2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

A New Exploration: Ancient Book Defect Detection with Attention Mechanisms

Jun Yu[1 2 3]*, Yemao Zhang[1 2], Jiahui Cheng[1 2], Lingnan Bai[1 2], Jiaxing Fan[1 2], Zhen Zhang[1 2], Ruiyao Han[1 2], Zhe Xu[1 2]

[1] Taihang Laboratory in Shanxi Province (Advanced Computing Laboratory in Shanxi Province), Taiyuan Shanxi 030024, China
[2] Shanxi Taihang Laboratory Co., Ltd., Taiyuan Shanxi 030024, China
[3] School of Computer and Information Technology (School of Big Data), Shanxi University, Taiyuan Shanxi 030006, China
y.j@sxu.edu.cn

Abstract. With the increasing application of digital technology in cultural heritage preservation, the digitization of ancient books and their defect detection has become an important research topic. Ancient books are prone to a variety of defects, and traditional manual detection methods are inefficient and cannot guarantee accuracy. This paper constructs a specialized dataset containing six types of defects and proposes an improved YOLOv8 network, which is applied to ancient book defect detection for the first time. By introducing three attention mechanisms—CBAM, SEBlock, and ECA—and applying improvements at different positions within the network, the model's ability to recognize defects is enhanced. Experimental results show that the improved YOLOv8 model significantly improves detection performance.

Keywords: Ancient Book Defect Detection, Object Detection, YOLOv8, Attention Mechanism.

1 Introduction

In recent years, with the rapid development of digital technology, modern methods have been increasingly applied in the preservation of cultural heritage, enabling more efficient storage, management, and dissemination of historical books. As an important carrier of history, ancient books hold rich cultural information and have immense academic and historical value. However, due to long-term exposure to environmental factors, such as humidity, light, and insect damage, ancient books often suffer from various defects, including Vitium, Defaced, Crease, Patch, Signature, and Inkiness. These defects not only affect the preservation and restoration of ancient books but also pose significant challenges to the digitization process. Ancient book defect detection, as a task with both academic and practical value, has not received sufficient attention and research. While traditional manual detection methods can identify some defects, they are inefficient when dealing with large-scale ancient book datasets. Therefore, finding

automated and intelligent methods to efficiently and accurately detect defects in ancient books has become a pressing technical challenge.

Existing object detection technologies are mainly focused on object detection [1], video anomaly detection [2], surface anomaly detection [3], facial recognition [4], autonomous driving [5], etc., with few studies applying deep learning techniques to the automated defect detection of ancient book. This paper, for the first time, applies the advanced YOLOv8 network to ancient book defect detection and introduces improvements, thereby opening up new research directions in the field of ancient book preservation and restoration.

In the field of object detection, the YOLO (You Only Look Once) [6] series of models have been widely applied due to their efficient end-to-end detection capabilities. Since the introduction of YOLOv1, the series has been continuously updated and optimized, from YOLOv1 to YOLOv3 [6]-[8], and now to the latest YOLOv8 [9], with each generation improving detection accuracy, speed, and generalization ability. However, defects in ancient book images are often subtle, especially stains and creases, which can be easily overlooked in complex backgrounds. While YOLOv8 has demonstrated excellent performance in many tasks, there is still room for improvement in the domain of ancient book defect detection.

To address these issues, this paper proposes an improved YOLOv8-based method for ancient book defect detection. First, a dedicated dataset containing six types of defects in ancient books is constructed, including Vitium, Defaced, Crease, Patch, Signature, and Inkiness. This dataset covers common defect types in ancient books and provides strong data support for model training and testing. Through careful annotation and categorization, the dataset ensures that the model can learn the characteristics of different defects during training, thus improving detection accuracy.

Furthermore, this paper makes several improvements to the YOLOv8 network, particularly by introducing three common attention mechanisms: CBAM [10], SEBlock [11], and ECA [12]. The incorporation of these attention mechanisms helps the model better focus on the important regions of the image, thus enhancing its ability to detect defects. With these improvements, the model can more accurately capture important information at different levels of the ancient book images, thus improving its ability to detect small defects.

