Oracle Bone Inscription Detection with Multi-Branch Feature Fusion and Efficient Attention*

Jiaze Cai[1], Qi Li[1,*], Xuexue Zhu[2], Bang Li[3], Xin Yan[4] and Xia Zhang[4]

[1]*College of Science and Engineering, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga, 525-8577, Japan*
[2]*School of computer & information engineering, Anyang Normal University, Anyang, 455000, China*
[3]*Key Laboratory of Oracle Bone Inscriptions Information Processing, Anyang Normal University, Anyang, 455000, China*
[4]*State Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing, 100876, China*

Abstract

Oracle bone inscription (OBI) detection is hindered by tiny character size, complex morphological variation, and severe interference from surface erosion, cracks, and background noise in ancient rubbings. To overcome these challenges, we present an enhanced YOLO framework that marries multi-scale feature fusion with efficient attention, enabling robust OBI character detection. Our core innovation is the TriFusion Block (TFB), which synergistically combines three parallel branches: spatial attention for global context modeling, global modeling for semantic feature extraction, and sequential processing for efficient dependency capture. This design enables the network to simultaneously extract fine-grained local details and long-range structural patterns with minimal computational overhead. Extensive experiments on the Oracle-Bone Inscriptions Multimodal Dataset show that the proposed method improves the baseline YOLOv8n by 2.33 % in recall, 8.03 % in precision, and 3.96 % in mAP@0.5, achieving final scores of 81.40 % recall, 91.30 % precision, and 86.35 % mAP@0.5.

Keywords

Oracle Bone Inscriptions, Multi-scale Feature Fusion, YOLOv8 Ancient Script Detection

1. Introduction

Oracle bone inscriptions (OBI) are the oldest attested form of Chinese writing, carved more than three millennia ago, mainly in turtle plastrons and animal scapulae, during the late Shang dynasty[1]. As rare physical artefacts, OBI embody a wealth of historical information and cultural significance. Their distinctive orthography and textual content offer first-hand evidence for tracing the evolution of Chinese characters and for exploring early Chinese social structures, religious practices, and historical events. Consequently, the accurate detection of OBIs is not

The 7th International Symposium on Advanced Technologies and Applications in the Internet of Things (ATAIT 2025), September 10-11, 2025, Kusatsu, Japan

[*] You can use this document as the template for preparing your publication. We recommend using the latest version of the ceurart style.

[*] Corresponding author.

caijiaze11@gmail.com (J. Cai); liqi24@fc.ritsumei.ac.jp (Q. Li); 2418516361@qq.com (X. Zhu); libang@aynu.edu.cn (B. Li); 2418516361@qq.com (X. Yan); 2418516361@qq.com (X. Zhang)
0009-0000-7171-8667 (J. Cai); 0000-0002-1963-5263 (Q. Li); 0009-0008-3445-0495 (X. Zhu); 0000-0002-7254-6569 (B. Li); 0009-0008-3445-0495 (X. Yan); 0009-0008-3445-0495 (X. Zhang)

merely a technical challenge in palaeography and archaeology, but a foundational task for data-driven reconstructions of Shang-era civilisation.

In recent years, computer vision and deep-learning techniques have advanced rapidly. Convolutional-neural-network (CNN)[2] object-detectors—especially the YOLO (You Only Look Once) series—have achieved notable success in image detection[3, 4]. Integrating these deep-learning detectors into OBI research can substantially improve detection accuracy and efficiency while reducing manual labour, thereby accelerating the large-scale digitisation of oracle-bone materials. These advances will support high-quality digital repositories, facilitate scholarly decipherment, and lay a foundation for cross-disciplinary data mining and intelligent analysis, giving the endeavour substantial theoretical and practical value.

Nevertheless, OBI detection still faces substantial challenges. Most available images are rubbings or fragmented pieces whose quality varies widely and often suffers from blur, surface damage, and geometric distortion. Each character is a minute target easily occluded by cracks, abrasion, and background noise; even lightweight detectors still produce numerous misses and false positives on high-resolution inputs. Recent advances in efficient attention mechanisms[5, 6] and state-space modeling[7] have demonstrated promising capabilities for handling such complex scenarios with reduced computational overhead, yet their application to oracle bone character detection remains underexplored.

