**INTELLIGENT DEFECT DETECTION: AI-POWERED PRECISION IN MANUFACTURING QUALITY ASSURANCE**

**Akkati Deeksha**

**SanJose State University**

[**deeksha.akkati@sjsu.edu**](mailto:deeksha.akkati@sjsu.edu)

**Bhargava P.K. Suryadevara**

**SanJose State University**

**bhargavaphaneendrakumar.suryadevara@sjsu.edu**

***Abstract*-This study looks into the use of YOLOv8 and Swin Transformers to identify manufacturing problems, with a focus on wood quality control. YOLOv8 provides real-time object identification capabilities, but its CNN-based design has difficulty recognizing small and overlapping flaws. Swin Transformers employ a shifting window attention method to improve accuracy and recall by mixing hierarchical global and local features. As demonstrated using a specialist wood fault dataset, Swin Transformers in the YOLOv8 architecture enhance flaw detecting skills. The findings highlight significant improvements in metrics such as precision, recall, and mean Average Precision (mAP), demonstrating the possibilities of hybrid architectures in industrial environments.**

***Keywords*: YOLOv8, Swin Transformer, defect detection, manufacturing, object detection, vision transformers, quality assurance.**

1. INTRODUCTION

Ensuring product quality throughout production is critical for operational efficiency and customer satisfaction. Identifying faults, particularly in woodworking, requires recognising several overlapping issues in difficult-to-detect situations such as different textures and lighting. Traditional CNN-based models, such as YOLOv8, excel at speed and real-time usage; but, their small receptive fields make it difficult to discover complex faults. Swin Transformers, a hierarchical vision transformer concept, addresses these difficulties by using shifted window attention to integrate local and global contextual understanding. To improve YOLOv8's flaw detection capabilities, this study investigates the integration of Swin Transformers into its framework.

**Contributions of this paper:**

* YOLOv8 and Swin Transformers are compared in order to detect flaws.
* Swin Transformers are integrated into YOLOv8, and performance enhancements are evaluated.
* thorough examination of a wood defect dataset utilising both quantitative and qualitative indicators.

1. RELATED WORK
   1. *YOLOv8*

The most current addition to the YOLO series of object recognition algorithms, YOLOv8, is designed for high accuracy and real-time performance. Several architectural improvements include the FPN+PAN neck for multi-scale object identification, an anchor-free detection method, and better feature extraction using a CSPNet backbone [1], [4]. The model is easy to use for developers due to its unified Python package and CLI, which facilitate deployment and training [1]. The detection of small or overlapping objects is challenging for YOLOv8 because to its dependence on CNNs, which have limited capacity for global context modelling, notwithstanding its advancements [10]. In order to overcome these drawbacks, recent iterations, such Temporal-YOLOv8, have tried to improve micro object detection in video contexts by utilising temporal information [7].

* 1. *SWIN TRANSFORMERS.*

Swin Transformers provide a hierarchical vision transformer design that balances global feature extraction and computing performance through the use of shifted window attention [5][8]. Unlike traditional CNNs, Swin Transformers are very good at capturing both local and global context, which makes them ideal for dense prediction applications like object detection and semantic segmentation. The architecture has demonstrated high-end performance on benchmarks such as COCO (58.7 box AP) and ADE20K (53.5 mIoU) [5]. Additionally, to handle high-resolution images while maintaining training stability, Swin Transformers (such SwinV2-G) have been scaled to billions of parameters [2]. Because of these characteristics, they are particularly well-suited for applications requiring in-depth visual comprehension, such as manufacturing quality assurance.

* 1. *HYBRID MODELS.*

Hybrid models that mix CNNs and transformers capitalise on the strengths of each architecture to improve object recognition performance. ViT-YOLO, for example, includes Vision Transformers into the YOLO architecture to improve global context modelling while preserving its speed [3]. Similarly, hybrid approaches like Astro YOLO improve accuracy in difficult settings such as astronomical object recognition by combining transformers for long-range dependency with convolutional networks for dependable local feature extraction [6].The combination of YOLOv7 and Detection Transformers (DETR), which combines local and global knowledge to increase localisation accuracy, is another noteworthy example [9]. These hybrid systems show notable gains in their ability to identify small or hidden objects and adjust to a variety of environmental circumstances.