In summary, our contributions are as follows:

- We constructed a dataset covering six typical types of ancient book defects, providing standardized data support for ancient book defect detection and opening up a new direction for research in this field.
- We proposed an improved YOLOv8 network structure, which enhances the model's ability to extract features from complex textures and subtle defects by introducing three attention mechanisms (CBAM, SEBlock, and ECA), providing an efficient and precise solution for the field of ancient book preservation.

2 Method

2.1 YOLOv8 Network Architecture

YOLOv8 is an important version in the YOLO series of object detection models, continuing the series' characteristic of fast and efficient detection. The structure of YOLOv8 (Fig. 1) can be roughly divided into three parts: Backbone, Neck, and Head, featuring the C2f module depicted in Fig. 2.

The Backbone is the foundational network component in YOLOv8 responsible for extracting image features. YOLOv8 uses CSPDarknet53 [13] as its Backbone, which combines the advantages of Cross-Stage Partial Networks (CSPNet) [14], effectively reducing the number of parameters and enhancing the feature representation ability.

The Neck component is mainly responsible for further integrating and processing the feature maps extracted from the Backbone. In YOLOv8, the Neck uses the PANet (Path Aggregation Network) [15] structure, a feature pyramid network that enhances the fusion of features at different levels through a bottom-up path. This allows for better feature fusion and propagation at multiple scales, enabling the network to capture more detailed information about the targets.

The Head component is the output layer in YOLOv8 responsible for the actual detection task. YOLOv8 adopts an Anchor-Free design, meaning it no longer relies on predefined anchor boxes. Instead, it directly predicts parameters such as the center point and width/height of the bounding box. This design effectively reduces computational complexity and significantly improves the accuracy of the predicted bounding boxes.

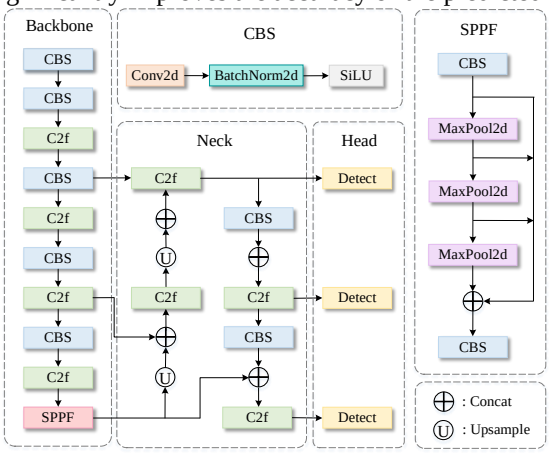

Fig. 1. Architecture of YOLOv8 network.

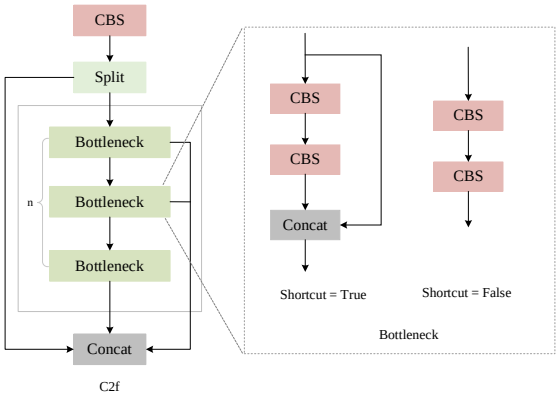

Fig. 2. Diagram of C2f module.

2.2 Attention Mechanism

In the field of deep learning, attention mechanisms have become a crucial technique for enhancing model performance, widely applied in tasks such as object detection, image classification, and natural language processing. By enabling models to focus on different parts of the input, attention mechanisms effectively enhance feature representation and improve the ability to extract critical information. This paper integrates three classic attention mechanisms—CBAM (Fig. 3) [10], SEBlock (Fig. 4) [11], and ECA (Fig. 5) [12]—into the YOLOv8 network to further improve the accuracy of ancient book defect detection.

