

## Research Article

# An Automated Insect Hole Detection System for Wood Carvings Utilizing Enhanced YOLOv11

Jun-Lin Wu, Chung-Wen Hung\*

National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C

Email: [M11312084@yuntech.edu.tw](mailto:M11312084@yuntech.edu.tw), [wenhung@yuntech.edu.tw](mailto:wenhung@yuntech.edu.tw)\*

\*Corresponding Author

## ARTICLE INFO

### Article History

Received 28 December 2025

Accepted 23 September 2026

### Keywords

micro-hole detection  
YOLOv11  
image super-resolution  
small object detection  
cultural heritage preservation

## ABSTRACT

Wood-carved artifacts often develop tiny insect holes during long-term preservation. Traditional manual inspection is a response to such defects, but it is inefficient and highly dependent on human judgment. A lightweight vision-based framework for automatic insect hole detection in complex wood-grain scenes is discussed in this paper. The proposed system is developed from an enhanced YOLOv11 model and combines image super-resolution with feature refinement to improve the recognition of extremely small targets. To reconstruct higher-resolution inputs for clearer representation of micro-holes, an Efficient Transformer for Single Image Super-Resolution (ESRT) is employed, while an Edge-Gaussian Aggregation (EGA) module is introduced to strengthen shallow edge information and reduce interference in deeper features. In addition, Slicing Aided Hyper Inference (SAHI) is adopted to preserve local target details in high-resolution images. Experimental results show that the proposed framework increases the Mean Average Precision (mAP) from 0.796 to 0.859 and the F1-score from 0.779 to 0.829, while the memory usage of the model parameters only increases slightly from 2.6M to 2.7M. The results demonstrate that the proposed method improves the reliability of micro-hole detection in wood carvings while maintaining a compact model design.

© 2026 The Author(s). Published by The Society of Artificial Life and Robotics.

This is an open access article distributed under the CC BY-NC 4.0 license

(<http://creativecommons.org/licenses/by-nc/4.0/>).

## 1. Introduction

Wood carvings preserved over long periods are vulnerable to insect-related biodeterioration. In conservation practice, wood-boring activity is often indicated by small exit holes and accumulated frass on object surfaces [1][2]. Manual inspection is still widely used for heritage damage assessment; however, this process is labor-intensive and strongly dependent on expert experience and judgment [3][4]. In addition, reliable detection of micro-holes remains challenging because these targets are extremely small and can be easily obscured by wood-grain texture, occlusion, and background interference [5]. To address these issues, this study adopts YOLOv11 as the baseline detector for automatic micro-hole detection in wood carvings [6]. An Efficient Transformer for Single Image Super-Resolution (ESRT) module is introduced before detection to reconstruct finer local details from low-resolution inputs [7]. In addition, an Edge-Gaussian Aggregation (EGA) module is incorporated to enhance boundary-related features and suppress background responses during feature extraction [8].

## 2. Methodology

### 2.1. Image Acquisition and Data Pre-processing

High-resolution images were acquired using an Arducam IMX519 camera module as the primary imaging device. According to the manufacturer's specifications, the sensor provides an active resolution of  $4656 \times 3496$  pixels with a pixel size of  $1.22 \mu\text{m}$  [9]. Although this imaging configuration is suitable for documenting wood-carving surfaces, the micro-holes still occupy only a very limited number of pixels because image acquisition must be performed at a practical working distance from the artifacts. Consequently, the target regions are visually weak and can be easily confused with surrounding wood-grain textures. To improve the visibility of these minute defects, the proposed framework incorporates ESRT for super-resolution reconstruction and EGA for feature refinement [7][8].

Before training and inference, the YOLO pipeline resizes the input images according to the specified image size (*imgsz*), which can further reduce the effective pixel area of very small targets [6]. Following the COCO evaluation protocol, objects with an area smaller than  $32 \times 32$  pixels are generally categorized as small objects [11].

As shown in Fig. 1, most annotated holes in the present dataset are concentrated within this scale range, indicating that preservation of local spatial details is essential for reliable detection. Therefore, a slicing-based preprocessing strategy was adopted using Slicing Aided Hyper Inference (SAHI), which partitions each high-resolution image into overlapping patches and then aggregates the predictions across slices [10]. Because slicing increases the relative pixel proportion of tiny targets within each patch, it is particularly suitable for improving the detectability of small objects represented by only a limited number of pixels [10].