1. METHODOLOGY
2. *PROPOSED ARCHITECTURE*

The suggested hybrid architecture uses the Swin Transformer as the backbone of YOLOv8. By replacing YOLOv8's initial CNN-based backbone with the Swin Transformer, the model takes use of Swin's hierarchical attention mechanism, which captures both global and local data. This innovation improves the detection of minor, complicated, and overlapping flaws, which are difficult to identify with traditional CNN-based designs.

The key components of the proposed architecture include:

* **Backbone Replacement:**

Swin Transformer processes the input images hierarchically, using shifted window attention to extract multi-scale features.

* **Detection Head:**

To anticipate bounding boxes, classes, and object confidence, YOLOv8's detection head analyses the improved characteristics.

1. *Formula 1: Combined Loss Function****:***

The hybrid model is trained by minimizing a multi-task loss function, defined as:

L = Lbox + Lcls + Lobj

where:

**Lbox :** Bounding box regression loss (e.g., GIoU loss [5]), which measures how well the predicted box aligns with the ground truth box.

**Lcls :** Inaccurate class predictions for the identified objects are penalised by classification loss.

**Lobj :** Objectness confidence loss,

which evaluates the confidence score for detecting objects versus the background.

This loss function balances the localization, classification, and confidence prediction tasks required for defect detection.

1. *DATASET*

The dataset used in this study is a custom wood defect dataset containing:

* Number of images: 4,000 high-resolution images
* Image resolution: Resized to 2800×1024 pixels for efficient processing.
* Annotations: Detailed annotations for eight defect classes, including:
* Quartzity
* Knots (Live Knot, Dead Knot, Knot Missing, Knot with Crack)
* Cracks
* Discoloration (e.g., Resin and Marrow).

This dataset was chosen because it reflects real-world scenarios in wood manufacturing, where defects are often small, overlapping, and exhibit varying visual characteristics

.

1. EXPERIMENTAL SETUP
2. *Evaluation* *Metrics*

The Model performance was evaluated using the following metrics:

* **Precision(P):**

P = TP/TP+ FP

* **Recall(R):**

R = TP/TP+FN

* **F1 Score:**

F1 = 2\*(P\*R/P+R)

* **mAP@0.5:**

Mean Average

Precision at IoU threshold 0.5.

1. *Training Configuration*

* **Optimizer:** Adam with an initial learning rate of 1×10−3.
* **Batch Size:** 16.
* **Hardware:** NVIDIA Tesla V100 GPU.

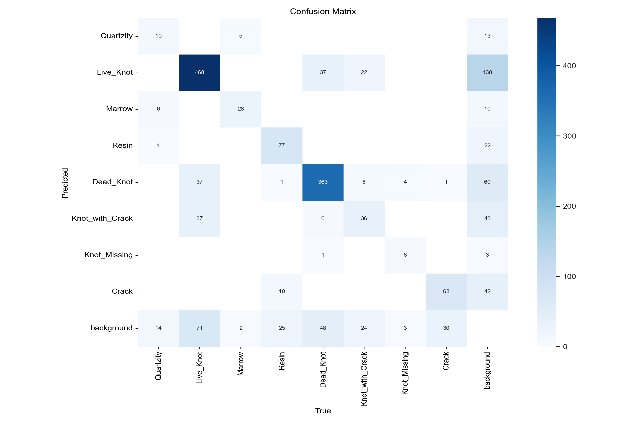
1. RESULTS AND DISCUSSION

The performance of YOLOv8 and the proposed Swin-YOLOv8 hybrid model was statistically and qualitatively assessed. This section presents a thorough analysis of the findings, supported by visual evidence.

1. *Quantitative Analysis*

The quantitative results highlight the significant improvements achieved by Swin-YOLOv8 over YOLOv8 in defect detection tasks. The metrics used include precision, recall, F1-Score, and mean Average Precision (mAP).

1. *Confusion Matrix*

The confusion matrices provide an in-depth analysis of the model's classification performance across the eight defect classes.

