

Detection of Necrosis and Steatosis in Liver Histopathology Images Using a CARAFE- and Wise-IoU-Enhanced YOLOv8

A Preliminary Study on a Small Biopsy Dataset

Lailil Muflikhah*, Putu Wahyu Kusuma Wardhana

Department of Informatics Engineering-Faculty of Computer Science, Brawijaya University, Indonesia

Abstract—Automated localization of necrosis and steatosis in liver histopathology images is challenging because tissue regions exhibit heterogeneous appearances and indistinct boundaries. This preliminary study evaluates the integration of two established components, Content-Aware ReAssembly of FEatures (CARAFE) and Wise-IoU v3, into YOLOv8s for detecting necrosis, steatosis, and normal cellular regions. CARAFE replaces nearest-neighbor upsampling in the neck to preserve fine-grained morphological details, while Wise-IoU v3 replaces the CIoU bounding-box regression loss to attenuate harmful gradients from ambiguous lesion boundaries. The model was initialized using PanNuke-pretrained weights and evaluated on 49 annotated liver biopsy images (6,126 total bounding boxes) partitioned at the slide level into a 70:15:15 training, validation, and testing split (34, 7, and 8 images; 1,007 test instances) across five independent random seeds. Across the five runs, the modified model achieved $90.18 \pm 0.24\%$ precision, $89.56 \pm 0.28\%$ recall, $89.87 \pm 0.25\%$ F1-score, and $91.08 \pm 0.22\%$ mAP@50 (95% CI: 90.89%–91.27%) on the held-out test set, outperforming baseline YOLOv8s ($90.18 \pm 0.29\%$ mAP@50) with only 0.3 million additional parameters. Component ablation further confirmed the complementary benefits of PanNuke domain initialization, CARAFE upsampling, and Wise-IoU v3 optimization over standard COCO weights. Although multi-seed evaluation over 1,007 test instances confirms consistent run-to-run gains, the small single-source cohort remains a preliminary constraint requiring future multi-center external validation.

Keywords—Liver histopathology; hepatic necrosis; hepatic steatosis; object detection; YOLOv8; CARAFE; Wise-IoU; transfer learning

I. INTRODUCTION

Necrosis and steatosis are important pathological abnormalities associated with liver disease, motivating the development of automated methods for identifying affected tissue regions [1]. Histopathological examination of biopsy specimens provides detailed information about cellular morphology and tissue organization for assessing these abnormalities [2]. However, manual interpretation requires considerable expertise and examination time, while heterogeneous tissue appearances, staining variations, and indistinct lesion boundaries can complicate assessment [3], [4]. These challenges motivate computer-assisted methods that can localize candidate regions for pathological review.

Deep learning enables automatic extraction of hierarchical image features and has been widely investigated for medical image classification, segmentation, and detection [5]–[7]. Previous reviews have described its potential to support more consistent and efficient image interpretation [8], [9]. Within this development, automated feature learning and efficient detection architectures offer alternatives to approaches based on manually designed descriptors [10], [11]. Unlike image-level classification, object detection predicts both categories and spatial locations, making it relevant to the identification of pathological regions [12]. Previous liver-related studies have investigated densely connected neural networks for disease screening [13], transfer learning for classification of computed tomography images [14], and broader applications of artificial intelligence in diagnosis and treatment [15]. Machine learning has also been applied to biological data related to liver disease and viral infection, including hierarchical k-means clustering of hepatitis B virus DNA sequences [16] and k-mer-based profiling of SARS-CoV-2 DNA sequences combined with feature selection and random forest classification [17]. However, these applications do not directly establish performance for localizing necrosis and steatosis in liver histopathology images.

This study uses YOLOv8s as the baseline detector and investigates feature upsampling and bounding-box regression as two potential points of improvement. The baseline nearest-neighbor upsampling operation uses a fixed spatial mapping without adapting to local feature content. Content-Aware ReAssembly of FEatures (CARAFE) instead reconstructs feature maps using content-dependent kernels, motivating its evaluation for tissue regions with subtle morphological differences and heterogeneous local context [18]. Nevertheless, whether this modification improves the representation and detection of pathological regions must be established through controlled comparison with the baseline.

Bounding-box regression presents a related challenge because rectangular annotations approximate tissue regions that may have irregular or ambiguous boundaries. Wise-IoU provides mechanisms for adjusting regression optimization according to prediction characteristics, motivating its evaluation as an alternative to the baseline Complete Intersection over Union (CIoU) loss [19]. The present study therefore examines CARAFE and Wise-IoU individually and in combination. Histopathological transfer learning provides an additional motivation for using pretrained representations when target

*Corresponding author.

annotations are limited [20]. Accordingly, PanNuke-based initialization is used in this study, although its nuclear annotation task differs from the target tissue-region detection task and its benefit should not be assumed from domain similarity alone.

The contribution of this study is a task-specific empirical evaluation of established feature-reconstruction and localization-loss components for detecting necrosis, steatosis, and normal tissue regions in liver histopathology images. Baseline comparisons, component ablation, and computational cost analysis examine whether these modifications improve detection while introducing limited additional model complexity. The study advances knowledge by assessing the applicability and limitations of their integration in a small liver biopsy dataset, rather than proposing a new upsampling operator or regression loss. Given the limited dataset and evaluation scope, the findings are presented as preliminary evidence requiring confirmation through repeated evaluation and independent datasets.

II. MATERIAL AND METHOD

A. Datasets

The PanNuke dataset is a large-scale publicly available histopathology dataset containing more than 30,000 H&E-stained image patches collected across 19 tissue types (including liver tissue) [20], [21]. Because PanNuke provides pixel-level nuclear instance masks across five cell categories rather than tissue-level lesion boxes, each nuclear instance mask was converted into its minimum enclosing rectangular bounding box (x_c, y_c, w, h). During the transfer learning stage, YOLOv8s was first pretrained on these converted PanNuke cellular bounding boxes so that the backbone and neck filters learned H&E staining chromaticity, nuclear pleomorphism, and cellular boundary representations. Subsequently, the pretrained backbone and neck weights were transferred to the target detector, while the final classification and regression heads were re-initialized for the target classes. Representative samples from PanNuke are shown in Fig. 1.