This study tackles the detection of minute, heavily eroded oracle-bone characters by introducing a multi-scale enhancement module that fuses global attention, content clustering, and state-space modeling. The module is inserted into both the YOLOv8n backbone and the SPPF block, enabling the network to capture global, local, and sequential cues with minimal computational overhead. A systematic study of insertion stages and fusion weights shows that the resulting model yields substantial accuracy gains in OBI detection, confirming the practicality and portability of this plug-and-play, module-level paradigm for low-resource ancient-script tasks.

Overall, our main contributions can be summarized below:

- Within YOLOv8n, we introduce a TriFusion Block (TFB) that fuses global attention, content clustering, and state-space cues in a single residual unit, enabling efficient feature mixing with minimal overhead.
- Building on TFB, we design TriFusion-SPPF (TF-SPPF)—an enhanced spatial-pyramid-pooling module that enlarges the receptive field via hierarchical pooling and fuses multi-scale features with Transformer attention, enabling the network's upper layers to unify local detail and global context.

The remainder of this article is organized as follows. In Section 2, we review related work on Oracle bone inscriptions detection. Section 3 elaborates our proposed TriFusion-YOLO framework, elaborating on the TFB and TF-SPPF as well as their attention formulations and training losses. Section 4 presents experimental settings, evaluation metrics, and both quantitative and qualitative analyses.Finally, Section 5 concludes the paper and discusses future directions.

2. Related Work

Oracle bone inscription detection has become a key research topic in computational archaeology and computer vision alike. Early studies relied chiefly on conventional computer-vision techniques—template matching, morphological processing, and graph-based reconstruction[8]. However, these methods showed limited stability and accuracy when confronted with the complex textures and damage patterns of oracle-bone rubbings.

Driven by advances in deep learning, researchers have increasingly adopted neural-network approaches for OBI detection. Zhen et al.[9] proposed an improved YOLOv8 framework that integrates a small-object head, revised loss functions, and attention modules to boost detection performance. Xu et al.[10] developed an intelligent detection model that couples Otsu thresholding with a modified YOLOv8 and employs a slim neck to improve small-object detection. The YOLO family has proved highly effective for OBI-detection tasks. Li et al.[11] proposed a lightweight oracle-character detector built on an improved YOLOv7-tiny architecture; it integrates partial convolution and an asymptotic feature-pyramid network, reducing computation while preserving accuracy. Li and Du[12] built a complete pipeline that employs YOLOv8 for character detection and ResNet-18 for classification.

Beyond single-model approaches, researchers have explored multi-stage recognition frameworks to address detection limitations. Fujikawa[13] proposed a two-model system combining YOLOv3-tiny for initial character detection and MobileNet for secondary recognition of missed characters, achieving 98.89% validation accuracy with significant computational efficiency. Similarly, Meng et al.[14] developed a two-stage recognition method that first extracts skeletal features using the Hough transform and then applies template matching with checkpoint hit rates, demonstrating nearly 90% recognition accuracy even under character inclination and damage conditions.

2.1. Multi-scale Feature Fusion and Attention Mechanisms in OBI Detection

Multi-scale feature fusion and attention mechanisms have been widely studied to improve OBI-detection accuracy. Liu et al.[15] proposed an oracle-character detection system built on an improved YOLOv7 that adds CoordConv layers and replaces classical NMS with matrix NMS, boosting both accuracy and inference speed. Tang et al.[16] built an intelligent system that employs YOLOv5 for character segmentation and ResNet-50 for classification, achieving robust results through extensive image pre-processing and transfer-learning strategies. Addressing the challenge of insufficient and imbalanced oracle bone datasets, Yue et al.[17] introduced Dynamic Data Augmentation (DDA) strategies that adaptively adjust augmentation policies based on real-time model performance during training. Their approach achieved 8.1% accuracy improvement over baseline Inception networks on the OBC306 dataset, demonstrating the importance of adaptive training strategies for handling incomplete character structures and damaged rubbings typical in oracle bone materials.