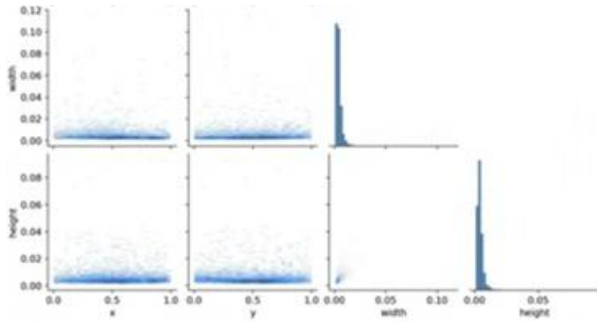


Figure 1 Distribution of Annotation Sizes

## 2.2. Image Super-Resolution

Because insect cavities occupy only a very limited number of pixels in the captured images, their visual characteristics are often weak and lack distinctive local details. When such images are directly resized to the standard detector input, the effective target area is further reduced, which increases the difficulty of reliable recognition. To alleviate this problem, this study introduces a single-image super-resolution strategy before the detection stage. Previous research has shown that super-resolution can improve object detection performance in imagery containing small targets, particularly when fine structural details are difficult to preserve at the original resolution [12].

Among available super-resolution models, the Efficient Transformer for Single Image Super-Resolution (ESRT) was adopted in this work. ESRT is a lightweight hybrid architecture that combines a Lightweight CNN Backbone with a Lightweight Transformer Backbone [7]. In this design, the convolutional component is used to extract low-cost local features, whereas the transformer component further refines image reconstruction through Efficient Transformers. In particular, the Efficient Multi-Head Attention (EMHA) mechanism reduces computational complexity and GPU memory usage while maintaining effective reconstruction capability [7]. These properties make ESRT suitable for integration with a downstream object detection framework.

In the proposed pipeline, each original image is first enhanced by ESRT before being sent to the detection

model. The corresponding ground-truth bounding boxes are then scaled accordingly to match the reconstructed images. To ensure a fair comparison with the baseline setting, the super-resolved images are subsequently divided into  $640 \times 640$  tiles before being used for YOLOv11 training and inference. Through this process, each insect cavity occupies a larger effective pixel area within the detector input, providing clearer local structures for subsequent feature extraction and boundary modeling.

A visual comparison is presented in (Figure 2), using the same region of interest. The figure includes the original image, a directly resized  $640 \times 640$  image, and an image enhanced by  $4\times$  ESRT before conversion to  $640 \times 640$ . The ESRT-enhanced result preserves clearer local textures and more recognizable boundary information. In addition, (Figure 3) shows the corresponding change in bounding-box scale distribution after super-resolution, where the annotated targets shift toward a larger pixel range.

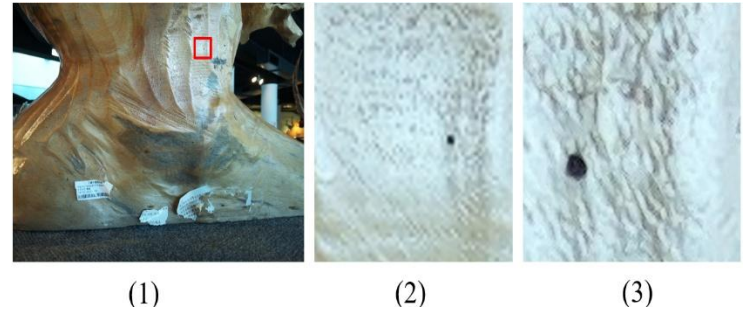


Figure 2 Multi-Scale Visual Comparison of the Same Region: (1) Original Image, (2) Standard  $640\times 640$  Input, and (3) ESRT-Enhanced  $640\times 640$  Input