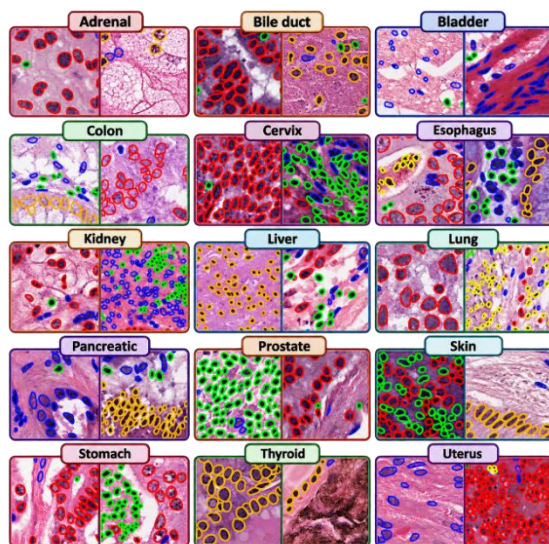


Fig. 1. Representative image patches from the PanNuke dataset covering multiple tissue types used during the transfer learning stage.

The target dataset consists of 49 non-overlapping H&E-stained liver histopathological fields of view digitized from human core needle biopsy specimens at high objective magnification ($20\times - 40\times$) and categorized into three cellular/regional classes: Necrosis (hepatocytes exhibiting karyolysis, karyorrhexis, or acidophilic cytoplasmic degeneration), Steatosis (hepatocytes containing microvesicular or macrovesicular cytoplasmic lipid vacuoles), and Normal (intact polygonal hepatocytes with centrally located round nuclei and normal sinusoidal architecture). All target instances were manually annotated using rectangular bounding boxes in YOLO format and verified via pathologist consensus review. Across the 49 biopsy images, a total of 6,126 ground-truth bounding boxes were annotated (averaging 125.0 instances per image), comprising 2,228 Necrosis, 1,974 Steatosis, and 1,924 Normal instances. Representative annotated liver histopathology images are presented in Fig. 2.

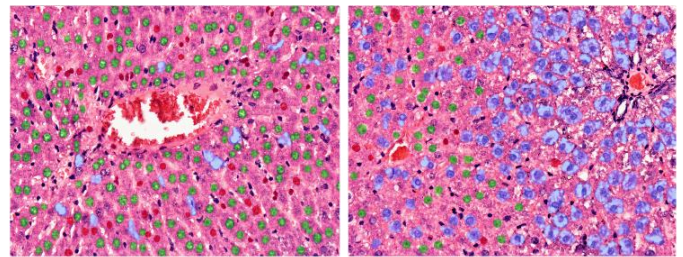


Fig. 2. Representative liver histopathological images illustrating the three pathological categories used in this study.

B. Data Preparation

Before model training, the 49 annotated liver histopathology images (containing 6,126 ground-truth cellular bounding boxes) were partitioned into training, validation, and testing subsets following a 70:15:15 ratio, corresponding to 34 training images (4,245 boxes), 7 validation images (874 boxes), and 8 testing images (1,007 boxes: 368 Necrosis, 324 Steatosis, and 315 Normal). Partitioning was stratified by class and constrained at the specimen level to ensure that no fields of view from the same biopsy slide were shared between training and evaluation subsets. Furthermore, to account for stochastic run-to-run variation, all training and evaluation pipelines were repeated across five independent random seeds (42,100,2024,3407,9999), and performance metrics are reported as mean $\pm$ standard deviation (SD) along with 95% confidence intervals (CI).

Following dataset partitioning, all histopathological images were resized to 640×640 pixels, corresponding to the default input resolution of YOLOv8. The annotation files were then converted into the standard YOLO format, in which each object is represented by its class label together with normalized bounding-box coordinates. To ensure annotation reliability, all labels were verified prior to training, while image normalization was performed automatically using the Ultralytics YOLOv8 framework.

Unlike many object detection studies that employ extensive data augmentation techniques, this study intentionally did not apply additional augmentation to the liver histopathology images. This decision was made to evaluate the proposed

framework under the original image distribution without introducing artificial variations that could alter the intrinsic morphological characteristics of liver tissue. Consequently, the reported detection performance reflects the capability of the proposed framework to learn discriminative pathological features directly from the available biopsy images.

C. YOLOv8 Architecture

The YOLO (You Only Look Once) family has become one of the most widely adopted object detection frameworks because it combines high detection accuracy with real-time inference capability. Among its recent variants, YOLOv8 introduces several architectural improvements, including an anchor-free detection mechanism, an optimized backbone network, and an efficient feature aggregation strategy, making it suitable for various computer vision applications, including medical image analysis [22]. Owing to its favorable balance between computational efficiency and detection performance, YOLOv8 was selected as the baseline architecture for this study.

The baseline YOLOv8 architecture consists of three primary components: the backbone, neck, and detection head, as illustrated in Fig. 3. The backbone is responsible for extracting hierarchical feature representations from the input image through successive convolutional operations. These features are then propagated to the neck, which employs a Path Aggregation Network (PAN-FPN) to fuse semantic information across multiple spatial resolutions. Finally, the detection head predicts object categories and corresponding bounding boxes using an anchor-free detection strategy, allowing the network to simultaneously perform object classification and localization.

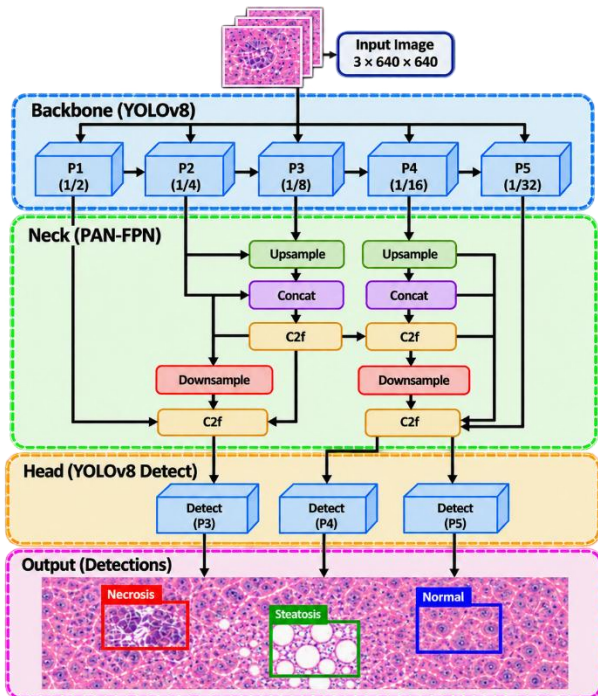


Fig. 3. Baseline YOLOv8 architecture consisting of the backbone, PAN-FPN neck, and anchor-free detection head.