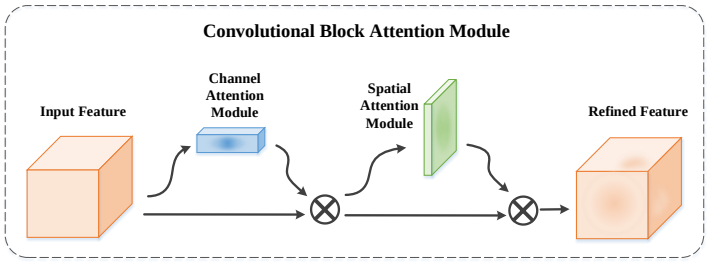

Fig. 3. Diagram of CBAM module.

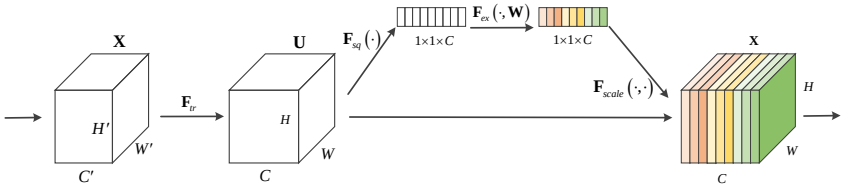

Fig. 4. Diagram of SEBlock module.

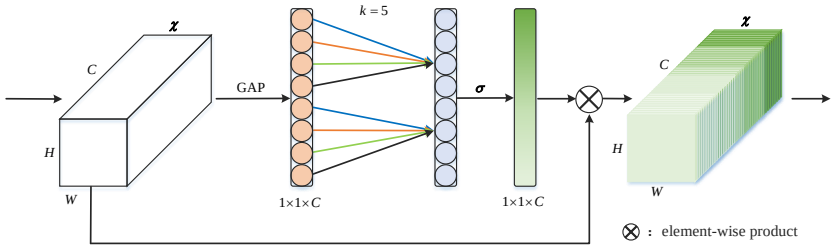

Fig. 5. Diagram of ECANet module.

CBAM is a lightweight attention module that combines channel attention and spatial attention. Channel attention adjusts the weights of each channel to focus on the most important features, while spatial attention enhances feature extraction by focusing on key regions in the image. CBAM first processes channel attention and then optimizes the feature map through spatial attention, which helps better detect subtle defects in complex backgrounds.

SEBlock enhances the network's focus on important features by dynamically adjusting the weight of each channel through the "Squeeze-and-Excitation" operation. It captures global information through global average pooling, then generates channel weights via fully connected layers and applies them to the input feature map, improving the model's ability to detect defects in ancient texts.

ECA is an efficient channel attention mechanism that calculates channel weights through local cross-channel interaction, avoiding fully connected layers and thus reducing computational complexity. ECA precisely captures inter-channel relationships, improving the accuracy of detecting subtle defects.

By integrating CBAM, SEBlock, and ECA, YOLOv8 significantly improves feature extraction accuracy, robustness, and efficiency in ancient text defect detection, especially in complex scenarios involving defect localization and classification.

2.3 Improved YOLOv8 Network

In ancient book defect detection, images often contain complex backgrounds and subtle, diverse defect features, such as damage, stains, and creases. Although YOLOv8 performs excellently in general object detection, it tends to be interfered with by background noise and irrelevant information when dealing with fine-grained defects. To address this issue, we introduce attention mechanisms to enhance the model's ability to capture key defect features in complex backgrounds. CBAM, SEBlock, and ECA are three lightweight attention mechanisms that help the model focus on relevant features while maintaining low computational overhead. These mechanisms are embedded at different positions within the YOLOv8 network to enhance its defect detection capability.

