YOLOV11 Wormhole Detection System based on ESRT and EGA Enhancements

Jun-Lin Wu, Chung-Wen Hung
National Yunlin University of Science and Technology, Taiwan
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C

Email: wenhung@yuntech.edu.tw, M11312084@yuntech.edu.tw

Abstract

Wood carving artifacts are prone to insect erosion during long-term preservation, resulting in tiny holes that cause structural damage. Traditional manual inspection methods are not only time-consuming but also susceptible to subjective judgment. This paper proposes an automated borer hole detection system based on YOLOv11 to address the challenge of detecting extremely small targets embedded in complex wood grain patterns. The proposed method integrates the Efficient Transformer for Single Image Super-Resolution (ESRT) and Edge-Gaussian Aggregation (EGA) to enhance detection performance for small objects. First, ESRT is utilized for super-resolution image reconstruction to resolve the issue of insufficient resolution for minute objects such as borer holes. Second, an EGA module is integrated into the model to enhance edge features in shallow layers and suppress noise in deep layers, thereby mitigating issues related to weak boundaries and background interference. Additionally, Slicing Aided Hyper Inference (SAHI) is incorporated to minimize pixel dilution effects on small objects during inference. The overall pipeline resulted in a marginal increase in parameters from 2.6M to 2.7M, while significantly improving the Mean Average Precision (mAP) from 0.796 to 0.859 and the F1-score from 0.779 to 0.829. These results demonstrate a marked improvement in the detectability and accuracy of tiny holes in complex wood grain scenes while maintaining a lightweight architecture.

Keywords: Object detection, YOLOv11, Super-resolution, Small object detection, Cultural heritage preservation

1. Introduction

Wood carving artifacts frequently suffer from biological deterioration during long-term storage, particularly minute holes caused by insect boring, which weaken structural and surface integrity, thereby affecting preservation decisions and subsequent restoration. Recent literature has introduced deep learning into museum pest monitoring and preventive conservation scenarios to assist detection, reducing manual subjectivity and facilitating large-scale deployment [1]. However, borer holes on wood surfaces are small objects with limited discriminative features and are often obscured by wood dust. In single-stage detection frameworks, the resolution degradation of small objects, combined with background interference and occlusion, simultaneously reduces Precision and Recall, representing a common bottleneck in current object detection [2]. The YOLO series is widely utilized in cultural heritage and industrial fields due to its excellent performance; this paper adopts Ultralytics YOLO11 as the baseline detector [3], balancing speed and accuracy while retaining modular expansion flexibility. To mitigate the dilemma of limited resolution and sparse features for small objects, we propose an integrated framework combining super-resolution and edge enhancement. Specifically, we introduce ESRT for image super-resolution during pre-processing to improve the discriminable details of tiny holes and reduce resolution loss during training and inference. Furthermore, the EGA module is integrated into the backbone network to enhance edges and filter noise, thereby improving the accuracy of tiny object detection [4][5].

2. Methodology

2.1. Image Acquisition and Data Pre-processing

This study employs the Arducam IMX519 as the primary imaging device. It features a single-image resolution of 4656 × 3496 with a pixel size of 1.22 μm; detailed specifications, including Field of View (FOV) and focus range, are available in the module documentation [6]. Due to constraints in working distance and lens focal length, borer holes occupy extremely few pixels in the captured images, presenting a scenario analogous to remote sensing (distant, blurred small objects). Consequently, this study integrates ESRT and EGA to enhance identifiable details and boundary representations.

Prior to training and inference, YOLO resizes input images to a specified imgsz, which causes further pixel reduction for small objects after down-sampling. This increases the likelihood of these objects falling into the small object category (defined by COCO metrics as an area less than 32×32 pixels) and being overlooked by detection heads. To mitigate signal loss caused by resizing, we

implemented data slicing. The slicing method adopts the concept of Slicing Aided Hyper Inference (SAHI), dividing high-resolution images into overlapping tiles to increase the effective pixel ratio of small objects. Empirical evidence demonstrates that this approach significantly improves Average Precision (AP) in small object scenarios such as aerial photography and remote sensing [7]. As shown in (Figure 1), the majority of annotations fall within the small object region, illustrating the impact on detectability before and after slicing.

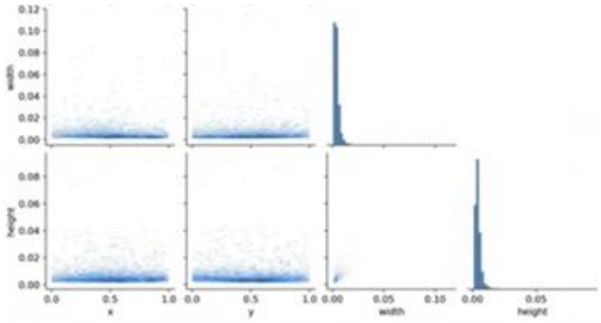

Figure 1 Distribution of annotation sizes.