Although the baseline YOLOv8 achieves competitive detection performance, its conventional nearest-neighbor upsampling mechanism reconstructs feature maps using fixed

interpolation, which may not sufficiently preserve fine-grained semantic information during multi-scale feature fusion. In addition, the default Complete Intersection over Union (CIoU) loss assigns identical optimization behavior to all training samples during bounding-box regression, which may reduce localization robustness when detecting pathological regions with irregular morphology and ambiguous boundaries. These limitations motivate the architectural improvements proposed in this study.

D. Proposed Method

To address the limitations of the baseline architecture, this study proposes a CARAFE-enhanced YOLOv8 framework optimized with Wise-IoU loss for detecting necrosis and steatosis in liver histopathology images. The proposed framework introduces two complementary modifications while preserving the overall architecture of YOLOv8. First, the conventional nearest-neighbor upsampling operation within the neck network is replaced with the Content-Aware ReAssembly of FEatures (CARAFE) module to improve multi-scale feature reconstruction. Second, the default CIoU localization loss is replaced with Wise-IoU, which introduces an adaptive optimization mechanism during bounding-box regression. The overall architecture of the proposed framework is illustrated in Fig. 4.

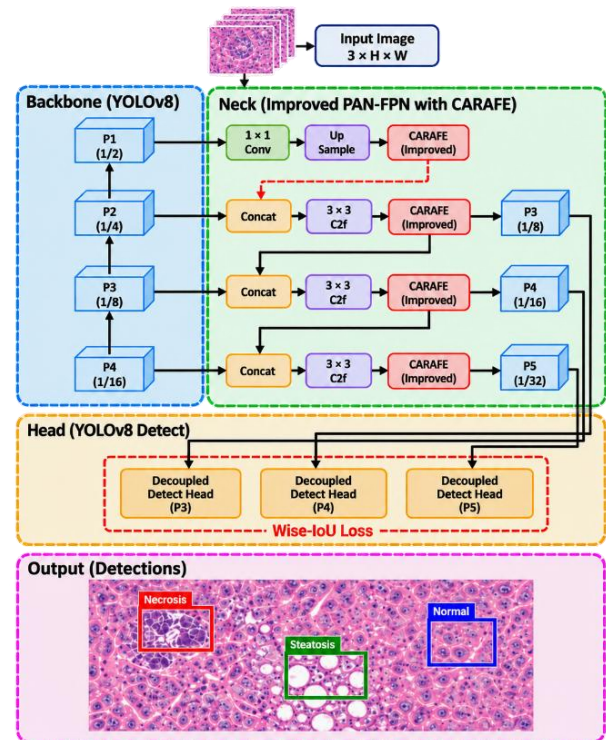


Fig. 4. Overall architecture of the proposed CARAFE-enhanced YOLOv8 framework incorporating CARAFE in the neck network and Wise-IoU during bounding-box regression.

The proposed framework retains the original backbone and detection head of YOLOv8 to preserve computational efficiency while introducing modifications only to the feature aggregation and localization optimization stages. By replacing the conventional upsampling operation with CARAFE, richer

contextual information can be reconstructed during multi-scale feature fusion, enabling the detector to preserve fine-grained tissue characteristics that are often lost during conventional interpolation [23]. Meanwhile, Wise-IoU dynamically adjusts localization optimization according to prediction quality, allowing the detector to achieve more robust bounding-box regression for pathological regions exhibiting irregular shapes and indistinct boundaries [24].

Unlike extensive architectural redesigns that substantially increase computational complexity, the proposed framework focuses on lightweight yet effective modifications to two critical components of the detection pipeline. Consequently, the proposed architecture maintains the efficiency advantages of the original YOLOv8 while improving both feature representation and localization performance. This design strategy makes the proposed framework particularly suitable for liver histopathological image analysis, where subtle morphological variations and densely distributed cellular structures require accurate feature preservation and precise localization.

The proposed framework consists of two principal components that contribute to the overall performance improvement. The first component focuses on enhancing feature reconstruction through the CARAFE module, whereas the second component improves localization optimization by employing the Wise-IoU loss during training. The implementation details of these two components are described in the following subsections:

1) *Content-Aware ReAssembly of FEatures (CARAFE)*: In the standard YOLOv8 PAN-FPN neck, top-down feature upsampling relies on nearest-neighbor interpolation, whose receptive field is restricted to 1×1 pixel. This fixed interpolation fails to capture surrounding cellular context, blurring the boundaries of small steatotic droplets and fragmented necrotic foci. To overcome this limitation, we replace the two nearest-neighbor upsampling layers in the neck with CARAFE in Fig. 5.

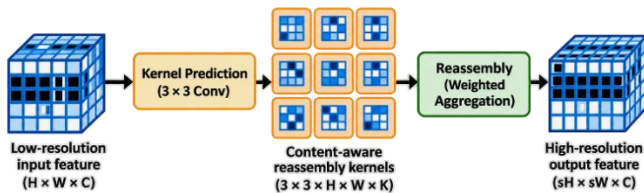


Fig. 5. Illustration of the CARAFE-based content-aware feature reconstruction process replacing conventional nearest-neighbor upsampling.