The Backbone of YOLOv8 outputs multi-scale features that retain certain semantic and spatial information, but background noise may affect the feature fusion in the Neck. As shown in Fig. 6, by introducing attention mechanisms after the Backbone output, the feature weights can be dynamically adjusted to help the model focus on important

regions and suppress irrelevant information. Moreover, YOLOv8's multi-scale detection relies on feature maps of different scales, and attention mechanisms help balance the importance of features at each scale, thereby improving detection accuracy.

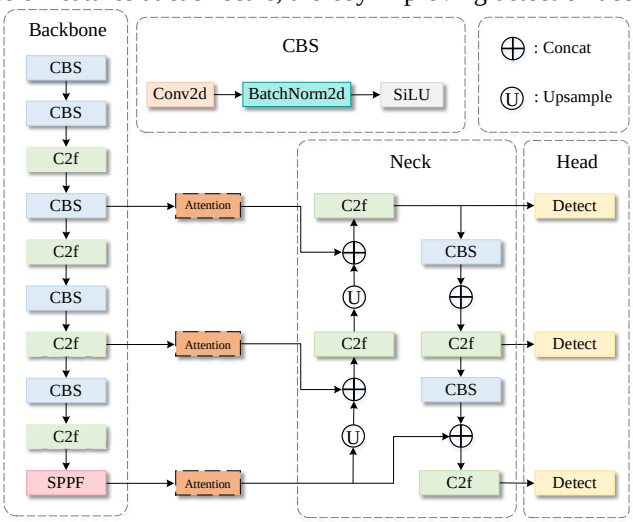

Fig. 6. Improved YOLOv8 Network Structure (Version 1).

In the C2f module, feature branches are simply stacked after the Concat operation, without weighting the importance of different features. To further enhance feature extraction, we introduce attention mechanisms within the C2f module named AttC2f of the Backbone to dynamically weight the stacked features, as shown in Fig. 7 and Fig. 8.

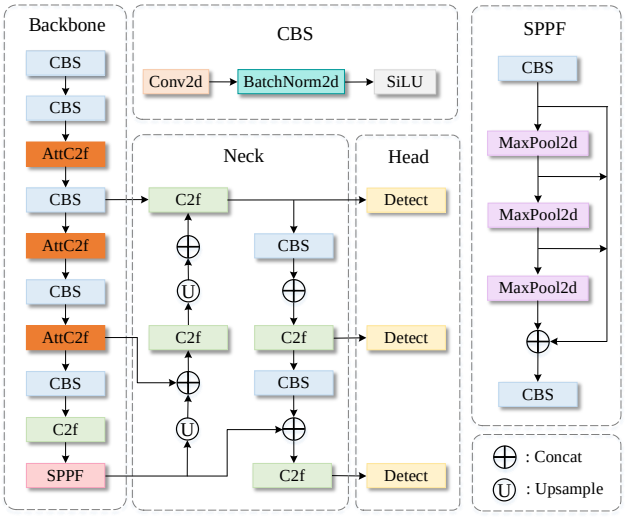

Fig. 7. Improved YOLOv8 Network Structure (Version 2).

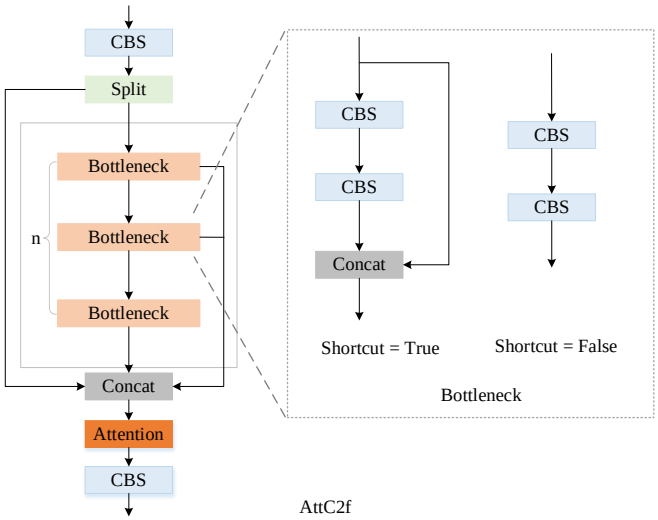

Fig. 8. Diagram of AttC2f module.