2.2. *Image Super-Resolution*

Given that borer holes are small objects with sparse discriminative features, direct resizing to the detection input size often results in insufficient effective pixels for instances, leading to unstable detection. Drawing on empirical evidence from remote sensing imaging, employing Single Image Super-Resolution (SISR) to enhance details prior to detection yields considerable gains in mAP and F1-score for small objects. Consequently, this study introduces SISR upstream of YOLOv11 as a front-end enhancement to improve the visualization and pixel coverage of small objects[8].

We selected the Efficient Transformer for Single Image Super-Resolution (ESRT) as the super-resolution backbone. ESRT features a lightweight hybrid architecture. It utilizes a lightweight Convolutional Neural Network (CNN) at the front end for low-cost feature extraction, followed by a stack of multiple Efficient Transformers at the back end. Its core component, Efficient Multi-Head Attention (EMHA), significantly reduces GPU memory usage and computational burden while maintaining competitive reconstruction quality, making it suitable for integration with downstream detection tasks.

For each original image, super-resolution is first performed using ESRT, and bounding box annotations are scaled accordingly. To ensure a fair comparison with the baseline, the output images are ultimately tiled to 640 px for YOLO input. Under the same input resolution, the image coverage of target objects is higher, which facilitates subsequent edge-oriented enhancement by the EGA module.

(Figure 2) presents a three-way comparison of the same region: (1) the original image, with the red box indicating the Region of Interest (ROI); (2) the standard 640×640 input; and (3) the input first upscaled ×4 by ESRT then resized back to 640×640. In comparison, the rightmost image exhibits clearer hole edges and textures, along with higher pixel coverage per instance. (Figure 3) illustrates the distribution of bounding box heights before and after ESRT; the distribution shifts toward larger pixel regions after Super-Resolution (SR).

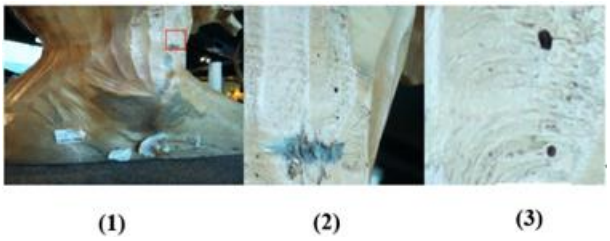

Figure 2 Multi-scale visual comparison of the same region: (1) Original image, (2) Standard 640×640 input, and (3) ESRT-enhanced 640×640 input.

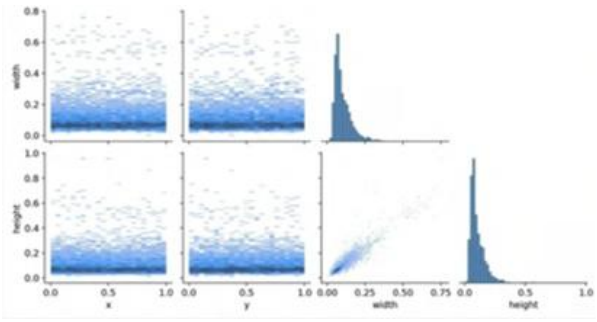

Figure 3 Distribution of annotation sizes after ESRT

2.3. *Edge-Gaussian Aggregation*

To enhance the separability of tiny holes against complex wood grain backgrounds, we introduced the Edge-Gaussian Aggregation (EGA) module into the multi-scale feature extraction layers of the YOLOv11 backbone. EGA employs a staged selection mechanism to enhance representation by adopting either edge extraction or Gaussian modeling at different layers. In this study, we configured the C3 layer (1/8 scale feature map) to utilize the Edge branch, while the C4 and C5 layers (1/16 and 1/32 scales) utilize the Gaussian branch. This configuration aligns with the feature pyramid allocation of the YOLO series, where the high-resolution P3 layer focuses on small objects, and P4/P5 progressively handle larger objects and semantic information. Therefore, enhancing boundaries at P3 directly improves the detectability of small objects.

The EGA module consists of two complementary branches. The first employs the Scharr edge operator to extract direction-aware first-order gradients, offering superior rotational robustness compared to the Sobel operator. The second is a Gaussian branch that applies

smoothing to low-confidence or noisy features, enhancing the contrast between foreground and background. As noted in the (Lightweight Edge-Gaussian Driven Network for Low-Quality Remote Sensing Image Object Detection, LEGNet), edge information is most effective in shallow layers; continuously adding high-frequency edges in deep layers can interfere with semantic representation. Conversely, using Gaussian modeling in deep layers for noise suppression and uncertainty modeling facilitates stable convergence and generalization. Based on these findings, we adopted the P3=Edge and C4/C5=Gaussian configuration, allowing shallow layers to first amplify hole edge features while deep layers focus on stabilizing semantics and suppressing wood grain noise.