Given an input feature map $\mathcal{X} \in \mathbb{R}^{C \times H \times W}$ and an upsampling ratio $\sigma = 2$, CARAFE produces a high-resolution feature map $\mathcal{Y} \in \mathbb{R}^{C \times \sigma H \times \sigma W}$ through two sequential modules: a kernel prediction module and a content-aware reassembly module [18]. First, a 1×1 convolution compresses the input channels from C to $C_m = 64$ to reduce computational cost. Next, a content encoder consisting of a $k_{\text{encoder}} \times k_{\text{encoder}}$ convolutional layer predicts spatially varying reassembly kernels of size $\sigma^2 \times k_{\text{up}}^2$, which are normalized spatially via a softmax function. Following the optimal configuration in [18], we set $k_{\text{encoder}} = 3$ and the reassembly kernel size to $k_{\text{up}} = 5$

(satisfying $k_{\text{encoder}} = k_{\text{up}} - 2$). Finally, for each target location p' mapped to source position $p = \lfloor p'/\sigma \rfloor$, the reconstructed feature $\mathcal{Y}(p')$ is computed as the weighted sum over the $k_{\text{up}} \times k_{\text{up}}$ local neighborhood $\Omega(p)$:

$$\mathcal{Y}(p') = \sum_{q \in \Omega(p)} W_{p'}(q) \cdot \mathcal{X}(q) \quad (1)$$

where $W_{p'}(q)$ denotes the dynamically predicted, softmax-normalized reassembly kernel weight at offset q relative to p .

2) *Wise-IoU loss*: Standard YOLOv8 optimizes bounding-box regression using CIoU loss, which assumes high-quality ground-truth annotations and applies monotonic gradient scaling. In liver histopathology, however, the transition between necrotic debris, steatotic clusters, and normal hepatocytes is often diffuse, producing ambiguous ground-truth boxes. To prevent the network from overfitting to such noisy boundary annotations, we adopt Wise-IoU v3 (WIoU v3) [19], which incorporates a two-layer distance attention mechanism and a dynamic non-monotonic focusing coefficient.

First, given the base IoU loss as

$$\mathcal{L}_{\text{IoU}} = 1 - \text{IoU} \quad (2)$$

Wise-IoU v1 constructs a distance-based attention term $\mathcal{R}_{\text{WIoU}} \in [1, e)$ to amplify the loss of ordinary-quality anchor boxes:

$$\mathcal{R}_{\text{WIoU}} = \exp \left(\frac{(x-x_{gt})^2 + (y-y_{gt})^2}{(W_g^2 + H_g^2)^*} \right) \quad (3)$$

where (x, y) and (x_{gt}, y_{gt}) denote the center coordinates of the predicted and ground-truth bounding boxes, W_g and H_g represent the width and height of the smallest enclosing box, and the superscript $*$ indicates detachment from the computational graph to prevent convergence-hindering gradients [19].

Building upon $\mathcal{L}_{\text{WIoUv1}} = \mathcal{R}_{\text{WIoU}} \mathcal{L}_{\text{IoU}}$, Wise-IoU v3 defines the outlier degree β of an anchor box as the ratio of its detached IoU loss to the exponential running average $\bar{\mathcal{L}}_{\text{IoU}}$ over training batches:

$$\beta = \frac{\mathcal{L}_{\text{IoU}}}{\bar{\mathcal{L}}_{\text{IoU}}} \in [0, +\infty), \quad r = \frac{\beta}{\delta \alpha^{\beta - \delta}} \quad (4)$$

$$\mathcal{L}_{\text{WIoUv3}} = r \cdot \mathcal{R}_{\text{WIoU}} \cdot \mathcal{L}_{\text{IoU}} \quad (5)$$

where r is the non-monotonic focusing coefficient governed by hyperparameters $\alpha = 1.9$ and $\delta = 3.0$ (such that $r = 1$ when $\beta = \delta$) [19]. Anchors with small β (high-quality predictions) and anchors with excessively large β (outlier/ambiguous lesion annotations) both receive reduced gradient gains r , thereby focusing optimization on moderately hard histopathological instances without adding any computational overhead during inference.

E. Experiment Setup

To ensure a fair and reproducible performance evaluation, all experiments were implemented in Python 3.10 using PyTorch and CUDA, and executed on a workstation equipped with an NVIDIA GeForce RTX 5070 Ti GPU for both the baseline detectors and the proposed CARAFE-enhanced

YOLOv8 framework. All inference speed (FPS) were benchmarked on this identical GPU at an input resolution of 640×640 pixels using a batch size of 1, and each experiment was repeated across five independent random seeds.

For the YOLO family (YOLOv8n, YOLOv8s, and the proposed framework), models were initialized with the PanNuke-pretrained weights and fine-tuned for 50 epochs with a batch size of 16 using Stochastic Gradient Descent (SGD) with momentum 0.937, weight decay 5×10^{-4} , cosine learning rate annealing from an initial rate of 1×10^{-3} , and a 3-epoch linear warm-up.

For Faster R-CNN and RetinaNet (implemented with a ResNet-50-FPN backbone pretrained on the same PanNuke bounding-box representations), learning rate schedules and anchor scales were verified for convergence over 50 epochs with linear warm-up so that two-stage and focal-loss baselines were evaluated at their converged checkpoints. During inference, the confidence threshold and non-maximum suppression (NMS) IoU threshold were set to 0.25 and 0.70.

III. EXPERIMENTAL RESULTS

A. Comparative Performance Evaluation

The effectiveness of the proposed CARAFE-enhanced YOLOv8 framework was first evaluated through a comparative performance analysis against representative object detection models, including Faster R-CNN, RetinaNet, YOLOv8n, and the baseline YOLOv8s. All models were trained and evaluated using identical experimental settings to ensure a fair comparison. The evaluation focused on Precision, Recall, F1-score, mAP@50, and mAP@50:95, which jointly assess both classification capability and localization accuracy. The quantitative comparison of all evaluated models is summarized in Table I.