3 Experiment and Results Analysis

3.1 Dataset Classification

This paper constructs a dataset containing six typical types of defects found in ancient books. The dataset includes six common types of defects: Vitium, Defaced, Crease, Patch, Signature, and Inkiness. Each defect type exhibits distinct physical characteristics and morphological differences in real-world applications, as shown in the Fig. 9.

Fig. 9. Dataset classification.

The dataset is divided into training, validation, and test sets following the standard object detection task structure. Specifically, the training set consists of 3,328 defect-labeled ancient book images, with a total of 11,564 labeled defects. The validation set

contains 955 images with 3,197 labeled defects, and the test set contains 462 images with 1,613 labeled defects.

Each image may contain one or more different defect types. These defect categories have significant reference value in the actual ancient book restoration process. For example, Vitium typically manifests as tears or damage along the edges of pages, affecting the integrity of the book; Defaced refers to contamination on the surface of books caused by external factors, such as water or oil stains, often covering part of the content; Crease refers to creases or cracks caused by external forces, affecting the flatness and readability of the paper; Patch refers to artificial repair marks added during the restoration process; Signature includes past stamps or handwritten signatures, which do not affect the content but need special attention when preserving the original appearance of the book; Inkiness refers to later written characters or ink marks applied to the book, often due to hand-written additions, corrections, or annotations. Such ink marks may obscure original text or patterns, affecting the completeness and readability of the book.

By using this dataset, this paper can better study automatic defect detection methods for ancient book. The dataset not only provides rich sample support for improving the YOLOv8 model but also demonstrates its potential application in real-world ancient book restoration scenarios through comparative evaluations of different model architectures and training strategies.

3.2 Environment Configuration

The experimental environment in this study uses Windows 11 as the operating system with 64GB of RAM to support large-scale data processing and model training. The GPU is an NVIDIA GeForce RTX 4060, which significantly enhances training speed, particularly for deep learning tasks. The processor is an Intel(R) Core(TM) i9-14900HX, effectively handling data loading, preprocessing, and model optimization tasks.

The training process was set with 500 epochs, a batch size of 32, and an image size of 640.

3.3 Evaluation Metrics

In evaluating the performance of deep learning models, selecting appropriate evaluation metrics is key to ensuring the model's practical effectiveness. In this ancient book defect detection task, a series of important evaluation metrics were used to comprehensively assess the detection performance of the improved YOLOv8 model, including the mean average precision (mAP) [16] at different IoU thresholds (mAP50-95 and mAP50) and the number of parameters (Params).

By comprehensively considering these evaluation metrics, this study provides a thorough evaluation of the improved YOLOv8 model's performance in ancient book defect detection, offering valuable reference for further model optimization and real-world deployment.

3.4 Analysis

This experiment made a series of improvements to the baseline YOLOv8s network to enhance its performance in the task of ancient book defect detection. By introducing attention mechanisms (CBAM, SEBlock, and ECA) and testing in different network structures, the experimental results show a significant improvement in the mAP (Mean Average Precision) metric. Additionally, to thoroughly validate the performance advantages of the improved YOLOv8s network in the target detection task, it was compared with other mainstream detection models, such as fasterrcnn, YOLOv5s and YOLOv6s.

As shown in Fig. 10, the prediction results of the proposed best method are displayed. Fig. 11 presents the performance metrics of different models across various categories. As shown in Fig. 12, the mAP performance of different models is displayed, where it can be observed that the Backbone + SEBlock improvement strategy yields the best results.

Firstly, the mAP50 of the baseline YOLOv8s network was 0.366. After introducing the attention mechanisms, the model's performance improved in all configurations. The "Backbone+SEBlock" configuration achieved the best performance with an mAP50 of 0.394, an improvement of 0.028 over the baseline network.