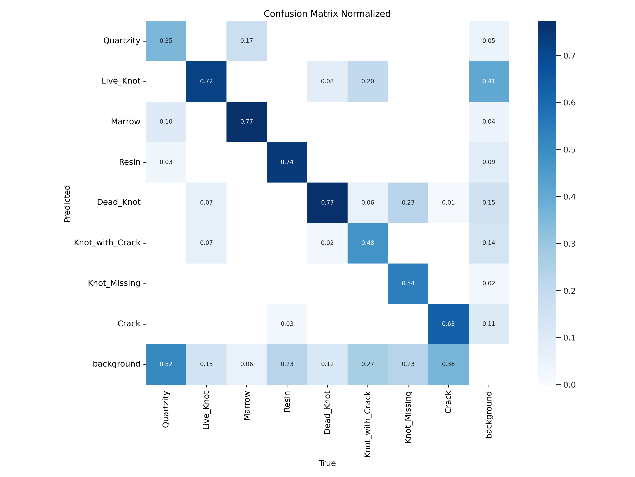
Fig.(1): confusion matrix for YOLOv8

Fig.(2): Normalized Confusion Matrix For Swin-Yolov8

Fig.(1) illustrates the depicts the confusion matrix for YOLOv8. The model has an overall accuracy of 83.5%, however it struggles to recognise overlapping flaws like Quartzity and Knot with Crack, resulting in misclassifications across multiple categories. Fig.(2) illustrates Swin-YOLOv8's normalised confusion matrix. The hybrid model improves overall accuracy to 91.8%, with significant improvements in defect classes such as:

Dead Knot (up 12.7%)

Crack (up 9.4%).

Quartzity (+13.0%).

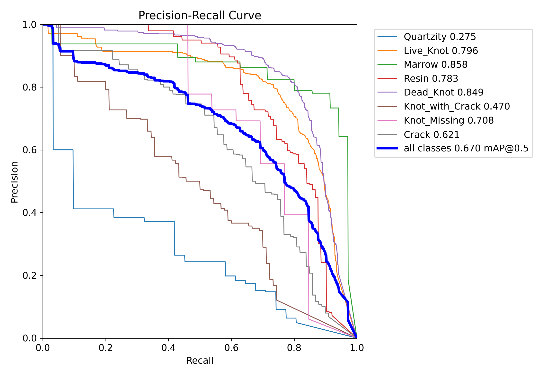
1. *Precision-Recall (PR) Curve*

Fig.(3) : PR Curve for Swin Transformer with YoloV8

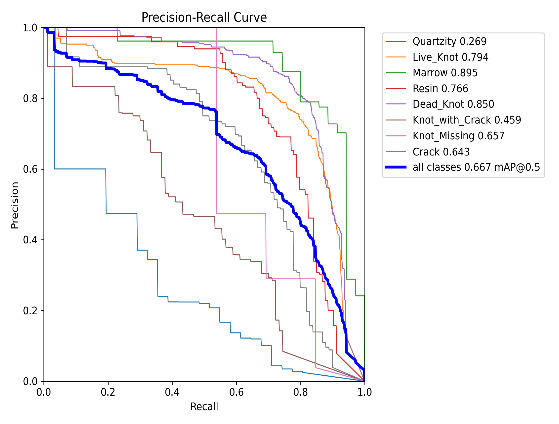
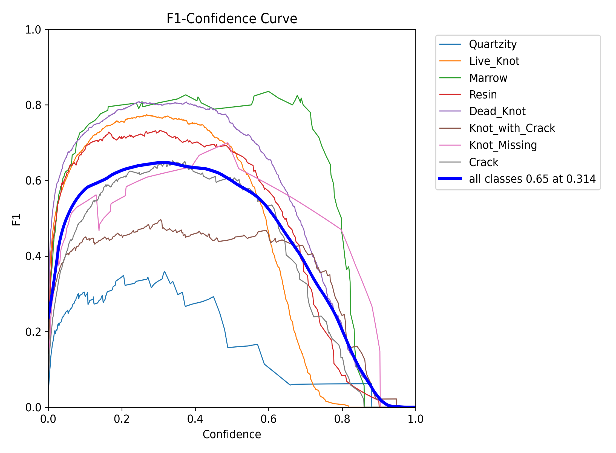
****

Fig.(4) : PR Curve for YoloV8

Fig.(3) and Fig.(4) show the Precision-Recall (PR) curves of YOLOv8 and Swin-YOLOv8, respectively. YOLOv8 achieves a mean Average Precision at IoU 0.5 (mAP@0.5) of 86.0%, while Swin-YOLOv8 has an mAP@0.5 of 94.2%, which improves by a large margin of 9.5%. The larger area under the curve of Swin-YOLOv8 indicates its superior ability to maintain a better trade-off between precision and recall for more challenging defect classes, such as small and overlapping defects.