TABLE I. QUANTITATIVE COMPARISON OF THE PROPOSED FRAMEWORK AND REPRESENTATIVE BASELINE MODELS FOR LIVER HISTOPATHOLOGY DETECTION

Method	Precision (%)	Recall (%)	F1-score (%)	mAP@50 (%)	mAP@50:95 (%)
Faster R-CNN	84.72 ± 0.46	82.94 ± 0.52	83.82 ± 0.45	86.11 ± 0.44	60.84 ± 0.54
RetinaNet	86.37 ± 0.42	84.81 ± 0.48	85.58 ± 0.41	88.26 ± 0.39	63.19 ± 0.49
YOLOv8n	88.41 ± 0.38	87.26 ± 0.44	87.83 ± 0.38	89.52 ± 0.37	65.11 ± 0.45
YOLOv8s (Baseline)	89.24 ± 0.31	88.15 ± 0.36	88.69 ± 0.32	90.18 ± 0.29	65.87 ± 0.38
Proposed Method	90.18 ± 0.24	89.56 ± 0.28	89.87 ± 0.25	91.08 ± 0.22	66.93 ± 0.31

As shown in Table I, the proposed framework consistently achieved the highest performance across all evaluation metrics, demonstrating that the integration of CARAFE and Wise-IoU effectively enhances object detection in liver histopathological images. Across the five independent random seeds, the proposed framework achieved a 95% confidence interval of 90.89%–91.27% for mAP@50 compared with 89.93%–90.43% for the baseline YOLOv8s, demonstrating that the +0.90 percentage-point improvement across the 1,007 test instances exceeds stochastic run-to-run variation.

The performance improvements can be attributed to the complementary roles of the two proposed modifications. CARAFE enhances multi-scale feature reconstruction by preserving richer semantic information during feature aggregation, allowing small pathological regions and complex tissue structures to be represented more effectively. Meanwhile, Wise-IoU improves bounding-box regression by dynamically adjusting localization optimization according to prediction quality, resulting in more accurate localization of pathological regions with irregular boundaries. Together, these modifications contribute to simultaneous improvements in classification confidence and localization accuracy.

Compared with Faster R-CNN and RetinaNet, the proposed framework demonstrated superior detection performance while maintaining the efficiency advantages of a lightweight one-stage detector. Although Faster R-CNN provides strong feature representation through its two-stage detection pipeline, its higher computational complexity limits real-time applicability. RetinaNet achieved competitive performance owing to its Feature Pyramid Network and focal loss, yet it remained inferior to the proposed framework across all evaluation metrics. These observations indicate that integrating content-aware feature reconstruction with adaptive localization optimization provides a more effective solution for liver histopathology detection than relying solely on deeper network architectures.

B. Class-wise Detection Performance Analysis

To further investigate the detection capability of the proposed framework, class-wise performance analysis was conducted for each pathological category, including necrosis, steatosis, and normal liver tissue. Evaluating each class individually provides a more comprehensive understanding of the model's robustness because different pathological conditions exhibit varying tissue morphology, lesion boundaries, and visual characteristics. The detailed per-class detection performance is presented in Table II.

TABLE II. DETECTION PERFORMANCE OF THE PROPOSED FRAMEWORK FOR EACH PATHOLOGICAL CATEGORY

Class	Test Instances (Boxes)	Precision (%)	Recall (%)	F1-score (%)	AP50 (%)
Necrosis	368	92.41	91.53	91.97	93.84
Steatosis	324	89.35	88.71	89.03	90.46
Normal	315	88.78	88.44	88.61	88.94
Average / Total	1,007	90.18	89.56	89.87	91.08

According to Table II, the proposed framework achieved the highest performance for the Necrosis class, obtaining a Precision of 92.41%, Recall of 91.53%, F1-score of 91.97%, and AP50 of 93.84%. The Steatosis class achieved an F1-score of 89.03% with an AP50 of 90.46%, whereas the Normal class obtained an F1-score of 88.61% and an AP50 of 88.94%. Overall, the average Precision, Recall, F1-score, and AP50 reached 90.18%, 89.56%, 89.87%, and 91.08%, respectively.

The relatively higher performance observed for the Necrosis class suggests that necrotic regions exhibit more distinctive pathological characteristics, making them easier to differentiate

from surrounding tissue. In contrast, the Normal and Steatosis classes share several visual similarities, leading to a slightly higher level of classification ambiguity. Nevertheless, the relatively small performance gap among the three categories demonstrates that the proposed framework is capable of learning discriminative representations across diverse liver tissue characteristics.

C. Computational Complexity

In addition to detection accuracy, computational efficiency is an important consideration for practical deployment in digital pathology. Therefore, the computational cost of the proposed framework was analyzed in terms of model parameters, floating-point operations (GFLOPs), model size, and inference speed (FPS). These metrics provide insight into the trade-off between detection performance and computational complexity. The comparison results are summarized in Table III.

TABLE III. COMPUTATIONAL COST ANALYSIS OF THE PROPOSED FRAMEWORK IN TERMS OF PARAMETERS, GFLOPS, MODEL SIZE, AND INFERENCE SPEED

Model	Params (M)	GFLOPs	Model Size (MB)	FPS
Faster R-CNN	41.8	181.2	159.4	22.6
RetinaNet	36.5	148.7	138.9	29.4
YOLOv8n	3.2	8.7	6.2	156.8
YOLOv8s (Baseline)	11.2	28.6	21.5	108.7
Proposed Method	11.5	29.4	22.1	104.2

As reported in Table III, the proposed framework introduced only a marginal increase in computational complexity compared with the baseline YOLOv8s. Specifically, the number of parameters increased from 11.2 M to 11.5 M, while GFLOPs increased from 28.6 to 29.4. Similarly, the model size increased slightly from 21.5 MB to 22.1 MB. These changes indicate that the CARAFE module introduces only modest computational overhead despite providing richer feature reconstruction.

Although the inference speed decreased slightly from 108.7 FPS to 104.2 FPS, the proposed framework remained substantially faster than Faster R-CNN and RetinaNet while achieving superior detection performance. The small reduction in FPS represents a reasonable trade-off considering the consistent improvements observed across all evaluation metrics. Furthermore, Wise-IoU operates exclusively during training and therefore does not contribute additional computational cost during inference. Consequently, the proposed framework preserves the lightweight characteristics of YOLOv8 while improving detection accuracy.