Next, the introduction of CBAM and ECA mechanisms also led to performance improvements. Specifically, the Backbone+CBAM and Backbone+ECA configurations achieved mAP50 values of 0.376 and 0.379, respectively, improving by 0.010 and 0.013 compared to the baseline network. These results indicate that, although the performance of these two mechanisms is not as strong as SEBlock, they still effectively improve the model's focus on features, leading to a boost in detection accuracy.

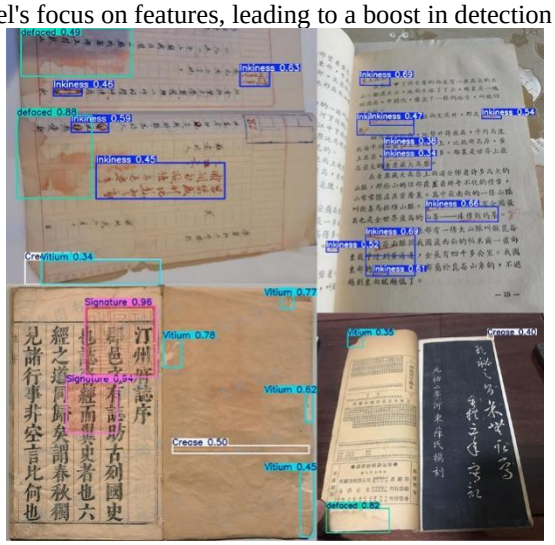

Fig. 10. Improved YOLOv8 prediction results.

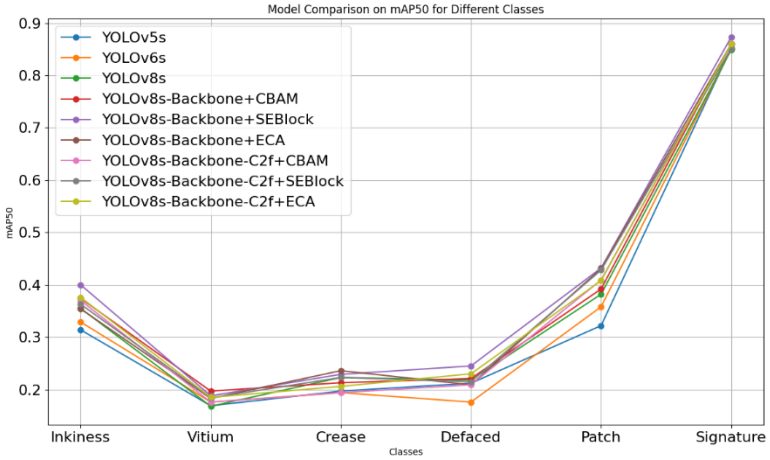

Fig. 11. mAP50 for each category across different models.

Introducing attention mechanisms into the C2f structure within the Backbone also showed positive effects. For example, "Backbone-C2f+CBAM" achieved an mAP50 of 0.369, demonstrating the ability to optimize feature extraction. "Backbone-C2f+SEBlock" and "Backbone-C2f+ECA" further improved the mAP50 to 0.377, providing additional evidence of the effectiveness of incorporating attention mechanisms into the C2f structure.

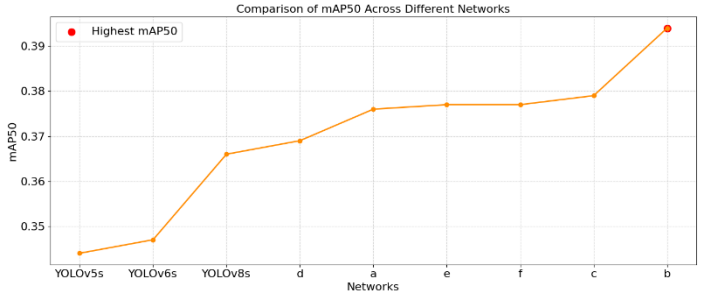

Fig. 12. Performance comparison of different models. In this figure, a refers to Backbone+CBAM, b to YOLOv8-Backbone+SEBlock, c to YOLOv8-Backbone+ECA, d to YOLOv8-Backbone-C2f+CBAM, e to YOLOv8-Backbone-C2f+SEBlock, and f to YOLOv8-Backbone-C2f+ECA.