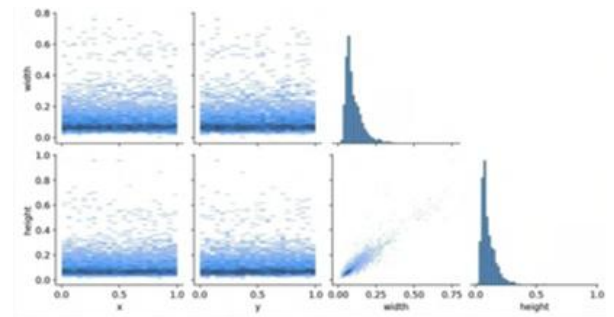


Figure 3 Distribution of Annotation Sizes after ESRT

## 2.3. Edge-Gaussian Aggregation

To improve the discrimination of minute insect cavities from complex wood-grain textures, an Edge-Gaussian Aggregation (EGA) module was incorporated into the YOLOv11-based detection framework. The EGA design combines two complementary mechanisms: orientation-aware edge enhancement and Gaussian-prior-based feature modeling [8]. This combination is intended to strengthen boundary-sensitive responses for visually weak targets while reducing interference from low-confidence

background patterns. In the present study, EGA was integrated into the multi-scale feature extraction process so that cavity-related structures could be represented more clearly under challenging texture conditions.

In our implementation, the edge-enhancement branch was assigned to the higher-resolution shallow stage, where fine boundary information is more directly preserved, whereas the Gaussian modeling branch was applied to deeper stages to suppress ambiguous responses and stabilize semantic features. This arrangement follows the general rationale of EGA, namely that shallow features benefit more from explicit edge cues, while deeper features benefit from noise attenuation and confidence-aware smoothing [8]. By combining these two pathways, the proposed design aims to preserve subtle cavity boundaries in early representations while reducing wood-grain interference in later feature abstraction.

From a computational perspective, the integration of EGA introduces only a limited increase in model complexity. As summarized in (Table 1), the EGA-enhanced YOLO11n model shows only a slight increase in parameters and FLOPs compared with the baseline configuration, while preserving a lightweight deployment profile. The qualitative effect of this design is illustrated in (Figure 4), where the EGA-based model produces more reliable detections in regions heavily affected by wood texture. In particular, several minute cavities that are missed by the baseline detector can still be identified after EGA enhancement, indicating that boundary-aware refinement is beneficial for this task.

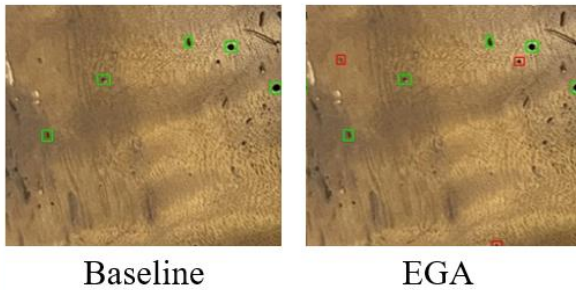


Figure 4 Qualitative Comparison of Detection Results

Table 1 Comparison of Computational Costs

| Model       | Params(M) | FLOPs(G) |
|-------------|-----------|----------|
| YOLO11n     | 2.6       | 6.5      |
| EGA-YOLO11n | 2.7       | 7.5      |

Table 2 Ablation Study Results

| Method                        | Precision | Recall | F1    | mAP@0.50:0.95 |
|-------------------------------|-----------|--------|-------|---------------|
| YOLO11n (w/o tiling)          | 0.435     | 0.077  | 0.130 | 0.089         |
| YOLO11n (tiling)              | 0.798     | 0.762  | 0.779 | 0.796         |
| YOLO11n + tiling + EGA        | 0.812     | 0.784  | 0.798 | 0.816         |
| YOLO11n + tiling + ESRT       | 0.855     | 0.780  | 0.816 | 0.843         |
| YOLO11n + tiling + ESRT + EGA | 0.867     | 0.794  | 0.829 | 0.859         |

### 3. Experiments and Evaluation

For experimental evaluation, a dataset containing 121 high-resolution images was established, with approximately 12,000 annotated insect cavities in total. The dataset was divided into training, validation, and testing subsets using a 7:2:1 ratio. To improve the robustness of model assessment, a 5-fold cross-validation procedure was further applied to the training portion, and the mean results are reported. Because the Ultralytics YOLO framework resizes input images according to the specified *imgsz* setting during training and inference [6], a slicing-based preprocessing strategy was adopted in this study to reduce the loss of effective pixels for extremely small targets. This design is consistent with prior small-object detection studies using Slicing Aided Hyper Inference (SAHI), where slicing has been shown to improve detection performance for targets represented by only a limited number of pixels [10].