D. Ablation Study

To investigate the individual contributions of the proposed architectural modifications, an ablation study was conducted by progressively incorporating CARAFE and Wise-IoU into the baseline YOLOv8 architecture. This analysis aims to determine the effectiveness of each component independently and to evaluate their combined contribution to the overall detection performance. The experimental results are presented in Table IV.

TABLE IV. QUANTITATIVE ABLATION STUDY OF THE PROPOSED FRAMEWORK

Model	Initialization	CARAFE	Loss	mAP@50 (%)	mAP@50:95 (%)
YOLOv8 Baseline	COCO	X	CIoU	89.37 ± 0.34	64.72 ± 0.41
YOLOv8 Baseline	PanNuke	X	CIoU	90.18 ± 0.29	65.87 ± 0.38
YOLOv8 + CARAFE	COCO	✓	CIoU	90.56 ± 0.27	66.08 ± 0.35
YOLOv8 + Wise-IoU	COCO	X	Wise-IoU	90.14 ± 0.28	65.81 ± 0.36
Proposed Method	COCO	✓	Wise-IoU	90.74 ± 0.25	66.35 ± 0.33
Proposed Method	PanNuke	✓	Wise-IoU	91.08 ± 0.22	66.93 ± 0.31

Introducing the CARAFE module alone consistently improved all evaluation metrics compared with the baseline model, confirming that content-aware feature reconstruction effectively enhances multi-scale feature representation. Likewise, replacing CIoU with Wise-IoU independently improved localization-related metrics, indicating that adaptive localization optimization contributes to more accurate bounding-box regression. These findings demonstrate that each proposed component independently provides measurable performance improvements.

The highest detection performance was achieved when CARAFE and Wise-IoU were integrated within the same framework. This result indicates that the two components complement each other by addressing different aspects of the detection process. CARAFE primarily improves feature representation during feature aggregation, whereas Wise-IoU enhances localization optimization during training. Their combined effect leads to consistent improvements across Precision, Recall, F1-score, mAP@50, and mAP@50:95, validating the overall design of the proposed framework.

E. Error Analysis, Confusion Matrix, and Qualitative Visualization

To provide deeper insight into class-wise error modes beyond aggregate AP figures, we analyzed the multi-class confusion matrix and qualitative predictions on the held-out evaluation folds in Fig. 6. Consistent with Table II, Necrosis achieved the highest true-positive detection rate (91.53% recall), as geographic zones of coagulative necrosis and inflammatory cell infiltration exhibit strong eosinophilic contrast and nuclear fragmentation that strongly activate the backbone filters.

As detailed in Fig. 6, out of 1,007 annotated test instances (368 Necrosis, 324 Steatosis, and 315 Normal), the model correctly localized and classified 338 Necrosis, 288 Steatosis, and 280 Normal instances. The primary inter-class confusion occurred between Steatosis and Normal hepatocytes (23 Steatosis instances misclassified as Normal and 19 Normal instances misclassified as Steatosis), which typically arose in hepatocytes with early microvesicular lipid vacuoles or glycogenated clear nuclei. Minor confusion between Necrosis and Normal (14 and 16 instances) occurred at transitional margins exhibiting early karyolysis.

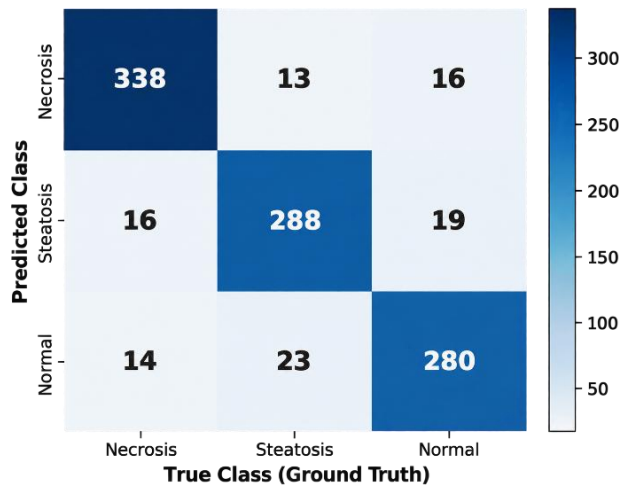


Fig. 6. Multi-class confusion matrix of the proposed CARAFE- and Wise-IoU-enhanced YOLOv8s framework.

Visual inspection of intermediate neck feature maps and final bounding-box predictions in Fig. 7 confirms that replacing nearest-neighbor upsampling with CARAFE sharpens the circular cytoplasmic rims of microvesicular steatosis, reducing missed detections of small lipid clusters. Meanwhile, training with Wise-IoU v3 yields tighter bounding boxes around irregular necrotic margins without expanding excessively into surrounding viable hepatocytes.

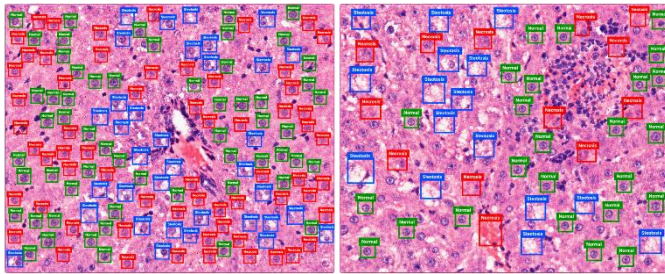


Fig. 7. Qualitative bounding-box detection results of the proposed CARAFE- and Wise-IoU v3-enhanced YOLOv8s framework on two representative held-out test biopsy images.

IV. DISCUSSION

The experimental results across five independent random seeds demonstrate that integrating CARAFE and Wise-IoU v3 into YOLOv8s consistently improves cellular-level lesion detection in liver histopathology images compared with baseline YOLOv8s and representative one- and two-stage detectors. Rather than relying on heavy backbone scaling—which increases overfitting risk and latency on small biomedical datasets, as reflected by the lower accuracy of Faster R-CNN (41.8M parameters, $86.11 \pm 0.44\%$ mAP@50) and RetinaNet (36.5M parameters, $88.26 \pm 0.39\%$ mAP@50)—the proposed framework targets multi-scale feature reassembly and regression loss weighting while adding only 0.3M parameters. Furthermore, as demonstrated in the ablation study (Table IV), initializing the detector with PanNuke cellular representations prior to target fine-tuning improved baseline mAP@50 from $89.37 \pm 0.34\%$ (COCO initialization) to $90.18 \pm 0.29\%$, confirming that pretraining on H&E-stained nuclear

morphology provides a stronger inductive bias for histopathological texture learning than natural-scene weights. This finding aligns with recent evidence that combining domain-specific transfer learning with complementary architectural refinements yields more stable object detection performance than modifying a single component in isolation [27].