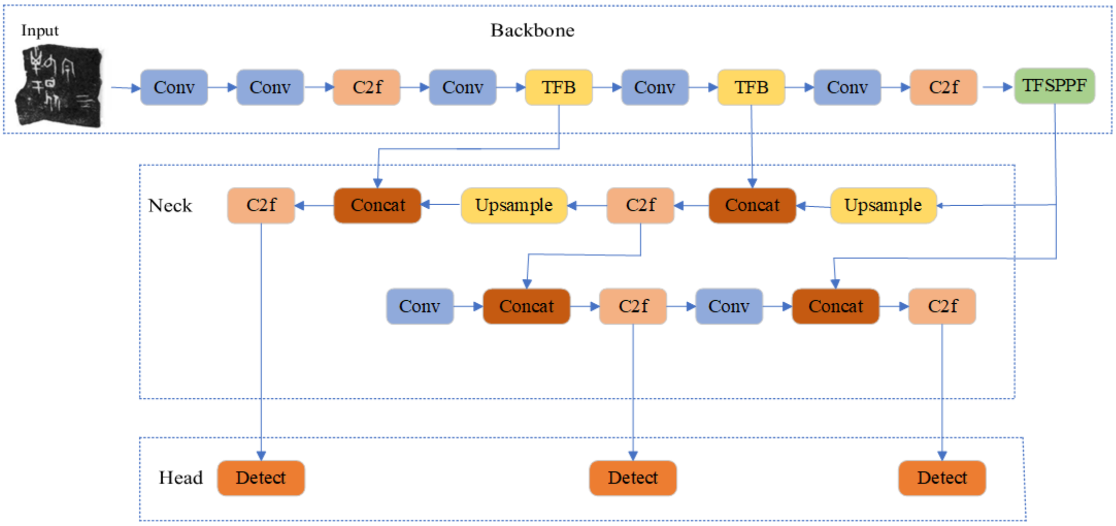

Figure 1: The figure depicts the improved YOLOv8n architecture: two TriFusion Blocks (TFB) and one TriFusion-SPPF (TF-SPPF) module are inserted into the backbone to strengthen multi-scale, multimodal feature representation. The neck and head retain the original YOLOv8n design, preserving its efficiency in object detection.

3. Methodology

In this study, we propose a YOLO architecture enhanced by modular multimodal feature fusion. The core innovation is the TriFusion Block (TFB), which integrates spatial attention, global modeling, and sequence-processing branches in parallel to capture local details, global structure, and spatial dependencies. We further design TriFusion-SPPF (TF-SPPF), which augments conventional spatial-pyramid pooling with TFB-based enhancement to strengthen multi-scale feature representation. The modules plug seamlessly into YOLOv8n, maintaining its efficiency while markedly improving detection accuracy, and offer strong scalability and transferability.The overall architecture of our improved YOLOv8n detector is illustrated in Fig. 2.

3.1. YOLOv8n

YOLO, first proposed by researchers at the University of Washington in 2015, is an efficient object-detection framework noted for its balance of speed and accuracy[18, 19]. Released by Ultralytics in 2023[20], YOLOv8 advances the series with several architectural innovations that noticeably improve feature extraction and detection performance.

In this study, we adopt YOLOv8n as the baseline owing to its lightweight design, which offers fast inference while maintaining strong detection accuracy. YOLOv8n follows a three-stage design: the backbone extracts multi-scale features, the neck fuses them, and the head performs localisation and classification. Relative to earlier versions, YOLOv8 introduces an anchor-free mechanism and a decoupled head, which respectively improve localisation flexibility and task-specific performance. In addition, YOLOv8 incorporates Task-Aligned Assigner for sample allocation and Distributive Focal Loss (DFL) for bounding-box regression, providing a robust foundation for oracle-bone character detection.

3.2. TriFusion Block

We propose the TFB, which integrates three parallel branches—a spatial-attention branch for global spatial modeling, a global-modeling branch for semantic-context extraction, and a sequential-processing branch for efficient dependency modeling. By combining these complementary pathways, TFB simultaneously captures fine-grained local features and global structural patterns, substantially improving the feature representation for oracle-bone character detection.

3.2.1. Spatial Attention Branch

To explicitly model long-range pixel dependencies in space, we embed a lightweight multi-head self-attention (MHSA) branch into the TriFusion Block. Traditional convolutional operations are constrained by local receptive fields, whereas this spatial attention mechanism enables direct interactions between any two spatial positions—an ability that is crucial for detecting sparsely distributed oracle-bone characters amid complex morphological variations and surface erosion. The branch first flattens the 2-D feature map $X \in \mathbb{R}^{B\times C\times H\times W}$ into a sequence of length $N = H \times W$, and then applies scaled dot-product attention with 32 heads to compute pairwise interactions. The attention-weighted features are finally reshaped back to their original spatial size, yielding context-rich representations for subsequent fusion stages.