YOLO11n was used as the baseline detector. All models were trained for 300 epochs with an input size of 640, while the remaining hyperparameters were kept at their default settings [6]. The experiments were conducted on a workstation equipped with an NVIDIA RTX 3070 GPU. To enhance the visual representation of minute cavities before detection, ESRT was introduced as a super-resolution preprocessing module [7]. In addition, the YOLO11n backbone was augmented with the EGA module to improve boundary-related feature representation and reduce interference from complex background textures [8]. In the implemented design, the edge-enhancement branch was assigned to the shallow high-resolution stage, whereas Gaussian modeling was applied to deeper stages for feature smoothing and stabilization [8].

Model evaluation was performed using the official Ultralytics validation framework, and the reported metrics include Precision, Recall, F1-score, and mAP@0.5:0.95 [6]. The quantitative results are summarized in (Table 2). Compared with the tiled YOLO11n baseline, the addition of ESRT is associated with improved Precision and mAP@0.5:0.95, suggesting that super-resolution preprocessing helps preserve more discriminative local details for tiny cavities. Meanwhile, the integration of EGA further reduces background false positives and enhances the Recall rate with negligible computational overhead.

#### 4. Conclusion

An enhanced YOLO11n framework for insect-cavity detection in wood-carving artifacts was developed in this study. By integrating slicing-based preprocessing, ESRT-based super-resolution, and EGA-based feature refinement, the proposed method was designed to improve the detectability of extremely small cavities in visually complex scenes. Experimental evaluation on 121 high-resolution images containing approximately 12,000 annotated cavities showed improved detection performance over the baseline YOLO11n model. The findings indicate that the combination of local-detail preservation and boundary-aware representation is beneficial for identifying minute defects under complex wood textures. These results support the feasibility of applying lightweight deep-learning models to automated inspection tasks in cultural heritage preservation.

#### References

1. Canadian Conservation Institute, Preventing Infestations: Control Strategies and Detection Methods.
2. English Heritage, Integrated Pest Management.
3. Mishra et al., Artificial intelligence-assisted visual inspection for cultural heritage: State-of-the-art review, *Journal of Cultural Heritage*, 2024.
4. Lee et al., Automatic Surface Damage Classification Developed Based on Deep Learning for Wooden Architectural Heritage, *ISPRS Annals*, 2023.
5. Nikouei et al., Small object detection: A comprehensive survey on challenges, techniques and real-world applications, 2025.
6. Ultralytics YOLO11 documentation
7. Lu et al., Transformer for Single Image Super-Resolution, *CVPR Workshops*, 2022.
8. Lu et al., LEGNet: A Lightweight Edge-Gaussian Network for Low-Quality Remote Sensing Image Object Detection, *ICCV Workshops*, 2025.
9. 16MP-IMX519-Arducam Wiki
10. Akyon, F. C., Altinuc, S. O., & Temizel, A. (2022). Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection. In 2022 IEEE International Conference on Image Processing (ICIP) (pp. 966–970).
11. Lin, T.-Y., Maire, M., Belongie, S., et al. Microsoft COCO: Common Objects in Context. In *ECCV*, 2014.
12. Shermeyer, J., & Van Etten, A. (2019, June). The Effects of Super-Resolution on Object Detection Performance in Satellite Imagery. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 1432–1441.

Dr. Chung-Wen Hung

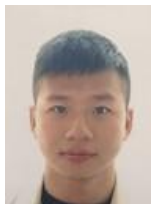


He received the Ph.D. degree in Electrical Engineering from National Taiwan University in 2006. Currently he is a Professor in National Yunlin University of Science & Technology. His research interests include the IoT, IIoT, and AI application.

---

#### Authors Introduction

Mr. Jun-Lin Wu



He received the B.S. degree in Electrical Engineering from Chung Hua University, Taiwan. He is currently pursuing the M.S. degree at the same university. His research interests include computer vision, deep learning, and cultural heritage preservation.