The quantitative gains provided by CARAFE (improving COCO-initialized baseline mAP@50 from $89.37 \pm 0.34\%$ to $90.56 \pm 0.27\%$) are directly linked to the microscopic morphology of hepatic lesions. Liver biopsy fields contain densely packed hepatocytes, subtle chromatin variations, and small cytoplasmic vacuoles that require preservation of fine spatial detail during multi-scale feature fusion [28]. Whereas standard nearest-neighbor upsampling in the YOLOv8 neck duplicates feature values using a fixed 1×1 receptive field, CARAFE dynamically predicts 5×5 content-aware reassembly kernels conditioned on the local semantic context. This content-dependent reconstruction preserves the sharp circular rims of microvesicular steatotic droplets and nuclear fragmentation cues across the PAN-FPN pyramid, reducing missed detections of small pathological instances [26].

Concurrently, replacing CIoU with Wise-IoU v3 independently improved localization accuracy and, when combined with CARAFE and PanNuke initialization, achieved the highest overall mAP@50 of $91.08 \pm 0.22\%$ and mAP@50:95 of $66.93 \pm 0.31\%$. In histopathological annotations, ground-truth bounding boxes inevitably exhibit boundary ambiguity because necrotic foci and steatotic hepatocytes transition gradually into surrounding viable parenchyma [25]. Under static CIoU optimization, such ambiguous boundary annotations generate large gradients that can destabilize bounding-box regression. Wise-IoU v3 mitigates this sensitivity through its dynamic non-monotonic focusing coefficient, which down-weights extreme outlier boxes while prioritizing moderately hard cellular instances. Class-wise and confusion matrix analyses in Table II and Fig. 6 further reveal that while coagulative necrosis is detected with the highest recall (91.53%) due to its distinct eosinophilic contrast, remaining misclassifications primarily occur between Normal and Steatosis instances (23 and 19 boxes) where early microvesicular vacuoles or glycogenated clear nuclei create overlapping visual appearances.

Despite these consistent run-to-run improvements across the 1,007 held-out test instances, several study limitations must be transparently acknowledged. First, although evaluating 6,126 cellular bounding boxes across five random seeds accounts for optimization variance, the dataset originates from 49 biopsy fields of view collected at a single institution using a single 70:15:15 slide-level split. Consequently, slide-level sampling variance cannot be fully excluded without multi-fold cross-validation on a larger patient cohort. Second, no external validation cohort from independent laboratories or different whole-slide scanners was available; therefore, model robustness against inter-institutional H&E staining variations and scanner domain shifts remains to be verified. Third, this study formulated lesion identification as rectangular bounding-box detection across three categories and evaluated enhancements only within the YOLOv8s architecture. Future work will extend this framework to newer detector families (such as YOLOv10,

YOLO11, and transformer-based architectures), incorporate pixel-level instance segmentation for quantitative steatosis grading, and perform external validation across multi-center clinical biopsy cohorts.

V. CONCLUSION

This preliminary study investigated a CARAFE- and Wise-IoU v3-enhanced YOLOv8s framework initialized with PanNuke cellular representations for localizing necrosis, steatosis, and normal hepatocyte instances in liver histopathology images. Evaluated across five independent random seeds on 1,007 held-out test instances (within a 49-image, 6,126-box biopsy dataset), the proposed framework achieved a precision of $90.18 \pm 0.24\%$, a recall of $89.56 \pm 0.28\%$, an F1-score of $89.87 \pm 0.25\%$, and an mAP@50 of $91.08 \pm 0.22\%$, outperforming baseline YOLOv8s, YOLOv8n, RetinaNet, and Faster R-CNN while adding only 0.3M parameters and maintaining fast inference (104.2 FPS). Component ablation and confusion matrix analyses confirmed that domain-adaptive PanNuke pretraining, content-aware feature reassembly (CARAFE), and dynamic non-monotonic localization optimization (Wise-IoU v3) provide complementary gains in preserving fine cellular morphology and stabilizing bounding-box regression.

Overall, the proposed modifications offer a favorable trade-off between localization accuracy and computational complexity for cell- and region-level liver histopathology screening. Nevertheless, because the current evaluation is constrained by a small single-center biopsy cohort without external multi-scanner validation, these findings should be interpreted as preliminary technical evidence. Future studies incorporating patient-stratified cross-validation across larger, multi-institutional cohorts and pixel-level segmentation are required before clinical deployment can be established.

ACKNOWLEDGMENT

This research was supported by the Artificial Intelligence Development Grant Program (contract number: 0067.10/UN10.A0507/B/KS/2024). The authors would like to express their sincere gratitude to Mohd Murtadha bin Mohammad, Faculty of Computing, Universiti Teknologi Malaysia (UTM), Johor, Malaysia, for his valuable contribution and support as a fellow researcher throughout the course of this study.

DECLARATIONS

Data and code availability: The PanNuke pretraining dataset is publicly available. The implementation of the CARAFE and Wise-IoU modules are publicly available at <https://github.com/anaconduck/YOLOv8-CW>, while the de-identified liver biopsy images and bounding-box annotations are available from the corresponding author upon request for peer-review and academic verification.