$$\mathbf{X} \in \mathbb{R}^{B\times C\times H\times W} \xrightarrow{\text{flatten}} \mathbf{X}_{\text{flat}} \in \mathbb{R}^{B\times N\times C} \tag{1}$$

$$\text{Attention} = \text{Softmax}\left(\frac{\mathbf{Q}_h\mathbf{K}_h^\top}{\sqrt{d}}\right) \tag{2}$$

Here, $N = H \times W$ denotes the flattened sequence length, and $d_k = C/32$ represents the dimensionality per attention head. The flatten operation transforms 2D spatial features into 1D sequences, enabling direct interactions between any two spatial positions. The scaled dot-product attention establishes long-range spatial dependencies, which are crucial for detecting sparsely distributed oracle-bone characters amid complex morphological variations and surface erosion.

3.2.2. Global Modeling Branch

To capture semantic context and long-range dependencies of oracle-bone characters, we employ a standard Transformer encoder in the global-modeling branch. Unlike pure attention mechanisms, this approach combines global contextual encoding with non-linear feature transformation, which is essential for understanding semantically sparse and visually ambiguous oracle-bone characters. The method utilizes a double-residual Transformer block design: Stage 1 applies multi-head self-attention for dependency modeling, and Stage 2 employs LayerNorm and MLP for semantic enhancement. The input feature map [B, C, H, W] is first reshaped to [B, HW, C] for sequence modeling, followed by residual connections and a final 1×1 convolution for feature refinement.

$$Y_1 = X_{\text{flat}} + \text{MHSA}(\text{LayerNorm}(X_{\text{flat}})) \quad (3)$$

$$Y_2 = Y_1 + \text{MLP}(\text{LayerNorm}(Y_1)) \quad (4)$$

Here, MHSA denotes multi-head self-attention with 16 heads, and MLP represents a feed-forward network with 3× channel expansion and GELU activation. The double-residual design ensures stable gradient flow during training while effectively capturing semantic features and long-range contextual information essential for oracle-bone character understanding.

3.2.3. Sequential Processing Block

To efficiently model long-range dependencies in oracle-bone character sequences with lower computational cost, we employ a simplified state-space model (SSM) in the sequential-processing branch. Unlike traditional attention mechanisms with $O(N^2)$ complexity, the SSM operates in O(N) linear time, making it especially suitable for high-resolution oracle-bone rubbing images with dense character distributions. This approach is vital for detecting visually ambiguous and semantically sparse oracle-bone characters, where sequential relationships provide crucial contextual cues. The method consists of two stages: Stage 1 applies SSM with LayerNorm for efficient dependency modelling; Stage 2 applies a feed-forward network (FFN) with LayerNorm for semantic enhancement. The input is reshaped from [B, C, H, W] to [B, HW, C] for sequence processing, followed by residual connections and a final 1 × 1 convolution for feature refinement.

$$Y_1 = X_{\text{flat}} + \text{SSM}(\text{LayerNorm}(X_{\text{flat}})) \quad (5)$$

$$Y_2 = Y_1 + \text{FFN}(\text{LayerNorm}(Y_1)) \quad (6)$$

Here, SSM denotes a simplified state-space model with Linear-GELU-Linear structure, and FFN represents a feed-forward network with 2× channel expansion. Unlike traditional attention mechanisms with $O(N^2)$ complexity, the SSM operates in O(N) linear time, making it particularly efficient for processing high-resolution oracle-bone rubbing images with dense character distributions.

3.2.4. Adaptive Multimodal Feature Fusion

To integrate the complementary information from the three parallel branches, we propose an adaptive multimodal fusion module. The input feature map x is fed concurrently to the spatial-attention, global-modeling, and sequential branches, producing three feature tensors. Learnable scalar weights adaptively aggregate these tensors, which are then concatenated along the channel dimension. A lightweight Conv–BatchNorm–GELU block further blends the concatenated features, after which a channel-attention unit—global average pooling, two 1 × 1 convolutions, and a sigmoid—re-calibrates the activations. Finally, a residual shortcut adds the fused features to the original input, delivering adaptive integration with negligible computational overhead.