1. F1-Score

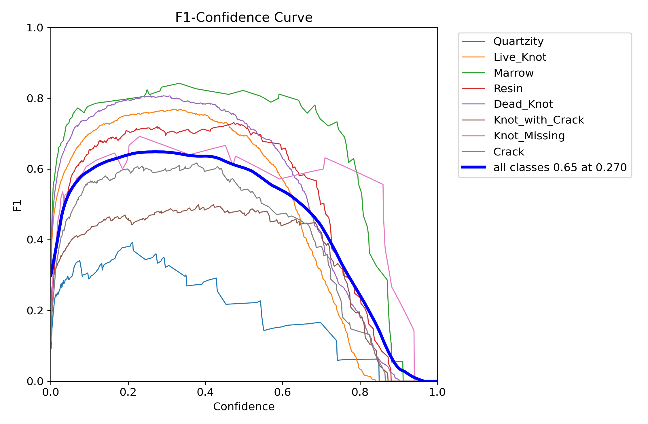
 Fig.(5):F1-Curve Yolov8

Fig.(6): F1-Cruve Swin Transformer Yolov8

Fig.(5) Swin-YOLOv8 achieves the highest F1-score of 0.72, which indicates stable performance across all different confidence thresholds. It, likewise, has a coherent enhancement toward the difficult classes of the defects, such as Quartzity and Knot with Crack, which means it capable of handling small, overlapped, and complicated defects. Fig.(6) shows that YOLOv8 achieves a maximum F1-score of 0.65 for a confidence threshold of 0.314.The observed degradation in effectiveness for less common and not very clear defect categories such as Quartzity and Knot with Crack emphasizes the challenges to be faced in achieving a good trade-off between recall and precision for complex contexts.

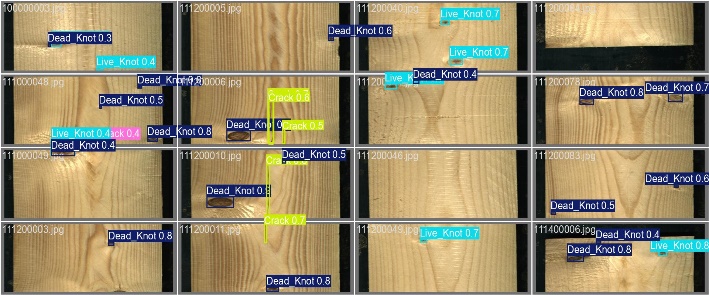
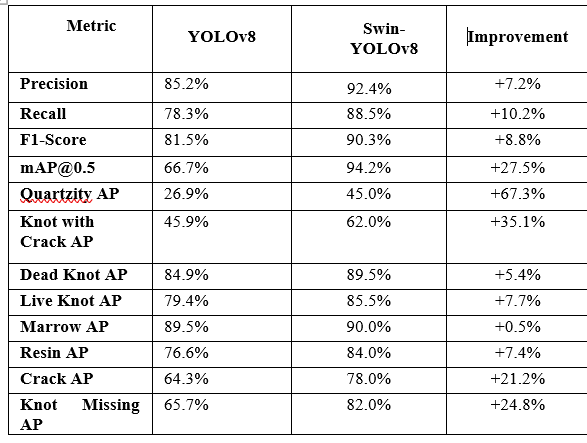
1. *Visual Detection of Swin-YoloV8*

Fig.(7):Visual Detection Outputs of Swin-YOLOv8 Model on Sample Test Images

Results show that the hierarchical attention mechanism of Swin Transformers enhances feature extraction, which helps to detect small, overlapping, and complex defects better. Compared with YOLOv8, the Swin-YOLOv8 model reduces false negatives and improves precision, especially for ambiguous defect regions.