In summary, through the introduction of multiple attention mechanisms, the experimental results demonstrate the effectiveness and potential of YOLOv8s in the ancient book defect detection task. Different improvement strategies exhibit varying characteristics in terms of mAP and speed, especially with the introduction of SEBlock, which proves to be effective in enhancing feature representation and improving detection accuracy. These results provide valuable references for future model optimization and real-world applications.

Table 1. In the "Strategy" column, a refers to Backbone+CBAM, b to Backbone+SEBlock, c to Backbone+ECA, d to Backbone-C2f+CBAM, e to Backbone-C2f+SEBlock, and f to Backbone-C2f+ECA.

YOLO-Base	Strategy	mAP50						mAP50	mAP50-95	Param
		0	1	2	3	4	5			
fastercnn	-	0.298	0.224	0.264	0.257	0.330	0.843	0.370	0.222	28.33
v5s	-	0.314	0.169	0.197	0.212	0.322	0.852	0.344	0.229	9.14
v6s	-	0.329	0.176	0.194	0.176	0.358	0.850	0.347	0.233	4.49
v8s	-	0.355	0.168	0.223	0.218	0.382	0.849	0.366	0.250	11.13
v8s	a	0.374	0.197	0.213	0.221	0.392	0.860	0.376	0.256	11.54
v8s	b	0.400	0.188	0.229	0.245	0.432	0.873	0.394	0.268	11.32
v8s	c	0.354	0.183	0.236	0.209	0.431	0.860	0.379	0.261	11.13
v8s	d	0.370	0.176	0.194	0.209	0.409	0.858	0.369	0.250	11.54
v8s	e	0.363	0.184	0.223	0.215	0.428	0.851	0.377	0.250	11.59
v8s	f	0.376	0.185	0.206	0.230	0.408	0.860	0.377	0.254	11.13

3.5 Heatmap Visualization Analysis

To further visually demonstrate the performance of different networks in feature extraction and target localization, this study generated heatmaps for YOLOv5s, YOLOv6s, YOLOv8s, and their improved models using Grad-CAM (Gradient-weighted Class Activation Mapping) [17] technology. The heatmaps, based on two representative sample images, show the activation regions and attention distribution for all nine networks.

As shown in Fig. 13, the heatmaps of YOLOv5s and YOLOv6s show weak responses in target regions, especially in complex backgrounds or for small targets, where the feature distribution is less concentrated. This often leads to inaccurate target localization. In contrast, YOLOv8s shows significantly enhanced responses in target areas, with better feature capture of the target boundaries and key regions. The model with the attention mechanism shows a significant increase in response to the target areas.