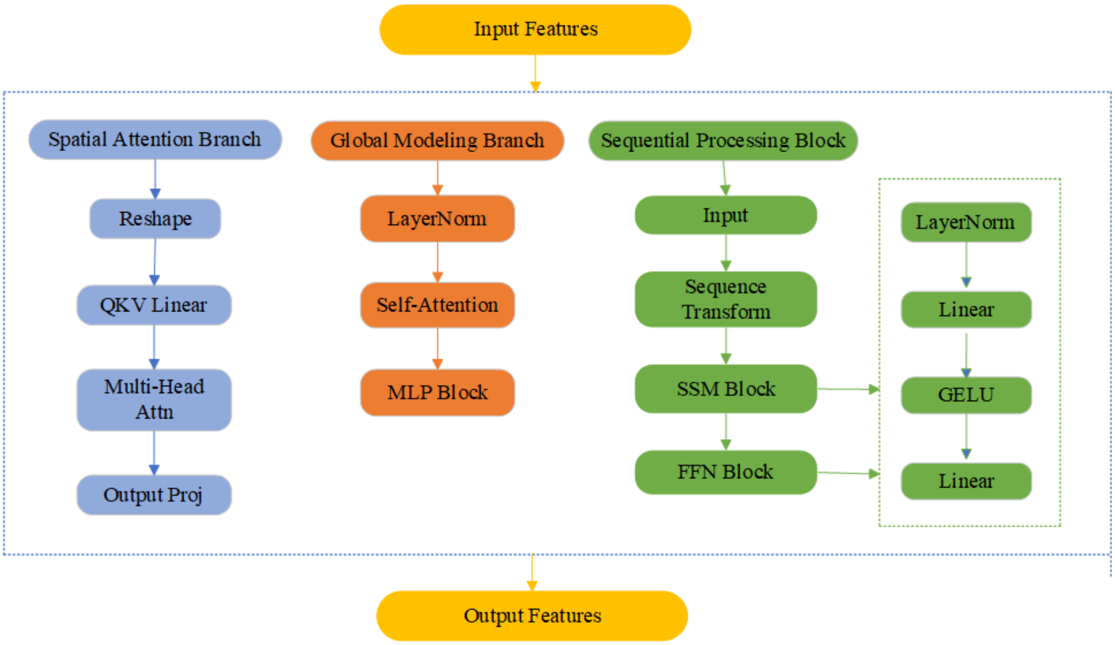

Figure 2: The figure illustrates the proposed TFB architecture, which integrates three parallel branches—spatial attention, global modeling, and sequential processing—to fuse multimodal features for improved oracle-bone character detection.

3.3. TF-SPPF Enhancement Design

To boost feature representation without sacrificing the original multi-scale capability, we design the TriFusion-SPPF (TF-SPPF) module as a non-intrusive wrapper around the standard SPPF. TF-SPPF preserves the cascaded triple 5 × 5 max-pooling stages and the multi-scale fusion of the original SPPF. After SPPF, the output features flow into the Combined-Enhancement module for multimodal refinement, which adaptively fuses three parallel branches—spatial attention, global modeling, and sequential processing—to enrich semantic representation. An adjustment layer of 1 × 1 convolution, BatchNorm, and SiLU activation further refines the fused features. To improve feature quality, a channel re-calibration block applies global average pooling, an eight-fold channel reduction, SiLU activation, expansion, and Sigmoid gating to produce channel-attention weights. Finally, a conditional residual shortcut is used when the input and output shapes match; otherwise, the recalibrated features are forwarded directly. This design retains the efficient multi-scale modeling of SPPF while markedly boosting feature discriminability through combined attention and re-calibration mechanisms.

4. Experiment

In this section, we conduct comprehensive experiments to evaluate the effectiveness of our proposed TriFusion-YOLO framework for oracle bone character detection. To demonstrate the superiority of our method, we perform extensive evaluations on the Oracle-Bone Inscriptions

Multimodal Dataset, comparing our approach against state-of-the-art object detection models including the baseline YOLOv8n and the latest YOLOv11n. Our experimental evaluation encompasses both quantitative and qualitative analyses, examining detection accuracy, computational efficiency, and visual performance across diverse oracle bone rubbing scenarios. We systematically investigate the contribution of each component through detailed ablation studies, analyzing the individual and combined effects of the spatial attention branch, global modeling branch, and sequential processing branch within the TriFusion Block. Additionally, we provide thorough implementation details, evaluation metrics, and performance comparisons to ensure reproducibility and fair assessment of our proposed method.

4.1. Dataset

We use the rubbing subset of the Oracle-Bone Inscriptions Multimodal Dataset (OBIMD)[21], which comprises 10,077 high-quality rubbing images sampled from five historical phases of Yinxu. Each image is professionally annotated by domain experts with bounding boxes and category labels for every character, fully reflecting real-world challenges in oracle-bone detection—character diversity, scale variation, and complex backgrounds. The annotation workflow combines AI-assisted pre-labelling with expert verification to ensure high data quality and accuracy. Overall, the dataset captures the diversity and complexity of real-world oracle-bone detection scenarios.