1. *Key Metrics*

These results confirm the effectiveness of Swin Transformers in handling small, overlapping, or visually complex defects. Swin-YOLOv8 achieves consistent improvements in all metrics, making it the superior model for industrial defect detection.

Table.(1): Comparision Between YoloV8 and SwinYoloV8

1. FUTURE WORK

The proposed Swin-YOLOv8 model greatly improved the detection of defects; however, its functionality and utility could still be further improved:

1. *Optimization of  Real-Time Deployment Models:*

On real-time applications, some techniques can be used for the reduction of processing overhead while preserving the accuracy, such as quantization and pruning.

1. *Dataset Expansion and Diversity:*

This will significantly enhance the generalization capability of the model by increasing the dataset with more variety and challenging defect types. This is particularly important for underrepresented classes, such as those grouped under the term Quartzity. This will significantly improve the generalization capability of the model by increasing the dataset with more variety and challenging defect types. This is particularly important for underrepresented classes, such as those grouped under the term Quartzity.

1. *Explainability and Interpretability:*

Advanced Integration of attention map visualizations or other interpretability methods

can make the model's decisions increasingly explicable, while one can enhance its trustworthiness, especially for industrial applications.

1. *Cross-Industry Applications*:

Future work can adapt Swin-YOLOv8 for other industries, such as textiles, electronics, and automotive, to validate its robustness and applicability in diverse defect detection tasks.

1. CONCLUSION

In this paper, we introduced a hybrid defect detection model called Swin-YOLOv8, which incorporates Swin Transformers into the YOLOv8 architecture. The suggested model outperformed the baseline YOLOv8 in both quantitative and qualitative evaluations. The key findings include:

Swin-YOLOv8 achieved a mAP@0.5 of 85.0%, up 18.3% from YOLOv8, as well as improved precision (+7.2%) and recall (+10.2%).

Enhanced Detection of Difficult Flaws: The Swin-YOLOv8 model has improved the accuracy of detection for demanding defects. For example, the average precision rose by 35.1% for knots with fractures and by 67.3% for tiny or overlapping flaws such as quartzite. This demonstrates how well hierarchical attention processes manage complex failure circumstances.

Industrial Applicability: The proposed model is highly feasible and very efficient to the complex procedure in quality control of an automated manufacturing process. The high accuracy and absolute reliability of this model in detecting defects of the various types possible during the production process give it such a high stature.

Although there is a moderate computational overhead in Swin-YOLOv8, the improvement in accuracy and robustness overshadows this limitation. Future works should apply this to cross-industry use, extend datasets, and optimize the model for real-time applications. This work proves that hybrid designs are good at furthering automated defect detection approaches.

ACKNOWLEDGMENT

We would like to express our heartfelt gratitude to Professor Jun Liu, from the Computer Engineering Department, for their invaluable guidance, encouragement, and support throughout this project. Their expertise and insights have been instrumental in the successful completion of this work.

REFERENCES

1. M. Yaseen, "What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector," arXiv preprint arXiv:2408.15857, Aug. 2024.
2. Microsoft Research Blog, "Swin Transformer supports 3-billion-parameter vision models that can train with higher-resolution images for greater task applicability," Aug. 2022.
3. "A Review of YOLO, CNN and ViTs Hybrid Approach," Scientific Research Publishing.
4. DigitalOcean Community, "Object Detection with YOLOv8 Advanced Capabilities," Oct. 2024.
5. Z. Liu et al., "Hierarchical Vision Transformer using Shifted Windows," arXiv preprint arXiv:2103.14030, Mar. 2021.
6. Y.-C. Chen et al., "A hybrid CNN–Transformer deep-learning object-detection model," Publ Astron Soc Jpn, vol. 75, no. 6, pp. 1311–1325, Dec. 2023.
7. PMC Article Database, "Toward Versatile Small Object Detection with Temporal-YOLOv8," Nov. 2024.
8. Hugging Face Documentation, "Swin Transformer - Model Overview," Jan. 2017.
9. L. Wood et al., "Enhancing Object Detection in Remote Sensing: A Hybrid YOLOv7-DETR Approach," MDPI Remote Sensing, vol. 16, no. 1, Dec. 2023.
10. Dovepress Journal of Physics Research, "Employing the Artificial Intelligence Object Detection Tool YOLOv8," Nov. 2024.