Regarding computational cost, EGA enhancement is a lightweight insertion. In our implementation, the increase in parameters and FLOPs is limited, consistent with the lightweight design emphasized in LEGNet. (Table 1) reports the comparison of parameters and FLOPs with and without EGA, showing only a minor overhead favorable for practical deployment. Qualitative effects are illustrated in (Figure 4), comparing (1) YOLOv11n and (2) EGA-YOLOv11n. We marked True Positives (TP) with green boxes and False Positives (FP) with red boxes for clarity. In regions with strong wood grain patterns, the proposed method successfully detects extremely small holes that were missed by the baseline model.

Table 1 Comparison of computational costs.

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

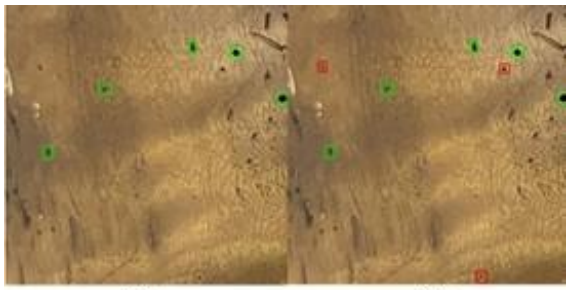

Figure 4 Qualitative comparison of detection results

3. Experiments and Evaluation

The dataset utilized in this study comprises 121 original images containing approximately 12,000 annotated borer holes. The data was partitioned into training, validation, and testing sets in a 7:2:1 ratio. A 5-fold cross-validation strategy was employed on the training set, and the average results are reported. Since YOLO rescales inputs to a specific imgsz prior to training and inference, we adopted data slicing (tiling) to prevent further reduction of effective pixels caused by resizing. This approach is consistent with standard practices for small object detection and has been shown to significantly improve detectability and Average Precision (AP) in scenarios involving large-scale images and minute objects.

The YOLOv11n model served as the baseline detector, trained for 300 epochs with an input size of 640; other hyperparameters remained at their default values. All experiments were conducted on an NVIDIA RTX 3070 GPU. To mitigate the dual challenges of limited resolution and feature sparsity, we introduced ESRT for Single Image Super-Resolution (SISR) prior to detection, thereby enhancing the distinguishable details of tiny holes. Furthermore, the EGA module was integrated into the YOLOv11 backbone for boundary-oriented enhancement and noise suppression. Specifically, the Scharr branch was applied at P3, while Gaussian branches were applied at P4 and P5. This configuration strengthens boundary features in higher-resolution shallow layers while stabilizing representations in lower-resolution deep layers.

We utilized the standard Ultralytics validation pipeline, reporting mAP@0.5:0.95, Precision, Recall, and F1-score to ensure reproducibility. Quantitative results are presented in (Table 2). Overall observations indicate that ESRT significantly boosts mAP and Precision due to increased pixel coverage and more stable detection head matching. Meanwhile, EGA further reduces background false positives and enhances Recall with negligible computational overhead.

4. Conclusion

Targeting the scenario of wood carving borer holes, which presents imaging characteristics analogous to remote sensing, this study proposes a reliable detection pipeline integrating data slicing, ESRT, and EGA with YOLOv11n. Evaluated on a dataset of 121 images containing approximately 12,000 annotations, the proposed combination achieved robust and consistent improvements in mAP and F1-score compared to the

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	0.78	0.816	0.843
YOLO11n + tiling + ESRT + EGA	0.867	0.794	0.829	0.859

baseline YOLOv11n. Specifically, slicing mitigates the pixel loss of small objects caused by proportional resizing; ESRT enhances the pixel coverage of low-quality targets; and EGA improves the separability of weak boundaries. These results demonstrate that the integration of ESRT and EGA offers a practical and scalable solution for automated inspection in the field of cultural heritage preservation.

References

1. Tsujino, J., Ueoka, T., Hasegawa, K., Fujita, Y., Ali, Irhamni and Ayu, Ellis Sekar. "Towards AI-Assisted Preventive Conservation in Libraries: Deep Learning for the Detection of Insect and Mold Damage in Ancient Manuscripts" Preservation, Digital Technology & Culture, 2025.
2. Nikouei, M., Baroutian, B., Nabavi, S., Taraghi, F., Aghaei, A., Sajedi, A., & Ebrahimi Moghaddam, M. (2025). Small object detection: A comprehensive survey on challenges, techniques and real-world applications. Intelligent Systems with Applications, 27, 200561.
3. Khanam, R., & Hussain, M. (2024, October 24). YOLOv11: An Overview of the Key Architectural Enhancements. arXiv.
4. Lu, Z., Li, J., Liu, H., Huang, C., Zhang, L., & Zeng, T. (2022, June). Transformer for Single Image Super-Resolution. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 456–465.
5. Lu, W., Chen, S.-B., Li, H.-D., Shu, Q.-L., Ding, C. H. Q., Tang, J., & Luo, B. (2025, June 2). LEGNet: Lightweight Edge-Gaussian Driven Network for Low-Quality Remote Sensing Image Object Detection (v2). arXiv.
6. 16MP-IMX519-Arducam Wiki
7. 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).
8. 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.

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.

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.