4.2. Implementation details

All experiments were run on a workstation equipped with an NVIDIA GeForce RTX 4070 GPU (12 GB). The experiments used PyTorch 2.6.0 with CUDA 11.8. The proposed method was implemented with Ultralytics 8.3.114, an official framework for YOLOv8n. OpenCV 4.11.0 handled image pre-processing.

Models were trained for 200 epochs with a batch size of 32. Input images were resized to 640 × 640. Optimisation used AdamW (initial LR = 0.001; final LR = 0.01; momentum = 0.937; weight decay = 1×10^{-4}). The learning-rate schedule followed a warm-up cosine-annealing pattern: LR increases during the first five epochs and then decays following a cosine curve. Early stopping (patience = 50 epochs) prevented over-fitting. Mixed-precision training (AMP) was enabled to improve efficiency and memory usage, and data-loading workers were set to 8.

4.3. Evaluation Metrics

The model's performance was evaluated using three critical metrics: Precision, Recall, and mAP50. These metrics provided key insights into the model's detection accuracy, recall capability, and overall performance on the validation set. Precision assessed the proportion of correct predictions among all positive predictions, Recall measured the ability to detect all relevant characters, and mAP0.5 offered a comprehensive measure of detection effectiveness. The metrics are defined as follows:

$$\text{Precision} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalsePositives}} \tag{7}$$

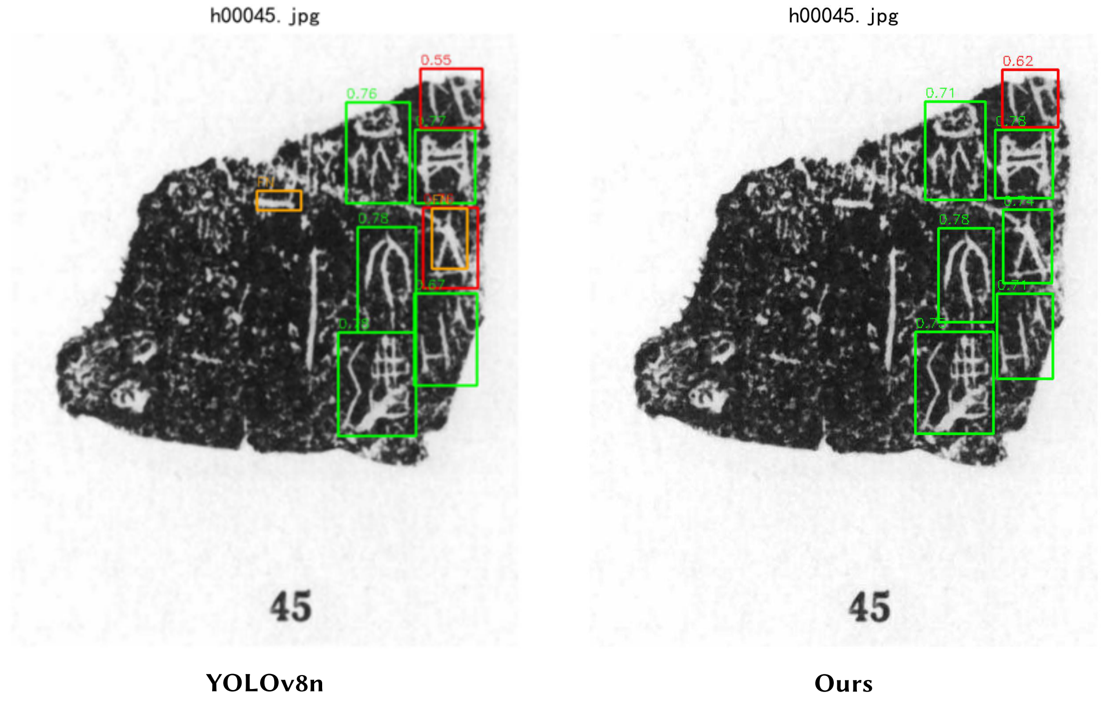

Figure 3: Qualitative comparison of oracle-bone character detection results. Our TriFusion-YOLO method demonstrates improved detection accuracy with reduced false positives compared to the YOLOv8n baseline.

