, 2022.

Pengjie Wang, Kaile Zhang, Xinyu Wang, Shengwei Han, Yongge Liu, Lianwen Jin, Xiang Bai, and Yuliang Liu. Puzzle pieces picker: Deciphering ancient chinese characters with radical reconstruction. In *International Conference on Document Analysis and Recognition*, pp. 169–187. Springer, 2024a.

Pengjie Wang, Kaile Zhang, Xinyu Wang, Shengwei Han, Yongge Liu, Jinpeng Wan, Haisu Guan, Zhebin Kuang, Lianwen Jin, Xiang Bai, et al. An open dataset for oracle bone character recognition and decipherment. *Scientific Data*, 11(1):976, 2024b.

Yin Wu, Zhengxuan Zhang, Jiayu Chen, Chang Xu, Yuyu Luo, Nan Tang, and Hui Xiong. Explainable oracle bone script recognition via multimodal pictographic reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 39460–39467, 2026.

Zhicong Wu, Qifeng Su, Ke Gu, and Xiaodong Shi. A cross-font image retrieval network for recognizing undeciphered oracle bone inscriptions. In *International Conference on Intelligent Computing*, pp. 196–208. Springer, 2025.

Jiang Yuan, Shanxiong Chen, Bofeng Mo, Yuqi Ma, Wenjun Zheng, and Chongsheng Zhang. R-gnn: recurrent graph neural networks for font classification of oracle bone inscriptions. *Heritage Science*, 12(1):1–14, 2024.

Jianing Zhang, Runan Li, Honglin Pang, Ding Xia, Zhou Zhu, Qian Zhang, Chuntao Li, and Xi Yang. Specializing large models for oracle bone script interpretation via component-grounded multimodal knowledge augmentation. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 35221–35236, 2026.

Tongjie Y Zhang and Ching Y. Suen. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 27(3):236–239, 1984.

Zhan Zhang, Hanbin Liu, Xingkun Zhang, Yiyuan Wang, Feng Gao, An Guo, Han Zhang, Qingju Jiao, Bang Li, and Yongge Liu. Glyph graph isomorphism network for structure recognition of oracle bone inscription. *Expert Systems with Applications*, pp. 129519, 2025.

Sheng Zhuo and Jiangshe Zhang. Attention-based deformable convolutional network for chinese various dynasties character recognition. *Expert Systems with Applications*, 238:121881, 2024.