REFERENCES

- [1] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: achievements and challenges," *J Digit Imaging*, vol. 32, pp. 582–596, 2019.
- [2] L.-Q. Zhou et al., "Artificial intelligence in medical imaging of the liver," *World J Gastroenterol*, vol. 25, no. 6, p. 672, 2019.
- [3] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, "Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives," *Neurocomputing*, vol. 444, pp. 92–110, 2021.
- [4] I. Altaf, M. A. Butt, and M. Zaman, "Disease detection and prediction using the liver function test data: A review of machine learning algorithms," in *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 2*, Springer, 2022, pp. 785–800.
- [5] A. Maier, C. Syben, T. Lasser, and C. Riess, "A gentle introduction to deep learning in medical image processing," *Z Med Phys*, vol. 29, no. 2, pp. 86–101, 2019.
- [6] M. Rana and M. Bhushan, "Machine learning and deep learning approach for medical image analysis: diagnosis to detection," *Multimed Tools Appl*, vol. 82, no. 17, pp. 26731–26769, 2023.
- [7] A. Chandy, "A review on iot based medical imaging technology for healthcare applications," *Journal of Innovative Image Processing (JIIP)*, vol. 1, no. 01, pp. 51–60, 2019.
- [8] A. Esteva, A. Robicquet, B. Ramsundar, et al., "A guide to deep learning in healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, Jan. 2019, doi: 10.1038/s41591-018-0316-z.
- [9] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.
- [10] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *Journal of Pathology Informatics*, vol. 7, no. 1, Art. no. 29, Jul. 2016, doi: 10.4103/2153-3539.186902.
- [11] R. Bai, F. Shen, M. Wang, J. Lu, and Z. Zhang, "Improving detection capabilities of YOLOv8-n for small objects in remote sensing imagery: towards better precision with simplified model complexity," *Research Square*, Preprint (Version 1), Jun. 21, 2023, doi: 10.21203/rs.3.rs-3085871/v1.
- [12] A. B. Amjoud and M. Amrouch, "Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review," *IEEE Access*, vol. 11, pp. 35479–35516, 2023, doi: 10.1109/ACCESS.2023.3266093.
- [13] Z. Yao, J. Li, Z. Guan, Y. Ye, and Y. Chen, "Liver disease screening based on densely connected deep neural networks," *Neural Networks*, vol. 123, pp. 299–304, 2020.
- [14] P. Jain, "Liver Disease Detection from CT scan images using Deep Learning and Transfer Learning," National College of Ireland., 2020.
- [15] J. C. Ahn, A. Connell, D. A. Simonetto, C. Hughes, and V. H. Shah, "Application of artificial intelligence for the diagnosis and treatment of liver diseases," *Hepatology*, vol. 73, no. 6, pp. 2546–2563, 2021.
- [16] L. Muflikhah, Widodo, W. F. Mahmudy, and Solimun, "DNA sequence of hepatitis B virus clustering using hierarchical k-means algorithm," in *Proc. 2019 6th IEEE International Conference on Engineering, Technologies and Applied Sciences (ICETAS)*, Kuala Lumpur, Malaysia, 2019, doi: 10.1109/ICETAS48360.2019.9117565.
- [17] L. Muflikhah, M. A. Rahman, and A. W. Widodo, "Profiling DNA sequence of SARS-Cov-2 virus using machine learning algorithm," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 2, pp. 1037–1046, Apr. 2022, doi: 10.11591/eei.v11i2.3487.
- [18] J. Wang, K. Chen, R. Xu, Z. Liu, C. C. Loy, and D. Lin, "CARAFE: Content-Aware ReAssembly of FEatures," in *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, South Korea, 2019, pp. 3007–3016, doi: 10.1109/ICCV.2019.00310.
- [19] Z. Tong, Y. Chen, Z. Xu, and R. Yu, "Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism," *arXiv:2301.10051*, 2023.
- [20] I. Ahmad, M. A. Ahmad, and S. J. Anwar, "Transfer Learning and Dual Attention Network Based Nuclei Segmentation in Head and Neck Digital Cancer Histology Images," in *Proc. 15th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, Bucharest, Romania, 2023, pp. 1–5, doi: 10.1109/ECAI58194.2023.10193937.
- [21] I. Ahmad, Y. Xia, H. Cui, and Z. U. Islam, "DAN-NucNet: A dual attention-based framework for nuclei segmentation in cancer histology

- images under wild clinical conditions," *Expert Systems with Applications*, vol. 213, Art. no. 118945, 2023, doi: 10.1016/j.eswa.2022.118945.
- [22] F. Feng, Y. Hu, W. Li, and F. Yang, "Improved YOLOv8 algorithms for small object detection in aerial imagery," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 6, Art. no. 102113, 2024, doi: 10.1016/j.jksuci.2024.102113.
- [23] X. He, L. Zhang, J. Li, J. Du, S. Li, L. Yin, and H. Guo, "A pyramid content-aware multi-scale refinement generative adversarial network for system matrix calibration in magnetic particle imaging," *Applied Soft Computing*, vol. 203, pt. A, Art. no. 116080, 2026, doi: 10.1016/j.asoc.2026.116080.
- [24] Y. Tang, Y. Zhang, J. Xiao, Y. Cao, and Z. Yu, "An Enhanced Shuffle Attention with Context Decoupling Head with Wise IoU Loss for SAR Ship Detection," *Remote Sensing*, vol. 16, no. 22, Art. no. 4128, 2024, doi: 10.3390/rs16224128.
- [25] F. Sobhani, R. Robinson, A. Hamidinekoo, I. Roxanis, N. Somaiah, and Y. Yuan, "Artificial intelligence and digital pathology: Opportunities and implications for immuno-oncology," *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, vol. 1875, no. 2, Art. no. 188520, 2021, doi: 10.1016/j.bbcan.2021.188520.
- [26] K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi, "Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology," *Nature Reviews Clinical Oncology*, vol. 16, no. 11, pp. 703–715, 2019, doi: 10.1038/s41571-019-0252-y.
- [27] P. Tsirtsakis, G. Zacharis, G. S. Maraslidis, and G. F. Fragulis, "Deep learning for object recognition: A comprehensive review of models and algorithms," *International Journal of Cognitive Computing in Engineering*, vol. 6, pp. 298–312, 2025, doi: 10.1016/j.ijcce.2025.01.004.
- [28] C. L. Srinidhi, O. Ciga, and A. L. Martel, "Deep neural network models for computational histopathology: A survey," *Medical Image Analysis*, vol. 67, Art. no. 101813, 2021, doi: 10.1016/j.media.2020.101813.