$$\text{Recall} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalseNegatives}} \tag{8}$$

$$\text{mAP}_{0.5} = \frac{1}{C} \sum_{c=1}^{C} \text{AP}_c \,|_{\text{IoU}=0.5} \tag{9}$$

For each class, Average Precision (AP) is defined as the area under the precision–recall curve, where precision is plotted against recall from 0 to 1. The mean Average Precision at an IoU threshold of 0.5 (mAP0.5) equals the average AP over all categories in the dataset. This metric jointly evaluates precision and recall, capturing the model's ability to detect and classify oracle-bone characters while limiting false positives; it is therefore well suited to complex archaeological document-analysis tasks.

4.4. Performance Comparison

To assess the effectiveness of the proposed TriFusion-YOLO model, we systematically compare its performance with the baseline YOLOv8n and the latest YOLOv11n on an oracle-bone character dataset. All models are trained under identical settings—including dataset splits, hyper-parameters, and evaluation metrics—to ensure a fair comparison. Performance is evaluated using Precision, Recall, and mAP@0.5, with special attention to detection accuracy and the reduction of false positives in complex archaeological documents.

Table 1
Performance comparison of different models on the oracle bone character detection dataset.

Model	Recall (%)	Precision (%)	mAP@0.5 (%)
YOLOv8n (baseline)	79.07	83.27	82.39
YOLOv11n	79.33	87.64	83.93
Ours	**81.40**	**91.30**	**86.35**

Table 1 compares the performance of different models on the oracle-bone character detection dataset. The baseline YOLOv8n achieves a recall of 79.07%, precision of 83.27%, and mAP@0.5 of 82.39%. YOLOv11n shows slight improvements, with a recall of 79.33%, precision of 87.64%, and mAP@0.5 of 83.93%. In contrast, the TriFusion-YOLO model significantly outperforms both baseline methods, attaining a recall of 81.40%, precision of 91.30%, and mAP@0.5 of 86.35%. Notably, our method demonstrates substantial gains over YOLOv8n, with recall improved by 2.33%, precision enhanced by 8.03%, and mAP@0.5 increased by 3.96%. These results confirm the effectiveness of the TriFusion Block and TF-SPPF modules in oracle-bone character detection, especially in reducing false positives while maintaining high detection accuracy.

Table 2
Ablation study of different modules on oracle-bone detection.

Model Config	SAB	GMB	SPB	Recall (%)	Precision (%)	mAP@0.5 (%)
YOLOv8n (baseline)				79.07	83.27	82.39
+ SAB	✓			80.62	85.95	83.29
+ GMB		✓		77.52	88.50	83.01
+ SPB			✓	79.07	84.30	81.68
+ SAB+GMB		✓	✓	79.84	87.29	83.57
+ SAB+GMB+SPB (Ours)	✓	✓	✓	**81.40**	**91.30**	**86.35**

Table 2 reports an ablation study of the TriFusion Block integrated into the YOLOv8n baseline. We systematically evaluate the contribution of each component—Spatial-Attention Branch (SAB), Global-Modeling Branch (GMB), and Sequential-Processing Block (SPB)—by adding them to the baseline one at a time. Introducing each branch individually yields consistent gains, and combining branches further improves performance. The full model attains the best recall (81.40%), precision (91.30%), and mAP0.5(86.35%), exceeding the YOLOv8n baseline by 3.96%. Best results are highlighted in bold.

5. Conclusion

In this work, we propose proposes the TriFusion Block (TFB), a multi-branch fusion module that integrates spatial attention, global modeling, and sequential processing to improve oracle-bone character detection, especially in complex and severely degraded rubbing images.

Integrated into the YOLOv8n backbone, our method leverages complementary capabilities from each branch to enhance both localization and classification performance. Extensive experiments on the Oracle-Bone Inscriptions Multimodal Dataset show that our approach achieves a recall of 81.40%, precision of 91.30%, and mAP@0.5 of 86.35%, surpassing the baseline model by 2.33, 8.03, and 3.96 percentage points, respectively. Ablation studies confirm that each component of TFB contributes positively, validating the effectiveness and generalizability of our multi-branch fusion strategy.