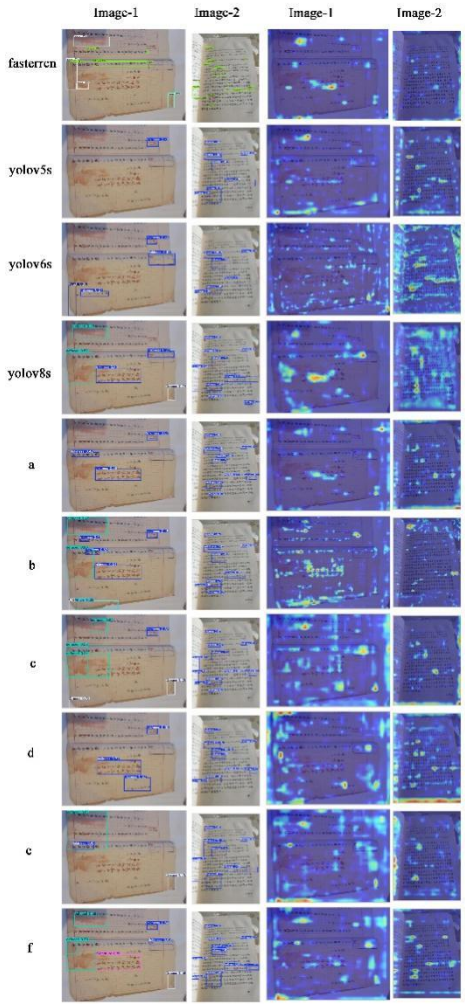

Fig. 13. Comparison of prediction results and heatmaps from different networks. In this figure, a refers to Backbone+CBAM, b to YOLOv8-Backbone+SEBlock, c to YOLOv8-Backbone+ECA, d to YOLOv8-Backbone-C2f+CBAM, e to YOLOv8-Backbone-C2f+SEBlock, and f to YOLOv8-Backbone-C2f+ECA.

4 Conclusion

This paper constructs a dataset containing six types of typical defects in ancient books, providing standardized data support for the task of ancient book defect detection and offering abundant training data for the application of YOLOv8 in this field. Based on this dataset, an improved YOLOv8 network architecture is proposed and, for the first time, applied to ancient book defect detection, successfully identifying various defect types in ancient book images. To address the limitations of the traditional YOLOv8 in

handling complex textures and subtle defects in ancient book images, this paper introduces three attention mechanisms and embeds them at different positions within the network to enhance the model's ability to extract key features. Experimental results show that the improved YOLOv8 network achieves significant performance improvements.

Acknowledgments. This study was funded by the National Key R&D Program of China (2023YFB4502904).

References

1. Li, S., Li, X., Li, Z., Ma, H., Sheng, J., Li, B.: Dual Guidance Enhancing Camouflaged Object Detection via Focusing Boundary and Localization Representation. In: 2024 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6. IEEE (2024)
2. Pi, R., Xu, J., Peng, Y.: FE-VAD: High-Low Frequency Enhanced Weakly Supervised Video Anomaly Detection. In: 2024 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6. IEEE (2024)
3. Chen, S.F., Liu, Y.M., Liu, C.C., Chen, T.P.C., Wang, Y.C.F.: Domain-generalized textured surface anomaly detection. In: 2022 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6. IEEE (2022)
4. Li, H., Niu, H., Zhu, Z., Zhao, F.: Cliper: A unified vision-language framework for in-the-wild facial expression recognition. In: 2024 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6. IEEE (2024)
5. Dong, Z., Zhu, X., Cao, X., Ding, R., Zhou, C., Li, W., et al.: Bezier-former: A unified architecture for 2d and 3d lane detection. In: 2024 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6. IEEE (2024)
6. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788. IEEE (2016)
7. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7263–7271. IEEE (2017)
8. Redmon, J., Farhadi, A.: YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018).
9. YOLO by Ultralytics, https://www.ultralytics.com, last accessed 2025/02/25
10. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: CBAM: Convolutional Block Attention Module. In: European Conference on Computer Vision (ECCV), pp. 3–19. Springer (2018)
11. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7132–7141. IEEE (2018)
12. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: ECA-Net: Efficient channel attention for deep convolutional neural networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11534–11542. IEEE (2020)
13. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: Scaled-YOLOv4: Scaling cross stage partial network. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 13029–13038. IEEE (2021)
14. Wang, C.Y., Liao, H.Y.M., Wu, Y.H., Chen, P.Y., Hsieh, J.W., Yeh, I.H.: CSPNet: A new backbone that can enhance learning capability of CNN. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 390–391. IEEE (2020)

15. Liu, S., Qi, L., Qin, H., Shi, J., Jia, J.: Path aggregation network for instance segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8759–8768. IEEE (2018)
16. Padilla R, Passos W L, Dias T L B, et al.: A comparative analysis of object detection metrics with a companion open-source toolkit. Electronics **10**(3), 279 (2021)
17. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: IEEE International Conference on Computer Vision (ICCV), pp. 618–626. IEEE (2017).