Acknowledgments

This research was supported by the Natural Science Foundation of China (Grant No. 62506007), the Natural Science Foundation of Henan Province (Grant No. 242300420680), the Paleography and Chinese Civilization Inheritance and Development Program (Grant Nos. G1807, G1806, G2821), the Henan Province Science and Technology Research Project (Grant Nos. 242102210116, 252102321071), Major Science and Technology Project of Anyang (Grant No. 2025A02SF007) and the Henan Province High-Level Talents International Training Program (Grant No. GCC2025028).

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] J. Li, X. Chi, Q. Wang, K. Huang, D. Wang, Y. Liu, C.-L. Liu, A comprehensive survey of oracle character recognition: Challenges, benchmarks, and beyond, Benchmarks, and Beyond (2024).

[2] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, L. Farhan, Review of deep learning: concepts, cnn architectures, challenges, applications, future directions, Journal of big Data 8 (2021) 1–74.

[3] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.

[4] A. Bochkovskiy, C.-Y. Wang, H.-Y. M. Liao, Yolov4: Optimal speed and accuracy of object detection, arXiv preprint arXiv:2004.10934 (2020).

[5] Z. Wu, T. Ding, Y. Lu, D. Pai, J. Zhang, W. Wang, Y. Yu, Y. Ma, B. D. Haeffele, Token statistics transformer: Linear-time attention via variational rate reduction, in: The Thirteenth International Conference on Learning Representations, 2024.

[6] X. Liu, J. Liu, J. Tang, G. Wu, Catanet: Efficient content-aware token aggregation for lightweight image super-resolution, in: Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 17902–17912.

[7] L. Kong, J. Dong, J. Tang, M.-H. Yang, J. Pan, Efficient visual state space model for image deblurring, in: Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 12710–12719.

[8] S. Gu, Identification of oracle-bone script fonts based on topological registration, Computer & Digital Engineering 10 (2016) 029.

[9] Q. Zhen, L. Wu, G. Liu, An oracle bone inscriptions detection algorithm based on improved yolov8, Algorithms 17 (2024) 174.

[10] H. Xu, Z. Zhang, Q. Liang, Z. Lin, Research on the intelligent oracle bone script recognition model based on otsu and improved yolov8, in: 2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI), IEEE, 2024, pp. 972–976.

[11] Y. Li, H. Chen, W. Zhang, W. Sun, Lightweight oracle bone character detection algorithm based on improved yolov7-tiny, in: 2024 IEEE International Conference on Mechatronics and Automation (ICMA), IEEE, 2024, pp. 485–490.

[12] D. Li, B. Du, Research on oracle bone inscription segmentation and recognition model based on deep learning, in: 2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI), IEEE, 2024, pp. 1309–1314.

[13] Y. Fujikawa, H. Li, X. Yue, C. Aravinda, G. A. Prabhu, L. Meng, Recognition of oracle bone inscriptions by using two deep learning models, International Journal of Digital Humanities 5 (2023) 65–79.

[14] L. Meng, Two-stage recognition for oracle bone inscriptions, in: International conference on image analysis and processing, Springer, 2017, pp. 672–682.

[15] J. Liu, T. Huang, R. Li, Z. Yang, Oracle recognition based on improved yolov7, in: International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2023), volume 13105, SPIE, 2024, pp. 521–527.

[16] H. Tang, R. Tang, H. Wang, Oracle bone script intelligent recognition: Automatic segmentation and recognition of original rubbing single characters, in: 2024 5th International Conference on Electronic Communication and Artificial Intelligence (ICECAI), IEEE, 2024, pp. 414–418.

[17] X. Yue, H. Li, Y. Fujikawa, L. Meng, Dynamic dataset augmentation for deep learning-based oracle bone inscriptions recognition, ACM Journal on Computing and Cultural Heritage 15 (2022) 1–20.

[18] P. Jiang, D. Ergu, F. Liu, Y. Cai, B. Ma, A review of yolo algorithm developments, Procedia computer science 199 (2022) 1066–1073.

[19] J. Terven, D.-M. Córdova-Esparza, J.-A. Romero-González, A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas, Machine learning and knowledge extraction 5 (2023) 1680–1716.

[20] M. Sohan, T. Sai Ram, C. V. Rami Reddy, A review on yolov8 and its advancements, in: International Conference on Data Intelligence and Cognitive Informatics, Springer, 2024, pp. 529–545.

[21] B. Li, D. Luo, Y. Liang, J. Yang, Z. Ding, X. Peng, B. Jiang, S. Han, D. Sui, P. Qin, et al., Oracle bone inscriptions multi-modal dataset, arXiv preprint arXiv:2407.03900 (2024).